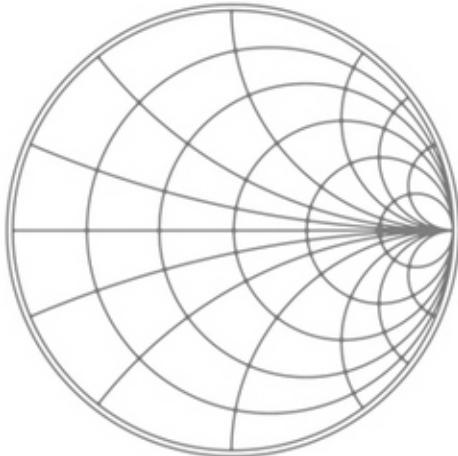


Wiley Series on Information and Communication Technology  
Series Editors, T. Russell Hsing and Vincent K. N. Lau

# RF circuit design

second edition



Richard Chi-Hsi Li

 WILEY



# RF CIRCUIT DESIGN





---

# *RF CIRCUIT DESIGN*

---

SECOND EDITION

Richard Chi Hsi Li



A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 2012 by John Wiley & Sons, Inc. All rights reserved

Published by John Wiley & Sons, Inc., Hoboken, New Jersey

Published simultaneously in Canada

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, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at [www.copyright.com](http://www.copyright.com). Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

**Limit of Liability/Disclaimer of Warranty:** While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at [www.wiley.com](http://www.wiley.com).

***Library of Congress Cataloging-in-Publication Data:***

Li, Richard Chi-Hsi, 1938-

RF circuit design [electronic resource] / Richard Chi-Hsi Li. – Second edition.

1 online resource. – (Information and communication technology series ; 102)

Includes bibliographical references and index.

Description based on print version record and CIP data provided by publisher; resource not viewed.

ISBN 978-1-118-30990-2 (Adobe PDF) – ISBN 978-1-118-30991-9 (ePub) –

ISBN 978-1-118-30993-3 (MobiPocket) – ISBN 978-1-118-20801-4 (cloth) –

ISBN 978-1-118-12849-7 (print) 1. Radio circuits—Design and construction. 2. Electronic circuit design. 3. Radio frequency. I. Title.

TK6560

621.384'12–dc23

2012011617

10 9 8 7 6 5 4 3 2 1

---

# CONTENTS

---

PREFACE TO THE SECOND EDITION	xix
<b>PART 1 DESIGN TECHNOLOGIES AND SKILLS</b>	<b>1</b>
<b>1 DIFFERENCE BETWEEN RF AND DIGITAL CIRCUIT DESIGN</b>	<b>3</b>
1.1 Controversy	3
1.1.1 Impedance Matching	4
1.1.2 Key Parameter	5
1.1.3 Circuit Testing and Main Test Equipment	6
1.2 Difference of RF and Digital Block in a Communication System	6
1.2.1 Impedance	6
1.2.2 Current Drain	7
1.2.3 Location	7
1.3 Conclusions	9
1.4 Notes for High-Speed Digital Circuit Design	9
Further Reading	10
Exercises	11
Answers	11
<b>2 REFLECTION AND SELF-INTERFERENCE</b>	<b>15</b>
2.1 Introduction	15
2.2 Voltage Delivered from a Source to a Load	16
2.2.1 General Expression of Voltage Delivered from a Source to a Load when $l \ll \lambda/4$ so that $T_d \rightarrow 0$	16
2.2.2 Additional Jitter or Distortion in a Digital Circuit Block	20
2.3 Power Delivered from a Source to a Load	23
2.3.1 General Expression of Power Delivered from a Source to a Load when $l \ll \lambda/4$ so that $T_d \rightarrow 0$	23
2.3.2 Power Instability	26
2.3.3 Additional Power Loss	27
2.3.4 Additional Distortion	28
2.3.5 Additional Interference	31
2.4 Impedance Conjugate Matching	33
2.4.1 Maximizing Power Transport	33
2.4.2 Power Transport without Phase Shift	35

2.4.3	Impedance Matching Network	37
2.4.4	Necessity of Impedance Matching	40
2.5	Additional Effect of Impedance Matching	42
2.5.1	Voltage Pumped up by Means of Impedance Matching	42
2.5.2	Power Measurement	49
	Appendices	51
2.A.1	VSWR and Other Reflection and Transmission Coefficients	51
2.A.2	Relationships between Power ( $\text{dB}_m$ ), Voltage (V), and Power (W)	58
	Reference	58
	Further Reading	58
	Exercises	59
	Answers	59
<b>3</b>	<b>IMPEDANCE MATCHING IN THE NARROW-BAND CASE</b>	<b>61</b>
3.1	Introduction	61
3.2	Impedance Matching by Means of Return Loss Adjustment	63
3.2.1	Return Loss Circles on the Smith Chart	63
3.2.2	Relationship between Return Loss and Impedance Matching	66
3.2.3	Implementation of an Impedance Matching Network	67
3.3	Impedance Matching Network Built by One Part	68
3.3.1	One Part Inserted into Impedance Matching Network in Series	68
3.3.2	One Part Inserted into the Impedance Matching Network in Parallel	70
3.4	Impedance Matching Network Built by Two Parts	74
3.4.1	Regions in a Smith Chart	74
3.4.2	Values of Parts	75
3.4.3	Selection of Topology	81
3.5	Impedance Matching Network Built By Three Parts	84
3.5.1	“ $\Pi$ ” Type and “T” Type Topologies	84
3.5.2	Recommended Topology	84
3.6	Impedance Matching When $Z_S$ Or $Z_L$ Is Not $50 \Omega$	85
3.7	Parts In An Impedance Matching Network	93
	Appendices	94
3.A.1	Fundamentals of the Smith Chart	94
3.A.2	Formula for Two-Part Impedance Matching Network	99
3.A.3	Topology Limitations of the Two-Part Impedance Matching Network	110
3.A.4	Topology Limitation of Three Parts Impedance Matching Network	114
3.A.5	Conversion between $\Pi$ and T Type Matching Network	122
3.A.6	Possible $\Pi$ and T Impedance Matching Networks	124
	Reference	124
	Further Reading	124

Exercises	125
Answers	127
<b>4 IMPEDANCE MATCHING IN THE WIDEBAND CASE</b>	<b>131</b>
4.1 Appearance of Narrow and Wideband Return Loss on a Smith Chart	131
4.2 Impedance Variation Due to the Insertion of One Part Per Arm or Per Branch	136
4.2.1 An Inductor Inserted into Impedance Matching Network in Series	137
4.2.2 A Capacitor Inserted into Impedance Matching Network in Series	139
4.2.3 An Inductor Inserted into Impedance Matching Network in Parallel	141
4.2.4 A Capacitor Inserted into Impedance Matching Network in Parallel	143
4.3 Impedance Variation Due to the Insertion of Two Parts Per Arm or Per Branch	145
4.3.1 Two Parts Connected in Series to Form One Arm	146
4.3.2 Two Parts Are Connected in Parallel to Form One Branch	148
4.4 Partial Impedance Matching for an IQ (in Phase Quadrature) Modulator in a UWB (Ultra Wide Band) System	151
4.4.1 Gilbert Cell	151
4.4.2 Impedances of the Gilbert Cell	153
4.4.3 Impedance Matching for LO, RF, and IF Ports Ignoring the Bandwidth	155
4.4.4 Wide Bandwidth Required in a UWB (Ultra Wide Band) System	159
4.4.5 Basic Idea to Expand the Bandwidth	160
4.4.6 Example 1: Impedance Matching in IQ Modulator Design for Group 1 in a UWB System	161
4.4.7 Example 2: Impedance Matching in IQ Modulator Design for Group 3 + Group 6 in a UWB System	172
4.5 Discussion of Passive Wideband Impedance Matching Network	174
4.5.1 Impedance Matching for the Gate of a MOSFET Device	175
4.5.2 Impedance Matching for the Drain of a MOSFET Device	177
Further Reading	179
Exercises	179
Answers	180
<b>5 IMPEDANCE AND GAIN OF A RAW DEVICE</b>	<b>181</b>
5.1 Introduction	181
5.2 Miller Effect	183
5.3 Small-Signal Model of a Bipolar Transistor	187
5.4 Bipolar Transistor with CE (Common Emitter) Configuration	190
5.4.1 Open-Circuit Voltage Gain $A_{v,CE}$ of a CE Device	190

5.4.2	Short-Circuit Current Gain $\beta_{CE}$ and Frequency Response of a CE Device	194
5.4.3	Primary Input and Output Impedance of a CE (common emitter) device	196
5.4.4	Miller's Effect in a Bipolar Transistor with CE Configuration	197
5.4.5	Emitter Degeneration	200
5.5	Bipolar Transistor with CB (Common Base) Configuration	204
5.5.1	Open-Circuit Voltage Gain $A_{v,CB}$ of a CB Device	204
5.5.2	Short-Circuit Current Gain $\beta_{CG}$ and Frequency Response of a CB Device	206
5.5.3	Input and Output Impedance of a CB Device	208
5.6	Bipolar Transistor with CC (Common Collector) Configuration	214
5.6.1	Open-Circuit Voltage Gain $A_{v,CC}$ of a CC Device	214
5.6.2	Short-Circuit Current Gain $\beta_{CC}$ and Frequency Response of the Bipolar Transistor with CC Configuration	217
5.6.3	Input and Output Impedance of a CC Device	218
5.7	Small-Signal Model of a MOSFET	221
5.8	Similarity Between a Bipolar Transistor and a MOSFET	225
5.8.1	Simplified Model of CS Device	225
5.8.2	Simplified Model of CG Device	228
5.8.3	Simplified Model of CD Device	230
5.9	MOSFET with CS (Common Source) Configuration	235
5.9.1	Open-Circuit Voltage Gain $A_{v,CS}$ of a CS Device	235
5.9.2	Short-Circuit Current Gain $\beta_{CS}$ and Frequency Response of a CS Device	237
5.9.3	Input and Output Impedance of a CS Device	239
5.9.4	Source Degeneration	240
5.10	MOSFET with CG (Common Gate) Configuration	244
5.10.1	Open-Circuit Voltage Gain of a CG Device	244
5.10.2	Short-Circuit Current Gain and Frequency Response of a CG Device	245
5.10.3	Input and Output Impedance of a CG Device	247
5.11	MOSFET with CD (Common Drain) Configuration	249
5.11.1	Open-Circuit Voltage Gain $A_{v,CD}$ of a CD Device	250
5.11.2	Short-Circuit Current Gain $\beta_{CD}$ and Frequency Response of a CD Device	250
5.11.3	Input and Output Impedance of a CD Device	251
5.12	Comparison of Transistor Configuration of Single-stage Amplifiers with Different Configurations	252
	Further Reading	256
	Exercises	256
	Answers	256
<b>6</b>	<b>IMPEDANCE MEASUREMENT</b>	<b>259</b>
6.1	Introduction	259
6.2	Scalar and Vector Voltage Measurement	260

6.2.1	Voltage Measurement by Oscilloscope	260
6.2.2	Voltage Measurement by Vector Voltmeter	262
6.3	Direct Impedance Measurement by a Network Analyzer	263
6.3.1	Direction of Impedance Measurement	263
6.3.2	Advantage of Measuring $S$ Parameters	265
6.3.3	Theoretical Background of Impedance Measurement by $S$ Parameters	266
6.3.4	$S$ Parameter Measurement by Vector Voltmeter	268
6.3.5	Calibration of the Network Analyzer	270
6.4	Alternative Impedance Measurement by Network Analyzer	272
6.4.1	Accuracy of the Smith Chart	272
6.4.2	Low- and High-Impedance Measurement	275
6.5	Impedance Measurement Using a Circulator	276
Appendices		277
6.A.1	Relationship Between the Impedance in Series and in Parallel	277
Further Reading		278
Exercises		278
Answers		279

## 7 GROUNDING 281

7.1	Implication of Grounding	281
7.2	Possible Grounding Problems Hidden in a Schematic	283
7.3	Imperfect or Inappropriate Grounding Examples	284
7.3.1	Inappropriate Selection of Bypass Capacitor	284
7.3.2	Imperfect Grounding	286
7.3.3	Improper Connection	288
7.4	'Zero' Capacitor	290
7.4.1	What is a Zero Capacitor	290
7.4.2	Selection of a Zero Capacitor	290
7.4.3	Bandwidth of a Zero Capacitor	293
7.4.4	Combined Effect of Multi-Zero Capacitors	295
7.4.5	Chip Inductor is a Good Assistant	296
7.4.6	Zero Capacitor in RFIC Design	298
7.5	Quarter Wavelength of Microstrip Line	300
7.5.1	A Runner is a Part in RF Circuitry	300
7.5.2	Why Quarter Wavelength is so Important	304
7.5.3	Magic Open-Circuited Quarter Wavelength of Microstrip Line	305
7.5.4	Testing for Width of Microstrip Line with Specific Characteristic Impedance	307
7.5.5	Testing for Quarter Wavelength	307
Appendices		309
7.A.1	Characterizing of Chip Capacitor and Chip Inductor by Means of $S_{21}$ Testing	309
7.A.2	Characterizing of Chip Resistor by Means of $S_{11}$ of $S_{22}$ Testing	319

Reference	321
Further Reading	322
Exercises	322
Answers	323
<b>8 EQUIPOTENTIALITY AND CURRENT COUPLING ON THE GROUND SURFACE</b>	<b>325</b>
8.1    Equipotentiality on the Ground Surface	325
8.1.1    Equipotentiality on the Grounded Surface of an RF Cable	325
8.1.2    Equipotentiality on the Grounded Surface of a PCB	326
8.1.3    Possible Problems of a Large Test PCB	327
8.1.4    Coercing Grounding	328
8.1.5    Testing for Equipotentiality	333
8.2    Forward and Return Current Coupling	335
8.2.1    Indifferent Assumption and Great Ignore	335
8.2.2    Reduction of Current Coupling on a PCB	336
8.2.3    Reduction of Current Coupling in an IC Die	338
8.2.4    Reduction of Current Coupling between Multiple RF Blocks	340
8.2.5    A Plausible System Assembly	341
8.3    PCB or IC Chip with Multimetallic Layers	344
Further Reading	346
Exercises	346
Answers	347
<b>9 LAYOUT</b>	<b>349</b>
9.1    Difference in Layout between an Individual Block and a System	349
9.2    Primary Considerations of a PCB	350
9.2.1    Types of PCBs	350
9.2.2    Main Electromagnetic Parameters	351
9.2.3    Size	351
9.2.4    Number of Metallic Layers	352
9.3    Layout of a PCB for Testing	352
9.4    VIA Modeling	355
9.4.1    Single Via	355
9.4.2    Multivias	359
9.5    Runner	360
9.5.1    When a Runner is Connected with the Load in Series	360
9.5.2    When a Runner is Connected to the Load in Parallel	363
9.5.3    Style of Runner	363
9.6    Parts	369
9.6.1    Device	369
9.6.2    Inductor	369
9.6.3    Resistor	370
9.6.4    Capacitor	370
9.7    Free Space	371

References	373
Further Reading	373
Exercises	373
Answers	374
<b>10 MANUFACTURABILITY OF PRODUCT DESIGN</b>	<b>377</b>
10.1 Introduction	377
10.2 Implication of $6\sigma$ Design	379
10.2.1 $6\sigma$ and Yield Rate	379
10.2.2 $6\sigma$ Design for a Circuit Block	382
10.2.3 $6\sigma$ Design for a Circuit System	383
10.3 Approaching $6\sigma$ Design	383
10.3.1 By Changing of Parts' $\sigma$ Value	383
10.3.2 By Replacing Single Part with Multiple Parts	385
10.4 Monte Carlo Analysis	386
10.4.1 A Band-Pass Filter	386
10.4.2 Simulation with Monte Carlo Analysis	387
10.4.3 Sensitivity of Parts on the Parameter of Performance	392
Appendices	392
10.A.1 Fundamentals of Random Process	392
10.A.2 Index $C_p$ and $C_{pk}$ Applied in $6\sigma$ Design	398
10.A.3 Table of the Normal Distribution	398
Further Reading	398
Exercises	399
Answers	399
<b>11 RFIC (RADIO FREQUENCY INTEGRATED CIRCUIT)</b>	<b>401</b>
11.1 Interference and Isolation	401
11.1.1 Existence of Interference in Circuitry	401
11.1.2 Definition and Measurement of Isolation	402
11.1.3 Main Path of Interference in a RF Module	403
11.1.4 Main Path of Interference in an IC Die	403
11.2 Shielding for an RF Module by a Metallic Shielding Box	403
11.3 Strong Desirability to Develop RFIC	405
11.4 Interference going along IC Substrate Path	406
11.4.1 Experiment	406
11.4.2 Trench	408
11.4.3 Guard Ring	409
11.5 Solution for Interference Coming from Sky	411
11.6 Common Grounding Rules for RF Module and RFIC Design	412
11.6.1 Grounding of Circuit Branches or Blocks in Parallel	412
11.6.2 DC Power Supply to Circuit Branches or Blocks in Parallel	413
11.7 Bottlenecks in RFIC Design	414
11.7.1 Low- $Q$ Inductor and Possible Solution	414
11.7.2 "Zero" Capacitor	419

11.7.3 Bonding Wire	419
11.7.4 Via	419
11.8 Calculating of Quarter Wavelength	420
Reference	423
Further Reading	423
Exercises	424
Answers	425
<b>PART 2 RF SYSTEM</b>	<b>427</b>
<b>12 MAIN PARAMETERS AND SYSTEM ANALYSIS IN RF CIRCUIT DESIGN</b>	<b>429</b>
12.1 Introduction	429
12.2 Power Gain	431
12.2.1 Basic Concept of Reflection Power Gain	431
12.2.2 Transducer Power Gain	434
12.2.3 Power Gain in a Unilateral Case	437
12.2.4 Power Gain in a Unilateral and Impedance-Matched Case	438
12.2.5 Power Gain and Voltage Gain	439
12.2.6 Cascaded Equations of Power Gain	439
12.3 Noise	441
12.3.1 Significance of Noise Figure	441
12.3.2 Noise Figure in a Noisy Two-Port RF Block	443
12.3.3 Notes on Noise Figure Testing	444
12.3.4 An Experimental Method to Obtain Noise Parameters	445
12.3.5 Cascaded Equations of Noise Figure	446
12.3.6 Sensitivity of a Receiver	448
12.4 Nonlinearity	453
12.4.1 Nonlinearity of a Device	453
12.4.2 IP (Intercept Point) and IMR (Intermodulation Rejection)	461
12.4.3 Cascaded Equations of Intercept Point	472
12.4.4 Nonlinearity and Distortion	479
12.5 Other Parameters	480
12.5.1 Power Supply Voltage and Current Drain	480
12.5.2 Part Count	482
12.6 Example of RF System Analysis	482
Appendices	485
12.A.1 Conversion between Watts, Volts, and $\text{dB}_m$ in a System with $50 \Omega$ Input and Output Impedance	485
12.A.2 Relationship between voltage reflection coefficient, $\Gamma$ , and Transmission coefficients when the load $R_o$ is equal to the standard characteristic resistance, $50 \Omega$ )	485
12.A.3 Definition of Powers in a Two-Port Block by Signal Flow Graph	488
12.A.4 Main Noise Sources	489

References	491
Further Reading	491
Exercises	493
Answers	494
<b>13 SPECIALITY OF "ZERO IF" SYSTEM</b>	<b>501</b>
13.1 Why Differential Pair?	501
13.1.1 Superficial Difference between Single-Ended and Differential Pair	501
13.1.2 Nonlinearity in Single-Ended Stage	503
13.1.3 Nonlinearity in a Differential Pair	505
13.1.4 Importance of Differential Configuration in a Direct Conversion or Zero IF Communication System	507
13.1.5 Why Direct Conversion or Zero IF?	508
13.2 Can DC Offset be Blocked out by a Capacitor?	508
13.3 Chopping Mixer	511
13.4 DC Offset Cancellation by Calibration	516
13.5 Remark on DC Offset Cancellation	517
Further Reading	517
Exercises	518
Answers	519
<b>14 DIFFERENTIAL PAIRS</b>	<b>521</b>
14.1 Fundamentals of Differential Pairs	521
14.1.1 Topology and Definition of a Differential Pair	521
14.1.2 Transfer Characteristic of a Bipolar Differential Pair	524
14.1.3 Small Signal Approximation of a Bipolar Differential Pair	527
14.1.4 Transfer Characteristic of a MOSFET Differential Pair	528
14.1.5 Small Signal Approximation of a MOSFET Differential Pair	530
14.1.6 What Happens If Input Signal Is Imperfect Differential	531
14.2 CMRR (Common Mode Rejection Ratio)	533
14.2.1 Expression of CMRR	533
14.2.2 CMRR in a Single-Ended Stage	539
14.2.3 CMRR in a Pseudo-Differential Pair	539
14.2.4 Enhancement of CMRR	541
Reference	542
Further Reading	542
Exercises	542
Answers	543
<b>15 RF BALUN</b>	<b>547</b>
15.1 Introduction	547
15.2 Transformer Balun	549

15.2.1	Transformer Balun in RF Circuit Design with Discrete Parts	550
15.2.2	Transformer Balun in RFIC Design	550
15.2.3	An Ideal Transformer Balun for Simulation	551
15.2.4	Equivalence of Parts between Single-Ended and Differential Pair in Respect to an Ideal Transformer Balun	555
15.2.5	Impedance Matching for Differential Pair by means of Transformer Balun	568
15.3	LC Balun	571
15.3.1	Simplicity of LC Balun Design	572
15.3.2	Performance of a Simple LC Balun	572
15.3.3	A Practical LC Balun	576
15.4	Microstrip Line Balun	580
15.4.1	Ring Balun	580
15.4.2	Split Ring Balun	582
15.5	Mixing Type of Balun	583
15.5.1	Balun Built by Microstrip Line and Chip Capacitor	583
15.5.2	Balun Built by Chip Inductors and Chip Capacitors	585
	Appendices	586
15.A.1	Transformer Balun Built by Two Stacked Transformers	586
15.A.2	Analysis of a Simple LC Balun	588
15.A.3	Example of Calculating of $L$ and $C$ Values for a Simple LC Balun	592
15.A.4	Equivalence of Parts between Single-Ended and Differential Pair with Respect to a Simple LC Balun	592
15.A.5	Some Useful Couplers	602
15.A.6	Cable Balun	603
	Reference	604
	Further Reading	604
	Exercises	605
	Answers	606

## 16 SOC (SYSTEM-ON-A-CHIP) AND NEXT

16.1	SOC	611
16.1.1	Basic Concept	611
16.1.2	Remove Bottlenecks in Approach to RFIC	612
16.1.3	Study Isolation between RFIC, Digital IC, and Analog IC	612
16.2	What is Next	612
	Appendices	615
16.A.1	Packaging	615
	References	621
	Further Reading	622
	Exercises	622
	Answers	623

<b>PART 3 INDIVIDUAL RF BLOCKS</b>	<b>625</b>
<b>17 LNA (LOW-NOISE AMPLIFIER)</b>	<b>627</b>
17.1 Introduction	627
17.2 Single-Ended Single Device LNA	628
17.2.1 Size of Device	629
17.2.2 Raw Device Setup and Testing	632
17.2.3 Challenge for a Good LNA Design	639
17.2.4 Input and Output Impedance Matching	646
17.2.5 Gain Circles and Noise Figure Circles	648
17.2.6 Stability	649
17.2.7 Nonlinearity	653
17.2.8 Design Procedures	655
17.2.9 Other Examples	656
17.3 Single-Ended Cascode LNA	662
17.3.1 Bipolar CE–CB Cascode Voltage Amplifier	662
17.3.2 MOSFET CS–CG Cascode Voltage Amplifier	666
17.3.3 Why Cascode?	669
17.3.4 Example	671
17.4 LNA with AGC (Automatic Gain Control)	684
17.4.1 AGC Operation	684
17.4.2 Traditional LNA with AGC	686
17.4.3 Increase in AGC Dynamic Range	688
17.4.4 Example	689
References	690
Further Reading	690
Exercises	691
Answers	692
<b>18 MIXER</b>	<b>695</b>
18.1 Introduction	695
18.2 Passive Mixer	698
18.2.1 Simplest Passive Mixer	698
18.2.2 Double-Balanced Quad-Diode Mixer	699
18.2.3 Double-Balanced Resistive Mixer	702
18.3 Active Mixer	706
18.3.1 Single-End Single Device Active Mixer	706
18.3.2 Gilbert Cell	708
18.3.3 Active Mixer with Bipolar Gilbert Cell	712
18.3.4 Active Mixer with MOSFET Gilbert Cell	715
18.4 Design Schemes	717
18.4.1 Impedance Measuring and Matching	717
18.4.2 Current Bleeding	718
18.4.3 Multi-tanh Technique	719

18.4.4	Input Types	722
Appendices		723
18.A.1	Trigonometric and Hyperbolic Functions	723
18.A.2	Implementation of $\tanh^{-1}$ Block	724
References		726
Further Reading		726
Exercises		726
Answers		727
<b>19</b>	<b>TUNABLE FILTER</b>	<b>731</b>
19.1	Tunable Filter in A Communication System	731
19.1.1	Expected Constant Bandwidth of a Tunable Filter	732
19.1.2	Variation of Bandwidth	732
19.2	Coupling between two Tank Circuits	733
19.2.1	Inappropriate Coupling	735
19.2.2	Reasonable Coupling	738
19.3	Circuit Description	738
19.4	Effect of Second Coupling	739
19.5	Performance	743
Further Reading		746
Exercises		747
Answers		747
<b>20</b>	<b>VCO (VOLTAGE-CONTROLLED OSCILLATOR)</b>	<b>749</b>
20.1	“Three-Point” Types of Oscillator	749
20.1.1	Hartley Oscillator	751
20.1.2	Colpitts Oscillator	753
20.1.3	Clapp Oscillator	753
20.2	Other Single-Ended Oscillators	755
20.2.1	Phase-Shift Oscillator	755
20.2.2	TITO (Tuned Input and Tuned Output) Oscillator	757
20.2.3	Resonant Oscillator	757
20.2.4	Crystal Oscillator	758
20.3	VCO and PLL (Phase Lock Loop)	759
20.3.1	Implication of VCO	759
20.3.2	Transfer Function of PLL	760
20.3.3	White Noise from the Input of the PLL	763
20.3.4	Phase Noise from a VCO	764
20.4	Design Example of a Single-Ended VCO	769
20.4.1	Single-Ended VCO with Clapp Configuration	769
20.4.2	Varactor	770
20.4.3	Printed Inductor	770
20.4.4	Simulation	773
20.4.5	Load-Pulling Test and VCO Buffer	776
20.5	Differential VCO and Quad-Phases VCO	778

Reference	783
Further Reading	783
Exercises	784
Answers	784
<b>21 PA (POWER AMPLIFIER)</b>	<b>789</b>
21.1 Classification of PA	789
21.1.1 Class A Power Amplifier	790
21.1.2 Class B Power Amplifier	790
21.1.3 Class C Power Amplifier	791
21.1.4 Class D Power Amplifier	791
21.1.5 Class E Power Amplifier	792
21.1.6 Third-Harmonic-Peaking Class F Power Amplifier	793
21.1.7 Class S Power Amplifier	794
21.2 Single-Ended PA	794
21.2.1 Tuning on the Bench	795
21.2.2 Simulation	796
21.3 Single-Ended PA IC Design	798
21.4 Push–Pull PA Design	799
21.4.1 Main Specification	799
21.4.2 Block Diagram	799
21.4.3 Impedance Matching	800
21.4.4 Reducing the Block Size	804
21.4.5 Double Microstrip Line Balun	808
21.4.6 Toroidal RF Transformer Balun	817
21.5 PA with Temperature Compensation	822
21.6 PA with Output Power Control	823
21.7 Linear PA	824
References	828
Further Reading	828
Exercises	829
Answers	829
<b>INDEX</b>	<b>833</b>



---

# PREFACE TO THE SECOND EDITION

---

I wrote the book titled *RF Circuit Design* in the United States, which was published by John Wiley & Sons, Inc. in 2009. It contains three parts:

1. Introduction to individual RF circuit block design. This part resembles existing books on RF circuit design. The topics concern main RF blocks such as the LNA (low-noise amplifier), Mixer, PA (power amplifier), VCO (voltage-controlled oscillator), PLL (phase lock loop), and so on. Most published RF books or articles focus largely on the description of operating principles of the circuitry. Distinctively, this part of the book emphasizes the actual engineering design procedures and schemes.

This part could be categorized as “longitudinal.”

2. Summary of skills and technologies in RF circuit design. Instead of describing circuit operating principles, the second part describes general design skills and technologies, such as impedance matching, RF grounding, layout, jeopardy in RFIC and SOC (system-on-a-chip) design,  $6\sigma$  design, and so on. This part is derived from my own design experience of over 20 years, highlighting both successes and failures. Therefore, it is unique among the published books on RF circuit design and represents the special feature of this book.

This part could be categorized as “transversal.”

3. Basic parameters of an RF system and the fundamentals of RF system design. This part considers a “must” theoretical background to an RF circuit designer, who should fully understand the basic RF parameters so that he can design the RF circuitry to serve the entire system.

Till date, more than 60 lectures on the subjects of this book have been held in mainland China, Taiwan, Hong Kong, and Singapore.

I received many precious comments and valuable inputs from readers after the first edition was published. This encouraged me and promoted the desire to work on a second edition. The following are the main changes in this book from the first edition:

1. Emphasis of the skills and technologies in RF circuit design. In order to emphasize the importance of the skills and technologies in the RF circuit design, the second part in the first edition that covers skills and technologies in RF circuit design is shifted as the first part in the second edition. To an RF circuit designer, no matter whether he or she would like to be a good engineer, a qualified professor, or an authoritative academic, the foremost objective is to master the design skills and technologies in the RF circuit design.
2. It is expected that this book can be adapted as a textbook for university courses. In order to help students further familiarize themselves with the topics of this

book, exercises are included at the end of each chapter. This may be convenient to those professors who would like to select this book as a textbook in their electrical engineering courses. In other words, it is expected that this book would be not only a science–technology–engineering reference but also a candidate for a textbook.

3. Expansion of topics. In addition to the rearranging of chapters or paragraphs, some chapters have been split up and new chapters have been inserted, increasing the number of chapters from 18 in the first edition to 21 in the second edition.

Finally, I express my deep appreciation to my lovely sons, Bruno Sie Li and Bruce Xin Li, who checked and corrected my English writing for this book. Also, it should be noted that unconventional descriptions, prejudices, or mistakes may inevitably appear in this book since most of the raw material comes from my own engineering designs and theoretical derivations. Comments or corrections from readers would be highly appreciated. My email address is [chihsili@yahoo.com.cn](mailto:chihsili@yahoo.com.cn).

*Fort Worth, TX, USA, 2011*

RICHARD CHI HSI LI

---

# PART 1

---

## DESIGN TECHNOLOGIES AND SKILLS

---



# DIFFERENCE BETWEEN RF AND DIGITAL CIRCUIT DESIGN

## 1.1 CONTROVERSY

For many years, there has been continued controversy between digital and RF circuit designers, some of which are given below:

- RF circuit designers emphasize impedance matching, whereas digital circuit designers are indifferent to it.
- RF circuit designers are concerned with frequency response, whereas digital circuit designers are interested in the waveform, or “eye’s diagram.” In other words, RF circuit designers prefer to work in the frequency domain, whereas digital circuit designers like to work in the time domain.
- As a consequence of the above, in a discussion of the budget for equipment, RF circuit designers like to purchase good network analyzers, whereas digital circuit designers prefer to buy the best oscilloscopes.
- RF circuit designers use the unit of  $\text{dB}_W$ , whereas digital circuit designers insist on using  $\text{dB}_V$ .
- Not only are the design methodologies different, so are their respective jargons. Digital circuit designers talk about AC bypass capacitors or DC blocking capacitors, but RF circuit designers rename those as “zero” capacitors.

It almost seems as if they were two different kinds of aliens from different planets. Even in some conferences or publications, these two kinds of “aliens” argue with

each other. Each tries to prove that their design methodology is superior to the others'. Eventually, nobody is the winner.

Let us outline the main controversies in the following.

### 1.1.1 Impedance Matching

The phrase “impedance matching” comes out of RF circuit designers’ mouths almost everyday. They were told by their supervisors that impedance matching is a “must” skill in circuit design. On the other hand, such terminology is never heard among digital circuit designers. Their supervisors tell them, “ignore that ‘foreign language’ just focus on the ‘eye diagram,’ or waveform.”

It is not just digital circuit designers who ignore impedance matching. Even some RF circuit designers “discovered” something new in their “advanced” RFIC (RF integrated circuit) design. While it was necessary to take care of impedance matching in RF module design or in RF blocks built by discrete parts, where the incident and reflected power in the circuit really existed, they thought it unnecessary to take care of impedance matching in an RFIC circuit design because the size of an IC die is so small as to render distinguishing the incident and reflective power or voltage redundant or meaningless. In agreement with their assertions, in the IC realm the design methodology for the RF circuit should be more or less the same as that for the digital circuit. Since then, they have been designing RF circuit blocks with the same method as used for digital circuit blocks. All the individual RF blocks are simply crowded together since “impedance matching between the individual blocks is not necessary.” Their design methodology for RF blocks is specially named as the “Combo” or “Jumbo” design. Theoretically, they thought that all kinds of circuitry must obey Ohm’s law and follow KCL (Kirchhoff’s current law) and KVL (Kirchhoff’s voltage law) rules without exception. So, why is the difference of design methodology? From their viewpoint, it seems unnecessary to divide the circuit design team into an RF and a digital circuit design group accordingly.

RF circuit designers would be very happy if impedance matching was unnecessary because impedance matching is the most difficult task in RF circuit design, especially in RFIC design for the UWB (ultrawide-band) system. Unfortunately, design experience indicates that the Combo or Jumbo design philosophy is absolutely wrong. For instance, without impedance matching, a LNA (low-noise amplifier) becomes a noisy attenuator or an oscillator in an RFIC chip. Without impedance matching, a mixer would become a “real” mixer indeed, blending all desired signals and undesired interference or noise together!

The key point to stop the controversy is whether the concept of voltage or power reflection is available in RF or digital circuitry. Should the reflection of voltage or power not exist in a practical circuitry, the idea of a Combo or Jumbo design could be a correct design methodology. On the contrary, if the reflection of voltage or power exists in a practical circuitry, impedance matching would be important for power transportation or manipulation in a circuitry, and then the idea of Combo or Jumbo design would be an incorrect design methodology.

As a matter of fact, the existence of power or voltage reflection can be deduced from a rough analysis of an RF block. For example, without impedance matching, the insertion loss of an LC passive filter could be significant. However, if the  $Q$  values of the inductors or capacitors are high, the LC passive filter itself should not conceivably produce a loss of power. This significant insertion loss demonstrates that quite a lot of power is reflected from the filter or load to the source. On the other hand, power or

voltage reflection is not related to the size of the block but to the impedance matching status between the source and the load. A simple example could illustrate the validity of such an assertion: light is reflected from a mirror in the same way no matter whether the light source is far from or very close to the mirror.

### 1.1.2 Key Parameter

There is a true story from a start-up company researching and developing a wireless communication system.

In spite of different opinions and various comments among his engineering teams, the engineering director asked both his RF and digital circuit design teams to work together for the system design of a communication system. He ruled that *voltage* must be taken as the key parameter to measure the performance of every block, including digital and RF blocks. In other words, the goal of the input and the output in every block, no matter RF or digital, must be specified with the voltage value. This engineering director hates the RF circuit designers' incessant "gossip" about power and impedance.

The engineers did try very hard to follow his instructions. There seemed to be no problem for the digital circuit blocks. However, the engineers were confused and did not know how to specify the goals for RF blocks by voltage instead of power.

By the RF engineers' understanding, all the parameters including G (power gain), NF (noise figure), IP<sub>3</sub> (3rd order intercept point), and IP<sub>2</sub> (2nd order intercept point) applied in RF circuit design were expressed by power but not voltage. In order to follow the director's instructions, they spent a lot of time to convert all the parameters from power to voltage, since power was the traditional unit and was read by most equipment. Sometimes, the conversion was meaningless or uncertain. For instance, by the unit of voltage, CNR (carrier-to-noise ratio) at the input of the demodulator was significantly dependent on the output impedance of the stage before the demodulator and the input impedance of the demodulator. Especially when the output impedance of the stage before the demodulator and the input impedance of the demodulator were different from each other, the conversion becomes impossible. Even more awkwardly, members of audience who attended the presentation meeting held by this system design team could not understand why the values appearing in the system plan were surprisingly higher or lower than those from other companies. Eventually, after they learnt of the extraordinary instructions given by the engineering director, the part of the audience equipped with calculators at hand could not but convert those values back from voltage to power!

Among this system design team, selection of a common key parameter for both RF and digital circuit designs became a hot topic. People argued with each other without result, while the director still insisted on his original instructions. After a couple of weeks, the system design still hung in the air and, finally, for unknown reasons the plan for the system design was dropped quietly. Some RF circuit designers felt upset and left the company despite the director's exhortations: "Nothing is Impossible!"

As a matter of fact, system design for a communication system must be divided into two portions: the digital portion and the RF portion. Yes, the key parameter in the digital circuit design is voltage or current. By means of voltage or current, all the intermediate parameters can be characterized. However, the key parameter in RF circuit design must be power or impedance. By means of power and impedance, all the intermediate parameters in a RF circuit block can be characterized. Impedance matching ensures the best performance of power transportation or manipulation in RF circuit blocks; therefore, impedance can be taken as the key parameter in RF circuit design.

Why? The answer can be found in the following sections.

### 1.1.3 Circuit Testing and Main Test Equipment

In addition to the arguments about impedance matching and the key parameters, the difference between digital and RF circuit design can also be found in circuit testing and test equipment.

In a digital test laboratory, the test objective is always voltage, and occasionally current. There are many pieces of test equipment available in a digital test laboratory; however, the main test equipment is the oscilloscope. The oscilloscope can sense the voltage at any node in the circuitry and display its eye diagram or a waveform on screen, which characterizes the performance of a digital circuit intuitively. In general, digital circuit designers prefer to analyze the circuitry in the time domain because the speed of response is important to the performance of a digital circuit block.

In an RF test laboratory, the test objective is always power. Most RF test equipment, such as the spectrum analyzer, noise meter, signal generator, and so on, measure the parameters of an RF circuit block in terms of power but not voltage. The main test equipment is the network analyzer. The performance of an RF circuit block can be characterized mainly by its frequency response on network analyzer screen, which is expressed by power gain or loss, in decibels. The RF circuit designer prefers to analyze the circuitry in the frequency domain because coverage of bandwidth is important to the performance of an RF block.

In the test laboratory, testing a digital circuit block is somewhat easier than testing an RF circuit block. In testing for a digital circuit block, the probe of an oscilloscope is usually a sensor with high impedance. It does not disturb the circuit performance when it touches a node in the circuitry.

On the other hand, while using a network analyzer, the circuit designers may worry about the difference of circuit performance before and after the tested equipment is connected to the desired test node, because the input and output impedance of the equipment is low, usually  $50\ \Omega$ . In most cases, it certainly will disturb the circuit performance.

Instead of voltage testing, the RF circuit designer is concerned with power testing. All power testing must be conducted under a good impedance matching condition so the test equipment must be well calibrated. Unlike the testing for a digital circuit block by an oscilloscope, a buffer connected between the desired test node and the input of the network analyzer is not allowed because all the power tests for the RF block must be conducted under the condition of impedance matching.

So far, the different methodology between RF and digital circuit design has been introduced only in terms of the three main aspects above. More differences exist but will not be listed. We are going to focus on the explanation of where these differences come from.

## 1.2 DIFFERENCE OF RF AND DIGITAL BLOCK IN A COMMUNICATION SYSTEM

### 1.2.1 Impedance

The input and output impedance of an RF circuitry are usually pretty low. In most cases, they are typically  $50\ \Omega$ . On the contrary, the input and output impedances in a digital circuitry are usually quite high. For example, the input and output impedances of an Op-Amp (operating amplifier) are mostly higher than  $10\ k\Omega$ .

The lower impedance in an RF circuitry is beneficial to deliver power to a block or a part. It is well known that the power of a signal delivered to a block or a part with impedance  $Z$  can be expressed by

$$P = vi = \frac{v^2}{Z}, \quad (1.1)$$

where

- $P$  = the power delivered to a block or a part,
- $v$  = the AC or RF voltage across the block or the part,
- $i$  = the AC or RF current flowing through the block or part, and
- $Z$  = the impedance of the block or the part.

For a given value of power,  $v^2$  is proportional to  $Z$ . This implies that, in order to deliver a given power to a block or a part, a higher voltage must be provided if its impedance is high. On the contrary, a lower voltage across the block is enough to deliver the same given power to a block or a part if its impedance is low. From the viewpoint of either cost or engineering design of the circuit, the application of a lower voltage is much better than that of a higher voltage. It is one of the reasons why the input and output impedance in the RF blocks are intentionally assigned to be low because only a lower voltage is needed in order to deliver the same given power to a block or part with low impedance.

However, it is just the opposite for a digital signal. The higher impedance in digital circuitry is beneficial to the voltage swing in a digital block or part. For a given current, a higher impedance can have a higher voltage swing across a block or a part, and then the signal can ON/OFF the device more effectively, because

$$v = iZ. \quad (1.2)$$

The question is: why is RF circuitry focused on the power while digital circuitry is concerned about voltage?

### 1.2.2 Current Drain

In RF circuit blocks, the current drains are usually in the order of milliamperes while in digital circuit blocks they are usually in the order of microamperes. That is, the difference of the current drain's magnitude between RF and digital circuit blocks is approximately 1000 times.

In RF circuit blocks, it is desirable to increase the power of the RF signal as much as possible. This implies that higher current drains are preferred in RF circuit blocks because they are beneficial to deliver power to the block or the part for a given voltage.

In digital circuit blocks, it is desirable to reduce the power of the digital signal as much as possible. This implies that lower current drains are preferred in digital circuit blocks as long as the voltage swing is high enough.

Again, the question is: why is RF circuitry focused on power while digital circuitry is concerned with voltage? The answer can be found in the following section.

### 1.2.3 Location

In a communication system, the demodulator is a remarkable demarcation in the receiver. As shown in Figure 1.1, before the demodulator, the blocks operate in the range of radio

frequency so that they are called *RF blocks*. They are sometimes called the *RF front end* in the receiver, where the RF circuit design is conducted. After demodulation, the blocks operate in the range of intermediate frequency or in the low digital data rate and are categorized as baseband blocks, or the digital/analog section. They are sometimes called the *back end* in the receiver, where digital/analog circuit design is conducted. The demodulator is a critical block in which both digital and RF design technology are needed.

The order of blocks in the transmitter is just opposite. Before the modulator, the blocks operate in the range of intermediate frequency or in the low digital data rate and are categorized as baseband blocks, or the digital/analog section. They are sometimes called the *front end* in the transmitter, where digital/analog circuit design is conducted. After the modulator, the blocks operate in the range of radio frequency so that they are called *RF blocks* and sometimes the *RF back end* in the transmitter, where RF circuit design is conducted. The modulator is also a critical block, in which both digital and RF design technology are needed.

A common feature can be seen from Figure 1.1. In either the receiver or the transmitter, the circuit portion close to the antenna side contains RF blocks and the portion farther from antenna side contains digital/analog circuit blocks.

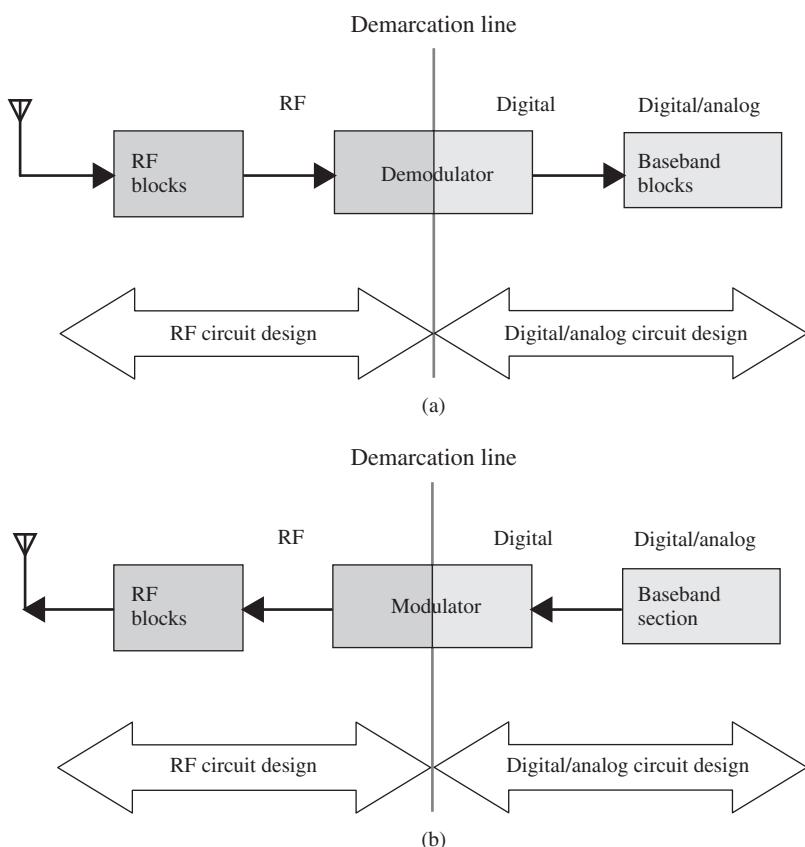


Figure 1.1. Demarcation line in a communication system. (a) Receiver. (b) Transmitter.

In the receiver, the received modulated carrier is usually very weak. After it is power-magnified by the LNA and its frequency mixed down by the mixer, the modulated carrier can be demodulated only if its power is strong enough to suppress the noise power at the input of the demodulator. Typically, the ratio of RF signal power to the noise power at the input of the demodulator is required to be more than 10 dB. It is therefore required that the RF signal be power-transported or power-operated before demodulation. After the demodulator, the digital-type message is demodulated from the RF to the base band. The digital signal is not required to be “power-transported” but is only “status-maintained” or voltage-transported between local blocks for digital signal processing. The voltage represents the status of the signal. For the sake of power saving, the power of the signal is reduced as much as possible. This answers the question why voltage transportation or manipulation or the “status” transportation or manipulation is a logical task in the digital circuit design.

Similarly, in the transmitter, the digital signal is only required to reach the “modulation-effective status level” before the modulator. This implies that the power or voltage of the input digital signal to the modulator could be as low as possible as long as the input voltage or power reaches a level by which the carrier can be effectively modulated. In this case, the digital signal is transported or manipulated between the local circuit blocks and is not required to be power-transported but only voltage-transported. However, the modulated carrier after the modulator must be power-magnified and delivered to the antenna so that the modulated carrier is powerful enough to propagate to a receiver located a long distance from the transmitter.

### 1.3 CONCLUSIONS

From the discussion in Section 1.2.3, it can be concluded that the power transportation or manipulation, but not voltage transportation or manipulation, is required in the RF blocks before the demodulator in a receiver and after the modulator in the transmitter. The voltage transportation or manipulation, but not power transportation or manipulation, is required in the digital blocks after the demodulator in a receiver or before the modulator in a transmitter.

This is the answer to the question in Sections 1.2.2 and 1.2.3: why is RF circuitry focused on the power while the digital circuitry is concerned about voltage?

It can be seen that the controversy between digital and RF designers is not necessary. The difference in circuit design methodology arises from the different circuit design tasks. The digital circuit is designed for voltage transportation or manipulation, while the RF circuit is designed for power transportation or manipulation.

### 1.4 NOTES FOR HIGH-SPEED DIGITAL CIRCUIT DESIGN

In digital communication, the synonym of “high speed” is high data rate.

In the case of high digital data rates, the tasks of both types of circuits are unchanged: the RF circuit still works for the power while digital circuit still works for the *status* transportation or operation. However, the design methodology in high data rate digital circuit designs is close to that of the RF circuit design methodology, because of the following reasons:

1. The remarkable variation of a digital circuit design from low speed to high speed is the change of the input and output impedance of a digital block. In the digital circuit design for low speeds, high input and output impedance of a digital block can be obtained from high input and output impedance of a transistor. The input and output capacitance of a transistor can impact on the input or output impedance so that they impact on the rising or dropping time. In the digital circuit design for high speeds, the input and output impedance of a transistor is reduced. The low input and output impedance leads to the necessity of impedance matching in the digital circuit design. Impedance matching is beneficial not only to power transportation but also to voltage transportation. Without impedance matching, the reflected voltage will appear and interfere with the incident voltage. This brings about additional attenuation, additional jitter, an additional cross talk, and eventually an additional bit error to the digital signal.
2. The traditional layout scheme is no longer reliable in the case of digital circuits with high data rates. For example, in the traditional layout for digital circuits with low data rates, the runners are always lined up in parallel so that the entire layout looks nice and neat. However, in the layout for digital circuits with high data rates, the runners lined up in parallel could cause appreciable coplanar capacitance and are coupled with each other, and, consequently, it may cause interference or cross talk. The layout for high-speed digital circuitry must be taken as seriously as for the RF circuitry.
3. The traditional AC grounding scheme for a digital circuit with low data rate is no longer reliable in the case of a digital circuit with high data rate. For high-speed digital circuitry, the AC grounding must be taken as seriously as for an RF circuitry.
4. In a digital circuitry with high data rate, isolation may become a serious problem. Usually, a digital signal is a rectangular pulse while an RF signal is sinusoidal. The former contains a wide-band spectrum while the latter contains a narrow-band spectrum. This implies that a digital circuit with a high data rate is easier to be interfaced with inside or outside interference sources because its frequency spectrum is much wider than that of a digital circuit with low data rate. Isolation between blocks becomes important and should be taken as seriously as for an RF circuitry.

## FURTHER READING

- Li RC-H. Key Issues in RF/RFIC Circuit Design. Beijing: Higher Education Press; 2005.
- Sakurai T., Low power digital circuit design, 31th European Solid-State Circuits Conference (ESS-CIRC), 2004, Leuven, Belgium, 11–18.
- Thomas H. Lee, The Design of CMOS Radio-Frequency, Integrated Circuit, Cambridge University Press, 1998.
- Turnbull JR, Transistor switching circuits. International Solid State Circuits Conference, Digest of Technical Papers; 1963 Feb 21; University of Pennsylvania-University Museum; 1963. pp. 58–59.
- Xiao E, Yuan JS. RF circuit design in reliability. IEEE Radio Frequency Integrated Circuits Symposium; Philadelphia (PA); 2003. pp. 575–578.

## EXERCISES

1. Specify the main difference in task between RF and digital circuit designs.
2. Compare and comment on the task between RF and digital circuit designs when the digital data rate is much less than the RF carrier frequency.
3. What is the difference in digital circuit design between low and high data rate cases?
4. High impedance is beneficial to voltage transportation or manipulation while low impedance is beneficial to power transportation or manipulation. Why?
5. Why is impedance the main parameter in RF circuit design?
6. Why is the normalized impedance regulated as  $50 \Omega$ ?
7. Assuming that the input capacitance of a MOSFET (metal–oxide–semiconductor field-effect transistor) is  $0.15915 \text{ pF} = 1/(2\pi) \text{ pF}$ , what is the input impedance for an RF signal if its operating frequency is 10, 100, 1000, and 10,000 MHz, respectively? Also, what is the input impedance for a digital signal to be considered if its digital data rate is 10, 100, 1000, 10,000 Mb/s, respectively, and if its second and third harmonics are as important as the main frequency corresponding to the digital data rate?
8. Why is power the transportation type of an RF signal while status (voltage or current) is the transportation type for a digital signal?
9. Why is impedance matching important in RF but not in digital circuit design when  $R \ll f_{RF}$ ?
10. Why is impedance matching important not only in RF but also in digital circuit design when  $R \approx f_{RF}$  ?

## ANSWERS

1. The main difference of task between RF and digital circuit design can be tabulated as follows:

<u>Item</u>	<u>RF Circuit Design</u>	<u>Digital Circuit Design</u>
Transportation type	Power	Status (voltage or current)

2. When the data rate is low, the difference between RF and digital circuit can be tabulated as follows:

<u>Item</u>	<u>RF Module/RFIC</u>	<u>Digital Circuit (Low Data Rate)</u>
Impedance	Low ( $50 \Omega$ typically)	High (infinity ideally)
Current	High (mA)	Low ( $\mu\text{A}$ )
<i>Location in a Communication System</i>		
• Rx	Front end (before demodulation)	Back end (after demodulation)
• Tx	Back end (after modulation)	Front end (before modulation)
Transportation type	Power (W)	Status (voltage or current)
Impedance matching	Important	Unimportant (usually)

3. In the low digital data rate case, the input or output impedance in the digital circuit is usually high. However, in the high data rate case, the input or output impedance

in the digital circuit is not high because of the existence of the input or output capacitance of the device. Impedance matching becomes effective in the voltage transportation. The voltage will be pumped up when the load resistance is greater than the source resistance, that is,  $R_L > R_S$ .

4. Low impedance is beneficial to power transportation or manipulation because,

$$P = vi = \frac{v^2}{Z}$$

for a definite voltage, the power becomes high if the impedance is low (say,  $50 \Omega$ ). On the contrary, high impedance is beneficial to status transportation because

$$v = iZ$$

for a definite current, and the voltage swing becomes large if the impedance is high.

5. The main task of the RF circuit is to transport or to manipulate the power of signal. This is the main reason that the impedance becomes the main parameter to be taken care of in an RF circuit design. The performance of power transportation or manipulation is mainly determined by the status of impedance matching.
6. For a coaxial cable with an air dielectric, in order to maximize the power-handling capability, one could increase the diameter of the inner wire. However, there are two competing effects: the breakdown voltage is increased but the characteristic impedance  $Z_0$  would be increased also, which would tend to reduce the power deliverable to a load. The maximum of the power-handling capability can be reached if the diameters of the inner wire and the outer conductor are well-selected so that a  $Z_0$  of  $30 \Omega$  is approached. On the other hand, in order to minimize the attenuation of a coaxial cable, there are also two competing effects due to the increase of the inner wire diameter: both of resistance per unit length and the characteristic impedance  $Z_0$  would be reduced at the same time, which would tend to an uncertainty of the attenuation. The minimum of the attenuation can be reached if the diameters of the inner wire and the outer conductor are well-selected so that a  $Z_0$  of  $77 \Omega$  is approached...'' Since  $77 \Omega$  gives us minimum loss and  $30 \Omega$  gives us maximum power-handling capability, a reasonable compromise is a round average value,  $Z_0$  of  $50 \Omega$ .
7. The input impedance for an RF signal is  $100 \text{ k}\Omega$ ,  $10 \text{ k}\Omega$ ,  $1000 \Omega$ , and  $100 \Omega$ , respectively. The input impedance for a digital signal is  $33.3 \text{ k}\Omega$ ,  $3.3 \text{ k}\Omega$ ,  $333.3 \Omega$ , and  $33.3 \Omega$ , respectively, if its second and third harmonics are as important as the main frequency corresponding to the digital data rate.
8. In the receiver, RF and digital circuits are located before and behind the demodulator, respectively. The performance of RF circuits must be good in power transportation so that the threshold of the CNR power ratio at the input of demodulator can be reached or exceeded and consequently the demodulation becomes available. On the contrary, the task of the digital circuits after the demodulator is to transport or manipulate the status (voltage or current) of digital signals. In the transmitter, digital and RF circuits are located before and behind the modulator, respectively. The task of the digital circuits before the modulator is to transport or manipulate the status (voltage or current) of digital signals. On the contrary, the performance of RF circuits after modulator must be good in power transportation

so that the CNR power ratio at the output of modulator can be radiated to remote places where the receiver antenna is located.

9. When  $R \ll f_{RF}$ , impedance matching is important in RF but not in digital circuit design because the carrier frequency  $f_{RF}$  of RF signal is high, whereas the frequency components of digital signals are low. The input/output impedances in RF circuits are low, while the input/output impedances in digital circuits are high. Impedance matching is important when the input/output impedances of a circuitry are low, but it is not important when the input/output impedances of a circuitry are high.
10. When  $R \approx f_{RF}$ , impedance matching is not only important in RF but also in digital circuit design because both the carrier frequency  $f_{RF}$  of RF signal and the frequency components of digital signals are high. Both of the input/output impedances in RF circuits and the input/output impedances in digital circuits are low. Impedance matching is important when the input/output impedances of a circuitry are low.



# REFLECTION AND SELF-INTERFERENCE

## 2.1 INTRODUCTION

Generally speaking, there are internal and external sources of interference to the signal in RF, digital, or analog circuitry.

The internal interference sources include the following:

1. Reflection at every node in the circuitry, which is called *self-interference*;
2. Nonlinearity of the device;
3. Additional radiation or interference due to improper grounding, layout, and packaging of the circuitry;
4. Electromagnetic radiation from a part, a runner, or a current loop on a PCB (printed circuit board) since any of these acts as a small antenna.

The external interference sources include the following:

5. There are two types of the artificial electromagnetic interference sources: one is electromagnetic waves transmitted from broadcasting stations, stations for navigation, remote control, exploration, detection, and so on. The other is electromagnetic emission from industrial equipment, home facilities, traffic vehicles, and so on.
6. Environmental interference sources such as lightning, cosmic rays, and so on.

All the interferences mentioned above must be the objectives to be studied in an EMC (electromagnetic compatibility) course. Unfortunately, in most EMC courses, the

main interferences 1 and 2 are excluded though 3 to 4 are covered. Consequently, many EMC problems remain unknown since many of these problems arise because of the interference sources 1 and 2.

The ideal case without voltage or power reflection never exists in a practical circuitry. Self-interference due to voltage or power reflection will be discussed in this chapter. Interference due to sources 3 and 4 would be analyzed in Chapters 7–9. The interference due to source 2, that is, nonlinearity of the device, will be explored in Chapter 12.

As to the interference due to the external interference sources, it is difficult to find a formal solution to handle the EMC problem. The solution definitely relies on the special features of both artificial and natural interference sources.

## 2.2 VOLTAGE DELIVERED FROM A SOURCE TO A LOAD

### 2.2.1 General Expression of Voltage Delivered from a Source to a Load when $l \ll \lambda/4$ so that $T_d \rightarrow 0$

Figure 2.1 shows that a voltage delivered from the source to the load and transported along a runner, which essentially is a micro strip line with characteristic impedance  $Z_0$ , when the length  $l \ll \lambda/4$ . The impedances of the source and the load are

$$Z_S = R_S + jX_S, \quad (2.1)$$

$$Z_L = R_L + jX_L, \quad (2.2)$$

where

- $Z_S$  = the impedance of the source,
- $R_S$  = the resistance of the source,
- $X_S$  = the reactance of the source,
- $Z_L$  = the impedance of the load,
- $R_L$  = the resistance of the load,
- $X_L$  = the reactance of the load,
- $Z_0$  = the characteristic impedance of the runner,
- $\Gamma_S$  = the voltage reflection coefficient at the source,
- $\Gamma_L$  = the voltage reflection coefficient at the load,
- $P_S$  = the power available at the source,
- $v_S$  = the voltage available at the source,
- $P_{R_S}$  = the power on  $R_S$ ,
- $v_{R_S}$  = the voltage across  $R_S$ ,
- $P_{R_L}$  = the power on  $R_L$ ,
- $v_{R_L}$  = the voltage across  $R_L$ .

In the expressions (2.1) and (2.2),  $X_S$  is the reactance of the source and  $X_L$  is the reactance of the load.

Before the derivation of the general expression of the voltage delivered from a source to a load, it is first necessary to clarify the meaning of “delivered from a source to a load.” As shown in Figure 2.1, the precise phrase must be “delivered from a source to the real part of a load,” which does not include “the imaginary part of a load” because of the following reasons:

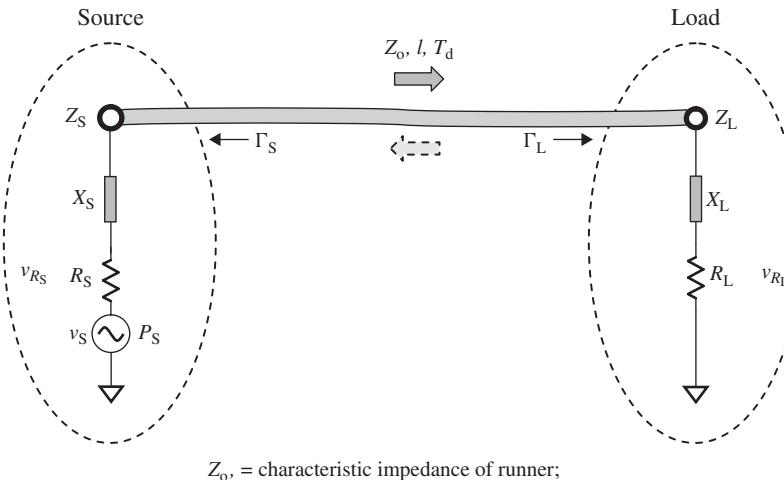


Figure 2.1. Voltage delivered from a source to a load.

1. It is well known that the average power across  $X_S$  or  $X_L$  over one period of the AC signal is zero. In other words, an ideal capacitor or inductor experiences only a process of charging and discharging but never receives or consumes any net power.
2. In practice, the attempt is to reduce the reactance of the source or load,  $X_S$  or  $X_L$ , and to have them neutralize each other as much as possible. In digital circuit design, the designers always try to select a device with low input capacitance, or try to neutralize it using an inductor or another method. In RF circuit design, the mutual neutralization of  $X_S$  or  $X_L$  is exactly one of the conditions of impedance conjugate matching.

Consequently, in the following discussion, the real meaning of voltage delivered from a source to a load is voltage delivered from a source to the real part of a load.

According to transmission line theory, the voltage reflection coefficients at the source and load are

$$\Gamma_S = \frac{Z_S - Z_0}{Z_S + Z_0}, \quad (2.3)$$

$$\Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0}. \quad (2.4)$$

Usually  $Z_0$  is  $50 \Omega$ , and, in general,

$$Z_S \neq Z_0, Z_L \neq Z_0, \quad (2.5)$$

and the voltage suffers a reflection at the source or load terminal: that is,

$$\Gamma_S \neq 0 \text{ and } \Gamma_L \neq 0. \quad (2.6)$$

The voltage delivered from source will bounce back and forth between the source and the load. The resultant voltage on  $R_L$  is the sum of all the remaining voltages on  $R_L$  after each bounce back and forth of the voltage between the source and the load.

Let us assume that the source delivers a pure sinusoidal voltage

$$v_S = v_{So} e^{j\omega t}, \quad (2.7)$$

where

$v_{So}$  = the amplitude of voltage, and  
 $\omega$  = the operating angular frequency.

The moment at which the source delivers this voltage is

$$t = 0, \quad (2.8)$$

and the arrival time of the delivered voltage at the load is the delayed time  $T_d$ .

Simultaneously there are many other voltages arriving at the load, which were delivered at the time moments  $t = \dots - 12T_d, -10T_d, -8T_d, -6T_d, -4T_d, -2T_d, 0$  where the negative sign indicates that it is past time. These voltages arrive at the load. Each voltage partially remains at the load and partially bounces back to the source. The resultant voltage at the load when  $t = T_d$ ,  $v_{R_L}|_{t=T_d}$  is the sum of all the remaining voltages on  $R_L$  after the reflected voltage bounces back and forth between source and load, and can be expressed as follows:

$$v_{R_L}|_{t=T_d} = v_{So} \frac{R_L}{Z_S + Z_L} (1 - \Gamma_L) \sum_{n=0}^{\infty} e^{j\omega[t-T_d-2nT_d]} (\Gamma_S \Gamma_L)^n, \quad (2.9)$$

or

$$\begin{aligned} v_{R_L}|_{t=T_d} &= v_{So} \frac{R_L}{Z_S + Z_L} (1 - \Gamma_L) e^{j\omega(t-T_d)} \\ &\quad + v_{So} \frac{R_L}{Z_S + Z_L} (1 - \Gamma_L) \sum_{n=1}^{\infty} e^{j\omega[t-(2n+1)T_d]} (\Gamma_S \Gamma_L)^n, \end{aligned} \quad (2.10)$$

where

$T_d$  = the delayed time of the voltage traveling from source to load, and  
 $v_{R_L}|_{t=T_d}$  = the resultant remaining voltage on the load.

The delayed time ( $t - T_d$ ) is explicitly kept in every exponential factor  $e^{j\omega(t-T_d-2nT_d)}$ . The time interval  $2nT_d$  appearing in the exponential factor  $e^{j\omega(t-T_d-2nT_d)}$  indicates the previous moment when the voltage term in the expression was delivered from the source.

Expression (2.10) is exactly the same as expression (2.9). In expression (2.10), the first term in expression (2.9) is extracted from the general summation expression so as to intentionally distinguish it from the others. The first term in expression (2.10) represents the remaining voltage on  $R_L$  after the first reflection from the load. The second summation term in expression (2.10) represents the sum of the remaining voltage on  $R_L$ , including all the reflected voltages bouncing back and forth between the source and the load in past time, which usually represents the self-interference due to the voltage reflection at the load.

It should be noted that all the voltage terms in expression (2.9) or (2.10) arrive at the load at the same time  $t = T_d$ .

The first term on the right-hand side of expression (2.9) or (2.10) represents the remaining voltage on  $R_L$  after the first reflection from the load with a delayed time  $T_d$ .

The second term on the right-hand side of expression (2.9) or (2.10) represents the remaining voltage on  $R_L$  after the second reflection from the load with a delayed time  $3T_d$ . The additional factor  $(\Gamma_S \Gamma_L)$  represents the reflected percentage after the voltage bounces back and forth once between the source and the load.

The third term ( $n = 2$ ) on the right-hand side of expression (2.9) or (2.10) represents the remaining voltage on  $R_L$  after the third reflection from the load with a delayed time  $5T_d$ . The additional factor  $(\Gamma_S \Gamma_L)^2$  represents the reflected percentage after the voltage bounces back and forth twice between the source and the load, and so on.

Figure 2.2 vividly depicts the voltages arriving at the load simultaneously when  $t = T_d$  but delivered from the source at different moments of  $t = \dots - 12T_d, -10T_d, -8T_d, -6T_d, -4T_d, -2T_d, 0, \dots$ . The number of voltages arriving at the load is infinite, and it is therefore impossible to depict them totally. Only eight of them are shown in Figure 2.2. The following points may be noted:

- The source and the load are located at the top left and top right, respectively.
- The distance from the source to the load is  $l$ , and the corresponding delay time for the delivered voltage moving from the source to the load is  $T_d$ .
- The time axis is directed vertically downwards.
- The voltages delivered from the source at different moments  $t = \dots - 12T_d - 10T_d, -8T_d, -6T_d, -4T_d, -2T_d, 0, \dots$  are remarked in left-hand side of Figure 2.2. In each remark box, the factor  $k_v$  denotes the portion of  $v_{So}$  that is delivered from the source to  $R_L$ .
- Finally, all the delivered voltages from the source at different moments that arrive at the load simultaneously are remarked with  $v_{R_L}|_{t=T_d}$  in the bottom-right corner of Figure 2.2.

If

$$T_d \rightarrow 0, \quad (2.11)$$

expression (2.9) becomes

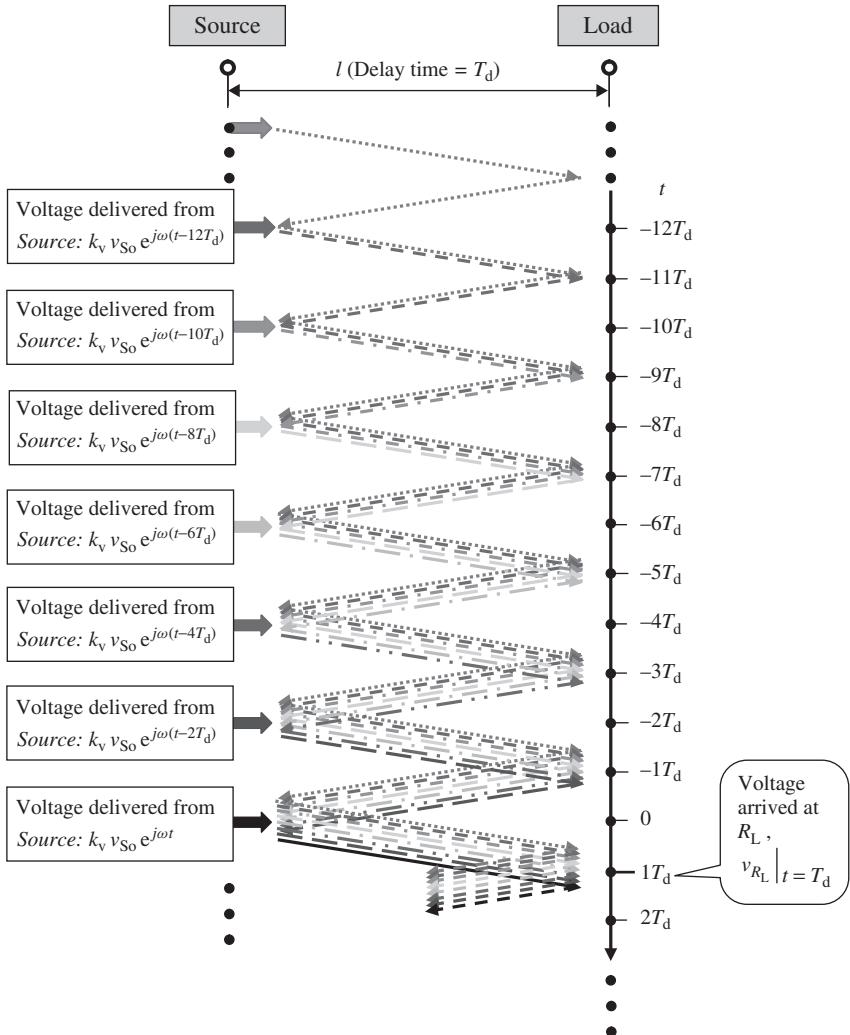
$$v_{R_L}|_{t=T_d} = v_{So} e^{j\omega t} \frac{R_L}{Z_S + Z_L} (1 - \Gamma_L) \sum_{n=0}^{\infty} (\Gamma_S \Gamma_L)^n. \quad (2.12)$$

Note that

$$\sum_{n=0}^{\infty} (\Gamma_S \Gamma_L)^n = \frac{1}{1 - \Gamma_S \Gamma_L}; \quad (2.13)$$

then

$$v_{R_L}|_{t=T_d} = v_{So} e^{j\omega t} \frac{R_L}{Z_S + Z_L} \frac{1 - \Gamma_L}{1 - \Gamma_S \Gamma_L}. \quad (2.14)$$



**Figure 2.2.** Voltages bouncing back and forth and arriving at the load when  $t = T_d$ .  
Note:  $k_v = R_L / (Z_S + Z_L)$ .

### 2.2.2 Additional Jitter or Distortion in a Digital Circuit Block

Now let us discuss the expression for the voltage transport (2.12) in the cases when the approximation (2.11) is satisfied. Again, the expression (2.12) can be rewritten as

$$v_{R_L}|_{t=T_d} = v_{S0} e^{j\omega t} \frac{R_L}{Z_S + Z_L} (1 - \Gamma_L) + v_{S0} e^{j\omega t} \frac{R_L}{Z_S + Z_L} (1 - \Gamma_L) \sum_{n=1}^{\infty} (\Gamma_S \Gamma_L)^n. \quad (2.15)$$

As mentioned earlier, the first term on the right-hand side of expression (2.15) represents the remaining voltage on  $R_L$  after the first reflection from the load with a delayed time  $T_d$ . It represents the desired voltage signal without reflection at the load. The second summation term in expression (2.15) represents the sum of the remaining voltages on  $R_L$ , including all the reflected voltages bouncing back and forth between the source and

the load in the past time, which usually represents the interference due to the voltage reflection at the load.

If the second summation term in expression (2.15) did not exist, the desired voltage signal would not be disturbed so that distortion or jitter would not occur. In other words, the expression (2.15) corresponds to the following expression:

$$v_{R_L}|_{t=T_d} = v_{R_L}|_{t=T_d,1st} + v_{R_L}|_{t=T_d,srv}, \quad (2.16)$$

where

- $v_{R_L}|_{t=T_d}$  = the summed voltage across  $R_L$ ,
- $v_{R_L}|_{t=T_d,1st}$  = the remaining voltage on  $R_L$  after the first reflection, and
- $v_{R_L}|_{t=T_d,srv}$  = the sequentially reflected voltages remaining on  $R_L$ .

By comparing (2.15) with (2.16), we have

$$v_{R_L}|_{t=T_d,1st} = v_{So}e^{j\omega t} \frac{R_L}{Z_S + Z_L} (1 - \Gamma_L), \quad (2.17)$$

$$v_{R_L}|_{t=T_d,srv} = v_{So}e^{j\omega t} \frac{R_L}{Z_S + Z_L} (1 - \Gamma_L) \frac{\Gamma_S \Gamma_L}{1 - \Gamma_S \Gamma_L}. \quad (2.18)$$

The sum of the reflected voltages  $v_{R_L}|_{t=T_d,srv}$  disturbs the incoming voltage  $v_{R_L}|_{t=T_d,1st}$ . The disturbance can be categorized into two types: additional distortion, or jitter, and additional interference. The additional distortion or jitter appears when the frequency of voltage is constant or when the sequentially reflected voltage has the same frequency as the incoming voltage.

The additional distortion can be evaluated as the ratio of the sum of reflected voltages  $v_{R_L}|_{t=T_d,srv}$  to the incoming voltage  $v_{R_L}|_{t=T_d,1st}$ , that is,

$$\Delta D_v|_{\%} = \frac{v_{R_L}|_{t=T_d,srv}}{v_{R_L}|_{t=T_d,1st}} = \frac{\Gamma_S \Gamma_L}{1 - \Gamma_S \Gamma_L}, \quad (2.19)$$

where  $\Delta D_v|_{\%}$  is the additional distortion in percent.

Alternatively, assuming that the additional jitter has the same percentage as that of the additional distortion, that is,

$$\Delta(\text{jitter})|_{\%} = \Delta D_v|_{\%} = \frac{\Gamma_S \Gamma_L}{1 - \Gamma_S \Gamma_L}, \quad (2.20)$$

the additional jitter can be calculated as

$$\Delta(\text{jitter})|_s = T \frac{v_{R_L}|_{t=T_d,srv}}{v_{R_L}|_{t=T_d,1st}} = \frac{1}{f} \frac{\Gamma_S \Gamma_L}{1 - \Gamma_S \Gamma_L}, \quad (2.21)$$

where

- $\Delta(\text{jitter})|_s$  = the additional jitter in s,
- $\Delta(\text{jitter})|_{\%}$  = the additional jitter in %,
- $T = 1/f$  = the operating period,
- $R$  = the digital data rate,

TABLE 2.1. Additional Distortion and Additional Jitter in Voltage Transport when  $f = 3.86$  GHz

$\Gamma_S$ , %	$\Gamma_L$ , %	$\Delta D$ , %	$f$ , GHz	$T$ , ns	Jitter, %	Jitter, ps
0	0	0.00	3.86	0.259	0.00	0
5	0	0.00	3.86	0.259	0.00	0
10	0	0.00	3.86	0.259	0.00	0
20	0	0.00	3.86	0.259	0.00	0
50	0	0.00	3.86	0.259	0.00	0
0	5	0.00	3.86	0.259	0.00	0
5	5	0.25	3.86	0.259	0.25	0.6
10	5	0.50	3.86	0.259	0.50	1.3
20	5	1.01	3.86	0.259	1.01	2.6
50	5	2.56	3.86	0.259	2.56	6.6
0	10	0.00	3.86	0.259	0.00	0
5	10	0.50	3.86	0.259	0.50	1.3
10	10	1.01	3.86	0.259	1.01	2.6
20	10	2.04	3.86	0.259	2.04	5.3
50	10	5.26	3.86	0.259	5.26	13.6
0	20	0.00	3.86	0.259	0.00	0
5	20	1.01	3.86	0.259	1.01	2.6
10	20	2.04	3.86	0.259	2.04	5.3
20	20	4.17	3.86	0.259	4.17	10.8
50	20	11.1	3.86	0.259	11.1	28.8
0	50	0.00	3.86	0.259	0.00	0
5	50	2.56	3.86	0.259	2.56	6.64
10	50	5.26	3.86	0.259	5.26	13.6
20	50	11.1	3.86	0.259	11.1	28.8
50	50	33.3	3.86	0.259	33.3	86.4

$f$  = the operating frequency, and  
 $\omega = 2\pi f$  is the angular frequency.

Let us see how serious the additional distortion or jitter of a digital voltage would be. Assume that the data rate is 3.86 Gbits/s or that the principal operating frequency is 3.86 GHz. From the expressions (2.19)–(2.21), the additional distortion and additional jitter can be calculated and is shown in Table 2.1.

It can be seen that

- when

$$\Gamma_L < 10\% \text{ and } \Gamma_S < 10\%, \quad (2.22)$$

$$\Delta D |_{\%} = \Delta(\text{jitter}) |_{\%} < 1.01\%, \quad (2.23)$$

and

$$\Delta(\text{jitter}) |_s < 2.6 \text{ ps}. \quad (2.24)$$

In the cases of (2.22), the voltage reflection does not seem to be too harmful to the digital voltage if impedance matching is ignored, as is usually done. However,

- when

$$\Gamma_L > 10\% \text{ and } \Gamma_S > 10\%, \quad (2.25)$$

$$\Delta D|_{\%} = \Delta(\text{jitter})_{\%} > 1.01\%, \quad (2.26)$$

and

$$\Delta(\text{jitter})|_s > 2.6 \text{ ps}. \quad (2.27)$$

The voltage reflection in this case becomes quite serious.

- When

$$\Gamma_L = 50\% \text{ and } \Gamma_S = 50\%, \quad (2.28)$$

$$\Delta D|_{\%} = \Delta(\text{jitter})_{\%} = 33.3\%, \quad (2.29)$$

and

$$\Delta(\text{jitter})|_s = 86.4 \text{ ps}. \quad (2.30)$$

The voltage reflection is horrible!

Similarly, the additional interference can also be evaluated on the basis of the expressions for (2.17) and (2.18) when the frequency of the incoming voltage is not the same as that of the sequentially reflected voltages. However, we are not going to do so, and leave this analysis to the reader as an exercise.

## 2.3 POWER DELIVERED FROM A SOURCE TO A LOAD

### 2.3.1 General Expression of Power Delivered from a Source to a Load when $|l| \ll \lambda/4$ so that $T_d \rightarrow 0$

Figure 2.3 shows the power delivered from a source to a load. Similar to the discussion on the voltage delivered from a source to a load earlier, the precise phrase must be “power delivered from a source to the real part of a load,” which does not include the imaginary part of the load.

Let us assume that the delivered power suffers a reflection at the source and the load, that is,

$$\gamma_S \neq 0, \quad (2.31)$$

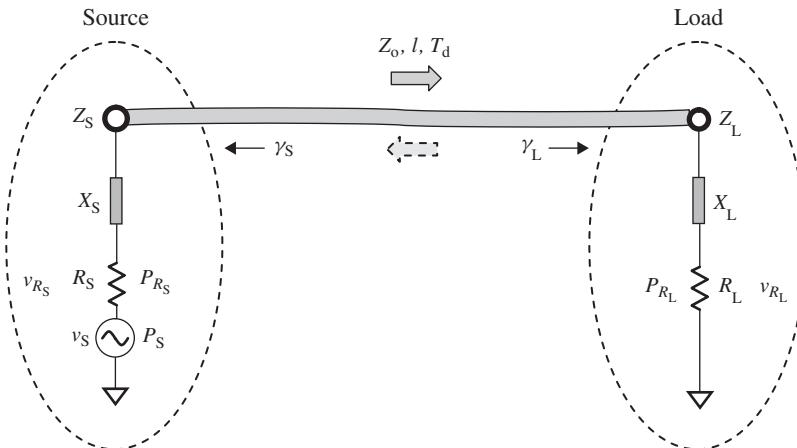
and

$$\gamma_L \neq 0, \quad (2.32)$$

where

$\gamma_S$  = the power reflection coefficient of source, and

$\gamma_L$  = the power reflection coefficient of load.



$Z_0$  = characteristic impedance of runner;

$l$  = length of runner, and  $l \ll \lambda / 4$ ;

$T_d$  = delayed time of signal transported from source to load.

Figure 2.3. Power delivered from a source to a load.

The relationships between the power reflection coefficients  $\gamma_S$  and  $\gamma_L$  at source and load and the voltage reflection coefficients  $\Gamma_S$  and  $\Gamma_L$  at source and load are

$$\gamma_S = \Gamma_S^2, \quad (2.33)$$

$$\gamma_L = \Gamma_L^2. \quad (2.34)$$

The impedances of source and load are not equal to the characteristic impedance of the runner as shown in expression (2.5). The delivered power from the source will bounce back and forth between the source and the load. The resultant power on  $R_L$  is the sum of all the remaining powers on  $R_L$  after each bounce back and forth of the voltage between the source and the load.

Also, let us assume that the source delivers a pure sinusoidal voltage,  $v_S = v_{So}e^{j\omega t}$  as shown in expression (2.7) and that the arrival time of the delivered voltage at the load is the delayed time  $T_d$ . The time reference point  $t = 0$  is defined as shown in expression (2.8) as the moment when the voltage is delivered from the source.

The general expression for the power remaining on  $R_L$  is the sum of all the remaining powers on  $R_L$  after the reflected voltage bounces back and forth between the source and the load, and can be expressed as follows:

$$P_{R_L}|_{t=T_d} = v_{So}^2 \frac{R_L}{|Z_S + Z_L|^2} (1 - \gamma_L) \sum_{n=0}^{\infty} e^{j2\omega[t-T_d-2nT_d]} (\gamma_S \gamma_L)^n, \quad (2.35)$$

or

$$P_{R_L}|_{t=T_d} = v_{So}^2 \frac{R_L}{|Z_S + Z_L|^2} (1 - \gamma_L) \sum_{n=0}^{\infty} e^{j2\omega[t-(2n+1)T_d]} (\gamma_S \gamma_L)^n, \quad (2.36)$$

where  $P_{R_L}|_{t=T_d}$  is the resultant remaining power.

The delayed time ( $t - T_d$ ) is explicitly kept in every exponential factor  $e^{j2\omega(t-T_d-2nT_d)}$ . The time interval  $2nT_d$  appearing in the exponential factor  $e^{j2\omega(t-T_d-2nT_d)}$  indicates the previous moment at which the power term in expression (2.35) or (2.36) was delivered from the source. It should be noted that all the power terms in expression (2.35) or (2.36) arrive at the load at the same time, that is,  $t = T_d$ .

The first term on the right-hand side of expression (2.35) or (2.36) represents the remaining power on  $R_L$  after the first reflection from the load with delayed time  $T_d$ .

The second term on the right-hand side of expression (2.35) or (2.36) represents the remaining power on  $R_L$  after the second reflection from the load with a delayed time  $3T_d$ . The additional factor  $(\gamma_S \gamma_L)$  represents the reflected percentage after the power bounces back and forth once between the source and the load.

The third term on the right-hand side of expression (2.35) or (2.36) represents the remaining power on  $R_L$  after the third reflection from the load with a delayed time  $5T_d$ . The additional factor  $(\gamma_S \gamma_L)^2$  represents reflected percentage after the power bounces back and forth twice between the source and the load, and so on.

Similar to Figure 2.2, Figure 2.4 vividly depicts the power arriving at the load simultaneously when  $t = T_d$  but delivered from the source at different moments when  $t = \dots -12T_d, -10T_d, -8T_d, -6T_d, -4T_d, -2T_d, 0, \dots$ . The number of power bounces arriving at the load is infinite, and it is therefore impossible to depict them totally; only eight are shown in Figure 2.4. They are sketched as follows:

- The source and load are located at top left and top right, respectively.
- The distance from the source to the load is  $l$ , and the corresponding delay time for the delivered power moving from the source to the load is  $T_d$ .
- The time axis is directed vertically downwards.
- The powers delivered from the source at different moments when  $t = \dots -12T_d, -10T_d, -8T_d, -6T_d, -4T_d, -2T_d, 0, \dots$  are marked in the left-hand side of Figure 2.4. In each remark box, the factor  $k_p$  denotes the portion of power at the source that is delivered from the source to  $R_L$ .
- Finally, all of the delivered power from the source at different moments arrive at the load simultaneously and are remarked with  $P_{R_L}|_{t=T_d}$  in the bottom-right corner of Figure 2.4.

Let us return to the expression (2.35):

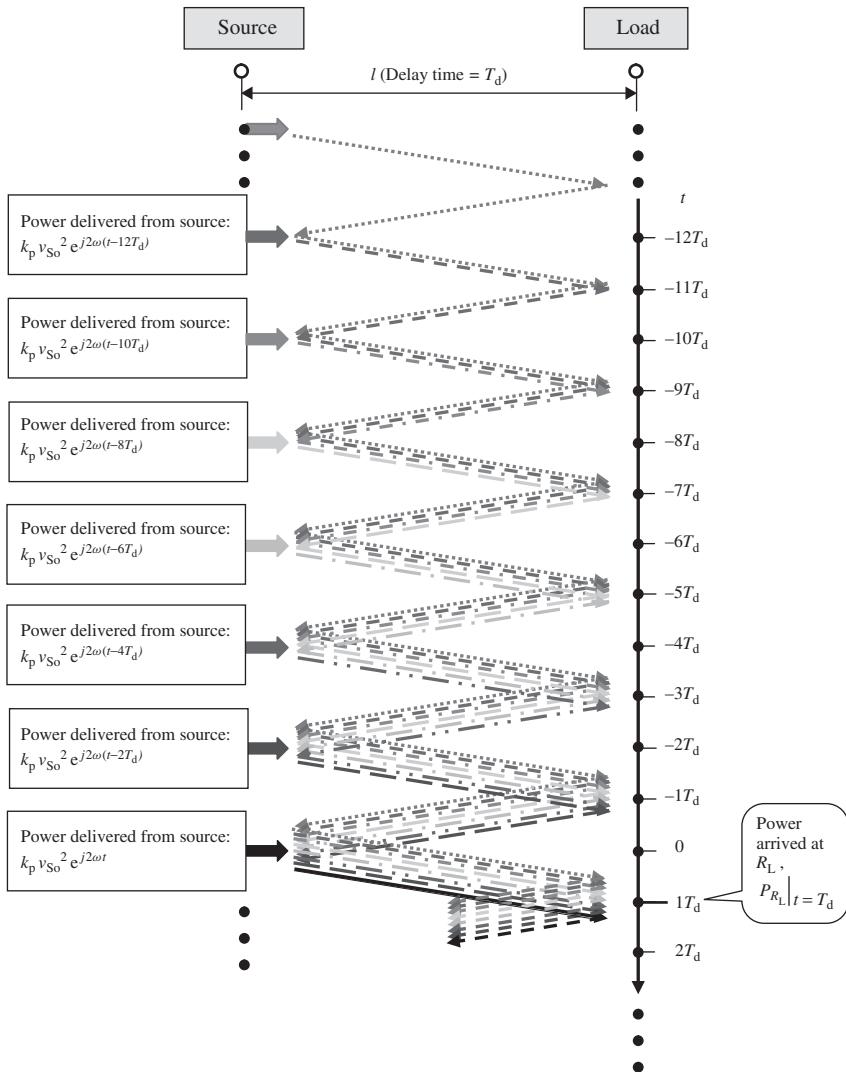
$$P_{R_L}|_{t=T_d} = v_{So}^2 \frac{R_L}{|Z_S + Z_L|^2} (1 - \gamma_L) \sum_{n=0}^{\infty} e^{j2\omega[t-T_d-2nT_d]} (\gamma_S \gamma_L)^n. \quad (2.37)$$

If

$$T_d \rightarrow 0, \quad (2.38)$$

then

$$P_{R_L}|_{t=T_d} = v_{So}^2 e^{j2\omega t} \frac{R_L}{|Z_S + Z_L|^2} \frac{1 - \gamma_L}{1 - \gamma_S \gamma_L}, \quad (2.39)$$



**Figure 2.4.** Power bouncing back and forth and arriving at load when  $t = T_d$ .

Note:  $k_p = R_L / |Z_S + Z_L|^2$ .

or

$$P_{R_L}|_{t=T_d} = v_S^2 \frac{R_L}{|Z_S + Z_L|^2} \frac{1 - \gamma_L}{1 - \gamma_S \gamma_L}. \quad (2.40)$$

### 2.3.2 Power Instability

In the case when

$$\gamma_L \neq 0, \quad (2.41)$$

and the condition (2.38) is not true, that is when power reflection at the load exists and the delay time is not negligible, the power delivered from the source to the load,  $P_{R_L}$ , is unstable. It can be found from expression (2.37) as long as  $T_d$  is not negligible. The power at the load varies from time to time even though the variation becomes smaller and smaller as the time elapses.

However, in most cases such a power instability can be only conceived in a very short period, as long as  $T_d$  is a tiny time slot. This is especially true in integrated circuits where the length of the runner from the source to the load is usually much less than the corresponding quarter wavelength. Consequently, such a power instability can be ignored, and thus, expression (2.38) is a good approximation indeed.

### 2.3.3 Additional Power Loss

The power reflection at the load looks like the power delivered from the source to the load after experiencing an additional power loss.

If there is no power reflection at the load, that is,

$$\gamma_L = 0, \quad (2.42)$$

then from expressions (2.39) and (2.40) we have

$$P_{R_L} \Big|_{t=T_d, \gamma_L=0} = v_{S0}^2 e^{j2\omega t} \frac{R_L}{|Z_S + Z_L|^2}, \quad (2.43)$$

or

$$P_{R_L} \Big|_{t=T_d, \gamma_L=0} = v_S^2 \frac{R_L}{|Z_S + Z_L|^2}. \quad (2.44)$$

The difference between (2.39) and (2.43) or (2.40) and (2.44) is the additional power loss, that is,

$$\Delta P_{R_L} \Big|_{t=T_d} = P_{R_L} \Big|_{t=T_d} - P_{R_L} \Big|_{t=T_d, \gamma_L=0} = -P_{R_L} \Big|_{t=T_d, \gamma_L=0} \frac{(1 - \gamma_S)\gamma_L}{1 - \gamma_S\gamma_L}. \quad (2.45)$$

The relative additional power loss is

$$\frac{\Delta P_{R_L} \Big|_{t=T_d}}{P_{R_L} \Big|_{t=T_d, \gamma_L=0}} = -\frac{(1 - \gamma_S)\gamma_L}{1 - \gamma_S\gamma_L}. \quad (2.46)$$

The additional power attenuation factor  $(1 - \gamma_S)\gamma_L / (1 - \gamma_S\gamma_L)$  appearing in expression (2.46) is always less than 1 if  $1 > \gamma_S > 0$  and  $1 > \gamma_L > 0$ .

As an example, Table 2.2 lists some calculated values of the additional power loss due to the unmatched cases in terms of expressions (2.45) and (2.46). The special features of the additional power loss are as follows:

TABLE 2.2. Additional Power Loss Due to the Unmatched Case when  $P_{R_L}|_{t=T_d, \gamma_L=0} = -30 \text{ dB}_m$

$\gamma_S, \%$	$\gamma_L, \%$	$\frac{(1-\gamma_S)\gamma_L}{(1-\gamma_S\gamma_L)}$	$P_{R_L} _{t=T_d, \gamma_L=0}, \text{dB}_m$	$\Delta P_{R_L} _{t=T_d}, \text{dB}_m$	$P_{R_L} _{t=T_d}, \text{dB}_m$
0	0	0.0000	-30	$-\infty$	-30
5	0	0.0000	-30	$-\infty$	-30
10	0	0.0000	-30	$-\infty$	-30
20	0	0.0000	-30	$-\infty$	-30
50	0	0.0000	-30	$-\infty$	-30
0	5	0.0500	-30	-43.01	-30.22
5	5	0.0476	-30	-43.22	-30.21
10	5	0.0452	-30	-43.45	-30.20
20	5	0.0404	-30	-43.94	-30.18
50	5	0.0256	-30	-45.91	-30.11
0	10	0.1000	-30	-40.00	-30.46
5	10	0.0955	-30	-40.20	-30.44
10	10	0.0909	-30	-40.41	-30.41
20	10	0.0816	-30	-40.88	-30.37
50	10	0.0526	-30	-42.79	-30.23
0	20	0.2000	-30	-36.99	-30.97
5	20	0.1919	-30	-37.17	-30.93
10	20	0.1837	-30	-37.36	-30.88
20	20	0.1667	-30	-37.78	-30.79
50	20	0.1111	-30	-39.54	-30.51
0	50	0.5000	-30	-33.01	-33.01
5	50	0.4872	-30	-33.12	-32.90
10	50	0.4737	-30	-33.25	-32.79
20	50	0.4444	-30	-33.52	-32.55
50	50	0.3333	-30	-34.77	-31.76

- When  $\gamma_L = 0$ , there is no additional power loss.
- For a given value of  $\gamma_L$ , the additional power loss is somewhat reduced as the value of  $\gamma_S$  is increased.
- For a given value of  $\gamma_S$ , the additional power loss is somewhat increased as the value of  $\gamma_L$  is increased.
- The value of the additional power loss in Table 2.2 is  $-3.01 \text{ dB}$  when  $\gamma_S = 0$  and  $\gamma_L = 50\%$ .

The additional power loss due to unmatched circuit design can seriously damage the performance of a communication or other system. For instance, a communication system with 64 QAMs (quadrant amplitude modulations) would require a power accuracy between the channels of less than 1 dB. The unmatched design for the RF circuit might be a serious killer in a dark corner.

### 2.3.4 Additional Distortion

Now let us discuss the general expression of the power transport (2.37) in the cases where the approximation (2.38) is satisfied. It is easy to understand that the sequentially

reflected powers between the source and the load are overlapped over the incoming power and hence disturb the incoming power at the load. It is therefore desirable to distinguish the incoming power from all the sequentially reflected powers.

The summed powers arriving at the load consist of two parts. One is the remaining power after first reflection. The second is the sequentially reflected powers between the source and the load that were delivered from the source at  $t = \dots -12T_d, -10T_d, -8T_d, -6T_d, -4T_d, -2T_d, 0$  as shown in Figure 2.4. The negative time signifies that the reflected powers happened before the moment  $t = 0$ . That is,

$$P_{R_L}|_{t=T_d} = P_{R_L}|_{t=T_d,1st} + P_{R_L}|_{t=T_d,srp} \quad (2.47)$$

where

- $P_{R_L}|_{t=T_d}$  = the summed power on load at the instant ( $t = T_d$ ),
- $P_{R_L}|_{t=T_d,1st}$  = the remaining power on the load after the first reflection, and
- $P_{R_L}|_{t=T_d,srp}$  = the sequentially reflected powers remaining on  $R_L$  after the second reflection.

Under the condition (2.38), the expression (2.37) can be rewritten as

$$P_{R_L}|_{t=T_d} = v_{So}^2 e^{j2\omega t} \frac{R_L}{|Z_S + Z_L|^2} (1 - \gamma_L) + v_{So}^2 e^{j2\omega t} \frac{R_L}{|Z_S + Z_L|^2} (1 - \gamma_L) \sum_{n=1}^{\infty} (\gamma_S \gamma_L)^n. \quad (2.47a)$$

The first term on the right-hand side of expression (2.47) is the remaining power on load after the first reflection. The second term on the right-hand side of expression (2.47) is the sequentially reflected powers remaining on  $R_L$  after the second reflection.

Note that

$$\sum_{n=1}^{\infty} (\gamma_S \gamma_L)^n = \sum_{n=0}^{\infty} (\gamma_S \gamma_L)^n - 1 = \frac{1}{1 - \gamma_S \gamma_L} - 1 = \frac{\gamma_S \gamma_L}{1 - \gamma_S \gamma_L}.$$

Then, (2.47a) becomes

$$P_{R_L}|_{t=T_d} = v_{So}^2 e^{j2\omega t} \frac{R_L}{|Z_S + Z_L|^2} (1 - \gamma_L) + v_{So}^2 e^{j2\omega t} \frac{R_L}{|Z_S + Z_L|^2} (1 - \gamma_L) \frac{\gamma_S \gamma_L}{1 - \gamma_S \gamma_L}. \quad (2.47b)$$

By comparing (2.47) and (2.47b), we have

$$P_{R_L}|_{t=T_d,1st} = v_{So}^2 e^{j2\omega t} \frac{R_L}{|Z_S + Z_L|^2} (1 - \gamma_L), \quad (2.48)$$

$$P_{R_L}|_{t=T_d,srp} = v_{So}^2 e^{j2\omega t} \frac{R_L}{|Z_S + Z_L|^2} (1 - \gamma_L) \frac{\gamma_S \gamma_L}{1 - \gamma_S \gamma_L}. \quad (2.49)$$

The sum of the sequentially reflected powers  $P_{R_L}|_{t=T_d,srp}$  disturbs the desired signal power  $P_{R_L}|_{t=T_d,1st}$ . The disturbance can be categorized into two types: additional distortion and

additional interference. The additional distortion appears when the frequency of the power is constant or when the sequentially reflected power has the same frequency as that of the incoming power. The additional interference appears when the frequency of the power is not constant or when the sequentially reflected power does not have the same frequency as that of the previous reflected power.

In the cases when the frequency of the power is constant or when the sequentially reflected power has the same frequency as that of the incoming power, the additional distortion can be evaluated as the ratio of the sum of sequentially reflected powers  $P_{R_L}|_{t=T_d, \text{srp}}$  to the desired signal power  $P_{R_L}|_{t=T_d, \text{1st}}$ . It should be noted that “the distortion of power” is meaningless, whereas it is the distortion of voltage that actually takes place in reality. The additional voltage distortion, therefore, is the square root of the ratio of  $P_{R_L}|_{t=T_d, \text{srp}}$  to  $P_{R_L}|_{t=T_d, \text{1st}}$  and can be calculated from (2.48) and (2.49), as

$$\Delta D_p|_{\%} = \sqrt{\frac{P_{R_L}|_{t=T_d, \text{srp}}}{P_{R_L}|_{t=T_d, \text{1st}}}} = \sqrt{\frac{\gamma_S \gamma_L}{1 - \gamma_S \gamma_L}}, \quad (2.50)$$

where  $\Delta D_p|_{\%}$  is the additional distortion in percentage.

TABLE 2.3. Additional Distortion in Power Transport

$\gamma_S, \%$	$\gamma_L, \%$	$\Delta D_p, \%$
0	0	0.00
5	0	0.00
10	0	0.00
20	0	0.00
50	0	0.00
0	5	0
5	5	5.01
10	5	7.09
20	5	10.05
50	5	16.01
0	10	0
5	10	7.09
10	10	10.05
20	10	14.29
50	10	22.94
0	20	0
5	20	10.05
10	20	14.29
20	20	20.41
50	20	33.33
0	50	0
5	50	16.01
10	50	22.94
20	50	33.33
50	50	57.74

Let us see how serious the additional distortion of a voltage would be. The additional distortion can be calculated, and some values are shown in Table 2.3. From the table, the following observations can be made:

- In the cases where  $\gamma_L = 0$ , there is no additional distortion. On the contrary, in the cases where  $\gamma_L \neq 0$ , the additional distortion is appreciable.
- The additional distortion is more sensitive to the value of  $\gamma_L$  than to the value of  $\gamma_S$ .
- For a given value of  $\gamma_L$ , the additional distortion is somewhat reduced as the value of  $\gamma_S$  is increased.
- For a given value of  $\gamma_S$ , the additional distortion is somewhat increased as the value of  $\gamma_L$  is increased.
- The highest value of the additional distortion in Table 2.3 is 70.71% when  $\gamma_S = 0$  and  $\gamma_L = 50\%$ .

### 2.3.5 Additional Interference

As mentioned earlier, the sequentially reflected powers disturb the desired signal power on the load. The additional interference appears when the frequency of the power is not constant or when the sequentially reflected power does not have the same frequency as that of the previous reflected power. Now let us recall expressions (2.48) and (2.49):

$$P_{R_L}|_{t=T_d,1st} = v_{So}^2 e^{j2\omega t} \frac{R_L}{|Z_S + Z_L|^2} (1 - \gamma_L), \quad (2.48)$$

$$P_{R_L}|_{t=T_d,srp} = v_{So}^2 e^{j2\omega t} \frac{R_L}{|Z_S + Z_L|^2} (1 - \gamma_L) \frac{\gamma_S \gamma_L}{1 - \gamma_S \gamma_L}. \quad (2.49)$$

The additional interference can be evaluated as the ratio of the sequentially reflected powers to the desired signal power. They are

$$S_{R_L}|_{t=T_d} = P_{R_L}|_{t=T_d,1st} = v_{So}^2 e^{j2\omega t} \frac{R_L}{|Z_S + Z_L|^2} (1 - \gamma_L), \quad (2.51)$$

$$\Delta I_{R_L}|_{t=T_d} = P_{R_L}|_{t=T_d,srp} = V_{So}^2 e^{j2\omega_2(t-T_d)} \frac{R_L}{|Z_S + Z_L|^2} (1 - \gamma_L) \frac{\gamma_S \gamma_L}{1 - \gamma_S \gamma_L}, \quad (2.52)$$

where

$S_{R_L}|_{t=T_d}$  = the desired signal power arriving on the load at the instant  $t = T_d$ ,

and

$\Delta I_{R_L}|_{t=T_d}$  = the additional interference power on the load at the instant  $t = T_d$ .

From (2.51) and (2.52) we have

$$\frac{\Delta I_{R_L}|_{t=T_d}}{S_{R_L}|_{t=T_d}} = \frac{\gamma_S \gamma_L}{1 - \gamma_S \gamma_L}. \quad (2.53)$$

The original SIR (signal-to-interference ratio)  $SIR_{R_{Lo}}$  without additional interference is

$$SIR_{R_{Lo}}|_{t=T_d} = \frac{S_{R_L}|_{t=T_d}}{I_{R_L}|_{t=T_d}}. \quad (2.54)$$

where  $I_{R_L}|_{t=T_d}$  = the the interference power on the load at the instant  $t = T_d$ . In addition, the subscript “ $t = T_d$ ” in equation (2.54) can be neglected.

From expressions (2.53) and (2.54), the SIR becomes

$$\text{SIR}_{R_L} = \frac{S_{R_L}}{I_{R_L} + \Delta I_{R_L}} = \frac{\text{SIR}_{R_{L0}}}{1 + \frac{\Delta I_{R_L}}{I_{R_L}}} = \frac{\text{SIR}_{R_{L0}}}{1 + \frac{\gamma_S \gamma_L}{1 - \gamma_S \gamma_L} \text{SIR}_{R_{L0}}}. \quad (2.55)$$

Finally,

$$\text{SIR}_{R_L} = \frac{\text{SIR}_{R_{L0}}}{1 + \frac{\gamma_S \gamma_L}{1 - \gamma_S \gamma_L} \text{SIR}_{R_{L0}}}, \quad (2.56)$$

where  $\text{SIR}_{R_L}$  is the resulting SIR with additional interference due to unmatched impedance.

Table 2.4 lists some calculated ratios of signal to interference due to the unmatched impedance. From Table 2.4, that the following observations can be made:

- In the cases where  $\gamma_L = 0$ , there is no additional interference, so the SIR remains unchanged.

TABLE 2.4. Calculated Ratio of Signal to Interference as the Reflection Coefficient  $\gamma$  is Varied

$\gamma_S$ , %	$\gamma_L$ , %	$\text{SIR}_{R_{L0}}$ , dB	$\text{SIR}_{R_{L0}}$ , W	$\gamma_S \gamma_L$ , %	$\gamma_S \gamma_L / (1 - \gamma_S \gamma_L)$ , W	$1 + \frac{\gamma_S \gamma_L}{1 - \gamma_S \gamma_L}$	$\text{SIR}_{R_L}$ , dB
0	0	15	31.62	0.0000	0.0000	1.00	15.00
5	0	15	31.62	0.0000	0.0000	1.00	15.00
10	0	15	31.62	0.0000	0.0000	1.00	15.00
20	0	15	31.62	0.0000	0.0000	1.00	15.00
50	0	15	31.62	0.0000	0.0000	1.00	15.00
0	5	15	31.62	0.0000	0.0000	1.00	15.00
5	5	15	31.62	0.0025	0.0025	1.08	14.67
10	5	15	31.62	0.0050	0.0050	1.16	14.36
20	5	15	31.62	0.0100	0.0101	1.32	13.80
50	5	15	31.62	0.0250	0.0256	1.81	12.42
0	10	15	31.62	0.0000	0.0000	1.00	15.00
5	10	15	31.62	0.0050	0.0050	1.16	14.36
10	10	15	31.62	0.0100	0.0101	1.32	13.80
20	10	15	31.62	0.0200	0.0204	1.65	12.84
50	10	15	31.62	0.0500	0.0526	2.66	10.74
0	20	15	31.62	0.0000	0.0000	1.00	15.00
5	20	15	31.62	0.0100	0.0101	1.32	13.80
10	20	15	31.62	0.0200	0.0204	1.65	12.84
20	20	15	31.62	0.0400	0.0417	2.32	11.35
50	20	15	31.62	0.1000	0.1111	4.51	8.45
0	50	15	31.62	0.0000	0.0000	1.00	15.00
5	50	15	31.62	0.0250	0.0256	1.81	12.42
10	50	15	31.62	0.0500	0.0526	2.66	10.74
20	50	15	31.62	0.1000	0.1111	4.51	8.45
50	50	15	31.62	0.2500	0.3333	11.54	4.38

- The additional interference is more sensitive to the value of  $\gamma_L$  than to the value of  $\gamma_S$ .
- For a given value of  $\gamma_L$ , the additional interference increases as the value of  $\gamma_S$  increases so that the SIR is reduced.
- For a given value of  $\gamma_S$ , the additional interference increases as the value of  $\gamma_L$  increases so that the SIR is reduced.
- The highest value of the additional interference in Table 2.4 is reached when  $\gamma_S = 50\%$  and  $\gamma_L = 50\%$ . At this point, the SIR drops from 15 dB to 4.38 dB.

## 2.4 IMPEDANCE CONJUGATE MATCHING

From the discussions in Sections 2.1 and 2.2, it can be seen that the voltage or power reflection is very harmful to the performance of a digital or RF circuit block. Both voltage and power reflections are due to the unmatched impedance condition existing in the source or the load. Impedance matching, therefore, becomes a key subject in all the circuit designs, including RF, digital, and analog.

### 2.4.1 Maximizing Power Transport

First, let us discuss the case of power delivered from a source to a load without reflection, that is, as shown in Figure 2.5,

$$\gamma_S = 0, \quad (2.57)$$

$$\gamma_L = 0. \quad (2.58)$$

The voltage and power delivered from source to the real part  $R_L$  of the load can be expressed by

$$v_{R_L}(t) = \frac{v_S}{Z_S + Z_L} R_L, \quad (2.59)$$

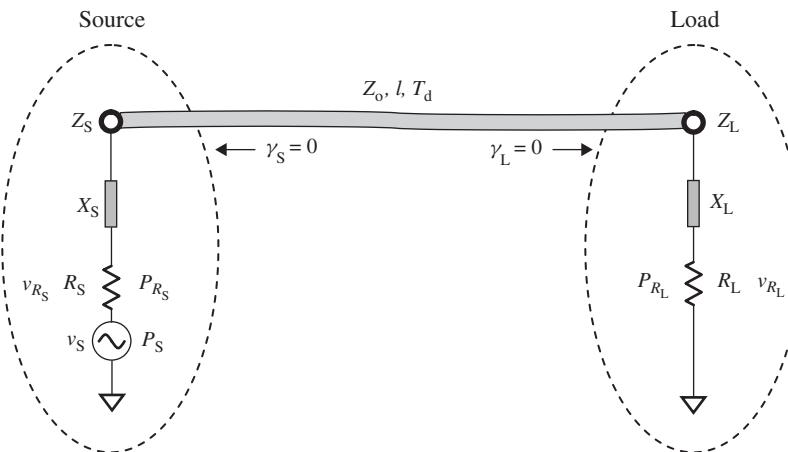
$$P_{R_L}(t) = v_S^2 \frac{R_L}{|Z_S + Z_L|^2} = v_S^2 \frac{R_L}{(R_S + R_L)^2 + (X_S + X_L)^2}. \quad (2.60)$$

For power transport, it is desirable to achieve maximum transfer of power from the source to the load resistor. From expression (2.60), it is easy to see that the maximum will be reached if the denominator in the expression (2.60) is a minimum. The two terms in the denominator cannot be negative but must be positive or zero because they are squares of real values. Theoretically, the smallest possible value is zero. The first term in the denominator is a term of the resistors' value, which cannot be zero. However, the second term in the denominator can be forced to become zero, that is,

$$X_S + X_L = 0,$$

or

$$X_S = -X_L. \quad (2.61)$$



**Figure 2.5.** Power delivered from a source to a load without reflection when  $\gamma_S = 0$ ,  $\gamma_L = 0$  and when  $l \ll \lambda/4$  so that  $T_d \rightarrow 0$ .

The relation (2.61) indicates that the reactance of the source and the load must have equal magnitudes but opposite signs. It implies that the load reactance  $X_L$  must be inductive if the source reactance  $X_S$  is capacitive, and vice versa. With the relation (2.61), equation (2.59) becomes

$$P_{R_L} = v_S^2 \frac{R_L}{(R_S + R_L)^2}. \quad (2.62)$$

Now, another relation to achieve maximum power transport from the source to the load resistor can be found from equation (2.61). Mathematically, by partially differentiating (2.62) with respect to  $R_L$ ,

$$\frac{\partial P_{R_L}}{\partial R_L} = v_S^2 \left[ \frac{1}{(R_S + R_L)^2} - 2 \frac{R_L}{(R_S + R_L)^3} \right] = v_S^2 \frac{R_S - R_L}{(R_S + R_L)^3}. \quad (2.63)$$

The condition to maximize  $P_{R_L}$  is

$$\frac{\partial P_{R_L}}{\partial R_L} = 0. \quad (2.64)$$

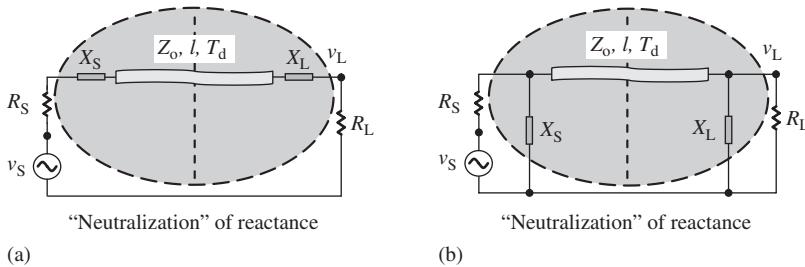
From relations (2.63) and (2.64), we have

$$R_S = R_L. \quad (2.65)$$

By combining of the relations (2.61) and (2.65), we obtain

$$Z_S^* = Z_L, \text{ or } Z_S = Z_L^*. \quad (2.66)$$

The relation (2.66) is the called the *condition of impedance conjugate matching* or simply the *condition of impedance matching*.



**Figure 2.6.** Two matching cases when the reactance of source is “neutralized” by reactance of load, or vice versa, that is,  $X_S = -X_L$ . (a)  $R_S$  in series with  $X_S$ ;  $R_L$  in series with  $X_L$ . (b)  $R_S$  in parallel with  $X_S$ ;  $R_L$  in parallel with  $X_L$ . Note: When  $l << \lambda/4$ ,  $T_d \rightarrow 0$ .

### 2.4.2 Power Transport without Phase Shift

Condition (2.61) indicates that, when the reactance of the source and the load,  $X_S$  and  $X_L$ , respectively, are neutralized by each other, then the schematic shown in Figure 2.5 can be redrawn as that shown in Figure 2.6. In the entire loop from the source to the load, there are only resistive parts and no parts with reactance. This implies that there is no phase shift when the voltage or power is transported from the source to the load, that is,

$$\angle v_L - \angle v_S = 0, \text{ or } \angle v_L = \angle v_S, \quad (2.67)$$

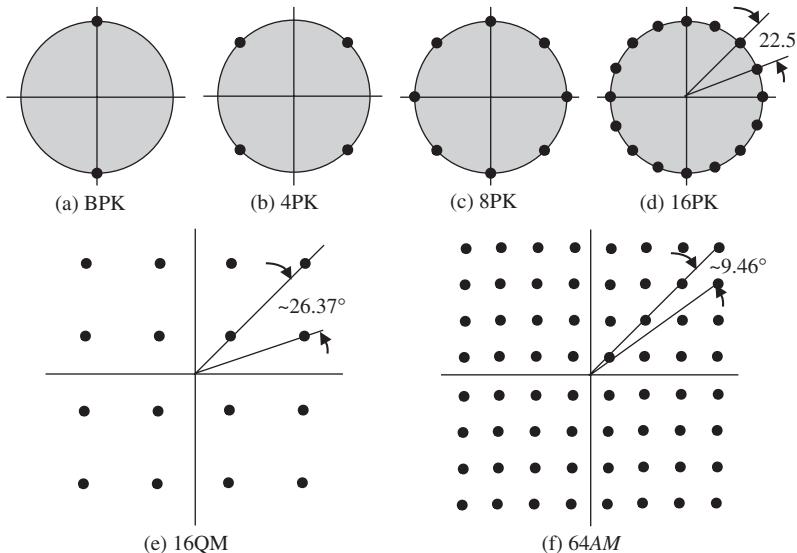
where the symbol “ $\angle x$ ” denotes the angle  $x$ .

To summarize, the significance of impedance conjugate matching is twofold:

1. To maximize the power transportation from source to load;
2. To transport the voltage from the source to the load without phase shift.

The second point, that is, voltage or power transport without phase shift, is another important feature of impedance conjugate matching. Unfortunately, it is very often ignored or mysteriously “swallowed” in the discussion on impedance conjugate matching.

In the past few decades, modulation technology has been developing in great leaps, and phase modulation plays an important role in the effective utilization of the frequency bandwidth. Figure 2.7 illustrates the progress in modulation technology in a communication system. Historically, PSK (phase-shift keying) technology has been developed from BPSK (Bi-phase shift keying) to QPSK (Quad-phase shift keying), 8PSK, 16PSK, etc. In a 16PSK system, as shown in Figure 2.7(d), the minimum phase difference between two adjacent symbols is only  $22.5^\circ$ . This is difficult for the circuit designer. Instead of the PSK system, the QAM system enhances the resolution to recognize two adjacent symbols considerably. In a QAM system, both the amplitude and the phase are applied for allocation of a symbol so that the system recognizes the symbol not only by its phase but also by its amplitude. It releases the phase difference between two adjacent symbols. In a 16QAM system, the minimum phase difference between two adjacent symbols is  $26.37^\circ$  as shown in Figure 2.7(e) which is greater than  $22.5^\circ$  in a 16PSK system as shown in Figure 2.7(d). In order to further enhance the utilization of bandwidth, the 64QAM system was developed after the 16QAM system. It, of course, is a more sophisticated



**Figure 2.7.** Progress of modulation technology from PSK to QAM. (a) BPSK, (b) 4PSK, (c) 8PSK, (d) 16PSK, (e) 16QAM, and (f) 64QAM.

system, but it brings about a  $9.46^\circ$  minimum phase difference between two adjacent symbols, which is a big challenge to the RF circuit designers.

Now, let us return to our topic of impedance conjugate matching.

It is clear from the above discussion that a phase shift exists in the loop from the source to the load if the impedance of the source and the load do not match well. If the phase shift is unstable and variable, it makes the designed RF block or system inoperable. As mentioned earlier, the minimum phase difference in a 64QAM system is only  $9.46^\circ$ , which is the upper limit of phase shift tolerance for the entire system. For an individual RF block, the tolerance of the phase shift variation must be much less than  $9.46^\circ$ . It can therefore be concluded that impedance matching is absolutely necessary in RF circuit designs.

Under the condition of impedance conjugate matching (2.66), the maximum power at the load can be obtained. By substituting equation (2.65) into (2.59) and (2.60) we have

$$v_{R_L} \Big|_{Z_S=Z_L^*} = v_{R_{L,\max}} = \frac{v_S}{2}, \quad (2.68)$$

$$P_{R_L} \Big|_{Z_S=Z_L^*} = P_{R_{L,\max}} = \frac{v_S^2}{4R_L}. \quad (2.69)$$

Similar to expressions (2.59) and (2.60), the voltage  $v_{R_S}$  and its corresponding power  $P_{R_S}$  on the source resistor  $R_S$  can be expressed as follows:

$$v_{R_S} = \frac{v_S}{Z_S + Z_L} R_S, \quad (2.70)$$

$$P_{R_S}(t) = v_S^2 \frac{R_S}{|Z_S + Z_L|^2} = v_S^2 \frac{R_S}{(R_S + R_L)^2 + (X_S + X_L)^2}. \quad (2.71)$$

By substituting the impedance conjugate matching condition (2.66) into (2.70) and (2.71), we have

$$v_{R_S} \Big|_{Z_S=Z_L^*} = v_{R_S,\max} = \frac{v_S}{2}, \quad (2.72)$$

$$P_{R_S} \Big|_{Z_S=Z_L^*} = P_{R_S,\max} = \frac{v_S^2}{4R_S}. \quad (2.73)$$

From equations (2.67), (2.68), (2.71), and (2.72), we can come to the following conclusions under the condition of impedance conjugate matching (2.66):

- the power delivered to the load reaches a maximum;
- the power delivered to the load is equal to the power remaining on the source.

The corresponding equations are

$$v_{R_L,\max} = v_{R_S,\max} = \frac{v_S}{2}, \quad (2.74)$$

$$P_{R_L,\max} = P_{R_S,\max} = \frac{v_S^2}{4R_L} = \frac{v_S^2}{4R_S}. \quad (2.75)$$

The total available power  $P_S$  in the loop from the source to the load is the sum of the powers on  $R_S$  and  $R_L$ , that is,

$$P_S = P_{R_S} + P_{R_L}. \quad (2.76)$$

Then,

$$P_{R_L} = P_{R_S} = \frac{1}{2}P_S. \quad (2.77)$$

The total power is divided into two. One half of the total power remains on  $R_S$ , and the other half is delivered to the load  $R_L$ .

### 2.4.3 Impedance Matching Network

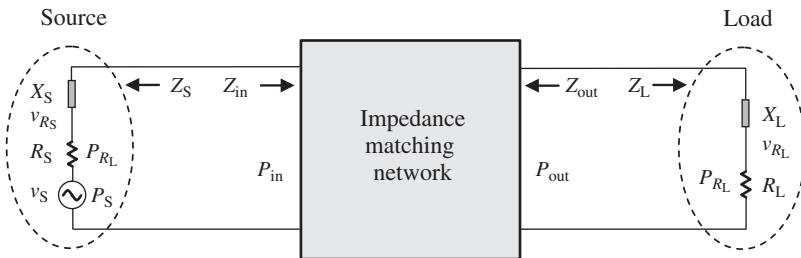
Usually, the impedance of a source does not conjugate-match with the impedance of the load, that is,

$$Z_S \neq Z_L^*. \quad (2.78)$$

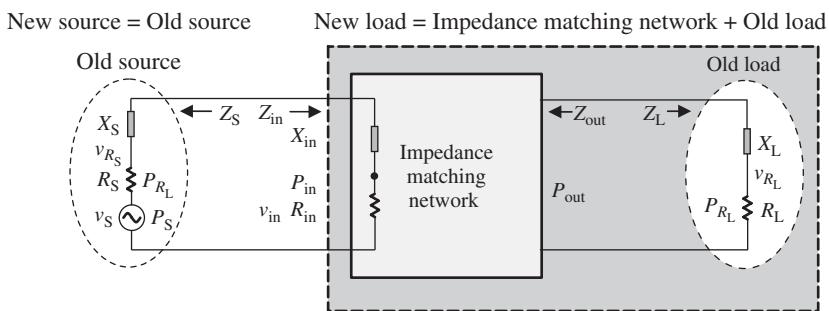
In order to maximize power transport without phase shift from the source to the load, the impedance conjugate matching condition must be satisfied. Therefore, an impedance matching network must be inserted between the source and the load. As shown in Figure 2.8, the input impedance of the impedance matching network must be equal to  $Z_S^*$  and the output impedance of the impedance matching network must be equal to  $Z_L^*$ , that is,

$$Z_{in} = Z_S^*, \quad (2.79)$$

$$Z_{out} = Z_L^*. \quad (2.80)$$



**Figure 2.8.** An impedance matching network inserted between the source and the load when  $Z_S \neq Z_L^*$ .



**Figure 2.9.** First sub-impedance matching loop: new source = old source; new load = impedance matching network + old load.

As a matter of fact, the entire system shown in Figure 2.8 can be recognized as three sub-impedance matching loops.

The first sub-impedance matching loop is shown in Figure 2.9.

The new source is the same as the old source shown in Figure 2.8. The new load consists of the impedance matching network and the old load shown in Figure 2.8. As shown in expression (2.79), the impedance of the new source is conjugate-matched with that of new load. Consequently, as shown in expression (2.75), the power delivered to  $R_{in}$  is equal to the power remaining on the  $R_S$ , that is,

$$P_{in} = \frac{v_{in}^2}{R_{in}} = \frac{v_S^2}{4R_{in}} = \frac{v_S^2}{4R_S} = \frac{v_{R_S}^2}{R_S} = P_{R_S}, \quad (2.81)$$

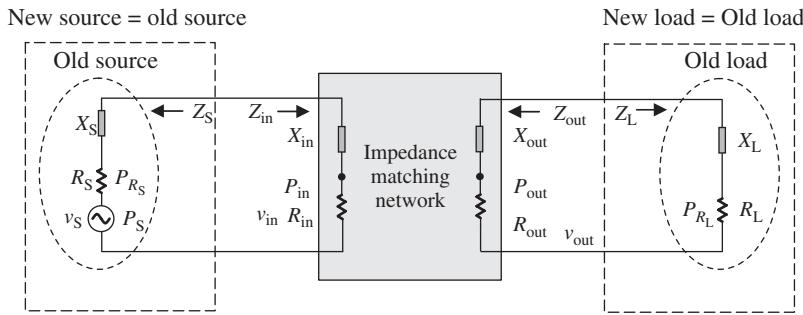
where

$P_{in}$  = the virtual or equivalent power resulting from the combination of the load and the impedance matching network, and

$R_{in}$  = the virtual or equivalent resistance resulting from the combination of the load and the impedance matching network.

The second sub-impedance matching loop is the impedance matching network itself, which is shown in Figure 2.10.

Theoretically, an impedance matching network could be constructed by either passive or active parts. The emitter follower, the source follower, and the buffer are examples of active impedance matching networks. However, implementation of an impedance matching network by active parts would increase noise and cost, degrade the linearity, and



**Figure 2.10.** Second sub-impedance matching loop: the impedance matching network itself.

expend more current. The LC impedance matching networks are passive ones and are used in most networks. In the implementation of a passive impedance matching network, the resistor is not a good choice because it causes noise and reduces power gain. Consequently, an impedance matching network usually consists only of capacitors and inductors.

There would be no power consumption in the impedance matching network if it consisted only of ideal inductors and capacitors. The output power on  $R_{\text{out}}$  will be equal to the input power on  $R_{\text{in}}$ .

$$P_{\text{in}} = \frac{v_{\text{in}}^2}{R_{\text{in}}} = \frac{v_{\text{out}}^2}{R_{\text{out}}} = P_{\text{out}}, \quad (2.82)$$

where

$P_{\text{in}}$  = the virtual or equivalent power resulting from the combination of the load and the impedance matching network,

$R_{\text{in}}$  = the virtual or equivalent resistance resulting from the combination of the load and the impedance matching network.

$P_{\text{out}}$  = the virtual or equivalent power resulting from the combination of the source and the impedance matching network, and

$R_{\text{out}}$  = the virtual or equivalent resistance resulting from the combination of the source and the impedance matching network.

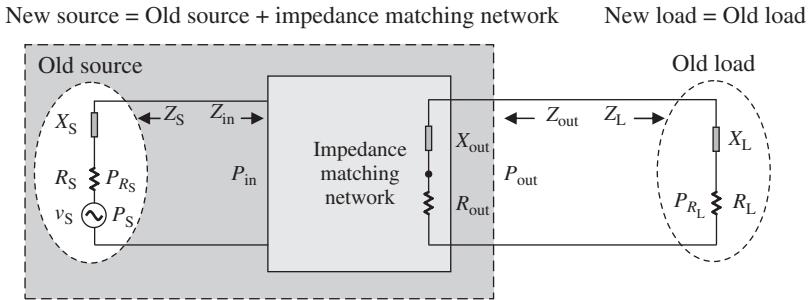
The third sub-impedance matching loop is shown in Figure 2.11.

The new source consists of the old source, as shown in Figure 2.8, and the impedance matching network. The new load is the same as the old load, as shown in Figure 2.8. As shown in expression (2.80), the impedance of the new source is conjugate-matched with the impedance of the new load. As shown in the expressions (2.74) and (2.75), the power delivered to  $R_L$  is equal to power remaining on the  $R_{\text{out}}$ , that is,

$$P_{R_L} = \frac{v_{R_L}^2}{R_L} = \frac{v_{\text{out}}^2}{R_{\text{out}}} = P_{\text{out}}. \quad (2.83)$$

Now, from equation (2.80)–(2.82) and (2.76), we have

$$P_{R_S} = P_{\text{in}} = P_{\text{out}} = P_{R_L} = \frac{1}{2}P_S. \quad (2.84)$$



**Figure 2.11.** Third sub-impedance matching loop: new source = old source + impedance matching network; new load = old load.

It should be pointed out that the simple expression (2.84) implies two important concepts:

1. The powers  $P_{R_S}$ ,  $P_{in}$ ,  $P_{out}$ , and  $P_{R_L}$  are equal. This is the basic theoretical background in practical power measurement because the power  $P_{R_S}$  at the source can be measured as its equal value  $P_{R_L}$  at the load if an impedance matching network is inserted between the source and the load.
2. The insertion of an impedance matching network without power consumption will ensure maximum power transport from the source to the load.

#### 2.4.4 Necessity of Impedance Matching

So far, we have emphasized the importance of impedance matching in RF circuit design. This is because the main task in RF circuit design is power transport or power manipulation. To maximize power transport without phase shift, impedance matching is an indispensable technology. For most RF circuit blocks, the input impedance must be matched with the output impedance of the preceding stage, and its output impedance must be matched with the input impedance of the next stage.

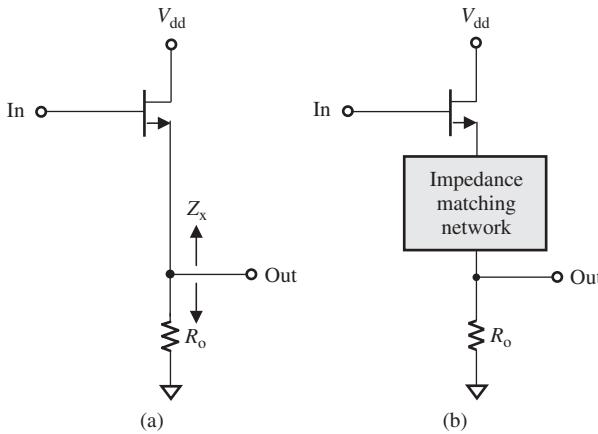
Nobody would doubt the importance of impedance matching in an RF or RFIC design. However, there are paradoxes about impedance matching. For example,

- Is it necessary to match the impedance from part to part?
- Within the impedance matching network, the impedance between parts is generally not matched. Is it necessary to insert a “sub-impedance matching network” between the two parts in the network itself?
- Is it necessary impedance-match all the RF blocks?

and so on.

To answer these questions, we should return to the goals of impedance matching: They are (i) maximizing power transport and (ii) eliminating phase shifts in power transport. Impedance matching is therefore required between power transport units but is not necessary between the individual parts.

If an impedance matching network consists of more than one part, it implies that all the parts in the network must cooperate so as to maximize power transport and to eliminate phase shift. In other words, the matching network, not the individual part, is



**Figure 2.12.** A source follower (DC bias is neglected). (a) Primary source follower. (b) “Improved” source follower.

an indivisible basic entity or a “minimum cell.” Their parts are indivisible in power transport. Impedance matching between the individual parts is meaningless.

In general, impedance must be matched between two basic power transport units. There are two exceptional cases in which we do not do so; they are the following:

1. The two basic power transport units are combined together for special performance.
2. Impedance is impossible to be matched or must not be matched from the output of a VCO (voltage-controlled oscillator) or oscillator to the input of next circuit block (this will be explained in Chapter 20).

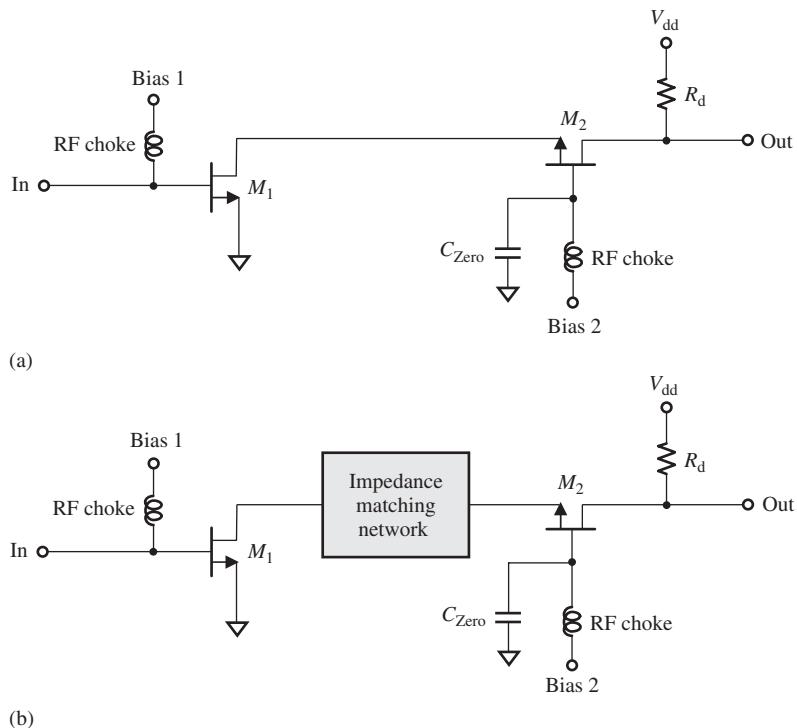
Let us illustrate this by examples.

**Example 2.1:** Figure 2.12(a) shows a raw source follower without impedance matching at its output (DC bias is neglected). It consists of a MOSFET (metal–oxide–semiconductor field-effect transistor) and an output resistor  $R_o$ . Looking at the emitter of the transistor in Figure 2.12(a), it can be seen that its impedance  $Z_x$  is usually not conjugate-matched to the impedance of the Remitter resistor,  $R_o$ . Is it beneficial for power transport if an impedance matching network is inserted between the source of the MOSFET and the output resistor  $R_o$ , as shown in Figure 2.12(b).

The answer is No, because an individual MOSFET or output resistor is not a basic or independent power transport unit. Both combined together form a basic power transport unit and are indivisible.

**Example 2.2:** Figure 2.13 shows a CS–CG cascade LNA (low-noise amplifier). Is it beneficial to the power transportation if a matching network is inserted between the collector of the CS transistor and the emitter of CG transistor?

In Figure 2.13(a), the first power transport unit consists of the CS transistor and the parts connected to its source and gate, and the second power transport unit consists of the CG transistor and the parts connected to its drain. These are independent units in the power transport. If an impedance matching network is inserted between the collector of the CS transistor and the emitter of the CG transistor as shown in Figure 2.13(b), it might be beneficial to maximize the power transport. However, the CS–CG cascade configuration will be seriously disturbed by the insertion of the impedance matching



**Figure 2.13.** A MOSFET cascade amplifier. (a) A primary MOSFET cascade amplifier. (b) An impedance matching network inserted between the two stages.

network: that is, its special feature, which is minimization of the input Miller capacitance, is removed or greatly degraded. It is advisable not to insert such an impedance matching network between the collector of the CS transistor and the emitter of the CG transistor.

## 2.5 ADDITIONAL EFFECT OF IMPEDANCE MATCHING

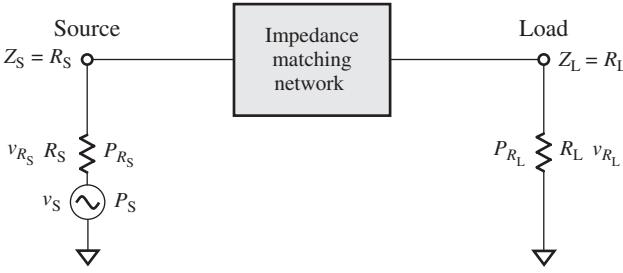
### 2.5.1 Voltage Pumped Up by Means of Impedance Matching

The voltage and power delivered from a source to a load was discussed in Sections 2.1 and 2.2, respectively. In this subsection, we discuss the cross-relationship between voltage and power transport, that is, voltage delivered from a source to a load through an impedance matching network.

Figure 2.14 shows a voltage delivered from a source to a load with an impedance matching network. For simplicity, the reactance of the source or load is assumed to be zero.

On the basis of equations (2.80), (2.82), and (2.83), the power delivered from the source to  $R_L$  is equal to half the total power  $P_S$  in the loop, while the other half remains at  $R_S$ , that is,

$$P_{R_L} = \frac{v_{R_L}^2}{R_L} = \frac{v_{R_S}^2}{R_S} = P_{R_S} = \frac{1}{2}P_S. \quad (2.85)$$



**Figure 2.14.** Voltage delivered from a source to a load through an impedance matching network.

**TABLE 2.5.** Relationship between  $v_{R_S}$ ,  $R_S$ , and  $v_{R_L}$ ,  $R_L$  for a Given Value of  $P_S = 0 \text{ dB}_m$  in an Impedance Matching Loop as shown in Figure 2.14

$P_S, \text{dB}_m$	$v_S, \text{V}$	$P_{R_S}, \text{dB}_m$	$R_S, \Omega$	$v_{R_S}, \text{V}$	$P_{R_L}, \text{dB}$	$R_L$	$v_{R_L}, \text{V}$
0	0.2236	-3	50	0.1583	-3	10 $\Omega$	0.0708
0	0.2236	-3	50	0.1583	-3	12 $\Omega$	0.0776
0	0.2236	-3	50	0.1583	-3	16 $\Omega$	0.0895
0	0.2236	-3	50	0.1583	-3	16 $\Omega$	0.1050
0	0.2236	-3	50	0.1583	-3	30 $\Omega$	0.1226
0	0.2236	-3	50	0.1583	-3	50 $\Omega$	0.1583
0	0.2236	-3	50	0.1583	-3	100 $\Omega$	0.2239
0	0.2236	-3	50	0.1583	-3	200 $\Omega$	0.3166
0	0.2236	-3	50	0.1583	-3	500 $\Omega$	0.5006
0	0.2236	-3	50	0.1583	-3	1 k $\Omega$	0.7079
0	0.2236	-3	50	0.1583	-3	2 k $\Omega$	1.0012
0	0.2236	-3	50	0.1583	-3	10 k $\Omega$	2.2387

Table 2.5 lists the correlated values of  $v_{R_S}$  and  $R_S$ ,  $v_{R_L}$  and  $R_L$ , calculated from equation (2.84) when impedance at source or load is well matched and when  $P_S = 0 \text{ dB}_m$ , which corresponds to a voltage  $v_S = 0.2236 \text{ V}$  across a  $50\text{-}\Omega$  resistor. In Table 2.5,  $R_S$  is kept  $50 \Omega$  unchanged while  $R_L$  is varied from  $10 \Omega$  to  $10 \text{ k}\Omega$ .

Three cases are shown in Table 2.5:

1. *Stepped-down case.*  $v_{R_L}$  is stepped down from  $v_{R_S}$ :

$$v_{R_L} < v_{R_S} \quad (2.86)$$

if

$$R_L < R_S. \quad (2.87)$$

2. *Unchanged case.*  $v_{R_L}$  is equal to  $v_{R_S}$ :

$$v_{R_L} = v_{R_S} = 0.1583 \text{ V} \quad (2.88)$$

if

$$R_L = R_S = 50 \Omega. \quad (2.89)$$

3. *Pumped-up case.*  $v_{R_L}$  is pumped up from  $v_{R_S}$ :

$$v_{R_L} > v_{R_S}, \quad (2.90)$$

if

$$R_L > R_S. \quad (2.91)$$

As shown in the last row in Table 2.5, when  $R_S = 50 \Omega$  and  $R_L = 10 \text{ k}\Omega$ , the value of  $v_{R_L}$  is equal to 2.2387 V, which is pumped up by more than 14 times from  $v_{R_S} = 0.1583 \text{ V}$ .

Now let us discuss another case when a voltage is directly delivered from the source to the load without an impedance matching network.

Figure 2.15 depicts such a simple loop with  $P_S = 0 \text{ dB}_m$ , which corresponds to a voltage  $v_{R_L} = 0.2236 \text{ V}$  across a 50- $\Omega$  resistor. The various voltages can be calculated by the following equations:

$$P_S = \frac{v_S^2}{50} \text{ or } v_S = \sqrt{P_S \times 50}, \quad (2.92)$$

$$v_{R_S} = \frac{R_S}{R_S + R_L} v_S, \quad (2.93)$$

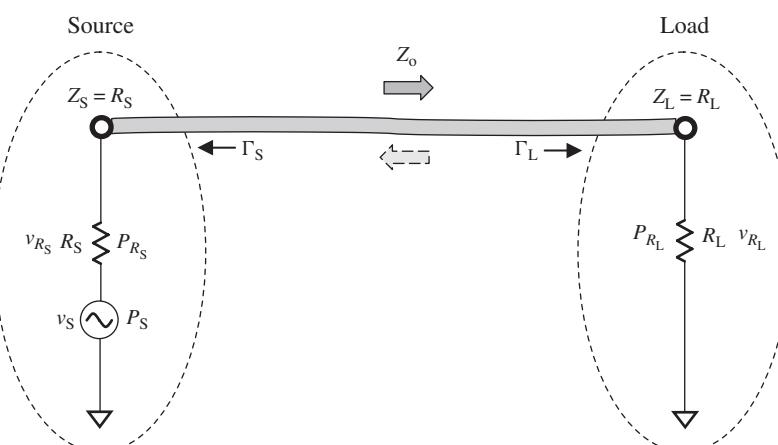
$$v_{R_L} = \frac{R_L}{R_S + R_L} v_S. \quad (2.94)$$

The calculated values of  $v_{R_S}$  and  $v_{R_L}$  are listed in Table 2.6, where  $R_S$  is kept at 50  $\Omega$  unchanged while  $R_L$  is varied from 10  $\Omega$  to 10 k $\Omega$ .

There are three cases shown in Table 2.6:

1. *Stepped-down case.*  $v_{R_L}$  is stepped down from  $v_{R_S}$ :

$$v_{R_L} < v_{R_S}, \quad (2.95)$$



**Figure 2.15.** Voltage delivered from a source to a load directly without an impedance matching network.

TABLE 2.6. Relationship between  $v_{R_S}$ ,  $R_S$  and  $v_{R_L}$ ,  $R_L$  for a given Value of  $P_S = 0 \text{ dB}_m$  in an Impedance Matching Loop as shown in Figure 2.15

$P_S, \text{dB}_m$	$v_S, \text{V}$	$R_S, \Omega$	$v_{R_S}, \text{V}$	$R_L$	$v_{R_L}, \text{V}$
0	0.2236	50	0.1863	10 $\Omega$	0.0373
0	0.2236	50	0.1803	12 $\Omega$	0.0433
0	0.2236	50	0.1694	16 $\Omega$	0.0542
0	0.2236	50	0.1553	20 $\Omega$	0.0683
0	0.2236	50	0.1398	30 $\Omega$	0.0839
0	0.2236	50	0.1118	50 $\Omega$	0.1118
0	0.2236	50	0.0745	100 $\Omega$	0.1491
0	0.2236	50	0.0447	200 $\Omega$	0.1789
0	0.2236	50	0.0203	500 $\Omega$	0.2033
0	0.2236	50	0.0106	1 k $\Omega$	0.2130
0	0.2236	50	0.0055	2 k $\Omega$	0.2182
0	0.2236	50	0.0011	10 k $\Omega$	0.2225

if

$$R_L < R_S. \quad (2.96)$$

2. *Unchanged case.*  $v_{R_L}$  or  $v_{R_S}$  is equal to  $v_{R_S}$ :

$$v_{R_L} = v_{R_S} = 0.1118 \text{ V}, \quad (2.97)$$

if

$$R_L = R_S = 50 \text{ } \Omega. \quad (2.98)$$

3. *Pumped-up case.*  $v_{R_L}$  is pumped up from  $v_{R_S}$ :

$$v_{R_L} > v_{R_S}, \quad (2.99)$$

if

$$R_L > R_S. \quad (2.100)$$

By comparing the two cases with and without impedance matching networks as shown in Tables 2.5 and 2.6, respectively, when a voltage is delivered from a source to a load, it is found that both cases are quite different.

The slight difference is that when  $R_S = R_L = 50 \text{ } \Omega$ ,  $v_{R_L} = v_{R_S} = 0.1583 \text{ V}$ , as shown in expression (2.92) for the case with the impedance matching network, but  $v_{R_L} = v_{R_S} = 0.1118 \text{ V}$ , as shown in the expression (2.101) for the case without the impedance matching network.

A large difference in voltage at the load  $v_{R_L}$  appears when  $R_L > R_S$ , especially when  $R_L \gg R_S$ , for example, when  $R_S = 50 \text{ } \Omega$  and  $R_L = 10 \text{ k}\Omega$ .

- In the case with impedance matching network, the value of  $v_{R_L}$  is dramatically pumped up to 2.2387 V from the source voltage  $v_S = 0.2236 \text{ V}$ , which is shown in the last row of Table 2.5.

- On the contrary, in the case without impedance matching network, the value of  $v_{R_L}$  is stepped down to 0.2225 V from the source voltage  $v_S = 0.2236$  V, which is shown in the last row of Table 2.6.

The voltage can be pumped up or stepped down if it is delivered from a source to a load through an impedance matching network. Special attention must be paid to the significantly pumped-up case. It is a very useful means in RF circuit design when a pumped-up voltage is needed.

Pumped-up or stepped-down voltage can be realized by an upward or downward resistance transformer. An upward resistance transformer is employed for delivering of a pumped-up voltage, while a downward resistance transformer is used for delivering a stepped-down voltage from the source to the load. Figure 2.16(a) and 2.16(b) shows the impedance matching network in Figure 2.14 by replacing the upward and downward resistance transformers, respectively.

Both the source and load impedance,  $Z_S$  and  $Z_L$ , are purely resistive, that is,

$$Z_S = R_S, \quad X_S = 0, \quad (2.101)$$

$$Z_L = R_L, \quad X_L = 0. \quad (2.102)$$

In Figure 2.16(a),

$$R_S < R_L \quad (2.103)$$

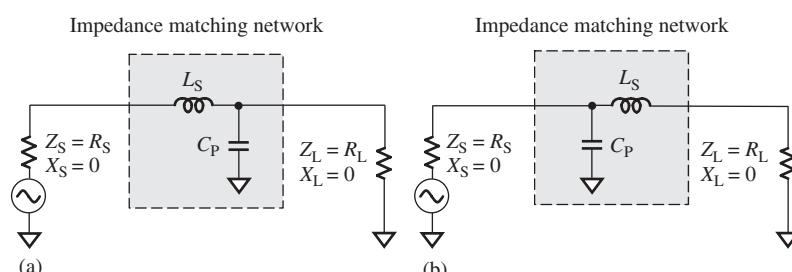
while in Figure 2.16(b),

$$R_S > R_L. \quad (2.104)$$

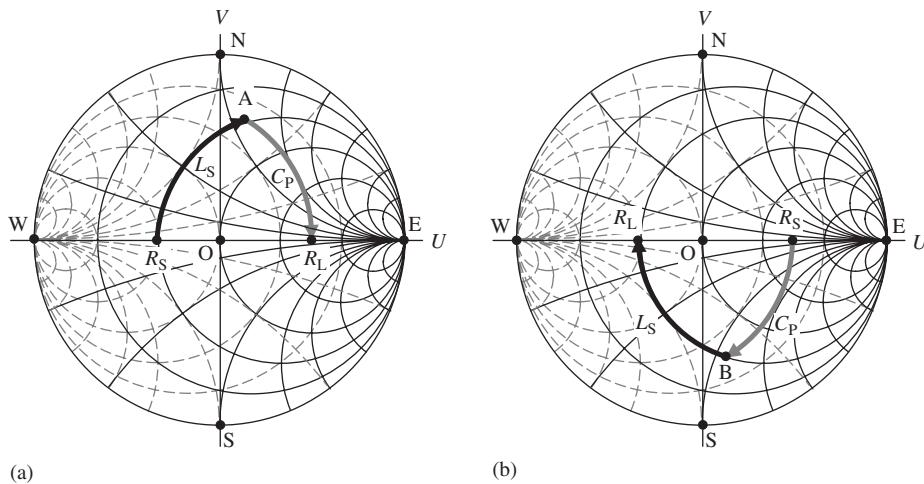
Figure 2.17 shows impedance matching of  $L_S - C_P$  upward and downward resistance transformers in the Smith Chart, respectively.

As a matter of fact, Figure 2.16(a) and 2.16(b) are images each other. They become the same if the positions of the load and the source,  $R_S$  and  $R_L$ , are exchanged, which corresponds to a change from an upward to a downward transformer.

Let us introduce the related formula from Lee (1998) for the downward impedance transformer as shown in Figure 2.16(b). The resistor  $R_L$  in series with  $L_S$  and the resistor



**Figure 2.16.** Voltage delivered from a source to a load through an impedance matching network, which is an  $L_S - C_P$  upward or downward resistance transformer. (a) Voltage is pumped up from  $R_S$  to  $R_L$  by the upward resistance transformer ( $R_S < R_L$ ). (b) Voltage is stepped down from  $R_S$  to  $R_L$  by the downward resistance transformer ( $R_S > R_L$ ).



**Figure 2.17.** Impedance matching of an  $L_S$ - $C_P$  upward and downward resistance transformer shown in a Smith Chart. (a) Upward resistance transformer consists of  $L_S$  and  $C_P$  ( $R_S < R_L$ ). (b) Downward resistance transformer consists of  $C_P$  and  $L_S$  ( $R_S > R_L$ ).

$R_S$  in parallel with  $C_P$  are impedance-matched if their  $Q$  values are equal to each other, that is

$$Q = \omega_0 C_P R_S = \frac{\omega_0 L_S}{R_L}, \quad (2.105)$$

where  $Q$  is called the *quality factor*, and  $\omega_0$  is the operating angular frequency

$$\omega_0^2 = \frac{1}{C_P L_S}, \quad (2.106)$$

$$R_S = R_L(1 + Q^2) = R_L \left[ 1 + \frac{1}{(\omega_0 R_L C_P)^2} \right] = R_L + \frac{1}{R_L} \frac{L_S}{C_P}. \quad (2.107)$$

Then,

$$R_L(R_S - R_L) = \frac{L_S}{C_P}. \quad (2.108)$$

From equations (2.105) and (2.107), we have

$$C_P = \frac{1}{\omega_0 \sqrt{R_L(R_S - R_L)}}, \quad (2.109)$$

$$L_S = \frac{\sqrt{R_L(R_S - R_L)}}{\omega_0}. \quad (2.110)$$

In these two equations,  $R_L$ ,  $R_S$ , and  $\omega_0$  are the given parameters. The values of  $C_P$  and  $L_S$  can be calculated from equations (2.108) and (2.109). They are real numbers because in a downward transformer  $R_S \gg R_L$ .

Now let us shift to the upward impedance transformer with  $L_S$  and  $C_P$  as shown in Figure 2.16(a). As mentioned earlier, the expressions for  $L_S$  and  $C_P$  can be obtained

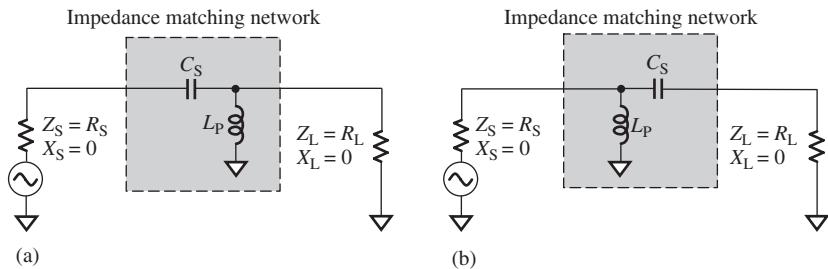
by simply exchanging the positions of the load and the source,  $R_S$  and  $R_L$ , in equations (2.108) and (2.109), that is,

$$L_S = \frac{\sqrt{R_S(R_L - R_S)}}{\omega_0}, \quad (2.111)$$

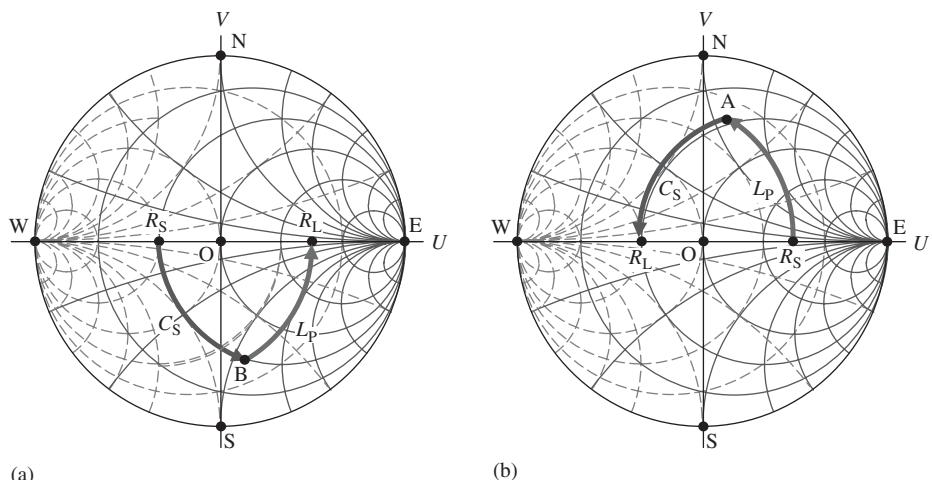
$$C_P = \frac{1}{\omega_0 \sqrt{R_S(R_L - R_S)}}. \quad (2.112)$$

In these two equations,  $R_L$ ,  $R_S$ , and  $\omega_0$  are the given parameters. The values of  $L_S$  and  $C_P$  can be calculated from equations (2.110) and (2.111). They are real numbers because in an upward transformer,  $R_S \ll R_L$ .

As a matter of fact, the upward and downward impedance transformer as shown in Figure 2.16(a) or (b) is a basic segment of a low-pass filter. We can have another



**Figure 2.18.** Voltage delivered from a source to a load through an impedance matching network, which is an  $C_S-L_P$  upward or downward resistance transformer. (a) Voltage is pumped up from  $R_S$  to  $R_L$  by the upward resistance transformer ( $R_S < R_L$ ). (b) Voltage is dropped down from  $R_S$  to  $R_L$  by the downward resistance transformer ( $R_S > R_L$ ).



**Figure 2.19.** Upward and downward resistance transformer consisting of  $C_S$  and  $L_P$ . (a) Upward resistance transformer consisting of  $C_S$  and  $L_P$  ( $R_S < R_L$ ). (b) Downward resistance transformer consisting of  $L_P$  and  $C_S$  ( $R_S > R_L$ ).

form of the matching network between  $R_L$  and  $R_S$ . Figure 2.18(a) and 2.18(b) shows an alternative  $C_S-L_P$  upward and downward impedance transformer, or, in other words, a matching network with a basic segment of a high-pass filter. Figure 2.19 shows the impedance matching of  $C_S-L_P$  upward and downward resistance transformers in the Smith Chart, respectively.

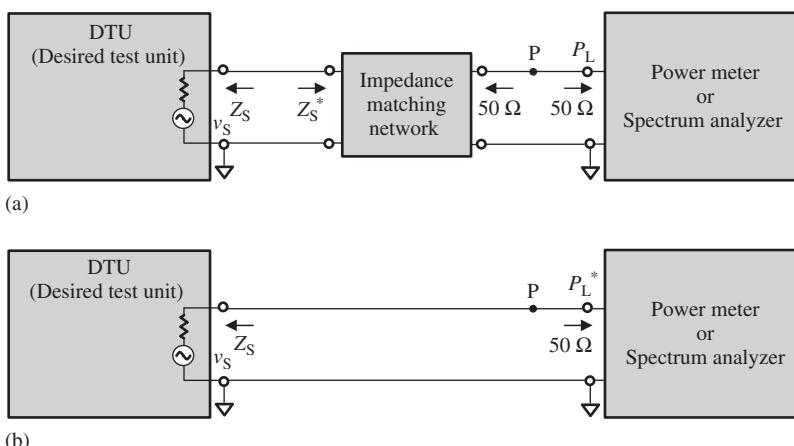
Again, as shown in Figure 2.18(a) and 2.18(b), these two types of upward and downward impedance transformers are images of each other. That is, both of them become the same if we exchange the position of the load and the source,  $R_L$  and  $R_S$ , from the upward to downward impedance transformer. The derivation of the two analytical expressions for  $C_S$  and  $L_P$ , which are similar to expressions (2.109)–(2.112), is given an exercise for the reader.

## 2.5.2 Power Measurement

Power measurement in an RF test laboratory is different from testing of other parameters. It is important to understand whether the test is done under matched or unmatched conditions since the test outcomes are determined by the impedance matching status between the tested point and the spectrum analyzer. The reading of power measurement under unmatched conditions is incorrect.

Usually, power measurement is conducted by means of a power meter or a spectrum analyzer. The impedance of the power meter or spectrum analyzer is typically  $50 \Omega$ , and the unit of power is decibel (dB). Figure 2.20 shows two cases of power testing at point P. Assume that the impedance and the voltage of the DTU (desired test unit) are  $Z_S$  and  $v_S$ , respectively, and that the input impedance of the power meter or spectrum analyzer is  $50 \Omega$ .

In the matched case (Fig. 2.20(a)), a matching network has been inserted between the tested point and the spectrum analyzer. As discussed in Section 2.3.2, the power reading from the spectrum analyzer is the same as shown in the expression (2.69),



**Figure 2.20.** Output power of a tested block measured by a power meter or a spectrum analyzer. (a) Testing conducted under the condition of impedance conjugate-matched. (b) Testing conducted under the condition of impedance unmatched.

that is

$$P_L = P_o = \frac{v_s^2}{4R_S}, \quad (2.113)$$

where  $P_o$  is maximum matched power that can be sensed by the spectrum analyzer. In the unmatched case (Fig. 2.20(b)), the spectrum analyzer directly measures the power at point P without any assistance, and we have

$$P_L^* = v_s^2 \frac{50}{|50 + Z_S|^2}, \quad (2.114)$$

where  $P_L^*$  is the measured power at point P.

The ratio of measured powers in the matched and unmatched cases is

$$\frac{P_L^*}{P_o} = 4R_S \frac{50}{|50 + Z_S|^2} = \frac{200R_S}{(50 + R_S)^2 + X_S^2}. \quad (2.115)$$

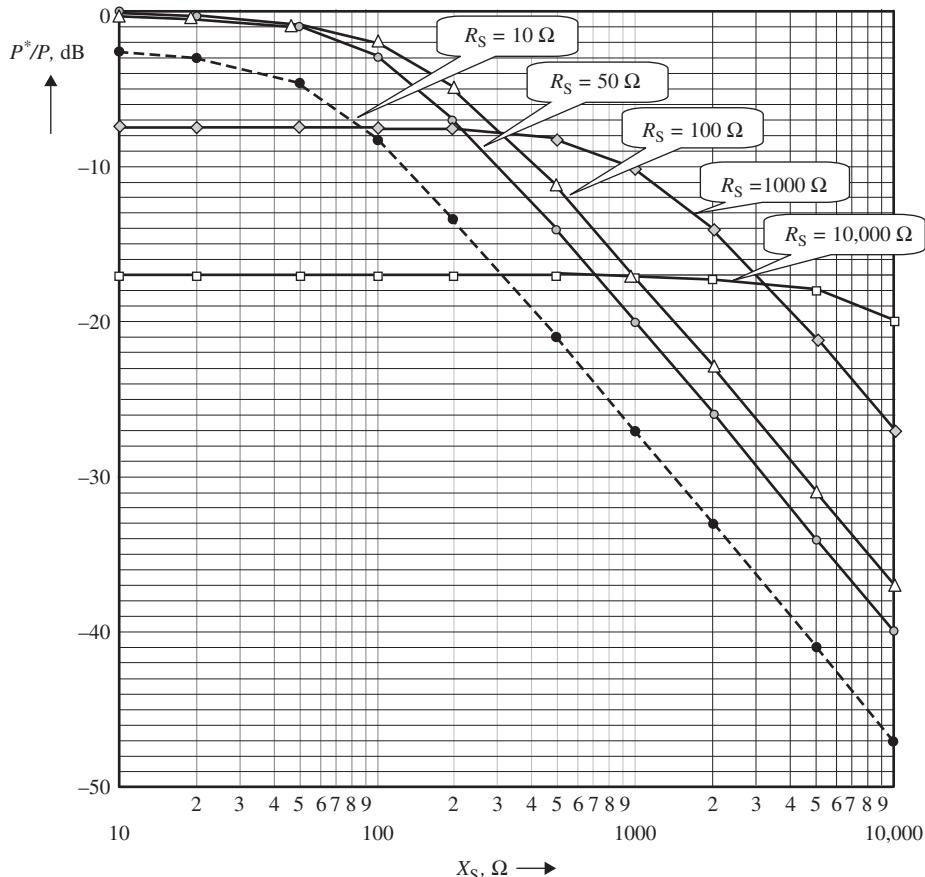


Figure 2.21. Plot of calculated power ratio,  $P_L^*/P_o$ , versus impedance of DTU,  $Z_S$ .

The expression (2.115) represents the difference between the measured powers under matched and unmatched cases. Table 2.7 lists the calculated results in decibels. It can be seen that in the 11th row of Table 2.7, where  $R_S = 50 \Omega$ ,  $X_S = 0 \Omega$ , the impedance matching condition is satisfied so that the impedance of the tested point matches with that of the spectrum analyzer. This results in the measured power  $P_L$  being equal to the expected power  $P_o$ , as shown in matched case (Fig. 2.20(a)). In general, in the unmatched case (Fig. 2.20(b)), the measured power  $P_L^*$  will deviate from the expected power  $P_o$  by a certain amount. The deviation becomes significant when the impedance is far from the matched condition, the power reading in the unmatched case being always lower than the reading under the impedance-matched case. In some cases, the difference is horrible! For instance, as shown in the 10th row of Table 2.7, where the impedance of the DTU is  $Z_S = 10 \Omega + j10 \text{ k}\Omega$  while the input impedance of power meter or spectrum analyzer is  $50 \Omega$ , the difference is 47 dB!

Figure 2.21 plots the calculated power ratio  $P_L^*/P_o$  versus impedance  $Z_S$  of the DTU. This “waterfall” diagram may give you an intuitive feeling for the inaccuracy of power measurement due to the unmatched impedance.

## APPENDICES

### 2.A.1 VSWR and Other Reflection and Transmission Coefficients

For the reader’s convenience, Table 2.A.1 lists the values calculated from the relationships between the impedance, reflection coefficient, transmission coefficient, and other parameters. The value of  $50 \Omega$  is taken as the reference impedance,  $R_o$  (Table 2.A.2).

The equations applied to calculate the various parameters are as follows:

- For voltage reflection coefficient

$$\Gamma = \frac{R - R_o}{R + R_o}. \quad (2.A.1)$$

- For VSWR (voltage standing wave ratio) in watt

$$\text{VSWR} = \frac{1 + |\Gamma|}{1 - |\Gamma|}. \quad (2.A.2)$$

- For VSWR in decibels

$$\text{VSWR}_{\text{dB}} = 20 \log(\text{VSWR}). \quad (2.A.3)$$

- For RL (return loss) in decibels

$$\text{RL}_{\text{dB}} = S_{11,\text{dB}} = 20 \log|\Gamma|. \quad (2.A.4)$$

- For TL (transmission loss) in decibels

$$\text{TL}_{\text{dB}} = 10 \log(1 - \Gamma^2). \quad (2.A.5)$$

TABLE 2.7. Calculated Power Ratio  $P_L^*/P_o$ , as the Impedance of DTU is Varied

$R_S, \Omega$	$X_S, \Omega$	$P_L^*/P_o$	$P_L^*/P_o, \text{ dB}$
10	0	0.555556	-2.6
10	20	0.500000	-3.0
10	50	0.327869	-4.8
10	100	0.147059	-8.3
10	200	0.500000	-13.4
10	500	0.327869	-21.0
10	1000	0.001993	-27.0
10	2000	0.500000	-33.0
10	5000	0.327869	-41.0
10	10,000	0.000020	-47.0
50	0	1.000000	0.0
50	20	0.961538	-0.2
50	50	0.800000	-1.0
50	100	0.500000	-3.0
50	200	0.200000	-7.0
50	500	0.038462	-14.1
50	1000	0.009901	-20.0
50	2000	0.002494	-26.0
50	5000	0.002494	-34.0
50	10,000	0.000100	-40.0
100	0	0.888889	-0.5
100	20	0.873362	-0.6
100	50	0.800000	-1.0
100	100	0.615385	-2.1
100	200	0.320000	-4.9
100	500	0.073394	-11.3
100	1000	0.019560	-17.1
100	2000	0.004972	-23.0
100	500	0.073394	-31.0
100	10,000	0.000200	-37.0
1000	0	0.181406	-7.4
1000	20	0.181340	-7.4
1000	50	0.180995	-7.4
1000	100	0.179775	-7.5
1000	200	0.175055	-7.6
1000	500	0.147874	-8.3
1000	1000	0.095125	-10.2
1000	2000	0.039196	-14.1
1000	5000	0.007662	-21.2
1000	10,000	0.001978	-27.0
10,000	0	0.019801	-17.0
10,000	20	0.019801	-17.0
10,000	50	0.019801	-17.0
10,000	100	0.019800	-17.0
10,000	200	0.019794	-17.0
10,000	500	0.019753	-17.0
10,000	1000	0.019607	-17.1
10,000	2000	0.019047	-17.2
10,000	5000	0.015873	-18.0
10,000	10,000	0.009950	-20.0

TABLE 2.A.1. Relationships between the Impedance, Reflection Coefficient, Transmission Coefficient, and other Parameters along the Axis  $U(V = 0)$  in the Complex Plane of the Voltage Reflection Coefficient  $\Gamma$

$R, \Omega$	$(R_o = 50 \Omega) r = R/R_o$	$ \Gamma $	VSWR	$VSWR_{dB}$	$(S_{11} \text{ or } S_{22}) RL_{dB}$	$TL_{dB}$	$\gamma$	$P_{T,\%}$	$P_{R,\%}$
0.2878	0.0056	0.99	178.57	45.04	-0.1000	-16.4280	0.98	2.22	97.78
1.4386	0.0290	0.94	34.48	30.75	-0.5000	-9.6361	0.89	10.96	89.04
2.8750	0.0574	0.89	17.42	24.82	-1.0000	-6.8683	0.79	20.53	79.47
5.7313	0.1146	0.79	8.73	18.82	-2.0000	-4.3292	0.63	36.90	63.10
8.5500	0.1710	0.71	5.85	15.34	-3.0000	-3.0206	0.50	49.88	50.12
11.3136	0.2263	0.63	4.42	12.91	-4.0000	-2.2048	0.40	60.19	39.81
14.0066	0.2801	0.56	3.57	11.06	-5.0000	-1.6509	0.32	68.36	31.64
16.6140	0.3323	0.50	3.01	9.57	-6.0000	-1.2563	0.25	74.88	25.12
19.1236	0.3825	0.45	2.62	8.35	-7.0000	-0.9665	0.20	80.04	19.96
21.5252	0.4305	0.40	2.32	7.32	-8.0000	-0.7494	0.16	84.15	15.85
23.8110	0.4762	0.35	2.10	6.44	-9.0000	-0.5843	0.13	87.42	12.58
25.9747	0.5195	0.32	1.93	5.69	-10.0000	-0.4576	0.10	89.99	10.01
29.9240	0.5986	0.25	1.67	4.46	-12.0000	-0.2830	0.06	93.70	6.30
34.9020	0.6980	0.18	1.43	3.12	-15.0000	-0.1396	0.03	96.84	3.16
40.9091	0.8182	0.10	1.22	1.74	-20.0000	-0.0436	0.01	99.00	1.00
44.6760	0.8935	0.06	1.12	0.98	-25.0000	-0.0138	0.00	99.68	0.32
46.9347	0.9387	0.03	1.07	0.55	-30.0000	-0.0043	0.00	99.90	0.10
48.2528	0.9651	0.02	1.04	0.31	-35.0000	-0.0014	0.00	99.97	0.03
49.0099	0.9802	0.01	1.02	0.17	-40.0000	-0.0004	0.00	99.99	0.01
49.4408	0.9888	0.01	1.01	0.10	-45.0000	-0.0001	0.00	100.00	0.00

(continued)

TABLE 2.A.1. (Continued)

$R, \Omega$	$(R_0 = 50 \Omega) r = R/R_0$	$ \Gamma $	VSWR	$VSWR_{dB}$	$(S_{11} \text{ or } S_{22}) RL_{dB}$	$TL_{dB}$	$\gamma$	$P_{T,\%}$	$P_{R,\%}$
49.6848	0.9937	0.00	1.01	0.05	-50.0000	0.0000	0.00	100.00	0.00
50.0000	1.0000	0.00	1.00	0.00	$-\infty$	0.0000	0.00	100.00	0.00
50.3172	1.0063	0.00	1.01	0.05	-50.0000	0.0000	0.00	100.00	0.00
50.5655	1.0113	0.01	1.01	0.10	-45.0000	-0.0001	0.00	100.00	0.00
51.0101	1.0202	0.01	1.02	0.17	-40.0000	-0.0004	0.00	99.99	0.01
51.8105	1.0362	0.02	1.04	0.31	-35.0000	-0.0014	0.00	99.97	0.03
53.2656	1.0653	0.03	1.07	0.55	-30.0000	-0.0043	0.00	99.90	0.10
55.9585	1.1192	0.06	1.12	0.98	-25.0000	-0.0138	0.00	99.68	0.32
61.1111	1.2222	0.10	1.22	1.74	-20.0000	-0.0436	0.01	99.00	1.00
71.6291	1.4326	0.18	1.43	3.12	-15.0000	-0.1396	0.03	96.84	3.16
83.5450	1.6710	0.25	1.67	4.46	-12.0000	-0.2830	0.06	93.69	6.31
96.2475	1.9244	0.32	1.93	5.69	-10.0000	-0.4576	0.10	90.01	9.99
104.9942	2.0999	0.35	2.10	6.44	-9.0000	-0.5844	0.13	87.41	12.59
116.1431	2.3229	0.40	2.32	7.32	-8.0000	-0.7494	0.16	84.15	15.85
130.7280	2.6146	0.45	2.61	8.35	-7.0000	-0.9665	0.20	80.05	19.95
150.4750	3.0095	0.50	3.01	9.57	-6.0000	-1.2563	0.25	74.88	25.12
178.4900	3.5698	0.56	3.57	11.05	-5.0000	-1.6509	0.32	68.38	31.62
220.9700	4.4194	0.63	4.42	12.91	-4.0000	-2.2048	0.40	60.19	39.81
292.4050	5.8481	0.71	5.85	15.34	-3.0000	-3.0207	0.50	49.88	50.12
436.2200	8.7244	0.79	8.72	18.81	-2.0000	-4.3293	0.63	36.90	63.10
869.5500	17.3910	0.89	17.39	24.81	-1.0000	-6.8683	0.79	20.57	79.43

TABLE 2.A.2. Relationships between Power (dB), Voltage (V), and Power (dB)

Power, dB <sub>m</sub>	Voltage, V	Power, W	Power, dB <sub>m</sub>	Voltage, mV	Power, mW	Power, dB <sub>m</sub>	Voltage, MV	Power, nW
50	70.71	100.00	0	223.61	1.0000	-50	707.11	10.0000
49	63.02	79.43	-1	199.29	0.794328	-51	630.21	7.943282
48	56.17	63.10	-2	177.62	0.630957	-52	561.67	6.309573
47	50.06	50.12	-3	158.30	0.501187	-53	500.59	5.011872
46	44.62	39.81	-4	141.09	0.398107	-54	446.15	3.981072
45	39.76	31.62	-5	125.74	0.316228	-55	397.64	3.162278
44	35.44	25.12	-6	112.07	0.251189	-56	354.39	2.511886
43	31.59	19.95	-7	99.88	0.199526	-57	315.85	1.995262
42	28.15	15.85	-8	89.02	0.158489	-58	281.50	1.584893
41	25.09	12.59	-9	79.34	0.125893	-59	250.89	1.258925
40	22.36	10	-10	70.71	0.1000	-60	223.61	1.0000
39	19.93	7.94	-11	63.02	0.079433	-61	199.29	0.794328
38	17.76	6.31	-12	56.17	0.063096	-62	177.62	0.630957
37	15.83	5.01	-13	50.06	0.050119	-63	158.30	0.501187
36	14.11	3.98	-14	44.62	0.039811	-64	141.09	0.398107
35	12.57	3.16	-15	39.76	0.031623	-65	125.74	0.316228
34	11.21	2.51	-16	35.44	0.025119	-66	112.07	0.251189
33	9.99	2.00	-17	31.59	0.019953	-67	99.88	0.199526
32	8.90	1.58	-18	28.15	0.015849	-68	89.02	0.158489
31	7.93	1.26	-19	25.09	0.012589	-69	79.34	0.125893
30	7.07	1	-20	22.36	0.0100	-70	70.71	0.1000
29	6.30	0.79	-21	19.93	0.007943	-71	63.02	0.079433
28	5.62	0.63	-22	17.76	0.006310	-72	56.17	0.063096
27	5.01	0.50	-23	15.83	0.005012	-73	50.06	0.050119
26	4.46	0.40	-24	14.11	0.003981	-74	44.62	0.039811
25	3.98	0.32	-25	12.57	0.003162	-75	39.76	0.031623
24	3.54	0.25	-26	11.21	0.002512	-76	35.44	0.025119
23	3.16	0.20	-27	9.99	0.001995	-77	31.59	0.019953
22	2.82	0.16	-28	8.90	0.001585	-78	28.15	0.015849

(continued)

TABLE 2.A.2. (Continued)

Power, dB <sub>m</sub>	Voltage, V	Power, W	Power, dB <sub>m</sub>	Voltage, mV	Power, mW	Power, dB <sub>m</sub>	Voltage, mV	Power, nW
21	2.51	0.13	-29	7.93	0.001259	-79	25.09	0.012589
20	2.24	0.1	-30	7.07	0.0010	-80	22.36	0.0100
19	1.99	0.0794	-31	6.30	0.000794	-81	19.93	0.007943
18	1.78	0.0631	-32	5.62	0.000631	-82	17.76	0.006310
17	1.58	0.0501	-33	5.01	0.000501	-83	15.83	0.005012
16	1.41	0.0398	-34	4.46	0.000398	-84	14.11	0.003981
15	1.26	0.0316	-35	3.98	0.000316	-85	12.57	0.003162
14	1.12	0.0251	-36	3.54	0.000251	-86	11.21	0.002512
13	1.00	0.0200	-37	3.16	0.000200	-87	9.99	0.001995
12	0.89	0.0158	-38	2.82	0.000158	-88	8.90	0.001585
11	0.79	0.0126	-39	2.51	0.000126	-89	7.93	0.001259
10	0.71	0.0100	-40	2.24	0.0001	-90	7.07	0.0010
9	0.63	0.0079	-41	1.99	0.000079	-91	6.30	0.000794
8	0.56	0.0063	-42	1.78	0.000063	-92	5.62	0.000631
7	0.50	0.0050	-43	1.58	0.000050	-93	5.01	0.000501
6	0.45	0.0040	-44	1.41	0.000040	-94	4.46	0.000398
5	0.40	0.0032	-45	1.26	0.000032	-95	3.98	0.000316
4	0.35	0.0025	-46	1.12	0.000025	-96	3.54	0.000251
3	0.32	0.0020	-47	1.00	0.000020	-97	3.16	0.000200
2	0.28	0.0016	-48	0.89	0.000016	-98	2.82	0.000158
1	0.25	0.0013	-49	0.79	0.000013	-99	2.51	0.000126
34	11.21	2.51	-16	35.44	0.025119	-66	112.07	0.251189
33	9.99	2.00	-17	31.59	0.019953	-67	99.88	0.199526
32	8.90	1.58	-18	28.15	0.015849	-68	89.02	0.158489
31	7.93	1.26	-19	25.09	0.012589	-69	79.34	0.125893
30	7.07	1	-20	22.36	0.0100	-70	70.71	0.1000

29	6.30	0.79	-21	19.93	0.007943	-71	63.02	0.079433
28	5.62	0.63	-22	17.76	0.006310	-72	56.17	0.063096
27	5.01	0.50	-23	15.83	0.005012	-73	50.06	0.050119
26	4.46	0.40	-24	14.11	0.003981	-74	44.62	0.039811
25	3.98	0.32	-25	12.57	0.003162	-75	39.76	0.031623
24	3.54	0.25	-26	11.21	0.002512	-76	35.44	0.025119
23	3.16	0.20	-27	9.99	0.001995	-77	31.59	0.019953
22	2.82	0.16	-28	8.90	0.001585	-78	28.15	0.015849
21	2.51	0.13	-29	7.93	0.001259	-79	25.09	0.012589
20	2.24	0.1	-30	7.07	0.0010	-80	22.36	0.0100
19	1.99	0.0794	-31	6.30	0.000794	-81	19.93	0.007943
18	1.78	0.0631	-32	5.62	0.000631	-82	17.76	0.006310
17	1.58	0.0501	-33	5.01	0.000501	-83	15.83	0.005012
16	1.41	0.0398	-34	4.46	0.000398	-84	14.11	0.003981
15	1.26	0.0316	-35	3.98	0.000316	-85	12.57	0.003162
14	1.12	0.0251	-36	3.54	0.000251	-86	11.21	0.002512
13	1.00	0.0200	-37	3.16	0.000200	-87	9.99	0.001995
12	0.89	0.0158	-38	2.82	0.000158	-88	8.90	0.001585
11	0.79	0.0126	-39	2.51	0.000126	-89	7.93	0.001259
10	0.71	0.0100	-40	2.24	0.0001	-90	7.07	0.0010
9	0.63	0.0079	-41	1.99	0.000079	-91	6.30	0.000794
8	0.56	0.0063	-42	1.78	0.000063	-92	5.62	0.000631
7	0.50	0.0050	-43	1.58	0.000050	-93	5.01	0.000501
6	0.45	0.0040	-44	1.41	0.000040	-94	4.46	0.000398
5	0.40	0.0032	-45	1.26	0.000032	-95	3.98	0.000316
4	0.35	0.0025	-46	1.12	0.000025	-96	3.54	0.000251
3	0.32	0.0020	-47	1.00	0.000020	-97	3.16	0.000200
2	0.28	0.0016	-48	0.89	0.000016	-98	2.82	0.000158
1	0.25	0.0013	-49	0.79	0.000013	-99	2.51	0.000126

mV represents millivolt and mW represents milliwatts

- For power reflection coefficient

$$\gamma = \Gamma^2. \quad (2.A.6)$$

- For transmitted power in percentage

$$P_{T,\%} = 100(1 - \Gamma^2). \quad (2.A.7)$$

- For reflected power in percentage

$$P_{R,\%} = 100 \Gamma^2. \quad (2.A.8)$$

## 2.A.2 Relationships between Power ( $\text{dB}_m$ ), Voltage (V), and Power (W)

The equations applied for calculation of voltage (mV) from power ( $\text{dB}_m$ ), and for power (mW) from voltage (mV) are as follows:

- Voltage (mV) calculated from power ( $\text{dB}_m$ ):

$$v_V = \sqrt{(P_{\text{dB}_m} - 30)R_\Omega}, \quad (2.A.9)$$

$$v_{\text{mV}} = 10^3 \sqrt{(P_{\text{dB}_m} - 30)R_\Omega}, \quad (2.A.10)$$

$$v_{\mu\text{V}} = 10^6 \sqrt{(P_{\text{dB}_m} - 30)R_\Omega}. \quad (2.A.11)$$

- Power (mW) calculated from voltage (mV):

$$P_W = v_V^2 \frac{1}{R_\Omega}, \quad (2.A.12)$$

$$P_{\text{mW}} = 10^3 \left( \frac{v_{\text{mV}}}{10^3} \right)^2 \frac{1}{R_\Omega}, \quad (2.A.13)$$

$$P_{\text{nW}} = 10^9 \left( \frac{v_{\mu\text{V}}}{10^6} \right)^2 \frac{1}{R_\Omega}. \quad (2.A.14)$$

## REFERENCE

Lee TH. *The Design of CMOS Radio-Frequency Integrated Circuits*. 2nd ed. Cambridge University Press; 1998.

## FURTHER READING

S-Parameters, circuit analysis and design. Hewlett-Packard Application Note 95. 1968 Sep.

Abrie PLD. *The Design of Impedance Matching Networks for Radio-Frequency and Microwave Amplifiers*. Norwood (MA): Artech House; 1985.

- Hayt WH Jr, Kemmerly JE. *Engineering Circuit Analysis*. New York: McGraw-Hill; 1971.
- Sifri JD. Matching Technique Yields Optimum LNA Performance. *Microwaves & RF*, 1986 Feb. pp. 87–90.
- Smith PH. *Electronic Applications of the Smith Chart*. New York: McGraw-Hill; 1969.
- Vendelin GD, Alexander W, Mock D. Computer analyzes RF circuits with generalized smith charts. *Electronics* 1974;45:102–109.

## EXERCISES

1. What are the main functions of impedance matching?
2. What happens to a digital circuit operating in an unmatched condition?
3. What happens to an RF circuit operating in an unmatched condition?
4. Assuming that the total power in the entire loop from the source to the load is  $P_S$ , what is the distribution of total power to the source and the load when the impedance of the loop is matched?
5. In the discussion of impedance matching, the condition  $Z_S = Z_L^*$  corresponds to the case without power reflection between the source and the load, whereas the condition  $Z_S = Z_L$  corresponds to the case without voltage reflection between the source and the load. What is the difference between the conditions  $Z_S = Z_L^*$  and  $Z_S = Z_L$ ?
6. When impedance is unmatched, all additional power loss, additional distortion, and SIR degradation disappear when  $\gamma_L = 0$  but  $\gamma_S \neq 0$ . Why?
7. What happens if one measures the power for a RF circuit under the impedance unmatched condition?
8. Assuming that
  - (a) the output power of a generator with  $50 \Omega$  output impedance is  $0 \text{ dB}_m$  so that its  $V_{\text{rms}}$  is  $0.2236 \text{ V}$ ,
  - (b) the impedance of a load is  $1000 \Omega$ , and
  - (c) an ideal impedance matching network is built between the generator and the load,
 what is the value of  $V_{\text{rms}}$  delivered to the load?
9. In the problem 8, if the impedance of a load were  $10 \Omega$ , what would be the value of  $V_{\text{rms}}$  delivered to the load?
10. In an impedance matching network, the impedances are not matched between two parts, usually. It is not necessary to insert an impedance matching network between these two parts. Why?

## ANSWERS

1. The main functions of impedance matching are
  - maximization of power transport, and
  - power transport without additional phase shift.
2. If a digital circuit operates in unmatched condition, the circuit will suffer from additional voltage attenuation, distortion, and jitter.
3. If an RF circuit operates is in unmatched condition, the circuit will suffer from additional power loss, distortion, and degradation of the SIR.

4. Assuming that the total power in the entire loop from the source to the load is  $P_S$  and when the impedance of the loop is matched, the total power delivered to the source and the load is 50:50, that is,

$$P_S = P_L = P_S/2.$$

5. The condition  $Z_S = Z_L^*$  corresponds to the case without power reflection between the source and the load, whereas the condition  $Z_S = Z_L$  corresponds to the case without voltage reflection between the source and the load. When condition  $Z_S = Z_L^*$ , there is only resistance in the loop so that these two conditions become identical, that is,  $R_S = R_L$ . Consequently, reflection of neither power nor voltage happens.
6. When impedance is unmatched, the additional power loss, additional distortion, and the SIR degradation disappear when  $\gamma_L = 0$  but  $\gamma_S \neq 0$  because there is no incident power from the load to the source since  $\gamma_L = 0$ .
7. Under conditions of unmatched impedance, the reading of measured power for an RF circuit is wrong. Usually, the reading will be much lower than what it should be under the impedance-matched condition.
8. Assuming that
- the output power of a generator with  $50 \Omega$  output impedance is  $0 \text{ dB}_m$  so that its  $V_{\text{rms}}$  is  $0.2236 \text{ V}$ ,
  - the impedance of a load is  $1000 \Omega$ , and
  - an ideal impedance matching network is built between the generator and the load, a voltage  $V_{\text{rms}} = 0.7079 \text{ V}$  is delivered to the load.
9. In the problem 8, if the impedance of a  $l_{\text{oad}}$  is  $10 \Omega$ , a voltage  $V_{\text{rms}} = 0.0708 \text{ V}$  is delivered to the load.
10. It is necessary to insert an impedance matching network between two independent units or blocks for power transport if these two independent units or blocks are not impedance-matched. However, it is not necessary to insert an impedance matching network between two parts because the individual part is not an independent unit or blocks power transport.

---

# IMPEDANCE MATCHING IN THE NARROW-BAND CASE

---

## 3.1 INTRODUCTION

As emphasized in the previous chapters, impedance matching is the key task in RF circuit design because the main function of the RF block is to ensure maximization of power transfer without any phase shift. The task of impedance matching is to change the original impedance to the standard reference impedance, that is,  $50 \Omega$ , or some other desired impedance by constructing an impedance matching network.

There are some basic engineering fundamentals that circuit designers must bear in mind in the process of impedance matching.

First, in real-life RF circuit block designs, the input or output impedance of an RF block may be required to match a certain impedance that is not  $50 \Omega$  for cost saving and performance improvement. However, in order to enable the testing of the said RF circuit blocks, the input and output impedances must nonetheless be matched to  $50 \Omega$ , which is the standard reference impedance that test equipment manufacturers adopt. Thus, a somewhat complicated design procedure must be used when the actual RF circuit design must match an impedance that is not  $50 \Omega$ . This will be discussed in Section 3.6.

Second, the impedance matching network could be built by either active or passive parts. It is more desirable to build an impedance matching network using passive parts rather than with active parts because the former is simpler, saves current, and is cost effective compared to the latter. The passive parts include inductors, capacitors, and resistors; however, resistors are usually excluded because they attenuate the signal and introduce considerable noise. Even though their use is allowed, almost no one does so.

Consequently, what we need to be familiar with is how to apply capacitors and inductors to the matching network. To build a narrow-band impedance matching network, one, two, or three parts are needed. However, the number of parts may be increased depending on the relative bandwidth.

Third, it is necessary to distinguish between a “narrow” and a “wide” band. As a matter of fact, this demarcation is not made strictly from theory, but from engineering design experience. Generally speaking, an RF circuit block is categorized as a narrow-band block if its relative bandwidth of operation is much less than 15%, while it is categorized as a wide-band block if its relative bandwidth of operation is more than 15%. When the relative bandwidth of operation is around 15%, it is better to treat an RF circuit block as a wide-band block so as to have a reasonable performance for both wide-band and narrow-band cases.

In the return loss RL ( $S_{11}$  or  $S_{22}$ ) testing, the trace of the impedance  $Z$  or return loss RL ( $S_{11}$  or  $S_{22}$ ) displayed on the Smith Chart corresponds to the response for a frequency bandwidth. As an example, Figure 3.1 shows the frequency response of the impedance  $Z$  or return loss RL ( $S_{11}$  or  $S_{22}$ ) from the lowest frequency to the highest frequency: that is,

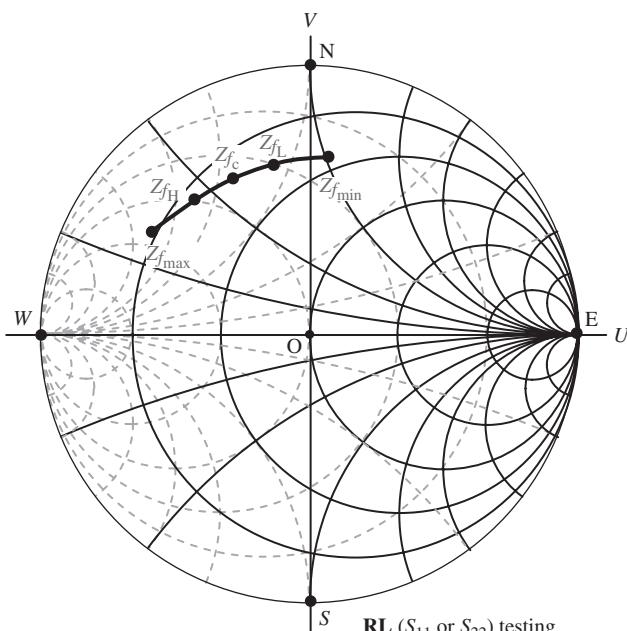
$$\delta f = f_{\max} - f_{\min}, \quad (3.1)$$

$$\text{BW} = f_H - f_L, \quad (3.2)$$

$$f_c = \frac{1}{2}(f_L + f_H). \quad (3.3)$$

where

$\delta f$  = frequency coverage of the trace,  
 $f_{\max}$  = highest frequency on the trace,



**Figure 3.1.** A trace displayed on a Smith Chart from RL ( $S_{11}$  or  $S_{22}$ ) testing covering a bandwidth  $\text{BW} = f_H - f_L$ .

$f_{\min}$  = lowest frequency on the trace,  
 $\Delta f$  = BW = frequency bandwidth,  
 $f_H$  = the high-frequency end of desired bandwidth,  
 $f_L$  = the low-frequency end of desired bandwidth, and  
 $f_c$  = the central frequency.

Each point on the displayed trace corresponds to a frequency. The values of the impedance  $Z$  and return loss RL ( $S_{11}$  or  $S_{22}$ ) can be read at each point simultaneously:

$Z_{f_{\min}}$  = the impedance at the lowest frequency on the trace,  
 $Z_{f_{\max}}$  = the impedance at the highest frequency on the trace,  
 $Zf_L$  = the impedance at the low-frequency end of the desired bandwidth,  
 $Zf_H$  = the impedance at the high-frequency end of the desired bandwidth, and  
 $Zf_c$  = the impedance at the central frequency point.

In wide-band RF circuit design, the response for all frequency bandwidths must be taken care of. However, in narrow-band RF circuit design, we could simply focus our attention on the central frequency because the frequency response at the central frequency is a good approximation of that for the entire bandwidth. Consequently, in all displays of the Smith Chart in this chapter, only one point corresponding to the central frequency is shown since only narrow-band cases are discussed in this chapter.

Fourth, the Smith Chart is a very powerful tool in an impedance matching work. For a simple impedance matching work, one can design the network through manual calculations by means of relevant equations. However, the task becomes much easier with the assistance of the Smith Chart. Here, the source or load impedance is no longer restricted to a pure resistance but can be a complex impedance, including both a real and an imaginary part, although the standard reference impedance, say  $50 \Omega$ , is always desired in the RF design.

There are many reference books on the applications of the Smith Chart in engineering design. Here we do not attempt to repeat the description, but only reproduce some important relations and equations between the parameters in Appendix 3.A.1.

The passive impedance matching network for the narrow-band case will be discussed in this chapter, while that for the wide-band case will be discussed in Chapter 4.

## 3.2 IMPEDANCE MATCHING BY MEANS OF RETURN LOSS ADJUSTMENT

### 3.2.1 Return Loss Circles on the Smith Chart

In the complex plane of voltage reflection coefficient, a family of power reflection coefficients  $\gamma = \Gamma^2$  and their corresponding return loss circles, either  $S_{11}$  or  $S_{22}$ , can be plotted using the Smith Chart with impedance coordination since

$$\Gamma = U + jV \quad (3.4)$$

and

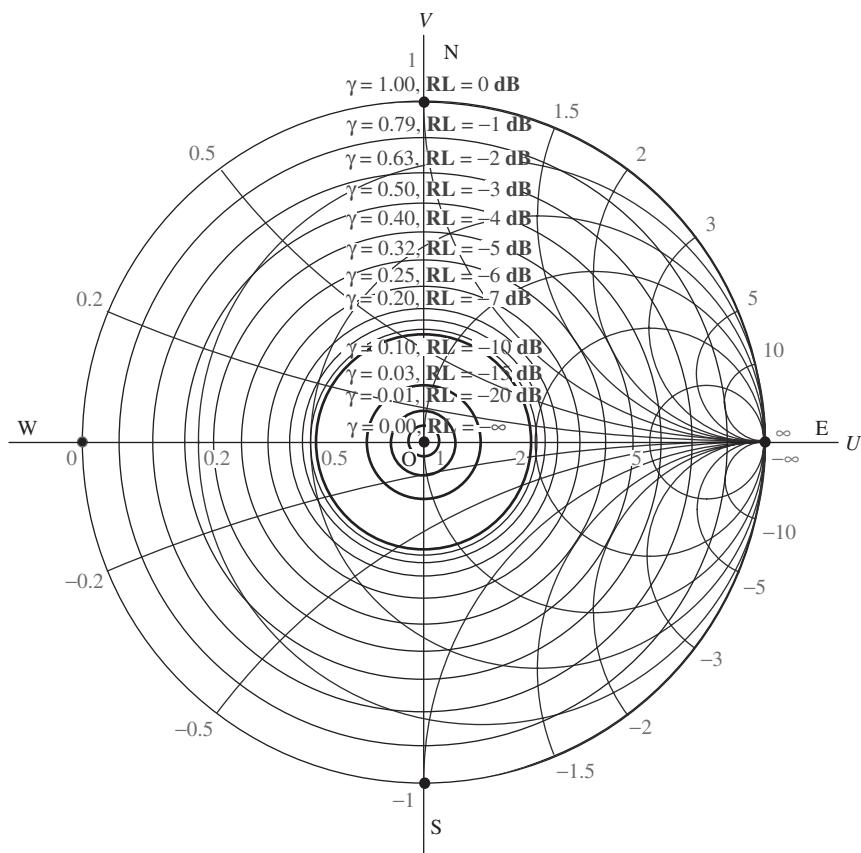
$$\gamma = \Gamma^2 = U^2 + V^2. \quad (3.5)$$

In the complex plane of the voltage reflection coefficient, the scale in both the  $U$ - and  $V$ -axes is linear so that the equation (3.5) is a circle centered at the reference impedance point  $50 \Omega$  with a radius  $|\Gamma|$ . In other words, all the points where  $|\Gamma| = \text{constant}$  are located on the same circle. Then, according to the definition of return loss,

$$\text{RL}(S_{11}\text{ or }S_{22}) = 10 \log |\Gamma|^2 \text{ in dB} \quad (3.6)$$

and all the points for which  $\text{RL} = \text{constant}$  are located on the same circle.

Figure 3.2 plots this family of circles. The center of these circles is the reference impedance point,  $50 \Omega$ : that is, the center of the Smith Chart. The constant value of each



**Figure 3.2.** Constant return loss  $S_{11,\text{dB}}$  or  $S_{22,\text{dB}}$ .

Note 1: Power reflection coefficient  $\gamma$  and return loss  $\text{RL}$  ( $S_{11}$  or  $S_{22}$ ) is bold-marked with values along the vertical axis,  $V$ , such as,  $\gamma = 1, \text{RL} = 0 \text{ dB}$ ,  $\gamma = 0.79, \text{RL} = -1 \text{ dB}$ ,  $\gamma = 0.63, \text{RL} = -2 \text{ dB}$ , and so on.

Note 2: Normalized resistance  $x$  is bold-marked with values around the biggest circle, such as, **0, 0.1, 0.2, 0.5, 1, 1.5, 2, 3, 5, 10,  $\infty$ ,  $-\infty$ , -10, -5, -3, -2, -1.5, -1, -0.5, -0.2, -0.1**.

Note 3: Normalized reactance  $r$  is bold-marked with values along the horizontal axis  $U$ , such as, **0, 0.2, 0.5, 1, 2, 5,  $\infty$** .

TABLE 3.1. Variation of Return Loss RL along the  $U$ -axis ( $V = 0$ )

Resistance, $R, \Omega$	Reference Resistance, $R_o, \Omega$	Normalized Resistance, $r, r = R/R_o$	Power Reflection Coefficient $\gamma = \Gamma^2$	Return Loss, RL ( $S_{11}$ or $S_{22}$ ), dB
2.8750	50	0.0575	0.79	-1
5.7313	50	0.1146	0.63	-2
8.5500	50	0.1710	0.50	-3
11.3136	50	0.2263	0.40	-4
14.0066	50	0.2801	0.32	-5
16.6140	50	0.3323	0.25	-6
19.1236	50	0.3825	0.20	-7
21.5252	50	0.4305	0.16	-8
23.8110	50	0.4762	0.13	-9
25.9747	50	0.5195	0.10	-10
29.9240	50	0.5986	0.06	-12
34.9020	50	0.6980	0.03	-15
40.9091	50	0.8182	0.01	-20
44.6760	50	0.8935	0.00	-25
46.9347	50	0.9387	0.00	-30
48.2528	50	0.9651	0.00	-35
49.0099	50	0.9802	0.00	-40
49.4408	50	0.9888	0.00	-45
49.6848	50	0.9937	0.00	-50
50.0000	50	1.0000	0.00	$-\infty$
50.3172	50	1.0063	0.00	-50
50.5655	50	1.0113	0.00	-45
51.0101	50	1.0202	0.00	-40
51.8105	50	1.0362	0.00	-35
53.2656	50	1.0653	0.00	-30
55.9585	50	1.1192	0.00	-25
61.1111	50	1.2222	0.01	-20
71.6291	50	1.4326	0.03	-15
83.5450	50	1.6709	0.06	-12
96.2475	50	1.9250	0.10	-10
104.9942	50	2.0999	0.13	-9
116.1431	50	2.3229	0.16	-8
130.7280	50	2.6146	0.20	-7
150.4750	50	3.0095	0.25	-6
178.4900	50	3.5698	0.32	-5
220.9700	50	4.4194	0.40	-4
292.4050	50	5.8481	0.50	-3
436.2200	50	8.7244	0.63	-2
869.5500	50	17.3910	0.79	-1

$\gamma$  or RL circle can be determined, as in Table 2.A.1. For convenience, let us relist the values of some columns from Table 2.A.1 in Table 3.1.

There are 39 rows in Table 3.1. In the last column, the values are symmetrical with respect to the 20th row where  $S_{11}$  or  $S_{22} = -\infty$ . This indicates that there are two normalized resistances corresponding to one value of  $S_{11}$  or  $S_{22}$  along the axis  $U$  ( $V = 0$ ) in the complex plane of the voltage reflection coefficient. In other words, there are two points on the  $U$ -axis ( $V = 0$ ) for the same value of  $S_{11}$  or  $S_{22}$ . One of them is located to the left of the center of the circle,  $50 \Omega$ , while the other is on the right side. The

distance between these two points is the diameter of the circle. For example, when the return loss is

$$RL = S_{11} \text{ or } S_{22} = -15 \text{ dB}, \quad (3.7)$$

the corresponding normalized resistances or the two resistances are

$$r = 0.6980 \text{ and } 1.4326, \quad (3.8)$$

$$R = 34.9020 \Omega \text{ and } 71.6291 \Omega. \quad (3.9)$$

It should be noted that the constant value of each  $\gamma$  or RL circle can be alternatively determined by a radially scaled ruler on a conventional Smith Chart.

### 3.2.2 Relationship between Return Loss and Impedance Matching

Figure 3.2 indicates one of the important features of the impedance on the Smith Chart. If an impedance is far from the center of the Smith Chart, it is far from the impedance matching state so that the return loss is high. In an extreme case, the impedance is located at the biggest outer circle, which implies a zero resistance. This leads to a maximum return loss, namely, 0 dB, in which case the reflection voltage is equal to the incident voltage. A terminal with such an impedance is in a full reflection state. On the contrary, another extreme case is where the impedance is equal to the reference impedance  $50 \Omega$ . It is a terminal with return loss  $= -\infty$ , which is an ideal impedance-matched state without reflection. In other cases, as the impedance approaches the center of the Smith Chart from the biggest circle, the power reflection coefficient  $\gamma$  and the return loss RL decrease simultaneously. Generally speaking, both the power reflection coefficient  $\gamma$  and the return loss RL are indicators of the impedance matching state. However, from the values of RL and  $\gamma$  as specified in Figure 3.2, it is easily seen that return loss RL is more sensitive to the impedance matching state than the power reflection coefficient  $\gamma$ . This is why return loss RL ( $S_{11}$  or  $S_{22}$ ), instead of the power reflection coefficient  $\gamma$ , is always taken as an indicator of the impedance matching state.

In practical engineering design, the ideal impedance matching state is never reached. RF circuit designers always try their best to match the original impedance to the ideal matched state. Does any rule exist to measure the impedance matching state? No! There is no formal regulation or rule suggested by anyone or by any international organization. RF circuit designers can have their own rules to judge the impedance matching state. However, an approximate rule is very often adopted by RF circuit designers in their simulations, that is,

$$RL = S_{11} \text{ or } S_{22} = -10 \text{ dB}; \quad (3.10)$$

from Table 3.1, the two corresponding normalized resistances or the two resistances are

$$r = 0.5195 \text{ and } 1.9250, \quad (3.11)$$

or,

$$R = 25.9747 \Omega \text{ and } 96.2475 \Omega. \quad (3.12)$$

In Figure 3.2, the circle with  $RL = S_{11}$  or  $S_{22} = -10$  dB is plotted with a bold line and its radius is approximately one-third of the radius of the biggest circle.

From now on, we will take the expression (3.10) as an approximate demarcation criterion. This implies the following:

- The impedance matching state is acceptable if

$$RL = S_{11} \text{ or } S_{22} < -10 \text{ dB}, \quad (3.13)$$

in which the impedance is close enough to the reference impedance of  $50 \Omega$ , or the center of the Smith Chart.

- On the contrary, the impedance matching state is unacceptable if

$$RL = S_{11} \text{ or } S_{22} > -10 \text{ dB}, \quad (3.14)$$

in which case the impedance is far from the reference impedance  $50 \Omega$ , or the center of the Smith Chart.

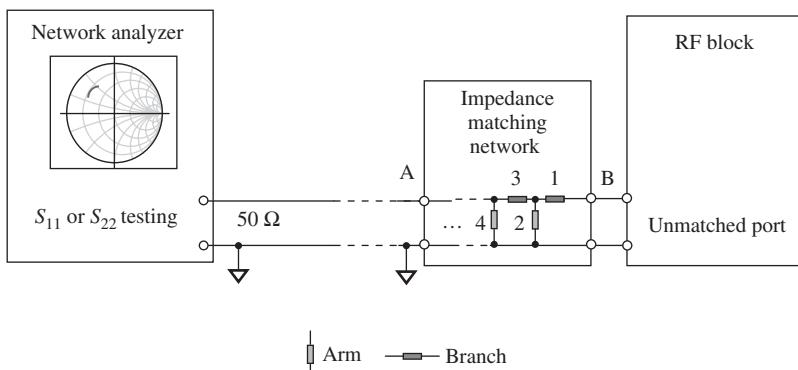
### 3.2.3 Implementation of an Impedance Matching Network

An original RF block usually has an impedance not equal to the reference impedance of  $50 \Omega$ , which is expected to be matched. Therefore, an impedance matching network must be inserted between the original RF block and a network analyzer that provides a  $50\text{-}\Omega$  reference impedance.

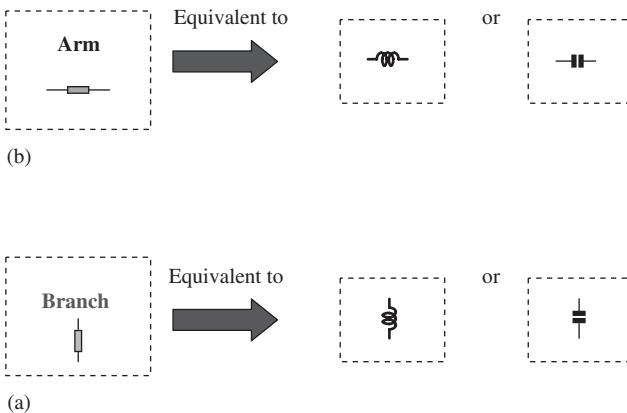
Figure 3.3 shows a setup for the implementation of the impedance matching network. It consists of three portions: a network analyzer to read the return loss or impedance; the original RF block to be matched; and the impedance matching network to be built.

Impedance matching is a process of constructing an impedance network and is executed through the adjustment of the return loss on the Smith Chart.

The impedance matching network consists of arms and branches. The arm is a circuit segment in series and looks like a “horizontal” component, while the branch is a circuit segment in parallel and looks like a “vertical” component. Each arm or branch contains a capacitor or inductor as shown in Figure 3.4. In the narrow-band case, each arm or branch contains only one part, either a capacitor or an inductor.



**Figure 3.3.** Setup for impedance matching from input impedance of an original RF block to  $50 \Omega$ .



**Figure 3.4.** Representation of an arm or a branch consisting of one part. (a) An arm consists of one part in the impedance matching network; it is a connected component in horizontal direction (in series). (b) A branch consists of one part in the impedance matching network; it is a connected component in vertical direction (in parallel).

In Figure 3.3, the arms or the branches, marked 1, 2, 3, 4, and so on, are inserted into the impedance matching network one by one sequentially. At the beginning, the impedance matching network is empty so that point A is directly connected to point B. The  $S_{11}$  or impedance trace is displayed on the screen of the network analyzer. It represents the original impedance of the RF block. As the first arm marked with 1 is inserted between points A and B, the  $S_{11}$  or impedance trace will move to another location on the Smith Chart. As further branches or arms are inserted into the impedance matching network, the  $S_{11}$  or impedance trace will move to other locations on the Smith Chart accordingly. The  $S_{11}$  or impedance trace is expected to ultimately move to the center of Smith Chart, that is,  $50 \Omega$ . In this case, the process of impedance matching is complete and the resultant entity of arms and branches is the desired impedance matching network. Usually more than one arm or branch must be added so that the final impedance, say,  $50 \Omega$ , at point A can be reached.

### 3.3 IMPEDANCE MATCHING NETWORK BUILT BY ONE PART

#### 3.3.1 One Part Inserted into Impedance Matching Network in Series

When the impedance matching network is inserted with one part in series, it is preferable to use impedance, instead of admittance, to describe the variation of its electrical characteristics. We have

$$Z_0 = R_0 + jX_0, \quad (3.15)$$

$$Z = Z_0 + \Delta Z = R + jX, \quad (3.16)$$

where

$Z_0$  = the original impedance before any part is inserted,

$R_0$  = the real part of  $Z_0$ ,

$X_0$  = the imaginary part of  $Z_0$ ,

$Z$  = the resultant impedance after the part is inserted,

$\Delta Z$  = the variation of impedance,

$R$  = the resistance of resultant impedance, or real part of  $Z$ , and

$X$  = the reactance of resultant impedance, or imaginary part of  $Z$ .

If the part inserted in the impedance matching network is either an ideal inductor or an ideal capacitor, then only the imaginary part of the resultant impedance is changed, that is,

$$\Delta R = R - R_o = 0, \quad (3.17)$$

$$\Delta X = X - X_o, \quad (3.18)$$

where

$\Delta R$  = the variation of resistance,

$\Delta X$  = the variation of reactance.

When the impedance matching network is inserted with an inductor in series,

$$\Delta X = \Delta X_L = +L\omega. \quad (3.19)$$

The resultant impedance is, of course, from (3.15) to (3.19),

$$Z = R + jX = Z_o + \Delta Z_L = R_o + j(X_o + \Delta X_L) \quad (3.20)$$

and when the impedance matching network is inserted with a capacitor in series,

$$\Delta X = \Delta X_C = -\frac{1}{C\omega}. \quad (3.21)$$

The resultant impedance is, of course, from (3.14) to (3.18),

$$Z = R + jX = Z_o + \Delta Z_C = R_o + j(X_o + \Delta X_C), \quad (3.22)$$

where

$L$  = the inductance of the part inserted to the impedance matching network,

$C$  = the capacitance of the part inserted to the impedance matching network,

$\omega$  = the angular frequency,

$\Delta Z_L$  = the inserted inductive impedance,

$\Delta Z_C$  = the inserted capacitive impedance,

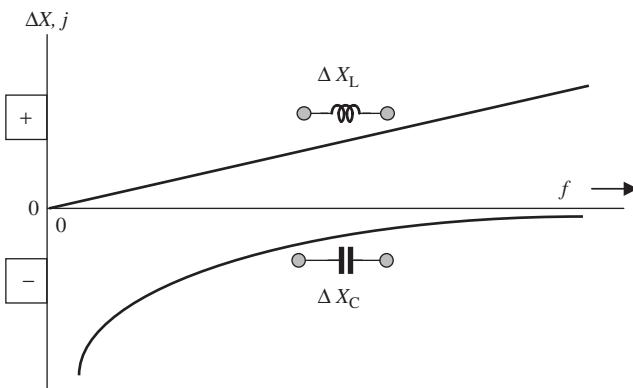
$\Delta X_L$  = the inserted inductive reactance, and

$\Delta X_C$  = the inserted capacitive reactance.

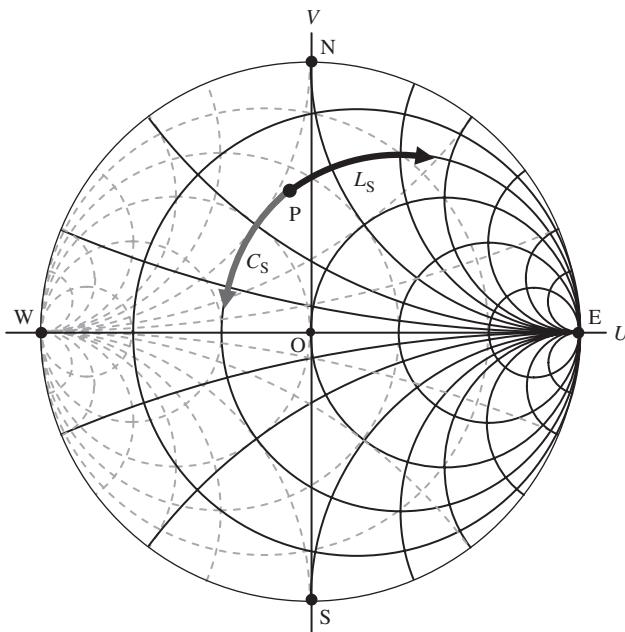
Figure 3.5 shows the inserted reactance  $\Delta X$  and Figure 3.6 shows the direction in which the impedance is pulled on the Smith Chart when the impedance matching network is inserted with  $L$  or  $C$  in series. It can be seen that the variation of impedance in the Smith Chart obeys the following rules of thumb:

When the impedance matching network is inserted with an inductor in series,

- the inserted inductive reactance  $\Delta X = \Delta X_L$  is positive and its magnitude increases as the operating frequency increases;
- the insertion of an inductor  $L_S$  in series results in the original impedance  $P$  moving clockwise along the  $r = \text{constant}$  impedance circle. The movement arc length depends on the value of the inductor.



**Figure 3.5.** Variation of the reactance when an impedance matching network is inserted with one part, either inductor or capacitor, in series.



**Figure 3.6.** Pulled directions of impedance by the insertion of an inductor or a capacitor in series on the Smith Chart.

When the impedance matching network is inserted with a capacitor in series,

- the inserted capacitive reactance  $\Delta X = \Delta X_C$  is negative and its magnitude decreases as the operating frequency increases;
- the insertion of a capacitor  $C_S$  in series results in the original impedance  $P$  moving counterclockwise along the  $r = \text{constant}$  impedance circle. The movement arc length depends on the value of the capacitor.

### 3.3.2 One Part Inserted into the Impedance Matching Network in Parallel

When an impedance matching network is inserted with one part in parallel, it is preferable to use admittance, instead of impedance, to describe the variation of its electrical

characteristics. We have

$$Y_o = G_o + jB_o, \quad (3.23)$$

$$Y = Y_o + \Delta Y = G + jB, \quad (3.24)$$

where

$Y_o$  = the original admittance before the part is inserted,

$G_o$  = the real part of  $Y_o$ ,

$B_o$  = the imaginary part of  $Y_o$ ,

$Y$  = the resultant admittance after the part is inserted,

$\Delta Y$  = the variation of admittance,

$G$  = the conductance of resultant admittance, or real part of  $Y$ , and

$B$  = the susceptance of resultant admittance, or imaginary part of  $Y$ .

If the part inserted into the impedance matching network is either an ideal inductor or an ideal capacitor, then only the imaginary part of the resultant impedance is changed, that is,

$$\Delta G = G - G_o = 0, \quad \Delta B = B - B_o. \quad (3.25)$$

When an impedance matching network is inserted with an inductor in parallel,

$$\Delta B = \Delta B_L = -\frac{1}{L\omega}. \quad (3.26)$$

The resultant admittance is, of course, from (3.23) to (3.25), and (3.28),

$$Y = Y_o + \Delta Y_L = G + jB = G_o + j(B_o + \Delta B_L). \quad (3.27)$$

When impedance matching network is inserted with a capacitor in parallel,

$$\Delta B = \Delta B_C = C\omega. \quad (3.28)$$

The resultant admittance is, of course, from (3.23) to (3.25),

$$Y = Y_o + \Delta Y_C = G + jB = G_o + j(B_o + \Delta B_C), \quad (3.29)$$

where

$L$  = the inductance of the part inserted to the block,

$C$  = the capacitance of the part inserted to the block,

$\omega$  = the angular frequency,

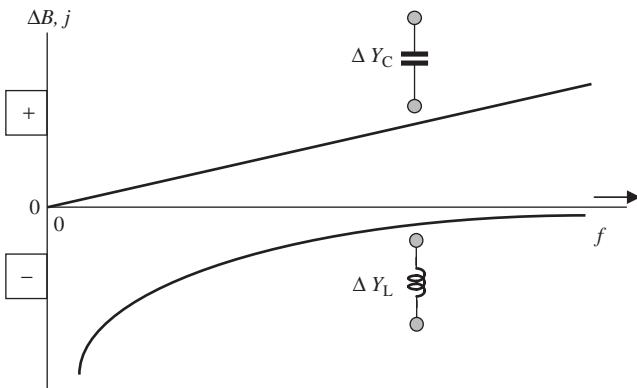
$\Delta Y_L$  = the inserted inductive admittance,

$\Delta Y_C$  = the inserted capacitive admittance,

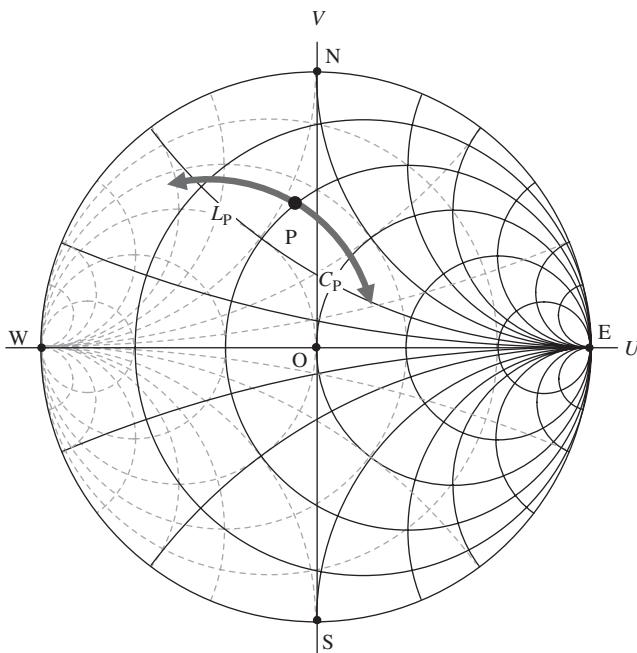
$\Delta B_L$  = the inserted inductive susceptance, and

$\Delta B_C$  = the inserted capacitive susceptance.

Figure 3.7 shows the inserted admittance  $\Delta Y$  and Figure 3.8 shows the pulled direction of impedance on Smith Chart when an impedance matching network is inserted with  $L$  or  $C$  in parallel. It can be seen that the variation of impedance on Smith Chart obeys the following rules of thumb:



**Figure 3.7.** Variation of susceptance when the impedance matching network is inserted with one part, either inductor or capacitor, in parallel.



**Figure 3.8.** Pulled directions of impedance by the insertion of an inductor or a capacitor in parallel on the Smith Chart.

When the impedance matching network is inserted with an inductor in parallel,

- the inserted admittance  $\Delta Y = \Delta Y_L$  is the negative susceptance  $\Delta B_L$  and its magnitude decreases as the operating frequency increases;
- the insertion of an inductor  $L_P$  in parallel results the original impedance  $P$  moving counterclockwise along the circle with  $g = \text{constant}$ . The movement arc length depends on the value of the inductor.

When the impedance matching network is inserted with a capacitor in parallel,

- the inserted admittance  $\Delta Y = \Delta Y_C$  is the positive susceptance  $\Delta B_C$  and its magnitude increases as the operating frequency increases;

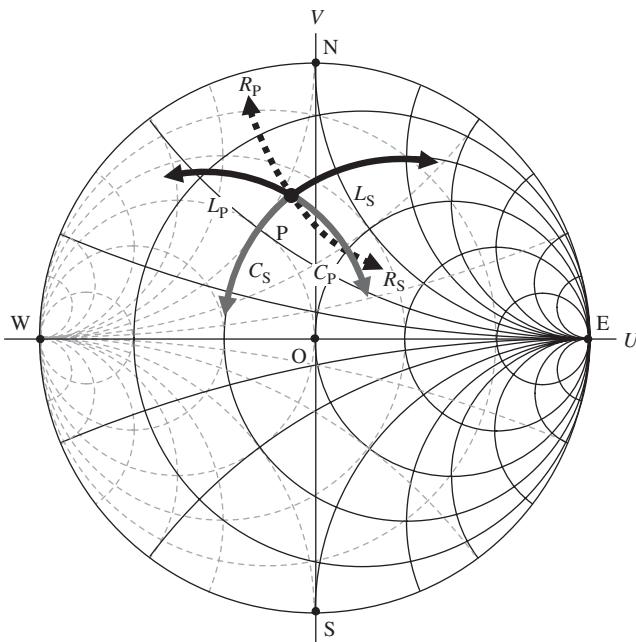


Figure 3.9. Pulled directions of impedance by the insertion of  $L$ ,  $C$ , or  $R$  on the Smith Chart.

- the insertion of a capacitor in parallel  $C_P$  results the original impedance  $P$  moving clockwise along the circle with  $g = \text{constant}$ . The movement arc length depends on the value of capacitor.

As a matter of fact, Figures 3.6 and 3.8 can be combined into Figure 3.9, where the pulled directions due to the insertion of a resistor in series and in parallel are also depicted, though it makes sense only theoretically and not in practical engineering design.

Theoretically,

- the insertion of a resistor  $R_S$  in series results in the original impedance  $P$  moving along the arc with  $x = \text{constant}$  to a higher resistance circle. The distance moved depends on the value of the resistor;
- the insertion of a resistor  $R_P$  in parallel results in the original impedance  $P$  moving along the arc with  $x = \text{constant}$  arc to a lower resistance circle. The distance moved depends on the value of the resistor.

Obviously, there is only a very slim chance of being able to build an impedance matching network containing only one part because it is impossible in most cases to pull the original impedance to the reference impedance of  $50 \Omega$  with only one part, either inductor or capacitor, unless the original impedance is located on both circles with  $r = 1$  and  $g = 1$ , or, in two narrow ring areas adjacent to both circles.

However, it is possible to use two parts to match the original impedance to a desired value if the block is operating for a narrow-band system. In most of RF circuit block design, it is quite conventional to implement an impedance matching network by two parts.

### 3.4 IMPEDANCE MATCHING NETWORK BUILT BY TWO PARTS

#### 3.4.1 Regions in a Smith Chart

The topology of a matching network depends on the target impedance to be matched and the original impedance. Usually, if the target impedance is the standard reference  $50 \Omega$ , then the topology of the impedance matching network depends only on the location of the original impedance to be matched on the Smith Chart.

In order to be able to formulize the values of the impedance matching parts, it is suggested that the Smith Chart be divided into four regions (Li, 2005). Not only the topology but also the value of the parts in an impedance matching network can be easily determined and calculated. Figure 3.10 shows the demarcation of these four regions on the Smith Chart and Table 3.2 lists the range of impedance in these four regions.

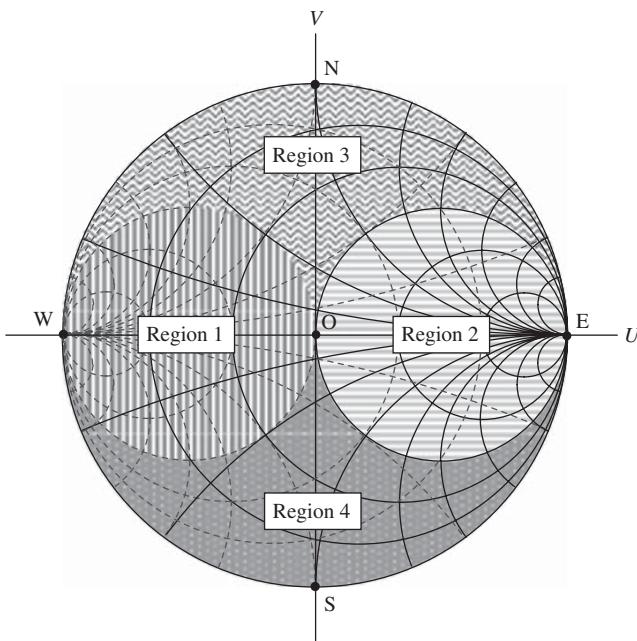


Figure 3.10. Demarcation of four regions on the Smith Chart.

TABLE 3.2. Range of Impedance in Four Regions on a Smith Chart

Region 1	Region 2	Region 3	Region 4
Low resistance or high conductance	High resistance or low conductance	Low resistance and low conductance	Low resistance and low conductance
$r < 1$	$r > 1$ ,	$r < 1$	$r < 1$
$x <  0.5 $	$-\infty < x < +\infty$	$x > 0$	$x < 0$
$g > 1$	$g < 1$	$g < 1$	$g < 1$
$-\infty < b < +\infty$	$b <  0.5 $	$b < 0$	$b > 0$

### 3.4.2 Values of Parts

Let us denote the original impedance to be matched as

$$Z_m = R_m + jX_m, \quad (3.30)$$

where

- $Z_m$  = the original impedance to be matched,
- $R_m$  = the resistance of original impedance, and
- $X_m$  = the reactance of original impedance.

Figure 3.11 shows that in all regions there are two ways to pull the original impedance, at  $P_1$ ,  $P_2$ ,  $P_3$ , and  $P_4$  to the center of the Smith Chart, that is, O, by the addition of two types of passive parts, the inductor and the capacitor. The subscript to P represents the region where the impedance is located.

From Figure 3.11 it can be seen that there are two ways to pull  $P_1$  to the center of Smith Chart, that is,  $50 \Omega$ :

1. In Figure 3.11(a),  $P_1$  is pulled to A by the addition of a capacitor  $C_S$  in series first, and then from A to O by the addition of an inductor  $L_P$  in parallel.
2. In Figure 3.11(a),  $P_1$  is pulled to B by the addition of an inductor  $L_S$  in series first, and then from B to O by the addition of a capacitor  $C_P$  in parallel.

There are two ways to pull  $P_2$  to the center of the Smith Chart,  $50 \Omega$ :

1. In Figure 3.11(b),  $P_2$  is pulled to C by the addition of an inductor  $L_P$  in parallel first, and then from C to O by the addition of a capacitor  $C_S$  in series.
2. In Figure 3.11(b),  $P_2$  is pulled to D by the addition of a capacitor  $C_P$  in parallel first, and then from D to O by the addition of an inductor  $L_S$  in series.

There are two ways to pull  $P_3$  to the center of the Smith Chart,  $50 \Omega$ :

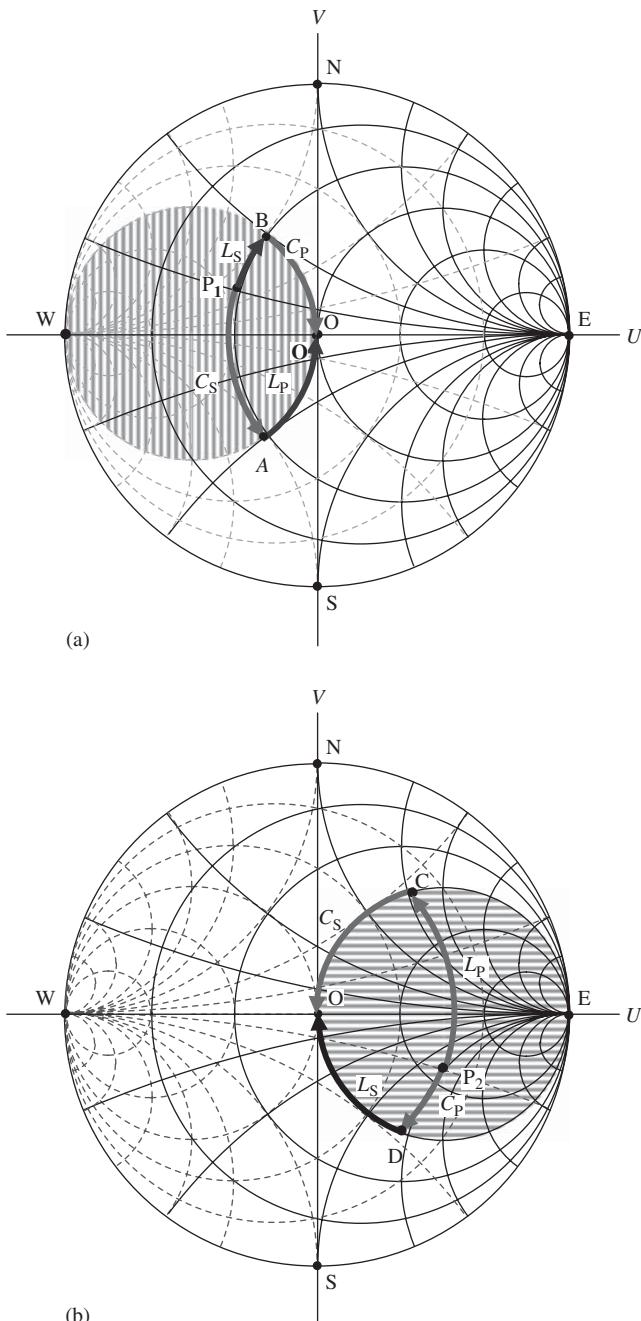
1. In Figure 3.11(c),  $P_3$  is pulled to B by the addition of a capacitor  $C_S$  in series first, and then from B to O by the addition of a capacitor  $C_P$  in parallel.
2. In Figure 3.11(c),  $P_3$  is pulled to C by the addition of a capacitor  $C_P$  in parallel first, and then from C to O by the addition of a capacitor  $C_S$  in series.

There are two ways to pull  $P_4$  to the center of the Smith Chart,  $50 \Omega$ :

1. In Figure 3.11(d),  $P_4$  is pulled to D by the addition of an inductor  $L_P$  in parallel first, and then from D to O by the addition of an inductor  $L_S$  in series.
2. In Figure 3.11(d),  $P_4$  is pulled to A by the addition of an inductor  $L_S$  in series first, and then from A to O by the addition of an inductor  $L_P$  in parallel.

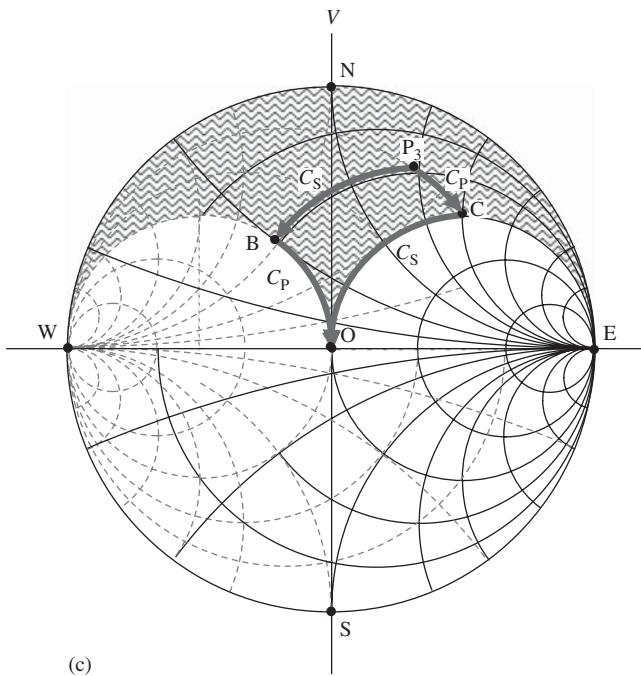
Three common rules or features can be summarized from the description above:

1. The first part in a two-part impedance matching network is to pull the original impedance to either the circle with  $g = 1$  or the circle with  $r = 1$ .

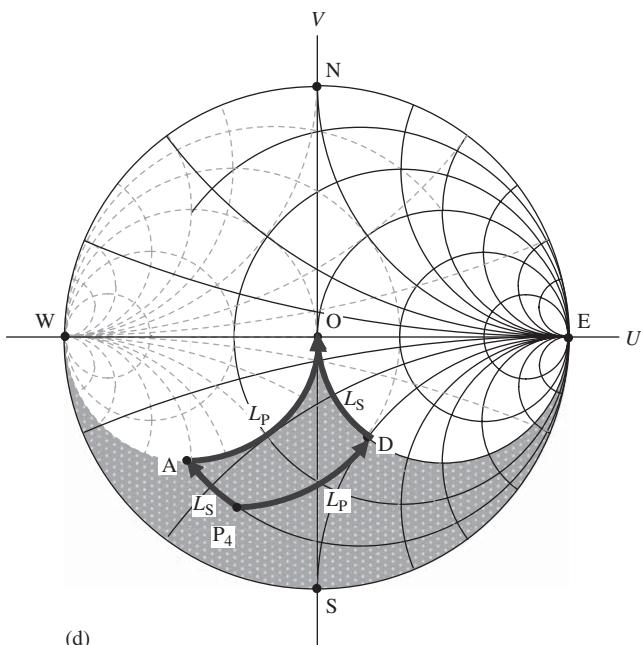


**Figure 3.11.** (a) Two ways to pull the original impedance (a)  $P_1$  in region 1 (b)  $P_2$  in region 2 (c)  $P_3$  in region 3 (d)  $P_4$  in region 4 to the center of the Smith Chart O by the addition of two passive parts.

2. The second part in a two-part impedance matching network is to pull the impedance from either the circle with  $g = 1$  or the circle with  $r = 1$  after it is pulled by first part, to the standard reference impedance  $50\ \Omega$ .
3. One of the two matching parts is in series and the other is in parallel, and vice versa. If the impedance to be matched is located in region 1, the designer must select the first part in series and the second part in parallel. On the contrary, if the



(c)



(d)

Figure 3.11. (Continued)

impedance to be matched is located in region 2, the designer must select the first part in parallel and the second part in series. However, the designer can select the first or second part in series or in parallel if the impedance to be matched is located in region 3 or 4.

On the basis of the common rules or features mentioned above, the type of the two parts can be selected and the values of the matching parts, one in series and the other in parallel, can be found. This is a tedious mathematical process and will be introduced in Appendix 3.A.2. We will introduce the appropriate formulas from Appendix 3.A.2 so that the values of the two impedance matching parts can be calculated when the original impedance is located in any of the four regions. The following four cases correspond to the four regions where the original impedance is located.

### 1. The original impedance to be matched is located in Region 1

There are two ways to pull  $P_1$  to the center of Smith Chart,  $50 \Omega$ :

1. By the addition of a capacitor  $C_S$  in series first, and then by the addition of an inductor  $L_P$  in parallel.

$$C_S = \frac{C_m}{C_m \omega \sqrt{R_m(Z_o - R_m)} - 1}, \quad (3.31)$$

$$L_P = \frac{R_m Z_o}{\omega \sqrt{R_m(Z_o - R_m)}}, \quad (3.32)$$

where

$C_S$  = the capacitance of the impedance matching part, a capacitor in series,

$L_P$  = the inductance of the impedance matching part, an inductor in parallel,

$C_m$  = the capacitance of original impedance,

$\omega$  = the angular operating frequency,

$Z_o$  = the standard reference resistance,  $50 \Omega$  usually, and

$R_m$  = the resistance of original impedance.

2. By the addition of an inductor  $L_S$  in series first, and then by the addition of a capacitor  $C_P$  in parallel.

$$L_S = \frac{\sqrt{R_m(Z_o - R_m)} - X_m}{\omega}, \quad (3.33)$$

$$C_P = \frac{\sqrt{R_m(Z_o - R_m)}}{R_m Z_o \omega}, \quad (3.34)$$

where

$L_S$  = the inductance of impedance matching part, an inductor in series,

$C_P$  = the capacitance of impedance matching part, a capacitor in parallel,

$\omega$  = the angular operating frequency,

$Z_o$  = the standard reference resistance,  $50 \Omega$  usually,

$R_m$  = the resistance of original impedance, and

$X_m$  = reactance of original impedance.

## 2. The original impedance to be matched is located in Region 2

There are two ways to pull  $P_2$  to the center of Smith Chart, 50  $\Omega$ :

1. By the addition of an inductor  $L_P$  in parallel first, and then by the addition of a capacitor  $C_S$  in series.

$$L_P = \frac{X_m Z_o + \sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{(R_m - Z_o) \omega}, \quad (3.35)$$

$$C_S = \frac{R_m / \omega}{\sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}. \quad (3.36)$$

2. By the addition of a capacitor  $C_P$  in parallel first, and then by the addition of an inductor  $L_S$  in series.

$$C_P = \frac{(R_m - Z_o) / \omega}{\sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)} - X_m Z_o}, \quad (3.37)$$

$$L_S = \frac{\sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{R_m \omega}. \quad (3.38)$$

## 3. The original impedance to be matched is located in Region 3

There are two ways to pull  $P_3$  to the center of Smith Chart, 50  $\Omega$ :

1. By the addition of a capacitor  $C_S$  in series first, and then by the addition of a capacitor  $C_P$  in parallel.

$$C_S = \frac{C_m}{C_m \omega \sqrt{R_m (Z_o - R_m)} - 1}, \quad (3.39)$$

$$C_P = \frac{\sqrt{R_m (Z_o - R_m)}}{R_m Z_o \omega}. \quad (3.40)$$

2. By the addition of a capacitor  $C_P$  in parallel first, and then by the addition of a capacitor  $C_S$  in series.

$$C_P = \frac{(R_m - Z_o) / \omega}{\sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)} - X_m Z_o}, \quad (3.41)$$

$$C_S = \frac{R_m / \omega}{\sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}. \quad (3.42)$$

#### 4. The original impedance to be matched is located in Region 4

There are two ways to pull  $P_4$  to the center of Smith Chart, 50  $\Omega$ :

1. By the addition of an inductor  $L_P$  in parallel first, and then by the addition of an inductor  $L_S$  in series.

$$L_S = \frac{\sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{R_m \omega}, \quad (3.43)$$

$$L_P = \frac{X_m Z_o + \sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{(R_m - Z_o) \omega}. \quad (3.44)$$

2. By the addition of an inductor  $L_S$  in series first, and then by the addition of an inductor  $L_P$  in parallel.

$$L_P = \frac{R_m Z_o}{\omega \sqrt{R_m (Z_o - R_m)}}, \quad (3.45)$$

$$L_S = \frac{\sqrt{R_m (Z_o - R_m)} - X_m}{\omega}. \quad (3.46)$$

Counting from expressions (3.31) to (3.46), there are sixteen expressions in total. In fact, they are only eight independent equations.

Before these equations were developed, the designer had to rely only on the Smith Chart and calculate the values of impedance matching parts by carefully estimating the location of the impedance and then measuring the arc length on the Smith Chart. These calculations always gave an inaccurate value, especially when the impedance to be measured was located in the high impedance area on the Smith Chart. In addition, the calculations were quite tedious.

In terms of these eight equations, it is quite convenient to calculate the values of the impedance matching parts for all cases no matter which region the original impedance  $Z_m$  to be matched is located in, as long as the original impedance  $Z_m$  and the operating frequency  $\omega$  are known. Armed with these eight equations, the designer does not need assistance from the Smith Chart.

Summarizing, in each region there are two options corresponding to the two topologies of the impedance matching networks as shown in Figure 3.11. It should be noted that the choice of whether to use an inductor or capacitor for  $X_1$  or  $X_2$  depends on the region where the original impedance  $Z_m$  or  $R_m$  and  $X_m$  are located. Figure 3.11 shows that

- if the original impedance is located in regions 1 or 2, the two matched parts consist of one capacitor and one inductor;
- if the original impedance is located in region 3, both impedance matching parts are capacitors;
- if the original impedance is located in region 4, both impedance matching parts are inductors.

### 3.4.3 Selection of Topology

In general, there are three factors that must be considered in the selection of the topology for an impedance matching network.

#### 1. Consideration of the availability of topology.

Table 3.3 lists eight possible topologies for an impedance matching network containing two parts.

As discussed above, only two topologies are available to match a specific original impedance in the Smith Chart to the center or the reference impedance of  $50 \Omega$ . In other words, not all possible topologies are available for the implementation of a two-part impedance matching network. The availability of the topology depends on the location of the original impedance on the Smith Chart. An impedance matching network with a specific topology is able to match an original impedance to  $50 \Omega$  only if the original impedance is located in some specific regions of the Smith Chart. These regions are called *applied regions* of the specific topology. On the contrary, an impedance matching network may be unable to match the original impedance to  $50 \Omega$  if this original impedance is located in some specific regions of the Smith Chart. These regions are called *prohibited regions* of the specific topology.

Appendix 3.A.3 presents the applied area and the prohibited area on a Smith Chart for a specific topology of an impedance matching network containing two parts, which is also summarized in Table 3.4.

These restrictions of impedance matching networks containing two parts can be avoided if the designer selects a correct topology based on Table 3.4.

#### 2. Considerations of cost.

Obviously, from the viewpoint of saving cost, the last two topologies 7 and 8 in Table 3.4 with two inductors are not good choices because they are too expensive. These two topologies should be avoided.

Fortunately, some other topologies are available if the original impedance to be matched is located in region 4. Instead of two inductors, the impedance matching network can be built by one inductor and one capacitor as shown in Figure 3.12, where the original impedance  $P_4$  can be pulled to the center of the Smith Chart by one inductor and one

TABLE 3.3. Eight Possible Topologies of an Impedance Matching Network Containing Two Passive Parts

(1)	$C_P - L_S$
(2)	$L_S - C_P$
(3)	$C_S - L_P$
(4)	$L_P - C_S$
(5)	$C_P - C_S$
(6)	$C_S - C_P$
(7)	$L_P - L_S$
(8)	$L_S - L_P$

Note 1: The first part is connected to the original impedance to be matched, and the second part is connected to the standard reference impedance,  $50 \Omega$ .

Note 2: Subscript "P" stands for "in parallel" and subscript "S" stands for "in series."

TABLE 3.4. Applied and Prohibited Regions of an Impedance Matching Network with Specific Topology

Topology	$Z_m$ Is Located in the Applied Region	$Z_m$ is Located in the Prohibited Region
(1) $C_P - L_S$	Regions 2,3	Regions 1,4
(2) $L_S - C_P$	<b>Regions 1,4</b>	Regions 2,3
(3) $C_S - L_P$	Regions 1,3	Regions 2,4
(4) $L_P - C_S$	<b>Regions 2,4</b>	Regions 1,3
(5) $C_P - C_S$	Region 3	Regions 1,2,4
(6) $C_S - C_P$	Region 3	Regions 1,2,4
(7) $L_P - L_S$	<b>Region 4</b>	Regions 1,2,3
(8) $L_S - L_P$	<b>Region 4</b>	Regions 1,2,3

Note 1:  $Z_m$  = the original impedance to be matched.

Note 2: In the column topology, the first part is connected to the original impedance to be matched and the second part is connected to the standard reference impedance,  $50 \Omega$ .

Note 3: Subscript P stands for in parallel and subscript S stands for in series.

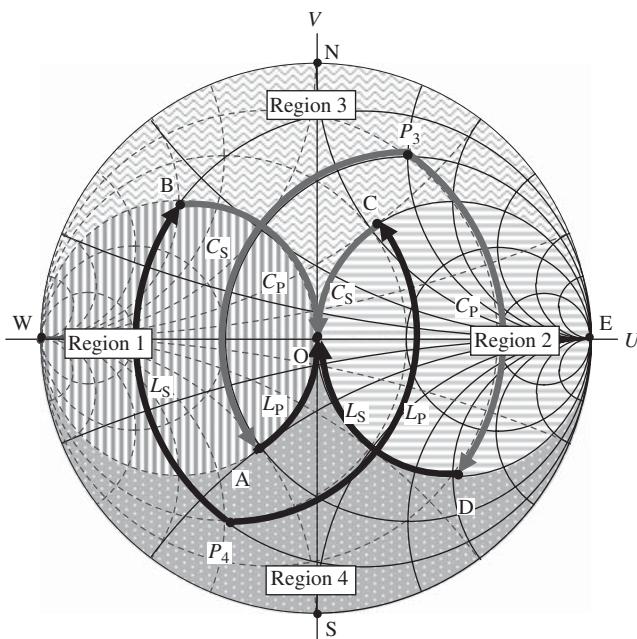


Figure 3.12. Alternative way to pull the original impedances  $P_3$  and  $P_4$  to the center of Smith Chart O (one capacitor and one small inductor).

capacitor by two options. This reduction in the number of inductors is a significant improvement in RF circuit design, especially in RFIC design. Figure 3.12 also shows that instead of two capacitors, one inductor and one capacitor can form an impedance matching network if the original impedance to be matched is located in region 3.

It provides an alternative method though it is not as good as the case of implementing an impedance matching network by only two capacitors.

In real-life engineering design, most impedance matching networks are built and inserted between a source and the input of a device, or between the output of a device and a load.

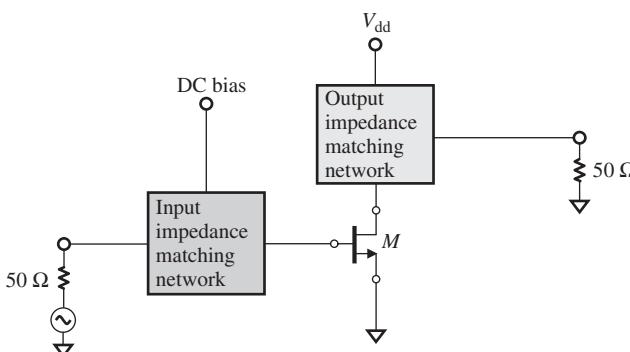
Figure 3.13 shows the input and output impedance matching networks located before and after the device, respectively. The selection of topology for input and output impedance matching networks in such a case is somewhat different from that in general cases.

### 3. Consideration of DC Blocking, DC Feeding, and DC short-Circuit.

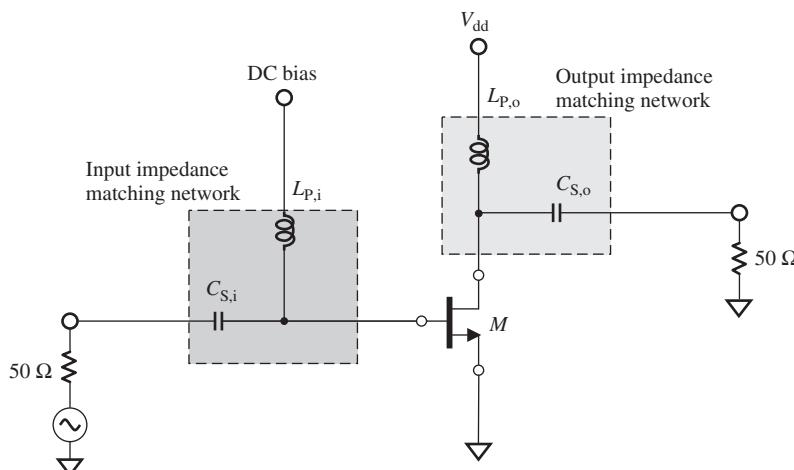
The input and output impedance matching network of a device must play the role not only of impedance matching but also of DC blocking and DC feeding. By looking at the topologies listed in Table 3.4, only topologies 4 ( $L_P - C_S$ ) and 7 ( $L_P - L_S$ ) are available for DC feeding. However, topology 7 ( $L_P - L_S$ ) cannot block DC; on the other hand, it consists of two inductors, so it is immediately discarded as an option considering DC blocking and cost saving.

It is therefore concluded that only topology 4 ( $L_P - C_S$ ) is the exclusive choice for the input and output impedance matching network of the device. Figure 3.14 illustrates the conclusions on implementing an input and output impedance matching network by two parts.

Also, it should be noted that the applied regions of topology 4 ( $L_P - C_S$ ) are regions 2 and 4. If the original impedance is not located in region 2 or 4, it can only implement



**Figure 3.13.** Location of input and output impedance matching network for a device.



**Figure 3.14.** Conclusions on implementing an input and output impedance matching network by two parts.

an impedance matching network containing three parts, which will be discussed in the following subsection.

### 3.5 IMPEDANCE MATCHING NETWORK BUILT BY THREE PARTS

An impedance matching network constructed by three parts, that is, a combination of capacitors and inductors, can reduce the restrictions in an impedance matching network built by two parts significantly. Of course, the price to pay is the cost of an extra part.

#### 3.5.1 “Π” Type and “T” Type Topologies

All the topologies of a matching network constructed by three parts can be categorized into Π and T types.

Π type or T type impedance matching network contains three parts, consisting of either capacitors or inductors. There are eight possible alternative topologies in Π -type or T-type impedance matching networks, which are plotted in Appendix 3.A.6 and listed in Table 3.5.

#### 3.5.2 Recommended Topology

##### 1. Consideration of cost

From the viewpoint of cost savings, the topologies listed in Table 3.5 containing two or three inductors can be ignored. In other words, only topologies 1, 2, 3, and 5 will be considered in the circuit design. They are listed in Table 3.6.

##### 2. Consideration of the availability of topology

As in the case of impedance matching networks built by two parts, there are also restrictions in three-part impedance matching networks.

An impedance matching network containing three parts with a specific topology is able to match an impedance to  $50 \Omega$  only if the original impedance to be matched is located in some specific regions of the Smith Chart. Appendix 3.A.4 presents the applied

TABLE 3.5. Possible Topologies of a Three-Part Impedance Matching Network

	Π Type	T Type
(1)	$C_{P1}-C_S-C_{P2}$	$C_{S1}-C_P-C_{S2}$
(2)	$L_{P1}-C_S-C_{P2}$	$L_{S1}-C_P-C_{S2}$
(3)	$C_{P1}-L_S-C_{P2}$	$C_{S1}-L_P-C_{S2}$
(4)	$L_{P1}-L_S-C_{P2}$	$L_{S1}-L_P-C_{S2}$
(5)	$C_{P1}-C_S-L_{P2}$	$C_{S1}-C_P-L_{S2}$
(6)	$L_{P1}-C_S-L_{P2}$	$L_{S1}-C_P-L_{S2}$
(7)	$C_{P1}-L_S-L_{P2}$	$C_{S1}-L_P-L_{S2}$
(8)	$L_{P1}-L_S-L_{P2}$	$L_{S1}-L_P-L_{S2}$

Note 1: In the topology list, the first part is connected to the original impedance to be matched, and the third part is connected to the standard reference impedance,  $50 \Omega$ .

Note 2: Subscript P stands for in parallel and subscript S stands for in series.

TABLE 3.6. Applied and Prohibited Regions of an Impedance Matching Network of Specific Topologies with the Consideration of Cost (from there the Impedance  $Z_m$  will be Matched to 50  $\Omega$ )

Topology	Applied Regions	Prohibited Regions
<i><math>\Pi</math> Type</i>		
(1) $C_{P1}-C_S-C_{P2}$	Region 3	Region 1, 2, 4
(2) $L_{P1}-C_S-C_{P2}$	<b>Region 2, 3, 4</b>	<b>Region 1</b>
(3) $C_{P1}-L_S-C_{P2}$	<b>Region 1,2, 3, 4</b>	None
(5) $C_{P1}-C_S-L_{P2}$	<b>Region 1, 3</b>	<b>Region 2, 4</b>
<i>T Type</i>		
(1) $C_{S1}-C_P-C_{S2}$	Region 3	Region 1, 2, 4
(2) $L_{S1}-C_P-C_{S2}$	<b>Region 1, 3, 4</b>	<b>Region 2</b>
(3) $C_{S1}-L_P-C_{S2}$	<b>Region 1, 2, 3, 4</b>	None
(5) $C_{S1}-C_P-L_{S2}$	<b>Region 2, 3</b>	<b>Region 1, 4</b>

Note 1: In the topology list, the first part is connected to the original impedance to be matched and the third part is connected to the standard reference impedance, 50  $\Omega$ .

Note 2: Subscript P stands for in parallel and subscript S stands for in series.

areas and the prohibited areas on a Smith Chart for a specific topology of the impedance matching network containing three parts; this is also summarized in Table 3.6. Usually, the original impedance  $Z_m$  is located in regions 1, 2, or 4. Consequently, the topologies that we are interested in are 2, 3, and 5 for both  $\Pi$  and T type of impedance matching networks. Topology 1 is immediately discarded because its prohibited regions are 1, 2, and 4.

### 3. Consideration of DC blocking, DC feeding, and DC short-circuit

Figures 3.15–3.20 plot the topologies 2, 3, and 5 for both  $\Pi$  and T type of impedance matching networks. Table 3.7 lists the main performance of these plots on DC blocking, DC feeding, and DC short-circuit.

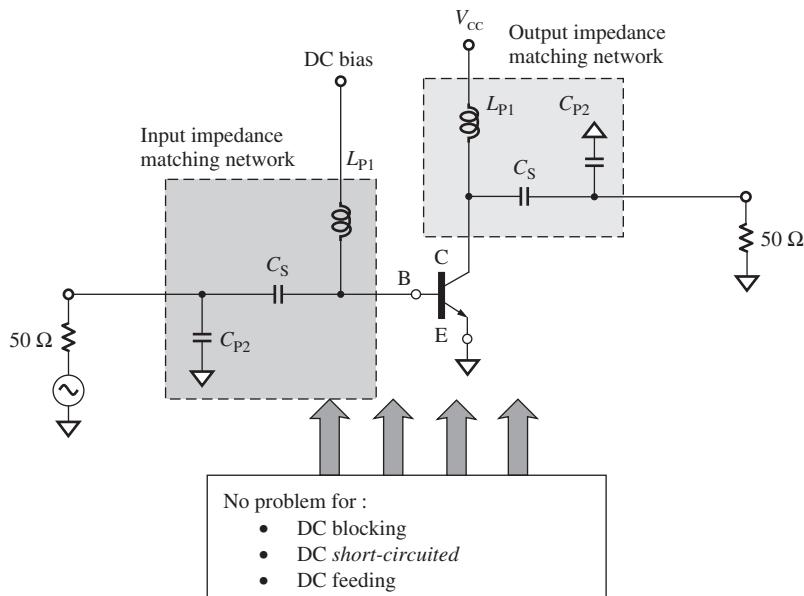
From Table 3.7 it can be seen that in  $\Pi$  type of impedance matching network, only topology 2 ( $L_{P1}-C_S-C_{P2}$ ) has no problem on DC blocking, DC feeding, and DC short-circuit. Topology 3 ( $C_{P1}-L_S-C_{P2}$ ) has problem on DC blocking and in addition, it needs two RF chokes, or two  $L_{\text{infinitive}}$  inductors. Topology 5 ( $C_{P1}-C_S-L_{P2}$ ) needs two RF chokes or two  $L_{\text{infinitive}}$  inductors. The impedance of a RF choke or a  $L_{\text{infinitive}}$  is an inductor approaches to infinitive over the operating frequency range.

In T type of impedance matching network, the topologies 2 ( $L_{P1}-C_S-C_{P2}$ ), 3 ( $C_{P1}-L_S-C_{P2}$ ), and 5 ( $C_{P1}-C_S-L_{P2}$ ) have no problem in DC blocking, DC feeding, and DC short-circuit. However, all of them need two RF chokes or two  $L_{\text{infinitive}}$  inductors so that they become a four- or five-part impedance matching network indeed.

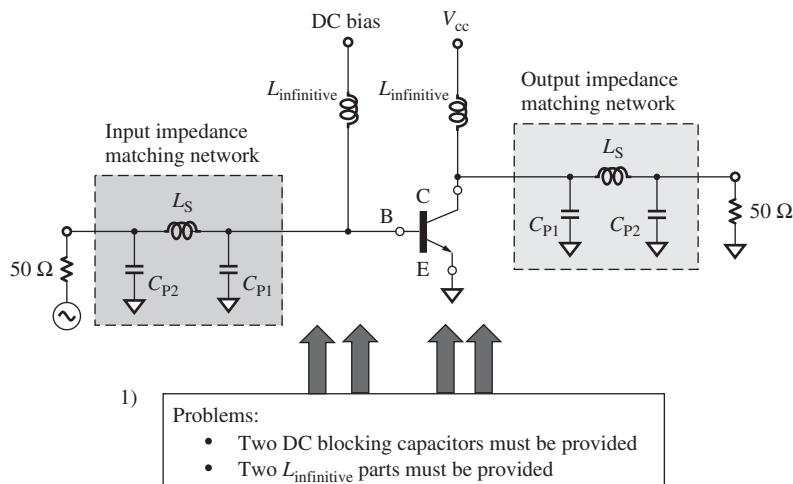
Obviously, the  $\Pi$  type of topology 2 ( $L_{P1}-C_S-C_{P2}$ ) is the best choice among the three-part impedance matching networks.

## 3.6 IMPEDANCE MATCHING WHEN $Z_s$ OR $Z_L$ IS NOT 50 $\Omega$

In the actual RF circuit design, most of input and output impedances of blocks are matched to 50  $\Omega$ . The main reason is that the standard reference impedance, 50  $\Omega$ , is adopted in almost all the equipment for RF block testing. In the testing of an RF block, the input or output impedance of the desired test RF unit or block must be matched to the input or output impedance of the test equipment. Otherwise, the test results will be



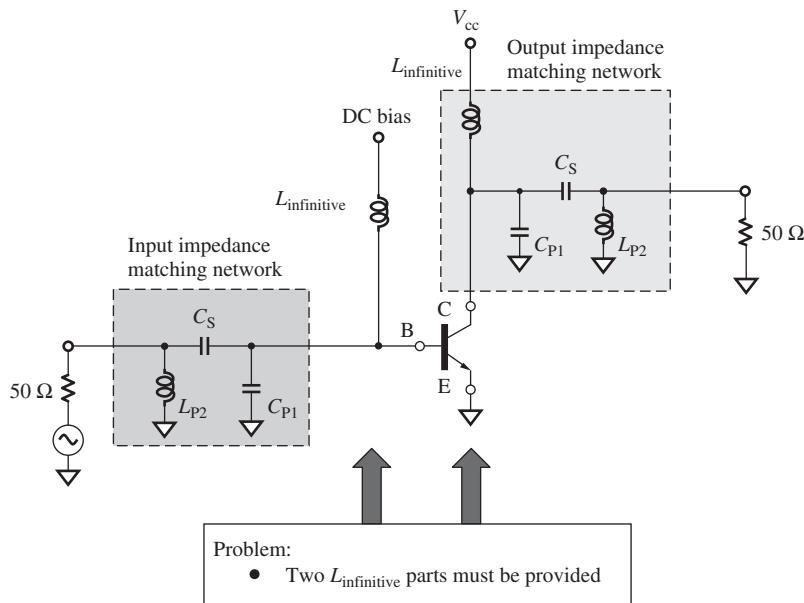
**Figure 3.15.**  $L_{P1}$ - $C_S$ - $C_{P2}$  topology of impedance matching network available for input and output of a device.



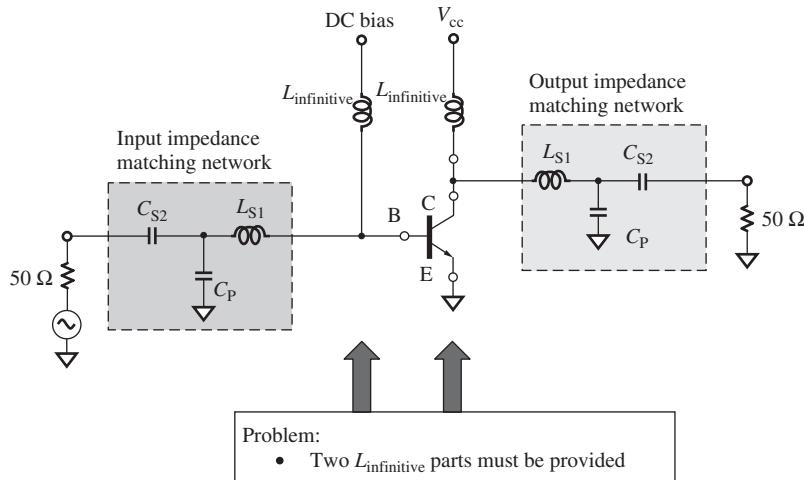
**Figure 3.16.**  $C_{P1}$ - $L_S$ - $C_{P2}$  topology of impedance matching network problematic for input and output of a device.

incorrect since the test objective in RF block testing is always to measure power instead of voltage. This is a special feature in testing RF blocks.

As an example, Figure 3.21(a) shows a block diagram for the combination of a LNA low (noise amplifierNoise Amplifier) and a mixer block. Obviously, it is a reasonable and typical plan. The input and output impedance matching networks are set before and after the LNA or mixer, respectively. The design or testing of the LNA or mixer can be conducted independently if the dashed line in Figure 3.21(a) is removed.

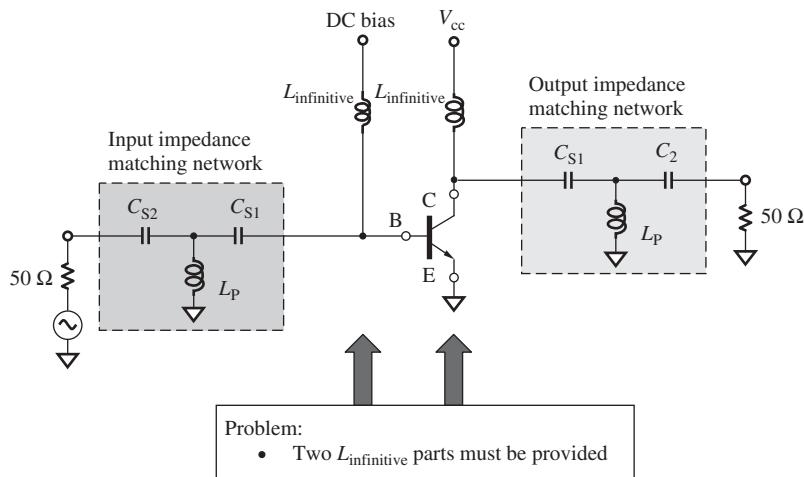


**Figure 3.17.**  $C_{P1}-C_S-L_{P2}$  topology of impedance matching network problematic for input and output of a device.

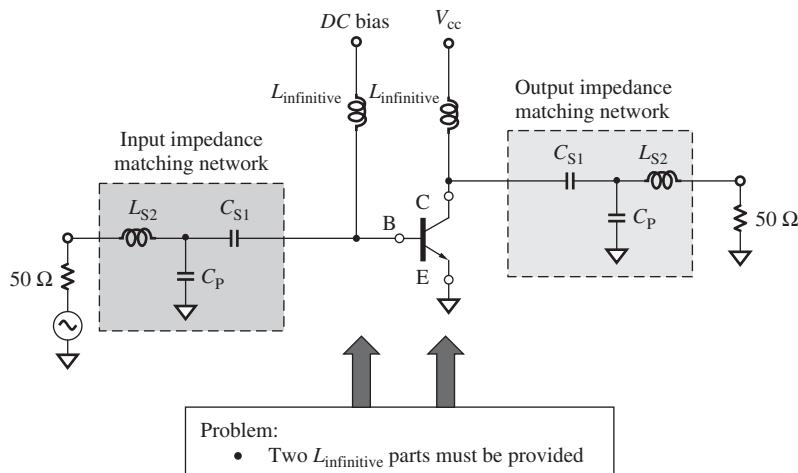


**Figure 3.18.**  $L_{S1}-C_P-C_{S2}$  topology of impedance matching network problematic for input and output of a device.

However, this is not an optimized plan, but rather a conservative one. As shown in Figure 3.21(a), there are two impedance matching networks between the LNA output and the mixer input. The first impedance matching network is to match the output impedance of LNA down to 50  $\Omega$  so that its input impedance is conjugate-matched to the output impedance of LNA,  $Z_{LNA,o}$ , and its output impedance is 50  $\Omega$ . The second impedance matching network is to match the input impedance of mixer from 50  $\Omega$  so that its input impedance is 50  $\Omega$  and its output impedance is conjugate-matched to the input impedance of mixer,  $Z_{mix,i}$ . Is it possible to combine these two impedance matching networks into



**Figure 3.19.**  $C_{S1}-L_P-C_{S2}$  topology of impedance matching network problematic for input and output of a device.



**Figure 3.20.**  $C_{S1}-L_P-C_{S2}$  topology of impedance matching network problematic for input and output of a device.

one? Yes, indeed. One of the networks can be omitted if the two are combined into one. Obviously, the advantages are the following:

- It is more cost effective.
- Insertion loss can be reduced.
- Also, less noise can be expected.

Figure 3.21(b) shows the block diagram after combining the two impedance matching networks between the LNA and mixer into one. The combined impedance matching network is to conjugate-match the output impedance of LNA directly to the input impedance of the mixer. Therefore, its input impedance must be  $Z_{\text{LNA},o}^*$  and its output impedance must be  $Z_{\text{mix},i}^*$ .

TABLE 3.7. Performance of the Topologies 2, 3, and 5, for Both  $\Pi$  and T Type of Impedance Matching Networks on DC blocking, DC feeding, and DC Short-Circuit

Topology	No Problem	Problems
$\Pi$ Type		
(2) $L_{P1}-C_S-C_{P2}$	<b>No problem</b>	
(3) $C_{P1}-L_S-C_{P2}$		<b>DC blocking</b> Two $L_{\text{infinitive}}$ are needed
(5) $C_{P1}-C_S-L_{P2}$		Two $L_{\text{infinitive}}$ are needed
T Type		
(2) $L_{S1}-C_P-C_{S2}$		Two $L_{\text{infinitive}}$ are needed
(3) $C_{S1}-L_P-C_{S2}$		Two $L_{\text{infinitive}}$ are needed
(5) $C_{S1}-C_P-L_{S2}$		Two $L_{\text{infinitive}}$ are needed

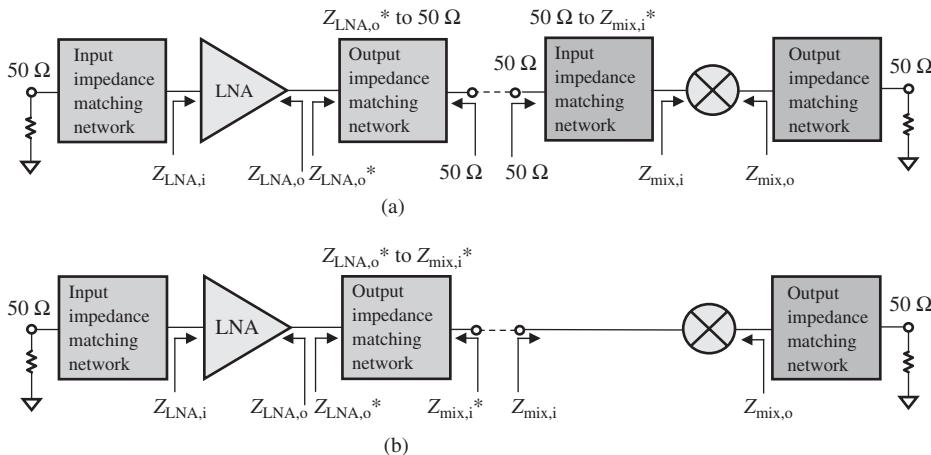


Figure 3.21. Two plans for building impedance matching networks for a low noise amplifier (LNA) and mixer. (a) A typical but conservative plan in building of impedance matching networks for LNA and mixer. (b) A better plan in building of impedance matching networks for LNA and mixer.

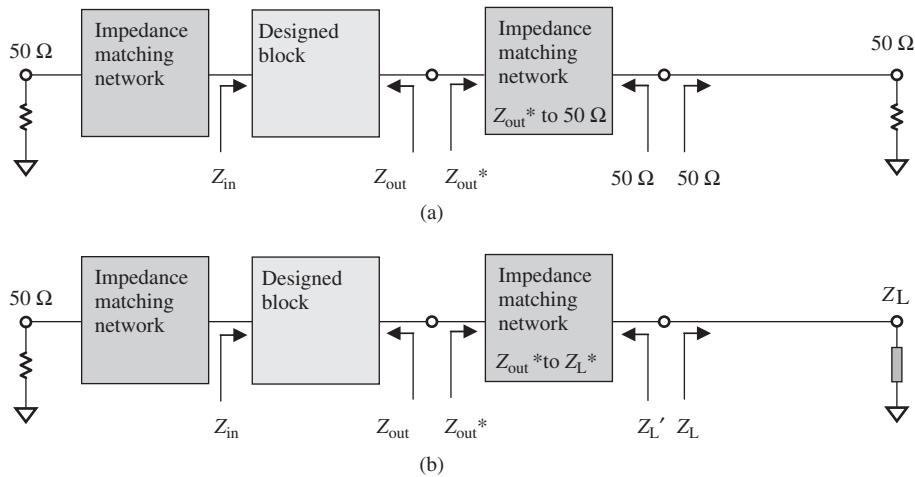
This is a special case of impedance matching when  $Z_S$  or  $Z_L$  of an RF block is not 50  $\Omega$ .

Figure 3.22 interprets this subject in a general form. The interpretation is focused on the first three blocks, or the LNA portion as shown in Figure 3.21. The corresponding relationships between Figures 3.21 and 3.22 are listed in Table 3.8.

Now let us copy Figure 3.22(b) as Figure 3.23(a) and denote it as block A.

Theoretically, to match the output impedance  $Z_{\text{out}}$  of the designed block to  $Z_L$ , but not to 50  $\Omega$ , is not a problem. It looks like a better plan. However, it brings about a new problem. The testing for block A is not available because its output is not 50  $\Omega$ . The testing is available only in a 50- $\Omega$  system since the input or output impedance of most test equipment in an RF laboratory is 50  $\Omega$ .

In order to enable testing or measurement, an additional impedance matching network from  $Z_L$  to 50  $\Omega$  must be added to block A. Figure 3.23(b) depicts the entire system of



**Figure 3.22.** Impedance matching when  $Z_L$  is not  $50 \Omega$ . (a) Typical impedance matching at the output of the designed block system. (b) Output impedance,  $Z_{out}$ , of a designed block is assigned to match  $Z_L$  but not  $50 \Omega$ .

TABLE 3.8. Relationship between Figures 3.21 and 3.22

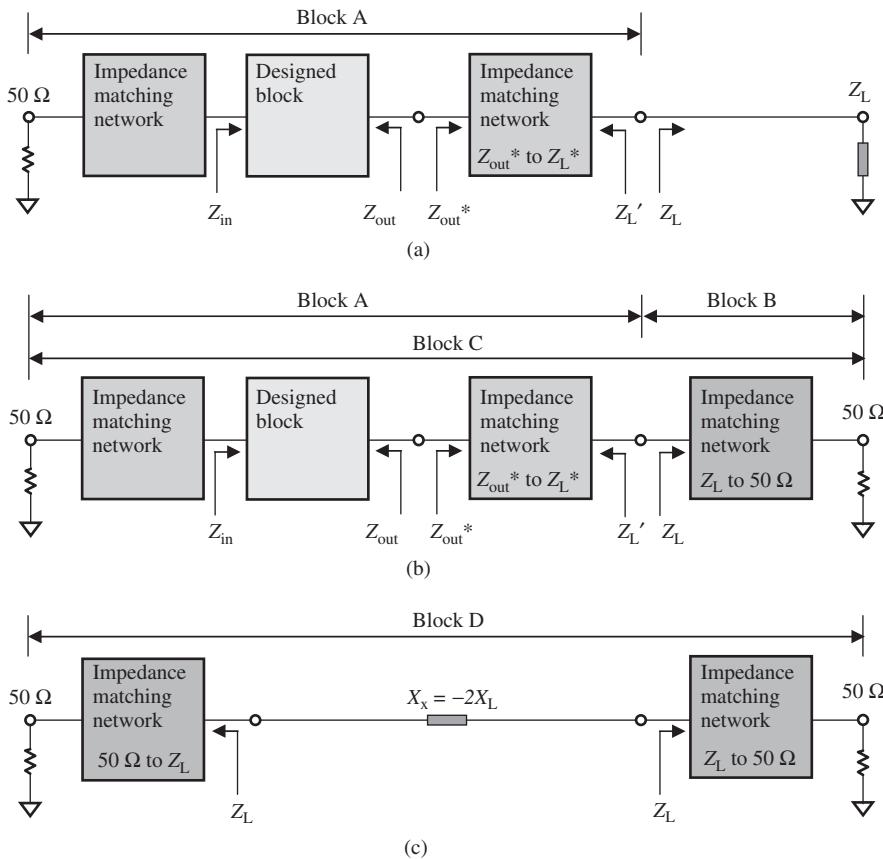
In Figure 3.21	In Figure 3.22
LNA	Designed block
$Z_{LNA,i}$	$Z_{in}$
$Z_{LNA,o}$	$Z_{out}$
$Z_{LNA,o}^*$	$Z_{out}^*$
$Z_{mix,i}^*$	$Z'_L$
$Z_{mix,i}$	$Z_L$

blocks. Let us denote the additional impedance matching block as block B and the entire block as C. The problems remaining are the following:

- One can test the performance for the combination of the LNA and block B, or block C, but not for the individual LNA or block B, because there is no node with  $50 \Omega$  impedance between the LNA output and block B, and so testing is not possible for the individual LNA or block B.
- Thus, it is also impossible to verify the performance of all the four individual blocks in block C if special means are not employed.

It must be noted that only block A will be applied and block B will be removed in the product. Alternatively, the performance of block A could be obtained by subtracting the performance contributed by block B from that of block C.

In order to evaluate the performance of block B, it must be tested or measured in the  $50\text{-}\Omega$  system as well. Figure 3.23(c) shows a simple but effective means. There are two identical blocks B piggy-backed together to form a new block D, so that they can be tested or measured exactly in the  $50\text{-}\Omega$  system. The performance of block D can be characterized by testing its power loss, noise figure, intercept points, and so on.



**Figure 3.23.** Design and testing an RF block when its output is to be matched with an impedance  $Z_L$  but not 50 Ω. (a) Output impedance,  $Z_{out}$ , of a designed block is assigned to match  $Z_L$  but not 50Ω. (b) Block B, an additional impedance matching network from  $Z_L$  to 50 Ω is added to block A. (c) Two piggy-backed blocks B form a new block D with 50 Ω. An additional reactance,  $X_x = -2X_L$ , must be added between the two piggy-backed blocks so that their impedances are conjugate-matched with each other.

The contribution of block B can be evaluated by testing block D in terms of cascaded equations. However, in order to reach the impedance matching condition for these two piggy-backed blocks, a part with reactance  $X_x = -2X_L$  must be inserted between the piggy-backed sides so as to “neutralize” the reactance existing between the two piggy-backed blocks.  $X_L$  is the reactance of the load impedance  $Z_L$  and is known as an input parameter for the circuit design.

Testing of these two piggy-backed blocks, or block D, will be conducted for all the required parameters to characterize its performance, such as power gain or loss, noise figure, IP<sub>3</sub> or IP<sub>2</sub>, and so on. Then, the value of the required parameters for one of the piggy-backed blocks or block B can be calculated in terms of cascaded equations. Instead of simply dividing the tested value by 2, the calculation is somehow tedious and will be demonstrated in the following exercise.

The next test is for block C, as shown in Figure 3.23(b). The tested value for all the required parameters is contributed by blocks A and B.

TABLE 3.9. Calculated Values for All the Required Parameters from Tested Values

Tested Values for the Blocks Shown in Figure 3.23(b)	Tested Values for the blocks Shown in Figure 3.23(c)	Calculated Values for the Last Block Shown in Figure 3.23(b)	Calculated Values for the Blocks Shown in Figure 3.23(a)
Column	1	2	3
Assigned Block C	Block D	Block B	Block A
$G(\text{gain}), \text{dB}$	12.5	-2.7	-1.35
$NF, \text{dB}$	2.8	2.7	1.35
$IIP_3, \text{dB}_m$	5	100	102.39
$IIP_2, \text{dB}_m$	32	150	155.37
			4
			Block A

Finally, the actual performance of block A can be obtained from the performance of blocks C and B if the contribution of block B is subtracted through the cascaded equations.

Here is an exercise to calculate the values of the required parameters for an assigned block A by subtracting the tested values for block B from the tested values for block C.

Assuming that the circuit block C, as shown in 3.23(b), has been well designed, the tested results for blocks C and D as shown in Figure 3.23(c) are listed in first and second column of Table 3.9, respectively, while the calculated values for the additional impedance matching block or block B and the assigned blocks or block A as shown in Figure 3.23(a) are listed in the third and fourth columns, respectively.

All the cascaded equations of NF, IP<sub>2</sub>, and IP<sub>3</sub>, applied in this section will be derived and verified in chapter 12. From Figure 3.23(c), it can be found that



The cascaded equations are therefore

$$G_B + G_B = G_D(\text{dB}), \quad (3.47)$$

or,

$$G_B = \frac{G_D}{2}(\text{dB}), \quad (3.48)$$

$$NF_B + \frac{NF_B - 1}{G_B} = NF_D(W), \quad (3.49)$$

or,

$$NF_B = \frac{1 + G_B NF_D}{1 + G_B}(W), \quad (3.50)$$

$$\frac{1}{(IIP_3)_B} + \frac{G_B}{(IIP_3)_B} = \frac{1}{(IIP_3)_D}(W), \quad (3.51)$$

or,

$$(IIP_3)_B = (IIP_3)_D(1 + G_B)(W), \quad (3.52)$$

$$\left[ \frac{1}{(IIP_2)_B} \right]^{1/2} + \left[ \frac{G_B}{(IIP_2)_B} \right]^{1/2} = \left[ \frac{1}{(IIP_2)_D} \right]^{1/2}(W), \quad (3.53)$$

or,

$$(IIP_2)_B = (IIP_2)_D (1 + G_B)^{1/2} (W) \quad (3.54)$$

where the subscript B and D denote the parameter of the corresponding block B and block D, respectively.

From Figure 3.23(b), it can be found that



$$G_A + G_B = G_C \text{ (dB)}, \quad (3.55)$$

or,

$$G_A = G_C - G_B \text{ (dB)}, \quad (3.56)$$

$$NF_A + \frac{NF_B - 1}{G_A} = NF_C(W), \quad (3.57)$$

or,

$$NF_A = NF_C - \frac{NF_B - 1}{G_A}(W), \quad (3.58)$$

$$\frac{1}{(IIP_3)_A} + \frac{G_A}{(IIP_3)_B} = \frac{1}{(IIP_3)_C}(W), \quad (3.59)$$

or,

$$(IIP_3)_A = \frac{(IIP_3)_C (IIP_3)_B}{(IIP_3)_B - G_A (IIP_3)_C}(W), \quad (3.60)$$

$$\left[ \frac{1}{(IIP_2)_A} \right]^{1/2} + \left[ \frac{G_A}{(IIP_2)_B} \right]^{1/2} = \left[ \frac{1}{(IIP_2)_C} \right]^{1/2}(W), \quad (3.61)$$

or,

$$(IIP_2)_A = \frac{1}{\left\{ \left[ \frac{1}{(IIP_2)_C} \right]^{1/2} - \left[ \frac{G_A}{(IIP_2)_B} \right]^{1/2} \right\}^2}(W), \quad (3.62)$$

where the subscripts A, B, and C denote the parameter of the corresponding block A, block B, and block C, respectively.

### 3.7 PARTS IN AN IMPEDANCE MATCHING NETWORK

In Sections 3.3–3.5, passive impedance matching networks built by one, two, and three parts have been discussed. It is very rare to see a passive impedance matching network built by only one part. In most practical engineering designs of the narrow-band case, it is usually built by two or three parts. A passive impedance matching network built by two parts is quite popular though there is a topology limitation. If a passive impedance

matching network is built by three parts, the topology constraints are significantly reduced and many choices of parts' values are available.

Empirically, in order to design a wide-band matching network, more than three parts may be necessary. It should be noted that, in reality, passive parts are not ideal. So they always attenuate the signal with their resistive components. Too many parts in an impedance matching network might bring about a serious attenuation of the signal.

Theoretically, resistors could be used in an impedance matching network. They are, however, never used in a practical impedance matching network because they bring about serious attenuation and noise. Usually, an impedance matching network is built by only capacitors and inductors, but not resistors; moreover, capacitors are preferable to inductors. In the circuit designed by discrete parts, the cost of the capacitor is much less than that of the inductor. In an IC circuit design, the area of a capacitor is much less than that of an inductor. Therefore, the cost of a capacitor is also much less than that of an inductor. In addition, the  $Q$  value of an inductor is much lower than that of a capacitor in IC circuits.

## APPENDICES

### 3.A.1 Fundamentals of the Smith Chart

The Smith Chart is a powerful tool in RF circuit design. It reflects the relationship between voltage reflection coefficient and impedance. On the basis of transmission line theory, an AC voltage transported along a transmission line and incident on a load will be partially reflected as long as the characteristic impedance of the transmission line  $Z_0$  is different from the impedance of the load  $Z$ . The ratio of the reflected voltage to the incident voltage is defined as the voltage reflection coefficient  $\Gamma$ , and can be expressed as

$$\Gamma = \frac{Z - Z_0}{Z + Z_0} \quad (3.A.1)$$

where

$\Gamma$  = the voltage reflection coefficient,

$Z$  = the impedance of the load, and

$Z_0$  = the characteristic impedance of transmission line

In general, the voltage reflection coefficient  $\Gamma$  is a complex parameter , that is,

$$\Gamma = U + jV \quad (3.A.2)$$

where

$\Gamma$  = the voltage reflection coefficient,

$U$  = the real part of voltage reflection coefficient, and

$V$  = the real part of voltage reflection coefficient.

The impedance  $Z$  is also a complex parameter, that is,

$$Z = R + jX \quad (3.A.3)$$

where

- $R$  = the real part or resistance of the impedance and  
 $X$  = the imaginary part or reactance of the impedance.

By introducing normalized impedance, that is,

$$z = \frac{Z}{Z_0}. \quad (3.A.4)$$

Expression (3.A.4) becomes

$$z = r + jx, \quad (3.A.5)$$

$$r = \frac{R}{Z_0}, \quad (3.A.6)$$

$$x = \frac{X}{Z_0} \quad (3.A.7)$$

where

- $r$  = the normalized resistance and  
 $x$  = the normalized reactance.

Equation (3.A.1) becomes

$$\Gamma = \frac{z - 1}{z + 1} = \frac{(r - 1) + jx}{(r + 1) + jx} = U + jV \quad (3.A.8)$$

where

$$U = \frac{r^2 - 1 + x^2}{(r + 1)^2 + x^2}, \quad (3.A.9)$$

$$V = \frac{2x}{(r + 1)^2 + x^2}. \quad (3.A.10)$$

By eliminating  $x$  from (3.A.9) and (3.A.10), we get

$$\left(U - \frac{r}{r + 1}\right)^2 + V^2 = \left(\frac{1}{r + 1}\right)^2. \quad (3.A.11)$$

By eliminating  $r$  from (3.A.9) and (3.A.10), we get

$$(U - 1)^2 + \left(V - \frac{1}{x}\right)^2 = \left(\frac{1}{x}\right)^2 \quad (3.A.12)$$

Equation (3.A.11) contains a family of circles centered at  $U = r/(r + 1)$ ,  $V = 0$  with radii  $1/(r + 1)$ , and equation (3.A.12) contains a family of arcs on circles centered at  $U = 1$ ,  $V = 1/x$  with radii  $1/x$ .

Figure 3.A.1 shows these circles and arcs on the Smith Chart. It presents the impedance coordination in the complex plane of the voltage reflection coefficient. The value of  $r$  for each circle is marked along the central horizontal line, while the value of  $x$  for each arc is marked along the biggest circle. At the leftmost point, denoted

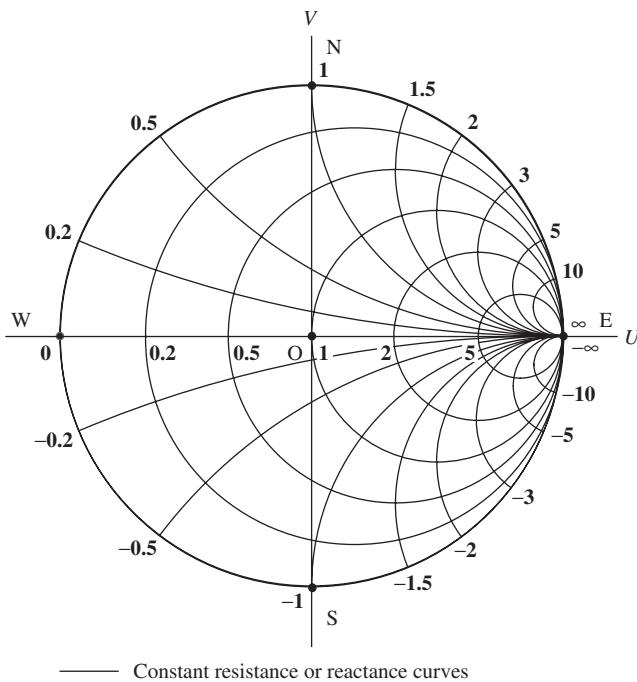


Figure 3.A.1. Impedance coordination of the Smith Chart.

Note 1: The normalized resistance  $x$  is bold-marked with values around the biggest circle,  $0, 0.1, 0.2, 0.5, 1, 1.5, 2, 3, 5, 10, \infty, -10, -5, -3, -2, -1.5, -1, -0.5, -0.2, -0.1$ .

Note 2: The normalized reactance  $r$  is bold-marked with values along the horizontal axis  $U$ ,  $0, 0.2, 0.5, 1, 2, 5, \infty$ .

by  $W$ ,  $r = x = 0$ , so that  $z = Z_0 = 0$ . It is a zero impedance point or a short-circuited point. At the rightmost point, denoted by  $E$ ,  $x = 0$  and  $0 \leq r < \infty$ , so that  $0 \leq z < \infty$ , or  $0 \leq Z < \infty$ . It is an open-circuited point or a singular point mathematically. At the uppermost point, denoted by  $N$ ,  $r = 0$ ,  $x = 1$ . It is the point where the impedance is purely inductive and  $X = Z_0$ . At the bottommost point, denoted by  $S$ ,  $r = 0$ ,  $x = -1$ . It is the point where the impedance is purely capacitive and  $X = -Z_0$ . At the center point, denoted by  $O$ ,  $r = 1$ ,  $x = 0$ , so that  $z = 1$  or  $Z = Z_0 = 50 \Omega$ . It is a standard reference impedance point.

The coordination of impedance  $z$  in the Smith Chart can be converted into the admittance coordination  $y$ .

$$y = \frac{1}{z} = g + jb \quad (3.A.13)$$

where

- $y$  = the normalized admittance,
- $g$  = the normalized conductance, and
- $b$  = the normalized susceptance.

The relationship between the admittance  $y$  and the voltage reflection coefficient  $\Gamma$  can be found from (3.A.1) and (3.A.4): that is,

$$z = \frac{Z}{Z_0} = \frac{1 + \Gamma}{1 - \Gamma}, \quad (3.A.14)$$

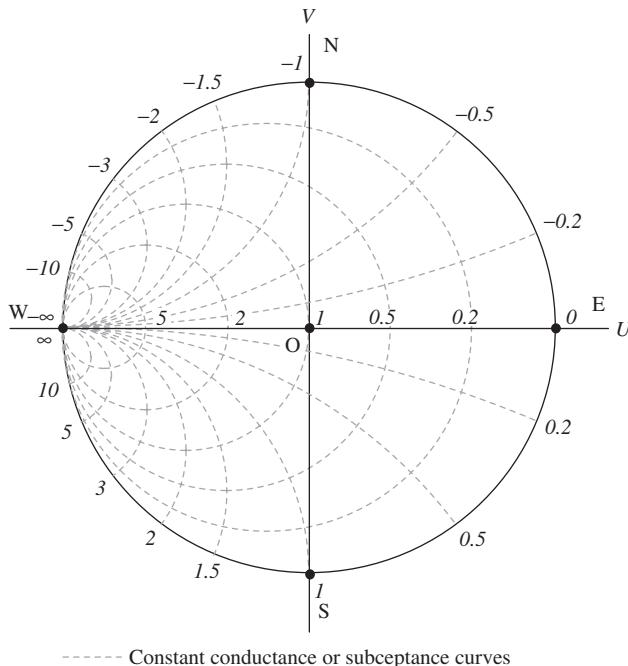
and then from the definition of  $y$  as shown in (3.A.13), we have

$$y = \frac{1}{z} = \frac{1 - \Gamma}{1 + \Gamma} = \frac{1 + \Gamma e^{j\pi}}{1 - \Gamma e^{j\pi}}. \quad (3.A.15)$$

In equation (3.A.15), we rewrite  $-\Gamma$  as  $\Gamma e^{j\pi}$  so that equations (3.A.14) and (3.A.15) both look similar in their mathematical expression. By comparing equations (3.A.14) and (3.A.15), it is found that the conversion from  $z$  to  $y$  in the Smith Chart can be done if the complex plane  $\Gamma$  is simply rotated by  $180^\circ$ .

Figure 3.A.2 presents the admittance coordination of the Smith Chart in the complex plane of the voltage reflection coefficient. The value of  $g$  for each circle is marked along the central horizontal line, while the value of  $b$  for each arc is marked along the biggest circle. At the leftmost point, denoted by  $W$ ,  $b = 0$  and  $0 \leq g < \infty$ , so that  $0 \leq y < \infty$ , or  $0 \leq Y < \infty$ . This is a short-circuited point, or, mathematically, a singular point. At the rightmost point, denoted by  $E$ ,  $g = b = 0$ , so that  $y = Y_o = 0$ . It is a zero admittance point or an open-circuited point. At the center point, denoted by  $O$ ,  $g = 1$  and  $b = 0$ , so that  $y = 1$ .

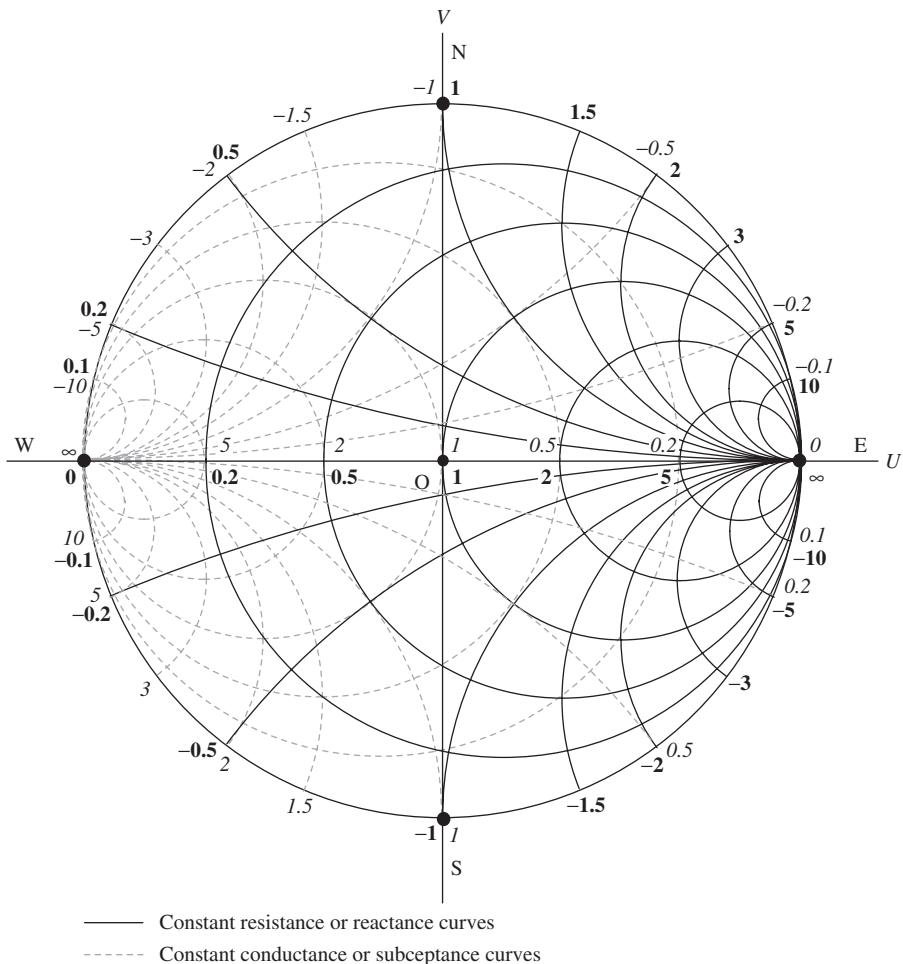
It is more convenient for an RF circuit designer if both impedance and admittance coordination can be marked on the Smith Chart simultaneously. Figure 3.A.3 shows such a combination of Figures 3.A.1 and 3.A.2, that is, a Smith Chart showing both impedance



**Figure 3.A.2.** Admittance coordination of the Smith Chart.

Note 1: Normalized conductance  $b$  is italicized with values around the biggest circle,  $0, 0.1, 0.2, 0.5, 1, 1.5, 2, 3, 5, 10, \infty, -10, -5, -3, -2, -1.5, -1, -0.5, -0.2, -0.1$ .

Note 2: Normalized susceptance  $g$  is italicized with values along the horizontal axis  $U, 0, 0.2, 0.5, 1, 2, 5, \infty$ .



**Figure 3.A.3.** Impedance and admittance coordination of the Smith Chart.

Note 1: Normalized resistance  $x$  is bold-marked with values around the biggest circle, such as, **0, 0.1, 0.2, 0.5, 1, 1.5, 2, 3, 5, 10,  $\infty$ , -10, -5, -3, -2, -1.5, -1, -0.5, -0.2, -0.1**.

Note 2: Normalized reactance  $r$  is bold-marked with values along the horizontal axis  $U$ , such as, **0, 0.2, 0.5, 1, 2, 5,  $\infty$** .

Note 3: Normalized conductance  $b$  is italicized with values around the biggest circle, such as, *0, 0.1, 0.2, 0.5, 1, 1.5, 2, 3, 5, 10,  $\infty$ , -10, -5, -3, -2, -1.5, -1, -0.5, -0.2, -0.1*.

Note 4: Normalized susceptance  $g$  is italicized with values along the horizontal axis  $U$ , such as, *0, 0.2, 0.5, 1, 2, 5,  $\infty$*

and admittance coordination. If a matching network consists of only passive parts either in series or in parallel, then the impedance coordination is applied when the part is added in series, while the admittance coordination is applied when the part is added in parallel. Consequently, with the help of the Smith Chart as shown in Figure 3.A.3, both variations of impedance and admittance for a matching network can be figured out simultaneously and without difficulty.

Usually, not only  $z$ ,  $y$ , and  $\Gamma$  but also other related parameters are marked on a Smith Chart. They are as follows:

- Reflection coefficient of power  $\gamma$ ,
- Return loss in dB,
- Reflection loss in dB,
- VSWR (voltage standing wave ratio),
- VSWR in dB, and
- transmission loss coefficient.

The multiple coordinates make the Smith Chart more convenient in RF circuit design. Most parameters can be read from same chart simultaneously.

### 3.A.2 Formula for Two-Part Impedance Matching Network

The following formulas for the two-part impedance matching network were developed and derived by the author. From the discussion on the two-part impedance matching network in Section 3.4.2, or directly from Figure 3.11, three common rules or features can be summarized as follows:

1. The first part in a two-part impedance matching network is to pull the original impedance to the circle either  $g = 1$  or  $r = 1$ .
2. The second part in a two-part impedance matching network is to pull the impedance on the circle either  $g = 1$  or  $r = 1$ , and then to the standard reference impedance,  $50 \Omega$ .
3. One of the two matching parts is in series and the other one is in parallel. However, the designer can select the first part in series or in parallel because, as described above, in all possible original impedance, there are two ways of pulling it to the center of the Smith Chart. The first part in one way is in series while in another way it is in parallel, and vice versa.

On the basis of the common rules or features mentioned above, the type of the two parts can be selected and the values of the matching parts, one in series and the other in parallel, can be formalized.

For the purpose of derivation of the formula, let us list the relations between impedance and admittance first. An impedance can be expressed as

$$Z = R + jX, \quad \text{and} \quad (3.A.16)$$

its normalized impedance and admittance are

$$z = r + jx \quad (3.A.17)$$

where

$$r = \frac{R}{Z_0}, \quad (3.A.18)$$

$$x = \frac{X}{Z_0} \quad (3.A.19)$$

where  $Z_0$  is the characteristic impedance,

$$y = g + jb = \frac{1}{z} = \frac{1}{r + jx} = \frac{r - jx}{r^2 + x^2}, \quad (3.A.20)$$

$$g = \frac{r}{r^2 + x^2}, \quad (3.A.21)$$

$$b = -\frac{x}{r^2 + x^2}. \quad (3.A.22)$$

The selection of the type and the calculation of value of the matching parts can be conducted for four cases.

Let us copy Figure 3.11 as Figure 3.A.4.

### Case 1: When the original impedance is located in region 1

The first impedance matching part,  $X_1$  or  $x_1$ , denoted by subscript 1, is added to the original impedance  $Z_m$  in series so that  $Z_m$  can be moved to the circle with  $g = 1$ .

Let us denote original impedance to be matched as  $Z_m$ . It consists of real and imaginary parts  $R_m$  and  $X_m$ , that is,

$$Z_m = R_m + jX_m, \quad (3.A.23)$$

and its normalized impedance is

$$z_m = r_m + jx_m, \quad (3.A.24)$$

where

$$r_m = \frac{R_m}{Z_0}, \quad (3.A.25)$$

$$x_m = \frac{X_m}{Z_0}. \quad (3.A.26)$$

After the first matching part  $X_1$  is added, the original impedance  $Z_m$  is moved to the circle with  $g = 1$  and becomes  $Z_{g1}$ , which is

$$Z_{g1} = R_{g1} + jX_{g1}, \quad (3.A.27)$$

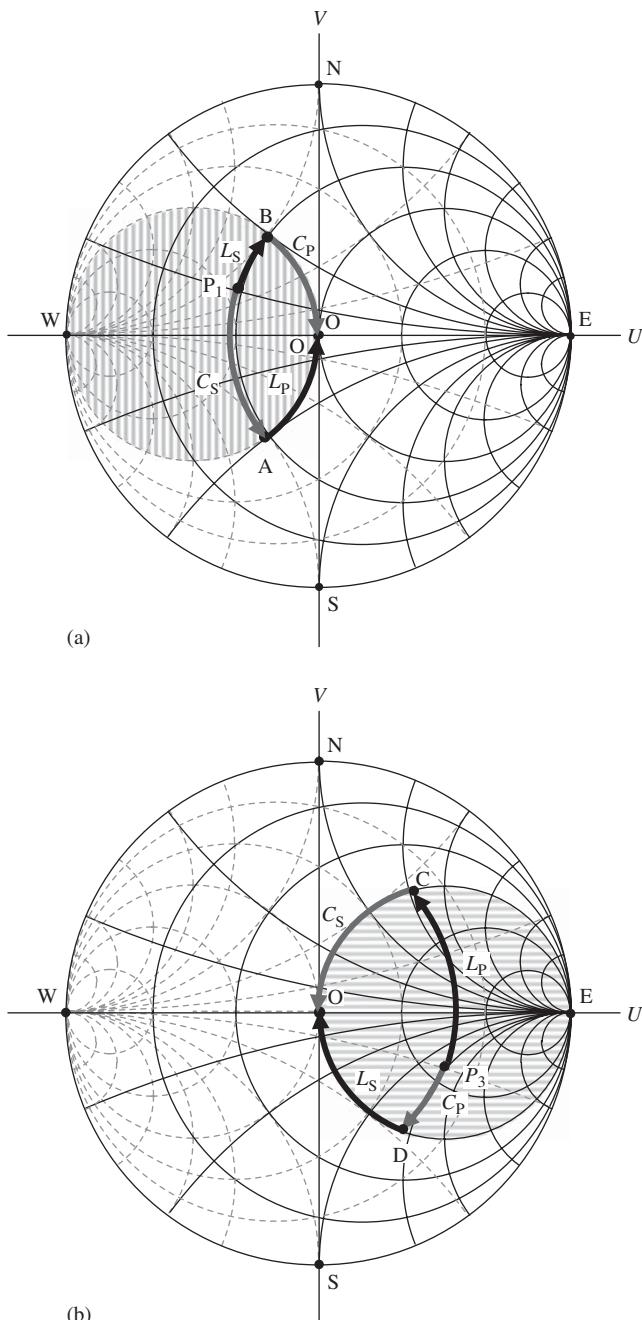
or its normalized impedance

$$z_{g1} = r_{g1} + jx_{g1}. \quad (3.A.28)$$

Correspondingly,

$$Y_{g1} = G_{g1} + B_{g1}, \quad (3.A.29)$$

$$y_{g1} = g_{g1} + b_{g1} \quad (3.A.30)$$



**Figure 3.A.4.** (a) Two ways to pull the original impedance  $P_1$  in region 1 to the center of Smith Chart,  $O$ , by addition of two passive parts (b) Two ways to pull the original impedance  $P_2$  in region 2 to the center of Smith Chart  $O$  by the addition of two passive parts. (c) Two ways to pull the original impedance  $P_3$  in region 3 to the center of Smith Chart  $O$  by the addition of two passive parts. (d) Two ways to pull the original impedance  $P_4$  in region 4 to the center of Smith Chart  $O$  by the addition of two passive parts.

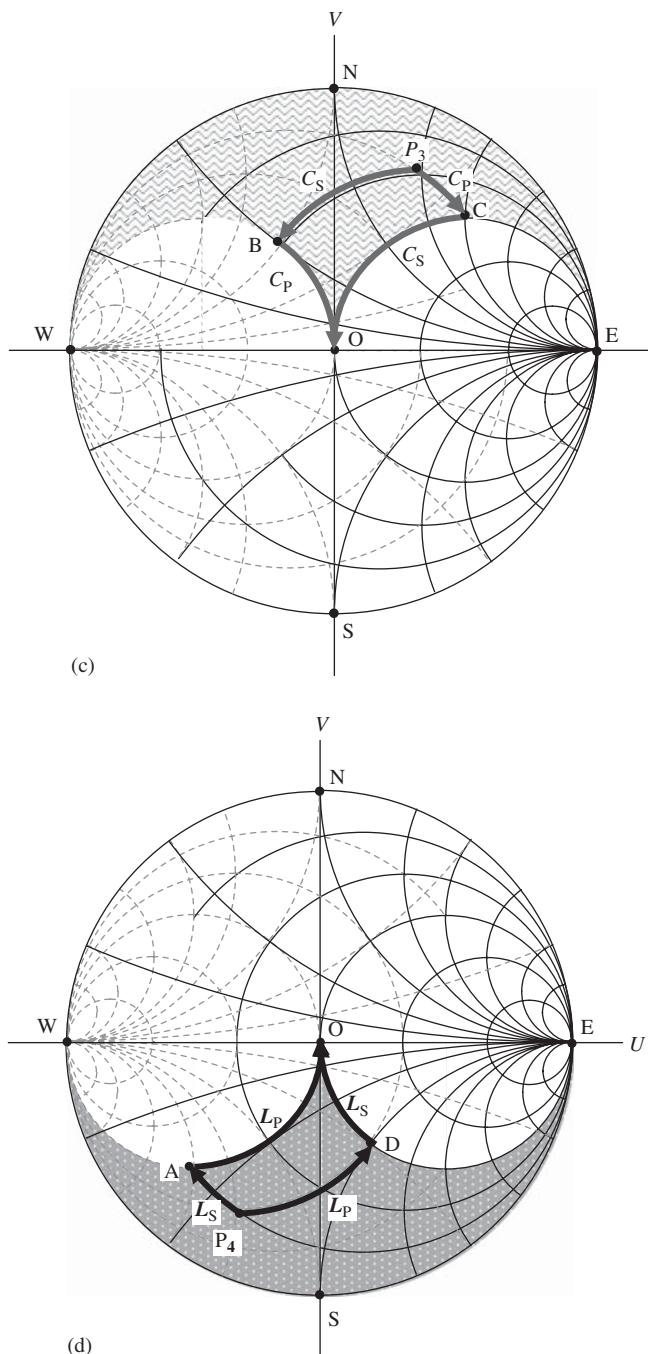


Figure 3.A.4. (Continued)

where

$$r_{g1} = \frac{R_{g1}}{Z_0}, \quad (3.A.31)$$

$$x_{g1} = \frac{X_{g1}}{Z_0}, \quad (3.A.32)$$

$$g_{g1} = \frac{G_{g1}}{Z_0} = \frac{r_{g1}}{r_{g1}^2 + x_{g1}^2}, \quad (3.A.33)$$

$$b_{g1} = \frac{B_{g1}}{Z_0} = -\frac{x_{g1}}{r_{g1}^2 + x_{g1}^2}. \quad (3.A.34)$$

The subscript “g1” denotes that the parameter is located at the circle with  $g = 1$  in the Smith Chart. The real part of  $Z_{g1}$  or  $z_{g1}$  is unchanged but the imaginary part of  $Z_{g1}$  or  $z_{g1}$  is changed as a result of the addition of  $X_1$  or  $x_1$ , that is,

$$r_{g1} = r_m, \quad (3.A.35)$$

and

$$x_{g1} = x_1 + x_m. \quad (3.A.36)$$

Consequently, from expressions (3.A.33), (3.A.35), and (3.A.36), we have

$$g_{g1} = \frac{r_{g1}}{r_{g1}^2 + x_{g1}^2} = \frac{r_m}{r_m^2 + (x_1 + x_m)^2} = 1. \quad (3.A.37)$$

Then

$$r_m^2 + (x_1 + x_m)^2 = r_m, \quad (3.A.38)$$

$$\left(\frac{R_m}{Z_0}\right)^2 + \left(\frac{X_1 + X_m}{Z_0}\right)^2 = \frac{R_m}{Z_0}, \quad (3.A.39)$$

$$X_1 + X_m = \pm\sqrt{R_m(Z_0 - R_m)}, \quad (3.A.40)$$

$$X_1 = \pm\sqrt{R_m(Z_0 - R_m)} - X_m. \quad (3.A.41)$$

The positive sign in equation (3.A.41) is for the selection of an inductive  $X_1$  while the negative sign is for the selection of a capacitive  $X_1$ .

The second matching part,  $X_2$  or  $x_2$ , denoted by subscript 2, in parallel is to pull  $Z_{g1}$  or  $z_{g1}$  on the circle  $g = 1$  to the center of the Smith Chart or to the reference impedance  $50 \Omega$ .

In other words, the second matching part  $X_2$  or  $x_2$  is to neutralize the imaginary part of  $Y_{g1}$  or  $y_{g1}$ , that is

$$\frac{1}{x_2} + (-b_{g1}) = 0. \quad (3.A.42)$$

From expressions (3.A.42) and (3.A.34) we have

$$\frac{1}{x_2} = b_{g1} = -\frac{x_1 + x_m}{r_m^2 + (x_1 + x_m)^2}. \quad (3.A.43)$$

From (3.A.41) and (3.A.43), we have

$$\frac{1}{X_2} = -\frac{X_1 + X_m}{R_m^2 + (X_1 + X_m)^2} = \frac{\mp\sqrt{R_m(Z_o - R_m)}}{R_m^2 + R_m(Z_o - R_m)}, \quad (3.A.44)$$

$$\frac{1}{X_2} = \frac{\mp\sqrt{R_m(Z_o - R_m)}}{R_m Z_o}. \quad (3.A.45)$$

The positive sign in equation (3.A.45) is for the selection of inductive  $X_2$ , while the negative sign is for the selection of capacitive  $X_2$ .

Also, a simple relationship between  $X_1$  and  $X_2$  can be found from equations (3.A.41) and (3.A.45):

$$-X_2(X_1 + X_m) = R_m Z_o. \quad (3.A.46)$$

Using equations (3.A.41) and (3.A.45), the values of two matching parts  $X_1$  and  $X_2$  can be calculated.

As shown in Figure 3.A.4, there are two options for impedance matching from the original value  $Z_m$  to the reference impedance  $50 \Omega$ . However, if the first impedance matching part is capacitive, the second impedance matching part must be inductive, or vice versa. Consequently, the two possible combinations to build the impedance matching network are as follows:

1. By adding a capacitor  $C_S$  in series first, and then an inductor  $L_P$  in parallel.

$$X_1 = -\sqrt{R_m(Z_o - R_m)} - X_m, \quad (3.A.47)$$

$$C_S = \frac{C_m}{C_m \omega \sqrt{R_m(Z_o - R_m)} - 1}, \quad (3.A.48)$$

and

$$\frac{1}{X_2} = \frac{+\sqrt{R_m(Z_o - R_m)}}{R_m Z_o}, \quad (3.A.49)$$

$$L_P = \frac{R_m Z_o}{\omega \sqrt{R_m(Z_o - R_m)}}. \quad (3.A.50)$$

2. By adding an inductor  $L_S$  in series first, and then a capacitor  $C_P$  in parallel.

$$X_1 = +\sqrt{R_m(Z_o - R_m)} - X_m, \quad (3.A.51)$$

$$L_S = \frac{\sqrt{R_m(Z_o - R_m)} - X_m}{\omega}, \quad (3.A.52)$$

and

$$\frac{1}{X_2} = \frac{-\sqrt{R_m(Z_o - R_m)}}{R_m Z_o}, \quad (3.A.53)$$

$$C_P = \frac{\sqrt{R_m(Z_o - R_m)}}{R_m Z_o \omega}. \quad (3.A.54)$$

These two options correspond to the two topologies of impedance matching networks as shown in Figure 3.A.4. It should be noted that the expressions are reasonable only if

$$\sqrt{R_m(Z_o - R_m)} > X_m. \quad (3.A.55)$$

The inductor and capacitor,  $L_S$  and  $C_P$ , respectively, in the first option would become a capacitor and an inductor,  $C_S$  and  $L_P$ , respectively, and vice versa in the second option if

$$\sqrt{R_m(Z_o - R_m)} < X_m. \quad (3.A.56)$$

### Case 2: When the original impedance is located in region 2

The first impedance matching part  $X_1$ , or  $x_1$ , denoted by subscript 1, is added with the original impedance  $Z_m$  in parallel so that  $Z_m$  can be moved to the circle with  $r = 1$ .

After the first matching part  $X_1$  is connected with the original impedance  $Z_m$  in parallel, the resultant impedance is

$$Z_{r1} = \frac{(R_m + jX_m)jX_1}{(R_m + jX_m) + jX_1} = \frac{(-X_m X_1 + jX_1 R_m)[R_m - j(X_m + X_1)]}{R_m^2 + (X_m + X_1)^2}, \quad (3.A.57)$$

$$R_{r1} = \frac{R_m X_1^2}{R_m^2 + (X_m + X_1)^2}, \quad (3.A.58)$$

$$X_{r1} = \frac{R_m^2 + X_m(X_m + X_1)}{R_m^2 + (X_m + X_1)^2} X_1. \quad (3.A.59)$$

The subscript “ $r1$ ” indicates that the denoted parameter is on the circle with  $r = 1$  on the Smith Chart.

The real part of the impedance  $Z_{r1}$  must be equal to  $Z_o$  because  $Z_{r1}$  is on the circle with  $r = 1$ , that is,

$$R_{r1} = \frac{R_m X_1^2}{R_m^2 + (X_m + X_1)^2} = Z_o, \quad (3.A.60)$$

$$(R_m - Z_o)X_1^2 - 2X_m Z_o X_1 - (R_m^2 + X_m^2)Z_o = 0, \quad (3.A.61)$$

$$X_1 = \frac{X_m Z_o \pm \sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{R_m - Z_o}. \quad (3.A.62)$$

The positive sign in equation (3.A.62) is for the selection of inductive  $X_1$ , while the negative sign is for the selection of capacitive  $X_1$ .

The second matching part  $X_2$  or  $x_2$ , denoted by subscript 2, in series is to pull  $Z_{g1}$  or  $z_{g1}$  on the circle  $r = 1$  to the center of the Smith Chart or to the reference impedance  $50 \Omega$ , that is,

$$X_2 = -X_{r1} = -\frac{R_m^2 + X_m(X_m + X_1)}{R_m^2 + (X_m + X_1)^2} X_p = \left[ \frac{(X_m + X_1)Z_o}{R_m X_1} - 1 \right] X_1, \quad (3.A.63)$$

$$X_2 = \frac{Z_o}{R_m} X_m + \frac{Z_o - R_m}{R_m} X_1. \quad (3.A.64)$$

Substituting of  $X_1$  from (3.A.62) into (3.A.64), we have

$$X_2 = \mp \frac{\sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{R_m}. \quad (3.A.65)$$

There are also two options. The positive sign in equation (3.A.65) is for the selection of inductive  $X_2$ , while the negative sign is for the selection of capacitive  $X_2$ .

Also, a simple relationship between  $X_1$  and  $X_2$  can be found from equations (3.A.62) and (3.A.65):

$$[X_m Z_o - X_1 (R_m - Z_o)] X_2 = Z_o (R_m^2 + X_m^2 - R_m Z_o). \quad (3.A.66)$$

In terms of equations (3.A.62) and (3.A.65), the values of the two matching parts  $X_1$  and  $X_2$  can be calculated.

As shown in Figure 3.A.4, there are two options for the impedance matching from the original value  $Z_m$  to the reference impedance  $50 \Omega$  in the expressions (3.A.62) and (3.A.65). However, if the first impedance matching part is capacitive, then the second part must be inductive, or vice versa. Consequently, two possible combinations to build the impedance matching network are as follows:

1. By adding an inductor  $L_p$  in parallel first, and then adding a capacitor  $C_s$  in series.

$$X_1 = \frac{X_m Z_o + \sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{R_m - Z_o}, \quad (3.A.67)$$

$$L_p = \frac{X_m Z_o + \sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{(R_m - Z_o) \omega}, \quad (3.A.68)$$

$$X_2 = -\frac{\sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{R_m}, \quad (3.A.69)$$

$$C_s = \frac{R_m / \omega}{\sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}. \quad (3.A.70)$$

2. By adding a capacitor  $C_P$  in parallel first, and then adding an inductor  $L_S$  in series.

$$X_1 = \frac{X_m Z_o - \sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{R_m - Z_o}, \quad (3.A.71)$$

$$C_P = \frac{(R_m - Z_o)/\omega}{\sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)} - X_m Z_o}, \quad (3.A.72)$$

$$X_2 = + \frac{\sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{R_m}, \quad (3.A.73)$$

$$L_S = \frac{\sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{R_m \omega}. \quad (3.A.74)$$

### Case<sup>#</sup> 3: When the original impedance is located in region 3

The first impedance matching part  $X_1$  or  $x_1$ , denoted by subscript 1, is added with the original impedance  $Z_m$  in series or in parallel so that  $Z_m$  can be moved to the circle with  $g = 1$  or  $r = 1$ .  $X_1$  is a capacitor.

In series

$$X_1 = -\sqrt{R_m (Z_o - R_m)} - X_m. \quad (3.A.75)$$

In parallel

$$X_1 = \frac{X_m Z_o - \sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{R_m - Z_o}. \quad (3.A.76)$$

The positive sign in equation (3.A.41) signifies the selection of an inductive  $X_1$ , while the negative sign signifies the selection of a capacitive  $X_1$ .

The second matching part  $X_2$  or  $x_2$ , denoted by subscript 2, in parallel or in series is to pull  $Z_{g1}$  or  $z_{g1}$  on the circle  $g = 1$  or  $r = 1$  to the center of the Smith Chart or to the reference impedance  $50 \Omega$ .  $X_2$  is a capacitor as well.

In parallel

$$\frac{1}{X_2} = \frac{-\sqrt{R_m (Z_o - R_m)}}{R_m Z_o}. \quad (3.A.77)$$

In series

$$X_2 = - \frac{\sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{R_m}. \quad (3.A.78)$$

As shown in Figure 3.A.4, there are two options for impedance matching from the original value  $Z_m$  to the reference impedance  $50 \Omega$ . The two impedance matching parts are capacitors.

Consequently, two possible combinations to build the impedance matching network are as follows:

1. By adding a capacitor  $C_S$  in series first, and then a capacitor  $C_P$  in parallel.

$$X_1 = -\sqrt{R_m(Z_o - R_m)} - X_m, \quad (3.A.79)$$

$$C_S = \frac{C_m}{C_m \omega \sqrt{R_m(Z_o - R_m)} - 1}, \quad (3.A.80)$$

and

$$\frac{1}{X_2} = \frac{-\sqrt{R_m(Z_o - R_m)}}{R_m Z_o}, \quad (3.A.81)$$

$$C_P = \frac{\sqrt{R_m(Z_o - R_m)}}{R_m Z_o \omega}. \quad (3.A.82)$$

2. By adding a capacitor  $C_P$  in parallel first, and then a capacitor  $C_S$  in series.

$$X_1 = \frac{X_m Z_o - \sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{R_m - Z_o}, \quad (3.A.83)$$

$$C_P = \frac{(R_m - Z_o)/\omega}{\sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)} - X_m Z_o}, \quad (3.A.84)$$

and

$$X_2 = -\frac{\sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{R_m}, \quad (3.A.85)$$

$$C_S = \frac{R_m/\omega}{\sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}. \quad (3.A.86)$$

#### **Case<sup>#</sup> 4: When the original impedance is located in region 4**

The first impedance matching part  $X_1$  or  $x_1$ , denoted by subscript 1, is added with the original impedance  $Z_m$  in series or in parallel so that  $Z_m$  can be moved to the circle with  $g = 1$  or  $r = 1$ .  $X_1$  is an inductor.

In series

$$X_1 = +\sqrt{R_m(Z_o - R_m)} - X_m. \quad (3.A.87)$$

In parallel

$$X_1 = \frac{X_m Z_o + \sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{R_m - Z_o}. \quad (3.A.88)$$

The second matching part,  $X_2$  or  $x_2$ , denoted by subscript 2, in parallel or in series is to pull  $Z_{g1}$  or  $z_{g1}$  on the circle  $g = 1$  or  $r = 1$  to the center of the Smith Chart or to the reference impedance  $50 \Omega$ .  $X_2$  is an inductor as well.

In parallel

$$\frac{1}{X_2} = \frac{+\sqrt{R_m(Z_o - R_m)}}{R_m Z_o}. \quad (3.A.89)$$

In series

$$X_2 = +\frac{\sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{R_m}. \quad (3.A.90)$$

As shown in Figure 3.A.4, there are two options for impedance matching from the original value  $Z_m$  to the reference impedance  $50 \Omega$ . The two impedance matching parts are inductors.

Consequently, the two possible combinations to build the impedance matching network are as follows:

1. By adding an inductor  $L_S$  in series first, and then a capacitor  $L_P$  in parallel.

$$X_1 = +\sqrt{R_m(Z_o - R_m)} - X_m, \quad (3.A.91)$$

$$L_S = \frac{\sqrt{R_m(Z_o - R_m)} - X_m}{\omega}, \quad (3.A.92)$$

$$\frac{1}{X_2} = \frac{+\sqrt{R_m(Z_o - R_m)}}{R_m Z_o}, \quad (3.A.93)$$

$$L_P = \frac{R_m Z_o}{\omega \sqrt{R_m(Z_o - R_m)}}. \quad (3.A.94)$$

2. By adding a capacitor  $L_P$  in parallel first, and then an inductor  $L_S$  in series.

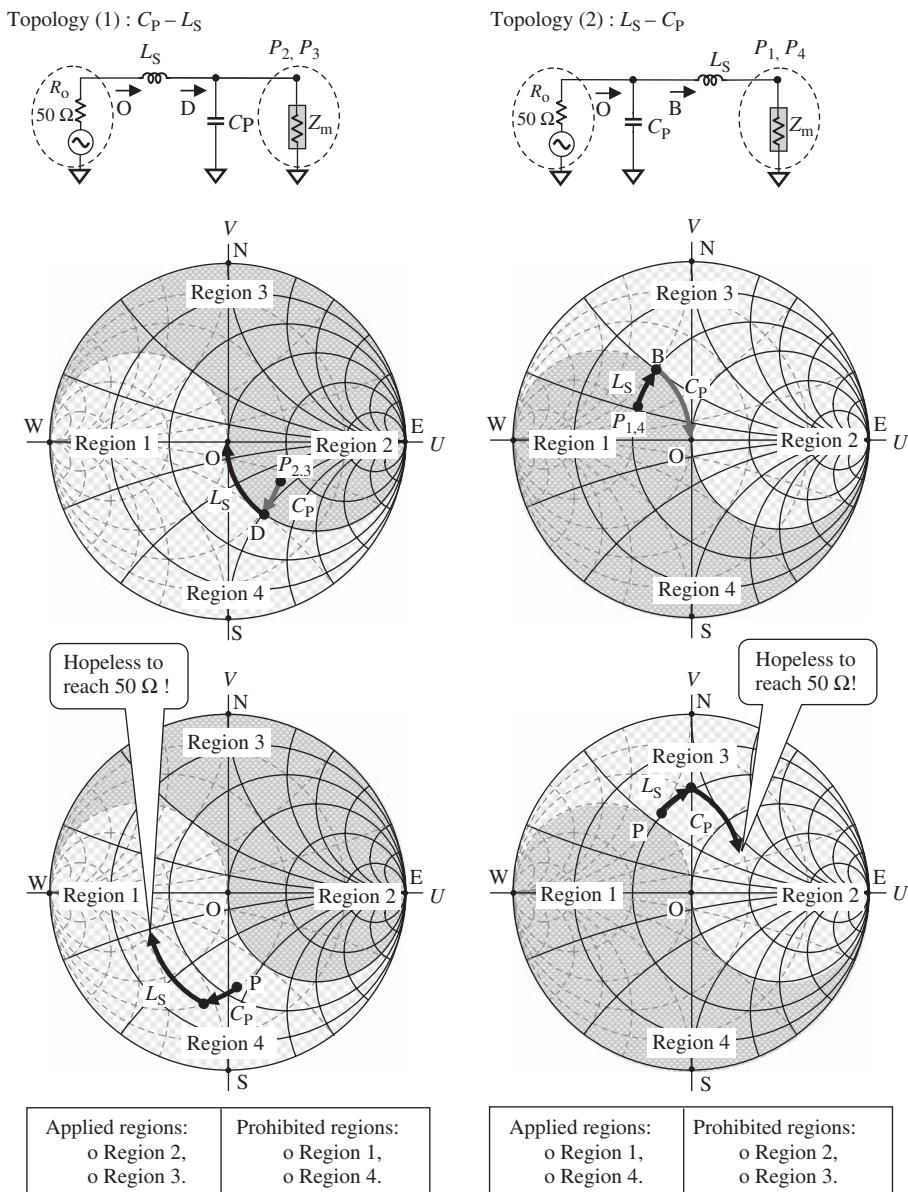
$$X_1 = \frac{X_m Z_o + \sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{R_m - Z_o}, \quad (3.A.95)$$

$$L_P = \frac{X_m Z_o + \sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{(R_m - Z_o) \omega}, \quad (3.A.96)$$

$$X_2 = +\frac{\sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{R_m}, \quad (3.A.97)$$

$$L_S = \frac{\sqrt{R_m Z_o (R_m^2 + X_m^2 - R_m Z_o)}}{R_m \omega}. \quad (3.A.98)$$

### 3.A.3 Topology Limitations of the Two-Part Impedance Matching Network

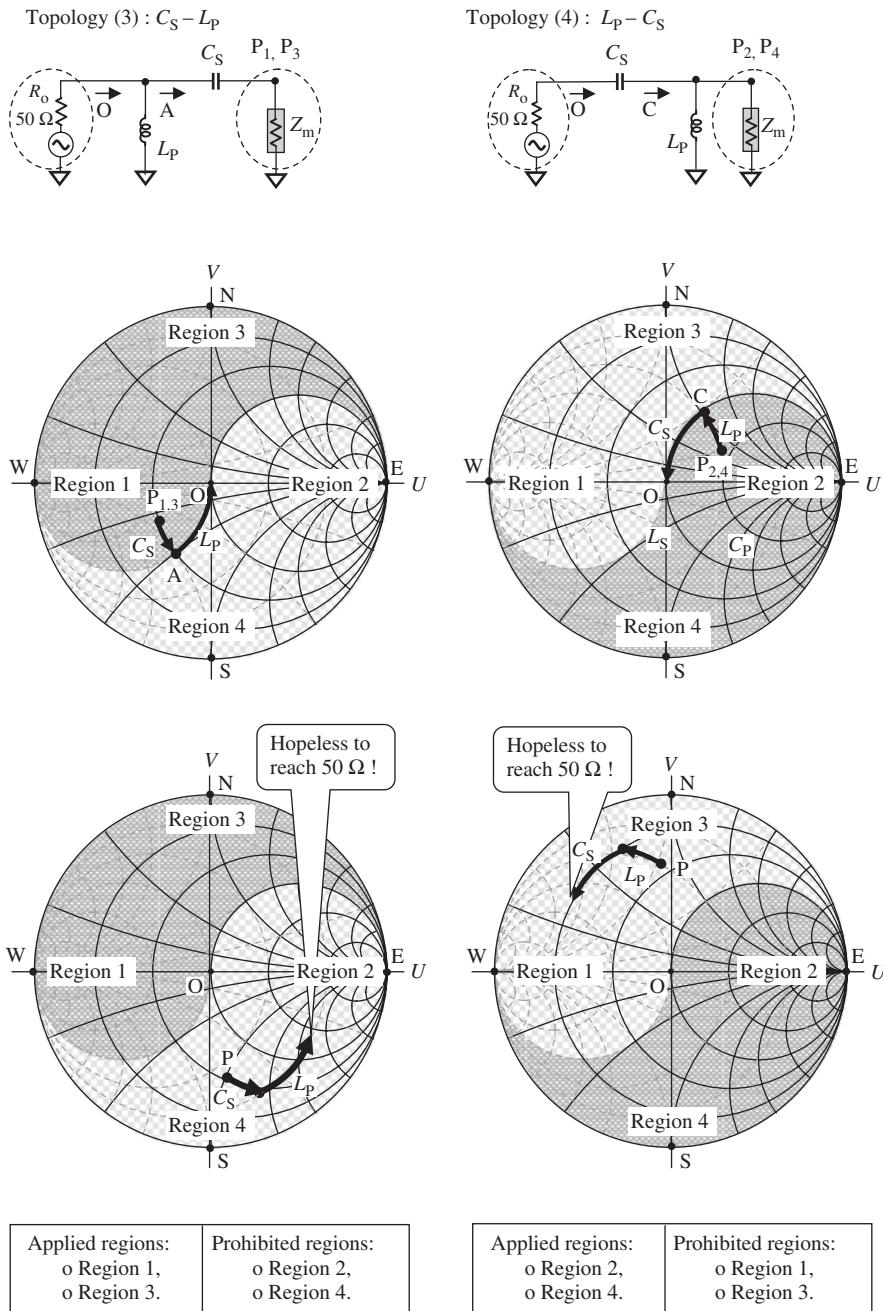


**Figure 3.A.5.** Applied and prohibitive region of the topology  $C_P - L_S$  or  $L_S - C_P$ .

Note 1:  $Z_m$  is the original impedance to be matched.

Note 2: In the title of the circuit topology " $C_P - L_S$ " or " $L_S - C_P$ ", the first part is connected to the original impedance to be matched and second part is connected to the reference impedance  $50 \Omega$ .

Note 3: Subscript P stands for in parallel and subscript S stands for in series.

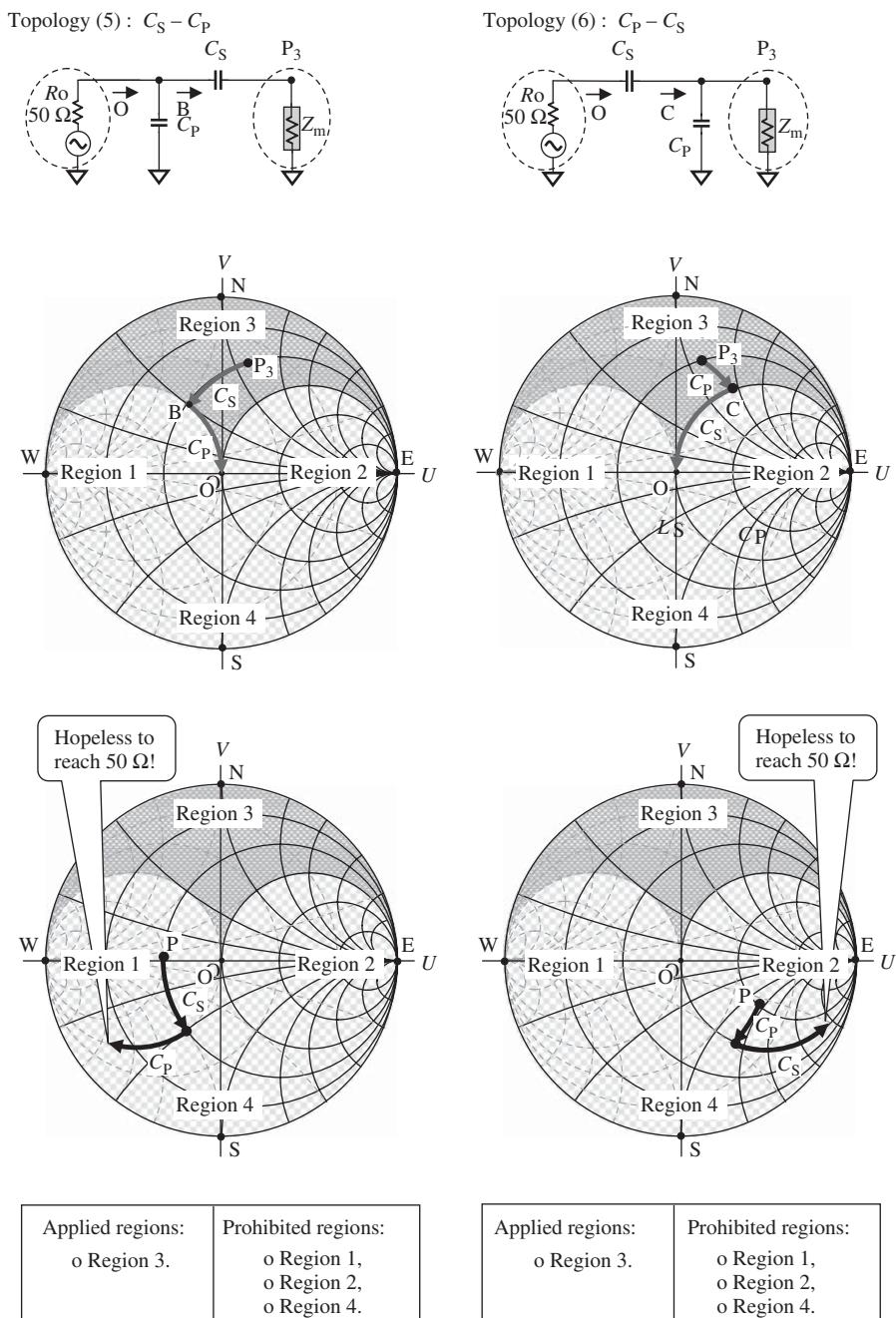


**Figure 3.A.6.** Applied and prohibitive region of the topology  $C_S - L_P$  or  $L_P - C_S$ .

Note 1:  $Z_m$  is the original impedance to be matched.

Note 2: In the title of circuit topology “ $C_S - L_P$ ” or “ $L_P - C_S$ ”, the first part is connected to the original impedance to be matched and second part is connected to the reference impedance  $50 \Omega$ .

Note 3: Subscript P stands for in parallel and subscript S stands for in series.

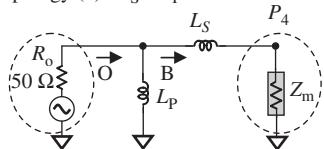
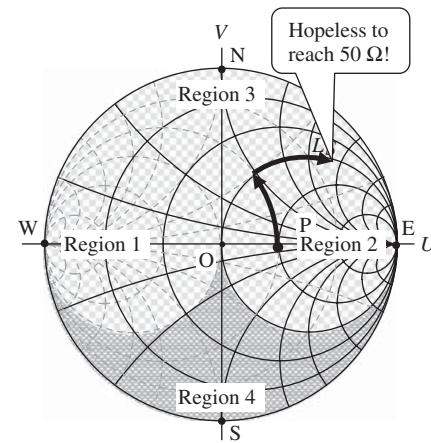
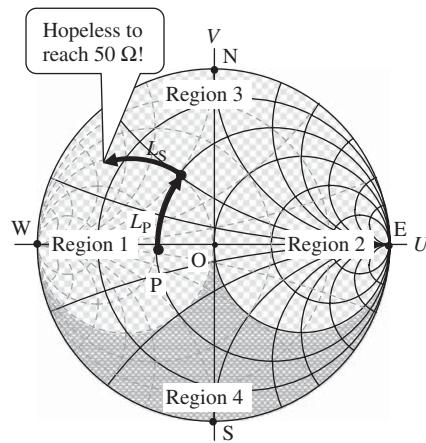
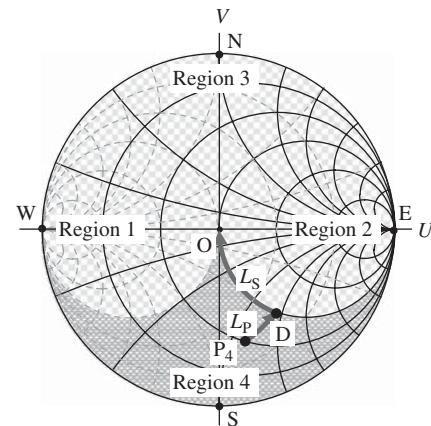
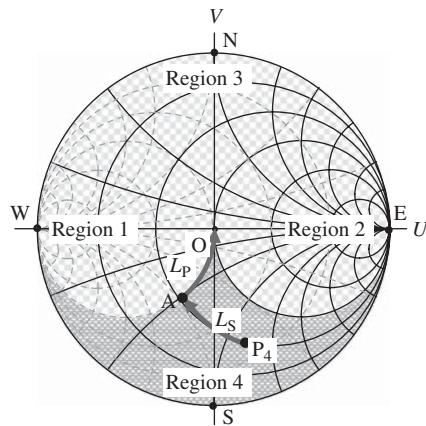
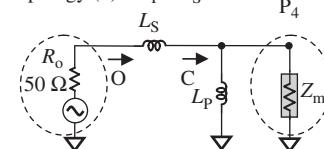


**Figure 3.A.7.** Applied and prohibitive region of the topology  $C_S - C_P$  or  $C_P - C_S$ .

Note 1:  $Z_m$  is the original impedance to be matched.

Note 2: In the title of circuit topology " $C_S - C_P$ " or " $C_P - C_S$ ", the first part is connected to the original impedance to be matched and the second part is connected to the reference impedance  $50 \Omega$ .

Note 3: Subscript P stands for in parallel and subscript S stands for in series.

Topology (7) :  $L_S - L_P$ Topology (8) :  $L_P - L_S$ 

Applied regions:  
o Region 4.

Prohibited regions:  
o Region 1,  
o Region 2,  
o Region 3.

Applied regions:  
o Region 4.

Prohibited regions:  
o Region 1,  
o Region 2,  
o Region 3.

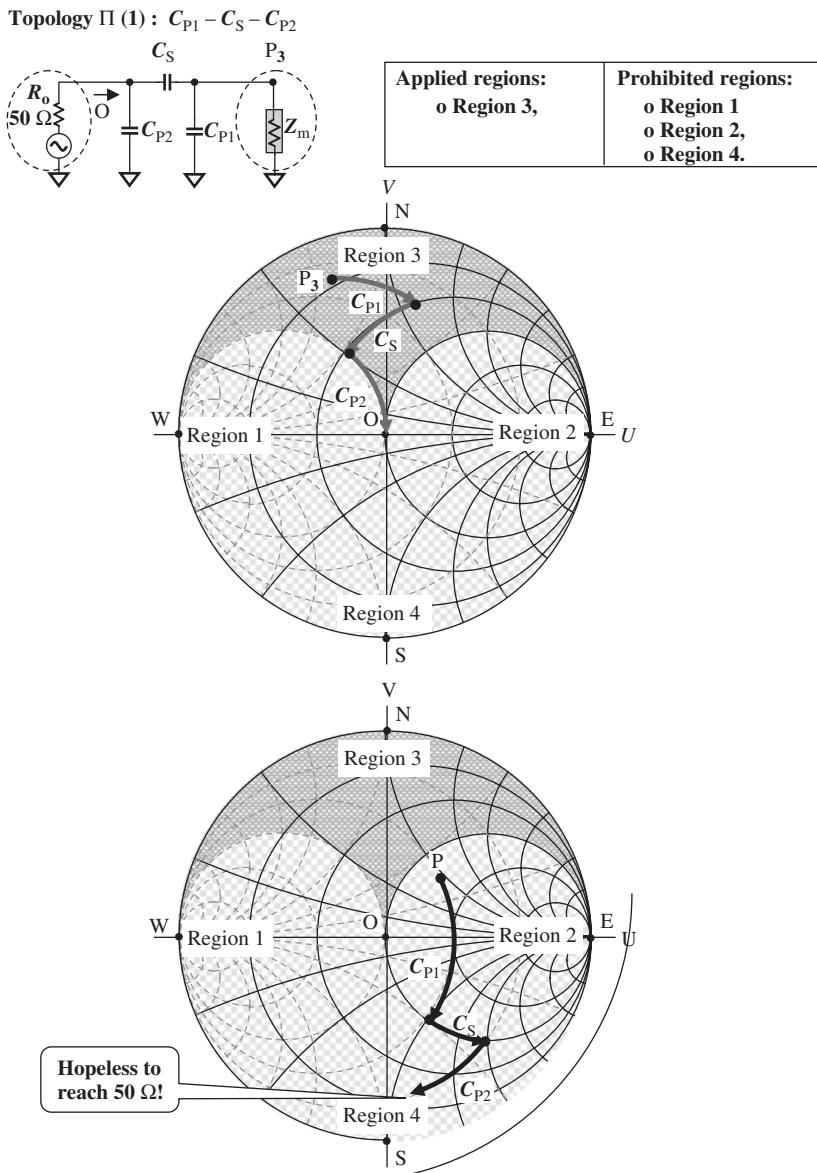
**Figure 3.A.8.** Applied and prohibitive region of the topology  $L_S - L_P$  or  $L_P - L_S$ .

Note 1:  $Z_m$  is the original impedance to be matched.

Note 2: In the title of circuit topology " $L_P - L_S$ " or " $L_S - L_P$ ," the first part is connected to the original impedance to be matched and the second part is connected to the reference impedance  $50 \Omega$ .

Note 3: Subscript P stands for in parallel and subscript S stands for in series.

### 3.A.4 Topology Limitation of Three Parts Impedance Matching Network



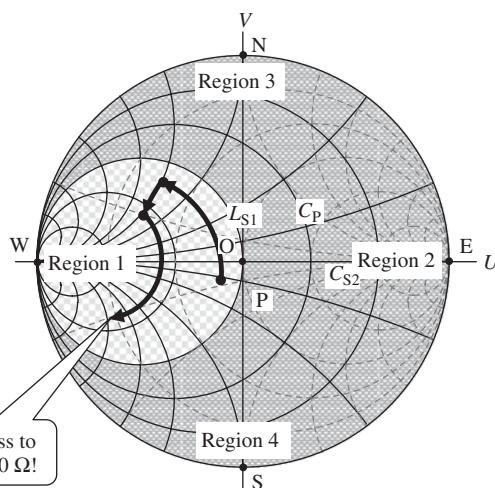
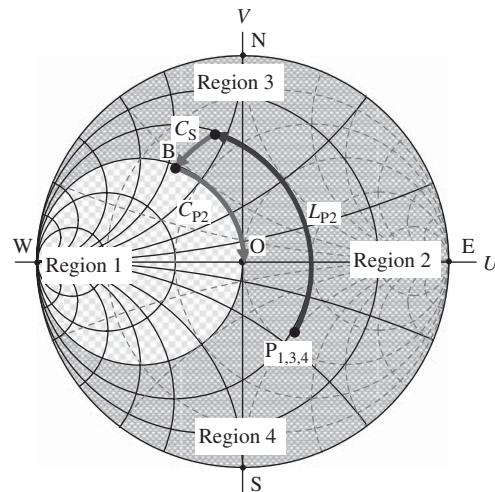
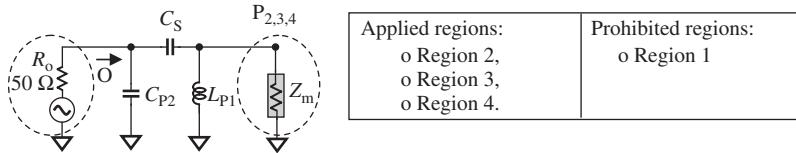
**Figure 3.A.9.** Applied and prohibitive area of the topology  $C_{P1} - C_S - C_{P2}$ .

Note 1:  $Z_m$  is original impedance to be matched.

Note 2: In the title of the circuit topology, first part marked with subscript 1 is connected to the original impedance to be matched and the second part marked with subscript 2 is connected to the reference impedance  $50 \Omega$ .

Note 3: Subscript P stands for in parallel and subscript S stands for in series.

Topology  $\Pi$  (2) :  $L_{P1} - C_S - C_{P2}$

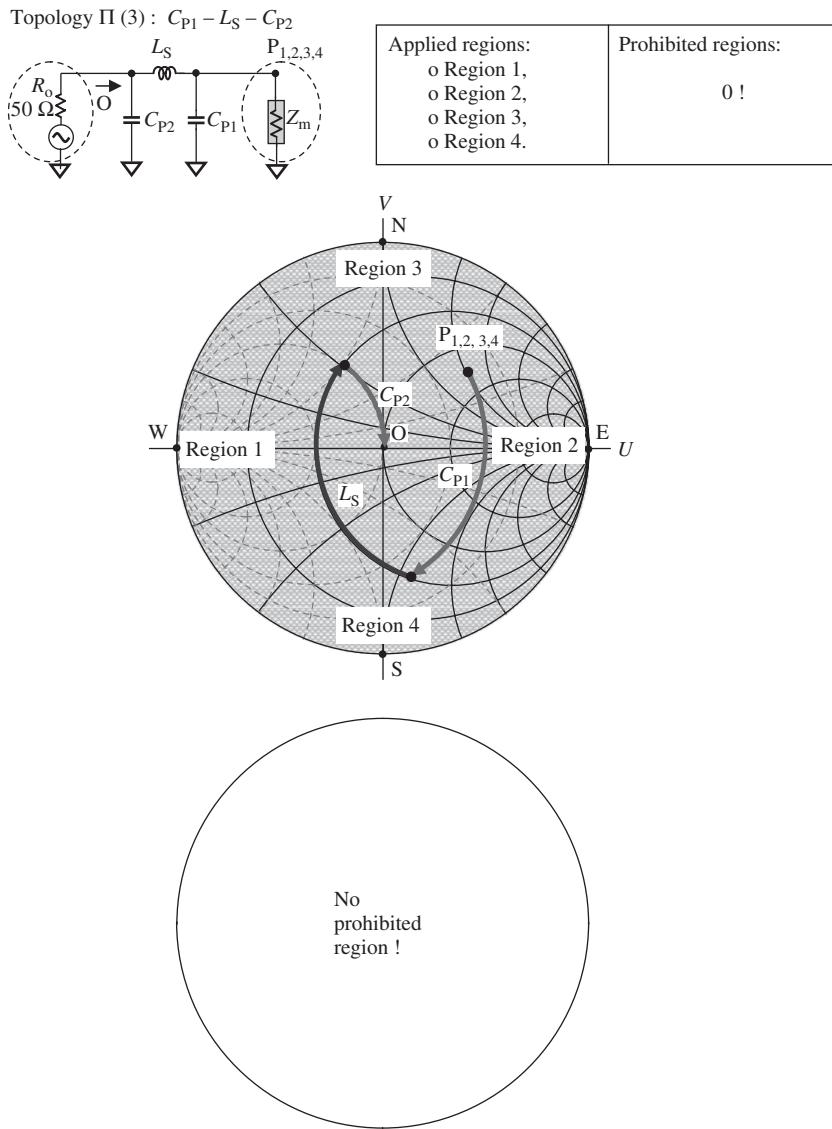


**Figure 3.A.10.** Applied and prohibitive area of the topology  $L_{P1} - C_S - C_{P2}$ .

Note 1:  $Z_m$  is the original impedance to be matched.

Note 2: In the title of the circuit topology, the first part marked with subscript 1 is connected to the original impedance to be matched and the second part marked with subscript 2 is connected to reference impedance  $50 \Omega$ .

Note 3: Subscript P stands for in parallel and subscript S stands for in series.



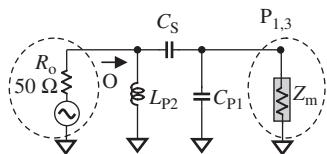
**Figure 3.A.11.** Applied and prohibitive area of the topology  $C_{P1}-L_S-C_{P2}$ .

Note 1:  $Z_m$  is the original impedance to be matched.

Note 2: In the title of the circuit topology, the first part marked with subscript 1 is connected to the original impedance to be matched and second part marked with subscript 2 is connected to the reference impedance  $50 \Omega$ .

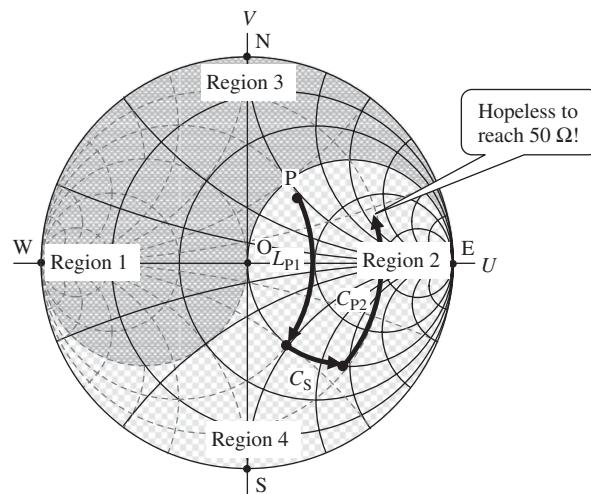
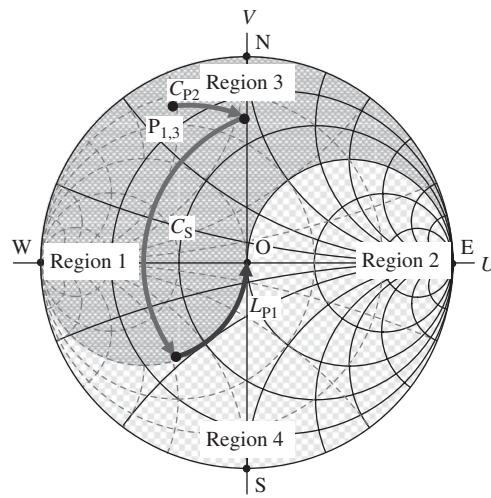
Note 3: Subscript P stands for in parallel and subscript S stands for in series.

Topology  $\Pi$  (5) :  $C_{P1} - C_S - L_{P2}$



Applied regions:  
o Region 1,  
o Region 3.

Prohibited regions:  
o Region 2,  
o Region 4.

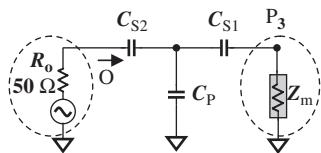


**Figure 3.A.12.** Applied and prohibitive area of the topology  $L_{P1}-C_S-C_{P2}$ .

Note 1:  $Z_m$  is the original impedance to be matched.

Note 2: In the title of the circuit topology, the first part marked with subscript 1 is connected to the original impedance to be matched and the second part marked with subscript 2 is connected to the reference impedance  $50 \Omega$ .

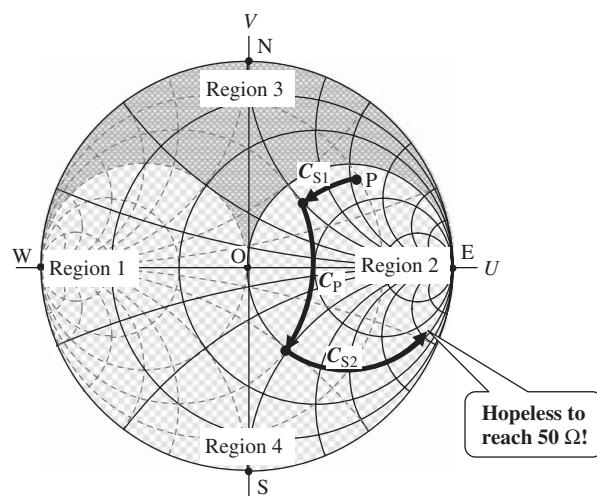
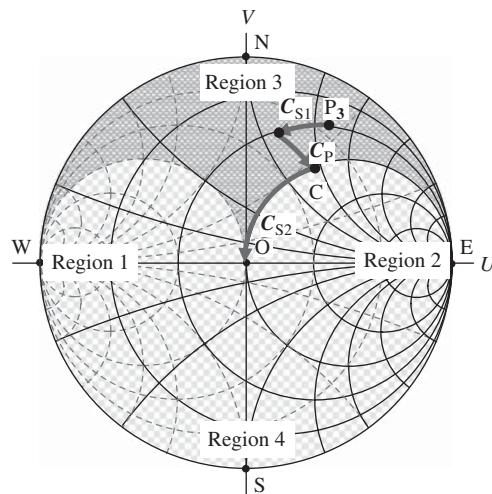
Note 3: Subscript P stands for in parallel and subscript S stands for in series.

**Topology T (1) :  $C_{S1} - C_P - C_{S2}$** **Applied regions:**

- o Region 3,

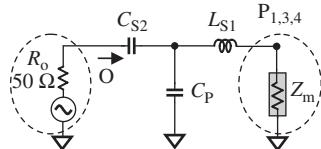
**Prohibited regions:**

- o Region 1
- o Region 2,
- o Region4.

**Figure 3.A.13. Applied and prohibitive area of the topology  $C_{S1}-C_P-C_{S2}$ .**Note 1:  $Z_m$  is the original impedance to be matched.Note 2: In the title of the circuit topology, the first part marked with subscript 1 is connected to the original impedance to be matched and the second part marked with subscript 2 is connected to the reference impedance  $50 \Omega$ .

Note 3: Subscript P stands for in parallel and subscript S stands for in series.

Topology T (2) :  $L_{S1} - C_P - C_{S2}$



Applied regions:	Prohibited regions:
o Region 1,	o Region 2
o Region 3,	
o Region4.	

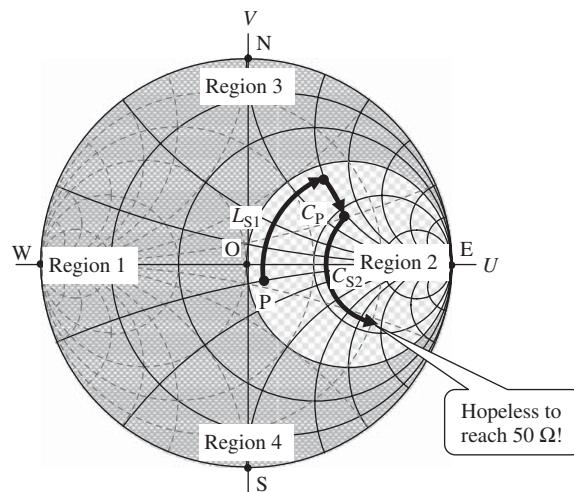
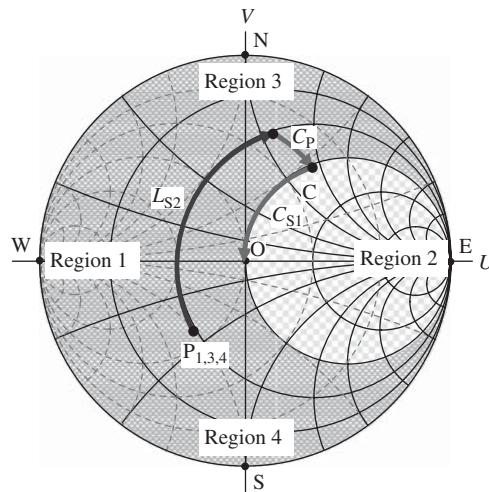


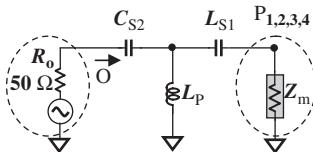
Figure 3.A.14. Applied and prohibitive area of the topology  $L_{S1} - C_P - C_{S2}$ .

Note 1:  $Z_m$  is the original impedance to be matched.

Note 2: In the title of the circuit topology, the first part marked with subscript 1 is connected to the original impedance to be matched and the second part marked with subscript 2 is connected to the reference impedance  $50 \Omega$ .

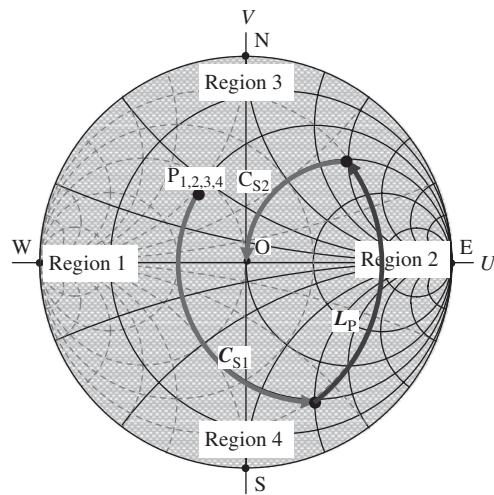
Note 3: Subscript P stands for in parallel and subscript S stands for in series.

**Topology T (3) :  $C_{S1} - L_P - C_{S2}$**



Applied regions:  
 Region 1,  
 Region 2,  
 Region 3,  
 Region 4.

Prohibited regions:  
**0 !**

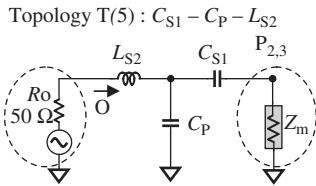


**Figure 3.A.15. Applied and prohibitive area of the topology  $C_{S1} - L_P - C_{S2}$ .**

Note 1:  $Z_m$  is the original impedance to be matched.

Note 2: In the title of the circuit topology, the first part marked with subscript 1 is connected to the original impedance to be matched and the second part marked with subscript 2 is connected to the reference impedance  $50 \Omega$ .

Note 3: Subscript P stands for in parallel and subscript S stands for in series.



Applied regions:	Prohibited regions:
o Region 2, o Region 3.	
	o Region 1 o Region 4

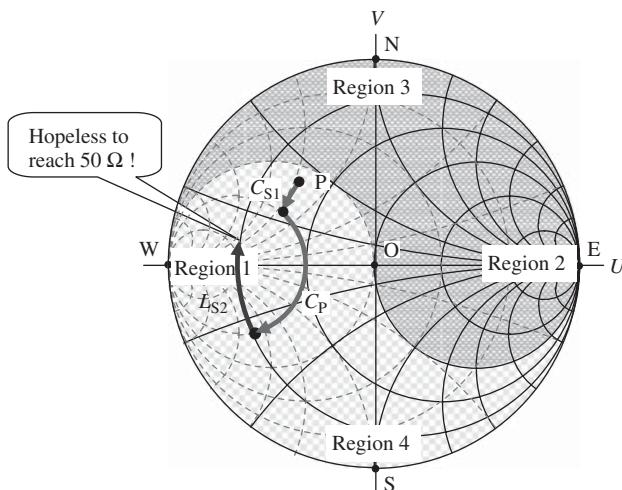
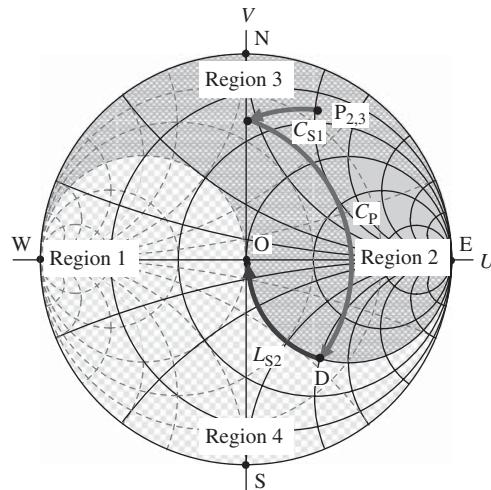


Figure 3.A.16. Applied and prohibitive area of the topology  $C_{S1} - C_P - L_{S2}$ .

Note 1:  $Z_m$  is the original impedance to be matched.

Note 2: In the title of circuit topology, the first part marked with subscript 1 is connected to the original impedance to be matched and second part marked with subscript 2 is connected to the reference impedance  $50 \Omega$ .

Note 3: Subscript P stands for in parallel and subscript S stands for in series.

### 3.A.5 Conversion between $\Pi$ and T Type Matching Network

If a matching network consists of three parts, it can be built with T or  $\Pi$  configuration as shown in Figure 3.A.17. It is very hard to say which one is better. Generally, the selection is based on the following considerations:

- Type T is better than type  $\Pi$  if the DC blocking to the next block is of concern.
- Type  $\Pi$  is better than type T if the stray reactance in the input of the next block is of concern.

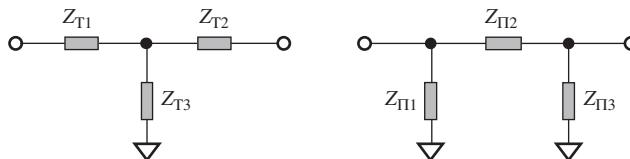


Figure 3.A.17. T and  $\Pi$  type matching networks with three parts.

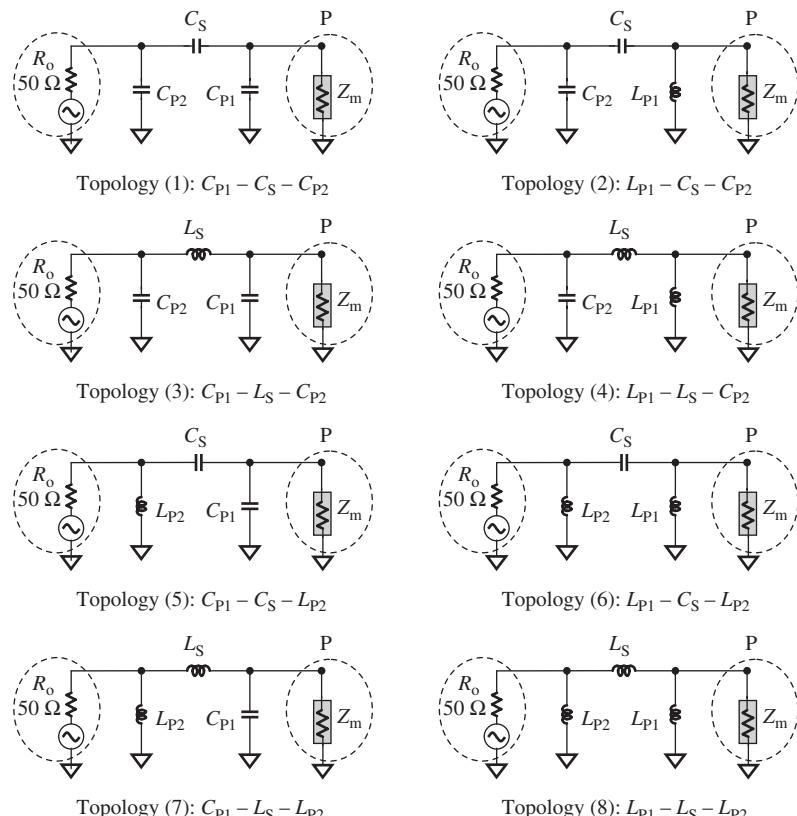


Figure 3.A.18. Topologies of a  $\Pi$ -type impedance matching network

Note 1:  $Z_m$  is the original impedance to be matched.

Note 2: In the title of the circuit topology, the first part marked with subscript 1 is connected to the original impedance to be matched, and the second part marked with subscript 2 is connected to the reference impedance  $50 \Omega$ .

Note 3: Subscript P stands for in parallel and subscript S stands for in series.

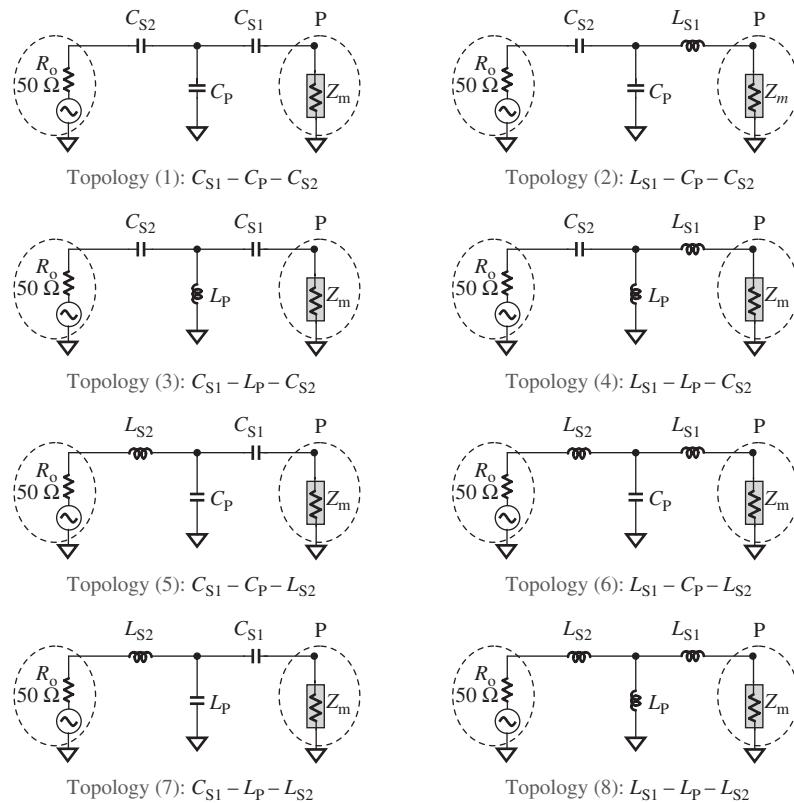


Figure 3.A.19. Topologies of a T-type impedance matching network.

Note 1:  $Z_m$  is the original impedance to be matched.

Note 2: In the title of the circuit topology, the first part marked with subscript 1 is connected to the original impedance to be matched, and the second part marked with subscript 2 is connected to the reference impedance  $50 \Omega$ .

Note 3: Subscript P stands for in parallel and subscript S stands for in series.

As a matter of fact, T and  $\Pi$  type of blocks can be converted from one to the other. The conversion equations from  $\Pi$  to T are

$$Z_{T1} = \frac{Z_{\Pi 1} Z_{\Pi 2}}{Z_{\Pi 1} + Z_{\Pi 2} + Z_{\Pi 3}}, \quad (3.A.99)$$

$$Z_{T2} = \frac{Z_{\Pi 2} Z_{\Pi 3}}{Z_{\Pi 1} + Z_{\Pi 2} + Z_{\Pi 3}}, \quad (3.A.100)$$

$$Z_{T3} = \frac{Z_{\Pi 3} Z_{\Pi 1}}{Z_{\Pi 1} + Z_{\Pi 2} + Z_{\Pi 3}}. \quad (3.A.101)$$

The conversion equations from T to  $\Pi$  are

$$Z_{\Pi 1} = Z_{T3} + Z_{T1} + \frac{Z_{T3} Z_{T1}}{Z_{T2}}, \quad (3.A.102)$$

$$Z_{\Pi 2} = Z_{T1} + Z_{T2} + \frac{Z_{T1}Z_{T2}}{Z_{T3}}, \quad (3.A.103)$$

$$Z_{\Pi 3} = Z_{T2} + Z_{T3} + \frac{Z_{T2}Z_{T3}}{Z_{T1}}. \quad (3.A.104)$$

$T-\Pi$  conversion is also called the “ $*-\Delta$ ” (star–delta) conversion because “ $T$ ” looks like “ $*$ ” and “ $\Pi$ ” looks like “ $\Delta$ .”

Figure 3.A.17 shows that the  $T-\Pi$  or  $*-\Delta$  conversion provides the flexibility of the topology for a matching network, while its input and output impedance can be kept unchanged. The  $T-\Pi$  or  $*-\Delta$  conversion is important to a designer. The reasons are the following:

- Sometimes the part’s value will be found to be too low and comparable with the parasitic parameters. At this point it makes no sense to talk about the reliability of its performance. On another extreme, its performance will also be not reliable if the part’s value is too high. In terms of  $T-\Pi$  or  $*-\Delta$  conversion, the values of the parts may become reasonable and appropriate.
- Sometimes the designer cannot find the part of the required value in the market. By means of the  $T-\Pi$  or  $*-\Delta$  conversion, the value of part can be changed.
- The total number of parts can be reduced or increased in the narrow-band cases.

### 3.A.6 Possible $\Pi$ and $T$ Impedance Matching Networks

The reader is referred to Figure 3.A.18 and 3.A.19.

## REFERENCE

Li Richard Chi-Hsi. *Key Issues in RF/RFIC Circuit Design*, Beijing: Higher Education Press; 2005. pp. 43–75.

## FURTHER READING

Abrie PLD. *The Design of Impedance Matching Networks for Radio-Frequency and Microwave Amplifiers*. Norwood (MA): Artech House; 1985.

Collantes JM, Pollard RD, Sayed M. Effects of DUT mismatch on the noise figure characterization: a comparative analysis of two Y-factor techniques. IEEE Trans Instrum Meas 2002; 51(6): 1150–1156.

Fidler JK, Sun Y. Computer-aided determination of impedance matching domain. IEE 12th Saraga Colloquium on Digital and Analogue Filters and Filtering Systems; London; 1992 Nov 6. pp. 1/1–1/6.

Girlando G, Palmisano G. Noise figure and impedance matching in RF cascode amplifiers. Circ Syst II: Analog Digit Signal Process, IEEE Transactions on [see also Circuits and Systems II: Express Briefs, IEEE Transactions on] 1999; 46(11): 1388–1396.

Gurley J. Impedance matching by means of nonuniform transmission lines. IRE Trans Anntenas Propag 1952; 4(1).

Lapinoja M, Rahkonen T. An active tuning and impedance matching element. Volume 1, Proceedings of the IEEE International Symposium on Circuits and Systems, ISCAS’98; Monterey, CA; 1998 May 31–Jun 3 pp. 559–562.

Rohde UL. Designing a matched low noise amplifier using CAD tools. Microw J 1986; 29(10): 154–160.

- Sun Y, Fidler JK. Design of  $\Pi$  impedance matching networks. Volume 5, IEEE International Symposium on Circuits and Systems, ISCAS '94; Eindhoven; 1994 May 30–Jun 2. pp. 5–8.
- Sun Y, Fidler JK. Design method for impedance matching networks. IEE Proc Circ Dev Syst [see also IEE Proceedings G- Circuits, Devices and Systems] 1996; 143(4): 186–194.
- Vendelin GD, Pavio AM, Rohde UL. *Microwave Circuit Design Using Linear and Nonlinear Techniques*. New York: John Wiley and Sons, Inc; 1990.
- Young GP, Scanlan SO. Matching network design studies for microwave transistor amplifiers. IEEE Trans Microw Theory Tech 1981; MTT-29(10): 1027–1035.

## EXERCISES

1. Why is the impedance matching conducted by means of return loss observation and adjustment?
2. Explain the physical meaning of a return loss of  $-10$  dB and the power reflection coefficient  $\gamma = 0.10$ .
3. What is the variation of impedance after one inductor is connected with it in series?
4. What is the variation of impedance after one capacitor is connected with it in series?
5. What is the variation of impedance after one inductor is connected with it in parallel?
6. What is the variation of impedance after one capacitor is connected with it in parallel?
7. Considering the part count and cost, how many parts (inductors and capacitors) would be a reasonable maximum in the construction of an impedance matching network in the narrow-band case?
8. Why is the Smith Chart divided into four regions in Figure 3.P.1?

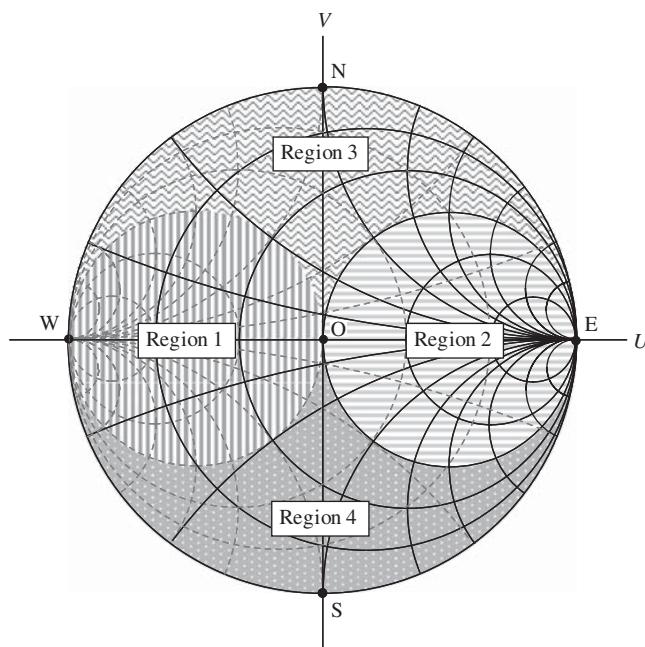


Figure 3.P.1. Divide Smith Chart into four regions.

9. In an impedance matching network with two parts (Fig. 3.P.2), examine the correctness of the two equations below for the calculation of the impedance matching parts' values under the conditions of  $R_m = 50 \Omega$  and  $X_m = 0 \Omega$ .

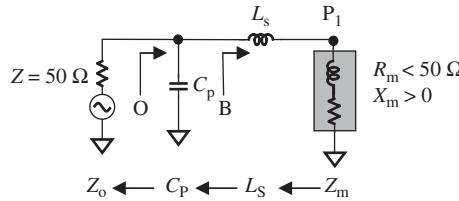


Figure 3.P.2. Impedance matching in region 1.

$$L_s = \frac{\sqrt{R_m(Z_o - R_m)} - X_m}{\omega},$$

$$C_p = \frac{\sqrt{R_m(Z_o - R_m)}}{R_m Z_o \omega}.$$

10. What is the recommended topology for an impedance matching network if it is built with two parts?
11. What is the recommended topology for an impedance matching network if it is built with three parts?
12. Compare the advantages and disadvantages of impedance matching networks with the  $\Pi$  and T configuration.
13. What is the purpose of  $\Pi$  to T conversion?
14. Based on the tested values of block D, calculate the values of block B for the required parameters in the following table:

	Tested values of block D	Calculated values of block B
$G$ (Gain, dB)	-1.5	?
$NF$ , dB	1.5	?
$IIP_3$ , dB <sub>m</sub>	50	?
$IIP_2$ , dB <sub>m</sub>	60	?



15. Based on the tested values of blocks B and C, calculate the values of block A for the required parameters in the following table:

	Tested values of block C	Calculated values of block B	Calculated values of block A
$G$ (Gain, dB)	12	-0.75	?
$NF$ , dB	3	0.75	?
$IIP_3$ , dB <sub>m</sub>	5	52.65	?
$IIP_2$ , dB <sub>m</sub>	40	65.65	?



16. Assume that the value of parts in a T impedance matching network (Figure 3.P.3) are

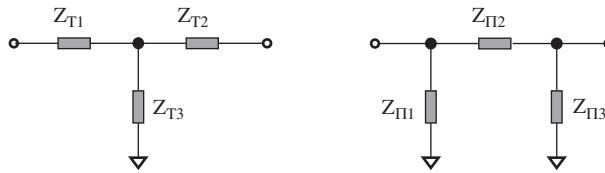


Figure 3.P.3. T and  $\Pi$  conversion.

$$Z_{T1} = jL_1\omega = j80 \Omega,$$

$$Z_{T2} = jL_2\omega = j120 \Omega, \text{ and}$$

$$Z_{T3} = 1/(jC_3\omega) = -j2000 \Omega,$$

calculate  $Z_{\Pi1}$ ,  $Z_{\Pi2}$ ,  $Z_{\Pi3}$ , based on the equations of T- $\Pi$  or \*- $\Delta$  transformation

$$Z_{T1} = (Z_{\Pi1} Z_{\Pi2})/(Z_{\Pi1} + Z_{\Pi2} + Z_{\Pi3}),$$

$$Z_{T2} = (Z_{\Pi2} Z_{\Pi3})/(Z_{\Pi1} + Z_{\Pi2} + Z_{\Pi3}),$$

$$Z_{T3} = (Z_{\Pi3} Z_{\Pi1})/(Z_{\Pi1} + Z_{\Pi2} + Z_{\Pi3}),$$

$$Z_{\Pi1} = Z_{T3} + Z_{T1} + Z_{T3}Z_{T1}/Z_{T2},$$

$$Z_{\Pi2} = Z_{T1} + Z_{T2} + Z_{T1}Z_{T2}/Z_{T3}, \text{ and}$$

$$Z_{\Pi3} = Z_{T2} + Z_{T3} + Z_{T2}Z_{T3}/Z_{T1}.$$

## ANSWERS

1. Return loss measures the attenuation of the return power with respect to the input power. It therefore represents the impedance matching state. In other words, a

stronger attenuation of the return power implies a better impedance matching state. On the contrary, a weaker attenuation of the return power implies a worse impedance matching state. Consequently, one can judge whether impedance matching is done or not on the basis of the observation and adjustment of  $S_{11}$  or  $S_{22}$ .

2. A return loss of  $-10$  dB and a power reflection coefficient  $\gamma = 0.10$  indicate that 10% of the incident power returns and 90% of incident power enters the terminal to be matched.
3. The addition of an inductor  $L_S$  in series results in the original impedance  $P$  moving clockwise along the  $r = \text{constant}$  impedance circle. The moved arc length depends on the value of the inductor.
4. The addition of a capacitor  $C_S$  in series results in the original impedance  $P$  moving counterclockwise along the  $r = \text{constant}$  impedance circle. The moved arc length depends on the value of the capacitor.
5. The addition of an inductor  $L_P$  in parallel results in the original impedance  $P$  moving counterclockwise along the  $g = \text{constant}$  admittance circle. The moved arc length depends on the value of inductor.
6. The addition of a capacitor  $C_P$  in parallel results in the original impedance  $P$  moving clockwise along the  $g = \text{constant}$  admittance circle. The moved arc length depends on the value of capacitor.
7. For an impedance matching network working in the narrow-band case, the reasonable maximum of part count for single-ended configuration is  $\leq 3$ .
8. Dividing the Smith Chart into four regions is beneficial to the construction of an impedance matching network with two parts. From the view point of mathematics, in all four regions the same mathematic language can be applied to the derivation of the equations for calculating the part's value.
9. The answer is:  $L_S = 0$  and  $C_P = 0$ . It implies that impedance is matched already. The parts  $L_S$  and  $C_P$  are not needed.
10. The first part is an inductor in parallel. The second part is a capacitor in series.
11. The first part is an inductor in parallel. The second part is a capacitor in series. The third part is a capacitor in parallel.
12. Type T is better than type  $\Pi$  if DC blocking to the next block is of concern. Type  $\Pi$  is better than type T if the stray reactance in the input of the next block is of concern.
13. The purpose is twofold: the value of parts might become more reasonable, or the number of parts may reduced.
14. Based on the tested values of block D, the table gives the following values for block B for the required parameters:

	Tested values of block D	Calculated values of block B
$G$ (Gain), dB	-1.5	-0.75
$NF$ , dB	1.5	-0.75
$IIP_3$ , dB <sub>m</sub>	50	52.65
$IIP_2$ , dB <sub>m</sub>	60	65.65



15. Based on the tested values of block B and C, the table gives the values of block A for the required parameters:

	Tested values of block C	Calculated values of block B	Calculated values of block A
$G$ (gain), dB	12	-0.75	12.75
$NF$ , dB	3	0.75	2.98
$IIP_3$ , dB <sub>m</sub>	5	52.65	5.00
$IIP_2$ , dB <sub>m</sub>	40	65.65	42.23



16.  $Z_{\Pi 1} = -j2000 \Omega + j80 \Omega - j2000 \Omega \cdot j80 \Omega / j120 \Omega = -j3253 \Omega$

$$Z_{\Pi 2} = j80 \Omega + j120 \Omega - j80 \Omega \cdot j120 \Omega / j2000 \Omega = j195.7 \Omega$$

$$Z_{\Pi 3} = j120 \Omega - j2000 \Omega - j120 \Omega \cdot j2000 \Omega / j80 \Omega = -j4880 \Omega.$$



---

# IMPEDANCE MATCHING IN THE WIDEBAND CASE

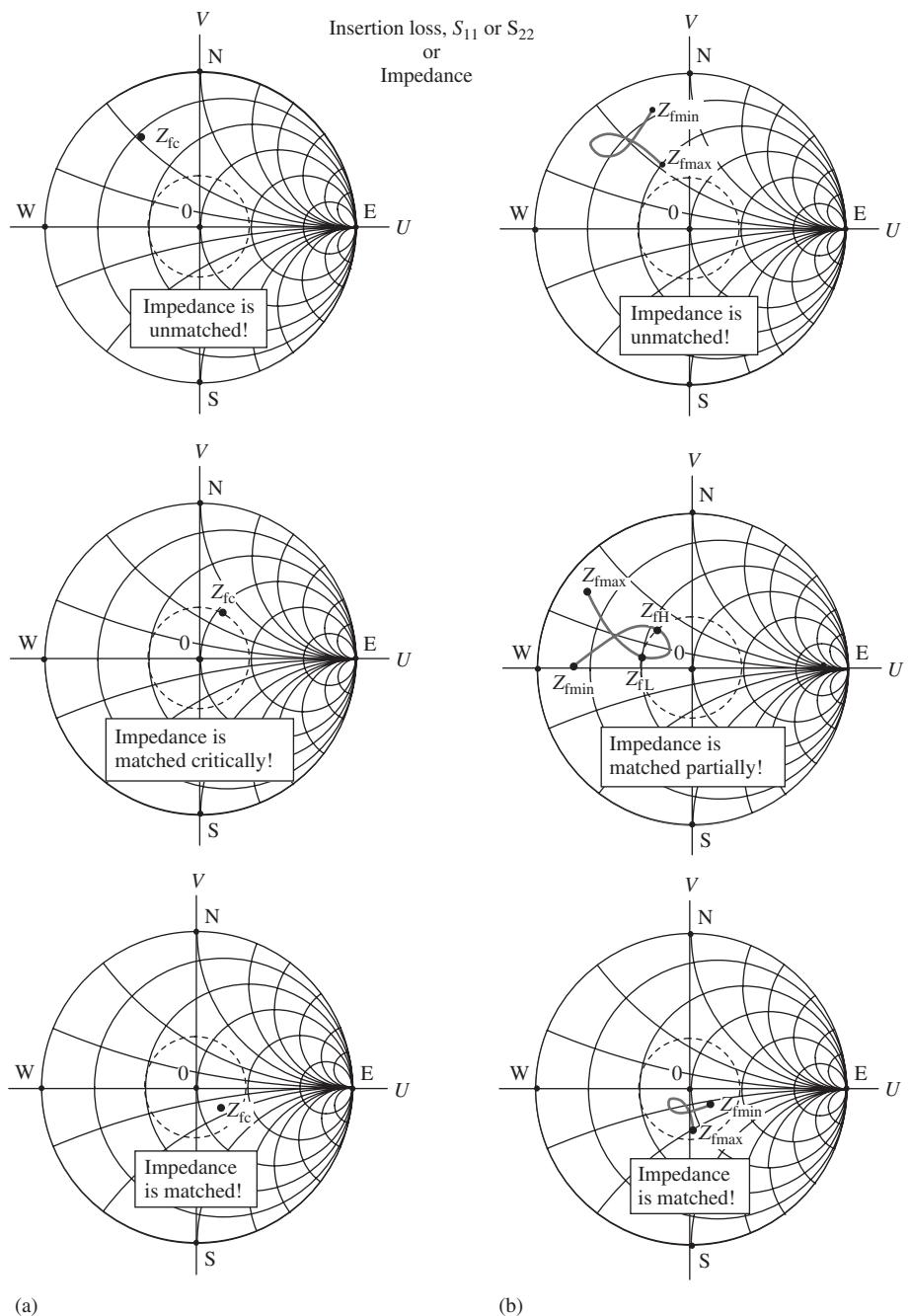
---

## 4.1 APPEARANCE OF NARROW AND WIDEBAND RETURN LOSS ON A SMITH CHART

The methodology of impedance matching is basically the same between narrow and wideband cases. However, in the narrowband case, only one frequency needs to be taken care of in the process of impedance matching because a narrow bandwidth can be represented approximately by the central frequency. In the wideband case, instead of a single frequency, the entire wide bandwidth must be taken care of, which makes impedance matching somewhat difficult. In this chapter, we will apply the same process of impedance matching as in the narrowband case, but the main effort will be to expand the skill from narrow to wide bandwidth. (The methodology of impedance matching in the wideband case introduced and discussed in this chapter was developed in 2005 by the author specially for a UWB (ultra wideband) system and has not been made public until now).

Figure 4.1 shows the difference of the return loss or impedance on the Smith Chart between the narrow and wideband cases. The dashed-line circle in each plot is the demarcation circle of return loss as mentioned in Chapter 3, where  $S_{11}$  or  $S_{22} = -10$  dB.

In the narrowband cases, as shown in Figure 4.1(a), the return loss or impedance is approximated by a central frequency, which is a single point on the Smith Chart denoted by  $Z_{fc}$ . The impedance matching state is clear if the critical circle of return loss  $S_{11}$  or  $S_{22} = -10$  dB is taken as the unique criterion.



**Figure 4.1.** Appearance of return loss in the narrow and wideband cases. (a) Narrowband case.  
(b) Wideband case.

- In the upper-left plot,  $Z_{fc}$  is located outside the demarcation circle so that the impedance is in the unmatched state.
- In the middle-left plot,  $Z_{fc}$  is exactly located on the demarcation circle so that the impedance is in a critically matched state.
- In the bottom-left plot,  $Z_{fc}$  is located inside the demarcation circle so that the impedance is in the matched state.

In the wideband case, as shown in Figure 4.1(b), the trace of  $S_{11}$ ,  $S_{22}$ , or impedance on the Smith Chart is not a point, but a line segment, which is the collective return loss or impedance responding for all the frequencies over the bandwidth. Instead of one point, the impedance matching state must be examined over the entire line segment of  $S_{11}$ ,  $S_{22}$ , or impedance. One end of the trace is marked  $Z_{f\min}$ , which denotes the impedance at the minimum frequency  $f_{\min}$ . The other end of the trace is marked with  $Z_{f\max}$ , which denotes the impedance at the maximum frequency  $f_{\max}$ . The impedance matching state can still be judged if the demarcation circle of return loss,  $S_{11}$  or  $S_{22} = -10$  dB, is taken as the unique criterion again: that is, as shown in Figure 4.1(b),

- in the upper-right plot, the entire trace of return loss or impedance is located outside the demarcation circle so that the impedances are entirely in the unmatched state;
- in the middle-right plot, the trace of return loss or impedance is located partially outside and partially inside the demarcation circle so that the impedances are in a partially matched state;
- in the bottom-right plot, the trace of return loss or impedance is located entirely inside the demarcation circle so that the impedances over the entire frequency bandwidth are in the matched state.

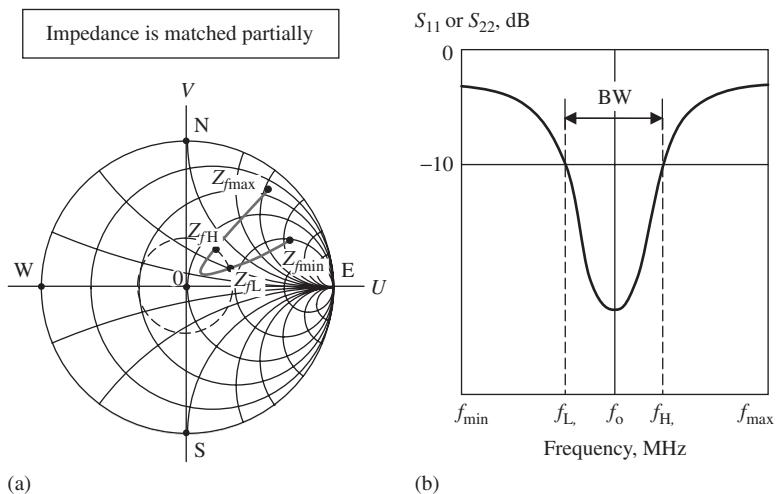
In order to explore the bandwidth from the appearance of return loss on the Smith Chart, let us focus on the middle-right plot in Figure 4.1(b). The return loss trace has two intercept points with the demarcation circle  $Z_{fL}$  and  $Z_{fH}$ . These are the impedances at the frequencies  $f_L$  and  $f_H$ , respectively. The impedances are in a matched state in the frequency interval from  $f_L$  to  $f_H$  so that the bandwidth is found as

$$BW = f_H - f_L, \quad (4.1)$$

whereas the impedances are in the unmatched state in the frequency interval from  $f_{\min}$  to  $f_L$  and from  $f_H$  to  $f_{\max}$ .

Obviously, the frequency bandwidth (4.1) is specially defined for impedance matching but not for the power gain or noise figure, even though they are closely related to each other. This definition is reasonable because those impedances within the bandwidth are closer to the center of the Smith Chart, that is  $50 \Omega$ , and are therefore in an acceptable impedance matching state. Although not a formal rule, engineering design experience has for many years proved that the demarcated return loss circle,  $S_{11}$  or  $S_{22} = -10$  dB, is quite a good criterion for acceptable impedance matching.

The return loss on the Smith Chart can be converted into a Cartesian plot, which provides a more intuitive understanding of the bandwidth with respect to the impedance matching status. Figure 4.2 shows the corresponding relationship between readings from the Smith Chart and from the curves of the return loss  $S_{11}$  or  $S_{22}$  versus the frequency using Cartesian coordinates. The impedances are in a partially matched state.



**Figure 4.2.** Return loss on the Smith Chart and in Cartesian coordinates. (a) Return loss on Smith Chart. (b) Return loss versus frequency in Cartesian coordinates.

The bandwidth for impedance matching thus expressed in Cartesian coordinates is much clearer than on the Smith Chart. Then, why do engineers prefer to use the Smith Chart rather than Cartesian coordinates in implementing impedance matching networks? The reason is that in the successive building process of impedance matching networks, one can judge whether to insert an inductor or capacitor into the impedance matching network from the Smith Chart, but not from the return loss versus frequency plot in Cartesian coordinates. Also, one can calculate the desired value of the necessary inductor or capacitor from the Smith Chart but not from the Cartesian plot. In other words, the implementation of impedance matching network relies on the Smith Chart but not on the return loss versus frequency plot in the Cartesian coordination.

It is therefore desirable to study how to evaluate and control the bandwidth on the basis of the appearance of  $S_{11}$  or  $S_{22}$  on the Smith Chart. What we are more interested in, additionally, is how to expand the bandwidth in the impedance matching process through the implementation of the impedance matching network.

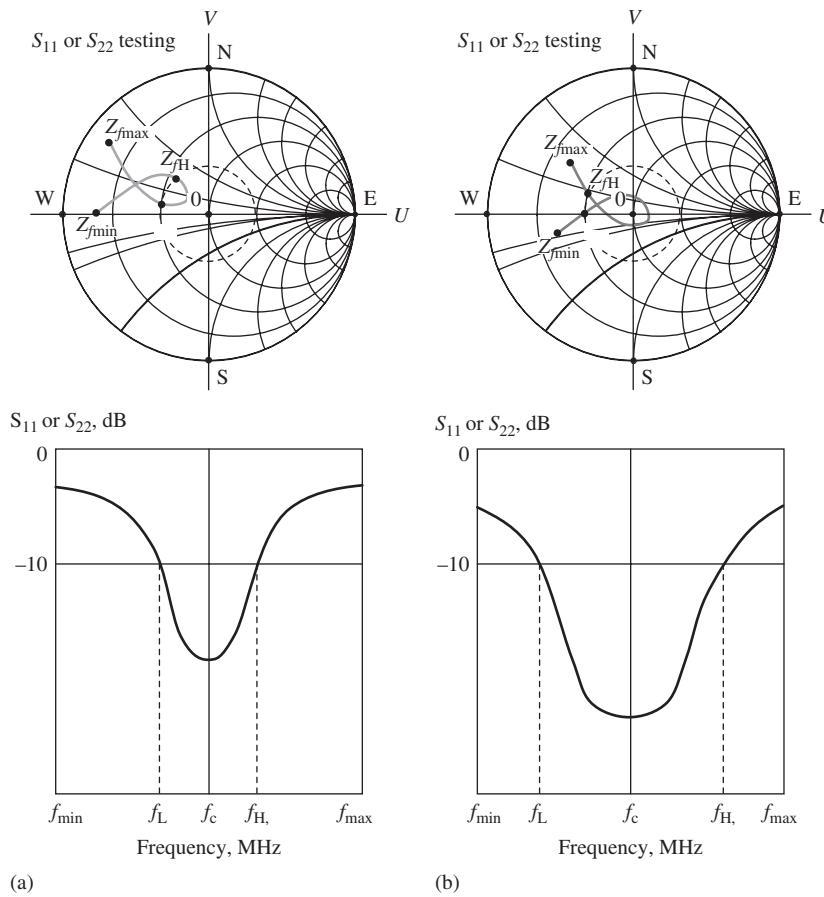
Let us observe the appearance of the trace  $S_{11}$  or  $S_{22}$  on the Smith Chart.

Figure 4.3 shows the dependence of the bandwidth on the location of the return loss on the Smith Chart or on the distance from the trace of  $S_{11}$  or  $S_{22}$  to the center of Smith Chart, that is,  $50 \Omega$ .

The return loss traces of  $S_{11}$  or  $S_{22}$  in both Figures 4.3(a) and 4.3(b) have the same shape and occupy the same area, but the trace in Figure 4.3(a) is farther from the center of Smith Chart than the trace in Figure 4.3(b). A large portion of the trace in Figure 4.3(b) is located inside the demarcation circle,  $S_{11} = -10$  dB, while most of the trace in Figure 4.3(a) is located outside the demarcation circle. Consequently, the trace in Figure 4.3(b) corresponds to a wideband case, while the trace in Figure 4.3(a) corresponds to a narrowband case.

Figure 4.4 shows the dependence of the bandwidth on the density of the return loss or on the occupied area by the trace of  $S_{11}$  or  $S_{22}$  on the Smith Chart.

The return loss traces of  $S_{11}$  or  $S_{22}$  in both Figures 4.4(a) and 4.4(b) have the same shape, but the occupied areas are different. The area occupied by the trace in Figure 4.4(a) is much larger than that occupied by the trace in Figure 4.4(b).



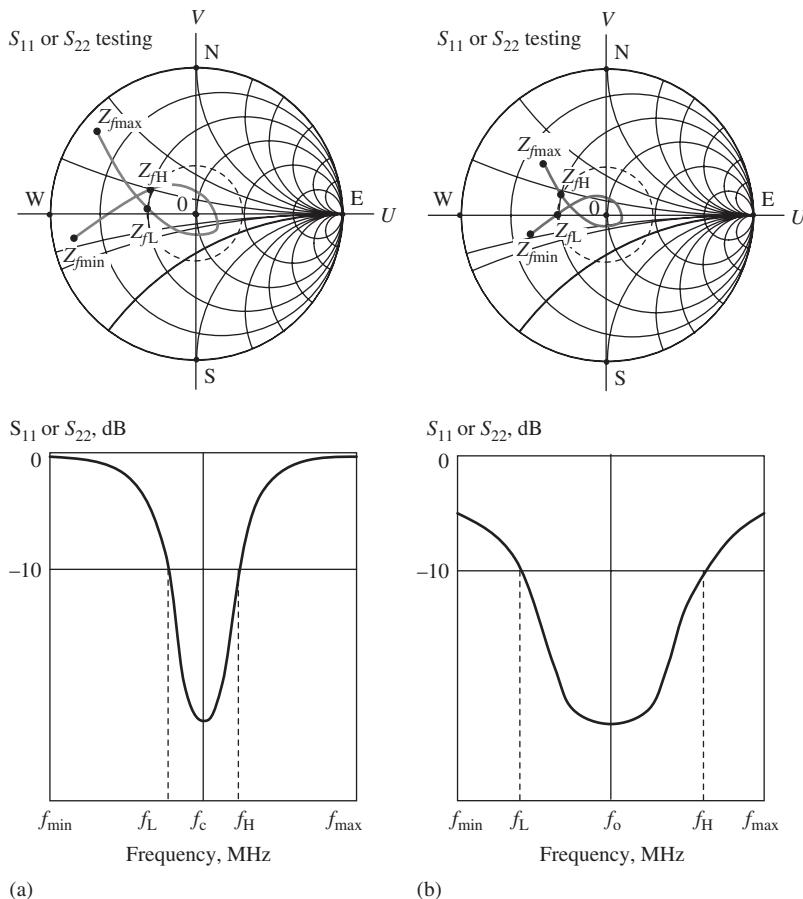
**Figure 4.3.** Bandwidth dependence on the location of return loss trace on the Smith Chart.  
(a) Narrowband response. (b) Wideband response.

The portion of the trace in Figure 4.4(b) located inside the demarcation circle,  $S_{11} = -10$  dB, is larger than that of the trace in Figure 4.4(a). Consequently, the trace in Figure 4.3(b) corresponds to a wideband case, while the trace in Figure 4.3(a) corresponds to a narrowband case.

From the two observations made in Figures 4.3 and 4.4, it is concluded that

- in order to change the impedance matching at a terminal from a narrowband to a wideband response, the distance from the trace of the return loss  $S_{11}$  or  $S_{22}$  (or impedances) to the center of Smith Chart must be as short as possible;
- in order to change the impedance matching at a terminal from a narrowband to a wideband response, the area the impedance trace or the return loss  $S_{11}$  or  $S_{22}$  on the Smith Chart must be shrunk or squeezed to as small an area as possible.

In short, in order to change the impedance matching at a terminal from a narrowband to a wideband response, “squeeze the return loss trace and move it to the center of Smith Chart!”



**Figure 4.4.** Bandwidth dependence on the density of the return loss trace on the Smith Chart.  
(a) Narrowband response. (b) Wideband response

## **4.2 IMPEDANCE VARIATION DUE TO THE INSERTION OF ONE PART PER ARM OR PER BRANCH**

In the passive impedance matching network discussed in Chapter 3, only one part was set forth on one branch (in parallel) or one arm (in series). In Chapter 3, we were only concerned with one frequency which approximated the narrowband case. In this chapter, we will extend our discussion from the narrowband to wideband case. The first thing to do is to expand our discussion of impedance matching in Chapter 3 from one original impedance, or only one point on the Smith Chart, to a collection of impedances contained within a frequency bandwidth. We also resume the discussion begun in Chapter 3 about the variation of impedance due to the insertion of one arm and one branch into the impedance matching network.

In the wideband case, the original impedance appearing on the Smith Chart is a segment of the impedance curve and is no longer a point. Each point on the segment of the impedance curve represents an impedance value corresponding to a frequency within the bandwidth. Each time when a part, either a capacitor or an inductor, is inserted into the impedance matching network, the impedance curve segment will be moved to

a new location on the Smith Chart and its shape, length, and orientation will usually change, depending on the part's value and the connection mode. As discussed in Chapter 3, there are four cases of impedance variation if only one part in one arm or one branch is inserted into the impedance matching network at a time.

#### 4.2.1 An Inductor Inserted into Impedance Matching Network in Series

When an inductor is inserted into the impedance matching network in series, the variation of the original impedance  $\Delta Z$  is equal to the variation of inductive reactance  $\Delta Z_L$ , that is,

$$\Delta Z = \Delta Z_L = j \Delta X_L = jL\omega. \quad (4.2)$$

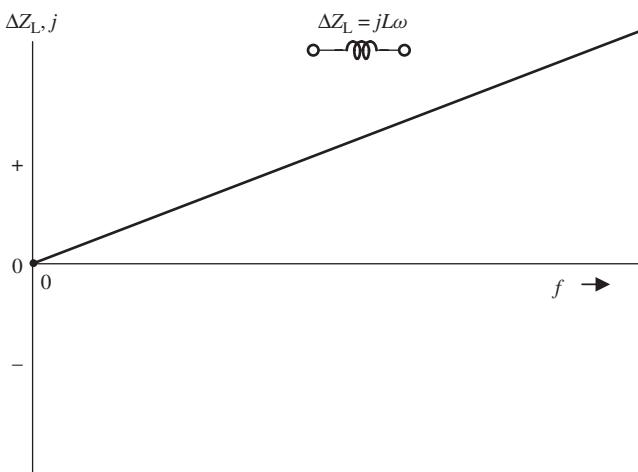
Figure 4.5 plots the curve of  $\Delta Z_L$  versus  $f$  from equation (4.2). The variation of the impedance is an increase of reactance as the frequency is increased.

On the Smith Chart, the impedance trace moves clockwise around the circle  $r = \text{constant}$ , where  $r$  is the resistance value of the original impedance. The reactance increases more at the high-frequency end than at the low-frequency end because the increase of reactance is proportional to the frequency as shown by equation (4.2) or in Figure 4.5.

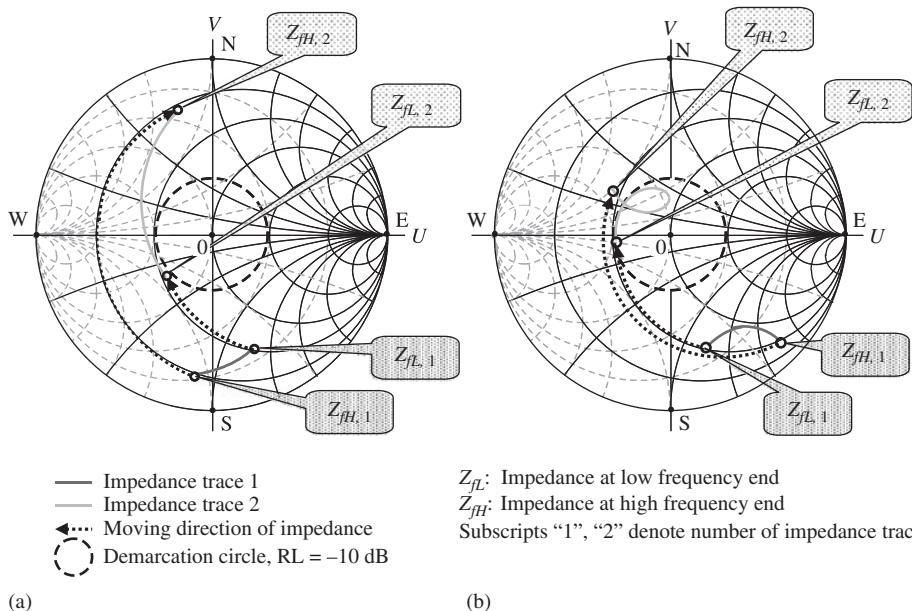
Figure 4.6 shows two examples, both in which the impedance changes from trace 1 to trace 2 after an inductor is inserted into the impedance matching network in series. Figure 4.6(a) shows an example in which the impedance trace has the same even shape before and after the inductor is inserted. Figure 4.6(b) shows an example in which the impedance trace changes its shape from the even shape, before the inductor is inserted, to a curly shape, after the inductor is inserted.

The symbols appearing in Figure 4.6,  $Z_{fL,1}, Z_{fH,1}, Z_{fL,2}$ , and  $Z_{fH,2}$ , denote the impedances at the low and high frequencies of trace 1 and trace 2, and are specified in Figure 4.6. The subscripts  $f_L$  and  $f_H$  denote the low- and high-frequency ends within the bandwidth respectively; subscripts "1" and "2" denote trace 1 and trace 2, respectively.

The variation of the impedance trace from trace 1 to trace 2 can be summarized as follows:



**Figure 4.5.** Variation of impedance when an inductor is inserted into impedance matching network in series.



**Figure 4.6.** Variation of impedance from trace 1 to trace 2 when an inductor is inserted into impedance matching network in series. (a) Impedance trace with an even shape. (b) Impedance trace changed from an even to a curly shape.

1. All the impedances on trace 1 basically move clockwise around the resistance circle, although their resistances increase a little due to the equivalent resistance of the inductor in series.
2. The variation of impedance at the high-frequency end is greater than the variation of impedance at the low-frequency end due to the relation (4.2).
3. The general purpose of inserting an inductor into the impedance matching network in series is usually to move trace 1 to trace 2, so that trace 2 is closer to the reference impedance or the center of the Smith Chart, as shown in Figure 4.6(a) and (b). In Figure 4.6, the circle drawn by the dashed line is the demarcation circle,  $RL = -10 \text{ dB}$ . Most of trace 2 in Figure 4.6(b) is located within the demarcation circle.
4. A remarkable event is the impedance trace’s change from an even shape to a curly shape as shown in Figure 4.6(b) so that the return loss trace is squeezed or rolled up. As mentioned in Section 4.1, this change indicates that the bandwidth of the impedance trace for impedance matching has changed from narrowband to wideband.

The reason for the “rolling” or “squeezing” of the trace is twofold:

- In Figure 4.6(b), the reactance of the original impedance at the high-frequency end is higher than that at the low-frequency end, that is,

$$X_{fL,1} > X_{fH,1}. \quad (4.3)$$

- The variation of impedance at the high-frequency end is much greater than the variation of impedance at the low-frequency end as shown in the relation (4.2) or Figure 4.5.

On the contrary, in Figure 4.6 (a),

$$X_{fL,1} < X_{fH,1}. \quad (4.4)$$

The trace of the original impedance is neither rolled up nor squeezed, and thus it retains its even shape before and after the inductor is inserted into the impedance matching network, even though the variation of impedance at the high-frequency end is much greater than the variation of impedance at the low-frequency end, due to the relation (4.2).

In order to expand the frequency bandwidth through the implementation of the impedance matching network, the designer always looks for or tries to create the original impedance trace satisfying the condition (4.3).

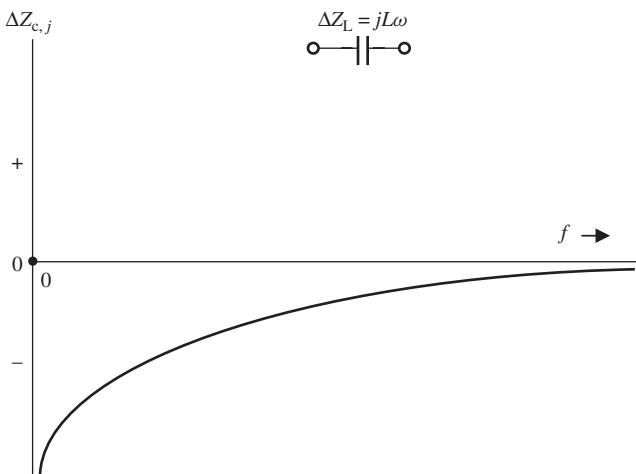
#### 4.2.2 A Capacitor Inserted into Impedance Matching Network in Series

When a capacitor is inserted into the impedance matching network in series, the variation of the original impedance  $\Delta Z$  is equal to the variation of capacitive reactance  $\Delta Z_C$ , that is,

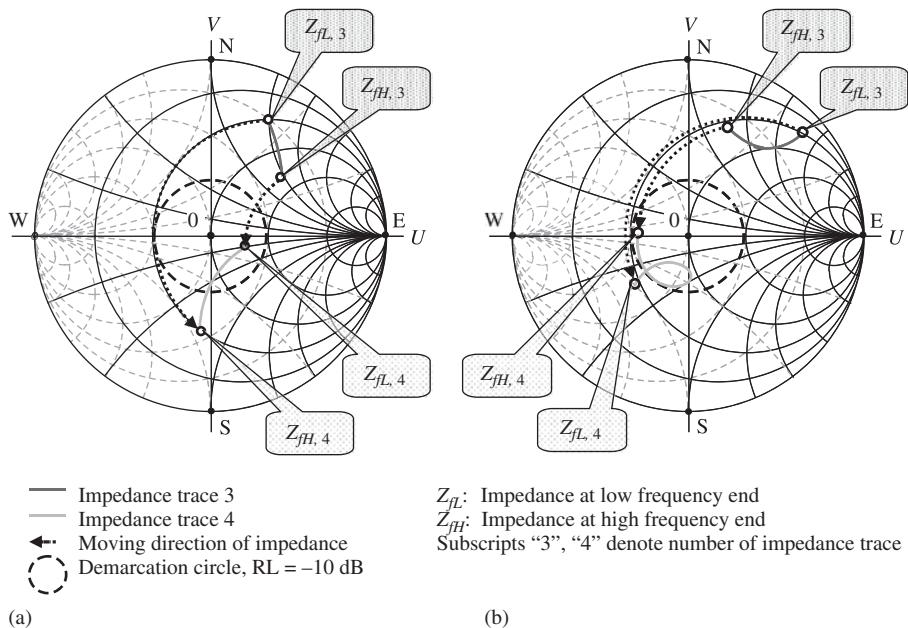
$$\Delta Z = \Delta Z_C = j \Delta X_C = \frac{1}{jC\omega} = -j \frac{1}{C\omega}. \quad (4.5)$$

Figure 4.7 plots the curve of  $\Delta Z_C$  versus  $f$  from equation (4.5). The variation of the impedance is a decrease of reactance in magnitude as the frequency is increased.

On the Smith Chart, the impedance trace moves counterclockwise along the circle of  $r = \text{constant}$ , where  $r$  is the resistance value of the original impedance. The reactance decreases more at the low-frequency end than at the high-frequency end because the decrease of reactance is inversely proportional to the frequency as shown by equation (4.5), or in Figure 4.7.



**Figure 4.7.** Variation of impedance when a capacitor is inserted into impedance matching network in series.



**Figure 4.8.** Variation of impedance from trace 3 to trace 4 when a capacitor is inserted into impedance matching network in series. (a) Impedance trace kept in an even shape. (b) Impedance trace changed from an even to a curly shape.

Figure 4.8 shows an example in which the impedance variation from trace 3,  $Z_{fL,3}$  and  $Z_{fH,3}$ , to trace 4,  $Z_{fL,4}$  and  $Z_{fH,4}$ , over the bandwidth  $f_L$  to  $f_H$ , before and after a capacitor is inserted into the impedance matching network in series.

The variation of the impedance trace from trace 3 to trace 4 can be summarized as follows:

1. All the impedances on trace 3 basically move counterclockwise around their own resistance circles.
2. The variation of impedance at the low-frequency end is greater than the variation of impedance at the high-frequency end due to the relation (4.5).
3. The general purpose of inserting the capacitor  $C_S$  into the impedance matching network in series is usually to move trace 3 to trace 4, so that trace 4 is closer to the reference impedance, that is, the center of the Smith Chart, as shown in Figure 4.8 (a) and (b). In Figure 4.8, the circle drawn by the dashed line is the demarcation circle,  $RL = -10 \text{ dB}$ . Most of trace 4 in Figure 4.8(b) is located within the demarcation circle.
4. A remarkable event is the impedance trace's change from an even shape to a curly shape as shown in Figure 4.8(b) so that the return loss trace is squeezed or rolled up. As mentioned in Section 4.1, this change indicates that the bandwidth of the impedance trace for impedance matching has changed from narrowband to wideband.

The reason for the rolling or squeezing of the trace is twofold:

- In Figure 4.8(b), the reactance of the original impedance at the low-frequency end is higher than that at high-frequency end, that is,

$$X_{fL,3} > X_{fH,3}; \quad (4.6)$$

- The variation of impedance at the low-frequency end is much greater than the variation of impedance at the high-frequency end as shown in the relation (4.5) or Figure 4.7.

On the contrary, in Figure 4.8(a),

$$X_{fL,3} < X_{fH,3}. \quad (4.7)$$

The trace of the original impedance is neither rolled up nor squeezed, and thus it retains its even shape before and after the capacitor is inserted into the impedance matching network, even though the variation of impedance at the high-frequency end is much greater than the variation of impedance at the low-frequency end due to the relation (4.5).

In order to expand the frequency bandwidth through the implementation of the impedance matching network, the designer always looks for or tries to create the original impedance trace satisfying the condition (4.6).

#### 4.2.3 An Inductor Inserted into Impedance Matching Network in Parallel

When an inductor is inserted into the impedance matching network in parallel, the variation of the original admittance  $\Delta Y$  is equal to the variation of inductive admittance  $\Delta Y_L$ , that is,

$$\Delta Y = \Delta Y_L = \frac{1}{\Delta Z_L} = j \Delta B_L = \frac{1}{jL\omega} = -j \frac{1}{L\omega}. \quad (4.8)$$

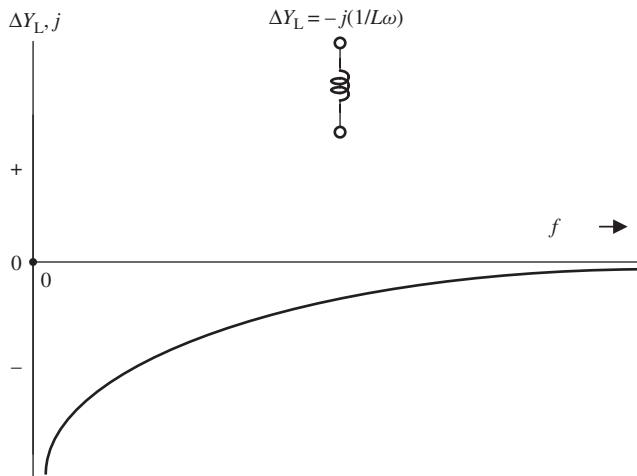
Figure 4.9 plots the curve of  $\Delta Y_L$  versus  $f$  from equation (4.8). The variation of impedance also results in a decrease of susceptance in magnitude as the frequency is increased.

On the Smith Chart, the impedance trace moves counterclockwise around the conductance circle. The susceptance decreases more at the low-frequency end than at the high-frequency end because the decrease of susceptance is inversely proportional to the frequency, as shown by equation (4.8), or in Figure 4.9.

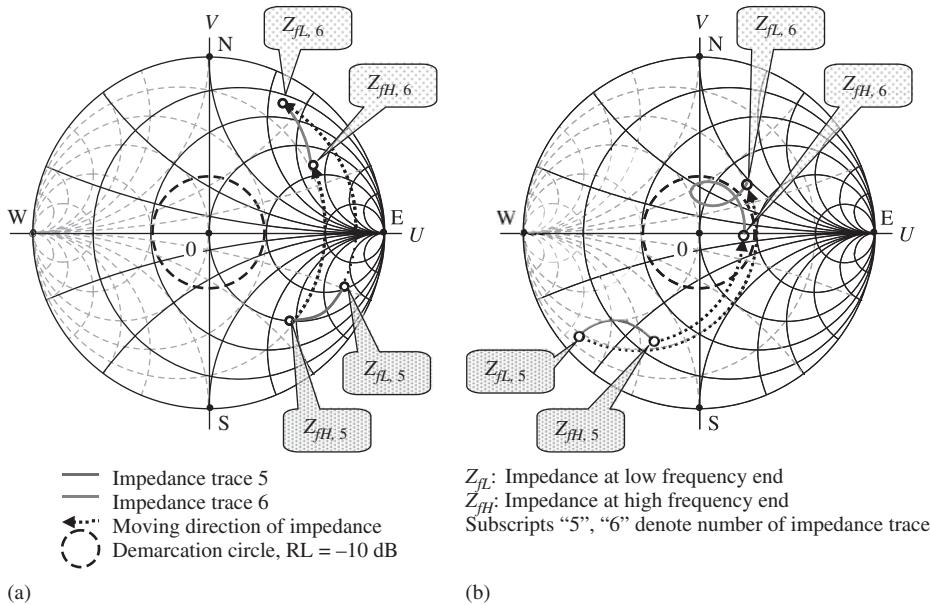
Figure 4.10 shows an example in which the impedance variation from trace 5,  $Z_{fL,5}$  and  $Z_{fH,5}$ , to trace 6,  $Z_{fL,6}$  and  $Z_{fH,6}$ , over the bandwidth  $f_L$  to  $f_H$ , before and after an inductor is inserted into the impedance matching network in parallel.

The variation of the impedance trace from trace 5 to trace 6 can be summarized as follows:

1. All the impedances on the trace 5 basically move counterclockwise around their own conductance circles.
2. The variation of impedance or admittance at the low-frequency end is greater than the variation of impedance at the high-frequency end due to the relation (4.8).



**Figure 4.9.** Variation of admittance when an inductor is inserted into impedance matching network in parallel.



**Figure 4.10.** Variation of impedance from trace 5 to trace 6 when an inductor is inserted into impedance matching network in parallel. (a) Impedance trace kept in even shape. (b) Impedance trace changed from an even to a curly shape.

3. The general purpose of inserting the inductor  $C_P$  into the impedance matching network in parallel is usually to move trace 5 to trace 6, so that trace 6 is closer to the reference impedance, that is, the center of the Smith Chart, as shown in Figure 4.10(a) and 4.10(b). In Figure 4.10, the circle drawn by the dashed line is the demarcation circle  $RL = -10 \text{ dB}$ . Most of trace 6 in Figure 4.10(b) is located within the demarcation circle.
4. A remarkable event is the impedance trace's change from an even shape to a curly shape as shown in Figure 4.10(b) so that the return loss trace is squeezed or

rolled up. As mentioned in Section 4.1, this change indicates that the bandwidth of the impedance trace for impedance matching has changed from narrowband to wideband.

The reason of rolling or squeezing of the trace is twofold:

- In Figure 4.10(b), the susceptance of the original impedance at low-frequency end is higher than that at high-frequency end, that is,

$$B_{fL,5} > B_{fH,5}. \quad (4.9)$$

- The variation of susceptance at the low-frequency end is much greater than the variation of impedance at the high-frequency end as shown by the relation (4.8), or in Figure 4.9.

On the contrary, in Figure 4.10 (a),

$$B_{fL,5} < B_{fH,5}. \quad (4.10)$$

The trace of the original impedance or admittance is neither rolled up nor squeezed, and thus it retains its even shape before and after the inductor is inserted into the impedance matching network, even though the variation of susceptance at the low-frequency end is much greater than that at the high-frequency end due to the relation (4.8).

In order to expand the frequency bandwidth through the implementation of the impedance matching network, the designer always looks for or tries to create the original trace of impedance or admittance satisfying of the condition (4.8).

#### 4.2.4 A Capacitor Inserted into Impedance Matching Network in Parallel

When a capacitor is inserted into the impedance matching network in parallel, the variation of the original admittance  $\Delta Y$  is equal to the variation of capacitive admittance  $\Delta Y_C$ , that is,

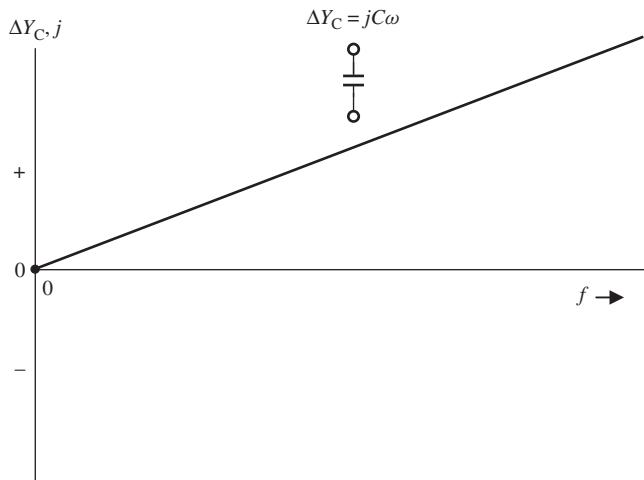
$$\Delta Y = \Delta Y_C = \frac{1}{\Delta Z_C} = j \Delta B_C = jC \omega. \quad (4.11)$$

Figure 4.11 plots the curve of  $\Delta Y_C$  versus  $f$  from equation (4.11). The variation of admittance also results in an increase of susceptance  $\Delta B_C$  as the frequency is increased.

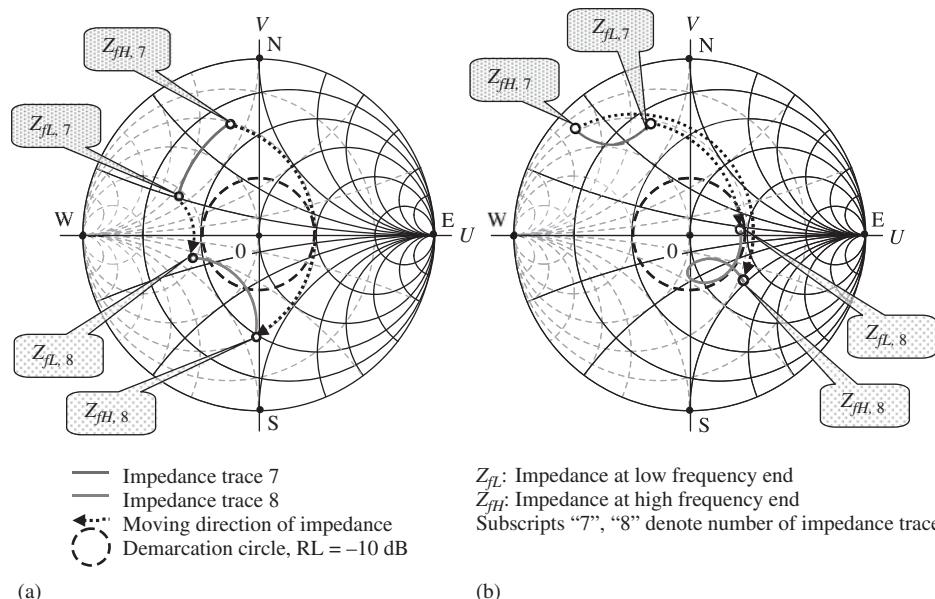
On the Smith Chart, the impedance trace moves clockwise along the circle  $g = \text{constant}$ , where  $g$  is the resistance value of the original impedance. The susceptance increases more at the high-frequency end than at the low-frequency end because the increase of susceptance is proportional to the frequency as shown by equation (4.11) or in Figure 4.11.

Figure 4.12 shows an example in which the impedance variation from trace 7,  $Z_{fL,7}$  and  $Z_{fH,7}$ , to trace 8,  $Z_{fL,8}$  and  $Z_{fH,8}$ , covering the bandwidth  $f_L$  to  $f_H$ , before and after a capacitor is inserted into the impedance matching network in series.

The variation of the impedance trace from trace 7 to trace 8 can be summarized as follows:



**Figure 4.11.** Variation of admittance when a capacitor is inserted into impedance matching network in parallel.



**Figure 4.12.** Variation of impedance from trace 7 to trace 8 when a capacitor is inserted into impedance matching network in parallel. (a) Impedance trace kept in an even shape. (b) Impedance trace changed from an even to a curly shape.

1. All the impedances on the trace 7 basically move clockwise around their own conductance circles.
2. The variation of admittance at the high-frequency end is greater than the variation of admittance at the low-frequency end due to the relation (4.10).
3. The general purpose of inserting the capacitor in parallel is to move trace 7 to trace 8 so that trace 8 is closer to the reference impedance, that is, the center of the Smith Chart, as shown in Figure 4.12(a) and (b). In Figure 4.12, the circle drawn by the dashed line is the demarcation circle,  $RL = -10 \text{ dB}$ . Most of trace 8 in Figure 4.12(b) is located within the demarcation circle.

4. A remarkable event is the impedance trace's change from an even shape to a curly shape as shown in Figure 4.12(b) so that the return loss trace is squeezed or rolled up. As mentioned in Section 4.1, this change indicates that the bandwidth of the impedance trace for impedance matching has changed from narrowband to wideband.

The reason for the rolling or squeezing of the trace is twofold:

- In Figure 4.12(b), the susceptance of the original impedance at the low-frequency end is higher than that at the high-frequency end, that is,

$$B_{fL,7} > B_{fH,7}. \quad (4.12)$$

- The variation of susceptance at the high-frequency end is much greater than the variation of susceptance at the low-frequency end as shown by the relation (4.11) or Figure 4.11.

On the contrary, in Figure 4.12(a),

$$B_{fL,7} < B_{fH,7}. \quad (4.13)$$

The trace of the original impedance or admittance is neither rolled up nor squeezed, and thus retains its even shape before and after the inductor is inserted into the impedance matching network, even though the variation of susceptance at the low-frequency end is much greater than that at the high-frequency end due to the relation (4.11).

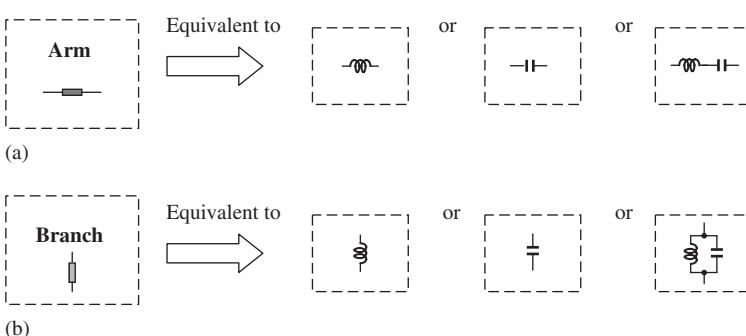
In order to expand the frequency bandwidth through the implementation of an impedance matching network, the designer always looks for or tries to create the original trace of impedance or admittance satisfying the condition (4.12).

### 4.3 IMPEDANCE VARIATION DUE TO THE INSERTION OF TWO PARTS PER ARM OR PER BRANCH

In order to expand the bandwidth in the process of impedance matching, rolling up or squeezing the trace of impedance on the Smith Chart is the right step.

In the examples shown in Figures 4.6(b), 4.8(b), 4.10(b), and 4.12(b), the trace of impedance is changed from an even to a curly shape. This indicates that it is possible to roll up or squeeze the trace of the impedance on the Smith Chart by inserting only one part per arm or per branch. However, the orientation of the original impedance trace on Smith Chart is restricted by any one of the conditions (4.3), (4.6), (4.9), or (4.12). As a matter of fact, this restriction results from the fact that the low- and high-frequency ends of the original impedance trace move in the same direction no matter whether the inserted part is an inductor or a capacitor, whether in series or in parallel. This can be easily verified from Figures 4.5, 4.7, 4.9, and 4.11. The variation of reactance or susceptance for all the frequencies has the same sign, either positive or negative, though the magnitude of the variation over the frequencies is different.

Should the low- and high-frequency ends at the impedance trace move in opposite directions after one arm or one branch is inserted into the impedance matching network, the impedance trace could be easily squeezed, rolled, or shrunk to a small area on the



**Figure 4.13.** Representation of an arm or a branch consisting of one part or two parts. (a) Arm consists of one part or two parts and in the impedance matching network it is a connected component in the horizontal direction (in series). (b) Branch consists of one part or two parts and in the impedance matching network it is a connected component in the vertical direction (in parallel).

Smith Chart and, therefore, the bandwidth could be expanded to what is expected. Instead of one part per arm or one part per branch, two parts per arm or two parts per branch can execute such a function. In other words, in order to expand the bandwidth in a more effective way, one new arm and one new branch of parts will be added to the list of arms and branches. Figure 3.4 now is modified to Figure 4.13. The new arm is an LC combination in series and the new branch is an LC combination in parallel, which is shown in the last column in Figure 4.13. Let us introduce their performances in the following section.

### 4.3.1 Two Parts Connected in Series to Form One Arm

When a combination of LC in series is inserted into the impedance matching network as an arm, the variation of the original impedance  $\Delta Z$  is equal to

$$\Delta Z = \Delta Z_S = jL_S\omega + \frac{1}{jC_S\omega} = jL_S\omega \left( 1 - \frac{1}{jL_S C_S \omega^2} \right) = jL_S\omega \left( 1 - \frac{\omega_S^2}{\omega^2} \right), \quad (4.14)$$

where the subscript “S” denotes that a part is connected “in series,”  $\Delta Z_S$  is the resultant impedance of LC combination in series, and

$$\omega_S^2 = (2\pi f_S)^2 = \frac{1}{L_S C_S}, \quad (4.14a)$$

where

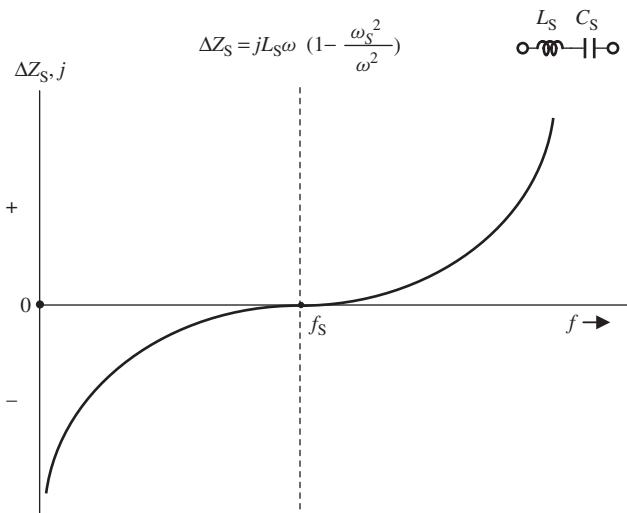
$\omega_S$  = the angular SRF (self-resonant frequency) of LC combination in series,

and

$f_S$  = SRF of the LC combination in series, which appears in Figure 4.14.

Figure 4.14 plots the curve of  $\Delta Z_S$  versus  $f$  from equation (4.14). It can be seen that

1. In the case of LC combination in series,
  - the variation of impedance is capacitive reactance when the operating frequency is lower than the SRF,
  - the variation of impedance is inductive reactance when frequency is the higher than the SRF, and
  - the variation of impedance is zero at the SRF;



**Figure 4.14.** Impedance of the LC combination in series.

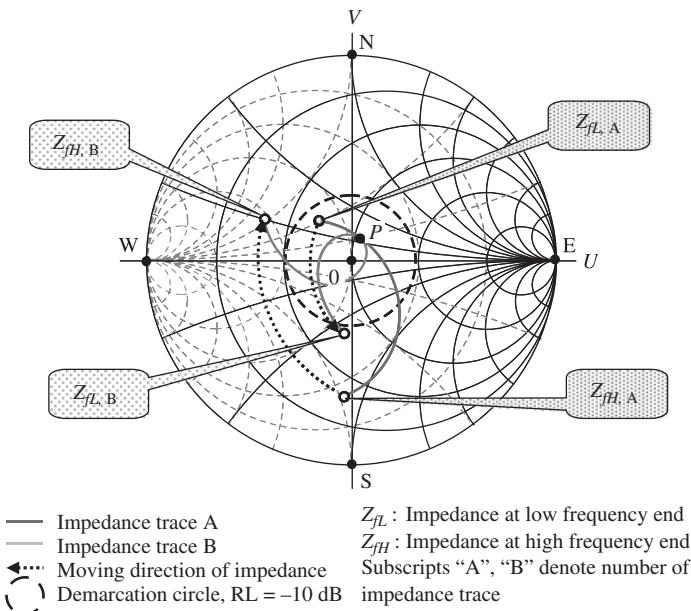
2. The LC combination in series can be inserted into an impedance matching network as an arm in series only. It is prohibited to apply it as a branch in parallel because the network would be short-circuited at the SRF where its resultant impedance would be zero.

In contrast to the previous case where one part per arm or one part per branch is inserted into the impedance matching network, the variation of impedance at low frequencies is *different* from that at high frequencies if the values of  $L_S$  and  $C_S$  are adjusted so that the SRF of the LC combination is the central frequency of the desired bandwidth. Then, the variation of impedance at the low-frequency end is capacitive, or negative, and the variation of impedance at the high-frequency end is inductive, or positive. This means that the low-frequency end of the impedance trace will be moved counterclockwise, whereas the high-frequency end will be moved clockwise on the Smith Chart. They move in opposite directions, so the impedance trace is expected to be rolled up into a smaller area. This is the goal we pursue.

Figure 4.15 shows an example before and after the LC combination in series is inserted into the impedance matching network as an arm. The impedance trace is moved from trace A,  $Z_{fL,A}$  and  $Z_{fH,A}$ , to trace B,  $Z_{fL,B}$  and  $Z_{fH,B}$ , over the bandwidth  $f_L$  to  $f_H$ .

The variation of the impedance trace from trace A to trace B can be summarized as follows:

1. The low-frequency end on the impedance trace A,  $Z_{fL,A}$ , basically moves counterclockwise along the reactance circle to  $Z_{fL,B}$ .
  2. The high-frequency end on the impedance trace A,  $Z_{fH,A}$ , moves clockwise along the reactance circle to  $Z_{fH,B}$ .
- A small deviation of the trace from the resistance circle exists because of a small resistance in the imperfect inductor  $L_S$ .
3. The common point P on trace A and trace B in Figure 4.15 corresponds to the impedance at  $\omega_S$ , or the SRF frequency, where the variation of impedance is zero on the basis of the equation (4.14).



**Figure 4.15.** Variation of impedance from trace A to trace B before and after the LC combination in series is inserted into impedance matching network as an arm.

- Consequently, trace B is rolled around the reference impedance, that is, the center of the Smith Chart, and therefore increases the bandwidth from trace A.

To insert an LC combination in series into an impedance matching network as an arm is a powerful scheme to squeeze the trace and, consequently, to expand the bandwidth.

### 4.3.2 Two Parts Are Connected in Parallel to Form One Branch

When a combination of LC in parallel is inserted into an impedance matching network as a branch, the variation of the original admittance  $\Delta Y$  is also equal to the variation of susceptance  $\Delta B_P$ , that is,

$$\Delta Z = j \Delta X_P = \frac{j L_P \omega \frac{1}{j C_P \omega}}{j L_P \omega + \frac{1}{j C_P \omega}} = \frac{1}{j C_P \omega \left( 1 - \frac{\omega_P^2}{\omega^2} \right)}, \quad (4.15)$$

$$\Delta Y = j \Delta B_P = \frac{1}{\Delta Z} = j C_P \omega \left( 1 - \frac{\omega_P^2}{\omega^2} \right), \quad (4.16)$$

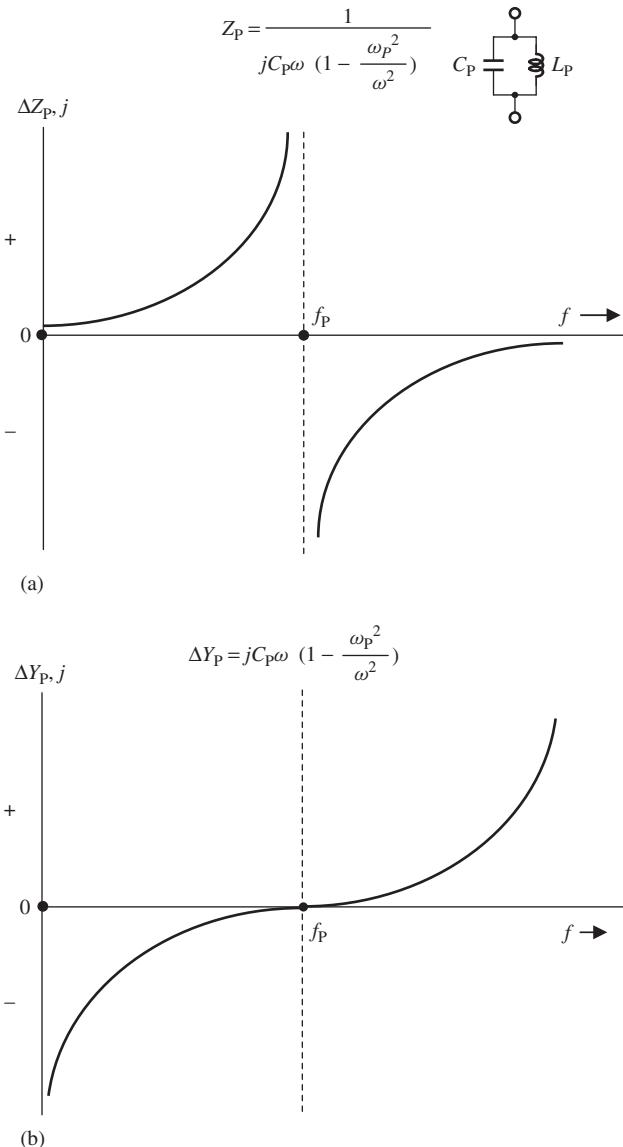
where

$\Delta Z$  = the variation of impedance due to insertion of LC combination in parallel,

$\Delta X_P$  = the variation of reactance due to insertion of LC combination in parallel,

$\Delta Y$  = the variation of admittance due to insertion of LC combination in parallel,

$\Delta Y_P$  = the variation of susceptance due to insertion of LC combination in parallel, and subscript P denotes that a part is connected in parallel,



**Figure 4.16.** Impedance or admittance of an LC combination in parallel. (a) Impedance of an LC combination in parallel. (b) Admittance of an LC combination in parallel.

and

$$\omega_p^2 = (2\pi f_p)^2 = \frac{1}{L_p C_p}, \quad (4.17)$$

where

$\omega_p$  = the angular SRF of LC combination in parallel, and

$f_p$  = the SRF of LC combination in parallel and is applied in Figure 4.16.

On the basis of equations (4.15) and (4.16), Figure 4.16(a) and 4.16(b) plots the curves of  $\Delta Z_p$  versus  $f$  and  $\Delta Y_p$  versus  $f$ , respectively. Figure 4.16(b) is similar to Figure 4.14.

In the case of an LC combination in parallel, admittance is a more convenient parameter than impedance. Therefore, only Figure 4.16(b) but not Figure 4.16(a) is used in our discussion. From Figure 4.16(b) it can be seen that

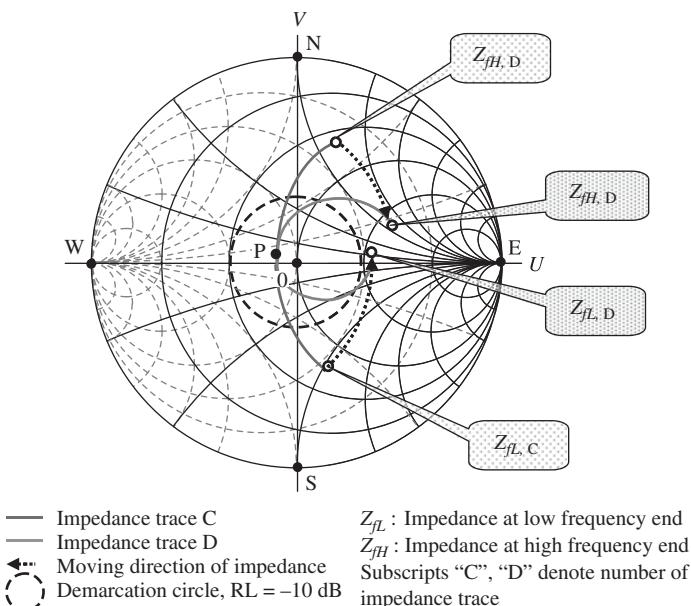
1. in the case of an LC combination in parallel,
  - the variation of admittance is a decrease of susceptance when the frequency is lower than  $f_p$  or the SRF,
  - the variation of admittance is an increase of susceptance when the frequency is higher than  $f_p$  or the SRF, and
  - the variation of susceptance is zero at the SRF;
2. LC combination in parallel can be inserted into an impedance matching network as a branch in parallel only. It is prohibited to apply it as an arm in series because the network would be open-circuited at  $f_p$  or SRF where its resultant impedance is infinite as shown in Figure 4.16(a).

Figure 4.17 shows an example before and after an LC combination in parallel is inserted into impedance matching network as a branch, in which the impedance varies from trace C,  $Z_{fL,C}$  and  $Z_{fH,C}$ , to trace D,  $Z_{fL,D}$  and  $Z_{fH,D}$ , over the bandwidth  $f_L$  to  $f_H$ .

The variation of impedance or admittance trace from trace C to trace D can be summarized as follows:

1. The low-impedance end on trace C,  $Z_{fL,C}$ , moves counterclockwise along the conductance circle to  $Z_{fL,D}$ .
2. The high-frequency end on the impedance trace C,  $Z_{fH,C}$ , moves clockwise along with the conductance circle to  $Z_{fH,D}$ .

A small deviation of the trace from the conductance circle exists because of a small resistance existing in the imperfect inductor  $L_p$ .



**Figure 4.17.** Variation of impedance from trace C to trace D before and after the LC combination in parallel is inserted into impedance matching network as a branch.

3. The common point P on trace C and trace D in Figure 4.17 corresponds to the impedance or admittance at  $\omega_p$  or the SRF frequency in parallel, where the variation of admittance is zero on the basis of the equation (4.16).
4. Consequently, trace C is squeezed to trace D and is now located around the reference impedance, which is the center of the Smith Chart. The bandwidth is therefore increased.

To insert an LC combination in parallel into the impedance matching network is another powerful scheme to squeeze the impedance trace and, consequently, to expand the bandwidth.

#### **4.4 PARTIAL IMPEDANCE MATCHING FOR AN IQ (IN PHASE QUADRATURE) MODULATOR IN A UWB (ULTRA WIDE BAND) SYSTEM**

So far we have been exploring how to expand the bandwidth of an RF circuit block through the implementation of an impedance matching network. This discussion has been conducted qualitatively but not quantitatively. Now, as a quantitative example, let us introduce a partial impedance matching work for an IQ modulator in a UWB system. This is enough to confirm the possibility of expanding the bandwidth of an RF block in a practical circuit design, although this example is only a partial but not a complete IQ modulator design.

##### **4.4.1 Gilbert Cell**

A Gilbert cell is selected as the core of the IQ modulator. As shown in Figure 4.18, it consists of six devices, which in this example are MOSFETs. It has two pairs of differential inputs and one pair of differential output. For a modulator design, it is supposed that

- the IF (intermediate frequency) signal goes to the gates of  $M_1$  and  $M_2$ , which are a pair of differential input;
- the LO (Local Oscillation) injection goes to the gates of  $M_3$  to  $M_6$ , which are another pair of differential input;
- the RF-modulated carrier comes from the drains of  $M_3$  to  $M_6$ , which are a pair of differential output.

To make the Gilbert cell operate, some necessary parts must be connected. Figure 4.19 shows the setup of an operative Gilbert cell for the purpose of circuit simulation. The DC power supply with output voltage  $V_{dd}$  provides DC electricity to the Gilbert cell through the inductors  $L_{P1}$  and  $L_{P2}$  while a bias voltage source also provides an appropriate voltage to the main switching transistors  $M_3$  to  $M_6$  through inductors  $L_{P3}$  and  $L_{P4}$ .

In order to avoid the disturbance of the Gilbert cell, the impedances of all the inductors  $L_{P1}$ ,  $L_{P2}$ ,  $L_{P3}$  and  $L_{P4}$  must approach infinity. The capacitor,  $C_0$ , is a “zero” capacitor, which approaches zero impedance at the operating frequency. In short, the combinations ( $L_{P1}$ ,  $L_{P2}$ , and  $C_0$ ), ( $L_{P3}$ ,  $L_{P4}$ , and  $C_0$ ) provide DC voltage or current but do not disturb the performance of the Gilbert cell.

On the other hand, for simplicity in the simulation stage, all three differential pairs are connected to an ideal transformer balun, respectively, so that the differential pairs become single-ended terminals. Each transformer balun consists of two identical transformers

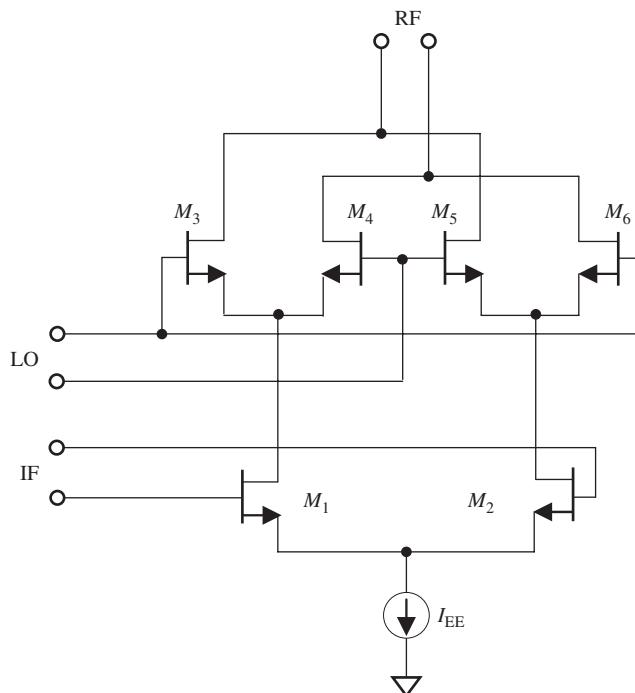


Figure 4.18. Gilbert cell.

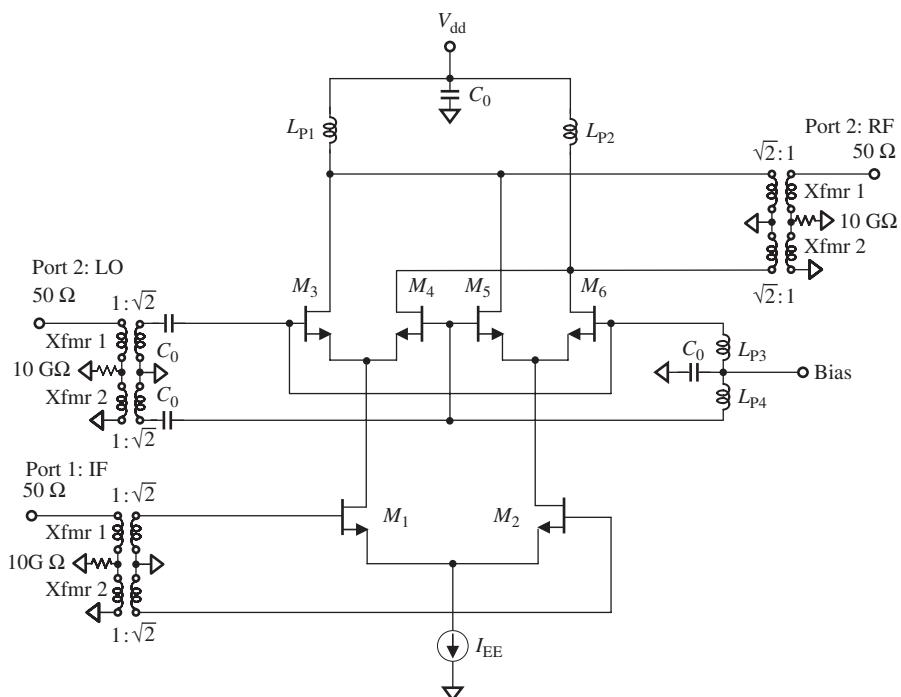
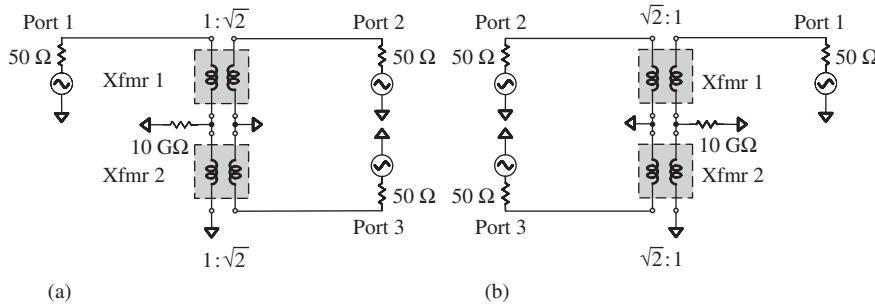


Figure 4.19. Setup of an operative Gilbert cell for simulation.



**Figure 4.20.** Transformer balun built by two transformers stacked together. (a) For LO and IF ports, the turns ratio =  $1:\sqrt{2}$ . (b) For RF ports, the turns ratio =  $\sqrt{2}:1$ .

stacked together. The turns ratio the of the transformer is  $1:(\sqrt{2})$  at the IF and LO inputs but is  $(\sqrt{2}):1$  at the RF output. Figure 4.20 depicts this special transformer balun, which is easily set up for the simulation of circuitry in a Cadence or ADS (Advanced Design System) system. Its special feature is that the impedances looking into the transformer either at the individual differential port or at the single-ended port are the same. The parts attached to the differential ports can be interpreted as the parts at the single-ended port or vice versa, which will be discussed in Chapter 15.

The transformer balun provides a convenient means to convert a differential port to a single-ended port. It reduces complexity a lot in the simulation stage of the design. However, it is only temporarily applied for the circuit simulation since ideal parts are never used in a product. It will be removed or replaced by a practical balun if needed.

In the setup for an operational Gilbert cell, inductors with infinite impedance or capacitors with zero impedance never exist in reality. As a reasonable approximation, the values of inductors and capacitors in Figure 4.19 are taken as

$$L_{P1} = L_{P2} = L_{P3} = L_{P4} = 1000 \text{ nH},$$

$$C_0 = 1000 \text{ pF},$$

where the operating frequency is assumed to be in the order of gigahertz (GHz).

#### 4.4.2 Impedances of the Gilbert Cell

The input and output impedances of a Gilbert cell can be obtained by the return loss  $S_{11}$  or  $S_{22}$ . The testing is conducted under the following conditions:

- Frequency range

$$f = 10 \text{ MHz to } 4.3 \text{ GHz.}$$

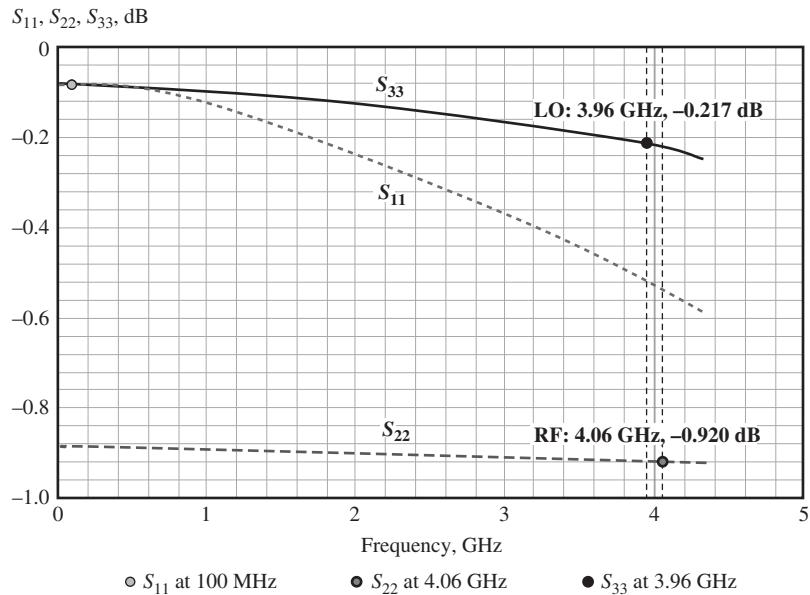
- DC power supply and bias

$$V_{dd} = 3 \text{ V,}$$

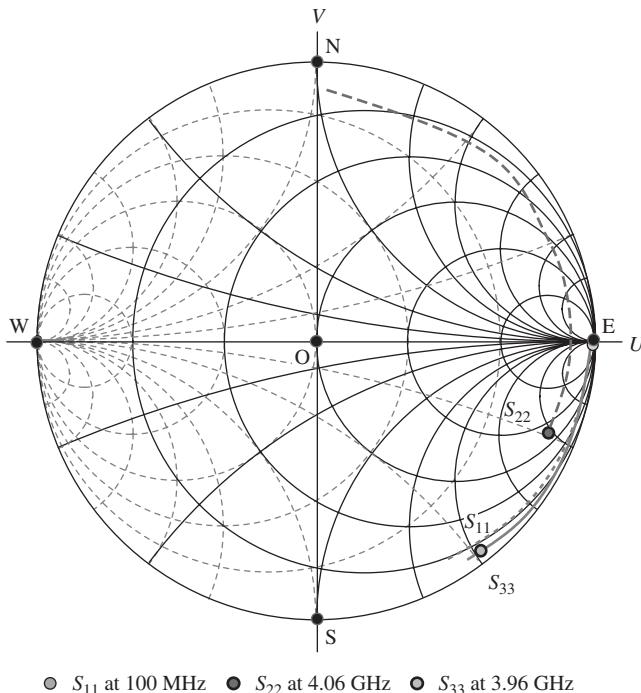
$$\text{Bias} = 0.685 \text{ V.}$$

- Current drain

$$I_{EE} = 2 \times 3.7 \text{ mA.}$$



**Figure 4.21.** Magnitudes (dB) of  $S_{11}$ ,  $S_{22}$ ,  $S_{33}$  when impedances at IF (port 1), RF (port 2), and LO (port 3) have not been matched.



**Figure 4.22.**  $S_{11}$ ,  $S_{22}$ ,  $S_{33}$  on the Smith Chart when impedances at IF (port 1), RF (port 2), and LO (port 3) have not been matched.

Figures 4.21 and 4.22 show the  $S$  parameters  $S_{11}$  (IF port),  $S_{22}$  (RF port), and  $S_{33}$  (LO port) of the Gilbert cell when the impedances at all three ports have not been matched.

It can be seen that the impedance matching state is very poor at all three ports. Figure 4.21 shows that the return losses  $S_{11}$ ,  $S_{22}$ , and  $S_{33}$  are all higher than  $-1$  dB in

the frequency range from 10 MHz to 4.3 GHz. Figure 4.22 shows that the traces of  $S_{11}$  and  $S_{33}$  at the IF and LO ports are located in regions 4 and 2 of the Smith Chart, where the reactances are capacitive and their values are mostly higher than  $50 \Omega$ . The impedances at the RF port are capacitive at high frequencies and are inductive at very low frequencies. The magnitudes of their reactances are much higher than  $50 \Omega$ . As a matter of fact, the impedances at IF (port 1), RF (port 2), and LO (port 3) can be directly read from Figure 4.22: they are partially listed in Table 4.1.

Within the frequency range as shown in Table 4.1,  $f = 3.696$  to  $7.392$  GHz, which is the frequency range UWB systems are concerned with, the variation of reactance approximately doubles from the low-frequency end to the high-frequency end at LO (port 3) and at RF (port 2). Impedance matching at these two ports is a difficult task because the frequency range to be covered is so wide. Impedance matching for wideband RF block is very much a challenge in the design of UWB systems.

#### 4.4.3 Impedance Matching for LO, RF, and IF Ports Ignoring the Bandwidth

Let us go ahead and design an IQ modulator for only band 2, group 1 in a UWB system, in which the frequencies and their bandwidth are

$$f_{\max} = 4224 \text{ MHz},$$

$$f_{\min} = 3696 \text{ MHz},$$

$$\text{BW} = f_{\max} - f_{\min} = 4224 - 3696 \text{ MHz} = 528 \text{ MHz}.$$

As the first try, in the implementation of input and output impedance matching network, ignore the bandwidth requirement and only take care of one frequency, the central

TABLE 4.1. Impedances at Three Ports of Gilbert Cell before Impedances Are Matched

Port	Frequency	Impedance
IF (port 1)	100 MHz	$1144 \Omega - j3171 \Omega$
	264 MHz	$186 \Omega - j1331.8 \Omega$
	528 MHz	$52.1 \Omega - j675.2 \Omega$
RF (port 2)	3.696 GHz	$78.6 \Omega - j249.1 \Omega$
	3.960 GHz	$66.9 \Omega - j234.2 \Omega$
	4.060 GHz	$61.9 \Omega - j225.8 \Omega$
	4.224 GHz	$60.8 \Omega - j222.6 \Omega$
	6.864 GHz	$26.3 \Omega - j138.6 \Omega$
	7.128 MHz	$25.2 \Omega - j134.0 \Omega$
	7.392 MHz	$24.0 \Omega - j129.2 \Omega$
LO (port 3)	3.696 GHz	$4.1 \Omega - j109.9 \Omega$
	3.960 GHz	$4.0 \Omega - j102.2 \Omega$
	4.224 MHz	$3.6 \Omega - j96.5 \Omega$
	6.864 MHz	$2.9 \Omega - j59.8 \Omega$
	7.128 MHz	$2.7 \Omega - j57.5 \Omega$
	7.392 MHz	$2.8 \Omega - j55.1 \Omega$

frequency of band 2 in the UWB system at each port, that is,

$$\begin{aligned}f_{\text{LO}} &= 3960 \text{ MHz}, \\f_{\text{RF}} &= 4060 \text{ MHz, and} \\f_{\text{IF}} &= 100 \text{ MHz.}\end{aligned}$$

This is easy work because the bandwidth is ignored. As a matter of fact, this is simply an extreme narrowband case. In terms of the design scheme discussed in Chapter 3, the impedance matching can be done in a short time since either the input or output impedance matching network can basically be implemented by only two parts, which are as follows:

- At IF (port 1):

$$C_p = 2 \text{ pF},$$

$$L_s = 1 \mu\text{F}.$$

- At RF (port 2):

$$L_{p1} = L_{p2} = 3.8 \text{ nH, } (r = 9.45\Omega, \text{ assumed } Q = 10),$$

$$C_s = 0.275 \text{ pF.}$$

- At LO (port 3):

$$C_p = 1.2 \text{ pF},$$

$$L_s = 5 \text{ nH } (r = 12.44\Omega, \text{ assumed } Q = 10).$$

Figure 4.23 shows the simulated result. In the input and output impedance matching network, the inductor is connected with a small resistor in series, which is calculated based on  $Q = 10$  and  $f = 3960 \text{ MHz}$ . The combination of the inductor and resistor approximates a practical spiral inductor on an RFIC chip. In Figure 4.23, it is circled by a dashed ellipse. In addition, the zero capacitor  $C_0$  does not count as a part of the impedance matching network.

Figures 4.24 and 4.25 show the parameters  $S_{11}, S_{22}, S_{33}$  after the impedances at the IF (port 1), RF (port 2), and LO (port 3) of the Gilbert cell have been matched.

Figure 4.24 shows that

$$BW_{\text{RF}} = 3.75 \text{ to } 4.33 \text{ GHz; } (S_{22} < -10 \text{ dB}),$$

$$BW_{\text{LO}} = 3.82 \text{ to } 4.15 \text{ GHz. } (S_{33} < -10 \text{ dB}),$$

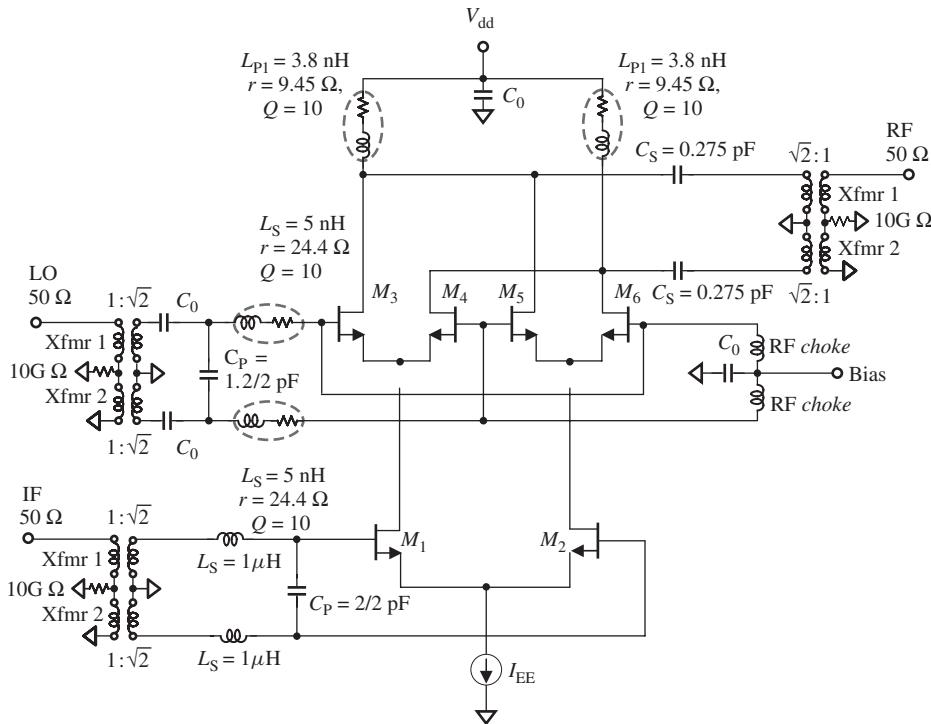
where

$BW_{\text{RF}}$  = the bandwidth at the RF port,

$BW_{\text{LO}}$  = the bandwidth at the LO port.

If the bandwidth is judged by the criterion of the return loss,

$$S_{11} = S_{22} = S_{33} = -10 \text{ dB.}$$



**Figure 4.23.** Input and output impedance matching network of the Gilbert cell for the simple try: only one frequency at each port is taken care, that is,  $f_{\text{LO}} = 3960 \text{ MHz}$ ,  $f_{\text{RF}} = 4060 \text{ MHz}$ ,  $f_{\text{IF}} = 100 \text{ MHz}$ .

In UWB system, the bandwidth of the IQ modulator designed for band 2, group 1, is expected to be

$$\text{BW}_{\text{RF}} = 3.596 \text{ to } 4.324 \text{ GHz},$$

$$\text{BW}_{\text{LO}} = 3.696 \text{ to } 4.224 \text{ GHz}.$$

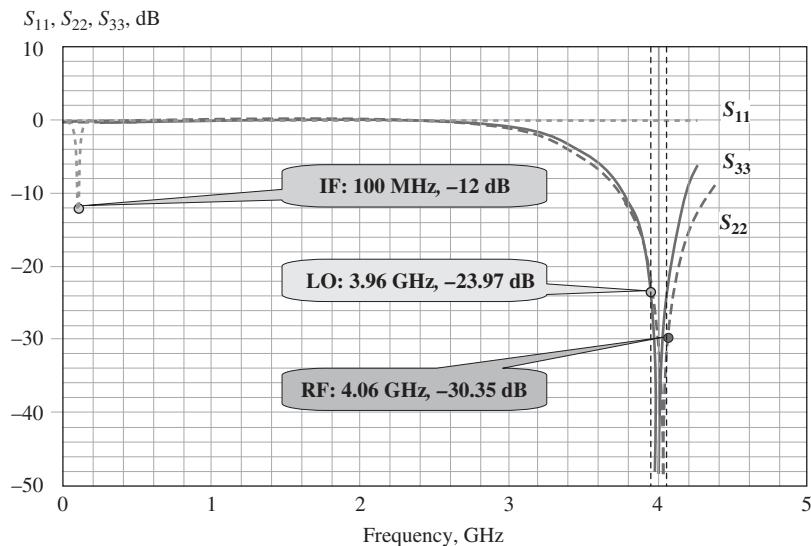
Obviously, the bandwidth is not wide enough to cover band 2, group 1, in a UWB system.

Figure 4.25 also displays a very narrow bandwidth response since most portions of the traces of  $S_{11}$ ,  $S_{22}$ , and  $S_{33}$  are outside the demarcation circle ( $S_{11} = S_{22} = S_{33} = -10 \text{ dB}$ ) and are scattered over a large area on the Smith Chart, although three impedances corresponding the specific frequencies,  $S_{11}$  at 100 MHz,  $S_{22}$  at 4.06 GHz, and  $S_{33}$  at 3.96 GHz, are inside the demarcation circle.

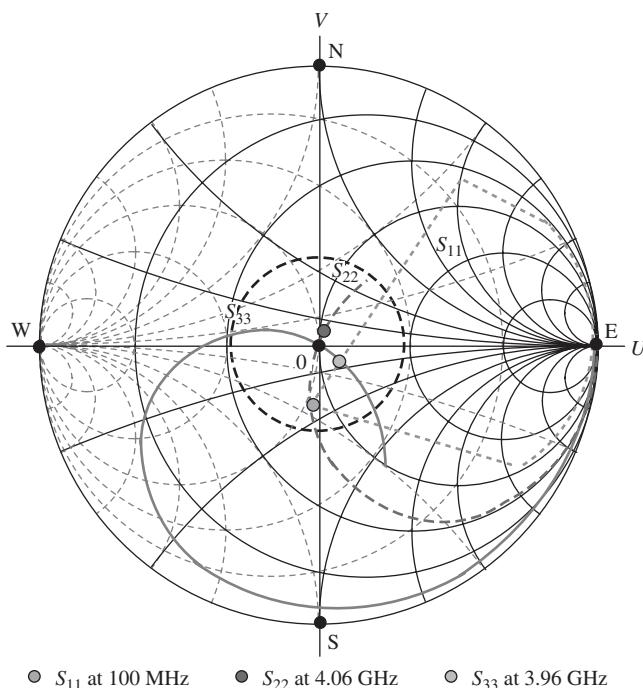
From now on we will ignore the IF port because of the following reasons:

- The input IF signal of the IQ modulator comes from the digital block. The design work for IF port must be discussed, negotiated, and cooperated with the digital circuit designer.
- The IF port is operated in the IF but not RF.

Instead, we will focus on the LO and RF ports because of the following reasons:



**Figure 4.24.** Magnitudes (dB) of  $S_{11}$ ,  $S_{22}$ ,  $S_{33}$  after the impedances at IF (port1), RF (port2), and LO (port3) have been matched.



**Figure 4.25.**  $S_{11}$ ,  $S_{22}$ ,  $S_{33}$  on the Smith Chart after the impedances at IF (port1), RF (port2), and LO (port3) have been matched.

- Usually, the LO port is assigned to the gate and the RF port is assigned to the drain of the transistor if the modulator is built by MOSFETs. Impedance matching with the gate and drain of the MOSFET represents two main techniques of impedance matching, which are widely demanded in the design of other RF blocks.
- The LO and RF ports operate in the RF range.

#### 4.4.4 Wide Bandwidth Required in a UWB (Ultra Wide Band) System

Portable radios or cellular phones are narrowband systems because their relative bandwidths are usually less than 15%. The emergence of the UWB system in recent years challenges the RF circuit designer to design RF circuits with a UWB response.

Table 4.2 lists the band group allocation in the UWB system. The frequency range in the UWB system is

$$f = 3.168 \text{ to } 10.560 \text{ GHz},$$

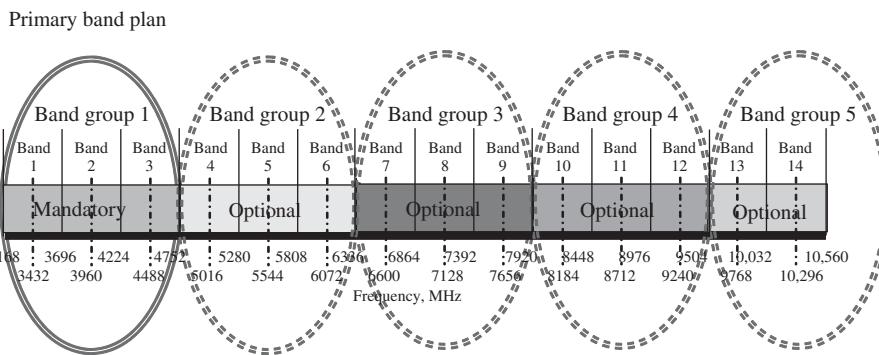
which is divided into 14 subbands, and the bandwidth of each subband is

$$\text{BW}_{\text{band}} = 528 \text{ MHz}.$$

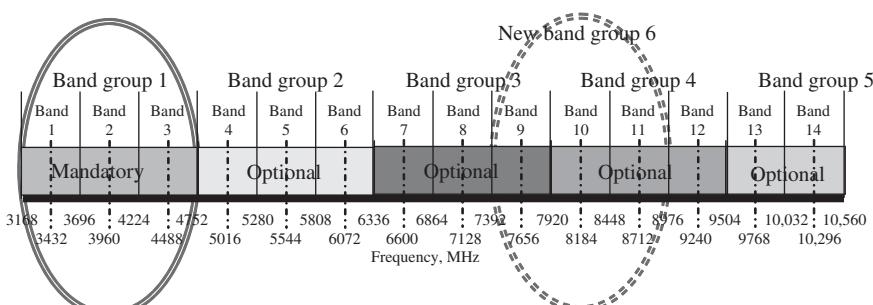
These 14 subbands are assigned as five groups. The first four groups each contains three subbands, while the fifth group contains only two subbands.

The first of these groups in the frequency range is a mandatory group to be covered in any UWB system, while the other groups are optional. The second group is always ignored because it is occupied by the WLAN (wireless local area network) communication system. Owing to restrictions made by government for military or other special agencies in different countries, the other subbands contained in third, fourth, and fifth groups are selected for use by UWB companies or enterprises in different ways. A new “sixth

TABLE 4.2. Frequency Plan of a UWB System



New world-wide compliant band plan



group,” containing subbands 9, 10, and 11, is found to have fewer restrictions in the world wireless market. Currently, most UWB companies and enterprises are focusing on the first and sixth groups.

As mentioned in Chapter 3, a system or a block is categorized as narrowband if its relative bandwidth is lower than 15%, whereas an RF block is categorized as a wideband system or block if its relative bandwidth is higher than 15%. The 15% relative bandwidth is a rough but reasonable value as a demarcation between narrow and wideband cases, though it is not a strict criterion. Table 4.3 lists the frequencies for each frequency band and calculates all of their absolute bandwidths,  $BW = \Delta f$ , and relative bandwidths,  $\Delta f/f_C$ . It can be seen that wideband design is required if the system is going to cover only group 1, in which the relative bandwidth is 40%. In order to lower the cost and simplify the circuit design, the relative bandwidth of a system should be expanded as much as possible. For example, it is desirable to combine group 3 and group 6 together, which results in a relative bandwidth of  $\Delta f/f_C = 34.48\%$  if the system is to cover the bandwidth containing bands 7–11 or to cover frequencies from 6336 to 8976 MHz.

The key issue in a wideband system or circuit design is, of course, the wideband impedance matching.

#### 4.4.5 Basic Idea to Expand the Bandwidth

As pointed out in Section 4.1, in order to change the RF block’s impedance response from narrowband to wideband, the following approaches are pursued in the implementation process of the impedance matching network:

1. The distance between the trace of return loss,  $S_{11}$  or  $S_{22}$ , and the center of the Smith Chart should be reduced to be as short as possible.
2. The area the impedance trace or the return loss  $S_{11}$  or  $S_{22}$  occupies on the Smith Chart must be squeezed or shrunk to as small as possible.

TABLE 4.3. Band Group Allocation UWB System

Band Group Allocation UWB System							
Group	Band	$f_L$ , MHz	$f_C$ , MHz	$f_H$ , MHz	$\Delta f$ , MHz	$\Delta f/f_C$	$\Delta f/f_C$
1	1	3168	3432	3696	528	15.38%	↑
	2	3696	3960	4224	528	13.33%	40.00%
	3	4224	4488	4752	528	11.76%	↓
2	4	4752	5016	5280	528	10.53%	
2	5	5280	5544	5808	528	9.52%	
2	6	5808	6072	6336	528	8.70%	
6	3	6336	6600	6864	528	8.00%	↑
	3	6864	7128	7392	528	7.41%	
	3	7392	7656	7920	528	6.90%	34.48%
6	4	7920	8184	8448	528	6.45%	
6	4	8448	8712	8976	528	6.06%	↓
	4	8976	9240	9504	528	5.71%	
5	13	9504	9768	10032	528	5.41%	
5	14	10032	10296	10560	528	5.13%	

A shorter distance implies that the RF block has a lower return loss and, therefore, correspondingly a wider bandwidth. If the trace is squeezed to the center of the Smith Chart, the RF block would definitely have wideband performance.

There are many ways to shrink the return loss,  $S_{11}$  or  $S_{22}$ , on the Smith Chart. The LC combination, either in series or in parallel, is quite an effective scheme to increase the bandwidth. The single part per arm or per branch scheme can effectively shrink the trace's covered area also if the original trace is located in certain positions.

The part count should be kept as low as possible. Above all, the number of inductors employed should be as small as possible because the inductor is much more expensive and has a lower  $Q$  value than the capacitor. However, in order to expand the impedance matching network from a narrow to a wide bandwidth, 1 to 5 inductors are usually required.

Keeping these ideas in mind, it has been proven that impedance matching for a UWB system can be well carried out through the reasonable implementation of the impedance matching networks. In the following subsections we will present two examples of partial works in the "T" or "Q" modulator design for a UWB system. Instead of the entire design works for the IQ modulator, we will show only the portion of how to expand the impedance response from narrowband to wideband at the LO and RF ports. In particular, we will show the evolution of the bandwidth as the impedance matching parts are inserted into the impedance matching network step by step.

#### 4.4.6 Example 1: Impedance Matching in IQ Modulator Design for Group 1 in a UWB System

Figure 4.26 shows an I or a Q modulator for a UWB system with impedance matching networks. In this example, the operating frequency range is from 3168 to 4752 MHz for group 1 in the UWB system so that the relative bandwidth is 40%. The modulator is supposed to be implemented on an RFIC chip.

Figures 4.27 and 4.28 plot the input impedance matching network at the LO port and the output impedance matching network at the RF port, respectively. The input impedance matching network at the LO port shown in Figure 4.27 consists of two arms and two branches. An inductor is always connected with a small resistor in series and is circled by a dashed-line ellipse. The small resistor represents the attenuation of the inductor, and its value is calculated on the basis of the assumed quality factor  $Q = 10$  and the operating frequency  $f = 3960$  MHz. There are two resistance values appearing in the  $r$  expression of  $L_{S1}$  and  $L_{P1}$ . The first value is the calculated resistance based on  $Q = 10$  and  $f = 3960$  MHz and the second value is the additional resistance for more bandwidth expansion. One of the two arms consists of one inductor and the other arm consists of an LC combination in series. Similarly, one of the two branches consists of one capacitor and another branch consists of an LC combination in parallel. The symbols "2×", appearing in  $L_{P1}$ , and "/2", appearing in  $C_P$  and  $C_{P1}$ , indicate that there are two identical parts connected together in series as a branch bridging the two differential nodes.

The output impedance matching network at the RF port is shown in Figure 4.28. It consists of two arms and three branches. Again, an inductor is always connected with a small resistor in series and is circled by a dashed-line ellipse. The small resistor represents the attenuation of the inductor and its value is calculated from the assumed quality factor  $Q = 10$  and the operating frequency  $f = 3960$  MHz. There is only one part contained in an arm or a branch. The symbol "/2" appearing in  $C_P$  and  $C_{P1}$  indicates that there are two identical parts connected together in series as a branch bridging two differential nodes. The DC power supply  $V_{dd}$  provides the current to the devices through two  $L_P$  inductors.

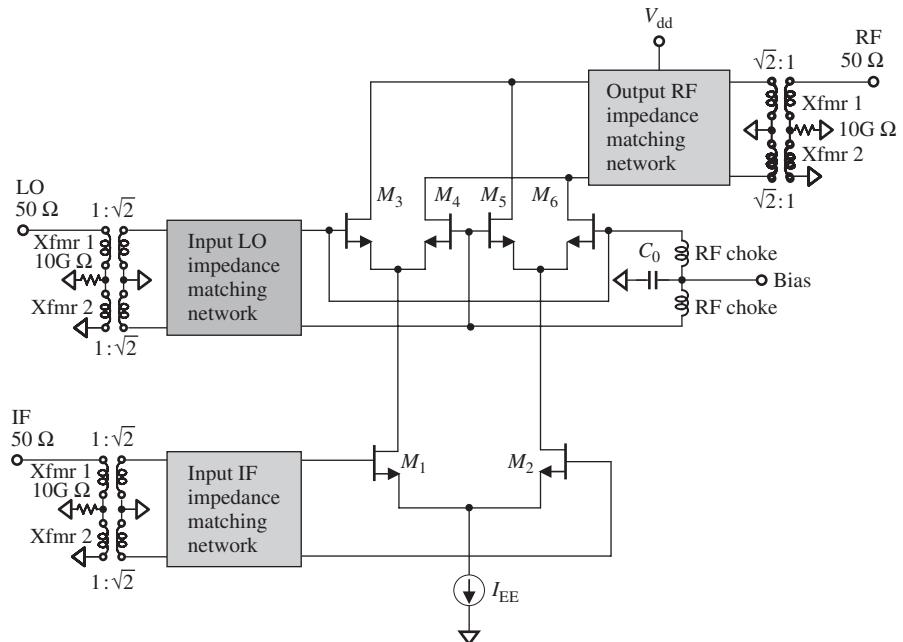


Figure 4.26. I or Q modulator for a UWB system with impedance matching networks.

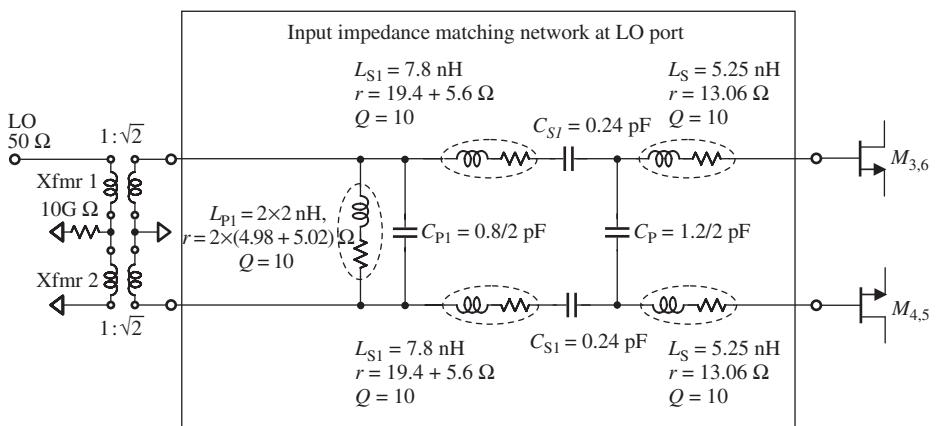
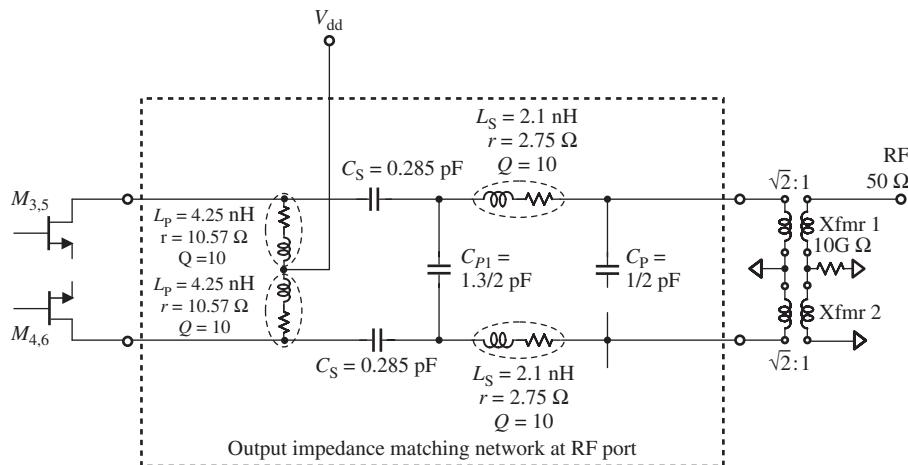


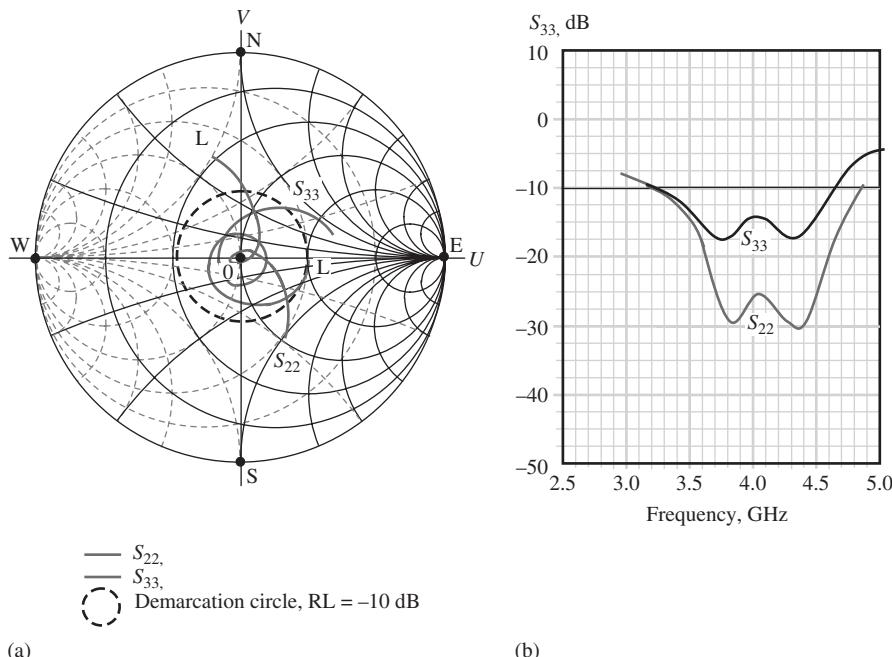
Figure 4.27. Input impedance matching network at LO ports for group 1 in a UWB system,  $f = 3168$  to  $4752$  MHz.

The input impedance matching network at the LO port, as shown in Figure 4.27, and the output impedance matching network at the RF port, as shown in Figure 4.28, successfully expand the bandwidth from very narrow to very wide. Figure 4.29 shows the expected value of the bandwidth being reached for an IQ modulator working in group 1 of a UWB system.

From Figure 4.29(a), it can be seen that most portions of the return loss trace, either  $S_{22}$  or  $S_{33}$ , are located within the demarcation circle. This indicates that the impedances at both port 2 and port 3 are matched to  $50 \Omega$  in the wideband case. The bandwidth can be read from Figure 4.29(b) accordingly: that is,



**Figure 4.28.** Output impedance matching network at RF port for group 1 in a UWB system,  $f = 3168$  to  $4752$  MHz.



**Figure 4.29.** At LO and RF ports, the relative bandwidth is expanded to the expected value for a group 1 UWB system through impedance matching. (a) Return loss on the Smith Chart after impedance matching is done at port RF ( $S_{22}$ ) and LO ( $S_{33}$ ). (b)  $S_{22}$  and  $S_{33}$  versus  $f$  after impedance matching is done at port RF ( $S_{22}$ ) and LO ( $S_{33}$ ).

At the RF port,

$$\text{BW}_{\text{RF}} = \Delta f = 4860 - 3100 \text{ MHz} = 1760 \text{ MHz},$$

$$(\text{BW}_{\text{RF}}) \text{ relative} = \Delta f / f_c = 1760 / [(4860 + 3100)/2] = 44.22\%.$$

At the LO port,

$$\text{BW}_{\text{RF}} = \Delta f = 4750 - 3150 \text{ MHz} = 1600 \text{ MHz},$$

$$(\text{BW}_{\text{RF}}) \text{ relative} = \Delta f / f_C = 1600 / [(4750 + 3150) / 2] = 41.56\%.$$

On the other hand, bandwidth required in the IQ modulator design for group 1 of the UWB system is

$$\text{BW}_{\text{RF or LO}} = \Delta f = 4752 - 3168 \text{ MHz} = 1584 \text{ MHz},$$

$$(\text{BW}_{\text{RF or LO}}) \text{ relative} = \Delta f / f_C = 1584 / [(4752 + 3168) / 2] = 40.00\%.$$

Obviously, the design of the input impedance matching network at the LO port as shown in Figure 4.27 and the output impedance matching network at the RF port as shown in Figure 4.28 are successful!

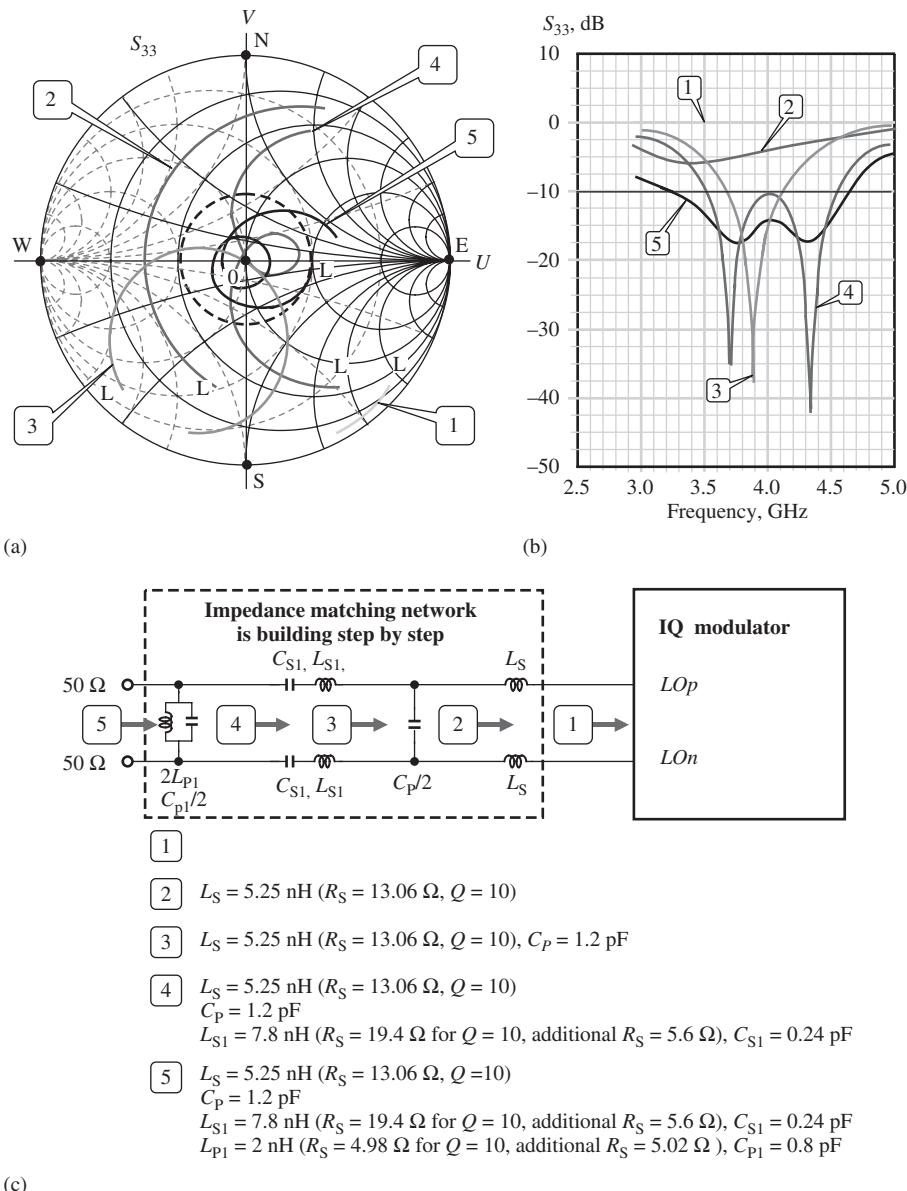
It is very desirable and of interest to study how to approach the goal of wideband impedance matching. In the following paragraphs we will demonstrate the evolution of bandwidth at the LO and RF ports as the impedance matching arms/branches are inserted one by one.

Let us demonstrate the evolution of bandwidth at the LO port first.

Figure 4.30(a) shows each impedance trace on the Smith Chart after an arm or a branch of parts is inserted into the impedance matching network. A letter “L” attached to each trace denotes the low-frequency end of the trace. Correspondingly, Figure 4.30(b) shows each curve of  $S_{33}$  versus frequency after an arm or a branch of parts is inserted into the impedance matching network. In Figure 4.30(c), the  $\text{LO}_p$  and  $\text{LO}_n$  are the LO differential inputs of the I or Q modulator and are assigned as port 3. These are, in fact, the gates of MOSFET devices. At the outputs of  $\text{LO}_p$  and  $\text{LO}_n$ , the round square frame marked with “1” corresponds to trace 1 in both the Smith Chart and in Cartesian coordinates with the magnitude of  $S_{33}$  (dB). From this point, an impedance matching network is constructed by successive insertion of parts. Each time an arm or a branch is added to the impedance matching network, the trace number of the return loss in Cartesian coordinates or the trace number of the impedance trace in the Smith Chart increases by 1. Figure 4.30 shows the evolution of impedance as the traces change from trace 1 all the way to trace 5 as individual arms or branches are inserted into the impedance matching network. At the bottom of Figure 4.30(c), the parts for all the arms or branches are listed for each corresponding trace.

It can be found that the return loss  $S_{33}$  of trace 1 is very high. The magnitude of  $S_{33}$  is almost equal to 0 dB in the entire frequency bandwidth of interest, from 3168 to 4752 MHz. Therefore, the first step to match the LO port is to insert an inductor in series,  $L_S$ , so that trace 1 at  $\text{LO}_p$  and  $\text{LO}_n$ , which is located in the high-impedance area, can be changed to trace 2, which is located in the low-impedance area. It should be noted that, as shown in Figure 4.31, all the impedances on trace 1 more or less move clockwise around their own resistance circles, though their resistances are increased because of the equivalent resistance of the inductor in series. On the other hand, the variation of impedance at the high-frequency end is much greater than the variation of impedance at the low-frequency end.

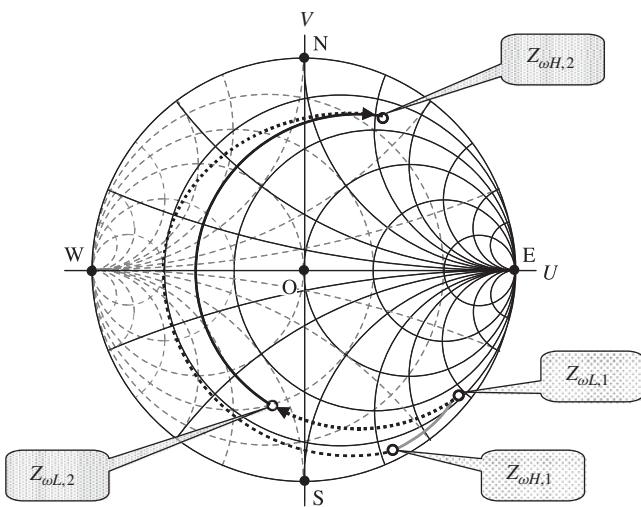
Trace 2 on the Smith Chart is closer to the reference impedance  $50 \Omega$ , the center of Smith Chart, than trace 1. Therefore, the magnitude of  $S_{33}$  is improved from zero to a couple of decibels as shown in the Cartesian coordinates of Figure 4.30. However, a couple



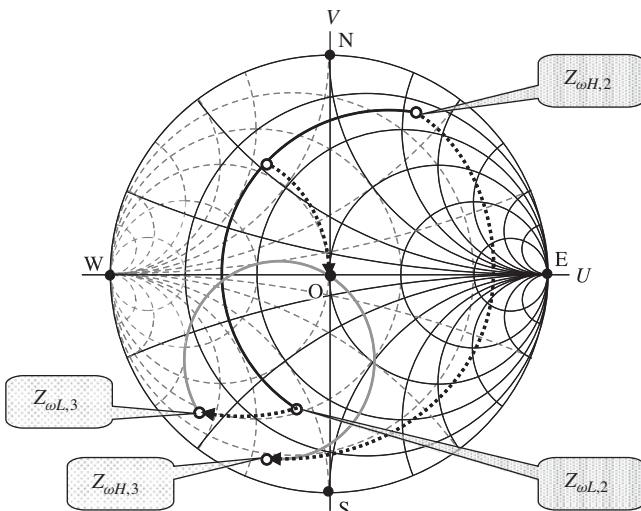
**Figure 4.30.** Evolution of bandwidth at the LO port as the parts are inserted into impedance matching network ( $f = 3168$  to  $4752$  MHz). (a)  $S_{33}$  relocated on the Smith Chart after one arm or branch is inserted into the impedance matching network. (b)  $S_{33}$  versus  $f$  curves for each location of  $S_{33}$  on the Smith chart in the adjacent left-hand side. (c) Impedance matching in the LO port.

negative decibel of  $S_{33}$  indicates too much return power and not enough to reach a good impedance matching performance, though it improved from trace 1. Typically, the return loss must be greater than  $10$  dB (or, say, the return loss must be lower than  $-10$  dB).

Consequently, the second step to match the LO port is to insert a capacitor in parallel,  $C_P$ , so that trace 2 is bent to trace 3 as shown in Figure 4.32. It is found that all the impedances on the trace 2 basically move clockwise along their own conductance circle.



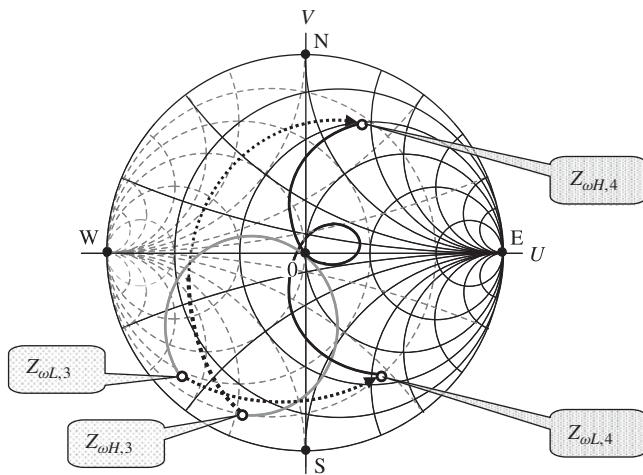
**Figure 4.31.** Variation of impedance from trace 1 to trace 2 at the LO port as the part  $L_s$  is inserted into impedance matching network ( $f = 3168$  to  $4752$  MHz).



**Figure 4.32.** Variation of impedance from trace 2 to trace 3 at the LO port as the part  $C_p$  is inserted into impedance matching network ( $f_L = 3168$  to  $4752$  MHz).

On the other hand, the variation of impedance at the high-frequency end is much greater than the variation of impedance at the low-frequency end. The impedance corresponding to the IF crosses the reference impedance, the center of the Smith Chart. The return loss  $S_{33}$  of trace 3 as shown in Cartesian coordinates in Figure 4.30 is greatly improved from that of trace 2. Especially at the IF, the magnitude of  $S_{33}$  reaches approximately  $-35$  dB. For narrowband RF block designs, the bandwidth in which the  $S_{33} < -10$  dB is easily satisfied. However, trace 3 is still not wide enough for wideband performance.

In order to further improve the bandwidth, a combination of LC in series is inserted to the impedance matching network in series. This moves trace 3 to trace 4 as shown in Figure 4.33. It should be noted that the high-frequency end of trace 3 moves clockwise while its low-frequency end moves counterclockwise. This special feature “twists” or “rolls” the trace and therefore widens the bandwidth since the twisting or rolling action shrinks the area covered by the trace on Smith Chart.



**Figure 4.33.** Variation of impedance from trace 3 to trace 4 at the LO port as the part  $L_{S1}$ ,  $C_{S1}$  is inserted into impedance matching network ( $f = 3168$  to  $4752$  MHz).

The rolling or twisting of the trace can be generated by a combination of LC in series. It could also be formed by a single inductor  $L$  in series, which will be shown later. The trace must be located on Smith Chart in a manner that

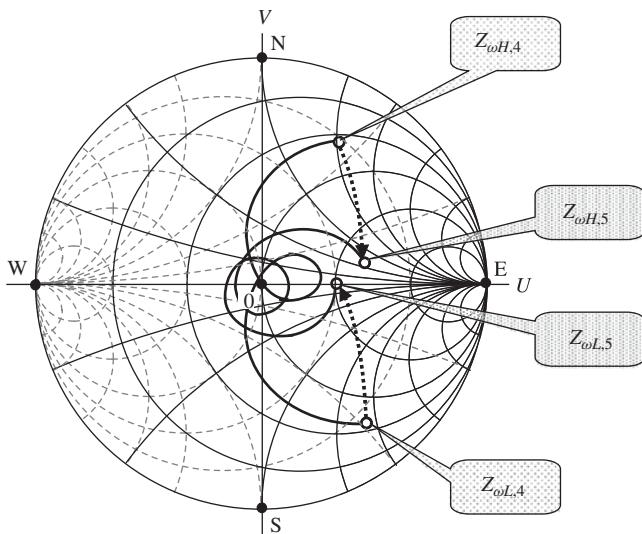
- the low- and high-frequency ends of the trace are located approximately on the same resistance circle, and
- the high-frequency end is “more capacitive” or “less inductive” than the low-frequency end.

The difference is that, in the case when a combination of LC in series is inserted, the entire trace does not move clockwise in such a simple way as is the case when a simple inductor  $L$  is inserted. It might instead be moved to a location where its resistance circles would be quite different from the original ones. It is therefore possible to move its intermediate portion to the area around the reference impedance,  $50 \Omega$ , in order to accomplish the widening of the bandwidth. In cases when a single inductor  $L$  is inserted, the entire trace would be moved clockwise to a new location, where its resistance circles and return loss would more or less remain unchanged and, consequently, its variation of bandwidth would be small.

Rolling or twisting of the trace by the insertion of a combination LC in series is undoubtedly a powerful scheme to increase the bandwidth.

Very often, the bandwidth of trace 4 is still not enough to a UWB system. Thus, we must widen the bandwidth even more. At this point, we do not wish to move the intermediate portion of the trace out of the area around the reference impedance,  $50 \Omega$ . Instead, it is preferable to squeeze both trace portions corresponding to the low and high ends of the frequency toward the reference impedance,  $50 \Omega$ , the center of Smith Chart. Figure 4.34 shows this step, in which a combination of LC in parallel is inserted into the impedance matching network in parallel. The high-frequency end of trace 4 is moved clockwise along the conductance circle while the low-frequency end of trace 4 is moved counterclockwise along the conductance circle. Trace 5 is now squeezed around the reference impedance more, and the bandwidth is increased so that it satisfies the desired goal.

Figure 4.35 shows the evolution of the bandwidth at the RF port as the parts are inserted into the impedance matching network.



**Figure 4.34.** Variation of impedance from trace 4 to trace 5 at the LO port as the part  $L_{P1}, C_{P1}$  is inserted into impedance matching network ( $f = 3168$  to  $4752$  MHz).

It can be seen that the squeezing of the trace by inserting of a combination LC in parallel into the impedance matching network is also an undoubtedly powerful scheme to increase the bandwidth even more. By now, the bandwidth is wide enough to satisfy application to a UWB system. As a matter of fact, Figures 4.30–4.34 show a typical wideband impedance matching process for the LO port of a Gilbert cell when MOSFET devices are applied, no matter whether the Gilbert cell is being constructed for a modulator or for a mixer.

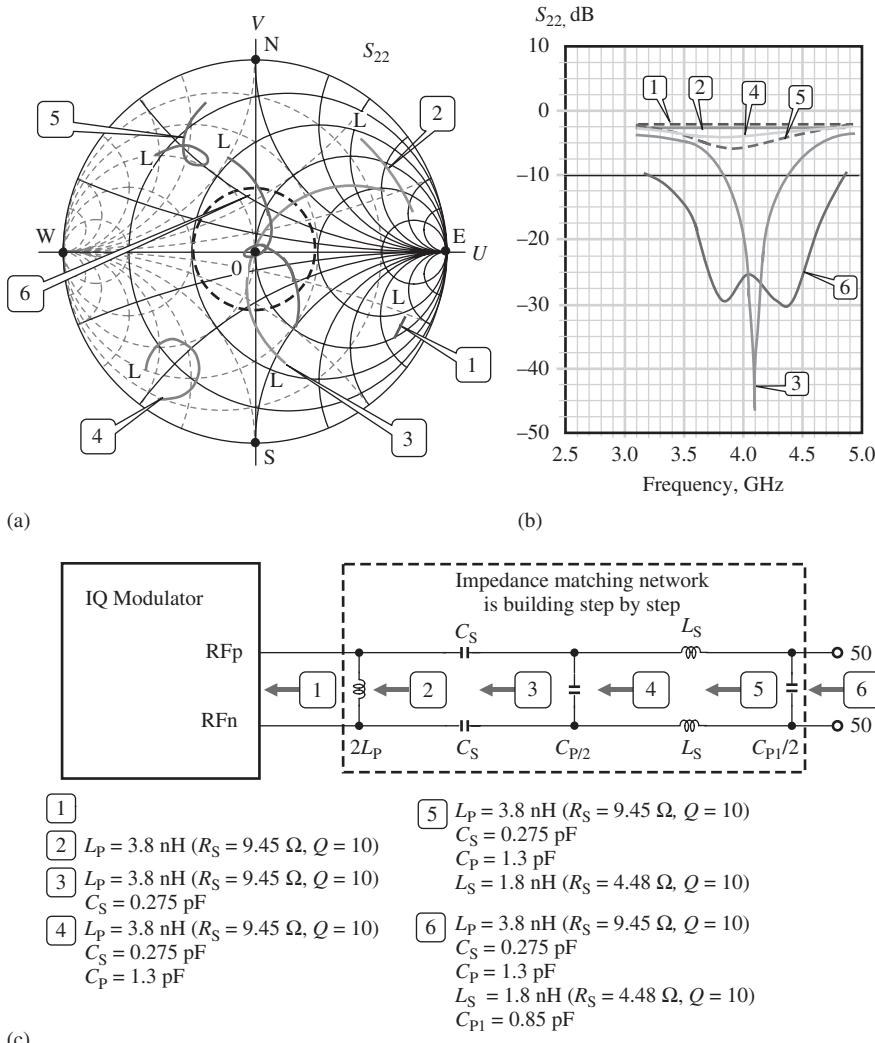
Now, let us demonstrate the evolution of bandwidth at RF port.

Figure 4.35(a) shows each impedance trace on the Smith Chart after an arm or a branch of parts is inserted into the impedance matching network. A letter “L,” attached to each trace, denotes the low-frequency end of the trace. Correspondingly, Figure 4.35(b) shows each curve of  $S_{22}$  versus frequency after an arm or a branch of parts is inserted into the impedance matching network. In Figure 4.35(c), the  $RF_p$  and  $RF_n$  are the RF differential outputs of the I or Q modulator and are assigned as port 2. These are, in fact, the drains of MOSFET devices. At the outputs of  $RF_p$  and  $RF_n$ , the round square frame marked with 1 corresponds to trace 1 in both the Smith chart and the Cartesian coordinates with the magnitude of  $S_{22}$  (dB). From this point, an impedance matching network is constructed by successive insertion of parts. Each time an arm or a branch is inserted into the impedance matching network, the trace number of the return loss in the Cartesian coordinates or the trace number of the impedance on the Smith Chart trace increases by 1. Figure 4.35 shows the evolution of impedance as the traces change from trace 1 to trace 6 as the individual arms or branches are inserted into the impedance matching network. At the bottom of Figure 4.35, the parts for all the arms or branches are listed for each corresponding trace.

It can be found that the return loss  $S_{22}$  of trace 1 is very high. The magnitude of  $S_{22}$  is almost equal to 0 dB in the entire frequency bandwidth of interest. The first step to match the RF port is to insert an inductor  $L_p$  in parallel, as shown in Figure 4.36.

The next step is to add a capacitor  $C_s$  in series, so that trace 2 is moved to trace 3 as shown in Figure 4.37.

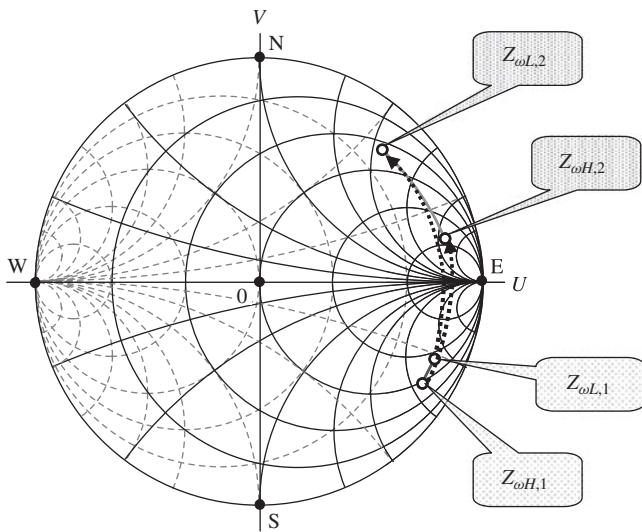
It is found that all the impedances on trace 2 are basically moved counterclockwise around their own resistance circles. On the other hand, the variation of impedance at the



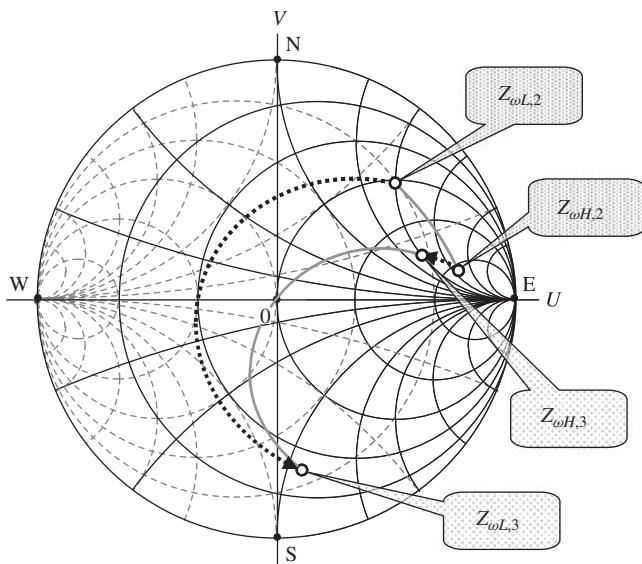
**Figure 4.35.** Evolution of bandwidth at the RF port as the parts are added to the impedance matching network. (a)  $S_{22}$  relocated on the Smith Chart after one arm or branch is inserted into the impedance matching network. (b)  $S_{22}$  and  $S_{33}$  versus  $f$  after impedance matching is done at port RF ( $S_{22}$ ) and LO ( $S_{33}$ ). (c) Impedance matching in RF port.

Note: The parts,  $C_p$ ,  $L_s$ , and  $C_{p1}$ , are added for the expansion of bandwidth. The resulting BW from simulation: 31004860 MHz. The desired BW for Group 1, Band 2 in a UWB system: 35964324 MHz. The desired BW for entire Group 1 in a UWB system: 31684752 MHz.

low-frequency end is much greater than the variation of impedance at the high-frequency end. The impedance corresponding to an IF crosses the reference impedance, the center of the Smith Chart. The return loss  $S_{22}$  of trace 3, as shown in Cartesian coordinates in Figure 4.35, is greatly improved from that of trace 2. Especially at the central frequency, the magnitude of  $S_{22}$  reaches approximately -46 dB. For narrowband RF block designs, the bandwidth in which the  $S_{22} < -10$  dB is easily satisfied. However, trace 3 is still not wide enough for wideband performance.



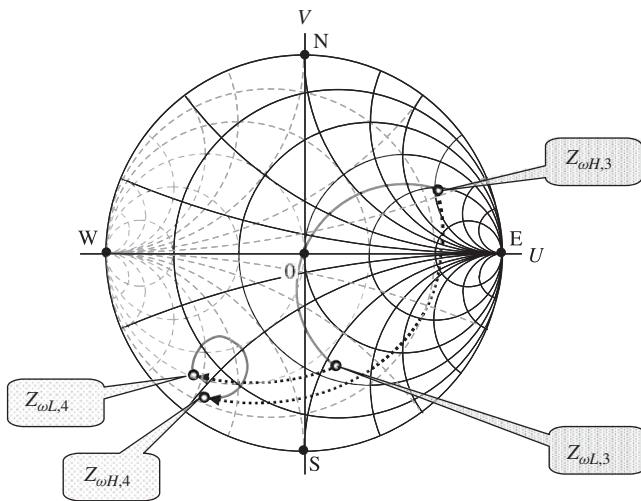
**Figure 4.36.** Variation of impedance from trace 1 to trace 2 at the RF port as the part  $L_p$  is inserted into impedance matching network ( $f = 31684752$  MHz).



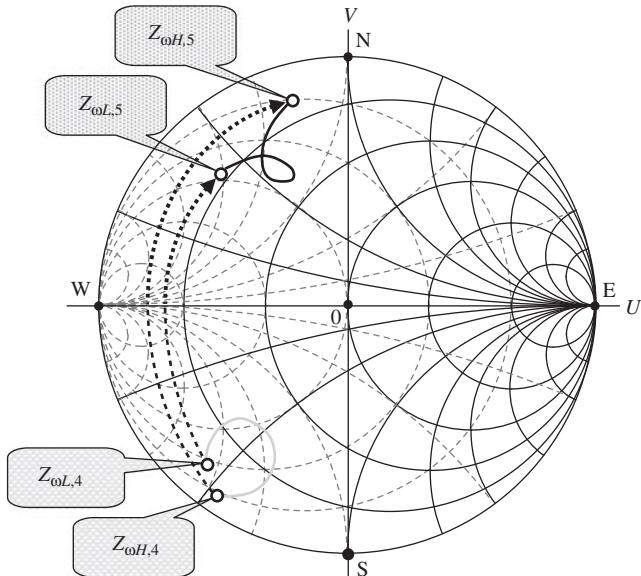
**Figure 4.37.** Variation of impedance from trace 2 to trace 3 at the RF port as the part  $C_s$  is inserted into impedance matching network ( $f = 31684752$  MHz).

In order to further improve the bandwidth, we might insert an LC combination in parallel to the impedance matching network to squeeze the trace as shown in Figure 4.34.

Instead, another alternative method for the third step is shown in Figure 4.38, in which a single capacitor  $C_p$  is inserted in parallel. This makes trace 3 move away from the reference impedance,  $50 \Omega$ , but it bends trace 3 to trace 4, which covers a much smaller area. Both the low- and high-frequency ends are moved clockwise along their respective susceptance circles. The variation of impedance at the low-frequency end is much greater than that at the high-frequency end. It should be noted that, in trace 4, the impedance at the high-frequency end is more “capacitive” than the impedance at the



**Figure 4.38.** Variation of impedance from trace 3 to trace 4 at the RF port as the part  $C_p$  is inserted into impedance matching network ( $f = 31684752$  MHz).

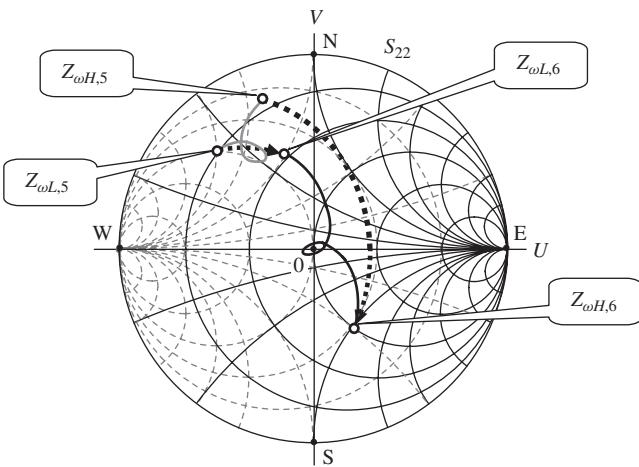


**Figure 4.39.** Variation of impedance from trace 4 to trace 5 at RF port as the part  $L_s$  is inserted into impedance matching network ( $f = 31684752$  MHz).

low-frequency end. As discussed above, this trace 4 could potentially be rolled up or twisted so as to further widen the bandwidth.

Figure 4.39 shows the movement from trace 4 to trace 5 after a single inductor  $L_s$  is inserted into the impedance matching network in series as the fourth step. As expected, trace 4 is rolled or twisted into trace 5. However, as shown in the Cartesian coordinates of Figure 4.35, the magnitudes of  $S_{22}$  of the traces 4 and 5 are almost in the same order and are unsatisfactory.

The fifth step is to insert a single capacitor into the impedance matching network in parallel, so that trace 5 is pulled to trace 6, which is located around the area of the reference impedance, as shown in Figure 4.40. The bandwidth can be read from the Cartesian coordinates of the  $S_{22}$  magnitude in Figure 4.35.

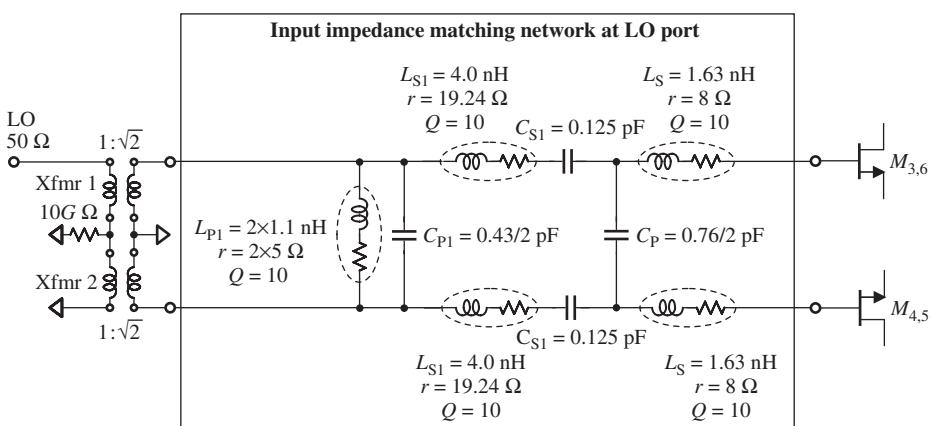


**Figure 4.40.** Variation of impedance from trace 5 to trace 6 at RF port as the part  $C_{P1}$  is inserted into impedance matching network, ( $f = 31684752$  MHz).

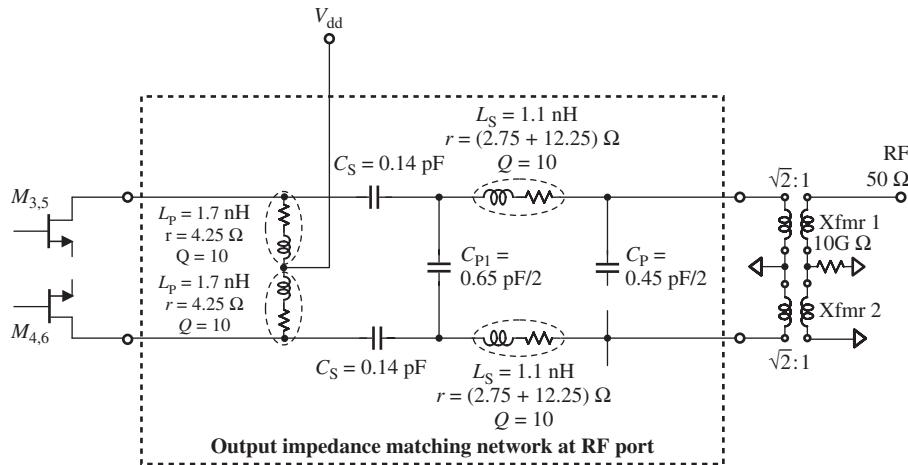
#### 4.4.7 Example 2: Impedance Matching in IQ Modulator Design for Group 3+ Group 6 in a UWB System

In Example 2, the topology of the I or Q modulator is the same as that in Example 1, as shown in Figure 4.26. The difference is that it is designed for band groups 3 and 6 together. The operating frequency range is from 6336 to 8976 MHz, so that the relative bandwidth is 34.48%. Again, the modulator is supposed to be implemented on an RFIC chip.

Figures 4.41 and 4.42 plot the input impedance matching network at the LO port and the output impedance matching network at the RF port, respectively. The input impedance matching network at the LO port shown in Figure 4.41 consists of two arms and two branches. An inductor is always connected with a small resistor in series and is circled by a dashed-line ellipse. The small resistor represents the attenuation of the inductor and its value is calculated on the basis of the assumed quality factor  $Q = 10$  and the operating frequency  $f = 7656$  MHz. There are two resistance values appearing in the  $r$  expression of  $L_{S1}$  and  $L_{P1}$ . The first value is the calculated resistance based on  $Q = 10$  and  $f = 7656$  MHz, and the second value is the additional resistance for more



**Figure 4.41.** Input impedance matching network at LO ports for group 3+ group 6 in a UWB system,  $f = 6336$  to  $8976$  MHz.



**Figure 4.42.** Output impedance matching network at RF port for group 3 + group 6 in a UWB system,  $f = 6336$  to  $8976$  MHz.

bandwidth expansion. One of the two arms consists of one inductor and the other arm consists of an LC combination in series. Similarly, one of the two branches consists of one capacitor and another branch consists of an LC combination in parallel. The symbols “ $2\times$ ”, appearing in  $L_{P1}$ , and “ $/2$ ,” appearing in  $C_p$  and  $C_{P1}$ , indicate that there are two identical parts connected together in series as a branch bridging two differential nodes.

The output impedance matching network at the RF port is shown in Figure 4.42. It consists of two arms and three branches. Again, an inductor is always connected with a small resistor in series and is circled by a dashed-line ellipse. The small resistor represents the attenuation of the inductor and its value is calculated from the assumed quality factor  $Q = 10$  and the operating frequency  $f = 7656$  MHz. There is only one part contained in an arm or a branch. The symbol “ $/2$ ” appearing in  $C_p$  and  $C_{P1}$  indicates that there are two identical parts connected together in series as a branch bridging two differential nodes. The DC power supply  $V_{dd}$  provides the current to the devices through two  $L_p$  inductors.

The input impedance matching network at the LO port as shown in Figure 4.41, and the output impedance matching network at the RF port, as shown in Figure 4.42, successfully expand the bandwidth from a very narrow to a very wide band. Figure 4.43 shows the expected value of the bandwidth being reached for an IQ modulator working in bandwidth groups 3 and 6 in a UWB system.

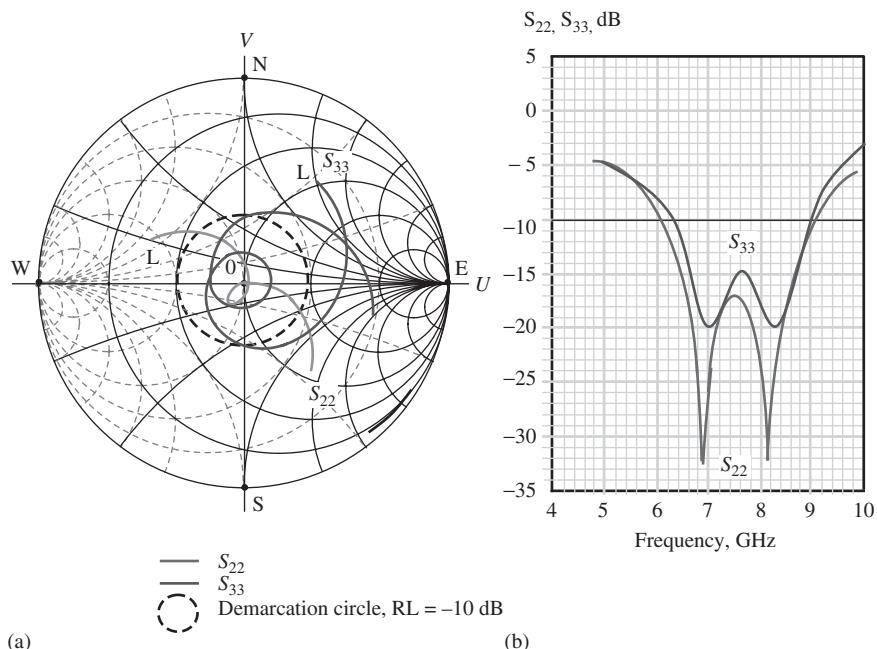
From Figure 4.43(a) it can be seen that most portions of the return loss trace, either  $S_{22}$  or  $S_{33}$ , are located within the demarcation circle. This indicates that the impedances at both port 2 and port 3 are matched to  $50 \Omega$  in the wideband case. The bandwidth can be read from Figure 4.43(b) accordingly, that is,

at the RF port,

$$\begin{aligned} \text{BW}_{RF} &= \Delta f = 9080 - 6100 \text{ MHz} = 2980 \text{ MHz}, \\ (\text{BW}_{RF})_{\text{relative}} &= \Delta f / f_C = 2980 / [(9080 + 6100)/2] = 39.26\%. \end{aligned}$$

At the LO port,

$$\begin{aligned} \text{BW}_{RF} &= \Delta f = 9000 - 6330 \text{ MHz} = 2670 \text{ MHz}, \\ (\text{BW}_{RF})_{\text{relative}} &= \Delta f / f_C = 2670 / [(9000 + 6330)/2] = 34.83\%. \end{aligned}$$



**Figure 4.43.** At the LO and RF ports, the relative bandwidth is expanded to the expected value for Group 3 + Group 6 UWB system through impedance matching. (a) Return loss on the Smith Chart after impedance matching is done at port RF ( $S_{22}$ ) and LO ( $S_{33}$ ). (b)  $S_{22}$  and  $S_{33}$  versus  $f$  after impedance matching is done at port RF ( $S_{22}$ ) and LO ( $S_{33}$ ).

On the other hand, bandwidth required in the IQ modulator design for Group 3 + Group 6 of the UWB system is

$$\text{BW}_{RF \text{ or } LO} = \Delta f = 8976 - 6336 \text{ MHz} = 2640 \text{ MHz},$$

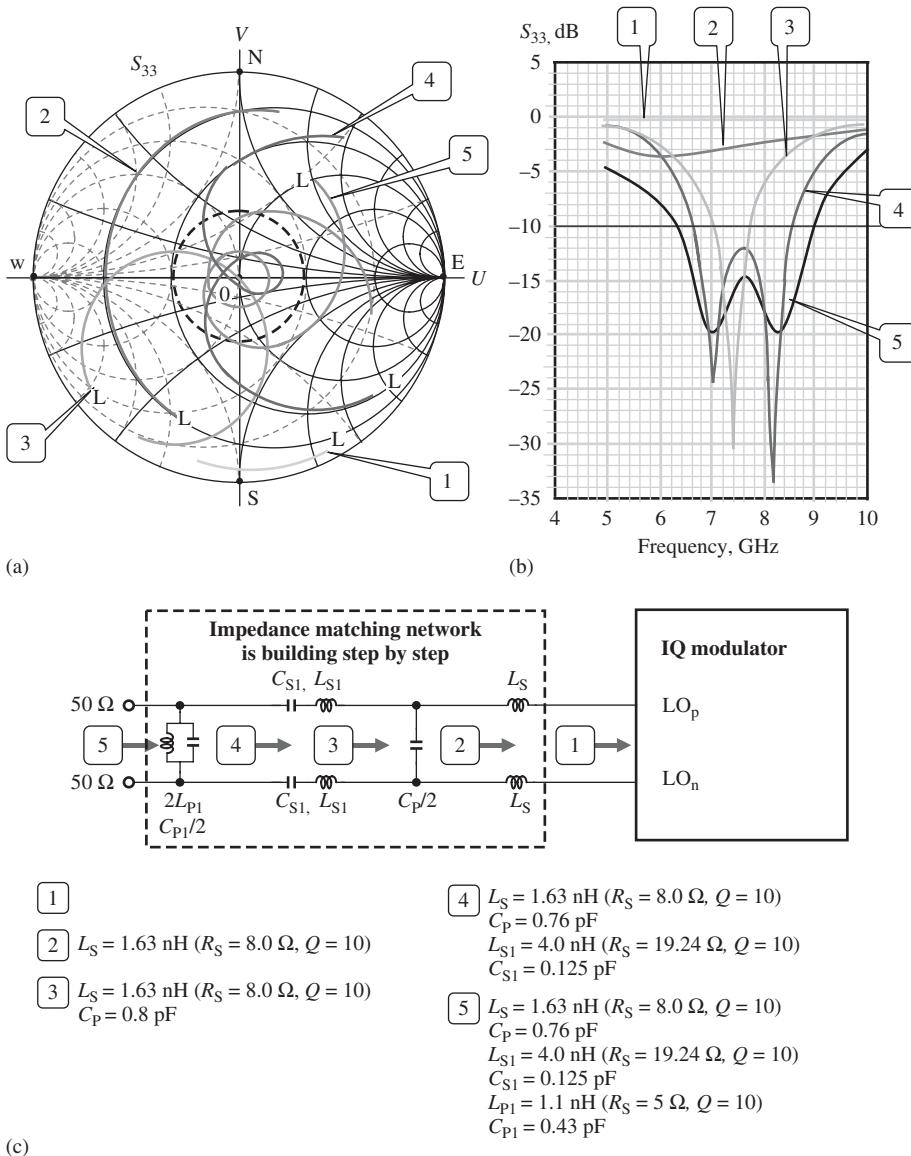
$$(\text{BW}_{RF \text{ or } LO})_{\text{relative}} = \Delta f / f_C = 2640 / [(8976 + 6336)/2] = 34.48\%.$$

Obviously, the design of the input impedance matching network at the LO port as shown in Figure 4.41 and the output impedance matching network at the RF port as shown in Figure 4.42 is successful.

Figure 4.44 shows the evolution of  $S_{33}$  at the LO port as the impedance matching parts are inserted successively. Figure 4.45 shows the evolution of  $S_{22}$  at the RF port as the impedance matching parts are inserted successively. By the same way as described in Example 1, readers can follow the evolution of the bandwidth accordingly.

## 4.5 DISCUSSION OF PASSIVE WIDEBAND IMPEDANCE MATCHING NETWORK

In Section 4.4, the designs of the IQ modulators with MOSFETs were taken as examples in the discussion of wideband impedance matching. As a matter of fact, they represent the most interesting cases because MOSFET devices have become more popular than the other devices nowadays. On the other hand, all the schemes and ideas about wideband impedance matching can be applied to circuit designs with other types of devices.

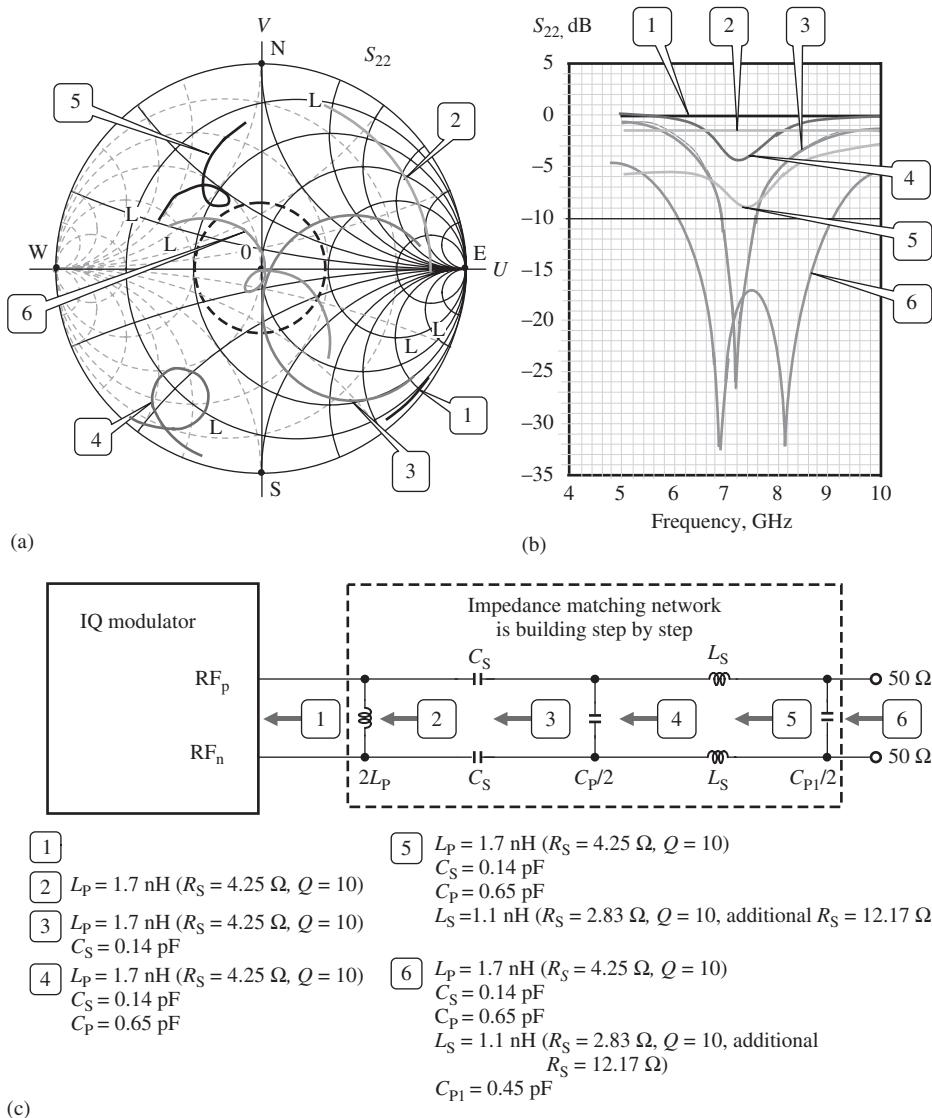


**Figure 4.44.** Evolution of bandwidth at the LO port as the parts are inserted into impedance matching network ( $f = 63368976 \text{ MHz}$ ). (a)  $S_{33}$  relocated on the Smith Chart after one arm or branch is inserted into the impedance matching network. (b)  $S_{33}$  versus  $f$  curves for each location of  $S_{33}$  on the Smith chart in the adjacent left-hand side. (c) Impedance matching at RF port.

On the basis of the discussion in the previous sections, a passive wideband impedance matching network for MOSFET devices can be formalized into two cases.

#### 4.5.1 Impedance Matching for the Gate of a MOSFET Device

An RF block built by MOSFET devices in which the gates of the devices are high-impedance terminals to be matched is quite a popular case. There are many examples



**Figure 4.45.** Evolution of bandwidth at RF port as the parts are inserted into impedance matching network ( $f = 63368976\text{ MHz}$ ). (a)  $S_{22}$  relocated on the Smith Chart after one arm or branch is inserted into the impedance matching network. (b)  $S_{22}$  versus  $f$  curves for each location of  $S_{22}$  on the Smith chart in the adjacent left-hand side. (c) Impedance matching at RF port.

available, such as the input of an LNA, the inputs of the RF and LO ports of a Gilbert cell, and so on.

In general, we have four choices for the first part applied to the impedance matching network:

1. *A capacitor in parallel.* This is not welcome because it will short the incoming RF signal to the ground and bring about unexpected attenuation.

2. *A capacitor in series.* This is not welcome because it will move the impedance at the gate of the MOSFET device to a higher impedance area because the impedance at the gate of a MOSFET device is originally capacitive with high impedance.
3. *An inductor in parallel.* This is in fact not qualified to be the first matching part because in such a case a DC blocking capacitor must be inserted between this inductor and the gate so as to avoid short-circuiting the DC bias.
4. *An inductor in series.* This is welcome because it can move the high-capacitive impedance at the gate to a point of low impedance.

Consequently, the topology of a wideband impedance matching network for the gates of MOSFET devices is suggested and is shown in Figure 4.46. The examples have been presented in Section 4.5 and the simulation results are shown in Figures 4.30 and 4.44, where  $LO_p$  and  $LO_n$  are the LO differential inputs of the I or Q modulator, which are equivalent to  $IN_p$  (Gate 1) and  $IN_n$  (Gate 2) as shown in Figure 4.46. In Section 4.5, this topology worked successfully in the designs of both frequency bands: group1 as well as the combination of group 3 and group 6 of the UWB system. This implies that this topology can expand the relative bandwidth by up to 40%.

The impedance matching network in Figure 4.46 contains five inductors and four capacitors. The five inductors will incur high cost and complicate the circuit design. However, this is the price that must be paid for the wide band requirement. As a matter of fact, if it were a narrowband design, two inductors and two capacitors,  $C_{S1}, L_{S1}, C_{P1}/2$  and  $2L_{P1}$ , could be omitted, and only two inductors,  $2 \times L_S$ , and one capacitor,  $C_P/2$ , would be required.

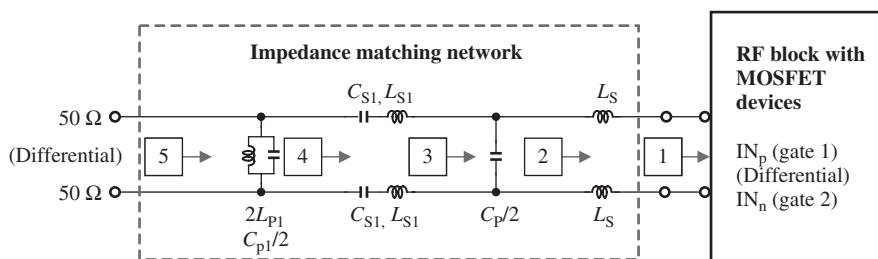
In order to lower the cost and simplify the circuit design, an alternative topology as shown in Figure 4.47 could be applied for the construction of a wideband impedance matching network.

This topology contains five capacitors but only three inductors, so the part count, especially the part count of inductors, is reduced.

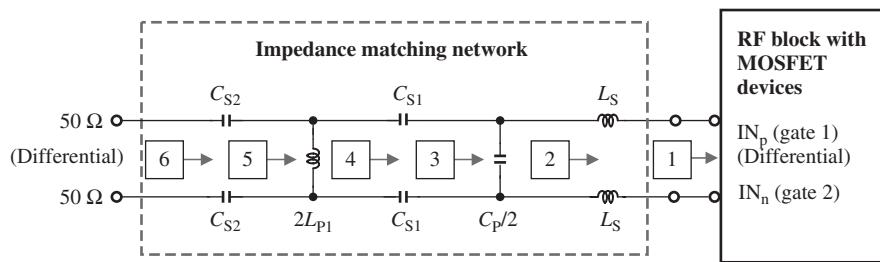
#### 4.5.2 Impedance Matching for the Drain of a MOSFET Device

An RF block built by MOSFET devices in which the drain of the MOSFET device is a high-impedance terminal to be matched is a quite popular case. There are many examples available, such as, the output of LNA, the output of a Gilbert cell for an IQ modulator, and so on.

In general, we cannot but apply an inductor as the first part to the impedance matching network.



**Figure 4.46.** Topology of impedance matching network at gates of MOSFET devices with five inductors and four capacitors, and with three inductors and five capacitors.



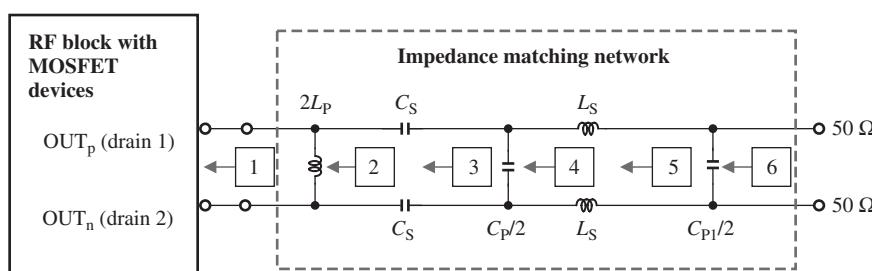
**Figure 4.47.** Alternative topology of an impedance matching network at the gates of MOSFET devices with three inductors and five capacitors.

- The first part must be connected between the drain and the DC power supply. This implies that the first part must be connected in parallel.
- Obviously, a capacitor is not a viable candidate.
- A resistor is also not a valid candidate because it forces the DC power supply to drop a certain amount of DC voltage, and then the DC voltage drop across the device may be much lower than the DC power supply voltage. Consequently, it increases the nonlinearity of the device, thereby degrading its performance.

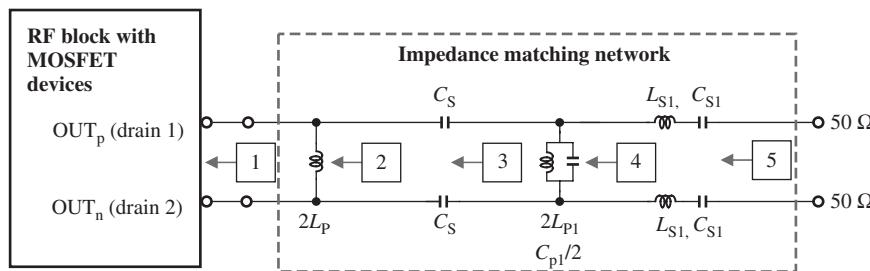
The topology of a wideband impedance matching network for the drain of MOSFET devices is suggested and is shown in Figure 4.48. The examples have been presented in Section 4.5 and the simulation results are shown in Figures 4.35 and 4.45, where  $RF_p$  and  $RF_n$  are the RF differential outputs of the I or Q modulator, which are equivalent to  $OUT_p$  (drain 1) and  $OUT_n$  (drain 2) as shown in Figure 4.48.

The impedance matching network in Figure 4.48 contains three inductors and four capacitors. The three inductors will incur a somewhat high cost and also complicate the circuit design. However, this price must be paid for the wideband requirement. As a matter of fact, if this were a narrowband design, two inductors and two capacitors,  $C_p/2$ ,  $2 \times L_s$ , and  $C_{p1}/2$ , could be omitted, and only one inductor,  $2L_p$ , and two capacitors,  $2 \times C_s$ , would be needed.

In Section 4.5, this topology worked successfully in the design for the combined frequency band of group 3 and group 6 in a UWB system. The alternative topology suggested is shown in Figure 4.49. The price of this topology shown in Figure 4.49 is the extra inductor and capacitor compared to the number of parts in the topology shown in Figure 4.48.



**Figure 4.48.** Topology of impedance matching network at the drain of MOSFET devices with three inductors and four capacitors.



**Figure 4.49.** Alternative topology of an impedance matching network at the drains of MOSFET devices with four inductors and five capacitors.

## FURTHER READING

- Carlin GJ, Amstutz P. On optimum broad-band matching. *IEEE Trans Circ Syst* 1981;CAS-28(5): 401–405.
- Cioffi GJ. Broad-band distributed amplifier impedance-matching techniques. *IEEE Trans Microw Theor Tech* 1989;37(12):1870–1876.
- Laughlin GJ. A new impedance-matched wide-band balun and magic tee. *IEEE Trans Microw Theor Tech* 1976;24(3):135–141.
- O'Meara T. Very-wide-band impedance-matching network. *IRE Trans Compon Parts* 1962;9(1): 38–44.
- Yarman GJ. New approaches to broadband matching problems. *IMS Workshop*; Boston, Massachusetts; 1983 Jun.
- Yarman GJ. Modern approaches to broadband matching problems. *IEE Proc* 1985;132(2):87–92.
- Youla DC. A new theory of broad-band matching. *IEEE Trans Circ Theor* 1964;CT-11:30–50.

## EXERCISES

1. What is the criterion of a narrowband and wideband block or system?
2. The return loss pattern moves around on the Smith Chart after a part is inserted into the impedance matching network. A return loss pattern is either squeezed into a smaller area or scattered over a larger area while its central position in Smith Chart remains almost the same. In which case is its bandwidth narrow? In which case is its bandwidth wide?
3. After a part is inserted into the impedance matching network, either the low- or high-frequency end of the return loss pattern moves in the same direction in Smith Chart, either clockwise or counterclockwise. How can we effect a movement of the low and high end in different directions, one clockwise and one counterclockwise?
4. What is the moving path of impedance in the Smith Chart if the reactance is increased?
5. What is the moving direction of impedance in the Smith Chart if the susceptance is increased?
6. The expansion of bandwidth of a block or system can be obtained by building an impedance matching network reasonably. Is it also possible to reduce the noise of a block or a system by building an impedance matching network reasonably?

7. What is the essential difference between a filter design and an impedance matching network design?
8. Considering part count and cost, what is a reasonable maximum part count (inductors and capacitors) to implement an impedance matching network in the wideband case?

## ANSWERS

1. The approximate criterion is
  - (a) For narrowband:  $\Delta f/f_c \ll 15\%$ ;
  - (b) For wideband:  $\Delta f/f_c \sim$  or  $>15\%$ .
2. If the central positions of two return loss patterns in the Smith Chart is almost the same, a return loss pattern squeezed into a smaller area means that its bandwidth is narrowed; a return loss pattern scattered over a larger area means that its bandwidth is widened.
3. If an LC combined part in series is inserted into the impedance matching network in series and its self-resonant frequency corresponds to one of the IFs, or, if an LC combined part in parallel is connected to the impedance matching network in parallel and its self-resonant frequency corresponds to one of the IFs, the low and the high end of the impedance pattern in Smith Chart will be moved in different directions, one clockwise and the other counterclockwise.
4. Moving along the circle with  $r = \text{constant}$  clockwise.
5. Moving along the circle with  $g = \text{constant}$  clockwise.
6. The expansion of bandwidth of a block or a system can be obtained by building an impedance matching network reasonably. However, it is impossible to reduce the noise of a block or a system by building an impedance matching network reasonably because the noise of a block is mainly generated by the device and its attached parts and not by the impedance matching network.
7. The input and output impedance of a filter usually is  $50 \Omega$ , which is independent of the frequency. On the contrary, the impedance of an impedance matching network usually is  $50 \Omega$  at one end. However, the impedance at other end is not  $50 \Omega$  but changes as the operating frequency is varied.
8. For an impedance matching network working in wideband cases,  
the reasonable maximum of part count for the single-ended case is <4;  
the reasonable maximum of part count for the differential configuration is <8.

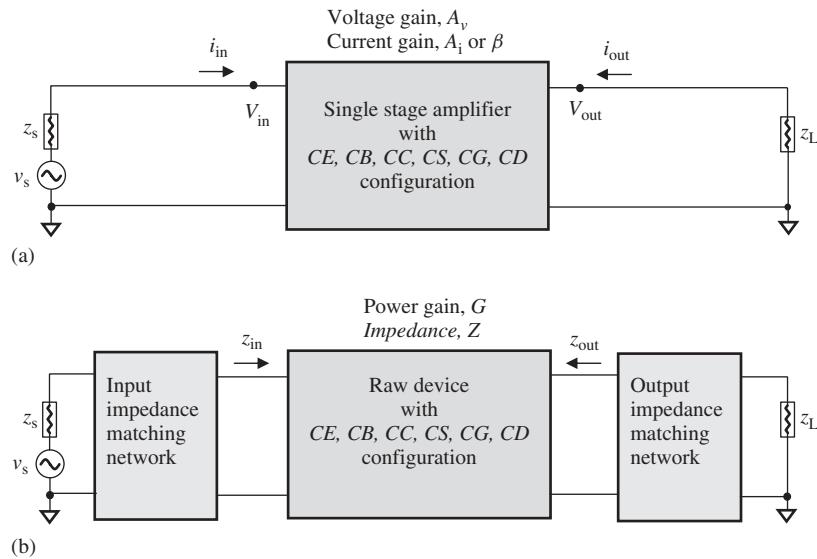
# IMPEDANCE AND GAIN OF A RAW DEVICE

## 5.1 INTRODUCTION

In electrical science and technology, the single-stage amplifier is almost indispensable. However, its implications in digital circuit design are different from those in RF circuit design. Figure 5.1 shows such a difference.

A single-stage amplifier in digital circuit design consists of a transistor and a couple of passive parts that provide the DC power, DC blocking, and DC bias, and offer an input and an output. Its performance is characterized by the input and output voltages or current, namely,  $V_{in}$ ,  $i_{in}$ ,  $V_{out}$ ,  $i_{out}$ , in which the voltage and current gain  $A_v$  and  $A_i$  (or  $\beta$ ), are of main concern.

A single-stage amplifier in an RF circuit design consists of three portions: the input impedance matching network, the raw device itself, and the output impedance matching network. A single-stage amplifier in the digital realm is, in fact, merely a raw device in the RF realm. A raw device without impedance matching cannot be an independently functioning block because the most important objective of an RF block is its power gain  $G$  and not its voltage or current gain. Impedance matching takes priority in order to ensure superior power transport or manipulation. The input impedance matching network must be matched between the source impedance  $z_S$  and the input impedance of the device  $z_{in}$ , while the output impedance matching network must be matched between the output impedance of the device  $z_{out}$  and the load impedance  $z_L$ . In RF circuit design, the main concerns of the raw device are its input and output impedances  $z_{in}$  and  $z_{out}$ . It must be noted that the input or output impedance matching networks should not contain the source or the load impedances  $z_S$  and  $z_L$  because our examination is concerned



**Figure 5.1.** Different viewpoints between single stage amplifier and raw device. (a) A single-stage amplifier in digital circuit design. (b) A single-stage amplifier in RF circuit design.

with the single-stage amplifier as shown in Figure 5.1(a) or the raw device block as shown in Figure 5.1(b), and not with the source or the load. In some published technical books, the equations of the input and output impedances do contain the source or the load impedances  $z_s$  and  $z_L$ . This is not a mistake, but is not suitable for RF circuit design.

For an RF circuit designer, the parameters of primary concern are the input and output impedances of the raw device,  $z_{in}$  and  $z_{out}$ , by which the input and output impedance matching networks can be built for power gain. Nevertheless, RF designers must also be familiar with voltage gain or current gain because, as a matter of fact, the power gain consists of voltage gain and current gain. Consequently, this chapter will cover voltage gain and current gain of a single-stage amplifier, but, since this book is aimed at RF designers, we will focus more on the input and output impedances. Study of the input and output impedances of a device with different configurations is a prerequisite in RF circuit design.

We will discuss six configurations of the raw device.

There are three configurations for the bipolar transistor:

- CE (common emitter), including CE with emitter degeneration;
- CB (common base);
- CC (common collector, or emitter follower).

Also, there are three configurations for the MOSFET (metal–oxide–semiconductor field-effect transistor):

- CS (common source), including CS with source degeneration;
- CG (common gate);
- CD (common drain or source follower).

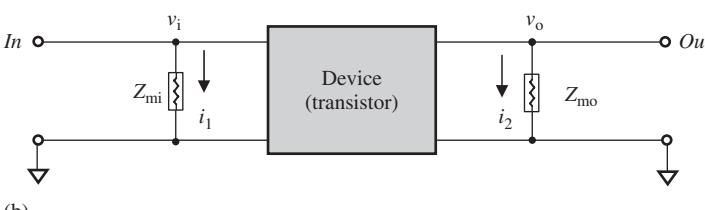
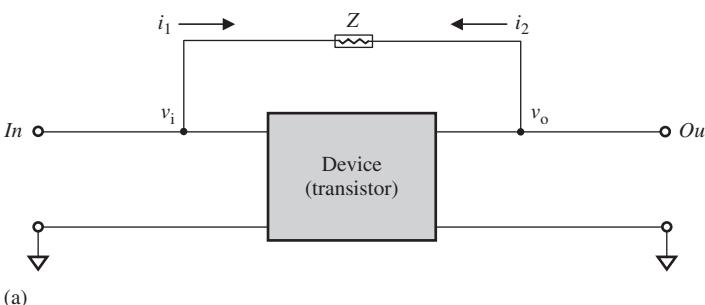
The characterization of the device is usually presented through the following parameters:

- *Open-circuit Voltage Gain.* This tests the voltage magnification capability of the device. The testing condition is that the output terminal is open or is not loaded, that is,  $Z_L \rightarrow \infty$ . “Open-circuit” implies that the testing represents the character of the device itself without other circuits connected to it, so that the testing indeed represents the character of the device itself.
- *Short-circuit Current Drain.* This tests the current magnification capability of the device. The test conditions are that the output terminal is shorted and the input terminal is provided with a small testing current source. Again, “short-circuit” implies that the testing represents the character of the device itself without other circuits to disturb it, so that the testing indeed represents the character of the device itself.
- *Input and Output Impedance Testing.* This tests the power transport capability of the device. From this testing, one can observe how well the device can be matched with its source and load so that the power transport or manipulation can be optimized. The tests of both input and output impedance are conducted inward to the device so that the results truly represent the character of the device itself.

## 5.2 MILLER EFFECT

Before discussing of the characterization of a device, let us introduce the Miller effect and its impact on a device.

A device operating at a high frequency has a feedback of the signal from the output to the input if its isolation between output and input is imperfect. Figure 5.2(a) shows a device block in which a feedback element with impedance  $Z$  is connected between the



**Figure 5.2.** Miller effect on a device (transistor). (a) A device with feedback element, impedance  $Z$ , connected between input and output. (b) Equivalent circuit of the device shown in (a),  $A_v = v_o/v_i$ .

output and the input. The open-circuit voltage gain of the device is

$$A_v = \frac{v_o}{v_i}, \quad (5.1)$$

where

- $A_v$  = the open-circuit voltage gain,
- $v_o$  = the output voltage at output node, and
- $v_i$  = the input voltage at input node.

Miller's theorem states that, if a device has an open-circuit voltage gain  $A_v$  between the input and the output node, then the feedback element with impedance  $Z$  connected between these two nodes can be replaced by an equivalent element connected to ground at each node. The impedance of the equivalent element at the input node is

$$Z_{mi} = \left| \frac{1}{A_v - 1} \right| Z, \quad (5.2)$$

and at the output node

$$Z_{mo} = \left| \frac{A_v}{A_v - 1} \right| Z, \quad (5.3)$$

where

- $Z_{mi}$  = the equivalent input Miller impedance at input node connected to the ground; and
- $Z_{mo}$  = the equivalent output Miller impedance at output node connected to the ground.

Let us verify the expressions (5.2) and (5.3). In Figure 5.2(a), a current  $i_1$  is drawn from the input to the output or a current  $i_2$  is drawn from the output to the input through the feedback element with impedance  $Z$ . The direction of the actual current flow depends on the magnitudes of  $v_i$  and  $v_o$ . In Figure 5.2(a), the direction of  $i_1$  corresponds to cases when  $v_i > v_o$ , while the direction of  $i_2$  corresponds to those cases when  $v_i < v_o$ . Let us set up a common rule to define the direction of either  $i_1$  or  $i_2$ . The current flows away from the node. Consequently, the mathematical derivations must be distinguished with two cases, that is,  $v_i > v_o$  and  $v_i < v_o$ .

1. When  $v_i > v_o$  or  $A_v < 1$ .

At the input node in Figure 5.2(a)

$$i_1 = \frac{v_i - v_o}{Z} = \frac{v_i - A_v v_i}{Z} = \frac{v_i}{\frac{1}{(1-A_v)}Z}, \quad (5.4)$$

and at the input node in Figure 5.2(b)

$$i_1 = \frac{v_i}{Z_{mi}}. \quad (5.5)$$

By comparing (5.4) with (5.5),

$$Z_{mi} = \frac{1}{1 - A_v} Z. \quad (5.6)$$

At the output node in Figure 5.2(a),

$$i_2 = -\frac{v_o - v_i}{Z} = -\frac{(v_o - v_o)/A_v}{Z} = \frac{v_o}{\frac{A_v}{(1-A_v)}Z}, \quad (5.7)$$

and at the output node in Figure 5.2(b)

$$i_2 = \frac{v_o}{Z_{mo}}. \quad (5.8)$$

By comparing (5.7) with (5.8),

$$Z_{mo} = \frac{A_v}{1 - A_v} Z. \quad (5.9)$$

2. When  $v_i < v_o$  or  $A_v > 1$ .

At the input node in Figure 5.2(a)

$$i_1 = -\frac{v_i - v_o}{Z} = -\frac{v_i - A_v v_i}{Z} = \frac{v_i}{\frac{1}{(A_v-1)}Z}. \quad (5.10)$$

By comparing (5.10) with (5.5)

$$Z_{mi} = \frac{1}{A_v - 1} Z. \quad (5.11)$$

At the output node in Figure 5.2(a)

$$i_2 = \frac{v_o - v_i}{Z} = \frac{(v_o - v_o)/A_v}{Z} = \frac{v_o}{\frac{A_v}{(A_v-1)}Z}. \quad (5.12)$$

By comparing (5.12) with (5.8),

$$Z_{mo} = \frac{A_v}{A_v - 1} Z. \quad (5.13)$$

As a matter of fact, expressions (5.6) and (5.9) and expressions (5.11) and (5.13) for the two cases can be combined into two expressions, which are

$$Z_{mi} = \left| \frac{1}{A_v - 1} \right| Z, \quad (5.2)$$

$$Z_{mo} = \left| \frac{A_v}{A_v - 1} \right| Z. \quad (5.3)$$

So, the derivation from (5.4) to (5.13) verifies Miller's theorem.

When the absolute value of the voltage gain  $A_v$  is much greater than 1, that is,

$$|A_v| \gg 1, \quad (5.14)$$

then,

$$Z_{mi} \approx \frac{Z}{|A_v|}, \quad (5.15)$$

$$Z_{mo} \approx Z. \quad (5.16)$$

This illustrates that the equivalent input Miller impedance  $Z_{mi}$  is  $A_v$  times less than the feedback impedance  $Z$ , while the equivalent output Miller impedance  $Z_{mo}$  is almost the same as the feedback impedance  $Z$ . In other words, the variation of the input impedance may be significant whereas the variation of the output impedance is approximately equal to the feedback impedance  $Z$ . For example, if the feedback element is a capacitor  $C$  and its impedance is

$$Z = \frac{1}{j\omega C}, \quad (5.17)$$

then, from (5.15), the equivalent input, the Miller impedance is

$$Z_{mi} \approx \frac{Z}{|A_v|} = \frac{1}{j\omega C |A_v|}, \quad (5.18)$$

which corresponds to an equivalent input Miller capacitance  $C_{mi}$ , that is,

$$C_{mi} \approx |A_v| C. \quad (5.19)$$

And, from (5.16) the equivalent input Miller impedance is

$$Z_{mo} \approx Z = \frac{1}{j\omega C}, \quad (5.20)$$

which corresponds to an equivalent output Miller capacitance  $C_{mo}$ , that is,

$$C_{mo} \approx C. \quad (5.21)$$

This implies that the existence of a feedback capacitor  $C$  in a device is equivalent to the addition of a capacitor with  $A_v$  times the feedback capacitance at the input terminal in parallel plus a capacitor with feedback capacitance at the output terminal in parallel ( $A_v$  being the open-circuit voltage gain of the device). This equivalent relationship is called *Miller effect*.

The Miller effect gives designers a serious warning that the feedback capacitor  $C$  may greatly degrade device performance because the input capacitance of the device could be significantly increased.

On the other hand, the Miller effect is very helpful to the circuit designer because the analysis of circuit performance is much easier if there is no feedback part. Through

the Miller effect described above, the device model can be modified so as to approach an equivalent model with no feedback element, as expected.

However, it must be noted that the impedance  $Z$  of the feedback element is dependent on the device configuration. Therefore, the Miller effect and the equivalent variations of input and output impedance due to the existence of the feedback element are dependent on the device configuration as well. By examining all the configurations of the device model, the feedback element in the configurations of CB, CG, CC, and CD is negligible. The Miller effect mainly exists in CE and CS configurations of the device. Consequently in the following discussion, the analysis of Miller effect will be conducted only for the configurations of CE and CS.

### 5.3 SMALL-SIGNAL MODEL OF A BIPOLAR TRANSISTOR

Figure 5.3 shows the DC characteristics of a bipolar transistor and Figure 5.4 shows those of a bipolar transistor with early voltage.

The DC characteristics of a bipolar transistor can be expressed by

$$I_c = I_s \left( 1 + \frac{V_{ce}}{V_A} \right) \exp \left( \frac{V_{be}}{V_T} \right) = I_s \left( 1 + \frac{V_{ce}}{V_A} \right) \exp \left( \frac{qV_{be}}{kT} \right). \quad (5.22)$$

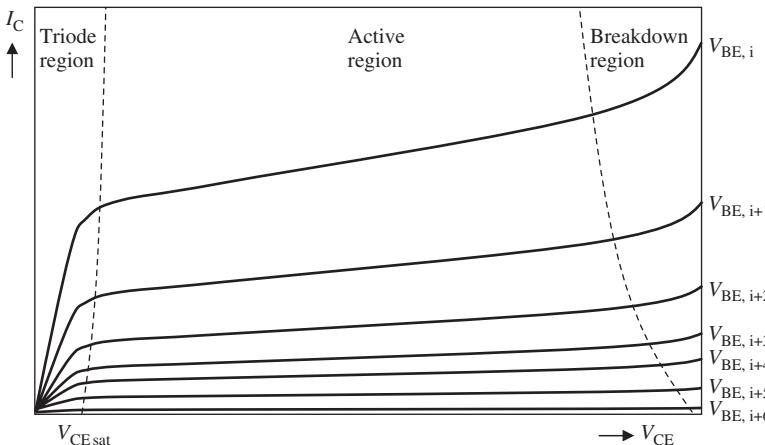


Figure 5.3. DC characteristics of a bipolar transistor ( $I_c$  vs  $V_{ce}$ ).

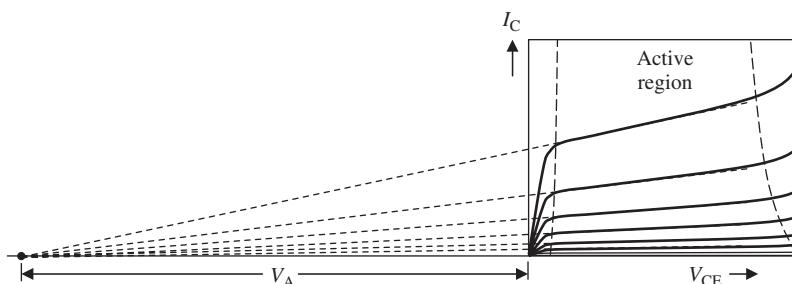


Figure 5.4. DC characteristics of a bipolar transistor ( $I_c$  vs  $V_{ce}$ ) and its early voltage.

Its transconductance is the derivative of the collector current  $I_c$  with respect to the voltage drop from base to emitter  $V_{be}$ , that is,

$$g_m = \frac{\partial I_c}{\partial V_{be}} = I_s \left( 1 + \frac{V_{ce}}{V_A} \right) \frac{q}{kT} \exp \left( \frac{qV_{be}}{kT} \right) = \frac{qI_c}{kT}, \quad (5.23)$$

$$C_\pi = C_b + C_{je}, \quad (5.24)$$

$$C_b = \tau_F g_m, \quad (5.25)$$

$$C_{je} = \frac{C_{jeo}}{\sqrt[3]{1 - \frac{V_D}{\psi_o}}}, \quad (5.26)$$

$$C_\mu = \frac{C_{\mu o}}{\left( 1 - \frac{V_{cb}}{\psi_o} \right)^n}, \quad (5.27)$$

$$r_\pi = \frac{\beta_o}{g_m}, \quad (5.28)$$

$$r_o = \frac{V_A}{I_c}, \quad (5.29)$$

$$r_\mu = (1 \rightarrow 10)\beta_o r_o, \quad (5.30)$$

where

$g_m$  = the transconductance ( $= 38 \text{ mA/V}$  if  $I_c = 1 \text{ mA}$  at room temperature),

$I_c$  = the total collector current,

$I_s$  = the collector current when  $V_{be} = 0$ ,

$Q$  = the charge of electron  $= 1.6 \times 10^{-19} \text{ Coulomb}$ ,

$k$  = the Boltzmann constant  $= 8.617 \times 10^{-5} \text{ eV/deg}$ ,

$T$  = the absolute temperature  $= 300^\circ\text{C}$  (room temperature),  $kT/q \approx 26 \text{ mV}$  at room temperature,

$V_{cb}$  = the voltage drop from collector to base,

$V_{be}$  = the voltage drop from base to emitter,

$V_{cb}$  = the forward bias on the collector-base junction,

$C_b$  = the base charging capacitance,

$\tau_F$  = the base transit time in the forward direction,

$C_{je}$  = the emitter depletion region capacitance

$C_{jeo}$  = the emitter depletion region capacitance when  $V_D = 0$ ,

$V_D$  = the bias on the junction,

$\Psi_o$  = the built-in potential, the voltage across the junction when bias is zero,

$C_\mu$  = the collector-base capacitance,

$C_{\mu o}$  = the value of  $C_\mu$  when  $V_{cb} = 0$ ,

$n$  = an exponent between 0.1 and 0.5,

$\beta_o$  = the small-signal current gain,

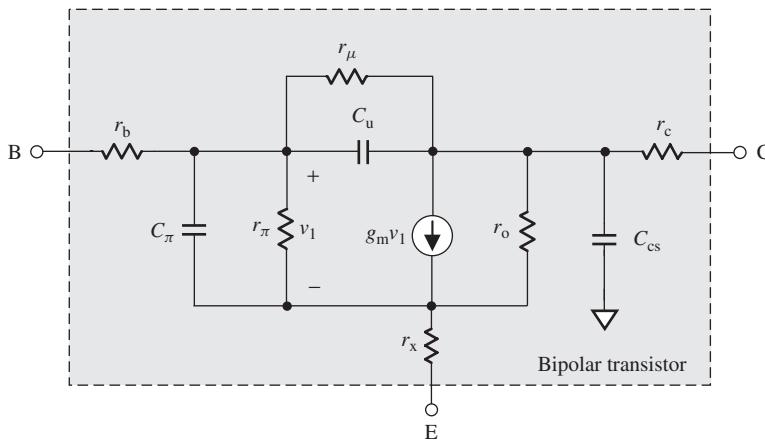
$V_A$  = the early voltage,

$r_\pi$  = the input resistance,

$r_o$  = the output resistance, and

$r_\mu$  = the collector-base resistance.

Figure 5.5 shows the small-signal equivalent model of a bipolar transistor, also called the *hybrid- $\pi$  model*,



**Figure 5.5.** Small-signal equivalent model of a bipolar transistor.

where

- E = the emitter of the bipolar transistor,
- B = the base of the bipolar transistor,
- C = the collector of the bipolar transistor,
- $r_b$  = the contact resistance of base and the resistance between base and emitter (about a few tens to a few hundred ohms),
- $C_\pi$  = the input capacitance, composed of the base charging capacitance  $C_b$  and the emitter-base depletion layer capacitance  $C_{je}$  (about 0.1–2 pF),
- $r_\pi$  = the input resistance (about a few tens of kilohms),
- $C_\mu$  = the collector-base capacitance (very small and is usually only a few femtofarads),
- $r_\mu$  = the collector-base resistance (has the highest value, and is usually in the range of tens of megohms),
- $r_x$  = the emitter lead resistance (only a few ohms and can be neglected if the emitter current is not too high),
- $C_{cs}$  = the collector-substrate capacitance (in the range of tens of femtofarads).
- $r_o$  = output resistance (about tens of kilohms), and
- $r_c$  = the collector resistance, composed of the resistance from the collector to the buried layer, resistance with the buried layer, and resistance from the buried layer to the base (about a few tens to few hundred ohms).

In most of circuit designs, the two resistors  $r_\mu$  and  $r_x$  are always neglected because usually

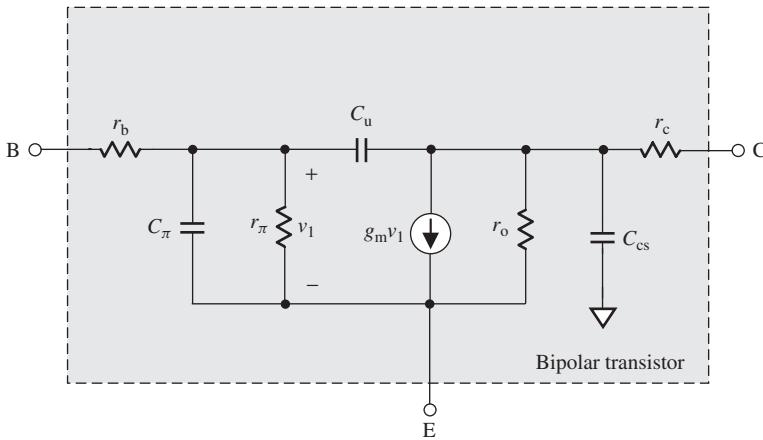
- $r_\mu$  is much higher than others, that is,

$$r_\mu \gg r_b, r_c, r_o, r_\pi, R_c. \quad (5.31)$$

On the contrary,  $r_x$  is much lower than of the others, that is,

$$r_x \ll r_b, r_c, r_o, r_\pi, R_c. \quad (5.32)$$

Then, the small-signal model shown in Figure 5.5 is simplified as the one shown in Figure 5.6 and will be applied to all the chapters in this book.



**Figure 5.6.** Simplified small-signal model of a bipolar transistor neglecting  $r_\mu$  and  $r_x$ .

## 5.4 BIPOLAR TRANSISTOR WITH CE (COMMON EMITTER) CONFIGURATION

A bipolar transistor with CE configuration indicates that the emitter is a common reference point or AC grounded point of the input and output terminals, which is plotted in Figure 5.7. In Figure 5.7(a), two “zero” capacitors, which approach zero impedance at the operating frequency, function as DC blocking capacitors and are denoted  $C_{\text{zero}}$ . The “infinite” inductor functions as an RF choke, which approaches infinite impedance at the operating frequencies and which is therefore is denoted  $L_{\text{infinite}}$ . The collector resistor  $R_c$  leads the DC power to the collector.

Figure 5.7(b) is a small-signal model of a CE device with an AC grounded emitter, in which all zerocapacitors and the infinite inductor have been ignored.

### 5.4.1 Open-Circuit Voltage Gain $A_{v,\text{CE}}$ of a CE Device

The open-circuit voltage gain  $A_v$  of a CE device can be analyzed on the basis of Figure 5.7(b), that is,

KCL at node C:

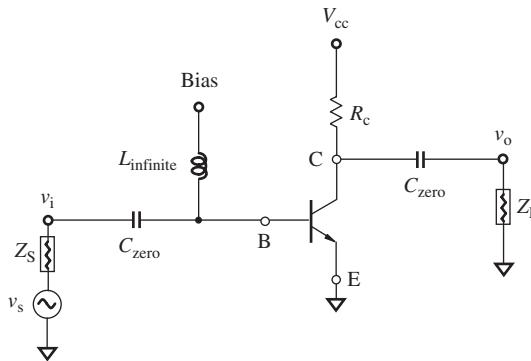
$$\frac{v_2 - v_o}{r_c} - \frac{v_o}{R_c // Z_L} = 0., \quad (5.33)$$

KCL at node  $P_1$ :

$$\frac{v_i - v_1}{r_b} = \frac{v_1 - v_2}{\frac{1}{j\omega C_\mu}} + \frac{v_1}{\frac{r_\pi}{1+j\omega C_\pi r_\pi}}. \quad (5.34)$$

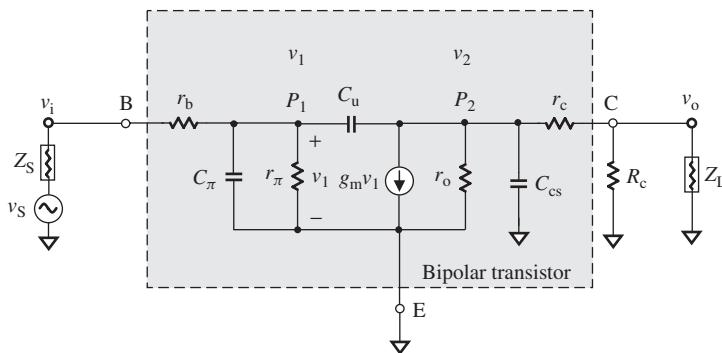
KCL at node  $P_2$ :

$$g_m v_1 + \frac{v_2}{(r_c + R_c // Z_L) // r_o // C_{cs}} + \frac{v_2 - v_1}{\frac{1}{j\omega C_\mu}} = 0. \quad (5.35)$$



- $C_{\text{zero}}$ : zero "capacitor approaches to zero capacitance in operating frequencies
- $L_{\text{infinite}}$ : "infinite" inductor approaches to infinite inductance in operating frequencies

(a)



- $Z_s$ : source impedance
- $Z_L$ : load impedance

(b)

**Figure 5.7.** A bipolar transistor with CE configuration. (a) Schematic of a bipolar transistor with CE configuration. (b) Small-signal model of a bipolar transistor with CE configuration.

These three equations can be rewritten as

$$v_2 = \left( 1 + \frac{r_c}{R_c // Z_L} \right) v_0, \quad (5.36)$$

$$v_i = \left[ 1 + \frac{r_b}{r_\pi} + j\omega(C_\mu + C_\pi)r_b \right] v_1 - j\omega C_\mu r_b v_2, \quad (5.37)$$

$$-(g_m - j\omega C_\mu)v_1 = \left[ \frac{1}{(r_c + R_c // Z_L) // r_o // C_{cs}} + j\omega C_\mu \right] v_2. \quad (5.38)$$

Substituting  $v_2$  in (5.36) into (5.38), we have

$$v_1 = \frac{1}{-(g_m - j\omega C_\mu)} \left[ \frac{1}{(r_c + R_c // Z_L) // r_o // C_{cs}} + j\omega C_\mu \right] \left( 1 + \frac{r_c}{R_c // Z_L} \right) v_0. \quad (5.39)$$

Replacing  $v_1$  in (5.39) by (5.37) and (5.36), we get

$$v_i = \left\{ \frac{\left[ 1 + \frac{r_b}{r_\pi} + j\omega(C_\mu + C_\pi)r_b \right]}{-(g_m - j\omega C_\mu)} \left[ \frac{1}{(r_c + R_c // Z_L) // r_o // C_{cs}} + j\omega C_\mu \right] - j\omega C_\mu r_b \right\} \\ \times \left( 1 + \frac{r_c}{R_c // Z_L} \right) v_o. \quad (5.40)$$

Finally, the open-circuit voltage gain of a CE device is

$$A_{v,CE} = \frac{v_o}{v_i} \\ = -\frac{R_c // Z_L}{r_c + R_c // Z_L} \frac{(g_m - j\omega C_\mu)}{\left[ 1 + \frac{r_b}{r_\pi} + j\omega(C_\mu + C_\pi)r_b \right] \left[ \frac{1}{(r_c + R_c // Z_L) // r_o // C_{cs}} + j\omega C_\mu \right]}, \\ + j\omega C_\mu r_b(g_m - j\omega C_\mu) \quad (5.41)$$

$$A_{v,CE} = \frac{v_o}{v_i} \\ = -\frac{R_c // Z_L}{r_c + R_c // Z_L} \frac{(g_m - j\omega C_\mu)}{\left[ 1 + \frac{r_b}{r_\pi} + j\omega(C_\mu + C_\pi)r_b \right] \left[ \frac{1}{r_o} + j\omega(C_\mu + C_{cs}) + \frac{1}{r_c + R_c // Z_L} \right]} \\ + j\omega C_\mu r_b(g_m - j\omega C_\mu) \quad (5.42)$$

The open-circuit voltage gain  $A_{v,CE}$  is in the case when

$$Z_L \rightarrow \infty \text{ is} \quad (5.43)$$

$$A_{v,CE} = -\frac{R_c}{r_c + R_c} \frac{(g_m - j\omega C_\mu)}{\left[ 1 + \frac{r_b}{r_\pi} + j\omega(C_\mu + C_\pi)r_b \right] \left[ \frac{1}{r_o} + j\omega(C_\mu + C_{cs}) + \frac{1}{r_c + R_c} \right]} \\ + j\omega C_\mu r_b(g_m - j\omega C_\mu) \quad (5.44)$$

The expression (5.44) can be further simplified by the following approximations:

- The resistor  $r_c$  is neglected because

$$r_c \ll R_c. \quad (5.45)$$

- The resistor  $r_b$  is not neglected but its value is much smaller than that of  $r_\pi$ , that is,

$$r_b \ll r_\pi. \quad (5.46)$$

- The value of capacitor  $C_\pi$  is much higher than the values of  $C_\mu$  and  $C_{cs}$ , that is,

$$C_\pi \gg C_\mu, \quad (5.47)$$

$$C_\pi \gg C_{cs}. \quad (5.48)$$

Then, the open-circuit voltage gain becomes

$$\begin{aligned} A_{v,CE} &= \frac{v_o}{v_i} \\ &= -\frac{(g_m - j\omega C_\mu)}{(1 + j\omega C_\pi r_b) \left[ \frac{r_o + R_c}{r_o R_c} + j\omega(C_\mu + C_{cs}) \right] + j\omega C_\mu r_b (g_m - j\omega C_\mu)}, \end{aligned} \quad (5.49)$$

Mathematically, the denominator can be written with two poles  $p_1$  and  $p_2$ , that is,

$$D(j\omega) = \left(1 - \frac{j\omega}{p_1}\right) \left(1 - \frac{j\omega}{p_2}\right) = 1 - j\omega \left(\frac{1}{p_1} + \frac{1}{p_2}\right) - \frac{\omega^2}{p_1 p_2} \approx 1 - \frac{j\omega}{p_1} - \frac{\omega^2}{p_1 p_2}, \quad (5.50)$$

if

$$|p_2| \gg |p_1|, \quad (5.51)$$

which is usually the case. Then, it is easy to write the expressions for the two poles from (5.49): that is,

$$p_1 \approx -\frac{1}{r_b} \frac{1}{C_\pi + C_\mu \left(1 + g_m \frac{r_o R_c}{r_o + R_c} + \frac{1}{r_b} \frac{r_o R_c}{r_o + R_c}\right) + C_{cs} \frac{1}{r_b} \frac{r_o R_c}{r_o + R_c}}, \quad (5.52)$$

$$p_2 \approx -\frac{1}{(C_\mu + C_{cs}) \frac{r_o R_c}{r_o + R_c}} - \frac{1}{\left(1 + \frac{C_{cs}}{C_\mu}\right) C_\pi \frac{1}{g_m + \frac{1}{r_b} + \frac{r_o + R_c}{r_o R_c}}} - \frac{1}{\left(1 + \frac{C_\mu}{C_{cs}}\right) C_\pi r_b}. \quad (5.53)$$

The first pole is the low-frequency pole and dominates over the second pole. The magnitude of  $p_1$  shows that the voltage gain is 3 dB below its low-frequency value at a frequency

$$\omega_{-3dB} = |p_1|. \quad (5.54)$$

The expression (5.49) is a general expression for the open-circuit voltage gain for both low and high frequencies.

In the low-frequency case, the terms containing capacitors in the expression (5.49) can be omitted, and the open-circuit voltage gain becomes

$$A_{v,CE}|_{\omega \rightarrow 0} = -g_m \frac{r_o R_c}{r_o + R_c}. \quad (5.55)$$

This is the low-frequency approximation of the open-circuit voltage gain.

The following should be noted:

- First, in some publications, the voltage gain is defined as

$$A_{v,CE} = \frac{v_o}{v_S}, \quad (5.56)$$

where  $v_S$  is the source voltage as shown in Figure 5.6(b). The expression (5.49) must be modified by multiplying by a factor of

$$\frac{(Z_S + r_b) // r_\pi}{Z_S + r_b} = \frac{r_\pi}{Z_S + r_b + r_\pi} \quad (5.57)$$

on its right-hand side. In an actual engineering circuit design, this factor is actually close to 1, so that the tolerance is acceptable.

However, the input voltage to the device should be  $v_i$  and not  $v_S$ , as shown in Figure 5.6(b), because the definition of voltage gain examines only the gain contributed by the device itself and not the gain contributed by the source impedance.

- Second, in some publications, the terms containing  $C_{cs}$  in the derivations for  $A_{v,CE}$  are neglected. This would seem contradictory, as the value of  $C_{cs}$  is usually higher than the value of  $C_\mu$ .

#### 5.4.2 Short-Circuit Current Gain $\beta_{CE}$ and Frequency Response of a CE Device

The frequency response of a transistor is most often specified by the transition frequency  $f_T$ , where the shorted-circuit current gain in CE configuration becomes 1. The short-circuit current gain is the ratio of  $i_o$  to  $i_i$  when the output is short-circuited to the ground, so  $v_o = 0$ . The transition frequency  $f_T$  is a measure of the maximum useful frequency of the transistor when used as an amplifier.

Figure 5.8 is drawn for the calculation of short-circuit current gain and frequency response of a bipolar transistor with a CE configuration.

Note that the output node is short-circuited,  $v_o = 0$ . Then, we have

$$v_1 \approx \frac{r_\pi}{1 + j\omega r_\pi (C_\pi + C_\mu)} i_i. \quad (5.58)$$

And, if the current  $i_o$  flowing through  $C_\mu$  is neglected,

$$i_o \approx g_m v_1 = \frac{g_m r_\pi}{1 + j\omega r_\pi (C_\pi + C_\mu)} i_i, \quad (5.59)$$

$$\beta_{CE}(j\omega) = \frac{i_o}{i_i} = \frac{\beta_o}{1 + j\beta_o \frac{(C_\pi + C_\mu)}{g_m} \omega}, \quad (5.60)$$

where  $\beta_{CE}(j\omega)$  is the short-circuit current gain of CE device.  $\beta_o = g_m r_\pi$  is the current gain at low frequencies.

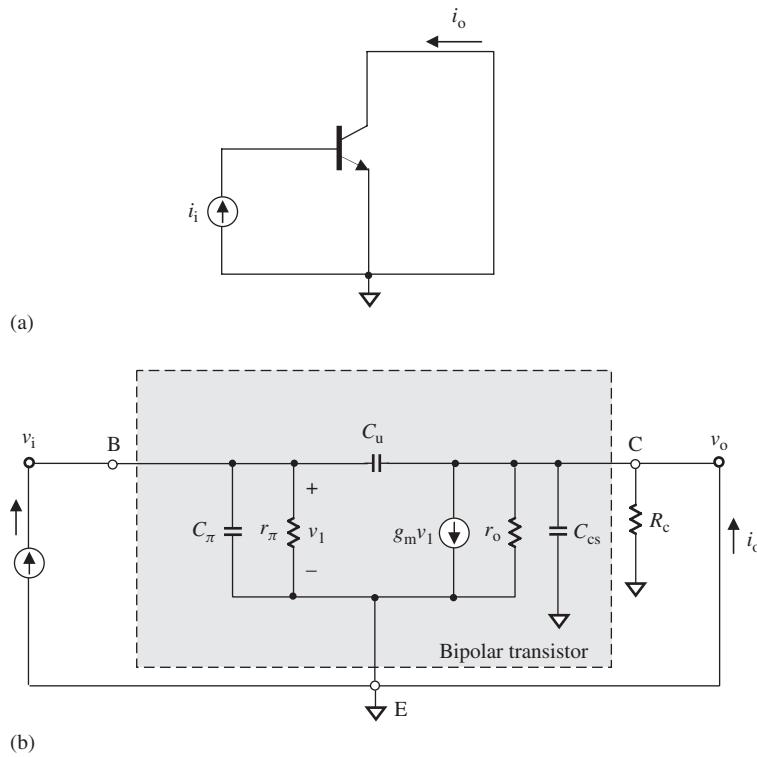
At high frequencies,

$$j\beta_o \frac{(C_\pi + C_\mu)}{g_m} \omega \gg 1, \quad (5.61)$$

$$\beta_{CE}(j\omega) \approx \frac{g_m}{j(C_\pi + C_\mu)\omega}. \quad (5.62)$$

When the input current is equal to the output current in magnitude, that is,

$$|\beta_{CE}(j\omega)| = 1, \quad (5.63)$$



**Figure 5.8.** Output short-circuited for calculating of current gain  $\beta_{CE}$  and frequency response of a bipolar with CE configuration. (a) Output is short-circuited for calculation of current gain  $\beta$  and frequency response. (b) Small-signal model for calculating of current gain  $\beta_{CE}$  and frequency.

then from (5.62) it can be found that the corresponding frequency  $\omega_T$  is

$$\omega_{T,CE} = \frac{g_m}{C_\pi + C_\mu}, \quad (5.64)$$

or

$$f_{T,CE} = \frac{1}{2\pi} \frac{g_m}{C_\pi + C_\mu}. \quad (5.65)$$

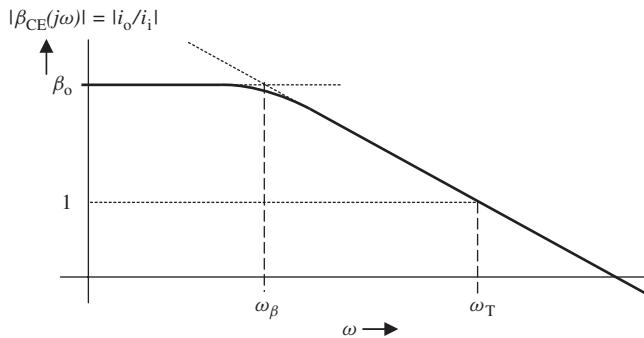
$\omega_{T,CE}$  is called *transition frequency* of the CE device.

On the other hand, when  $\beta_{CE}$  is 3 dB less than the low-frequency value  $\beta_o$ , that is,

$$|\beta_{CE}(j\omega)| = \frac{\beta_o}{\sqrt{2}}, \quad (5.66)$$

then from (5.60) it can be found that the beta cutoff, or corner frequency,  $\omega_{\beta,CE}$ , is

$$\omega_{\beta,CE} = \frac{1}{\beta_o} \frac{g_m}{C_\pi + C_\mu} = \frac{\omega_{T,CE}}{\beta_o}, \quad (5.67)$$



**Figure 5.9.** Frequency response of short-circuit current gain in a bipolar transistor with CE configuration.

or

$$f_{\beta,CE} = \frac{1}{2\pi\beta_0} \frac{g_m}{C_\pi + C_\mu} = \frac{f_{T,CE}}{\beta_0}. \quad (5.68)$$

At low frequencies, from (5.60)

$$\beta_{CE}(j\omega)|_{\omega \rightarrow 0} = \beta_0. \quad (5.69)$$

Figure 5.9 shows an example of the frequency response of a CE device.

### 5.4.3 Primary Input and Output Impedance of a CE (common emitter) device

The primary input and output impedances of a CE device can be calculated on the basis of Figure 5.10. The input impedance is

$$Z_{in} = r_b + c_\pi // r_\pi = r_b + \frac{r_\pi}{1 + j\omega C_\pi r_\pi} = \frac{r_b + r_\pi + j\omega C_\pi r_\pi r_b}{1 + j\omega C_\pi r_\pi}, \quad (5.70)$$

or

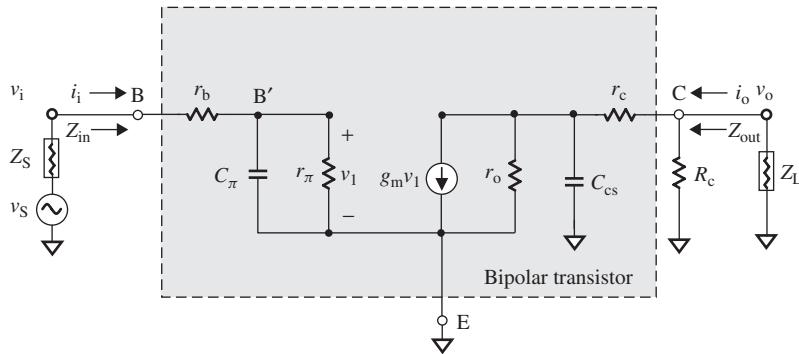
$$Z_{in} = \left[ r_b + \frac{r_\pi}{(r_\pi C_\pi \omega)^2 + 1} \right] - j \frac{r_\pi^2 C_\pi \omega}{(r_\pi C_\pi \omega)^2 + 1}. \quad (5.71)$$

The output impedance is

$$Z_{out} = (r_c + r_o // C_{cs}) // R_c = \left( r_c + \frac{r_o}{1 + j\omega C_{cs} r_o} \right) // R_c, \quad (5.72)$$

$$Z_{out} = \frac{r_c + r_o + jC_{cs}\omega r_c r_o}{R_c + r_c + r_o + jC_{cs}\omega r_o(r_c + R_c)} R_c. \quad (5.73)$$

It can be seen that the input impedance of a CE device is mainly determined by  $r_\pi$  and  $C_\pi$ , and  $r_b$  is negligible because usually  $r_b \ll r_\pi$ , and that the output impedance of a CE device is mainly determined by  $r_o$ ,  $R_c$ , and  $C_{cs}$  because usually  $r_c \ll r_o$  and  $R_c$ .



**Figure 5.10.** Simplified small-signal model of a CE device for impedance calculation.

In low frequencies, the capacitances  $C_\pi$  and  $C_{cs}$  are negligible. Therefore, expressions (5.71) and (5.73) become

$$Z_{in}|_{\omega \rightarrow 0} = r_b + r_\pi, \quad (5.74)$$

and

$$Z_{out}|_{\omega \rightarrow 0} = (r_c + r_o) // R_c. \quad (5.75)$$

#### 5.4.4 Miller's Effect in a Bipolar Transistor with CE Configuration

In order to simplify the discussion of the Miller effect, the value of the open-circuit voltage gain  $A_{v,CE}$  for the CE device is taken from the low-frequency cases, that is,

$$A_{v,CE} = \frac{v_o}{v_i} = -g_m \frac{r_o}{r_o + R_c} R_c. \quad (5.76)$$

This is, of course, a rough approximation.

Note that the feedback element in a CE device is the capacitor  $C_\mu$ . Then, from expressions (5.19), (5.21), and (5.76), the input and output Miller capacitances can be found as

$$C_{i,miller} = C_\mu \left( 1 + g_m \frac{r_o}{r_o + R_c} R_c \right), \quad (5.77)$$

and

$$C_{o,miller} = C_\mu \left( 1 + \frac{r_o + R_c}{g_m r_o R_c} \right). \quad (5.78)$$

The total input capacitance is therefore

$$C_{in} = C_\pi + C_{i,miller} = C_\pi + C_\mu \left( 1 + g_m \frac{r_o}{r_o + R_c} R_c \right). \quad (5.79)$$

And the total output capacitance is

$$C_{\text{out}} = C_{\text{cs}} + C_{\text{o,miller}} = C_{\text{cs}} + C_{\mu} \left( 1 + \frac{r_{\text{o}} + R_{\text{c}}}{g_{\text{m}} r_{\text{o}} R_{\text{c}}} \right). \quad (5.80)$$

From now on, in the analysis of circuitry involving a CE device, it will be more convenient to replace the equivalent model of the CE device shown in Figure 5.7(b) by Figure 5.11, in which the feedback capacitance  $C_{\mu}$  is included in the input and output capacitances  $C_{\text{in}}$  and  $C_{\text{out}}$ .

By means of the model modified with the Miller effect, shown in Figure 5.11, let us recalculate the open-circuit voltage gain.

The output voltage is

$$v_{\text{o}} = -g_{\text{m}} v_1 R_{\text{c}} // (r_{\text{c}} + r_{\text{o}} // C_{\text{out}}). \quad (5.81)$$

Note that

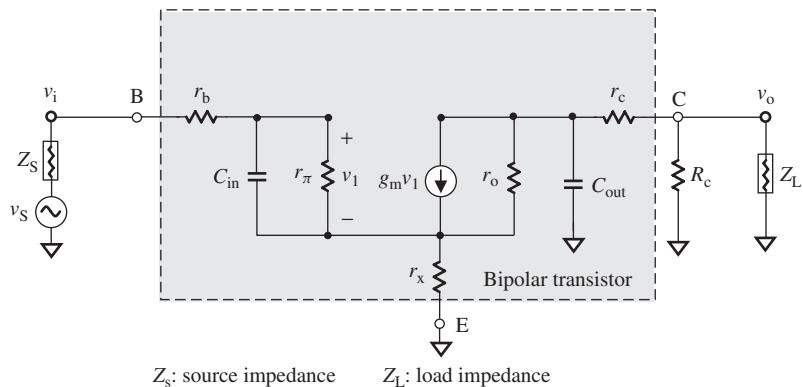
$$r_{\text{c}} + \frac{r_{\text{o}}}{1 + j\omega C_{\text{out}} r_{\text{o}}} = \frac{r_{\text{c}} + j\omega C_{\text{out}} r_{\text{c}} r_{\text{o}} + r_{\text{o}}}{1 + j\omega C_{\text{out}} r_{\text{o}}} \approx \frac{1 + j\omega C_{\text{out}} r_{\text{c}} r_{\text{o}}}{1 + j\omega C_{\text{out}} r_{\text{o}}}, \quad (5.82)$$

$$R_{\text{c}} // (r_{\text{c}} + r_{\text{o}} // C_{\text{out}}) = R_{\text{c}} // \left( \frac{1 + j\omega C_{\text{out}} r_{\text{c}} r_{\text{o}}}{1 + j\omega C_{\text{out}} r_{\text{o}}} r_{\text{o}} \right) = \frac{R_{\text{c}} \frac{1 + j\omega C_{\text{out}} r_{\text{c}}}{1 + j\omega C_{\text{out}} r_{\text{o}}} r_{\text{o}}}{R_{\text{c}} + \frac{1 + j\omega C_{\text{out}} r_{\text{c}}}{1 + j\omega C_{\text{out}} r_{\text{o}}} r_{\text{o}}}, \quad (5.83)$$

$$\begin{aligned} R_{\text{c}} // (r_{\text{c}} + r_{\text{o}} // C_{\text{out}}) &= \frac{R_{\text{c}} r_{\text{o}} (1 + j\omega C_{\text{out}} r_{\text{c}})}{R_{\text{c}} (1 + j\omega C_{\text{out}} r_{\text{o}}) + r_{\text{o}} (1 + j\omega C_{\text{out}} r_{\text{c}})} \\ &\approx \frac{R_{\text{c}} r_{\text{o}}}{R_{\text{c}} + r_{\text{o}}} \frac{1 + j\omega C_{\text{out}} r_{\text{c}}}{1 + j\omega C_{\text{out}} \frac{R_{\text{c}} r_{\text{o}}}{R_{\text{c}} + r_{\text{o}}}}. \end{aligned} \quad (5.84)$$

If

$$r_{\text{c}} \ll R_{\text{c}}, \quad (5.85)$$



**Figure 5.11.** Model of bipolar transistor with CE configuration with Miller capacitors included in  $C_{\text{in}}$  and  $C_{\text{out}}$ .

substituting (5.84) into (5.81), we have

$$v_o = -g_m v_i R_c // (r_c + r_o // C_{out}) \approx -g_m v_i \frac{r_o R_c}{r_o + R_c} \frac{1 + j\omega C_{out} r_c}{1 + j\omega C_{out} \frac{r_o R_c}{r_o + R_c}}. \quad (5.86)$$

The voltage across the capacitor  $C_{in}$  is

$$v_1 = \frac{r_\pi // C_{in}}{r_b + r_\pi // C_{in}} v_i = \frac{1}{1 + j\omega C_{in} r_b} v_i, \quad (5.87)$$

if

$$r_b \ll r_\pi. \quad (5.88)$$

The open-circuit voltage gain of CE device then becomes

$$A_{v,CE} = \frac{v_o}{v_i} = -g_m \frac{r_o R_c}{r_o + R_c} \frac{1}{1 + j\omega C_{in} r_b} \frac{1 + j\omega C_{out} r_c}{1 + j\omega C_{out} \frac{r_o R_c}{r_o + R_c}}. \quad (5.89)$$

It can be seen that the two poles exist in (5.89); they are

$$p_1 = -\frac{1}{C_{in} r_b} = -\frac{1}{r_b} \frac{1}{C_\pi + C_\mu \left(1 + g_m \frac{r_o}{r_o + R_c} R_c\right)}, \quad (5.90)$$

$$p_2 = -\frac{1}{C_{out} \frac{r_o R_c}{r_o + R_c}} = -\frac{1}{C_{cs} \frac{r_o R_c}{r_o + R_c} + C_\mu \left(\frac{r_o R_c}{r_o + R_c} + \frac{1}{g_m}\right)}. \quad (5.91)$$

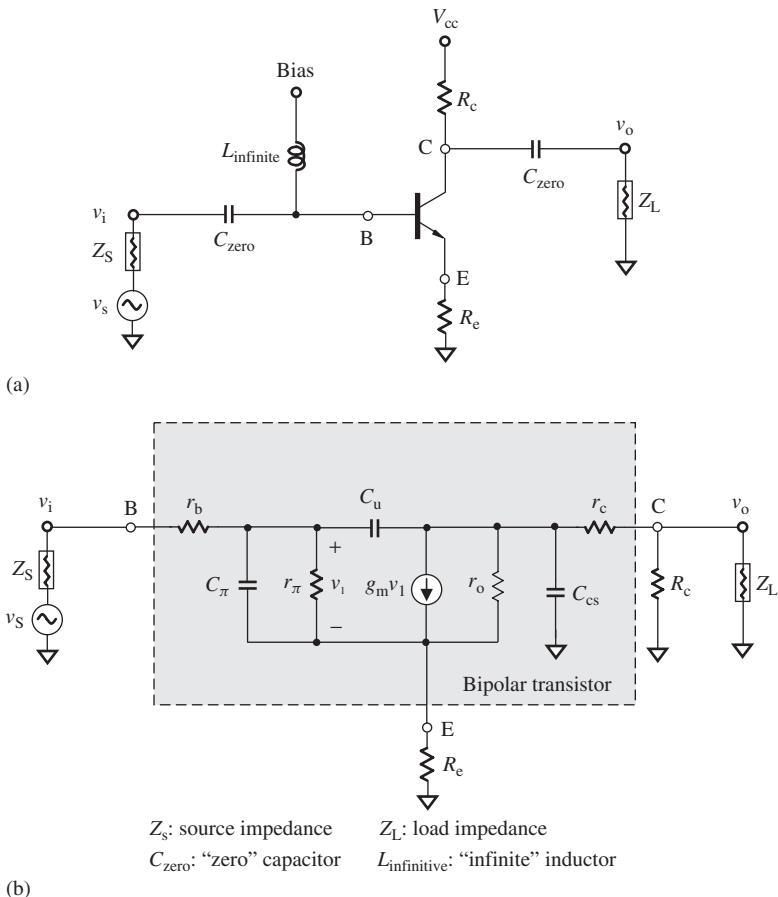
By comparing (5.90) with (5.52), we see that the  $p_1$  values are almost identical to each other. The difference is not too much. This implies that the Miller approximation is a good approach in circuit analysis.

By means of the equivalent model modified with the Miller effect as shown in Figure 5.11, the input and output impedances of a CE device can be modified from expressions (5.71) and (5.73) to the following:

$$Z_{in} = \frac{r_b + r_\pi + j\omega C_{in} r_\pi r_b}{1 + j\omega C_{in} r_\pi}, \quad (5.92)$$

$$Z_{out} = \frac{r_c + r_o + jC_{out} \omega r_c r_o}{R_c + r_c + r_o + jC_{out} \omega r_o (r_c + R_c)} R_c, \quad (5.93)$$

in which the capacitances  $C_\pi$  and  $C_{cs}$  in (5.71) and (5.73) are simply replaced by  $C_{in}$  and  $C_{out}$ , respectively. In low-frequency cases, the expressions (5.92) and (5.93) revert to the previous equations (5.74) and (5.75).



**Figure 5.12.** Bipolar transistor with CE configuration and emitter degeneration. (a) Schematic of a bipolar transistor with CE configuration and emitter degeneration. (b) Model of a bipolar transistor with CE configuration and emitter degeneration.

### 5.4.5 Emitter Degeneration

Figure 5.12 shows a bipolar device with a CE configuration and emitter degeneration. Its model is simplified by additionally neglecting  $C_\mu$ .

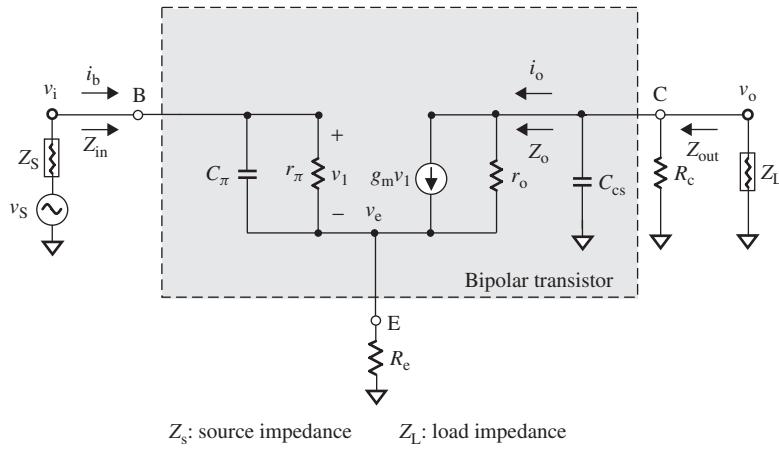
Then, the equivalent model of the bipolar transistor with a CE configuration shown in Figure 5.12(b) is simplified to that in Figure 5.13.

KCL at node E:

$$\frac{v_e}{R_e} + \frac{v_e + i_o(R_c // C_{cs})}{r_o} = g_m v_1 + i_b. \quad (5.94)$$

KCL at node C:

$$i_o + \frac{v_e + i_o(R_c // C_{cs})}{r_o} = g_m v_1. \quad (5.95)$$



**Figure 5.13.** Simplified model of bipolar transistor with CE configuration and emitter degeneration.

KVL at the input loop:

$$i_b = \frac{v_i - v_e}{r_\pi // C_\pi}, \quad (5.96)$$

$$v_1 = (r_\pi // C_\pi) i_b. \quad (5.97)$$

Solving (5.94) for  $i_o$ , substituting into (5.95), and applying (5.97) for  $v_1$ , we have

$$\begin{aligned} v_e &= \frac{R_e R_c}{R_c + R_e (1 + j\omega C_{cs} R_c) + r_o (1 + j\omega C_{cs} R_c)} \\ &\times \left[ 1 + \left( \frac{\beta_o}{1 + j\omega C_\pi r_\pi} + 1 \right) (1 + j\omega C_{cs} R_c) \frac{r_o}{R_c} \right] i_b, \end{aligned} \quad (5.98)$$

$$\begin{aligned} Z_{in} &= \frac{v_i}{i_b} = \frac{1}{1 + j\omega C_\pi r_\pi} \\ &\times \left( r_\pi + \frac{r_o (1 + j\omega C_{cs} R_c) (\beta_o + 1 + j\omega C_\pi r_\pi) + R_c (1 + j\omega C_\pi r_\pi) R_e}{R_c + (r_o + R_e) (1 + j\omega C_{cs} R_c)} R_e \right). \end{aligned} \quad (5.99)$$

If

$$r_o \gg R_c, \quad \text{and} \quad (5.100)$$

$$r_o \gg R_e, \quad (5.101)$$

$$Z_{in} = \frac{v_i}{i_b} = \frac{1}{1 + j\omega C_\pi r_\pi} [r_\pi + (\beta_o + 1 + j\omega C_\pi r_\pi) R_e]. \quad (5.102)$$

And at low frequencies

$$Z_{in}|_{\omega \rightarrow 0} \approx r_\pi + (\beta_o + 1) R_e. \quad (5.103)$$

At low frequencies, the input impedance is increased by an amount of  $(\beta_o + 1)R_e$  from  $r_\pi$ . At high frequencies, the input impedance is somewhat reduced from (5.102) to (5.99) as a result of the existence of a finite  $r_o$ .

Let us calculate the transconductance  $G_m = i_o/v_i$  with output short-circuited: Substituting (5.98) into (5.94) with  $R_c = 0$  and applying (5.97) for  $v_1$ , we have

$$v_e = \frac{\left( \frac{\beta_o}{1+j\omega C_\pi r_\pi} + 1 \right)}{\left( \frac{1}{R_e} + \frac{1}{r_o} \right) r_\pi // C_\pi + \left( \frac{\beta_o}{1+j\omega C_\pi r_\pi} + 1 \right)} v_i. \quad (5.104)$$

Substituting (5.98) and (5.104) into (5.97) with  $R_c = 0$  and applying (5.97) for  $v_1$ , we have

$$G_m = \frac{i_o}{v_i} = g_m \frac{1 - \frac{R_e(1+j\omega C_\pi r_\pi)}{\beta_o r_o}}{1 + g_m R_e \left[ 1 + \frac{(1+j\omega C_\pi r_\pi)}{\beta_o} + \frac{1}{g_m r_o} \right]}. \quad (5.105)$$

Usually

$$\beta_o \gg 1, \quad (5.106)$$

$$r_o \gg R_e, \quad (5.107)$$

$$g_m r_o \gg 1, \quad (5.108)$$

so

$$G_m \approx g_m \frac{1 - \frac{R_e j\omega C_\pi r_\pi}{\beta_o r_o}}{1 + g_m R_e \left[ 1 + \frac{j\omega C_\pi r_\pi}{\beta_o} \right]}. \quad (5.109)$$

At low frequencies,

$$G_m|_{\omega \rightarrow 0} \approx \frac{g_m}{1 + g_m R_e}. \quad (5.110)$$

From (5.105), (5.109), and (5.110) it can be seen that  $G_m$  drops from  $g_m$  as a result of the emitter resistor  $R_e$ . At low frequencies,  $G_m$  drops to half of  $g_m$  if  $R_e = 1/g_m$ . The output impedance is calculated by using Figure 5.13 when the input is short-circuited, that is,

$$v_i = 0, \quad (5.111)$$

then

$$v_1 = -i_o(r_\pi // C_\pi // R_e). \quad (5.112)$$

The current flowing through  $r_o$  is

$$i_o|_{r_o} = i_o - g_m v_1 = i_o [1 + g_m (r_\pi // C_\pi // R_e)], \quad (5.113)$$

$$v_o = -v_1 + i_o|_{r_o} r_o. \quad (5.114)$$

Substituting (5.112) and (5.113) into (5.114), we have

$$v_o = i_o(r_\pi // C_\pi // R_e) + i_o[1 + g_m(r_\pi // C_\pi // R_e)]r_o, \quad (5.115)$$

$$\begin{aligned} Z_o &= \frac{v_o}{i_o} = (r_\pi // C_\pi // R_e) + [1 + g_m(r_\pi // C_\pi // R_e)]r_o \\ &= r_o + (1 + g_m r_o)(r_\pi // C_\pi // R_e), \end{aligned} \quad (5.116)$$

$$\begin{aligned} Z_{out} &= Z_o // C_{cs} // R_c \\ &= [r_o + (1 + g_m r_o)(r_\pi // C_\pi // R_e)] // C_{cs} // R_c, \end{aligned} \quad (5.117)$$

$$Z_{out} = r_o \frac{1 + \left(\frac{1}{r_o} + g_m\right) \frac{r_\pi R_e}{r_\pi + R_e(1+j\omega C_\pi r_\pi)}}{1 + \frac{r_o}{R_c}(1+j\omega C_{cs} R_c) \left[1 + \left(\frac{1}{r_o} + g_m\right) \frac{r_\pi R_e}{r_\pi + R_e(1+j\omega C_\pi r_\pi)}\right]}. \quad (5.118)$$

Usually

$$r_o \gg 1, \quad (5.119)$$

so

$$Z_{out} \approx r_o \frac{1 + \frac{\beta_o R_e}{r_\pi + R_e(1+j\omega C_\pi r_\pi)}}{1 + \frac{r_o}{R_c}(1+j\omega C_{cs} R_c) \left[1 + \frac{\beta_o R_e}{r_\pi + R_e(1+j\omega C_\pi r_\pi)}\right]}. \quad (5.120)$$

In the cases of infinite  $R_c$ , that is,

$$R_c \rightarrow 0, \quad (5.121)$$

$$Z_{out}|_{R_c \rightarrow \infty} \approx r_o \frac{1 + \frac{\beta_o R_e}{r_\pi + R_e(1+j\omega C_\pi r_\pi)}}{1 + j\omega C_{cs} r_o \left[1 + \frac{\beta_o R_e}{r_\pi + R_e(1+j\omega C_\pi r_\pi)}\right]}. \quad (5.122)$$

At low frequencies,

$$Z_{out}|_{\omega \rightarrow 0} \approx r_o \frac{1 + \frac{\beta_o R_e}{r_\pi + R_e}}{1 + \frac{r_o}{R_c} \left[1 + \frac{\beta_o R_e}{r_\pi + R_e}\right]}. \quad (5.123)$$

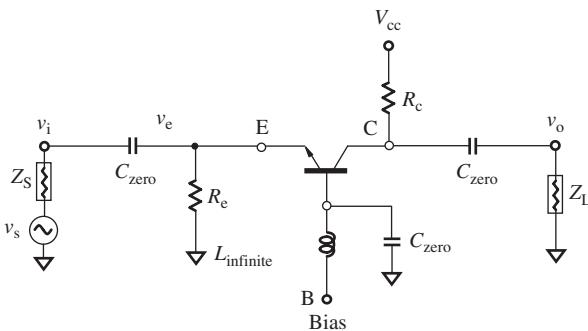
In the cases of infinite  $R_c$  and low frequencies,

$$Z_{out}|_{\omega \rightarrow 0, R_c \rightarrow \infty} \approx r_o \frac{1 + (\beta_o + 1) \frac{R_e}{r_\pi}}{1 + \frac{R_e}{r_\pi}}. \quad (5.124)$$

As is usual, if

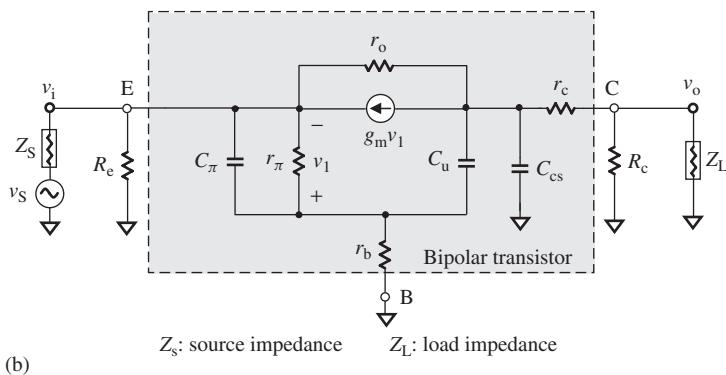
$$r_\pi \gg R_e, \quad (5.125)$$

$$Z_{out}|_{\omega \rightarrow 0, R_c \rightarrow \infty, r_\pi \gg R_e} \approx r_o(1 + g_m R_e). \quad (5.126)$$



- $C_{\text{zero}}$ : “zero” capacitor approaches to zero capacitance in operating frequencies
- $L_{\text{infinite}}$ : “infinitive” inductor approaches to infinitive inductance in operating frequencies

(a)



(b)

**Figure 5.14.** Equivalent circuit of bipolar transistor with CB configuration. (a) Schematic of a bipolar transistor with CB configuration. (b) Model of a bipolar transistor with CB configuration.

## 5.5 BIPOLAR TRANSISTOR WITH CB (COMMON BASE) CONFIGURATION

A bipolar transistor with CB configuration can be drawn as Figure 5.14. In Figure 5.14(a), two zero capacitors at the input and output, which approach zero impedance at the operating frequency, function as DC blocking capacitors and are denoted as  $C_{\text{zero}}$ . Another zero capacitor  $C_{\mu}$  is connected between the base and the ground so that the base is AC-grounded. The infinite inductor functions as an RF choke, which approaches infinite impedance at the operating frequencies and is denoted  $L_{\text{infinite}}$ . The collector resistor  $R_c$  leads the DC power to the collector. The emitter resistor  $R_e$  leads the DC power to the ground.

### 5.5.1 Open-Circuit Voltage Gain $A_{v,\text{CB}}$ of a CB Device

In the calculation of open-circuit voltage of a CB transistor, Figure 5.14 is simplified to Figure 5.15 by an additional approximation: the capacitor  $C_{\mu}$  is neglected because

$$C_{\mu} \ll C_{\text{cs}}, C_{\pi}. \quad (5.127)$$

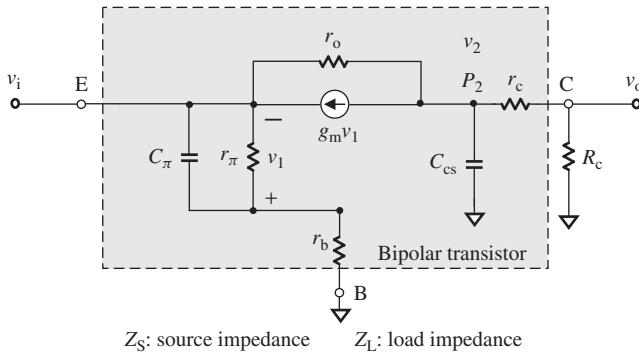


Figure 5.15. Equivalent circuit for calculating the open-circuit voltage gain of a CB device.

By applying KCL at node  $v_o$ ,

$$\frac{v_2}{(R_c + r_c) // C_{cs}} + g_m v_1 + \frac{v_o - v_i}{r_o} = 0. \quad (5.128)$$

Note that

$$v_i = -v_1. \quad (5.129)$$

$$v_2 = \frac{R_c + r_c}{R_c} v_o. \quad (5.130)$$

By replacing  $v_1$  by  $v_i$  in (5.128), from (5.129), and replacing  $v_2$  by  $v_o$  in (5.130), the expression (5.128) becomes

$$\frac{\frac{R_c + r_c}{R_c} v_o}{\frac{R_c + r_c}{1 + j\omega C_{cs}(R_c + r_c)}} - g_m v_i + \frac{v_o - v_i}{r_o} = 0, \quad (5.131)$$

$$\left( \frac{1 + j\omega C_{cs}(R_c + r_c)}{R_c} + \frac{1}{r_o} \right) v_o = \left( g_m + \frac{1}{r_o} \right) v_i, \quad (5.132)$$

$$A_{v,CB} = \frac{v_o}{v_i} = \frac{g_m R_c \left( 1 + \frac{1}{g_m r_o} \right)}{1 + j\omega C_{cs}(R_c + r_c) + \frac{R_c}{r_o}}. \quad (5.133)$$

Finally,

$$A_{v,CB} = \frac{r_o}{R_c + r_o} \frac{g_m R_c \left( 1 + \frac{1}{g_m r_o} \right)}{1 + j\omega C_{cs} \frac{(R_c + r_c)}{(R_c + r_o)} r_o}. \quad (5.134)$$

At low frequencies,

$$A_{v,CB}|_{\omega \rightarrow 0} = g_m \left( 1 + \frac{1}{g_m r_o} \right) \frac{R_c r_o}{R_c + r_o}. \quad (5.135)$$

If

$$r_o \gg R_c, \quad \text{and} \quad (5.136)$$

$$g_m r_o \gg 1, \quad (5.137)$$

then

$$A_{v,CB}|_{\omega \rightarrow 0} \approx g_m R_c. \quad (5.138)$$

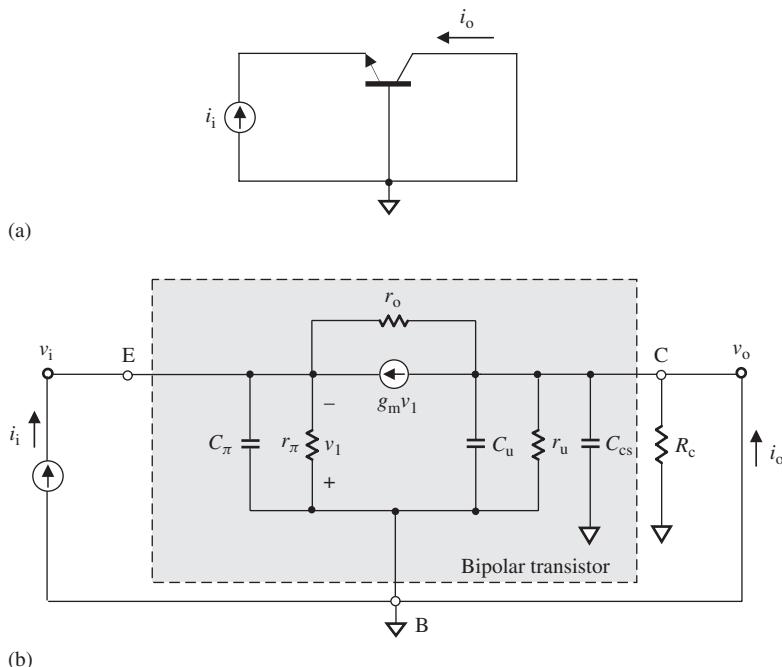
### 5.5.2 Short-Circuit Current Gain $\beta_{CG}$ and Frequency Response of a CB Device

Figure 5.16 is drawn for the calculation of the short-circuit current gain and frequency response of a bipolar transistor with a CB configuration, in which the following additional approximations have been made: the resistors  $r_b$  and  $r_c$ , are neglected because their values are much lower than the value of  $r_\pi$ , that is,

$$r_b, r_c \ll r_\pi. \quad (5.139)$$

At node C,

$$i_o = -g_m v_1. \quad (5.140)$$



**Figure 5.16.** Output short-circuited for calculating the current gain  $\beta$  and frequency response of a bipolar with CB configuration. (a) Output is short-circuited for calculation of current gain  $\beta$  of a CB device. (b) Calculation of current gain  $\beta$  and frequency response for a CB device.

At node E,

$$i_i + g_m v_1 + \frac{v_1}{C_\pi // r_\pi} = 0, \quad (5.141)$$

$$i_i = - \left( g_m + \frac{1}{r_\pi} + j\omega C_\pi \right) v_1. \quad (5.142)$$

From (5.140) and (5.142), we have

$$\beta_{CB}(j\omega) = \frac{i_o}{i_i} = \frac{g_m r_\pi}{g_m r_\pi + 1} \frac{1}{1 + j\omega C_\pi \frac{r_\pi}{g_m r_\pi + 1}}. \quad (5.143)$$

Note that in a CE device at low frequencies, as shown in (5.69),

$$\beta_{CE}(j\omega)|_{\omega \rightarrow 0} = \beta_o = g_m r_\pi = \frac{\alpha_o}{1 - \alpha_o}, \quad (5.144)$$

where

$\alpha_o$  = the base transport factor and

$\beta_o$  = the small-signal current gain of a CE device in low frequencies.

Then

$$\beta_{CB}(j\omega) = \frac{\beta_o}{\beta_o + 1} \frac{1}{1 + j\omega \frac{C_\pi}{g_m} \frac{\beta_o}{\beta_o + 1}} = \frac{\alpha_o}{1 + j\omega \frac{C_\pi}{g_m} \alpha_o}. \quad (5.145)$$

It can be seen that the current gain of a CB device is less than 1, that is,

$$|\beta_{CB}(j\omega)| < 1. \quad (5.146)$$

At low frequencies,

$$\beta_{CB}(j\omega)|_{\omega \rightarrow 0} \approx \alpha_o. \quad (5.147)$$

The short-circuit current gain of a CB device in low frequencies is close to but less than 1. Rewriting (5.145) in terms of (5.147), when  $\beta_{CB}(j\omega)$  is 3 dB down from the low-frequency value  $\beta_{CB}(j\omega)|_{\omega \rightarrow 0}$  that is,

$$|\beta_{CB}(j\omega)| = \frac{\beta_{CB}(j\omega)|_{\omega \rightarrow 0}}{\sqrt{2}} = \frac{\alpha_o}{\sqrt{2}}, \quad (5.148)$$

$$\omega = \omega_{\beta,CB} = \frac{g_m}{\alpha_o C_\pi}, \quad (5.149)$$

or

$$f = f_{\beta,CB} = \frac{1}{2\pi} \frac{g_m}{\alpha C_\pi}. \quad (5.150)$$

In expressions (5.149) or (5.150),  $\omega_\beta$  or  $f_\beta$  is determined by  $C_\pi$  only. This means that  $\omega_{\beta,CB}$  or  $f_{\beta,CB}$  in a CB device is much higher than that in a CE device as shown in expression (5.70) or (5.71), which is reproduced below:

$$\omega_{\beta,CE} = \frac{1}{\beta_0} \frac{g_m}{C_\pi + C_\mu} = \frac{\omega_T}{\beta_0}, \quad (5.67)$$

or

$$f_{\beta,CE} = \frac{1}{2\pi\beta_0} \frac{g_m}{C_\pi + C_\mu} = \frac{f_T}{\beta_0}. \quad (5.68)$$

It must be noted that it is impossible to define a transition frequency  $\omega_{T,CB}$  like that in expressions (5.64) to (5.65) for the case of a CE device, because the current gain of a CB device is always less than 1 as shown in (5.146). Reluctantly, we might have a special alternate definition of  $\omega_{T,CB}$  for the transition frequency of a CB device, such as

$$|\beta_{CB}(j\omega)| = 0.5, \quad (5.151)$$

for which an international agreement must be reached before it could be applied to engineering design.

Figure 5.17 shows an example of the frequency response of a CB device.

### 5.5.3 Input and Output Impedance of a CB Device

Figure 5.14(b) is an equivalent circuit of a CB device in terms of its small-signal hybrid- $\pi$  model, in which all the zero capacitors and the infinite inductor are ignored. As a matter of fact, it is a copy of Figure 5.7(b), but with the positions of the emitter and the base exchanged, so that the emitter is at the input position which usually is drawn in the left side of the figure while the output is drawn in the right side of the figure.

From Figure 5.14(b) it is found that the circuit analysis would be somewhat difficult because the dependent current source is replaced by two identical current sources, one from the collector to the base and the other from the base to the emitter as shown in Figure 5.18. The reason for this is that the currents fed into and drawn from the base are equal. The dependent current source from the base to the emitter is connected with the input resistor  $r_\pi$  and  $C_\pi$  in parallel and forms a special combined part, in which

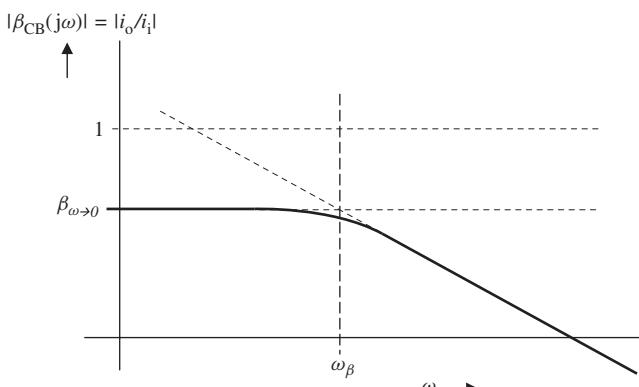
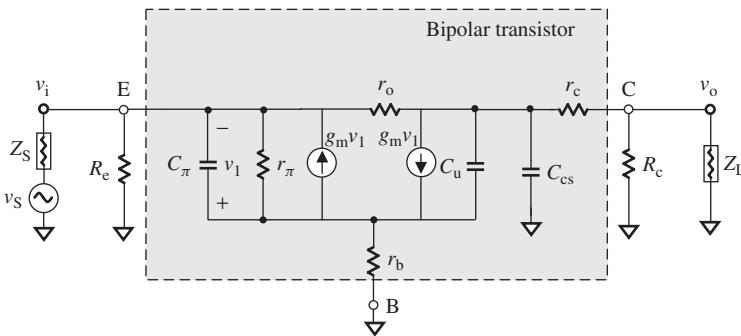


Figure 5.17. Frequency response of short-circuit current gain in a bipolar transistor with CB configuration.



**Figure 5.18.** Equivalent circuit of a bipolar transistor with CB configuration.

the current of the dependent current source depends on the voltage drop between two terminals of this special combined part.

Now we are going to apply Ohm's law to this special combined part. By applying Ohm's law, this dependent current source can be replaced by a resistor with a value of  $1/g_m$ , which is connected with either  $r_\pi$  and  $C_\pi$  in parallel. This resistor is called the *emitter resistor*  $R_e$ , that is,

$$r_e = r_\pi // \frac{1}{g_m} = \frac{1}{g_m + \frac{1}{r_\pi}} = \frac{\alpha_o}{g_m}, \quad (5.152)$$

$$\alpha_o = \frac{1}{1 + \frac{1}{\beta_o}} = \frac{\beta_o}{\beta_o + 1}, \quad (5.153)$$

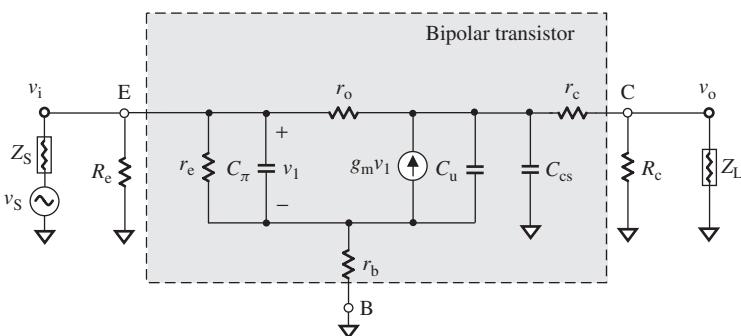
where

$\alpha_o$  = the base transport factor and

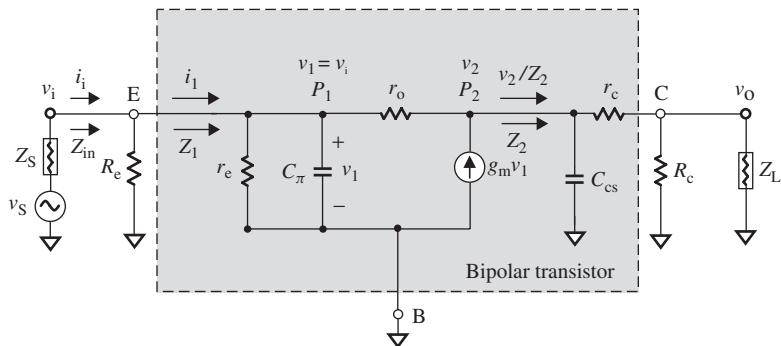
$\beta_o$  = the small-signal current gain in low frequencies.

Consequently, Figure 5.18 can be redrawn as Figure 5.19. It should be noted that the direction of the dependent current source in Figure 5.19 has been changed to upwards from downwards in Figure 5.15, while the direction of dependent voltage  $v_1$  is correspondingly upside down. The following analysis of the CB device will be conducted on the basis of Figure 5.19.

In the calculation of input impedance of the CB transistor, Figure 5.19, is simplified to Figure 5.20 by the additional approximations:  $C_\mu$  and  $r_b$  are neglected because the



**Figure 5.19.** Equivalent circuit for the calculation the input and output impedances of a CB device.



**Figure 5.20.** Equivalent circuit for the calculation of the input impedance of a CB device.

value of resistor  $r_b$  is much lower than that of the other resistors, that is,

$$r_b \ll r_\pi, r_o, \quad (5.154)$$

and the value of capacitor is much lower than the values of  $C_\pi$  and  $C_{cs}$ , that is,

$$C_\mu \ll C_{cs}, C_\pi. \quad (5.155)$$

KCL at node  $P_2$ :

$$\frac{v_2}{Z_2} + \frac{v_2 - v_1}{r_o} = g_m v_1. \quad (5.156)$$

KCL at node  $P_1$ :

$$i_1 = \frac{v_1}{r_e // C_\pi} + \frac{v_1 - v_2}{r_o}. \quad (5.157)$$

Solving for  $v_2$  from (5.156) and substituting it into (5.157) gives

$$v_2 = \frac{\left(g_m + \frac{1}{r_o}\right)}{\left(\frac{1}{Z_2} + \frac{1}{r_o}\right)} v_1, \quad (5.158)$$

$$\frac{i_1}{v_1} = \frac{1}{r_e // C_\pi} + \frac{1}{r_o} \left(1 - \frac{g_m + \frac{1}{r_o}}{\frac{1}{Z_2} + \frac{1}{r_o}}\right), \quad (5.159)$$

$$Z_1 = \frac{v_1}{i_1} = \frac{r_o + Z_2}{1 - g_m Z_2 + \frac{r_o + Z_2}{r_e // C_\pi}} = \frac{(r_e // C_\pi)(r_o + Z_2)}{r_e // C_\pi + r_o + [1 - g_m(r_e // C_\pi)]Z_2}, \quad (5.160)$$

where

$$Z_2 = C_{cs} // (r_c + R_c) = \frac{r_c + R_c}{1 + j\omega C_{cs}(r_c + R_c)}, \quad (5.161)$$

$$v_1 = v_i, \quad (5.162)$$

$$v_2 = \frac{r_c + R_c}{R_c} v_o, \quad (5.163)$$

$$r_e = \frac{\alpha_o}{g_m}, \quad (5.164)$$

$$r_e // C_\pi = \frac{\alpha_o}{g_m} // C_\pi = \frac{\alpha_o}{g_m + j\omega C_\pi \alpha_o} = \frac{1}{\frac{\beta_o + 1}{r_\pi} + j\omega C_\pi}. \quad (5.165)$$

Substituting  $r_e // C_\pi$  from (5.165) into (5.160), we have

$$\begin{aligned} Z_1 &= \frac{\frac{1}{\frac{\beta_o + 1}{r_\pi} + j\omega C_\pi} (r_o + Z_2)}{\frac{1}{\frac{\beta_o + 1}{r_\pi} + j\omega C_\pi} + r_o + \left(1 - \frac{g_m}{\frac{\beta_o + 1}{r_\pi} + j\omega C_\pi}\right) Z_2} \\ &= \frac{r_o + Z_2}{1 + \left(\frac{\beta_o + 1}{r_\pi} + j\omega C_\pi\right) r_o + \left(\frac{\beta_o + 1}{r_\pi} + j\omega C_\pi - g_m\right) Z_2}, \end{aligned} \quad (5.166)$$

$$\begin{aligned} Z_1 &= \frac{r_o + Z_2}{1 + (1 + \beta_o) \frac{r_o}{r_\pi} + j\omega C_\pi r_o + \left(\frac{1}{r_\pi} + j\omega C_\pi\right) Z_2} \\ &= \frac{r_\pi}{1 + \beta_o} \frac{1 + \frac{Z_2}{r_o}}{1 + \frac{r_\pi}{1 + \beta_o} \left[ \frac{1}{r_o} + j\omega C_\pi + \left(\frac{1}{r_\pi} + j\omega C_\pi\right) \frac{Z_2}{r_o} \right]}. \end{aligned} \quad (5.167)$$

Substituting  $Z_2$  from (5.161) into (5.167), we have

$$Z_1 = \frac{r_\pi}{1 + \beta_o} \frac{1 + \frac{1}{r_o} \frac{r_c + R_c}{1 + j\omega C_{cs}(r_c + R_c)}}{1 + \frac{r_\pi}{1 + \beta_o} \left[ \frac{1}{r_o} + j\omega C_\pi + \frac{1}{r_o} \left( \frac{1}{r_\pi} + j\omega C_\pi \right) \frac{r_c + R_c}{1 + j\omega C_{cs}(r_c + R_c)} \right]}. \quad (5.168)$$

At low frequencies

$$Z_1|_{\omega \rightarrow 0} = \frac{r_\pi}{1 + \beta_o} \frac{1 + \frac{r_c + R_c}{r_o}}{1 + \frac{(r_\pi + r_c + R_c)}{r_o(1 + \beta_o)}}, \quad (5.169)$$

$$Z_1|_{\omega \rightarrow 0} \approx \frac{r_\pi}{1 + \beta_o} \left( 1 + \frac{r_c + R_c}{r_o} \right), \quad (5.170)$$

if

$$r_o(1 + \beta_o) \gg (r_\pi + r_c + R_c). \quad (5.171)$$

Finally, the input impedance of the CB device is

$$\begin{aligned} Z_{\text{in}} &= \frac{v_i}{i_i} = R_e // Z_1 \\ &= R_e // \frac{r_\pi}{1 + \beta_o} \frac{1 + \frac{1}{r_o} \frac{r_c + R_c}{1 + j\omega C_{cs}(r_c + R_c)}}{1 + \frac{r_\pi}{1 + \beta_o} \left[ \frac{1}{r_o} + j\omega C_\pi + \frac{1}{r_o} \left( \frac{1}{r_\pi} + j\omega C_\pi \right) \frac{r_c + R_c}{1 + j\omega C_{cs}(r_c + R_c)} \right]}. \end{aligned} \quad (5.172)$$

It can be seen that the input impedance of a CB device is about  $\beta_o$  times lower than  $r_\pi$ . This is one of the special features of the CB device.

It should be noted that in Figures 5.18–5.20 the resistor  $R_e$  might not be needed in some applications, while the resistor  $R_c$  is, however, an indispensable part because it leads the DC power supply to the device. For instance, in a CE–CB cascade amplifier, the implementation of the resistor  $R_e$  is unnecessary, so that in such a case we have

$$Z_{\text{in}}|_{R_e \rightarrow \infty} = Z_1 = \frac{r_\pi}{1 + \beta_o} \frac{1 + \frac{1}{r_o} \frac{r_c + R_c}{1 + j\omega C_{cs}(r_c + R_c)}}{1 + \frac{r_\pi}{1 + \beta_o} \left[ \frac{1}{r_o} + j\omega C_\pi + \frac{1}{r_o} \left( \frac{1}{r_\pi} + j\omega C_\pi \right) \frac{r_c + R_c}{1 + j\omega C_{cs}(r_c + R_c)} \right]}, \quad (5.173)$$

which corresponds to the case of

$$R_e \rightarrow \infty. \quad (5.174)$$

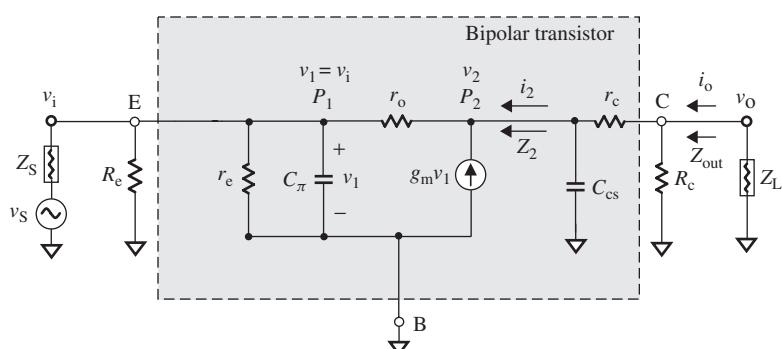
In the cases of low frequencies

$$Z_{\text{in}}|_{\omega \rightarrow 0, R_e \rightarrow \infty} = r_\pi \frac{r_o + r_c + R_c}{(1 + \beta_o)r_o + r_\pi + r_c + R_c} = r_\pi \frac{1}{1 + (1 + g_m r_o) \frac{r_\pi}{r_o + r_c + R_c}}. \quad (5.175)$$

Now let us calculate the output impedance. Figure 5.21 shows the equivalent model for calculation of the output impedance of a CB device.

KCL at node  $P_1$ :

$$\frac{v_1}{R_e // C_\pi} + \frac{v_1 - v_2}{r_o} = 0. \quad (5.176)$$



**Figure 5.21.** Equivalent circuit for the calculation of the output impedance of a CB device.

KCL at node P<sub>2</sub>:

$$i_2 = -g_m v_1 + \frac{v_2 - v_1}{r_o} = \frac{v_2}{r_o} - \left( g_m + \frac{1}{r_o} \right) v_1. \quad (5.177)$$

Solving for  $v_1$  from (5.192) and substituting it into (5.193) gives

$$v_1 = \frac{1/r_o}{\left( \frac{1}{R_e} + \frac{1}{r_e // C_\pi} + \frac{1}{r_o} \right)} v_2 = \frac{R_e(r_e // C_\pi)}{R_e r_o + (R_e + r_o)(r_e // C_\pi)} v_2, \quad (5.178)$$

$$\frac{i_2}{v_2} = \frac{1}{r_o} - \frac{\left( g_m + \frac{1}{r_o} \right) R_e(r_e // C_\pi)}{R_e r_o + (R_e + r_o)(r_e // C_\pi)}, \quad (5.179)$$

$$\frac{i_2}{v_2} = \frac{1}{r_o} \left[ 1 - \frac{(1 + g_m r_o) R_e(r_e // C_\pi)}{R_e r_o + (R_e + r_o)(r_e // C_\pi)} \right] = \frac{1}{r_o} - \frac{1}{r_o} \frac{\left( \frac{1}{r_o} + g_m \right) R_e(r_e // C_\pi)}{R_e + \left( \frac{R_e}{r_o} + 1 \right) (r_e // C_\pi)}, \quad (5.180)$$

$$\frac{i_2}{v_2} = \frac{1}{r_o} \frac{R_e + (1 - g_m R_e)(r_e // C_\pi)}{R_e + \left( \frac{R_e}{r_o} + 1 \right) (r_e // C_\pi)}, \quad (5.181)$$

$$Z_2 = \frac{v_2}{i_2} = r_o \frac{R_e + \left( \frac{R_e}{r_o} + 1 \right) (r_e // C_\pi)}{R_e + (1 - g_m R_e)(r_e // C_\pi)} = r_o \frac{\frac{1}{(r_e // C_\pi)} + \left( \frac{1}{r_o} + \frac{1}{R_e} \right)}{\frac{1}{(r_e // C_\pi)} + \frac{1}{R_e} - g_m}. \quad (5.182)$$

Note that

$$r_e = \frac{\alpha_o}{g_m}, \quad (5.183)$$

$$r_e // C_\pi = \frac{\alpha_o}{g_m} // C_\pi = \frac{\alpha_o}{g_m + j\omega C_\pi \alpha_o} = \frac{1}{\frac{\beta_o + 1}{r_\pi} + j\omega C_\pi}. \quad (5.184)$$

Substituting  $r_e // C_\pi$  from (3.184) into (3.182), we have

$$Z_2 = r_o \frac{\frac{1}{\frac{\alpha_o}{g_m + j\omega C_\pi \alpha_o}} + \left( \frac{1}{r_o} + \frac{1}{R_e} \right)}{\frac{1}{\frac{\alpha_o}{g_m + j\omega C_\pi \alpha_o}} + \frac{1}{R_e} - g_m} = r_o \frac{1 + \left( \frac{1}{r_o} + \frac{1}{R_e} \right) \frac{\alpha_o}{g_m + j\omega C_\pi \alpha_o}}{1 + \left( \frac{1}{R_e} - g_m \right) \frac{\alpha_o}{g_m + j\omega C_\pi \alpha_o}}, \quad (5.185)$$

$$Z_2 = r_o r_\pi \frac{\left[ 1 + R_e \left( \frac{1}{r_o} + \frac{\beta_o + 1}{r_\pi} + j\omega C_\pi \right) \right]}{(r_\pi + R_e + j\omega C_\pi R_e r_\pi)}. \quad (5.186)$$

Finally, the output impedance is

$$Z_{\text{out}} = [Z_2 // C_{\text{cs}} // (r_c + R_c)] \frac{R_c}{r_c + R_c} = \frac{R_c}{r_c + R_c} \frac{(r_c + R_c)}{[1 + j\omega C_{\text{cs}}(r_c + R_c)] + \frac{(r_c + R_c)}{Z_2}}, \quad (5.187)$$

$$Z_{\text{out}} = R_c \frac{1}{1 + j\omega C_{\text{cs}}(r_c + R_c) + (r_c + R_c) \frac{(r_\pi + R_e + j\omega C_\pi R_e r_\pi)}{r_o r_\pi + R_e [r_\pi + r_o (\beta_o + 1) + j\omega C_\pi r_\pi r_o]}}. \quad (5.188)$$

In the low-frequency cases,

$$Z_{\text{out}}|_{\omega \rightarrow 0} = R_c \frac{1}{1 + (r_c + R_c) \frac{(r_\pi + R_e)}{r_o r_\pi + R_e [r_\pi + (\beta_o + 1) r_o]}}. \quad (5.189)$$

In cases without the resistor  $R_e$ , that is,  $R_e \rightarrow \infty$ ,

$$Z_{\text{out}}|_{R_e \rightarrow \infty} = R_c \frac{1}{1 + j\omega C_{\text{cs}}(r_c + R_c) + \frac{(r_c + R_c)(1 + j\omega C_\pi r_\pi)}{r_\pi + (1 + \beta_o + j\omega C_\pi r_\pi) r_o}}. \quad (5.190)$$

In low-frequency cases,

$$Z_{\text{out}}|_{\omega \rightarrow 0, R_e \rightarrow \infty} = R_c \frac{1}{1 + \frac{r_c + R_c}{r_\pi + (1 + \beta_o) r_o}}, \quad (5.191)$$

## 5.6 BIPOLAR TRANSISTOR WITH CC (COMMON COLLECTOR) CONFIGURATION

Figure 5.22 shows the equivalent circuit of a bipolar transistor with a CC configuration, in which the collector is AC-grounded, the input port is the base and the output port is the emitter of the bipolar transistor. The notation “CC” originates from the fact that the collector is a common ground terminal of the input (base) port and the output (emitter) port.

In Figure 5.22(a), two zero capacitors which approach zero impedance at the operating frequencies function as DC blocking capacitors, and are symbolized as  $C_{\text{zero}}$ . The infinite inductor functions as an RF choke, which approaches infinite impedance at operating frequencies and is therefore denoted  $L_{\text{infinite}}$ . The emitter resistor  $R_e$  is connected from the emitter to the ground.

Figure 5.22(b) is the equivalent circuit of a CC device in terms of its small-signal hybrid- $\pi$  model, in which all the zero capacitors and the infinite inductor are ignored.

### 5.6.1 Open-Circuit Voltage Gain $A_{v,CC}$ of a CC Device

The voltage gain  $A_v$  of a CC device can be derived with the additional approximations:

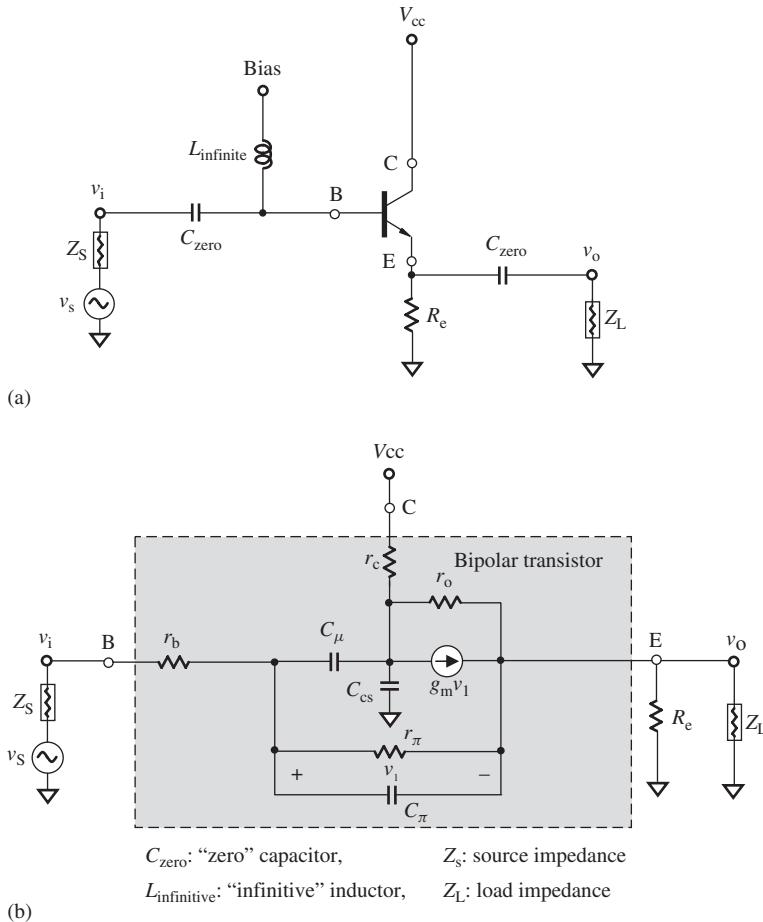
the resistor  $r_c$  is neglected because its value is much lower than the value of  $R_c$ , that is,

$$r_c \ll R_c, \quad (5.192)$$

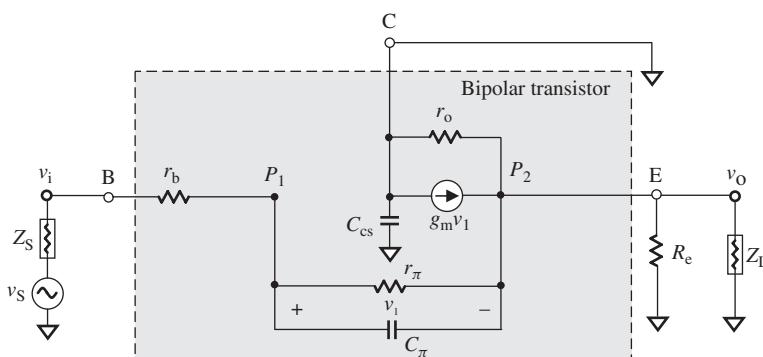
and the value of the capacitor  $C_\pi$  is much higher than the values of  $C_\mu$  and  $C_{cs}$ , that is,

$$C_\pi \gg C_\mu, C_{cs}. \quad (5.193)$$

Then, the model of bipolar transistor with the CC configuration shown in Figure 5.22(b) can be redrawn as Figure 5.23.



**Figure 5.22.** Bipolar transistor with CC configuration. (a) Schematic of a bipolar transistor with CC configuration. (b) Model of bipolar transistor with CC configuration.



**Figure 5.23.** Equivalent model for the calculation of the open-circuit voltage gain of a CC device.

KCL at node P<sub>2</sub>:

$$\frac{v_i - v_o}{r_b + r_\pi // C_\pi} + g_m v_1 - \frac{v_o}{R_e // Z_L} - \frac{v_o}{r_o} = 0. \quad (5.194)$$

Note that the terminal V<sub>cc</sub> is AC-grounded and

$$g_m = \frac{\beta_o}{r_\pi}, \quad (5.195)$$

$$v_1 = \frac{r_\pi // C_\pi}{r_b + r_\pi // C_\pi} (v_i - v_o). \quad (5.196)$$

Substituting  $g_m$  in (5.195) and  $v_1$  in (5.196) into (5.194), we have

$$\frac{v_i - v_o}{r_b + r_\pi // C_\pi} + \frac{\beta_o}{r_\pi} \frac{r_\pi // C_\pi}{r_b + r_\pi // C_\pi} (v_i - v_o) - \frac{v_o}{R_e // Z_L} - \frac{v_o}{r_o} = 0. \quad (5.197)$$

Finally the voltage gain is

$$A_{v,CC} = \frac{v_o}{v_i} = \frac{\left( \frac{1}{r_b + r_\pi // C_\pi} + \frac{\beta_o}{r_\pi} \frac{r_\pi // C_\pi}{r_b + r_\pi // C_\pi} \right)}{\left( \frac{1}{r_b + r_\pi // C_\pi} + \frac{\beta_o}{r_\pi} \frac{r_\pi // C_\pi}{r_b + r_\pi // C_\pi} \right) + \left( \frac{1}{R_e // Z_L} + \frac{1}{r_o} \right)}. \quad (5.198)$$

Note that

$$\frac{1}{r_b + r_\pi // C_\pi} + \frac{\beta_o}{r_\pi} \frac{r_\pi // C_\pi}{r_b + r_\pi // C_\pi} = \frac{\beta_o + 1 + j\omega C_\pi r_b}{r_\pi \left[ 1 + \frac{r_b}{r_\pi} (1 + j\omega C_\pi r_b) \right]}, \quad (5.199)$$

$$\frac{1}{R_e // Z_L} + \frac{1}{r_o} = \frac{r_o (R_e + Z_L) + R_e Z_L}{r_o R_e Z_L}, \quad (5.200)$$

$$A_{v,CC} = \frac{v_o}{v_i} = \frac{\frac{\beta_o + 1 + j\omega C_\pi r_b}{r_\pi \left[ 1 + \frac{r_b}{r_\pi} (1 + j\omega C_\pi r_b) \right]}}{\frac{\beta_o + 1 + j\omega C_\pi r_b}{r_\pi \left[ 1 + \frac{r_b}{r_\pi} (1 + j\omega C_\pi r_b) \right]} + \frac{R_e Z_L + Z_L r_b + r R_e}{R_e Z_L r_b}}, \quad (5.201)$$

$$A_{v,CC} = \frac{v_o}{v_i} = \frac{1}{1 + \frac{(R_e Z_L + Z_L r_b + r R_e)(r_\pi + r_b + j\omega C_\pi r_b r_b)}{R_e Z_L r_b (\beta_o + 1 + j\omega C_\pi r_b)}}. \quad (5.202)$$

The open-circuit voltage gain  $A_{v,CC}$  is in the case when

$$Z_L \rightarrow \infty. \quad (5.203)$$

Then,

$$A_{v,CC} = \frac{v_o}{v_i} = \frac{1}{1 + \frac{(R_e + r_b)(r_\pi + r_b + j\omega C_\pi r_b r_b)}{R_e r_b (\beta_o + 1 + j\omega C_\pi r_b)}}. \quad (5.204)$$

In low-frequency cases,

$$A_{v,CC}|_{\omega \rightarrow 0} = \frac{1}{1 + \frac{(R_e + r_b)(r_\pi + r_b)}{R_e r_b (\beta_o + 1)}}. \quad (5.205)$$

Usually, we have

$$r_\pi \gg r_b, \quad \text{and} \quad (5.206)$$

$$\beta_o \gg 1, \quad (5.207)$$

so

$$A_{v,CC}|_{\omega \rightarrow 0} \approx \frac{1}{1 + \frac{1}{g_m} \left( \frac{1}{R_e} + \frac{1}{r_o} \right)}. \quad (5.208)$$

### 5.6.2 Short-Circuit Current Gain $\beta_{CC}$ and Frequency Response of the Bipolar Transistor with CC Configuration

Figure 5.24 is drawn for the calculation of the short-circuit current gain and frequency response of a bipolar transistor with a CC configuration, in which the resistor  $r_c$  and  $C_\mu$  are neglected because

$$r_c \ll R_c, \quad (5.209)$$

$$C_\pi \gg C_{cs}, C_\mu. \quad (5.210)$$

Also note that  $C_{cs}$  is shorted to the ground because in a CC device (emitter follower) the terminal  $C$  is grounded. Then,

KCL at node E:

$$i_o = g_m v_1 + i_i. \quad (5.211)$$

Note that

$$v_1 = i_i C_\pi // r_\pi, \quad (5.212)$$

so

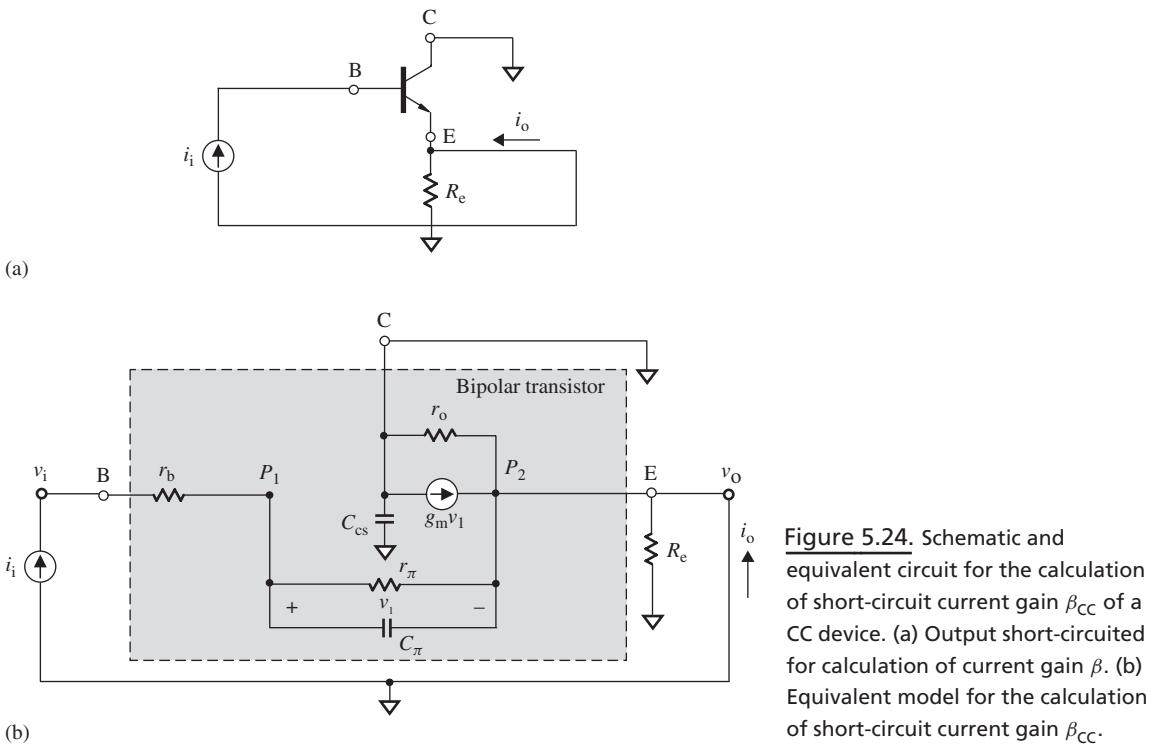
$$i_o = (g_m C_\pi // r_\pi + 1) i_i = \beta_o \frac{\left(1 + \frac{1+j\omega C_\pi r_\pi}{\beta_o}\right)}{(1 + j\omega C_\pi r_\pi)} i_i, \quad (5.213)$$

$$\beta_{CC}(j\omega) = \frac{i_o}{i_i} = \beta_o \frac{\left(1 + \frac{1+j\omega C_\pi r_\pi}{\beta_o}\right)}{(1 + j\omega C_\pi r_\pi)} = \frac{1 + \beta_o + j\omega C_\pi r_\pi}{1 + j\omega C_\pi r_\pi}. \quad (5.214)$$

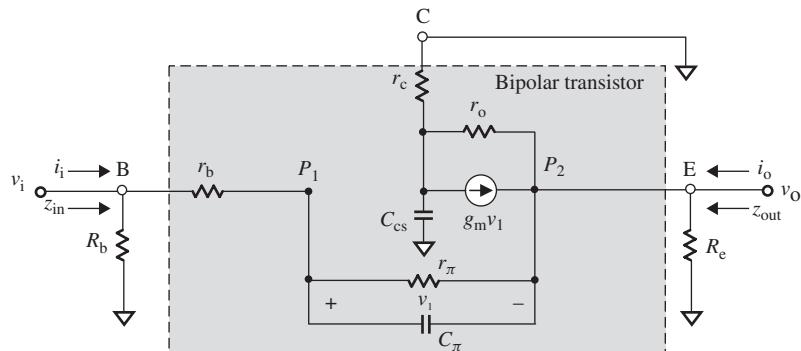
At low frequencies,

$$\beta_{CC}(j\omega)|_{\omega \rightarrow 0} = 1 + \beta_o. \quad (5.215)$$

The short-circuit current gain of a CC device at low frequencies is  $(1 + \beta_o)$ , which is in the same order as that in a CE device.



**Figure 5.24.** Schematic and equivalent circuit for the calculation of short-circuit current gain  $\beta_{CC}$  of a CC device. (a) Output short-circuited for calculation of current gain  $\beta$ . (b) Equivalent model for the calculation of short-circuit current gain  $\beta_{CC}$ .



**Figure 5.25.** Equivalent circuit for the calculation of the input and output impedance of a CC device.

### 5.6.3 Input and Output Impedance of a CC Device

In the calculation of input impedance of a CC transistor, Figure 5.25 is simplified by the additional approximation: the capacitor  $C_\mu$  is neglected because

$$C_\mu < C_{cs} \ll C_\pi. \quad (5.216)$$

Then, from the input to output node,

$$v_i = i_i(r_b + r_\pi // C_\pi) + v_o. \quad (5.217)$$

KCL at node P<sub>2</sub>:

$$v_o = (i_i + g_m v_1) [R_e // (r_o + r_c // C_{cs})]. \quad (5.218)$$

Ohm's law at  $r_\pi // C_\pi$  gives

$$v_1 = i_i (r_\pi // C_\pi). \quad (5.219)$$

From the input to output node,

$$v_i = i_i (r_b + r_\pi // C_\pi) + v_o. \quad (5.220)$$

KCL at node P<sub>2</sub>:

$$v_o = (i_i + g_m v_1) [R_e // (r_o + r_c // C_{cs})]. \quad (5.221)$$

Ohm's law at  $r_\pi // C_\pi$  gives

$$v_1 = i_i (r_\pi // C_\pi), \quad (5.222)$$

Substituting (5.222) into (5.221),

$$v_o = i_i [1 + g_m (r_\pi // C_\pi)] [R_e // (r_o + r_c // C_{cs})]. \quad (5.223)$$

Replacing  $v_o$  in (5.220) by (5.223),

$$v_i = i_i (r_b + r_\pi // C_\pi) + i_i [1 + g_m (r_\pi // C_\pi)] [R_e // (r_o + r_c // C_{cs})]. \quad (5.224)$$

Then,

$$Z_{in} = \frac{v_i}{i_i} = (r_b + r_\pi // C_\pi) + [1 + g_m (r_\pi // C_\pi)] [R_e // (r_o + r_c // C_{cs})]. \quad (5.225)$$

Note that

$$r_\pi // C_\pi = \frac{1}{1 + jC_\pi \omega r_\pi} r_\pi \quad (5.226)$$

$$r_b + r_\pi // C_\pi = \frac{r_b + r_\pi + jC_\pi \omega r_b r_\pi}{1 + jC_\pi \omega r_\pi}, \quad (5.227)$$

$$r_o + r_c // C_{cs} = \frac{r_o + r_c + jC_{cs} \omega r_o r_c}{1 + jC_{cs} \omega r_c} \quad (5.228)$$

$$R_e // (r_o + r_c // C_{cs}) = \frac{r_o + r_c + jC_{cs} \omega r_o r_c}{R_e + r_o + r_c + jC_{cs} \omega r_c (r_o + R_e)} R_e. \quad (5.229)$$

Then,

$$Z_{in} = \frac{r_\pi + r_b (1 + jC_\pi \omega r_\pi) + (1 + \beta_o + jC_\pi \omega r_\pi) \frac{r_o + r_c + jC_{cs} \omega r_o r_c}{R_e + r_o + r_c + jC_{cs} \omega r_c (r_o + R_e)} R_e}{1 + jC_\pi \omega r_\pi}. \quad (5.230)$$

In low-frequency cases,

$$Z_{\text{in}}|_{\omega \rightarrow 0} = r_\pi + r_b + (1 + \beta_o) \frac{r_o + r_c}{R_e + r_o + r_c} R_e. \quad (5.231)$$

A special feature of a CC device is that the input resistance is equal to  $r_\pi$  plus  $(1 + \beta_o)$  times the incremental resistance connected from the emitter to the ground.

Now let us discuss the output impedance of a CC device.

The output current consists of three portions when an output voltage is applied to the terminal E. The first portion is the current flowing through  $r_o$  and  $r_c$  and  $C_{cs}$  to the AC ground due to the output voltage. The second portion is the current generated by the device  $g_m v_1$ , in which  $v_1$  is the voltage drop on  $r_\pi$  and  $C_\pi$  due to the output voltage. The third portion is the current flowing through  $r_\pi$ ,  $r_b$ , and  $R_b$  to the ground due to the output voltage. We then have

$$i_o = \frac{v_o}{r_o + (r_c // C_{cs})} + g_m \frac{(r_\pi // C_\pi)v_o}{R_b + r_b + (r_\pi // C_\pi)} + \frac{v_o}{R_b + r_b + (r_\pi // C_\pi)}, \quad (5.232)$$

$$i_o = \frac{1 + j\omega C_{cs} r_c}{r_c + r_o(1 + j\omega C_{cs} r_c)} v_o + \frac{1 + \beta_o + j\omega C_\pi r_\pi}{r_\pi + (R_b + r_b)(1 + j\omega C_\pi r_\pi)} v_o, \quad (5.233)$$

$$\frac{1}{Z_{\text{out}}} = \frac{i_o}{v_o} = \frac{1 + j\omega C_{cs} r_c}{r_c + r_o(1 + j\omega C_{cs} r_c)} + \frac{1 + \beta_o + j\omega C_\pi r_\pi}{r_\pi + (R_b + r_b)(1 + j\omega C_\pi r_\pi)}, \quad (5.234)$$

$$Z_{\text{out}} = \frac{r_\pi + (R_b + r_b)(1 + j\omega C_\pi r_\pi)}{(1 + \beta_o + j\omega C_\pi r_\pi) + (1 + j\omega C_{cs} r_c) \frac{r_\pi + (R_b + r_b)(1 + j\omega C_\pi r_\pi)}{r_c + r_o(1 + j\omega C_{cs} r_c)}}. \quad (5.235)$$

At a low-frequency range, the capacitors  $C_\pi$  and  $C_{cs}$  could be neglected; thus, we have

$$Z_{\text{in}}|_{\omega \rightarrow 0} \approx r_\pi + r_b + (1 + \beta_o) \frac{1}{1 + \frac{R_e}{r_o + r_c}} R_e \approx r_\pi + r_b + \beta_o R_e, \quad (5.236)$$

$$Z_{\text{out}}|_{\omega \rightarrow 0} \approx \frac{(r_\pi + R_b + r_b)}{(1 + \beta_o) + \frac{(r_\pi + R_b + r_b)}{(r_c + r_o)}} \approx \frac{1}{g_m} + \frac{R_b + r_b}{\beta_o}. \quad (5.237)$$

It can be seen that the input impedance  $Z_{\text{in}}$  is dependent on the output resistors and load impedance  $R_e$ . Similarly, the output impedance  $Z_{\text{out}}$  is dependent on the input resistors and source impedance  $R_b$ .

A further approximation is that if

$$r_o \gg Z_L, \quad (5.238)$$

then

$$Z_{\text{in}} \approx r_\pi + r_b + (1 + \beta) Z_L, \quad (5.239)$$

and if

$$r_\pi \gg Z_s, \quad (5.240)$$

then

$$Z_{\text{out}} \approx \frac{1}{1 + \beta + \frac{r_{\pi}}{r_{\text{c}} + r_{\text{o}}}} r_{\pi}. \quad (5.241)$$

Furthermore, if

$$r_{\text{o}} \gg r_{\pi}, \quad (5.242)$$

then

$$z_{\text{out}} \approx \frac{1}{g_{\text{m}}} \quad (5.243)$$

## 5.7 SMALL-SIGNAL MODEL OF A MOSFET

Figures 5.26 and 5.27 show the DC characteristics of a MOSFET.

The DC characteristics of a MOSFET can be expressed as follows:  
At the triode region 1:

$$I_{\text{d}} = \mu_{\text{n}} C_{\text{ox}} \frac{W}{L} (V_{\text{gs}} - V_{\text{th}}) V_{\text{ds}}, \quad (5.244)$$

where

$I_{\text{d}}$  = the drain current,

$g_{\text{m}}$  = the transconductance of the MOSFET,

$W$  = the width of the MOSFET,

$L$  = the length of the MOSFET,

$V_{\text{gs}}$  = the gate–source voltage for an  $n$ -channel MOSFET,

$V_{\text{th}}$  = the threshold voltage for an  $n$ -channel MOSFET, that is, the minimum gate-to channel voltage needed for  $n$  carriers in the channel to exist,

$V_{\text{ds}}$  = the drain–source voltage for an  $n$  channel MOSFET,

$\mu_{\text{n}}$  = the channel mobility, typically  $700 \text{ cm}^2/\text{V s}$ ,

$C_{\text{ox}}$  = the capacitance per unit area of the gate oxide

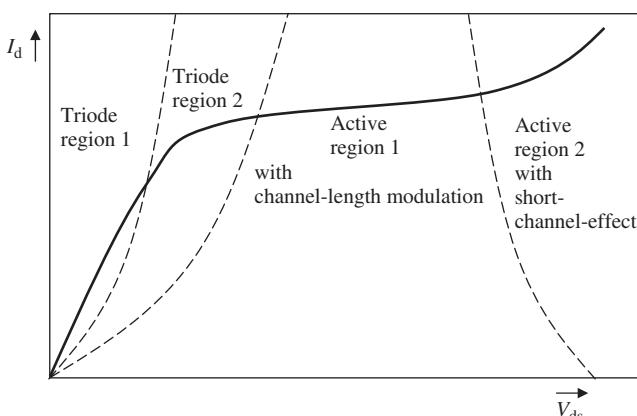
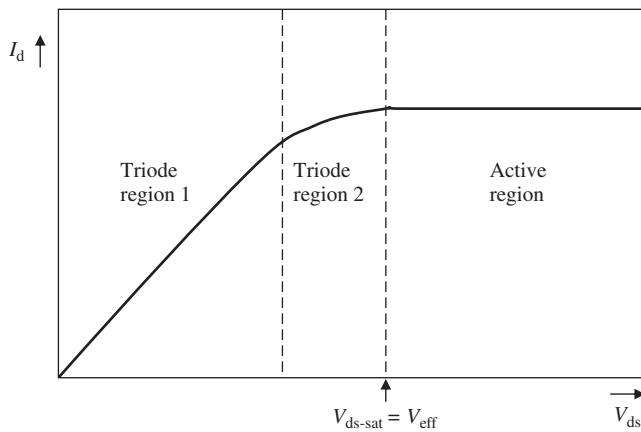


Figure 5.26. DC characteristics of a MOSFET ( $I_{\text{d}}$  vs  $V_{\text{ds}}$ ) with short-channel effect.



**Figure 5.27.** DC characteristics of a MOSFET ( $I_d$  vs  $V_{ds}$ ) without short-channel effect.

and

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}, \quad (5.245)$$

where

$t_{ox}$  = the thickness of the gate oxide, and  
 $\epsilon_{ox}$  = the permittivity of oxide.

For example,  
if

$$t_{ox} = 400 \text{ \AA}^\circ = 0.4 \times 10^{-5} \text{ cm}, \quad (5.246)$$

$$\mu_n = 700 \text{ cm}^2/V \text{ sec}, \quad (5.247)$$

$$\epsilon_{ox} = 3.45 \times 10^{-13} \text{ F/cm}, \quad (5.248)$$

then

$$C_{ox} = 0.86 \text{ fF}/\mu^2, \quad (5.249)$$

$$\mu_n C_{ox} = 60.2 \mu\text{A}/\text{V}^2. \quad (5.250)$$

At triode region 2:

$$I_d = \mu_n C_{ox} \frac{W}{L} \left[ (V_{gs} - V_{th}) V_{ds} - \frac{V_{ds}^2}{2} \right]. \quad (5.251)$$

As  $V_{ds}$  increases,  $I_d$  increases until the drain end of the channel becomes pinched-off, and then levels off. This pinchoff occurs for

$$V_{dg} = -V_{th}, \quad (5.252)$$

or approximately,

$$V_{ds} = V_{eff} = V_{gs} - V_{th}, \quad (5.253)$$

where  $V_{eff}$  = the effective gate-source voltage.

Therefore, the active region equation can be obtained by substituting (5.253) into (5.251), that is,

at the active region:

$$I_d = \frac{\mu_n C_{ox}}{2} \frac{W}{L} (V_{gs} - V_{th})^2 = I_{d-sat}. \quad (5.254)$$

For  $V_{ds} > V_{eff}$ , the current stays constant at the value given by equation 5.283 and is denoted by  $I_{d-sat}$ . This implies that the drain current is independent of the drain-source voltage  $V_{ds}$ .

In the active region,

$$g_m = \frac{\partial I_d}{\partial V_{gs}} = \mu_n C_{ox} \frac{W}{L} (V_{gs} - V_{th}) = \mu_n C_{ox} \frac{W}{L} V_{eff}. \quad (5.255)$$

Replacing  $V_{eff}$  by  $I_d$ , from (5.282) and (5.283) we have

$$V_{eff} = \sqrt{\frac{2I_d}{\mu_n C_{ox} (W/L)}}. \quad (5.256)$$

Consequently,

$$g_m = \sqrt{2\mu_n C_{ox} \frac{W}{L} I_d} = \frac{2I_d}{V_{eff}} = \frac{2I_d}{V_{gs} - V_{th}}. \quad (5.257)$$

It must be noted that the independence between  $I_d$  and  $V_{ds}$ , expressed by (5.254), is only true to a first-order approximation. As a matter of fact, the channel is shrunk as  $V_{ds}$  is increased. A pinched-off region with very little charge exists between the drain and the channel. The voltage at the end of the channel closest to the drain is fixed at  $V_{gs} - V_{th} = V_{eff}$ . The voltage difference between the drain and the near end of the channel lies across a short depletion region often called the *pinch-off region*. As  $V_{ds}$  becomes larger than  $V_{eff}$ , this depletion region surrounding the drain junction increases its width in a square-root relationship with respect to  $V_{ds}$ . This increase in the width of the depletion region surrounding the drain junction decreases the effective channel length. In turn, this decrease in effective channel length increases the drain current, resulting in what is commonly referred to as *channel-length modulation*. Then, expression (5.254) must be modified:

$$I_d = \frac{\mu_n C_{ox}}{2} \frac{W}{L} (V_{gs} - V_{th})^2 [1 + \lambda(V_{ds} - V_{eff})], \quad (5.258)$$

$$\lambda = \frac{k_{ds}}{2L\sqrt{V_{ds} - V_{eff} + \Phi_o}}, \quad (5.259)$$

$$k_{ds} = \sqrt{\frac{2K_S \varepsilon_0}{qN_A}}, \quad (5.260)$$

$$\Phi_o = V_T \ln \frac{N_A N_D}{n_i^2}, \quad (5.261)$$

where

- $\lambda$  = the output impedance constant (in units of  $V^{-1}$ ),
- $\Phi_o$  = the built-in voltage of an open-circuit PN junction (in units of V),
- $V_T = kT/q$  = the thermal voltage (at room temperature it is approximately 26 mV),
- $K_S$  = the relative permittivity of silicon (typically it equals to 11.8),
- $\epsilon_o$  = the permittivity of free space equals to  $8.854 \times 10^{-12} \text{ F/m}$ ,
- $q$  = the charge of an electron ( $1.602 \times 10^{-19} \text{ C}$ ),
- $N_A$  = the acceptor concentration (in units of holes/ $\text{m}^3$ ),
- $N_D$  = the donor concentration (in units of electrons/ $\text{m}^3$ ), and
- $n_i$  = the carrier concentration in intrinsic silicon (in unit of carriers/ $\text{m}^3$ ).

Furthermore, the expression (5.283) is accurate only in the cases without velocity saturation. Velocity saturation occurs when  $V_{ds}$  is so large that the carrier speed no longer increases as the electric field increases. This second-order effect is often called the *short-channel effect*, in which  $I_d$  increases more than that calculated from (5.283) as  $V_{ds}$  is increased. Figure 5.24 shows the regions affected by the short-channel effect.

In the actual device construction, the voltage of the source is sometimes different from that of the substrate (also known as the *bulk*). In other words, a voltage  $V_{sb}$  exists between the source of the device and the substrate, or in other words, it is nonzero. The threshold voltage of an *n*-channel transistor is now given by

$$V_{tn} = V_{tno} + \gamma(\sqrt{V_{sb} + 2|\Phi_F|} - \sqrt{2|\Phi_F|}), \quad (5.262)$$

$$\gamma = \sqrt{\frac{2qN_A K_S \epsilon_o}{C_{ox}}}, \quad (5.263)$$

$$\Phi_F = -\frac{kT}{q} \ln \left( \frac{N_A}{n_i} \right), \quad (5.264)$$

where

$V_{tno}$  = the threshold voltage with zero  $V_{sb}$  and

$\Phi_F$  = the built-in voltage of an open-circuit PN junction (in units of V).

It causes the body effect. For a small signal in the active region, this corresponds to having another current source  $v_s$ , and its corresponding transconductance  $g_s$ , which is

$$g_s = \frac{\gamma g_m}{2\sqrt{V_{sb} + 2|\Phi_F|}}, \quad (5.265)$$

$$V_s = V_{sb}, \quad (5.266)$$

$$\frac{1}{r_{ds}} = \lambda I_d. \quad (5.267)$$

Figure 5.28 shows the small-signal model of a MOSFET when the body effect is considered. This is most commonly used hybrid- $\pi$  model, where

$S$  = the source of the MOSFET,

$G$  = the gate of the MOSFET,

$D$  = the drain of the MOSFET,

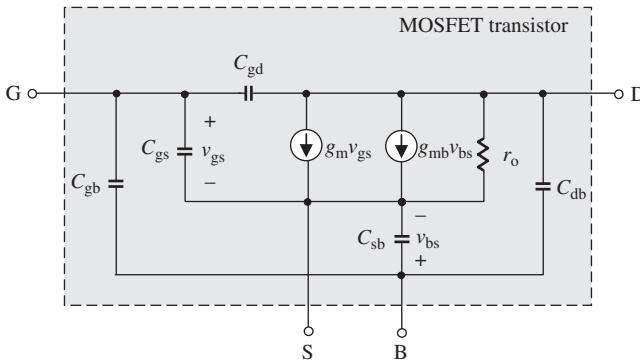


Figure 5.28. Small-signal model of a MOSFET.

$C_{gs}$  = the gate–source capacitance,  
 $C_{gd}$  = the gate–drain capacitance,  
 $C_{db}$  = the drain–body depletion capacitance,  
 $C_{gb}$  = the gate–body capacitance,  
 $C_{sb}$  = the source–body depletion capacitance,  
 $r_o$  = is the output resistance,  
 $g_m$  = the top-gate transconductance, and  
 $g_{mb}$  = the body effect transconductance.

## 5.8 SIMILARITY BETWEEN A BIPOLAR TRANSISTOR AND A MOSFET

A similarity exists between the small-signal device models of bipolar and MOSFET. By means of this similarity, the conclusions or rules acquired from the bipolar device may be extended to the MOSFET device so as to create a shortcut to the derivation of the theory or formula. However, a discussion of the MOSFET model will be conducted before MOSFET models become comparable with bipolar transistor models.

### 5.8.1 Simplified Model of CS Device

Looking at the MOSFET small-signal model as shown in Figure 5.28, there are three small capacitors,  $C_{gb}$ ,  $C_{sb}$ , and  $C_{db}$ , which are connected between G (gate), S (source), D (drain), and B (substrate) with a T configuration as sketched in Figure 5.29(a). In order to make the MOSFET model similar to the bipolar model, it is best to divide the capacitor  $C_{sb}$  into two capacitors,  $C_{sb,g}$  and  $C_{sb,d}$ , so that it forms two piggy-back  $\Gamma$  configurations

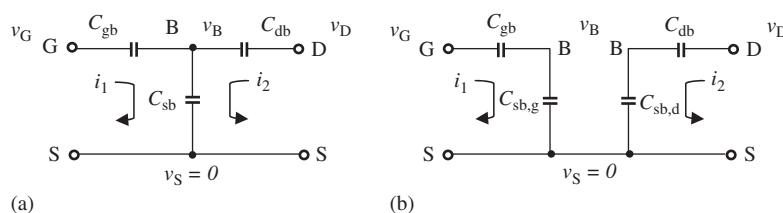


Figure 5.29. T network and its equivalent two  $\Gamma$  networks in CS device model,  $v_S = 0$ . (a) T network. (b)  $\Gamma$  network.

as shown in Figure 5.29(b). Then, one  $\Gamma$  branch on the left-hand side of Figure 5.29(b) is combined with  $C_{gs}$  in parallel to form a new capacitance  $C_{gs}$ , while the other  $\Gamma$  branch on the right-hand side of Figure 5.29(b) forms another new capacitance  $C_{ds}$ .

The question is: How will we divide the capacitance  $C_{sb}$ , and what are the values of  $C_{sb,g}$  and  $C_{sb,d}$ ?

In order to maintain the equivalence between Figure 5.29(a) and 5.29(b), the voltages  $v_G$ ,  $v_B$ ,  $v_D$ ,  $i_1$ ,  $i_2$ ,  $C_{gb}$ , and  $C_{db}$  must be kept constant from Figure 5.29(a) to 5.29(b).

In Figure 5.29(a),

$$i_1 = \frac{v_G - v_B}{\frac{1}{j\omega C_{gb}}} = j\omega C_{gb}(v_G - v_B), \quad (5.268)$$

$$i_2 = \frac{v_D - v_B}{\frac{1}{j\omega C_{db}}} = j\omega C_{db}(v_D - v_B), \quad (5.269)$$

$$v_B = \frac{i_1 + i_2}{j\omega C_{sb}} = \frac{j\omega C_{gb}(v_G - v_B) + j\omega C_{db}(v_D - v_B)}{j\omega C_{sb}} = \frac{C_{gb}(v_G - v_B) + C_{db}(v_D - v_B)}{C_{sb}}. \quad (5.270)$$

In Figure 5.29(b),

$$v_B = \frac{i_1}{j\omega C_{sb,g}} = \frac{i_2}{j\omega C_{sb,d}}, \quad (5.271)$$

$$v_B = \frac{C_{gb}(v_G - v_B)}{C_{sb,g}} = \frac{C_{db}(v_D - v_B)}{C_{sb,d}}. \quad (5.272)$$

From (5.270) and the first equation of (5.272), we have

$$\frac{C_{gb}(v_G - v_B)}{C_{sb,g}} = \frac{C_{gb}(v_G - v_B) + C_{db}(v_D - v_B)}{C_{sb}}, \quad (5.273)$$

$$\left( \frac{1}{C_{sb,g}} - \frac{1}{C_{sb}} \right) C_{gb}(v_G - v_B) = \frac{C_{db}(v_D - v_B)}{C_{sb}}, \quad (5.274)$$

$$\frac{C_{sb} - C_{sb,g}}{C_{sb,g} C_{sb}} C_{gb}(v_G - v_B) = \frac{C_{db}(v_D - v_B)}{C_{sb}}, \quad (5.275)$$

$$\frac{C_{gb}(v_G - v_B)}{C_{sb,g}} = \frac{C_{db}(v_D - v_B)}{(C_{sb} - C_{sb,g})}. \quad (5.276)$$

Comparing (5.272) and (5.276), we have

$$C_{sb} = C_{sb,g} + C_{sb,d}. \quad (5.277)$$

From (5.272)

$$C_{sb,d} = C_{sb,g} \frac{C_{db}(v_D - v_B)}{C_{gb}(v_G - v_B)}. \quad (5.278)$$

Combining (5.277) and (5.278),

$$C_{sb,g} = \frac{1}{\left[ 1 + \frac{C_{db}(v_D - v_B)}{C_{gb}(v_G - v_B)} \right]} C_{sb} = \frac{C_{gb}(v_G - v_B)}{C_{gb}(v_G - v_B) + C_{db}(v_D - v_B)} C_{sb}. \quad (5.279)$$

From (5.277) and (5.279), we have

$$C_{sb,d} = C_{sb} - C_{sb,g} = \frac{C_{db}(v_D - v_B)}{C_{gb}(v_G - v_B) + C_{db}(v_D - v_B)} C_{sb}. \quad (5.280)$$

As an approximation, the values of  $C_{gb}$  and  $C_{db}$  are considered to be in the same order, that is,

$$C_{gb} \Leftrightarrow C_{db}, \quad (5.281)$$

and usually in a CS device,

$$v_B \ll v_D, \text{ and } v_B \ll v_G. \quad (5.282)$$

Incorporating these two approximations into (5.279) and (5.280), we have

$$C_{sb,g} \approx \frac{C_{gb}}{A_{v,CS} C_{db}} C_{sb}, \quad (5.283)$$

$$C_{sb,d} \approx \frac{1}{1 + \frac{C_{gb}}{A_{v,CS} C_{db}}} C_{sb}, \quad (5.284)$$

where  $A_{v,CS}$  is the open-circuit voltage gain of a CS device, that is,

$$A_{v,CS} = \frac{v_D}{v_G}. \quad (5.285)$$

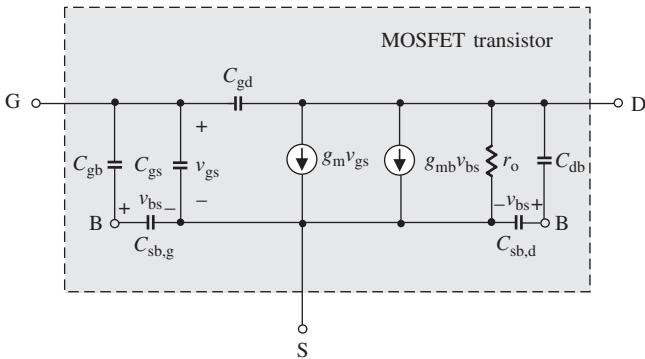
Now, in terms of expressions (5.283) and (5.284), the small-signal model as shown in Figure 5.28 can be modified to a new one as shown in Figure 5.30, in which the capacitance  $C_{sb}$  is split into two parts,  $C_{sb,g}$  and  $C_{sb,d}$ . The value of  $C_{sb,g}$  is much lower than  $C_{sb}$ , while the value of  $C_{sb,d}$  is lower than but almost equal to that of  $C_{sb}$ .

A further simplification of the modified model in Figure 5.30 is to combine  $C_{gb}$ ,  $C_{sb,g}$ , and  $C_{gs}$  as one capacitor  $C'_{gs}$ , and combine  $C_{db}$  and  $C_{sb,d}$  as another capacitor  $C'_{ds}$ , that is,

$$C'_{gs} = C_{gs} + \frac{1}{1 + A_{v,CS} \frac{C_{db}}{C_{sb}}} C_{gb}, \quad (5.286)$$

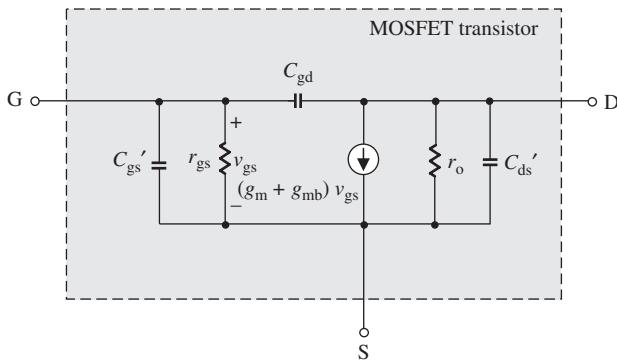
$$C'_{ds} = \frac{1}{1 + \frac{C_{db}}{C_{sb}} + \frac{C_{gb}}{A_{v,CS} C_{sb}}} C_{db}. \quad (5.287)$$

Expressions (5.286) and (5.287) show that the three small capacitors,  $C_{gb}$ ,  $C_{sb}$ , and  $C_{db}$ , increase the capacitance of  $C_{gs}$  up to that of  $C'_{gs}$  with an amount of  $C_{gb}$  and  $C_{sb,g}$  in series and forms a new capacitor  $C'_{ds}$  by  $C_{db}$  and  $C_{sb,d}$  in series.



$$\text{In CS device: } C_{\text{sb},g} \approx \frac{C_{\text{gb}}}{A_{v,\text{CS}} C_{\text{db}}} C_{\text{sb}}, \quad C_{\text{sb},d} \approx \frac{1}{1 + \frac{C_{\text{gb}}}{A_{v,\text{CS}} C_{\text{db}}}} C_{\text{sb}}$$

**Figure 5.30.** Modified model of a CS device when the capacitor  $C_{\text{sb}}$  is split into two capacitors  $C_{\text{sb},g}$  and  $C_{\text{sb},d}$ .



$$\text{In CS Device: } C'_{\text{gs}} = C_{\text{gs}} + \frac{1}{1 + A_{v,\text{CS}} \frac{C_{\text{db}}}{C_{\text{sb}}}} C_{\text{gb}}, \quad C'_{\text{ds}} = \frac{1}{1 + \frac{C_{\text{db}}}{C_{\text{sb}}} + \frac{C_{\text{gb}}}{A_{v,\text{CS}} C_{\text{sb}}}} C_{\text{db}}$$

**Figure 5.31.** Simplified model of a CS device with combined capacitors  $C'_{\text{gs}}$  and  $C'_{\text{ds}}$ .

Figure 5.31 shows the final simplified model of a CS device with the combined capacitors  $C'_{\text{gs}}$  and  $C'_{\text{ds}}$ . It looks very similar to the model of the CE device shown in Figure 5.6 (reproduced below).

It must be noted that the discussion in this subsection might not be important to RF circuit design if the designed block is operated in a relatively low frequency range because the three capacitances  $C_{\text{gb}}$ ,  $C_{\text{sb}}$ , and  $C_{\text{db}}$  have much lower values than  $C_{\text{gs}}$ . A simple workaround is to just forget or erase them from the schematic. However, in a high-frequency range, say, in gigahertz or above, the difference between  $C'_{\text{gs}}$  and  $C_{\text{gs}}$ , and the value of the new capacitor  $C'_{\text{ds}}$  is not negligible but appreciable.

## 5.8.2 Simplified Model of CG Device

Figure 5.32 is a copy of Figure 5.28 with the positions of S and G exchanged.

There are three small capacitors,  $C_{\text{gb}}$ ,  $C_{\text{sb}}$ , and  $C_{\text{db}}$ , which are connected between G (gate), S (source), D (drain), and B (substrate) with a T configuration as sketched in

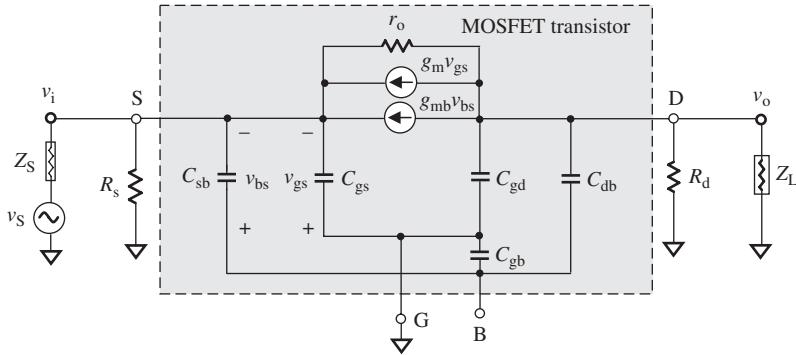


Figure 5.32. Model of a MOSFET with CG configuration.

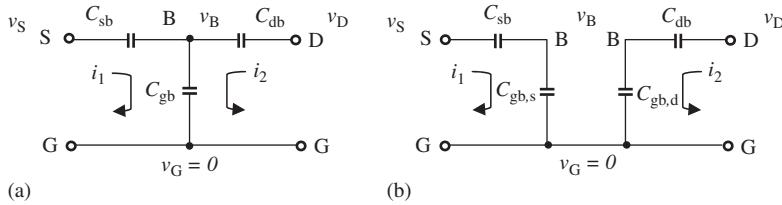
Figure 5.33. T network and its equivalent two  $\Gamma$  networks in CG device model,  $v_G = 0$ . (a) T network. (b)  $\Gamma$  network.

Figure 5.33(a). In order to make the MOSFET model similar to the bipolar one, it is preferable to divide the capacitor  $C_{gb}$  into two capacitors  $C_{gb,s}$  and  $C_{gb,d}$ , forming two piggy-back  $\Gamma$  configurations as shown in Figure 5.33(b). Then, one  $\Gamma$  branch on the left-hand side of Figure 5.33(b) is combined with  $C_{gs}$  in parallel to form a new capacitor  $C'_{gs}$ , while another  $\Gamma$  branch on the right-hand side of Figure 5.33(b) is combined with  $C_{gd}$  in parallel to form a new capacitor  $C'_{gd}$ .

The mathematical derivation for the values of  $C_{gb,s}$  and  $C_{gb,d}$  is similar to that for the capacitors  $C_{sb,g}$  and  $C_{sb,d}$  in the last subsection, (5.8.1). By comparing Figure 5.33 with Figure 5.29, one can find that the Figure 5.33 becomes Figure 5.29 if the subscripts “s” or “S” and the symbol “S” are exchanged with the subscripts “g” or “G” and the symbol “G,” respectively.

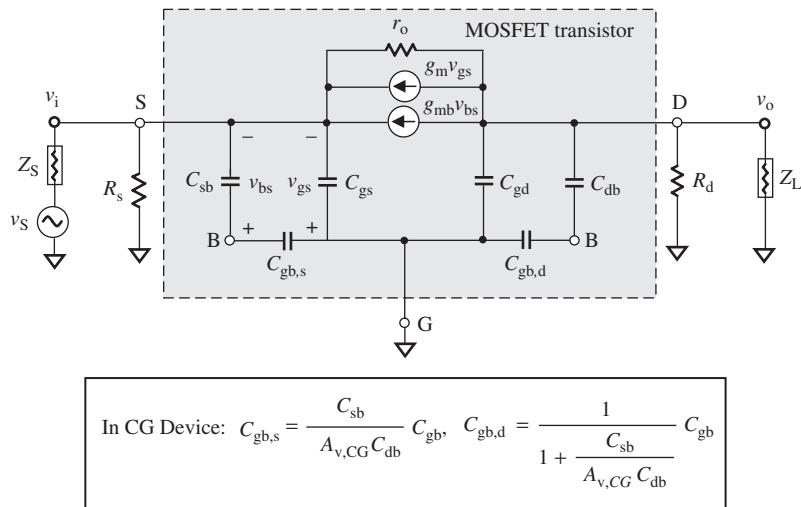
Therefore, it is not necessary to repeat the mathematical derivation. We need only copy expressions (5.283) to (5.285) with the subscripts and symbol exchanged as described; we then have

$$C_{gb,s} \approx \frac{C_{sb}}{A_{v,CG} C_{db}} C_{gb}, \quad (5.288)$$

$$C_{gb,d} \approx \frac{1}{1 + \frac{C_{sb}}{A_{v,CG} C_{db}}} C_{gb}, \quad (5.289)$$

where  $A_{v,CG}$  is open-circuit voltage gain of CG device, that is,

$$A_{v,CG} = \frac{v_D}{v_S}. \quad (5.290)$$



**Figure 5.34.** Modified model of a CS device when the capacitor  $C_{gb}$  is split into two capacitors  $C_{gb,s}$  and  $C_{gb,d}$ .

By replacing one T network by two  $\Gamma$  networks as shown in Figure 5.33 in Figure 5.32, the equivalent model of CG device becomes Figure 5.34. To further simplify the model, we can combine the capacitors  $C_{sb}$ ,  $C_{gb,s}$ , and  $C_{gs}$  as one capacitor  $C'_{gs}$ , and the capacitors  $C_{db}$ ,  $C_{gb,d}$ , and  $C_{gd}$  as one capacitor  $C'_{gd}$ , in which the node B is “swallowed” or “submerged” inside  $C'_{gs}$  and  $C'_{gd}$ :

$$C'_{gs} = C_{gs} + \frac{C_{sb} C_{gb,s}}{C_{sb} + C_{gb,s}} = C_{gs} + \frac{C_{sb}}{1 + A_{v,CG} \frac{C_{db}}{C_{gb}}}, \quad (5.291)$$

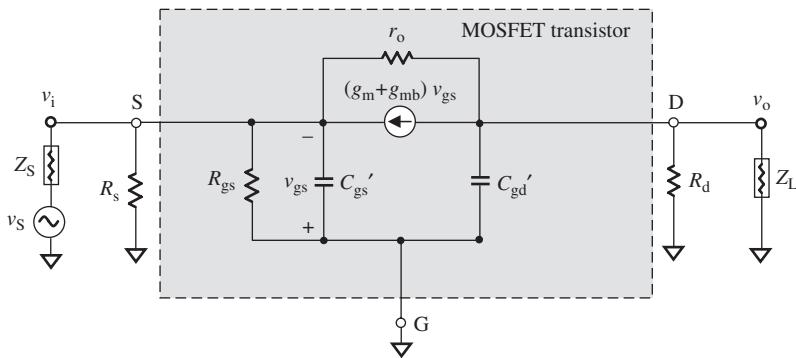
$$C'_{gd} = C_{gd} + \frac{C_{db} C_{gb,d}}{C_{db} + C_{gb,d}} = C_{gd} + \frac{C_{db}}{1 + \frac{C_{db}}{C_{gb}} + \frac{1}{A_{v,CG} \frac{C_{sb}}{C_{gb}}}}. \quad (5.292)$$

Meanwhile, the two current sources are combined as one for same reasons mentioned in the discussion of the CS device. In addition, a resistor  $r_{gs}$  is added to increase the similarity, though its value approaches infinity.

The simplified model of a MOSFET with a CG configuration with two combined capacitors  $C'_{gs}$  and  $C'_{gd}$  and a combined current source is shown in Figure 5.35, which is quite similar to the CB transistor model shown in Figure 5.36.

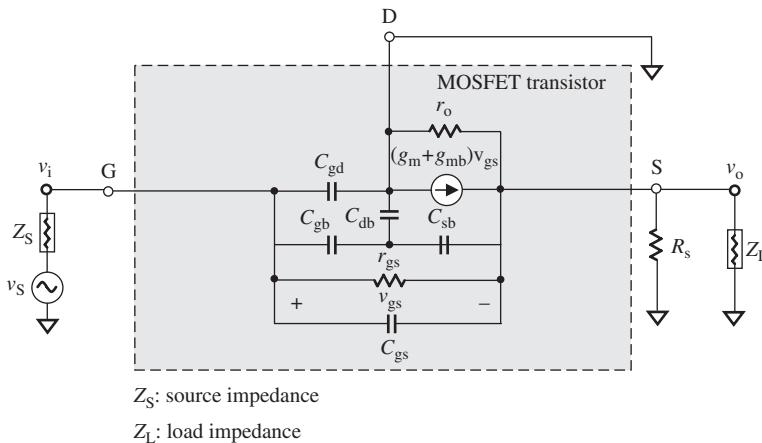
### 5.8.3 Simplified Model of CD Device

Figure 5.36 is a copy of Figure 5.28 with a  $90^\circ$  counterclockwise turning of the positions of D and S. The three small capacitors  $C_{gb}$ ,  $C_{sb}$ , and  $C_{db}$  are connected between G (gate), S (source), D (drain), and B (substrate) with a T configuration as sketched in Figure 5.37(a). In order to make the MOSFET model similar to the bipolar one, we divide the capacitor  $C_{db}$  into two capacitors  $C_{db,g}$  and  $C_{db,s}$ , so that they form two back-to-back  $\Gamma$  configurations as shown in Figure 5.37(b). Then, the  $\Gamma$  branch on the left-hand



$$\text{In CG Device: } C'_{gs} = C_{gs} + \frac{C_{sb}}{1 + A_{v,CG} \frac{C_{db}}{C_{gb}}}, \quad C'_{gd} = C_{gd} + \frac{C_{db}}{1 + \frac{C_{db}}{C_{gb}} + \frac{C_{sb}}{A_{v,CG} C_{gb}}}$$

**Figure 5.35.** Simplified model of a CS device with two combined capacitors  $C'_{gs}$  and  $C'_{gd}$  and a combined current source.

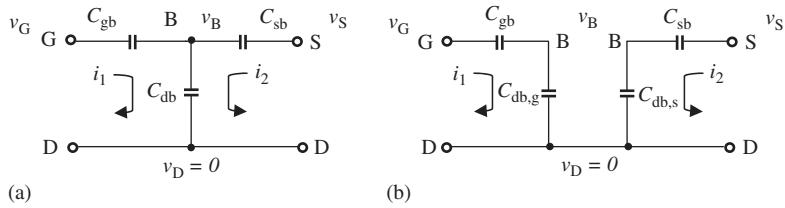


**Figure 5.36.** A MOSFET with CD configuration,  $r_{gs} \rightarrow \infty$ .

side of Figure 5.37(b) is combined with  $C_{gd}$  in parallel to form a new capacitor  $C'_{gd}$ , while the other  $\Gamma$  branch on right-hand side of Figure 5.37(b) forms a new capacitor  $C'_{ds}$ .

The mathematical derivation for the values of  $C_{db,g}$  and  $C_{db,s}$  is similar to that for the capacitors  $C_{gb,g}$  and  $C_{gb,d}$  in Section (5.8.2). By comparing Figures 5.37 and 5.33, one can find that Figure 5.33 becomes Figure 5.37 if

- the subscripts or symbol s and S are changed with the subscripts or symbol g and G, respectively;
- the subscripts or symbol g and G are changed to the subscripts or symbol d and D, respectively, and



**Figure 5.37.** T network and its equivalent two  $\Gamma$  networks in CD device model,  $v_D = 0$ . (a) T network. (b)  $\Gamma$  network.

- the subscripts or symbol d and D are changed to the subscripts or symbol s and S, respectively.

Therefore, it is not necessary to repeat the mathematical derivation. We need only copy expressions (5.288) to (5.290) with the subscripts and symbols exchanged as described; we then have

$$C_{db,g} \approx \frac{C_{gb}}{C_{gb} + A_{v,CD}C_{sb}} C_{db}, \quad (5.293)$$

$$C_{db,s} \approx \frac{1}{1 + \frac{C_{gb}}{A_{v,CD}C_{sb}}} C_{db}, \quad (5.294)$$

where  $A_{v,CD}$  is open-circuit voltage gain of CD device, that is,

$$A_{v,CD} = \frac{v_S}{v_G}. \quad (5.295)$$

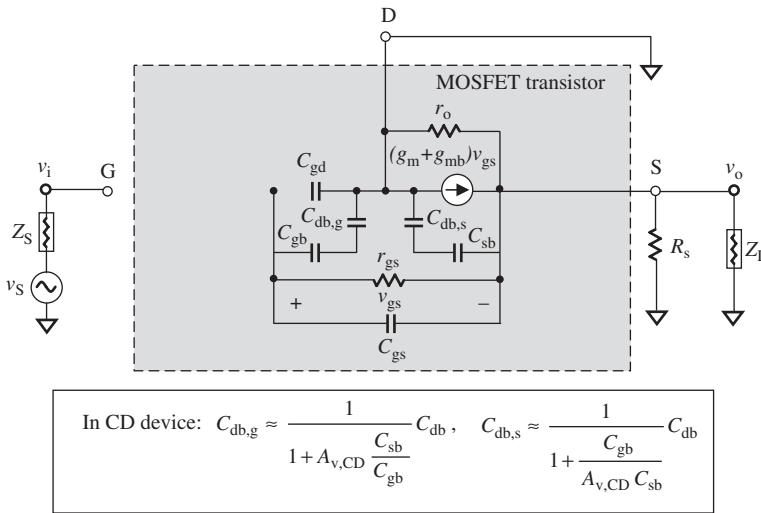
It should be noted that expression (5.293) is not exactly translated from expression (5.288) or (5.283). Instead, a capacitor  $C_{gb}$  is added in the denominator of expression (5.293). This term is recovered by neglecting in the derivation of expression (5.283) where  $A_{v,CS} \gg 1$ . In the case of the CD device,  $A_{v,CD}$  is not much greater than 1 but approximately equal to 1, that is,

$$A_v = \frac{v_S}{v_G} \approx 1. \quad (5.296)$$

Therefore, the neglected term due to the high open-circuit voltage gain must be recovered. On the other hand, from (5.293), (5.294), and (5.296), we have

$$C_{db,s} \approx C_{db,g} \approx \frac{1}{2} C_{db}. \quad (5.297)$$

By replacing one T network by two  $\Gamma$  networks as shown in Figure 5.37(b) in Figure 5.36, the equivalent model of the CG device becomes Figure 5.38. A further simplification of the model is to combine the capacitors  $C_{gb}$ ,  $C_{db,g}$ , and  $C_{gd}$  as one capacitor



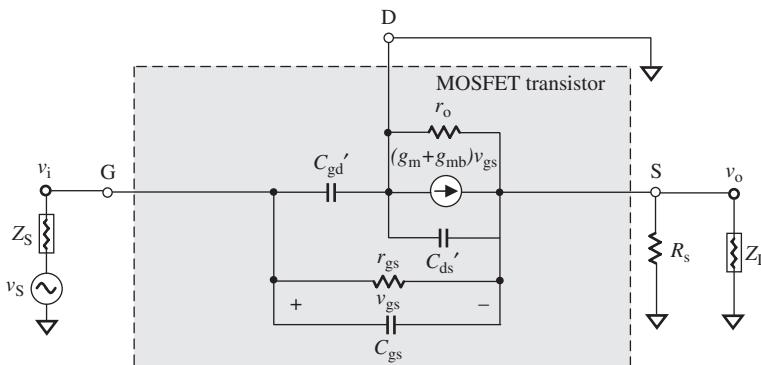
**Figure 5.38.** Modified model of a CD device when the capacitor  $C_{db}$  is split into two capacitors  $C_{db,g}$  and  $C_{db,s}$ .

$C'_{gd}$ , and the capacitors  $C_{db,s}$  and  $C_{sb}$  as one new capacitor  $C'_{ds}$ , in which the node B is swallowed or submerged inside  $C'_{gd}$  and  $C'_{ds}$ , which are

$$C'_{gd} = C_{gd} + \frac{C_{gb} C_{db,s}}{C_{gb} + C_{db,s}} = C_{gd} + \frac{C_{db}}{1 + 2 \frac{C_{gb}}{C_{db}}} C_{gb}, \quad (5.298)$$

$$C'_{ds} = \frac{C_{sb} C_{db,s}}{C_{sb} + C_{db,s}} = \frac{1}{1 + 2 \frac{C_{sb}}{C_{db}}} C_{sb}. \quad (5.299)$$

The simplified model of a MOSFET with a CD configuration when two combined capacitors  $C'_{gd}$  and  $C'_{ds}$  and a combined current source is shown in Figure 5.39, which is quite similar to the CB transistor model in Figure 5.22(b), reproduced above. The



$$\text{In CD Device: } C'_{gd} = C_{gd} + \frac{1}{1 + 2 \frac{C_{gb}}{C_{db}}} C_{gb}, \quad C'_{ds} = \frac{1}{1 + 2 \frac{C_{sb}}{C_{db}}} C_{sb}$$

**Figure 5.39.** Simplified model of a CD device with two combined capacitors  $C'_{gd}$  and  $C'_{ds}$  and a combined current source.

apparent difference is that the capacitor  $C'_{ds}$  in the CS model, due to the existence of  $C_{sb}$  and  $C_{db}$ , is connected with  $r_o$  in parallel, while  $C_{cs}$  in the CE model is a capacitor connected from the collector to the substrate.

Now let us compare the models of the MOSFET and bipolar transistor. The apparent differences can be outlined and some modifications must be made as follows:

- In the MOSFET model, there are two current sources  $g_m v_1$  and  $g_{mb} v_{bs}$ , while there is only one current source  $g_m v_1$  in the bipolar model. However, the two current sources can be combined as one, as

$$g_m v_{gs} + g_{mb} v_{bs} \rightarrow (g_m + g_{mb}) v_{gs} \quad (5.300)$$

because  $v_{bs} = v_{gs}$  if the substrate or body connection is assumed to operate at AC ground. In most cases, the value of  $g_{mb}$  is much lower than the value of  $g_m$  and is neglected.

- In order to easily compare the MOSFET and bipolar transistor model, a resistor with a very high value  $r_{gs}$  is added to the MOSFET model; therefore its impact on the model is negligible.
- On the contrary, there is a resistor  $r_\mu$  in the bipolar transistor model, while there is none in the corresponding position in the MOSFET model. This difference does not impact their similarity because the resistor  $r_\mu$  in the bipolar transistor model is usually neglected or is a resistor with a very high value.
- There is an emitter resistor  $r_x$  in the bipolar transistor model which does not exist in the MOSFET model. However, its value is very low, say, a couple ohms, and is very often neglected.
- The substrate symbol  $B$  is swallowed or submerged in the combined capacitors  $C'_{gs}$ ,  $C'_{ds}$ , and  $C'_{gd}$  in the simplified MOSFET model; however, this does not impact the analysis of the main parameters.
- On the other hand, the substrate in the bipolar collector substrate capacitor  $C_{cs}$  is usually grounded, whereas the substrate in the MOSFET drain–body depletion capacitor  $C_{db}$  can be grounded or connected by other methods. In our discussed MOSFET model, we connected the substrate to the terminal S. Actually, the substrate is grounded in the same way as in the MOSFET CS case and the bipolar CE case. However, the topologies of  $C_{cs}$  and  $C_{db}$  are slightly different because, in the MOSFET CG and CD cases, the substrate is connected to terminal S, while in the bipolar CB and CC cases the substrate is grounded. This slight difference might somewhat impact the calculation of the input and output impedances, but could simply be taken as a tolerance or an allowable approximation.

The similarity between the bipolar and MOSFET models thus provides a shortcut to translate most of the formulas, such as the open-circuit voltage gain, the short-circuit current gain, and the input and output impedances, from the bipolar transistor to the MOSFET as long as we perform the replacements shown in Table 5.1.

In a CS device:

$$C'_{gs} = C_{gs} + \frac{1}{1 + A_{v,CS} \frac{C_{db}}{C_{sb}}} C_{gb}, \quad C'_{ds} = \frac{1}{1 + \frac{C_{db}}{C_{sb}} + \frac{C_{gb}}{A_{v,CS} C_{sb}}} C_{db}.$$

TABLE 5.1. Corresponding Relationship of Parameters between Bipolar and MOSFET Devices

CE/CB/CD	From Bipolar		To MOSFET model	
	CS		CG	CD
$r_x$	0		0	0
$r_b$	0		0	0
$r_c$	0		0	0
$r_\pi$	$r_{gs}$		$r_{gs}$	$r_{gs}$
$r_o$	$r_o$		$r_o$	$r_o // C'_{ds}$
$C_\mu$	$C_{gd}$			$C'_{gd}$
$C_\mu + C_{cs}$			$C'_{gd} (\text{CG})$	
$C_{cs}$	$C'_{ds}$			
$C_\pi$	$C'_{gs}$		$C'_{gs}$	$C_{gs}$
$C_{cs}/R_c$				$R_d$
$R_c$	$R_d$		$R_d$	
$g_m$	$(g_m + g_{mb})$		$(g_m + g_{mb})$	$(g_m + g_{mb})$
$\beta_o$	$(g_m + g_{mb})r_{gs}$		$(g_m + g_{mb})r_{gs}$	$(g_m + g_{mb})r_{gs}$

Note 1:  $r_{gs} \rightarrow \infty$ .

Note 2: Capacitors in MOSFET models.

Note 3:  $A_{v,CS}$  and  $A_{v,CG}$  are open-circuit voltage gain of CS and CG devices, respectively.

In a CG Device:

$$C'_{gs} = C_{gs} + \frac{1}{1 + A_{v,CG} \frac{C_{db}}{C_{gb}}} C_{sb},$$

$$C'_{gd} = C_{gd} + \frac{1}{1 + \frac{C_{db}}{C_{gb}} + \frac{C_{sb}}{A_{v,CG} C_{gb}}} C_{db}.$$

In CD Device:

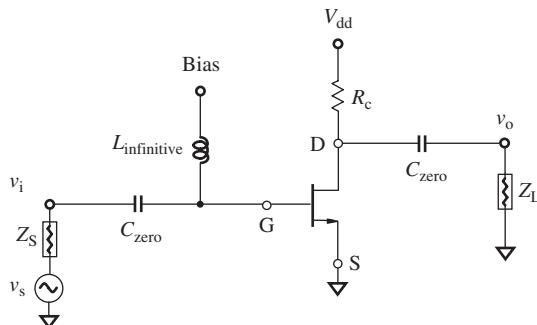
$$C'_{gd} = C_{gd} + \frac{1}{1 + 2 \frac{C_{gb}}{C_{db}}} C_{gb}, C'_{ds} = \frac{1}{1 + 2 \frac{C_{sb}}{C_{db}}} C_{sb}.$$

## 5.9 MOSFET WITH CS (COMMON SOURCE) CONFIGURATION

A MOSFET with a CS configuration can be drawn as Figure 5.40. Figure 5.40(a) plots its schematic, which contains two zero capacitors  $C_{zero}$  and one infinite  $L_{infinite}$ . The drain resistor  $R_d$  leads DC power to the drain. Figure 5.40(b) is the simplified model of the CS device with the combined capacitors  $C'_{gs}$  and  $C'_{ds}$ , in which all the zero capacitors and the infinite inductor are ignored.

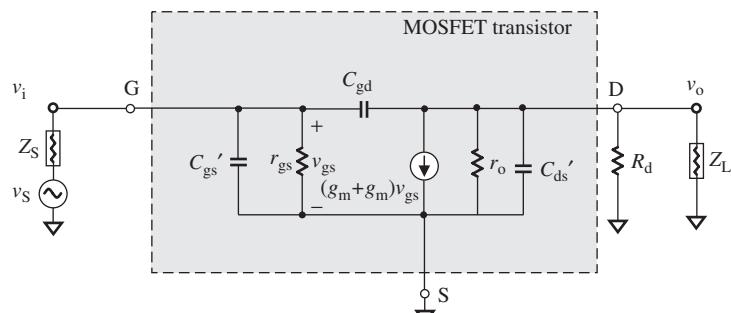
### 5.9.1 Open-Circuit Voltage Gain $A_{v,CS}$ of a CS Device

The simplified model of the CS device with the combined capacitors  $C'_{gs}$  and  $C'_{ds}$  shown in Figure 5.40(b) is quite similar to the model of the CE device shown in Figure 5.7(b).



- $C_{\text{zero}}$ : “zero” capacitor approaches to zero capacitance in operating frequencies
- $L_{\text{infinite}}$ : “infinite” inductor approaches to infinitive inductance in operating frequencies

(a)



(b)

$$\text{In CS Device: } C_{gs}' = C_{gs} + \frac{1}{1 + A_{v,CS}} C_{gb}, \quad C_{ds}' = \frac{1}{1 + \frac{C_{db}}{C_{sb}} + \frac{C_{gb}}{A_{v,CS} C_{sb}}} C_{db}$$

 $Z_s$ : source impedance     $Z_L$ : load impedance

**Figure 5.40.** A MOSFET with CS configuration. (a) Schematic of a MOSFET with CS configuration. (b) Modified model of a MOSFET with combined capacitors  $C'_{gs}$  and  $C'_{ds}$ .

Based on the similarity between these two models, the open-circuit voltage gain of the CS device can be obtained by translating expression (5.49) in terms of similar relationships shown in Table 5.1. Table 5.2 lists those relationships that must be particularly taken care of in the translation from the CE to the CS configuration.

$$A_{v,CS} = -(g_m + g_{mb} - j\omega C_{gd}) \frac{r_o R_d}{r_o + R_d} \frac{1}{1 + j\omega(C_{gd} + C_{ds}') \frac{r_o R_d}{r_o + R_d}}. \quad (5.301)$$

It can be seen that this expression is simpler than (5.52) because there is no gate resistor as the base resistor  $r_b$  in the CE device. In addition, instead of two poles in the expression (5.52), there is only one pole in the expression (5.301):

$$\omega_{-3 \text{ dB}} = |p_1| = \frac{1}{(C_{gd} + C_{ds}') \frac{r_o R_d}{r_o + R_d}}. \quad (5.302)$$

TABLE 5.2. Special Relationships that must be taken care in the Translation from CE to CS Configuration

From Bipolar model	To MOSFET model
CE	CS
$r_\pi$	$r_{gs}$
$r_o$	$r_o$
$C_\mu$	$C_{gd}$
$C_{cs}$	$C'_{ds}$
$C_\pi$	$C'_{gs}$
$R_c$	$R_d$
$g_m$	$(g_m + g_{mb})$
$\beta_o$	$(g_m + g_{mb})r_{gs}$

In low-frequency cases, the terms containing capacitors in expression (5.301) can be omitted, and then the open-circuit voltage gain becomes

$$A_{v,CS}|_{\omega \rightarrow 0} = -(g_m + g_{mb}) \frac{r_o R_d}{r_o + R_d}, \quad (5.303)$$

which is similar to the expression (5.55).

If Miller effect is considered, the voltage gain can be translated from expression (5.89) as

$$A_{v,CS} = \frac{v_o}{v_i} = -(g_m + g_{mb}) \frac{r_o R_d}{r_o + R_d} \frac{1}{1 + j\omega C_{out} \frac{r_o R_d}{r_o + R_d}}. \quad (5.304)$$

It can again be seen that, instead of two poles as in the CE device, there is only one pole existing in (5.304) because there is no the gate resistor as the base resistor  $r_b$  in the CE device, and

$$\omega_{-3 \text{ dB}} = |p_1| = -\frac{1}{C_{in} r_b} = \frac{1}{C_{out} \frac{r_o R_d}{r_o + R_d}}. \quad (5.305)$$

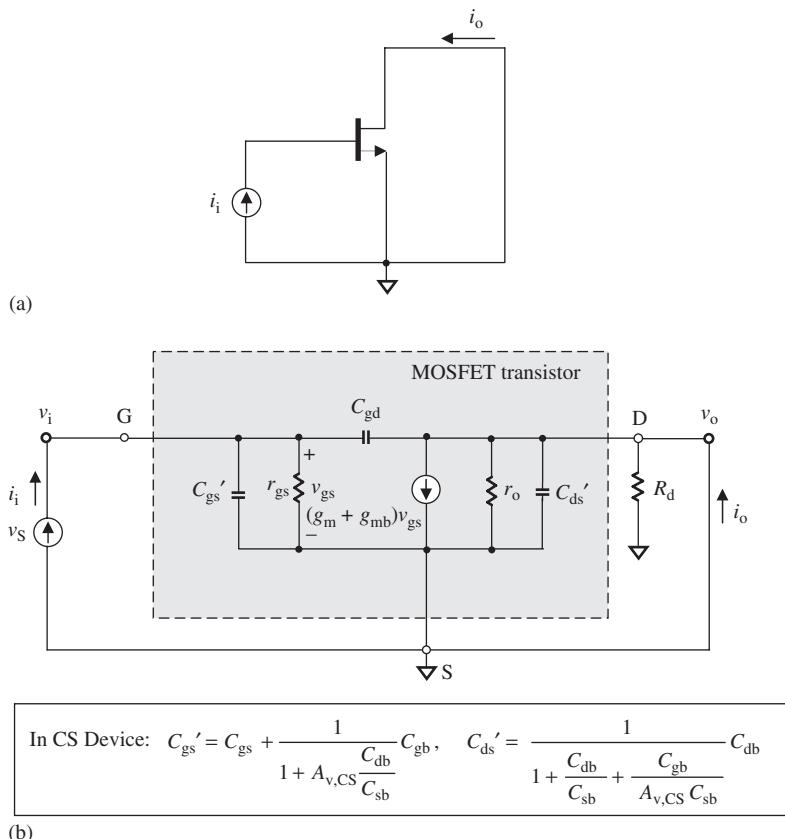
In low-frequency cases, the terms containing capacitors in expression (5.304) can be taken away; then, the open-circuit voltage gain becomes

$$A_{v,CS}|_{\omega \rightarrow 0} = -(g_m + g_{mb}) \frac{r_o R_d}{r_o + R_d}. \quad (5.306)$$

Expressions (5.304)–(5.306) are quite similar to expressions (5.301)–(5.303) respectively.

## 5.9.2 Short-Circuit Current Gain $\beta_{CS}$ and Frequency Response of a CS Device

Figure 5.41 is drawn for the calculation of the short-circuit current gain and frequency response of a MOSFET with a CS configuration.



**Figure 5.41.** Output short-circuited for calculating of current gain  $\beta_{CS}$  and frequency response of a MOSFET with CS configuration. (a) Output short-circuited for calculation of current gain  $\beta_{CS}$  and frequency response. (b) Modified model of a MOSFET with combined capacitors  $C_{gs}'$  and  $C_{ds}'$ .

Similar to the mathematical derivation in the corresponding section about CE device, if the current going through the capacitor  $C_{gd}$  is neglected, and note that

$$\beta_o = (g_m + g_{mb})r_{gs}, \quad (5.307)$$

then translated from (5.60), we have

$$\begin{aligned} \beta_{CS}(j\omega) &= \frac{i_o}{i_i} = \frac{(g_m + g_{mb})r_{gs}}{1 + j(g_m + g_{mb})r_{gs} \frac{(C_{gs}' + C_{gd})}{g_m + g_{mb}} \omega} \\ &= \frac{\beta_o}{1 + j(\omega C_{gs}' + C_{gd})r_{gs}}. \end{aligned} \quad (5.308)$$

It can be seen that  $\beta_{CS}(j\omega) = \beta_o$  in low frequency cases, when  $\omega \rightarrow 0$ , and in high frequency cases

$$\beta_{CS}(j\omega) \approx \frac{(g_m + g_{mb})}{j\omega(C_{gs}' + C_{gd})}. \quad (5.309)$$

When the input current is equal to the output current in magnitude, that is,  $\beta_{\text{CS}}(j\omega) = 1$ , then from (5.309) it can be found that the corresponding frequency  $\omega_{T,\text{CS}}$  is

$$\omega = \omega_{T,\text{CS}} = \frac{g_m + g_{mb}}{C'_{gs} + C_{gd}}, \quad (5.310)$$

$\omega_{T,\text{CS}}$  is called the transition frequency of the CS device.

On the other hand, when  $\beta_{\text{CS}}$  is 3 dB lower than the low-frequency value  $\beta_o$ , that is,

$$|\beta_{\text{CS}}(j\omega)| = \frac{\beta_{\text{CS}}(j\omega)|_{\omega \rightarrow 0}}{\sqrt{2}} = \frac{\beta_o}{\sqrt{2}}. \quad (5.311)$$

then the corresponding transition frequency of the CS device,  $\omega_{\beta,\text{CS}}$ , is

$$\omega = \omega_{\beta,\text{CS}} = \left( \frac{g_m + g_{mb}}{C'_{gs} + C_{gd}} \right) \cdot \frac{1}{\beta} = \frac{1}{\beta} \omega_{T,\text{CS}}. \quad (5.312)$$

### 5.9.3 Input and Output Impedance of a CS Device

As we did for the CE device in the calculation of input and output impedances, the feedback capacitor  $C_{gd}$  will be omitted, and the input capacitor  $C'_{gs}$  and output capacitor  $C'_{ds}$  will be replaced by  $C_{i,\text{miller}}$  and  $C_{o,\text{miller}}$ , which can be translated from the expressions (5.77) and (5.78), respectively, as follows:

$$C_{i,\text{miller}} = C_{gd} \left[ 1 + (g_m + g_{mb}) \frac{r_o}{r_o + R_d} R_d \right], \quad (5.313)$$

and

$$C_{o,\text{miller}} = C_{gd} \left[ 1 + \frac{r_o + R_d}{(g_m + g_{mb}) r_o R_d} \right]. \quad (5.314)$$

The total input capacitance is therefore

$$C_{in} = C'_{gs} + C_{i,\text{miller}} = C'_{gs} + C_{gd} \left[ 1 + (g_m + g_{mb}) \frac{r_o}{r_o + R_d} R_d \right]. \quad (5.315)$$

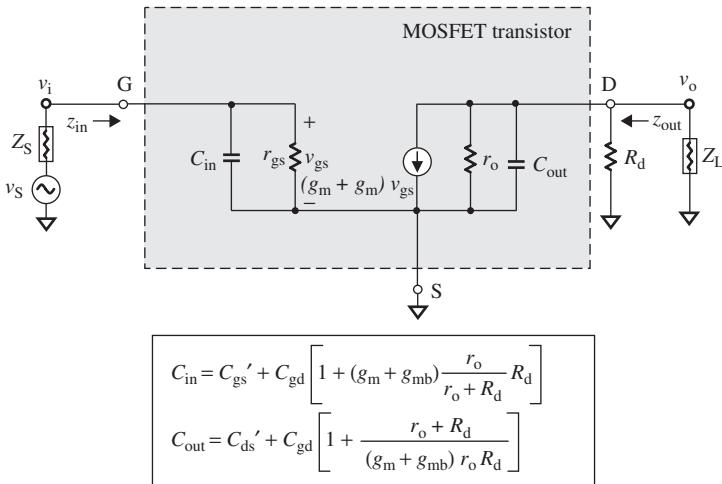
And the total output capacitance is

$$C_{out} = C'_{ds} + C_{o,\text{miller}} = C'_{ds} + C_{gd} \left[ 1 + \frac{r_o + R_d}{(g_m + g_{mb}) r_o R_d} \right]. \quad (5.316)$$

Figure 5.42 shows the modified model of a CS device. The input and output impedances are

$$Z_{in} = \frac{1}{jC_{in}\omega}, \quad (5.317)$$

$$Z_{out} = \frac{r_o}{R_d + r_o + jC_{out}\omega r_o R_d} R_d. \quad (5.318)$$



**Figure 5.42.** Modified model of a CS device with combined capacitors  $C_{in}$  and  $C_{out}$  for the calculation of input and output impedances.

In a low-frequency range, all the capacitors can be neglected. Thus, we have

$$Z_{in}|_{\omega \rightarrow 0} \rightarrow \infty, \quad (5.319)$$

$$Z_{out}|_{\omega \rightarrow 0} = r_o // R_d = \frac{r_o R_d}{r_o + R_d}. \quad (5.320)$$

#### 5.9.4 Source Degeneration

Figure 5.43 shows a MOSFET device with a CS configuration and source degeneration. Its model would be simplified with the additional approximation that

- the capacitor  $C_{gd}$  can be neglected because its value is much lower than that of  $C'_{gs}$ , that is,

$$C_{gd} \ll C'_{gs}. \quad (5.321)$$

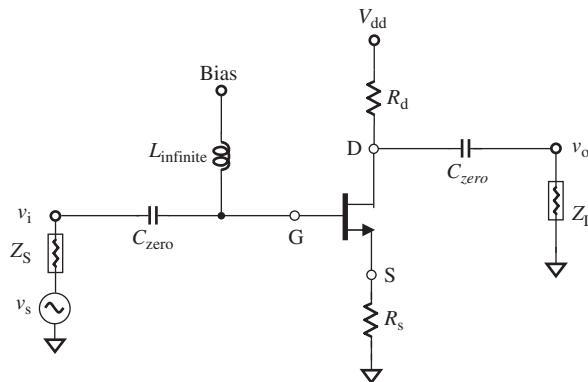
Then, the equivalent model of the CS device with source degeneration shown in Figure 5.43(b) is simplified as shown in Figure 5.44. Comparing Figures 5.44 and 5.13, we see that the capacitor  $C_{ds}'$  in Figure 5.44 does not correspond to the capacitor  $C_{cs}$  in Figure 5.13. Therefore, those relationships to be specially taken care of in the translation from the CE device with emitter degeneration to the CS device with source degeneration are listed in Table 5.3.

Consequently, from (5.99) we have

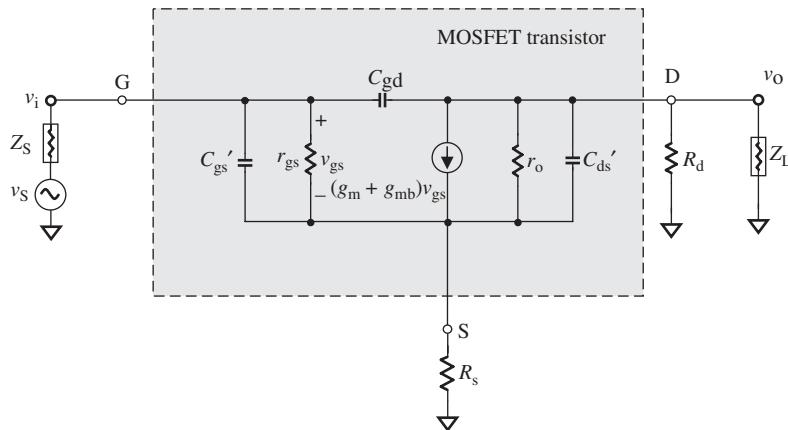
$$Z_{in} \approx \frac{1}{j\omega C'_{gs}} \left\{ 1 + \frac{r_o[(g_m + g_{mb}) + j\omega C'_{gs}] + j\omega C'_{gs} R_d (1 + j\omega C_{ds}' r_o)}{r_o + (R_d + R_s)(1 + j\omega C_{ds}' r_o)} R_s \right\}, \quad (5.322)$$

under the condition

$$r_{gs} \rightarrow \infty. \quad (5.323)$$



$C_{\text{zero}}$ : "zero" capacitor approaches to zero capacitance in operating frequencies,  
 $L_{\text{infinite}}$ : "infinite" inductor approaches to infinite inductance in operating frequencies.  
(a)



$$\text{In CS Device: } C_{\text{gs}}' = C_{\text{gs}} + \frac{1}{1 + A_v \frac{C_{\text{db}}}{C_{\text{sb}}}} C_{\text{gb}}, \quad C_{\text{ds}}' = \frac{1}{1 + \frac{C_{\text{db}}}{C_{\text{sb}}} + \frac{C_{\text{gb}}}{A_v C_{\text{sb}}}} C_{\text{db}}$$

(b)  $Z_s$ : source impedance     $Z_L$ : load impedance

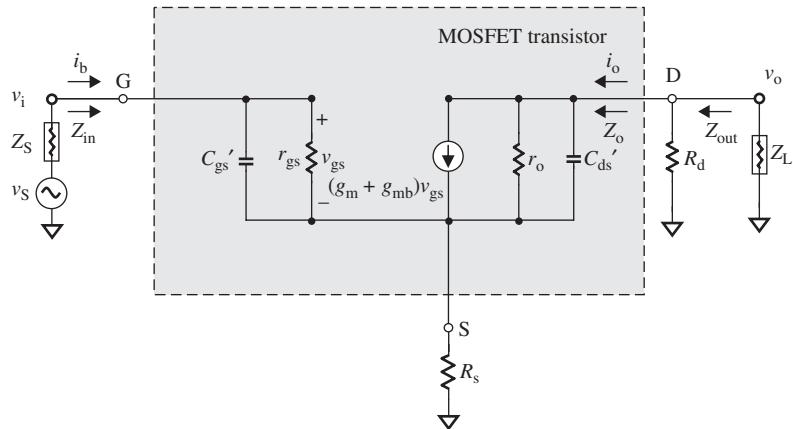
**Figure 5.43.** Bipolar transistor with CE configuration and emitter degeneration. (a) Schematic of a MOSFET with CS configuration and emitter degeneration. (b) Modified model of a MOSFET with CS configuration and source degeneration and with combined capacitors  $C'_{\text{gs}}$  and  $C'_{\text{ds}}$ .

If

$$r_o \gg R_d, \quad (5.324)$$

$$r_o \gg R_s, \text{ and} \quad (5.325)$$

$$Z_{\text{in}} \approx \frac{1}{j\omega C'_{\text{gs}}} \left\{ 1 + \frac{[(g_m + g_{\text{mb}}) - \omega^2 C'_{\text{gs}} C'_{\text{ds}} R_d + j\omega C'_{\text{gs}}] R_s}{1 + j\omega C'_{\text{ds}} (R_d + R_s)} \right\}. \quad (5.326)$$



$$\text{In CS Device: } C'_{gs} = C_{gs} + \frac{1}{1 + A_{v,CS} \frac{C_{db}}{C_{sb}}} C_{gb}, \quad C'_{ds} = \frac{1}{1 + \frac{C_{db}}{C_{sb}} + \frac{C_{gb}}{A_{v,CS} C_{sb}}} C_{db}$$

**Figure 5.44.** Modified model of a CS device with source degeneration and with combined capacitors  $C'_{gs}$  and  $C'_{ds}$ .

**TABLE 5.3.** Special Relationships that must be taken care in the Translation of Equations from CE with Emitter Degeneration to CS with Source Degeneration

From Bipolar model CE	To MOSFET model CS
$r_\pi$	$r_{gs}$
$r_o$	$r_o // C'_{ds}$
$C_\mu$	$C_{gd}$
$C_{cs}$	$C'_{ds}$
$C_\pi$	$C'_{gs}$
$C_{cs}/R_c$	$R_d$
$g_m$	$(g_m + g_{mb})$
$\beta_o$	$(g_m + g_{mb})r_{gs}$

And at low frequencies,

$$Z_{in}|_{\omega \rightarrow 0} \rightarrow \infty. \quad (5.327)$$

So, at low frequencies, the input impedance approaches infinity. At high frequencies, the input impedance is capacitive. Its capacitance is reduced because of the existence of  $R_s$  but is increased because of the existence of  $R_d$ .

Let us translate the transconductance  $G_m = i_o/v_i$  from (5.105) with output short-circuited, that is,  $R_d = 0$ .

Then

$$G_m = \frac{i_o}{v_i} = (g_m + g_{mb}) \frac{1 - \frac{R_s(1+j\omega C'_{gs}r_{gs})}{(g_m+g_{mb})r_{gs}\frac{r_o}{1+j\omega C'_{ds'}r_o}}}{1 + (g_m + g_{mb})R_s \left[ 1 + \frac{(1+j\omega C'_{gs}r_{gs})}{(g_m+g_{mb})r_{gs}} + \frac{1+j\omega C'_{ds'}r_o}{(g_m+g_{mb})r_o} \right]}. \quad (5.328)$$

Note that if

$$r_{\text{gs}} \rightarrow \infty,$$

$$G_m \approx (g_m + g_{mb}) \frac{1 - \frac{j\omega C'_{\text{gs}} R_s (1+j\omega C_{\text{ds}}' r_o)}{(g_m + g_{mb}) r_o}}{1 + R_s \left[ (g_m + g_{mb}) + j\omega (C'_{\text{gs}} + C_{\text{ds}}') + \frac{1}{r_o} \right]}. \quad (5.329)$$

And usually

$$r_o \gg R_s, \quad (5.330)$$

$$(g_m + g_{mb}) r_o \gg 1, \quad (5.331)$$

and so

$$G_m \approx (g_m + g_{mb}) \frac{1 + \frac{\omega^2 C'_{\text{gs}} C_{\text{ds}}' R_s}{(g_m + g_{mb})}}{1 + R_s [(g_m + g_{mb}) + j\omega (C'_{\text{gs}} + C_{\text{ds}}')]} \quad (5.332)$$

At low frequencies,

$$G_m|_{\omega \rightarrow 0} \approx \frac{(g_m + g_{mb})}{1 + (g_m + g_{mb}) R_s}. \quad (5.333)$$

From (5.328), (5.329), (5.332), and (5.333) it can be seen that  $G_m$  drops from  $(g_m + g_{mb})$  because of the source resistor  $R_s$ . At low frequencies,  $G_m$  will drop to half of  $(g_m + g_{mb})$  if  $R_e = 1/(g_m + g_{mb})$ .

The output impedance can be translated from (5.118), that is,

$$Z_{\text{out}} = \frac{r_o}{1 + j\omega C_{\text{ds}}' r_o} \frac{1 + \left( \frac{1+j\omega C_{\text{ds}}' r_o}{r_o} + g_m + g_{mb} \right) \frac{r_{\text{gs}} R_s}{r_{\text{gs}} + R_s (1+j\omega C'_{\text{gs}} r_{\text{gs}})}}{1 + \frac{r_o}{R_d} \frac{1}{(1+j\omega C_{\text{ds}}' r_o)} \left[ 1 + \left( \frac{1+j\omega C_{\text{ds}}' r_o}{r_o} + g_m + g_{mb} \right) \frac{r_{\text{gs}} R_s}{r_{\text{gs}} + R_s (1+j\omega C'_{\text{gs}} r_{\text{gs}})} \right]}. \quad (5.334)$$

Note that as

$$r_{\text{gs}} \rightarrow \infty,$$

$$Z_{\text{out}} = \frac{r_o}{1 + j\omega C_{\text{ds}}' r_o} \frac{1 + \left( \frac{1+j\omega C_{\text{ds}}' r_o}{r_o} + g_m + g_{mb} \right) \frac{R_s}{1+j\omega C'_{\text{gs}} R_s}}{1 + \frac{r_o}{R_d} \frac{1}{(1+j\omega C_{\text{ds}}' r_o)} \left[ 1 + \left( \frac{1+j\omega C_{\text{ds}}' r_o}{r_o} + g_m + g_{mb} \right) \frac{R_s}{1+j\omega C'_{\text{gs}} R_s} \right]}. \quad (5.335)$$

Usually

$$r_o \gg 1, \quad (5.336)$$

and so

$$Z_{\text{out}} = \frac{1}{j\omega C_{\text{ds}}'} \frac{1 + (j\omega C_{\text{ds}}' + g_m + g_{mb}) \frac{R_s}{1+j\omega C'_{\text{gs}} R_s}}{1 + \frac{r_o}{R_d} \frac{1}{(1+j\omega C_{\text{ds}}' r_o)} \left[ 1 + (j\omega C_{\text{ds}}' + g_m + g_{mb}) \frac{R_s}{1+j\omega C'_{\text{gs}} R_s} \right]}. \quad (5.337)$$

In cases of infinite  $R_d$ , that is,

$$R_d \rightarrow \infty, \quad (5.338)$$

$$Z_{\text{out}}|_{R_d \rightarrow \infty} = \frac{1}{j\omega C_{\text{ds}}'} \left[ \frac{1 + j\omega C'_{\text{gs}} R_s + (j\omega C'_{\text{ds}} + g_m + g_{\text{mb}})R_s}{1 + j\omega C'_{\text{gs}} R_s} \right]. \quad (5.339)$$

At low frequencies,

$$Z_{\text{out}}|_{\omega \rightarrow 0} = r_o \frac{1 + (g_m + g_{\text{mb}})R_s}{1 + \frac{r_o}{R_d} \left( 1 + \frac{1}{r_o} + g_m + g_{\text{mb}} \right) R_s}. \quad (5.340)$$

## 5.10 MOSFET WITH CG (COMMON GATE) CONFIGURATION

A MOSFET with CG configuration is shown in Figure 5.45.

### 5.10.1 Open-Circuit Voltage Gain of a CG Device

The modified model of a CG device with the combined capacitors  $C'_{\text{gs}}$  and  $C'_{\text{gd}}$  shown in Figure 5.44 is quite similar to the model of the CB device shown in Figure 5.14. On the basis of the similarity between these two models, the open-circuit voltage gain of the CG device can be obtained by translation from the corresponding expressions in Section 5.5.1 in terms of the similar relationships listed in Table 5.1. Table 5.4 lists those relationships that must be particularly taken care of in the translation from a CB to a CG configuration.

The open-circuit voltage gain of a CG device can be translated from expression (5.133) as

$$A_{v,CG} = \frac{r_o}{R_d + r_o} \frac{(g_m + g_{\text{mb}})R_d \left[ 1 + \frac{1}{(g_m + g_{\text{mb}})r_o} \right]}{1 + j\omega C'_{\text{gd}} \frac{(R_d + r_o)}{(R_d + r_o)} r_o}. \quad (5.341)$$

At low frequencies,

$$A_{v,CG}|_{\omega \rightarrow 0} = \frac{R_d r_o}{R_d + r_o} (g_m + g_{\text{mb}}) \left[ 1 + \frac{1}{(g_m + g_{\text{mb}})r_o} \right]. \quad (5.342)$$

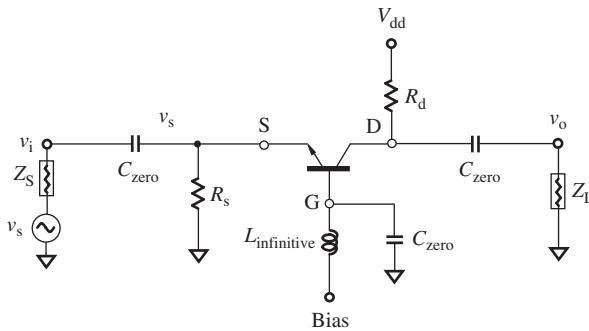
If

$$r_o \gg R_d, \quad (5.343)$$

$$(g_m + g_{\text{mb}})r_o \gg 1, \quad (5.344)$$

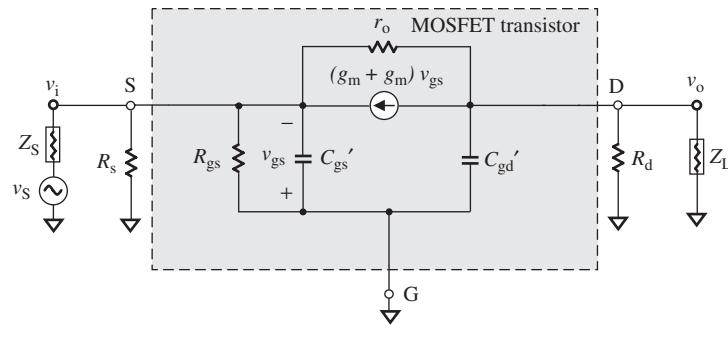
and so

$$A_{v,CG}|_{\omega \rightarrow 0} \approx (g_m + g_{\text{mb}})R_d. \quad (5.345)$$



$C_{zero}$ : “zero” capacitor approaches to zero capacitance in operating frequencies,  
 $L_{infinitive}$ : “infinitive” inductor approaches to infinitive inductance in operating frequencies.

(a)



$$\text{In CS Device: } C_{gs}' = C_{gs} + \frac{1}{1 + A_v \frac{C_{db}}{C_{gb}}} C_{sb}, \quad C_{gd}' = C_{gd} + \frac{1}{1 + \frac{C_{db}}{C_{gb}} + \frac{C_{sb}}{A_v C_{gb}}} C_{db}$$

(b)  $Z_s$ : source impedance     $Z_L$ : load impedance

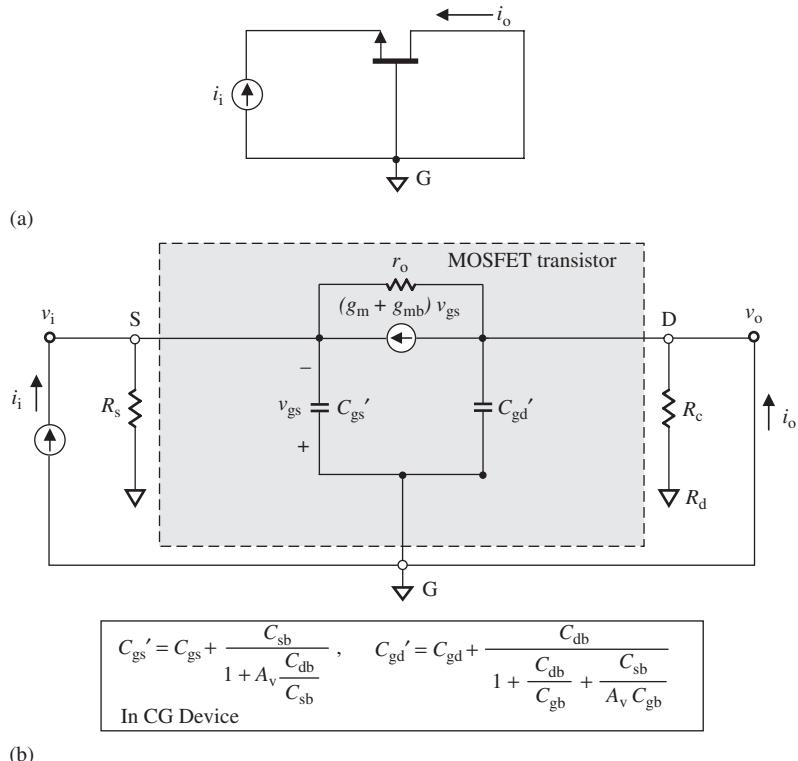
**Figure 5.45.** MOSFET with CG configuration. (a) Schematic of a MOSFET with CG configuration. (b) Simplified model of a MOSFET with CG configuration when two combined capacitors  $C'_{gs}$  and  $C'_{gd}$  and a combined current source.

**TABLE 5.4.** Special Relationships that must be taken care in the Translation from CB to CG Configuration

From Bipolar model CB	To MOSFET model CG
$r_\pi$	$r_{gs}$
$r_o$	$r_o'$
$C_\mu + C_{cs}$	$C'_{gd}$
$C_\pi$	$C'_{gs}$
$R_c$	$R_d$
$g_m$	$(g_m + g_{mb})$
$\beta_o$	$(g_m + g_{mb})r_{gs}$

### 5.10.2 Short-Circuit Current Gain and Frequency Response of a CG Device

Figure 5.46 is drawn for the calculation of the short-circuit current gain and frequency response of a MOSFET with the CG configuration. It is similar to Figure 5.14, which



**Figure 5.46.** Output short-circuited for calculating of current gain  $\beta$  and frequency response of a MOSFET with CG configuration. (a) Output short-circuited for calculation of current gain  $\beta$  of a CG device. (b) Model of a MOSFET with CG configuration when two combined capacitors  $C'_{gs}$  and  $C'_{gd}$  and a combined current source.

is a bipolar transistor with the CB configuration. Therefore, its short-circuit current gain and frequency response can be translated from expression (5.143) as

$$\beta_{CG}(j\omega) = \frac{i_o}{i_i} = \frac{(g_m + g_{mb})r_{gs}}{(g_m + g_{mb})r_{gs} + 1} \frac{1}{1 + j\omega C'_{gs} \frac{r_{gs}}{(g_m + g_{mb})r_{gs} + 1}} \approx \frac{1}{1 + j\omega \frac{C'_{gs}}{(g_m + g_{mb})}}, \quad (5.346)$$

because

$$r_{gs} \rightarrow \infty. \quad (5.347)$$

It can be seen that the current gain of the CG device is less than 1, that is,

$$|\beta_{CG}(j\omega)| < 1. \quad (5.348)$$

At low frequencies,

$$\beta_{CG}(j\omega)|_{\omega \rightarrow 0} \approx 1. \quad (5.349)$$

The current gain of the CG device at low frequencies is approximately equal to 1. Rewriting (5.417) in terms of (5.420), when  $\beta_{CG}(j\omega)$  is 3 dB lower than the low frequency value  $\beta_{CG}$ , that is,

$$|\beta_{CG}(j\omega)|_{-3 \text{ dB}} = \frac{\beta_{CG}(j\omega)|_{\omega \rightarrow 0}}{\sqrt{2}} = \frac{1}{\sqrt{2}}, \quad (5.350)$$

$$\omega = \omega_{\beta,CG} = \frac{g_m + g_{mb}}{C'_{gs}}, \quad (5.351)$$

or

$$f = f_{\beta,CG} = \frac{1}{2\pi} \frac{g_m + g_{mb}}{C'_{gs}}. \quad (5.352)$$

Figure 5.47 shows an example of the frequency response of a CG device.

### 5.10.3 Input and Output Impedance of a CG Device

Similar to the discussion for the CB device, from Figure 5.46(b) we discuss the input and output impedances somewhat reluctantly because the dependent current source is located between the drain and the source. By the same scheme as in the discussion for the CB device, the equivalent model of the CG device can be modified to Figure 5.48.

$$|\beta_{CG}(j\omega)| = |i_o / i_i|$$

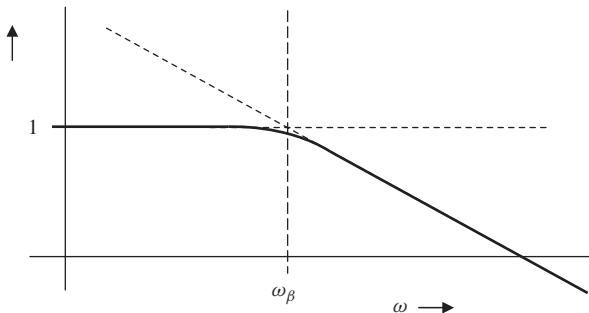
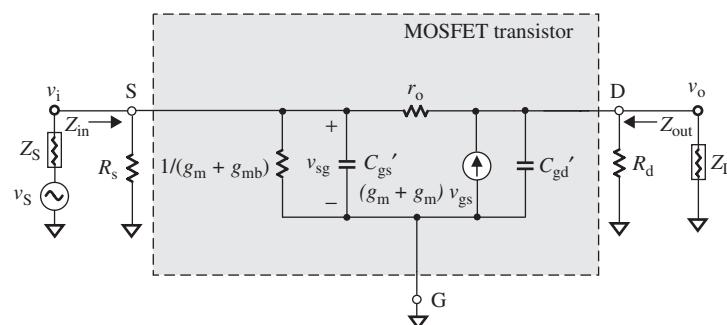


Figure 5.47. Frequency response of short-circuit current gain in a CG device.



In CG Device: $C_{gs}' = C_{gs} + \frac{1}{1 + A_v \frac{C_{db}}{C_{gb}}} C_{sb}, \quad C_{gd}' = C_{gd} + \frac{1}{1 + \frac{C_{db}}{C_{gb}} + \frac{C_{sb}}{A_v C_{gb}}} C_{db}$
--

Figure 5.48. Equivalent model of MOSFET with CG configuration for the calculation of input and output impedances.

Figure 5.48 is similar to Figure 5.20; the input impedances can be translated from the expressions (5.172) to (5.175), such that

$$Z_{\text{in}} = R_s // \frac{1}{(g_m + g_{mb})} \frac{1 + \frac{1}{r_o} \frac{R_d}{1+j\omega C'_{gd} R_d}}{1 + \frac{1}{(g_m + g_{mb})} \left[ \frac{1}{r_o} + j\omega C'_{gs} + \frac{1}{r_o} j\omega C'_{gs} \frac{R_d}{1+j\omega C'_{gd} R_d} \right]}. \quad (5.353)$$

It can be seen that the input impedance of a CG device is low and is about  $1/(g_m + g_{mb})$ . It should be noted that the resistor  $R_s$  might be unnecessary in some applications, while the resistor  $R_d$  is an indispensable part because it leads the DC power supply to the device. For instance, in the CS-CG cascade amplifier, implementation of the resistor  $R_s$  is not necessary, so that in such a case, we have

$$Z_{\text{in}}|_{R_s \rightarrow \infty} = \frac{1}{(g_m + g_{mb})} \frac{1 + \frac{1}{r_o} \frac{R_d}{1+j\omega C'_{gd} R_d}}{1 + \frac{1}{(g_m + g_{mb})} \left[ \frac{1}{r_o} + j\omega C'_{gs} + \frac{1}{r_o} j\omega C'_{gs} \frac{R_d}{1+j\omega C'_{gd} R_d} \right]}, \quad (5.354)$$

which corresponds to the case of

$$R_s \rightarrow \infty. \quad (5.355)$$

In low-frequency cases,

$$Z_{\text{in}}|_{\omega \rightarrow 0, R_s \rightarrow \infty} = \frac{1 + \frac{R_d}{r_o}}{(g_m + g_{mb}) + \frac{1}{r_o}}. \quad (5.356)$$

The output impedance is

$$Z_{\text{out}} = R_d \frac{1}{1 + j\omega C'_{dg}' R_d + R_d \frac{(1+j\omega C'_{gs} R_s)}{r_o + R_s [1 + (g_m + g_{mb} + j\omega C'_{gs}) r_o]}}. \quad (5.357)$$

In low-frequency cases,

$$Z_{\text{out}}|_{\omega \rightarrow 0} = R_d \frac{1}{1 + R_d \frac{1}{r_o + R_s [1 + (g_m + g_{mb}) r_o]}}. \quad (5.358)$$

In cases without the resistor  $R_s$ , that is,  $R_s \rightarrow \infty$ ,

$$Z_{\text{out}}|_{R_s \rightarrow \infty} = R_d \frac{1}{1 + j\omega C'_{dg}' R_d + \frac{j\omega C'_{gs} R_d}{1 + (g_m + g_{mb} + j\omega C'_{gs}) r_o}}. \quad (5.359)$$

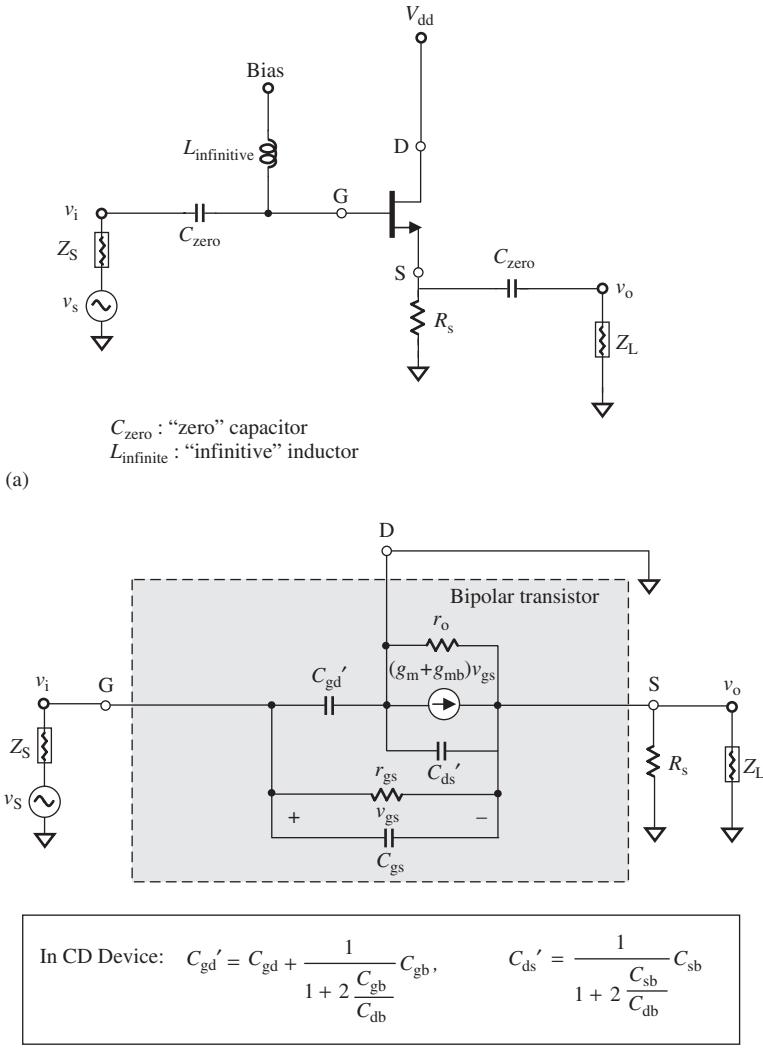
In low-frequency cases,

$$Z_{\text{out}}|_{\omega \rightarrow 0, R_s \rightarrow \infty} = R_d. \quad (5.360)$$

## 5.11 MOSFET WITH CD (COMMON DRAIN) CONFIGURATION

Figure 5.49 shows the equivalent circuit of a MOSFET with CD configuration, in which the drain is AC-grounded, the input port is the base, and the output port is the emitter of the bipolar transistor. The notation “CD” indicates that the drain is a common ground terminal of the input (gate) port and the output (source) port.

The modified model of the CD device with the combined capacitors  $C'_{gd}$  and  $C'_{ds}$  shown in Figure 5.49 is quite similar to the model of the CC device presented in Figure 5.22. On the basis of the similarity between these two models, most of the CD device's parameters can be translated from the corresponding expressions of CC parameters in terms of the similar relationships listed in Table 5.1. Comparing Figures 5.49 and 5.22, the capacitor  $C'_{ds}$  in the former does not correspond to the capacitor  $C_{cs}$  in the latter. Therefore, in the translation of equations from the bipolar CC device with



**Figure 5.49.** MOSFET with CD configuration,  $r_{gs} \rightarrow \infty$ . (a) Schematic of a MOSFET with CD configuration. (b) model of bipolar transistor with CC configuration.

TABLE 5.5. Special Relationships that must be taken care in the Translation from CC to CD Configuration

From Bipolar model CC	To MOSFET model CD
$r_\pi$	$r_{gs}$
$r_o$	$r_o/C'_{ds}$
$C_\pi$	$C'_{gs}$
$C_\mu$	$C'_{gd}$
$C_{cs}/R_c$	$R_d$
$g_m$	$(g_m + g_{mb})$
$\beta_o$	$(g_m + g_{mb})r_{gs}$

emitter degeneration to the MOSFET CS device with source degeneration, the special relationships listed in Table 5.5 must be taken care of.

### 5.11.1 Open-Circuit Voltage Gain $A_{v,CD}$ of a CD Device

The voltage gain  $A_v$  of a CD device is derived with the additional approximation that the capacitor  $C'_{gd}$  in Figure 5.49(b) is neglected.

The open-circuit voltage gain of a CD device can be translated from expression (5.204), that is,

$$A_{v,CD} = \frac{1}{1 + \frac{1}{R_s(g_m + g_{mb})} + \frac{(1+j\omega C_{ds'} r_o)}{r_b(g_m + g_{mb})}}. \quad (5.361)$$

In low-frequency cases,

$$A_{v,CD}|_{\omega \rightarrow 0} = \frac{1}{1 + \frac{1}{(g_m + g_{mb})} \frac{r_b + R_s}{r_b R_s}}. \quad (5.362)$$

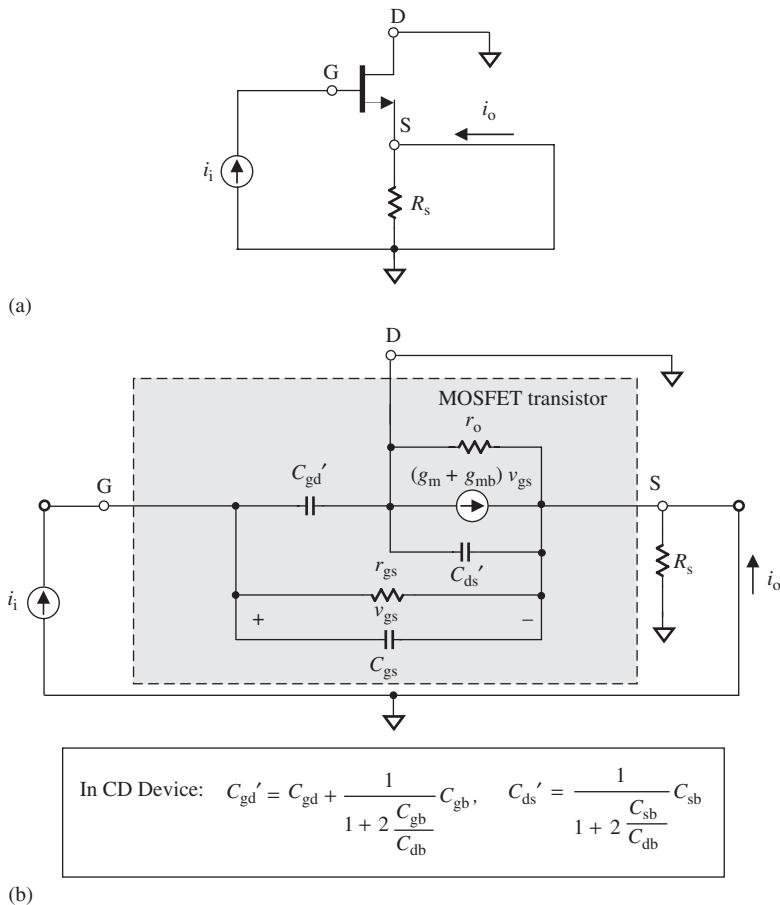
### 5.11.2 Short-Circuit Current Gain $\beta_{CD}$ and Frequency Response of a CD Device

Figure 5.50 is drawn for the calculation of the short-circuit current gain and frequency response of a MOSFET with a CD configuration, in which the additional approximation  $C'_{gd} \rightarrow 0$  is applied. It is similar to that of the CC device presented in Figure 5.24. Therefore, the short-circuit current gain and the frequency response can be translated from expression (5.214), that is,

$$\beta_{CD}(j\omega) = \frac{i_o}{i_i} = \frac{1 + (g_m + g_{mb})r_{gs} + j\omega C_{gs}r_{gs}}{1 + j\omega C_{gs}r_{gs}} \approx 1 + \frac{(g_m + g_{mb})}{j\omega C_{gs}}. \quad (5.363)$$

In low-frequency cases,

$$\beta_{CD}(j\omega)|_{\omega \rightarrow 0} \rightarrow \infty. \quad (5.364)$$



**Figure 5.50.** Schematic and equivalent circuit for the calculation of short-circuit current gain  $\beta$  of a CD device. (a) Output short-circuited for the calculation of current gain. (b) Modified model of a MOSFET with CD configuration,  $r_{gs} \rightarrow \infty$ .

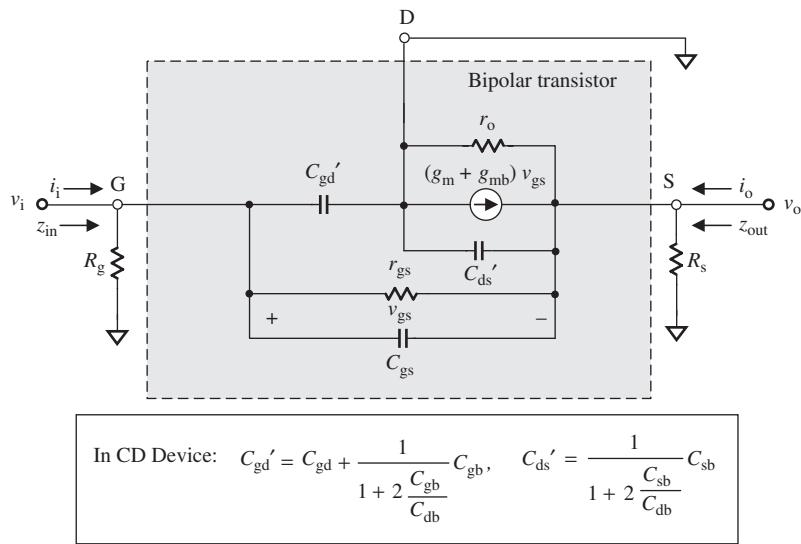
### 5.11.3 Input and Output Impedance of a CD Device

Figure 5.51 shows the schematic and its equivalent model of a MOSFET with a CD configuration for the calculation of the input and output impedances. It is similar to that for the CC device as shown in Figure 5.25. Therefore, the input and output impedances can be translated from the expression (5.230), that is,

$$Z_{in} = \frac{1 + \frac{R_s r_o}{R_s(1+j\omega C_{ds}') + r_o} [(g_m + g_{mb}) + j\omega C_{gs}]}{jC_{gs}\omega}. \quad (5.365)$$

In low-frequency cases,

$$Z_{in}|_{\omega \rightarrow 0} \rightarrow \infty. \quad (5.366)$$



**Figure 5.51.** Equivalent circuit for the calculation of input and output impedance of a CD device.

The output impedance of the CD device can be translated from (5.235), that is,

$$Z_{\text{out}} = \frac{1}{1 + \left[ \frac{(g_m + g_{mb} + j\omega C_{gs})}{(1 + j\omega C_{gs} R_g)} + j\omega C_{ds}' \right] r_o} r_o. \quad (5.367)$$

In low-frequency cases,

$$Z_{\text{out}}|_{\omega \rightarrow 0} = \frac{1}{1 + r_o(g_m + g_{mb})} r_o, \quad (5.368)$$

$$Z_{\text{out}}|_{\omega \rightarrow 0} \approx \frac{1}{g_m + g_{mb}}, \quad (5.369)$$

if

$$r_o(g_m + g_{mb}) \gg 1. \quad (5.370)$$

## 5.12 COMPARISON OF TRANSISTOR CONFIGURATION OF SINGLE-STAGE AMPLIFIERS WITH DIFFERENT CONFIGURATIONS

From Table 5.6 we can see the following:

- Devices with a CE or CS configuration have a high open-circuit voltage gain and a high short-circuit current gain; the input and output voltages are phase-shifted by  $180^\circ$ .
- Devices with a CB or CG configuration have the highest open-circuit voltage gain and lowest short-circuit current gain; their current gain is less than but close to 1.

TABLE 5.6. List of Open-Circuit Voltage and Short-Circuit Current Gain of a Device with Different Configurations in the very Low Frequency Cases, that is,  $\omega \rightarrow 0$

Configuration	Voltage gain, $A_v$	Current gain, $\beta$ or $A_i$
CE	$-g_m \frac{R_c r_o}{R_c + r_o}$	$\beta_o$
CB	$g_m \left(1 + \frac{1}{g_m r_o}\right) \frac{R_c r_o}{R_c + r_o}$	$\alpha_o$
CC	$\frac{1}{1 + \frac{1}{g_m} \left(\frac{1}{R_c} + \frac{1}{r_o}\right)}$	$1 + \beta_o$
CS	$-(g_m + g_{mb}) \frac{R_d r_o}{R_d + r_o}$	$\rightarrow \infty$
CG	$(g_m + g_{mb}) \left[1 + \frac{1}{(g_m + g_{mb}) r_o}\right] \frac{R_d r_o}{R_d + r_o}$	$\rightarrow 1$
CD	$\frac{1}{1 + \frac{1}{(g_m + g_{mb})} \left(\frac{1}{R_s} + \frac{1}{r_b}\right)}$	$\rightarrow \infty$

TABLE 5.7. Input and Output Impedances of a Device with Different Configurations in the very Low Frequency Cases, that is,  $\omega \rightarrow 0$

Configuration	Input impedance	Output impedance
CE	$r_\pi + r_b$	$(r_c + r_o) // R_c$
CB	$r_\pi \frac{1}{1 + (1 + g_m r_o) \frac{r_\pi}{r_o + r_c + R_c}}$	$R_c \frac{1}{1 + \frac{r_c + R_c}{r_\pi + (1 + \beta_o) r_o}}$
CC	$r_\pi + r_b + \beta_o R_e$	$\frac{1}{g_m} + \frac{R_b + r_b}{\beta_o}$
CS	$\rightarrow \infty$	$(r_c + r_o) // R_c \frac{r_o R_d}{r_o + R_d}$
CG	$\frac{1 + \frac{R_d}{r_o}}{(g_m + g_{mb}) + \frac{1}{r_o}}$	$R_d$
CD	$\rightarrow \infty$	$\frac{1}{(g_m + g_{mb}) + \frac{1}{r_o}}$

- Devices with a CC or CD configuration have the lowest open-circuit voltage gain and highest short-circuit current gain.

From Table 5.7 we can be seen that the following:

- Devices with a CE or CS configuration have high input and output impedances.
- Devices with a CB or CG configuration have the lowest input impedance and the highest output impedance. These are the very special features of a CB or CG device.
- Devices with a CC or CD configuration have the highest input impedance and lowest output impedance.

For the reader's convenience, the formulae of the transition frequency and the general expressions of input and output impedances are listed in Table 5.8.

TABLE 5.8. Transition Frequency and General Expressions for the Input and Output Impedance  
CE:

$$\omega_{T,CE} = \frac{g_m}{C_\pi + C_\mu}, \quad (5.64)$$

$$\omega_{\beta,CE} = \frac{1}{\beta_o} \frac{g_m}{C_\pi + C_\mu} = \frac{\omega_{T,CE}}{\beta_o}, \quad (5.67)$$

$$Z_{in} = \frac{r_b + r_\pi + j\omega C_\pi r_\pi r_b}{1 + j\omega C_\pi r_\pi}, \quad (5.70)$$

$$Z_{out} = \frac{r_c + r_o + jC_{cs}\omega r_c r_o}{R_c + r_c + r_o + jC_{cs}\omega r_o(r_c + R_c)} R_c. \quad (5.73)$$

CE with emitter degeneration:

$$Z_{in} = \frac{v_i}{i_b} = \frac{1}{1 + j\omega C_\pi r_\pi} \left( r_\pi + \frac{r_o(1 + j\omega C_{cs}R_c)(\beta_o + 1 + j\omega C_\pi r_\pi) + R_c(1 + j\omega C_\pi r_\pi)}{R_c + (r_o + R_e)(1 + j\omega C_{cs}R_c)} R_e \right), \quad (5.99)$$

$$Z_{out} = r_o \frac{1 + \left( \frac{1}{r_o} + g_m \right) \frac{r_\pi R_e}{r_\pi + R_e(1 + j\omega C_\pi r_\pi)}}{1 + \frac{r_o}{R_c}(1 + j\omega C_{cs}R_c) \left[ 1 + \left( \frac{1}{r_o} + g_m \right) \frac{r_\pi R_e}{r_\pi + R_e(1 + j\omega C_\pi r_\pi)} \right]}. \quad (5.118)$$

CB:

$$\omega = \omega_{\beta,CB} = \frac{g_m}{\alpha_o C_\pi}, \quad (5.149)$$

$$Z_{in} = \frac{v_i}{i_i} = R_e // \frac{r_\pi}{1 + \beta_o} \frac{1 + \frac{1}{r_o} \frac{r_c + R_c}{1 + j\omega C_{cs}(r_c + R_c)}}{1 + \frac{r_\pi}{1 + \beta_o} \left[ \frac{1}{r_o} + j\omega C_\pi + \frac{1}{r_o} \left( \frac{1}{r_\pi} + j\omega C_\pi \right) \frac{r_c + R_c}{1 + j\omega C_{cs}(r_c + R_c)} \right]}, \quad (5.172)$$

$$Z_{out} = R_c \frac{1}{1 + j\omega C_{cs}(r_c + R_c) + (r_c + R_c) \frac{(r_\pi + R_e + j\omega C_\pi R_e r_\pi)}{r_o r_\pi + R_e [r_\pi + r_o(\beta_o + 1) + j\omega C_\pi r_\pi r_o]}}. \quad (5.188)$$

CC:

$$Z_{in} = \frac{r_\pi + r_b(1 + jC_\pi \omega r_\pi) + (1 + \beta_o + jC_\pi \omega r_\pi) \frac{r_o + r_c + jC_{cs}\omega r_o r_c}{R_e + r_o + r_c + jC_{cs}\omega r_c(r_o + R_e)} R_e}{1 + jC_\pi \omega r_\pi}, \quad (5.230)$$

$$Z_{out} = \frac{r_\pi + (R_b + r_b)(1 + j\omega C_\pi r_\pi)}{(1 + \beta_o + j\omega C_\pi r_\pi) + (1 + j\omega C_{cs}r_c) \frac{r_\pi + (R_b + r_b)(1 + j\omega C_\pi r_\pi)}{r_c + r_o(1 + j\omega C_{cs}r_c)}}. \quad (5.235)$$

(continued)

TABLE 5.8. (Continued)

CS:

$$\omega_{T,CS} = \frac{g_m + g_{mb}}{C'_{gs} + C_{gd}}, \quad (5.310)$$

$$\omega = \omega_{\beta,CS} = \left( \frac{g_m + g_{mb}}{C'_{gs} + C_{gd}} \right) \cdot \frac{1}{\beta} = \frac{1}{\beta} \omega_{T,CS}, \quad (5.312)$$

$$Z_{in} = \frac{1}{jC_{in}\omega}, \quad (5.317)$$

$$Z_{out} = \frac{r_o}{R_d + r_o + jC_{out}\omega r_o R_d} R_d. \quad (5.318)$$

CS with emitter degeneration:

$$Z_{in} \approx \frac{1}{j\omega C'_{gs}} \left\{ 1 + \frac{r_o[(g_m + g_{mb}) + j\omega C'_{gs}] + j\omega C'_{gs}R_d(1 + j\omega C'_{ds}r_o)}{r_o + (R_d + R_s)(1 + j\omega C'_{ds}r_o)} R_s \right\}, \quad (5.322)$$

$$Z_{out} = \frac{r_o}{1 + j\omega C'_{ds}r_o} \frac{1 + \left( \frac{1 + j\omega C'_{ds}r_o}{r_o} + g_m + g_{mb} \right) \frac{R_s}{1 + j\omega C'_{gs}R_s}}{1 + \left[ 1 + \left( \frac{1 + j\omega C'_{ds}r_o}{r_o} + g_m + g_{mb} \right) \frac{R_s}{1 + j\omega C'_{gs}R_s} \right]}. \quad (5.335)$$

CG:

$$\omega = \omega_{\beta,CG} = \frac{g_m + g_{mb}}{C'_{gs}}, \quad (5.351)$$

$$Z_{in} = R_s // \frac{1}{(g_m + g_{mb})} \frac{1 + \frac{1}{r_o} \frac{R_d}{1 + j\omega C'_{gd}R_d}}{1 + \frac{1}{(g_m + g_{mb})} \left[ \frac{1}{r_o} + j\omega C'_{gs} + \frac{1}{r_o} j\omega C'_{gs} \frac{R_d}{1 + j\omega C'_{gd}R_d} \right]}, \quad (5.353)$$

$$Z_{in}|_{R_s \rightarrow \infty} = \frac{1}{(g_m + g_{mb})} \frac{1 + \frac{1}{r_o} \frac{R_d}{1 + j\omega C'_{gd}R_d}}{1 + \frac{1}{(g_m + g_{mb})} \left[ \frac{1}{r_o} + j\omega C'_{gs} + \frac{1}{r_o} j\omega C'_{gs} \frac{R_d}{1 + j\omega C'_{gd}R_d} \right]}. \quad (5.354)$$

CD:

$$Z_{in} = \frac{1 + \frac{R_s r_o}{R_s(1 + j\omega C'_{ds}) + r_o} [(g_m + g_{mb}) + j\omega C'_{gs}]}{jC_{gs}\omega}, \quad (5.365)$$

$$Z_{out} = \frac{1}{1 + \left[ \frac{(g_m + g_{mb} + j\omega C_{gs})}{(1 + j\omega C_{gs}R_g)} + j\omega C'_{ds} \right] r_o} r_o. \quad (5.367)$$

## FURTHER READING

Gray PR, Hurst PJ, Lewis SH, Meyer RG. *Analysis and Design of Analog Integrated Circuits*. 4th ed. New York: John Wiley & Sons, Inc; 2001.

Richard CHL. *Key Issues in RF/RFIC Circuit Design*. Beijing: Higher Education Press; 2005.

Smith J. *Modern Communication Circuits*. New York: McGraw-Hill Publishing Company; 1986.

## EXERCISES

1. What is Miller effect? How does it impact circuit design?
2. Qualitatively compare the input impedances between CE, CB, and CC configuration when  $\omega \rightarrow 0$ .
3. Qualitatively compare the output impedances between CE, CB, and CC configuration when  $\omega \rightarrow 0$ .
4. Qualitatively compare the input impedances between CS, CG, and CD configuration when  $\omega \rightarrow 0$ .
5. Qualitatively compare the output impedances between CS, CG, and CD configuration when  $\omega \rightarrow 0$ .
6. Qualitatively compare the input impedances between CE–CB, CB–CC, and CC–CE configuration when  $\omega \rightarrow 0$ .
7. Qualitatively compare the output impedances between CE–CB, CB–CC, and CC–CE configuration when  $\omega \rightarrow 0$ .
8. Qualitatively compare the input impedances between CS–CG, CD–CD, and CD–CS configuration when  $\omega \rightarrow 0$ .
9. Qualitatively compare the output impedances between CS–CG, CD–CD, and CD–CS configuration when  $\omega \rightarrow 0$ .

## ANSWERS

1. Miller's theorem states that if a device has open-circuit voltage gain  $A_v$  between the input and the output nodes, then the feedback element with impedance  $Z$  connected between these two nodes can be replaced by an equivalent element connected to ground at each node. If

$$Z = \frac{1}{j\omega C},$$

then

$$Z_{mi} \approx \frac{Z}{|A_v|} = \frac{1}{j\omega C |A_v|},$$

$$Z_{mo} \approx Z = \frac{1}{j\omega C},$$

$$C_{mi} \approx |A_v|C,$$

$$C_{mo} \approx C.$$

The existence of a feedback capacitor  $C$  in a device is equivalent to the addition of an input equivalent Miller capacitor  $C_{mi}$ , which is a capacitor with  $A_v$  times the feedback capacitance at input terminal in parallel, plus an output equivalent Miller capacitor  $C_{mo}$ , which is a capacitor with the feedback capacitance at the output terminal in parallel ( $A_v$  being the open-circuit voltage gain of the device).

This equivalent relationship is called the *Miller effect*.

The main impact to the circuit design is the existence of the input equivalent Miller capacitor  $C_{mi}$ , which is a capacitor with  $A_v$  times the feedback capacitance. It slows the rise time of a pulse so that it makes the device unable to work well in the high digital data rate case.

2.		CE	CB	CC
	Input impedance	Mid	Low	High
3.		CE	CB	CC
	Output impedance	Mid	High	Low
4.		CS	CG	CD
	Input impedance	High	Low	High
5.		CS	CG	CD
	Output impedance	Mid	High	Low
6.		CE–CB	CB–CC	CC–CE
	Input impedance	Mid	Low	High
7.		CE–CB	CB–CC	CC–CE
	Output impedance	High	Low	Mid
8.		CS–CG	CG–CD	CD–CS
	Input impedance	High	Low	High
9.		CS–CG	CG–CD	CD–CS
	Output impedance	High	Low	Mid



---

# 6

---

## IMPEDANCE MEASUREMENT

---

### 6.1 INTRODUCTION

As mentioned in Chapter 1, the task of a digital circuit block in a communication system is status transport or manipulation. The so-called status of a digital signal is characterized by “0” or “1.” Direct observation of the signal waveform, including voltage at a node or current flowing through a part, is the best way to judge the performance of the circuit block, including its sequence of digits, repetition frequency, voltage level, rise and fall time, jitter and stability, and so on. Therefore, the main test equipment to be used is the oscilloscope, since a digital circuit block is tested in the time domain.

On the contrary, the task of an RF circuit block in a communication system is power transport or manipulation. In order to ensure effective transport and manipulation of power, the input and output impedances of the RF block must be matched with the output impedance of the previous block and the input impedance of the following block, respectively. In addition, the frequency bandwidth of the block must also be taken care of. The observation of the impedance matching state, power intensification, or attenuation over the entire frequency bandwidth is the best way to judge the performance of the circuit block, including the genuineness of the product, noise, nonlinearity, and so on. Therefore, the main test equipment to be used is the network analyzer, since the testing of the RF circuit block is carried out in the frequency domain.

Impedance measurement takes the first priority in RF circuit testing. It is the starting point for the implementation of an impedance matching network. Also, it is a powerful means to examine the performance of an RF block. The performance of power transport

or manipulation of an RF block is directly related to how well its impedance matching has been taken care of.

In the case of a small signal, the impedance is usually tested by a network analyzer. In the case of a large signal, the impedance is usually tested with the assistance of a circulator. The impedance of discrete parts can be tested by a network analyzer as well as an impedance meter.

The RF circuit is always tested through power measurement by various test equipment, such as the network analyzer, spectrum analyzer, power meter, and so on. Voltage measurement by an oscilloscope is basically not helpful to characterize the performance of an RF circuit. Almost all the parameters that specify the performance, such as the power gain, noise figure, and intercept points, are computed through power measurements. However, in RF circuit design, voltage measurement is not to be neglected. As matter of fact, there are two kinds of voltage measurement. One is the measurement of the scalar voltage by an oscilloscope as mentioned above. A scalar voltage is the resultant of the incident and reflected voltages at one node. The other measurement is that of the vector voltage by a vector voltmeter. RF designers are not interested in the former, but are in the latter. The vector voltmeter can distinguish between the incident and reflected voltages at one node instead of just the resultant voltage. As a matter of fact, the operating principle of a network analyzer is based on the measurement of the vector voltage, in which the incident and reflected voltages are measured simultaneously.

## 6.2 SCALAR AND VECTOR VOLTAGE MEASUREMENT

### 6.2.1 Voltage Measurement by Oscilloscope

An oscilloscope can test or measure the voltage at any node in a circuitry and display its waveform on its screen. This intuitive feeling is very nice to everyone. As electronic technology moves forward, the frequency response of an oscilloscope has approached up to tens of gigahertz today, and more than 10 waveforms can be displayed on the same screen simultaneously. It is indeed a powerful tool in digital circuit design.

This powerful tool raises several questions: why is an oscilloscope not to be found in any advanced RF circuit test laboratory? Why do RF circuit designers never use the oscilloscope in RF circuit testing?

The answer can be explained by Figure 6.1. For simplicity, the actual layout on a PCB (printed circuit board) is replaced by a simple schematic. There are three circuit branches coming together at node P: a branch with a MOSFET (metal–oxide–semiconductor field-effect transistor)  $M$ , a branch with an inductor  $L$ , and a branch with a capacitor  $C$ , while the rest of the circuitry is connected between the capacitor  $C$  and the output SMA (SubMiniature version A) connector in series. The impedance of a probe is usually quite high, so it will not disturb the performance of the circuitry when it touches the test node P. When it senses the voltage  $V$  at node P, the oscilloscope magnifies and displays the voltage on its screen as a waveform in the time domain. The waveform at node P is a resultant voltage contributed by the three circuit branches. For a digital circuit block, this does indeed describe the “status” at node P, either that of high voltage, corresponding to 1, or that of low voltage, corresponding to 0.

If the input impedance of the oscilloscope is not high enough, a buffer with high input and output impedances can be inserted between the oscilloscope and the DTU (desired test unit) as shown in Figure 6.2.

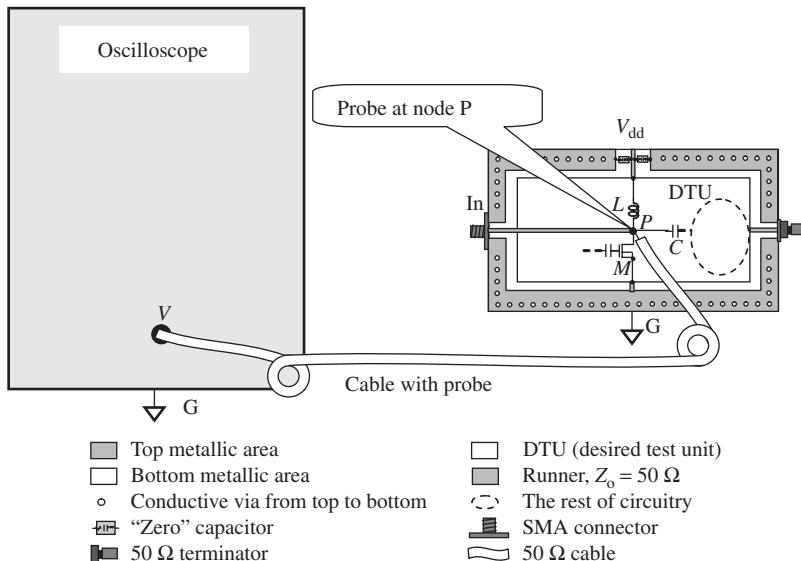


Figure 6.1. Voltage measured at node P by an oscilloscope or a regular voltmeter.

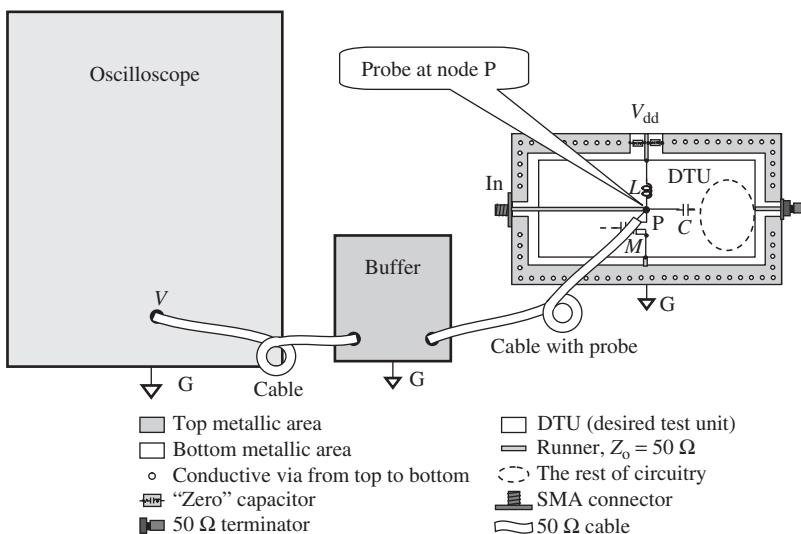


Figure 6.2. Voltage measured at node P with assistance of a buffer.

The oscilloscope can read only the resultant voltage; it cannot distinguish between the incident and reflected voltages in the cable with the probe. Such a test or measurement is nondirective.

In RF circuit design, such a test is insufficient. The voltage measurement for an RF block must distinguish between the incident and reflected voltages so that the impedance can be computed. Therefore, voltage measurement by an oscilloscope is not what an RF circuit designer asks for, which in most cases is of limited usefulness in RF circuit design.

### 6.2.2 Voltage Measurement by Vector Voltmeter

Voltage can also be tested or measured by a vector voltmeter. The vector voltmeter can distinguish between the incident and the reflected voltages at one node, so the test or measurement is directional.

Figure 6.3 shows the test setup. In addition to a vector voltmeter, a signal generator and a bidirectional coupler are also needed. The signal generator provides a voltage signal to the input of the bidirectional coupler. The output of the bidirectional coupler feeds the voltage signal to the input SMA connector of the PCB with the DTU.

An oscilloscope simply magnifies and displays the resultant voltage, which is sensed by a cable with a probe at node P; this is a passive test or measurement. On the contrary, in the voltage test or measurement by a vector voltmeter, a voltage signal is provided by the signal generator and is fed to the test node P. It is an active test or measurement.

Another remarkable aspect is that, in the voltage test by a vector voltmeter, a microstrip line with  $50 \Omega$  of characteristic impedance must be connected from the input SMA connector to the tested node P, so that the vector voltmeter can sense the incident voltage from the signal generator at port A and the reflected voltage returned from the tested node P at port B without considerable additional attenuation. On the contrary, in testing or measurement using an oscilloscope, the microstrip line with  $50 \Omega$  is not needed. Its high-impedance probe touches the tested node P directly.

Now let us take a look at the various voltages around the test node P in a little more detail. Figure 6.4 depicts only the DTU with its various incident and reflected voltages.

There are two resultant voltages on the  $50\text{-}\Omega$  microstrip line: the resultant incident voltage  $v_{\text{in}}$  and the resultant reflected voltage  $v_{\text{ref}}$ , that is

$$v_{\text{in}} = v_{\text{in}1} + v_{\text{in}2} + v_{\text{in}3}, \quad (6.1)$$

$$v_{\text{ref}} = v_{\text{ref}1} + v_{\text{ref}2} + v_{\text{ref}3}, \quad (6.2)$$

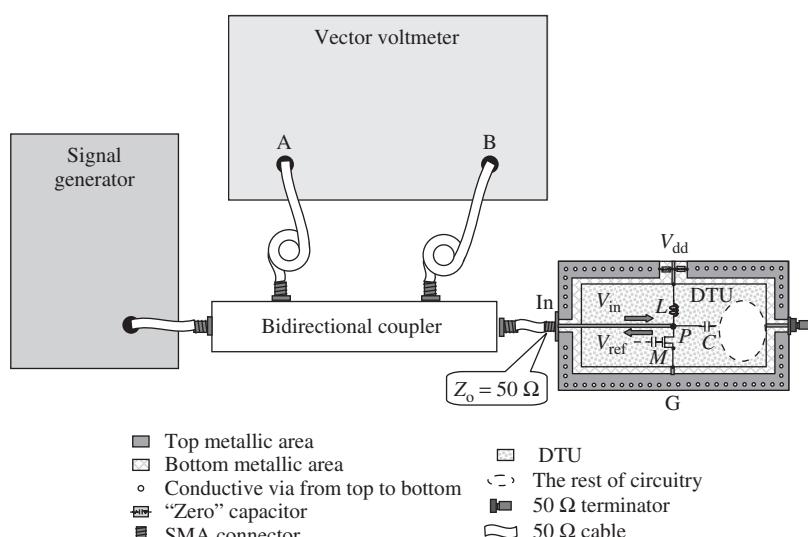


Figure 6.3. An example of a test PCB and the test setup.

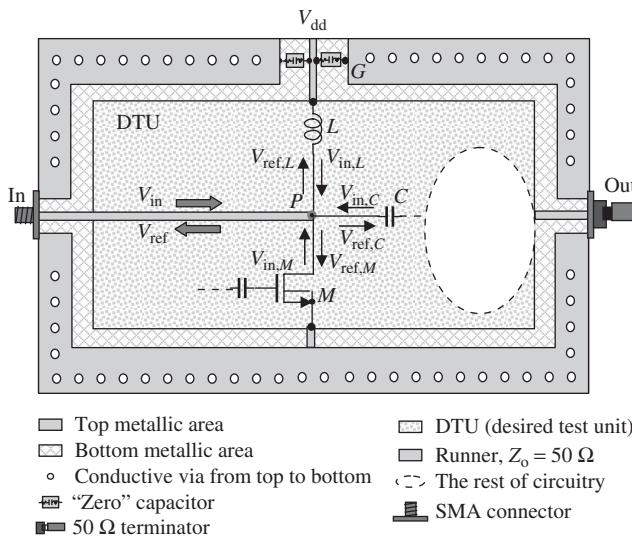


Figure 6.4. Various incident and reflected voltages at node P.

where

- $v_{in}$  = the incident voltage at node P,
- $v_{ref}$  = the reflected voltage at node P,
- $v_{in,L}$  = the incident voltage from branch L at node P,
- $v_{ref,L}$  = the reflected voltage to branch L at node P,
- $v_{in,C}$  = the incident voltage from branch C at node P,
- $v_{ref,C}$  = the reflected voltage to branch C at node P,
- $v_{in,M}$  = the incident voltage from branch M at node P, and
- $v_{ref,M}$  = the reflected voltage to branch M at node P.

The vector voltmeter can distinguish between and read the resultant incident and reflected voltage  $v_{in}$  and  $v_{ref}$ , respectively, but not the individual incident and reflective voltages  $v_{in,L}$ ,  $v_{ref,L}$ ,  $v_{in,C}$ ,  $v_{ref,C}$ ,  $v_{in,M}$ , or  $v_{ref,M}$ . However, this is enough to calculate the impedance at node P, such as

$$\Gamma = \frac{v_{ref}}{v_{in}}, \quad (6.3)$$

$$z = \frac{1 + \Gamma}{1 - \Gamma}, \quad (6.4)$$

where

- $\Gamma$  = the voltage reflection coefficient at node P, and
- $z$  = the normalized impedance looking into the node P.

## 6.3 DIRECT IMPEDANCE MEASUREMENT BY A NETWORK ANALYZER

### 6.3.1 Direction of Impedance Measurement

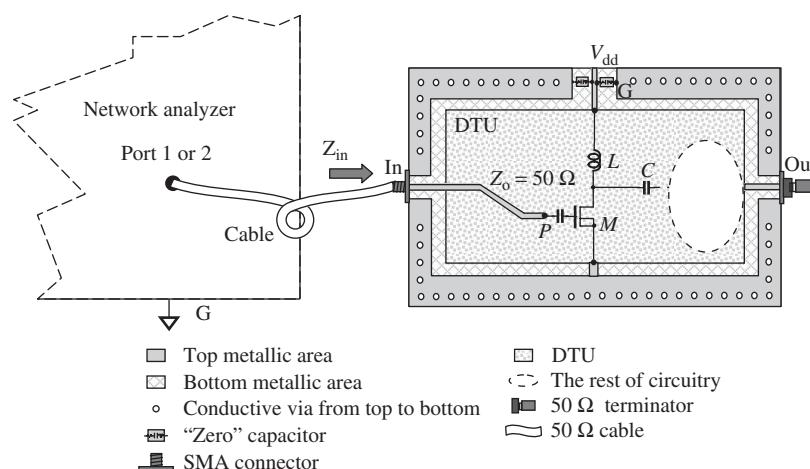
As mentioned above, the operation of the network analyzer is based on the operating principle of the vector voltmeter. The impedance is calculated from the directional voltage

measurement. The direction of the voltage is referenced to the node: that is whether the voltage is entering or reflected from the node.

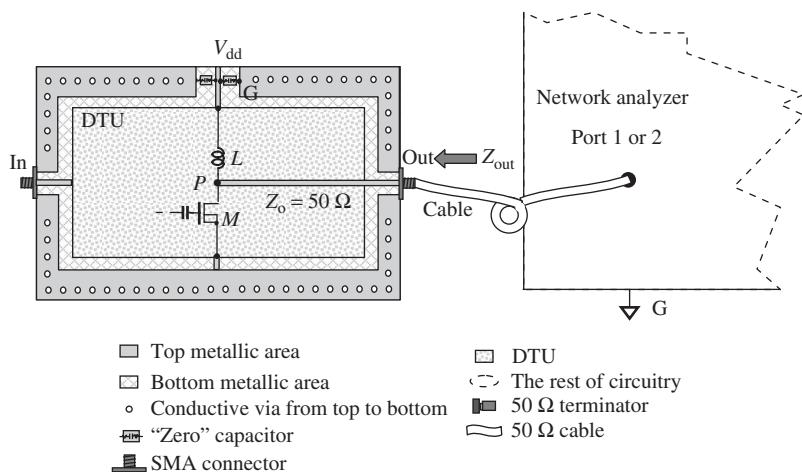
The direction of impedance can be based on another reference: that is, referenced to the input or output of a circuit stage, instead of a node. Figures 6.5–6.7 show three impedance measurements at node P.

Figure 6.5 shows an input impedance measurement. The impedance  $Z_{in}$  is looking into the node P, which is connected to only the input subcircuit branch.

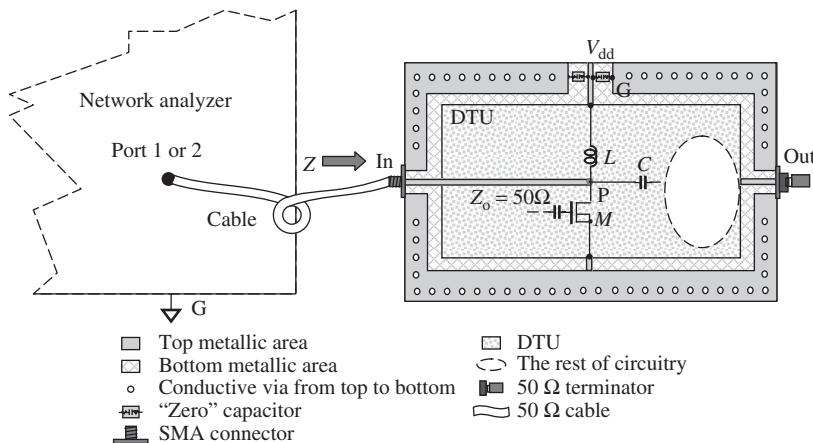
Figure 6.6 shows an output impedance measurement. The impedance  $Z_{out}$  is looking into the node P, which is connected to two output subcircuit branches. One is the inductor L between the drain of the MOSFET and the DC power supply  $v_{dd}$ , and the other is the MOSFET M. They are connected together in parallel.



**Figure 6.5.** Directional impedance measurement for input impedance  $Z_{in}$ . At node P, only input subcircuit is connected.



**Figure 6.6.** Output impedance measurement for output impedance  $Z_{out}$ . At node P, two subcircuit branches are connected in parallel.



**Figure 6.7.** Nondirectional impedance measurement for resultant impedance  $Z$ . At node P, three subcircuit branches are connected in parallel.

Figure 6.7 shows a nondirectional impedance measurement. The impedance  $Z$  is looking into the node P, which is connected to three subcircuit branches. The first one is the inductor  $L$  between the drain of the MOSFET and the DC power supply  $v_{dd}$ , the second is the MOSFET  $M$ , and the third is the branch with the capacitor  $C$  at the output. They are connected together in parallel.

### 6.3.2 Advantage of Measuring $S$ Parameters

In digital circuit design, or in the early stages of RF circuit design, there are many sets of parameters to describe a two-port network, including impedance  $Z$  parameters, admittance  $Y$  parameters, and hybrid  $h$  parameters. They are outlined in Figure 6.8 with both matrix and block diagrams.

Parameters	Matrix	Coefficients	Block diagram
$Z$	$\begin{aligned} v_1 &= z_{11}i_1 + z_{12}i_2 \\ v_2 &= z_{21}i_1 + z_{22}i_2 \end{aligned}$	$\begin{aligned} z_{11} &= v_1/i_1 \mid i_2 = 0 \\ \dots & \end{aligned}$	
$Y$	$\begin{aligned} i_1 &= y_{11}v_1 + y_{12}v_2 \\ i_2 &= y_{21}v_1 + y_{22}v_2 \end{aligned}$	$\begin{aligned} y_{11} &= i_1/v_1 \mid v_2 = 0 \\ \dots & \end{aligned}$	
$h$	$\begin{aligned} v_1 &= h_{11}i_1 + h_{12}v_2 \\ i_2 &= h_{21}i_1 + h_{22}v_2 \end{aligned}$	$\begin{aligned} h_{11} &= v_1/i_1 \mid v_2 = 0 \\ h_{12} &= v_1/v_2 \mid i_1 = 0 \\ \dots & \end{aligned}$	
$S$	$\begin{aligned} b_1 &= s_{11}a_1 + s_{12}a_2 \\ b_2 &= s_{21}a_1 + s_{22}a_2 \end{aligned}$	$\begin{aligned} s_{11} &= b_1/a_1 \mid a_2 = 0 \\ \dots & \end{aligned}$	

**Figure 6.8.** Various parameters that characterize a two-port network.

The first three parameters in Figure 6.8 are based on the measurements of voltage and currents at the input or output terminals. In all three sets of parameters, inaccuracy occurs because of the measurement of their coefficients, as shown in the third column of Figure 6.8. These coefficients must be measured or tested under conditions of either open-circuit or short-circuit, which is never absolutely possible in reality. In the RF frequency range, the voltage of a short-circuited terminal is not absolutely equal to zero because of the isolation problem between the terminal and the ground. For the same reason, the current of an open-circuit terminal is not absolutely equal to zero as well.

The fourth set of parameters in Figure 6.8, called *spread S parameters*, was adopted in the second half of the twentieth century. Instead of the resultant voltage or current, this set distinguishes between the incident or reflected voltage or current at the input or output terminals in a two-port network. This idea, in fact, comes from the concept of a traveling wave. The incident and reflected voltage or current is fully dependent on the impedances of the source and the load. In other words, the voltage, current, and impedance at a node or terminal are the directional parameters.

By means of the *S* parameters, the conditions of open- or short-circuit in the determination of their coefficients of the matrix are not required, and therefore, the inaccuracy introduced in those three sets of parameters mentioned above is eliminated. Today, the *S* parameters are widely applied in RF circuit design, though other parameter sets are still applied in digital circuit design. The *S* parameters of a two-port network are measured by a network analyzer; hence, the network analyzer has become the most important test equipment in an RF circuit design laboratory.

Very often, RF designers read or calculate the input or output impedances of a two-port network directly from a conversion of *S* parameters, either  $S_{11}$  or  $S_{22}$ , through the relations

$$z_{\text{in}} = \frac{1 + S_{11}}{1 - S_{11}}, \quad (6.5)$$

$$z_{\text{out}} = \frac{1 + S_{22}}{1 - S_{22}}. \quad (6.6)$$

It should be noted that equations (6.5) and (6.6) are approximate expressions valid only under certain conditions, which will be discussed in the following sections.

### 6.3.3 Theoretical Background of Impedance Measurement by *S* Parameters

In RF laboratories, the impedance of a basic part, such as a capacitor, inductor, or resistor, or the impedance of a block, subsystem, or an entire system can be measured by either an impedance meter or a network analyzer. In the simulation phase, the impedance measurement for a DTU is usually executed by a network analyzer. In the early stages of development, the network analyzer was used for low-power measurements, based on the principle of small signal measurement. Today, the network analyzer is able to test or measure a two-port network with high or large signals. This is, however, outside the scope of this book. In the following discussion, the impedance measured directly by the network analyzer through *S* parameters is based on the principle of small signal measurement.

The network analyzer measures the *S* parameter at one or two ports, and then the impedance is calculated from the *S* parameters or directly read from the Smith Chart.

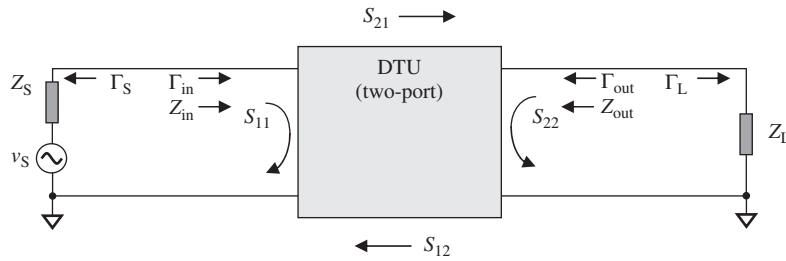


Figure 6.9.  $S$  parameters and voltage reflection coefficients of a two-port network.

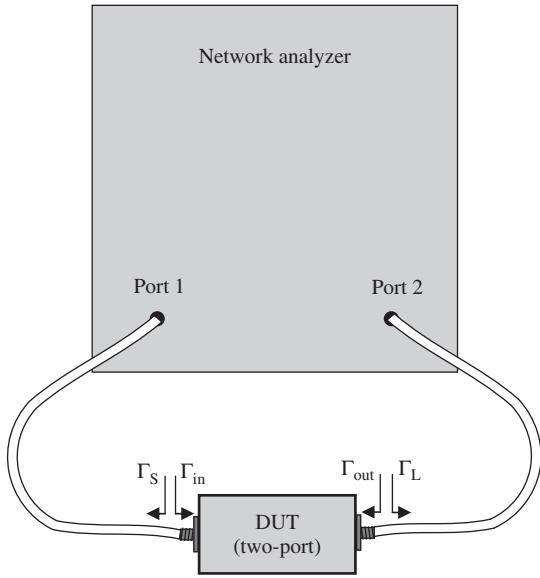


Figure 6.10. Impedance measured by network analyzer.

Figure 6.9 shows the  $S$  parameters and voltage reflection coefficients of a two-port network, and Figure 6.10 shows the impedance measured by a network analyzer.

Theoretically, the relations between the input and output voltage reflection coefficients  $\Gamma_{\text{in}}$  and  $\Gamma_{\text{out}}$  and the input and output  $S$  parameters  $S_{11}$  and  $S_{22}$  are

$$\Gamma_{\text{in}} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L}, \quad (6.7)$$

$$\Gamma_{\text{out}} = S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1 - S_{11}\Gamma_S}. \quad (6.8)$$

If

$$S_{12} = 0, \quad (6.9)$$

or

$$S_{21} = 0, \quad (6.10)$$

or

$$\Gamma_S = \Gamma_L = 0. \quad (6.11)$$

then,  $\Gamma_{in}$ ,  $\Gamma_{out}$  would be equal to  $S_{11}$ ,  $S_{22}$  respectively, that is

$$\Gamma_{in} = S_{11}, \quad \Gamma_{out} = S_{22}. \quad (6.12)$$

Condition (6.9) or (6.10) implies that the DTU is in an ideal state of either forward or backward isolation, which is usually not true in reality. Relation (6.11) then becomes the necessary and sufficient condition of the approximations in (6.12). Condition (6.11) means that the cables connected to port 1 and port 2 of the network analyzer must be well calibrated. This condition (6.11) enables us to obtain the input and output impedances  $Z_{in}$  and  $Z_{out}$  simply converted from the  $S_{11}$  and  $S_{22}$  measurements, respectively, as shown in Figure 6.10:

$$z_{in} = \frac{1 + \Gamma_{in}}{1 - \Gamma_{in}} = \frac{1 + S_{11}}{1 - S_{11}}, \quad (6.13)$$

$$z_{out} = \frac{1 + \Gamma_{out}}{1 - \Gamma_{out}} = \frac{1 + S_{22}}{1 - S_{22}}. \quad (6.14)$$

These values can be directly read from the Smith Chart displayed on the screen of the network analyzer.

### 6.3.4 S Parameter Measurement by Vector Voltmeter

The test fixture shown in Figure 6.11 consists of three portions. The upper portion  $X_1 - Y_1$  tests the  $S$  parameters for a transistor. A bipolar transistor is soldered on the pads marked B, E, and C, which denote the base, emitter, and collector, respectively. A MOSFET can be also tested with this fixture if its G, S, and D (which denote the gate, source, and drain) are instead soldered at the B, E, and C pads on the test fixture, respectively. The intermediate portion  $X_2$  is for short-circuited calibration, while the bottom portion  $X_3 - Y_3$  is for the double length line calibration. It should be noted that the characteristic impedance of all the runners connected to the SMA connector must be  $50 \Omega$ .

The bias  $T_1$  combines the DC voltage  $v_1$  from the DC power supply 1 and the RF signal  $v_S$  from the signal generator, and then delivers them to the base of the transistor through the bidirectional coupler 1. The bias  $T_2$  passes the DC voltage  $v_2$  from the DC power supply 2 and then delivers it onto the collector of the transistor through the bidirectional coupler 2. A  $50\Omega$  terminator must be connected to the bias  $T_2$  output terminal.

The vector voltmeter is calibrated as follows:

1. *Short-Circuit Calibration.* Port B of the vector voltmeter is connected to  $B_1$  of the bidirectional coupler 1, and the output terminal X of the bidirectional coupler 1 is connected to the input terminal  $X_2$  of the intermediate portion on the test fixture.

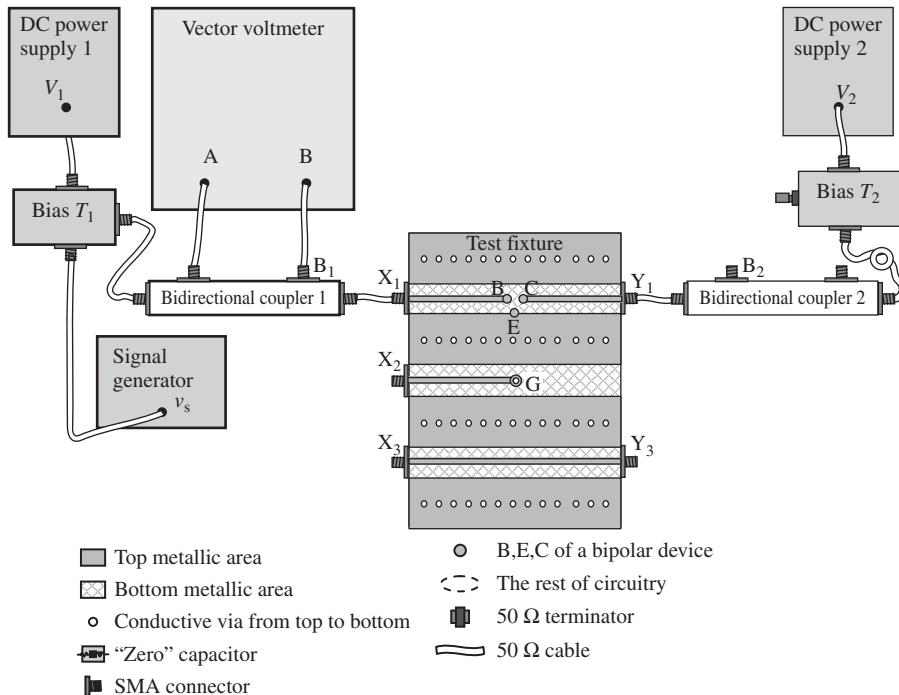


Figure 6.11. Setup of  $S$  parameter test for a transistor by a vector voltmeter.

Then,  $1 \cdot 180^\circ$  (Relative magnitude of voltage = 1, Relative phase of voltage =  $180^\circ$ ) must be indicated in the vector voltmeter after appropriate adjustment.

2. *Double Length Line Calibration.* Port B of the vector voltmeter is connected to  $B_2$  of the bidirectional coupler 2. The output terminal X of the bidirectional coupler 1 is connected to  $X_3$ , which is the input terminal of bottom portion on the test fixture, whose output terminal is connected to the input terminal of the bidirectional coupler 2. Then,  $1 \cdot 0^\circ$  (Relative magnitude of voltage = 1, Relative phase of voltage =  $0^\circ$ ) must be indicated in the vector voltmeter after an appropriate adjustment.

After the above calibrations have been done, the test is conducted with the following procedures step by step:

1. Set the reference RF input power of the signal generator at the required level. Set  $v_1$  and  $v_2$  of the DC power supplies 1 and 2 to the required value, and the base of the transistor is connected to  $X_1$  and the collector of the transistor is connected to  $Y_1$ .
2. Connect port B of the vector voltmeter to  $B_1$  of the bidirectional coupler 1, and read  $S_{11}$ .
3. Connect port B of the vector voltmeter to  $B_2$  of the bidirectional coupler 2, and read  $S_{21}$ .

4. Exchange  $v_1$  and  $v_2$  of DC power supplies 1 and 2, and reverse the test fixture: that is, connect the base of the transistor to  $Y_1$  and the collector of the transistor to  $X_1$ .
5. Connect port B of the vector voltmeter to  $B_1$  of the bidirectional coupler 1, and read  $S_{22}$ .
6. Connect port B of the vector voltmeter to  $B_2$  of the bidirectional coupler 2, and read  $S_{12}$ .

### 6.3.5 Calibration of the Network Analyzer

The network analyzer is a powerful tool in the measurement of an RF circuit block or system; calibration is an important step before the test or measurement is conducted.

Figure 6.12 depicts the layout for a DTU on a PCB. Its input terminal is point A and its output terminal is point B. For simplicity, the detailed layout of the DTU is replaced by a blank rectangular block with the DC power supply terminal  $v_{dd}$  and “zero” capacitors. At the input and output ports, a microstrip line leads from the input and output SMA connector to the circuit block. The characterized impedance of the microstrip line is  $50 \Omega$ .

One can carry out calibration by means of the standard calibration kit provided by the manufacturer. There are four basic calibration procedures: “open,” “short,” “ $50 \Omega$ ,” and “through.” It should be noted that the calibration is “on the air” without considering of the effect of the test environment and the tested PCB. The calibration accuracy may decline in the high RF frequency range.

Instead of using the standard calibration kit provided by the manufacturer, the RF design engineer may develop a custom calibration kit for more precise measurement. Figure 6.13 shows a set of self-supporting calibration kits: with the same PCB adapted for the DTU as shown in Figure 6.12, there are four subcalibration kits for open, short,  $50 \Omega$ , and through calibration purposes. Each subcalibration kit has the same grounding pattern and size as the DTU. As shown in Figures 6.12 and 6.13, the input terminal is calibrated at point A and the output terminal is calibrated at point B. By this method, the calibrations represent the actual environment of the DTU better than the standard calibration kit.

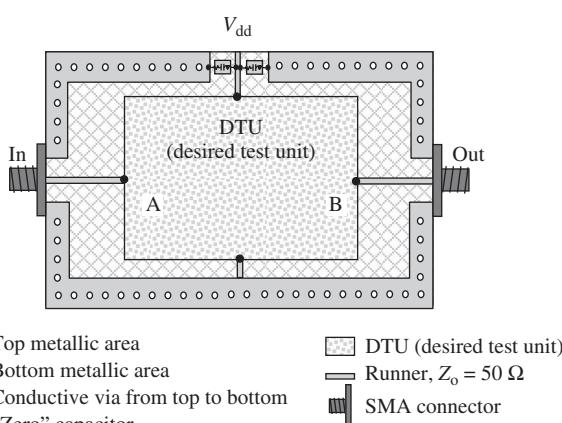
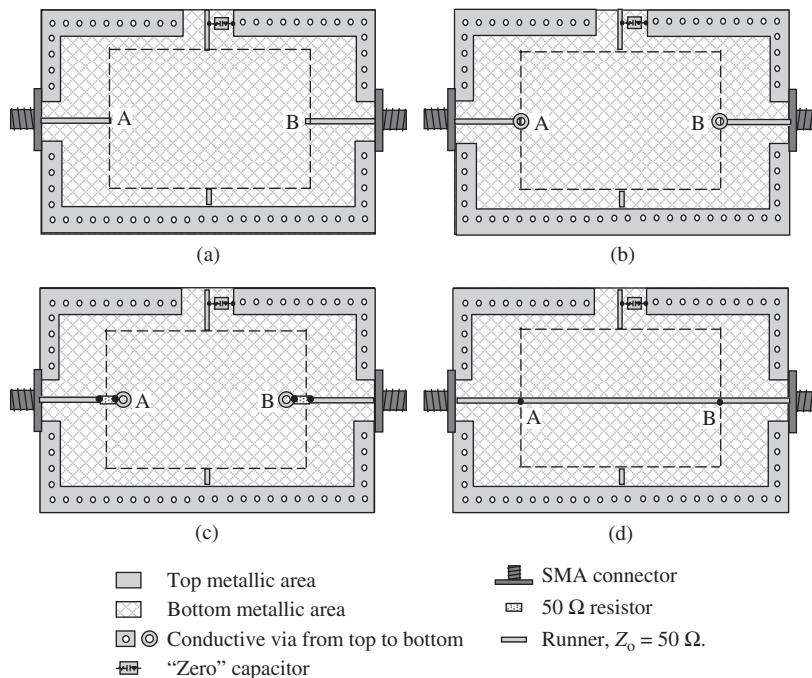


Figure 6.12. Layout of test PCB.



**Figure 6.13.** Layout of a self-supporting calibration kit. (a) Calibration kit: open. (b) Calibration kit: short. (c) Calibration kit:  $50 \Omega$ . (d) Calibration kit: through.

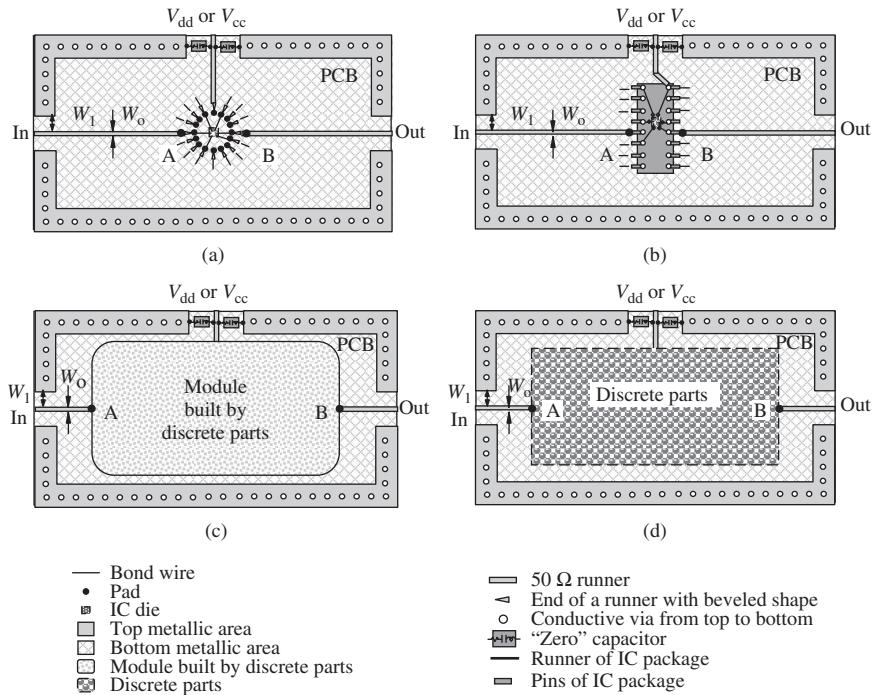
Experiments indicate that the values of measured values of  $S_{11}$  and  $S_{22}$  could differ by a couple of decibels at the frequency range of gigahertz when calibrations are made by the standard calibration kit and by the self-supporting calibration kit.

It should be noted that in order to ensure the constancy of the  $50\text{-}\Omega$  resistor within the desired frequency bandwidth, a combination of resistors might be necessary. For instance, a  $50\text{-}\Omega$  resistor could be substituted by two  $50\text{-}\Omega$  resistors connected in series and then connected with another  $100\text{-}\Omega$  resistor in parallel. This combination of multiple resistors tends to cancel the frequency variation of the resistance in the resulting resistor. Empirically, the value of  $50\ \Omega$  can be kept unchanged by a combination of multiple resistors from DC up to 6 GHz.

There are four basic types of PCB:

1. COB (chip on board). Cut the block or system from a wafer as an IC die first, and then bond the IC die onto a test PCB for testing.
  2. POB (package on board). Cut the block or system from a wafer as an IC die first, then package it, and finally use it as a part on the test PCB for testing.
  3. MOB (Module on board On Board as one part). The module is built by discrete parts and then packaged as one part on the test board.
  4. DOB (Module directly built by discrete parts on board).

Another method is to test the RFIC on the wafer directly in an IC probe station. The IC die could be either an individual block, multiple blocks, or an SOC (system-on-one-chip); the test PCB, of course, is not necessary when testing is conducted directly on the wafer.



**Figure 6.14.** Basic types of test PCB. (a) COB (chip on board). (b) POB (IC package on board). (c) MOB (module on board). (d) DOB (discrete parts on board).

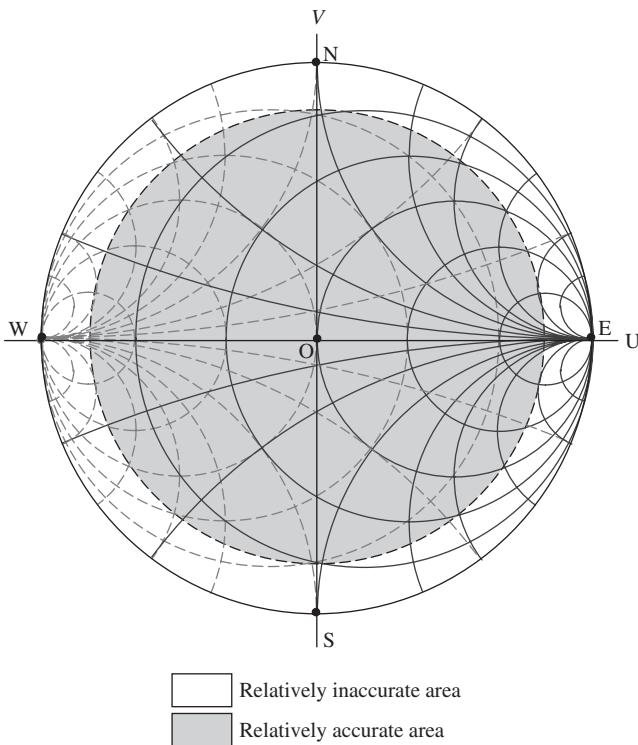
Figure 6.14 illustrates the four basic types of PCB, in which one subfigure contains only one type of PCB built by either RFIC or RF discrete parts. The test PCBs shown in Figure 6.14(a) and (b) contain an RFIC die and RFIC package, respectively, while the test PCBs shown in Figure 6.14(c) and (d) contain an RF module or RF discrete parts. For simplicity, only the main part and the main terminals, including the input, output, and DC power supply, are drawn, while other additional parts are neglected since we are only concerned with the main part and the main terminals.

If the PCB is tested by the network analyzer, calibration must be conducted up to points A and B as shown in Figure 6.14, but not just up to the input and output terminal. Unlike the MOB and DOB types of PCB shown in Figure 6.14, in the COB and POB types of PCB the calibration points A and B are quite far from the input and output terminals. Using the calibration kit provided by the manufacturer, calibration may be reluctantly accepted for the MOB and DOB types of PCB, where the calibration points are close to the input and output terminals; however, this is definitely not acceptable for the COB and POB boards. Instead of the standard calibration kit, a self-supporting calibration kit must be designed and applied for calibration work.

## **6.4 ALTERNATIVE IMPEDANCE MEASUREMENT BY NETWORK ANALYZER**

#### 6.4.1 Accuracy of the Smith Chart

It should be noted that the accuracy of a reading from a Smith Chart depends on the location of the reading. Figure 6.15 shows the relatively inaccurate and accurate areas



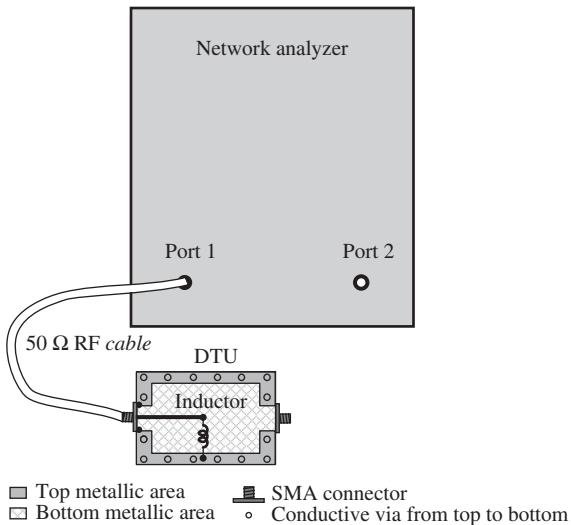
**Figure 6.15.** Relatively accurate and inaccurate areas of a Smith Chart.

on a Smith Chart. In areas where the impedance is not too low and not too high, the accuracy of an impedance reading is acceptable and reliable. However, in areas where the impedance is very low or very high, the accuracy of the impedance reading might be questionable.

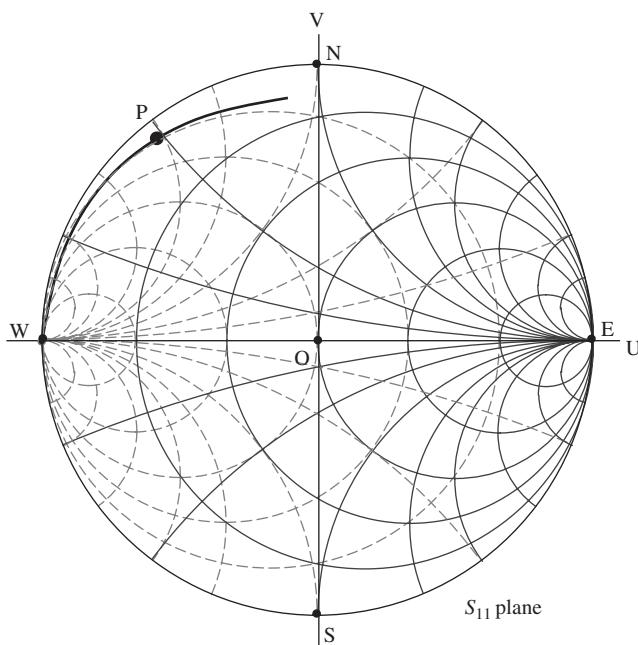
An evidence for the poor accuracy in the low-impedance area can be found from the measurement of the  $Q$  value of an inductor. One can test and read the impedance of an inductor by a single-port test as shown in Figure 6.16; its display on the Smith Chart is shown in Figure 6.17. The trace on the Smith Chart is the measured result of  $S_{11}$  or impedance  $Z$  in the frequency range from DC to 10 GHz. When  $f = 0$ , the impedance or its resistance and reactance are zero because the behavior of an inductor under DC operation is equivalent to short-circuiting. In general, its inductance increases as the operating frequency increases, although it is not necessarily proportional. On the other hand, its resistance also increases as the operating frequency increases. On the trace, each point P corresponds to an operating frequency, and from each point P, the reactance  $x_S$  and the resistance  $r_S$  can be directly read from the Smith Chart. Finally, from the reactance the inductance of the inductor  $L$  can be calculated.

The single-port impedance measurement directly by the network analyzer is very simple and easy. Unfortunately, if the resistance or reactance of the inductor has a very low or very high value as shown on the Smith Chart in either Figure 6.15 or 6.17, then the readings of the reactance  $x_S$  or the resistance  $r_S$  are quite inaccurate, and hence the calculated  $L$  values can deviate from the actual values by around 10–50%. The actual values, of course, can be obtained from other test means with higher accuracy.

The second evidence can be found in the impedance measurement of a short whip antenna that is directly connected to the port 1 of the network analyzer as shown in Figure 6.18.



**Figure 6.16.** Single-port impedance test for an inductor.

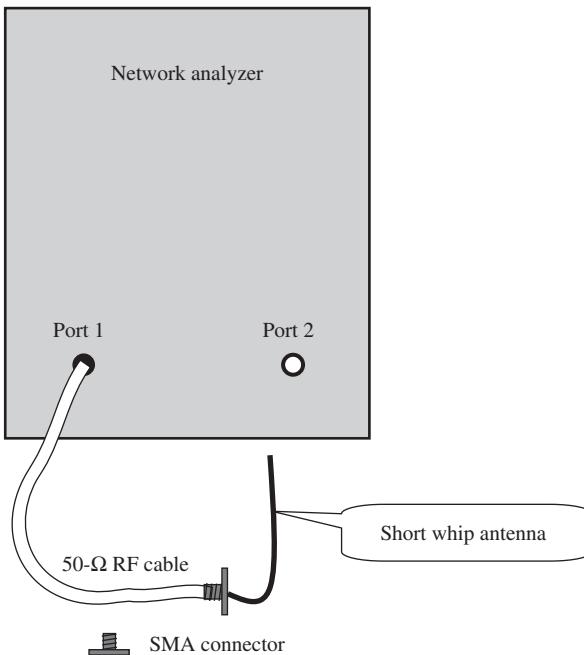


**Figure 6.17.** One-port test for the impedance of an inductor ( $f = 0$  to 10 GHz).

The antenna operates at around 27 MHz, which is a low RF frequency. Its impedance can be directly displayed on the screen of a network analyzer via the Smith Chart, say,

- Real part:  $-1000$  to  $+1500 \Omega$ ,
- Imaginary part:  $3.0 \text{ pF}$ .

The real part of an impedance is unstable and rapidly fluctuates between  $-1000 \Omega$  and  $+1500 \Omega$ ; therefore it cannot be determined definitely. The reactance is kept at about  $3 \text{ pF}$ .



**Figure 6.18.** Single-port impedance test for a short whip antenna.

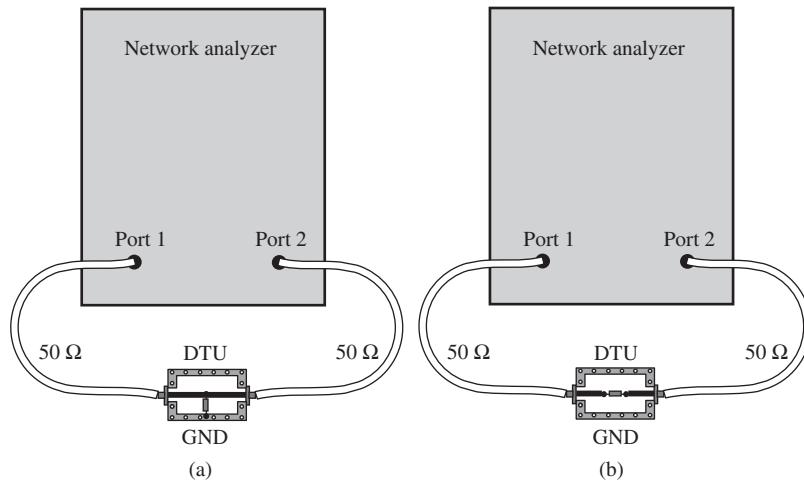
These two examples prove the truth of the assertion about the measurement inaccuracy of impedance on the Smith Chart as shown in Figure 6.15. Consequently, other alternative measuring or testing methods must be sought.

#### 6.4.2 Low- and High-Impedance Measurement

As a matter of fact, the developed RF impedance meter can be applied to measure or test parts that have very low or very high values of resistance or reactance. However, the developed meter brings about other inaccuracies and inconveniences. In addition, it is not a cost-effective equipment.

One of the alternative ways is to keep using the network analyzer for testing or measuring impedance. However, the method of testing or measuring must be changed.

Instead of the single-port test through  $S_{11}$  measurement, it is possible to test or measure the impedance by a two-port test through  $S_{21}$  measurement. Figure 6.19(a) and (b) shows how to test or measure the impedance when the impedance is very low and very high, respectively. In Figure 6.19(a), the desired test part is soldered on the test PCB in parallel. If the impedance of the desired test part is very low, the power delivered from port 1 to port 2 will be considerably attenuated, and therefore its resistance can be calculated from the insertion loss  $S_{21}$  while its reactance can be calculated from its SRF (self-resonance frequency) in parallel. In Figure 6.19(b), the desired test part is soldered onto the test PCB in series. If the impedance of the desired test part is very high, the power delivered from port 1 to port 2 will be considerably attenuated, and therefore its resistance can also be calculated from the insertion loss  $S_{21}$  while its reactance can be calculated from its SRF in series. We will continue discussion of this subject in more detail in Appendix 7.A.1.



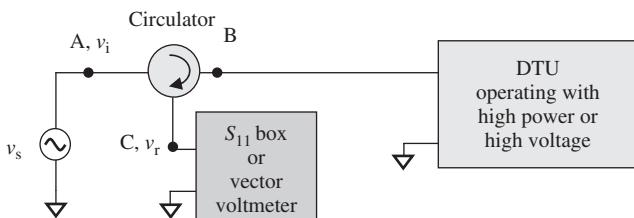
**Figure 6.19.** Test setup for parts with very low and very high impedance DTU (desired test unit). (a) Testing for the parts with very low impedance. (b) Testing for the parts with very high impedance.

## 6.5 IMPEDANCE MEASUREMENT USING A CIRCULATOR

So far, we have discussed impedance measurement by means of  $S$  parameter testing. It is well known that the  $S$  parameters are accurate only for a linear part, in which the test signal is small. For the impedance measurement of a DTU which operates under high power or high voltage,  $S$  parameter testing may be no longer suitable. The impedance measurement in this case can be conducted by means of a circulator as shown in Figure 6.20.

The special feature of a circulator is that the input power or voltage can be transported only in one direction, either clockwise or counterclockwise. As shown in Figure 6.20, the input voltage or power at point A can reach point B as the incident voltage or power for the DTU. The reflected voltage or power from DTU can be transported only to point C and cannot be returned to point A, so it will not disturb the measurement at the input, point A. Consequently, at point C, the reflected voltage or power from the DTU can be correctly measured by the  $S_{11}$  box or vector voltmeter. The impedance can be calculated from the incident voltage  $v_i$  measured at point A and the reflected voltage  $v_r$  measured at point C by the relationship

$$\Gamma = \frac{v_r}{v_i}, \quad (6.15)$$



**Figure 6.20.** Impedance testing of a desired test unit with high power or high voltage.

and then

$$z = \frac{1 + \Gamma}{1 - \Gamma}. \quad (6.16)$$

In a practical power amplifier design, the amplifier operates with a high power input and output. In a practical mixer design, its RF input and IF output are usually treated as low power ports, whereas the LO injection must be treated as a high power port.

## APPENDICES

### 6.A.1 Relationship Between the Impedance in Series and in Parallel

The impedance reading from a test includes both real and imaginary parts, which are usually expressed in series as shown in expression (6.A.1). Sometimes the impedance with its real and imaginary parts in series must be converted into the impedance with its real and imaginary parts in parallel. Figure 6.A.1 sketches the real and imaginary parts of the impedance expressed (a) in series or (b) in parallel. Their relations are as follows:

$$R_S + jX_S = R_P // jX_P = \frac{X_P^2 R_P + jX_P R_P^2}{R_P^2 + X_P^2}, \quad (6.A.1)$$

$$Q = \frac{|X_S|}{R_S} = \frac{R_P}{|X_P|}, \quad (6.A.2)$$

$$R_P = R_S(Q^2 + 1) \approx \frac{X_S^2}{R_S}, \quad \text{if } Q \gg 1, \quad (6.A.3)$$

$$X_P = X_S \frac{Q^2 + 1}{Q^2} \approx X_S, \quad \text{if } Q \gg 1, \quad (6.A.4)$$

where  $Q$  = the quality factor.

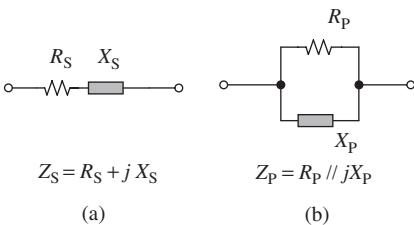


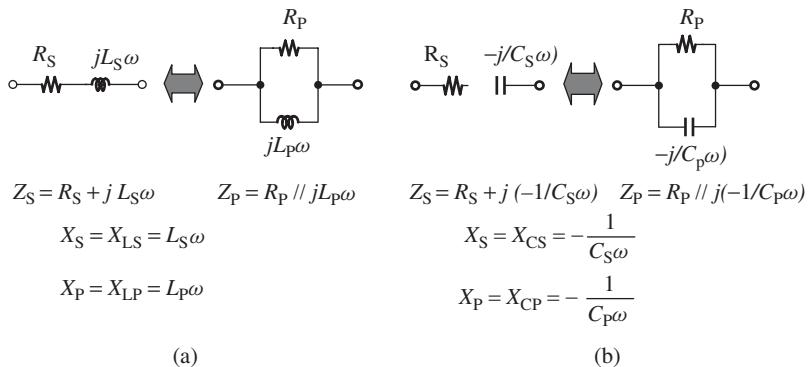
Figure 6.A.1. Real and imaginary parts of impedance expressed (a) in series or (b) in parallel.

When the reactance is inductive, as shown in Figure 6.A.2(a), we have

$$R_S + jL_S\omega = R_P // jL_P\omega = \frac{L_P^2 \omega^2 R_P + jL_P\omega R_P^2}{R_P^2 + L_P^2 \omega^2}. \quad (6.A.5)$$

From (6.A.5), the  $Q$  factor is

$$Q = \frac{L_S\omega}{R_S} = \frac{R_P}{L_P\omega}, \quad (6.A.6)$$



**Figure 6.A.2.** Conversion of impedance between in series and in parallel. (a) Reactance is inductive. (b) Reactance is capacitive.

$$R_P = R_S(Q^2 + 1), \quad (6.A.7)$$

$$L_P = L_S \frac{Q^2 + 1}{Q^2}. \quad (6.A.8)$$

When the reactance is capacitive, as shown in Figure 6.A.2(b), we have

$$R_S + j \left( -\frac{1}{C_S\omega} \right) = R_P // j \left( -\frac{1}{C_P\omega} \right) = \frac{R_P - jC_P\omega R_P^2}{1 + R_P^2 C_P^2 \omega^2}, \quad (6.A.9)$$

and from (6.A.9), the  $Q$  factor is

$$Q = \frac{1}{R_S C_S \omega} = R_P C_P \omega, \quad (6.A.10)$$

$$R_P = R_S(Q^2 + 1), \quad (6.A.11)$$

$$C_P = C_S \frac{Q^2}{Q^2 + 1}. \quad (6.A.12)$$

## FURTHER READING

Lapinoja M, Rahkonen T. An active tuning and impedance matching element. In: Proceedings of the 1998 IEEE International Symposium on Circuits and Systems, 1998. ISCAS '98; 1998 May-3 June, Volume 1. 1998. pp. 559–562.

Laughlin GJ. A new impedance-matched wide-band balun and magic tee. IEEE Trans Microw Theory Tech 1976;24(3):135–141.

Wadell BC. Smith charts are easy. I. Instrum Meas Mag IEEE 1999;2(1):37–40.

Wadell BC. Smith charts are easy. II. Instrum Meas Mag IEEE 1999;2(2):45–47.

Wadell BC. Smith charts are easy. III. Instrum Meas Mag IEEE 1999;2(3):38–42.

## EXERCISES

- What is the essential difference between RF and digital voltage testing?
- Besides voltage testing, what are the other differences between RF and digital circuit testing?

3. What is the advantage of measuring impedance by means of  $S$  parameters, instead of other parameters such as the  $Z$ ,  $Y$ , or  $H$  parameters?
4. What is essential difference between power and impedance measurement in the RF testing laboratory?
5. What is difference between  $S_{11}$  and  $\Gamma_{\text{in}}$  and  $S_{22}$  and  $\Gamma_{\text{out}}$ ? How does this difference impact the impedance measurement?
6. Why is the frequency domain preferred in testing RF circuitry whereas the time domain is preferred in testing digital circuitry?
7. Why is the Smith Chart so powerful in the testing of an RF circuit block or system?
8. Describe the principle of impedance testing by means of a network analyzer.

## ANSWERS

1. In digital voltage testing, the voltage is measured as a scalar quantity without direction. It is the resultant of the incident and reflected voltages. On the contrary, in RF voltage testing, the voltage is measured as a vector quantity with direction. The incident and reflected voltage are distinguished and recorded separately.
2. Digital circuit testing usually works in the time domain and the main test equipment is an oscilloscope with high impedance. On the contrary, RF voltage testing usually works in the frequency domain and the main test equipment is a network analyzer with low impedance ( $50 \Omega$ ).
3. Measurement of impedance by means of  $S$  parameters avoids the difficulty of short-circuiting or open-circuiting, which exist in the testing by means of other parameters such as  $Z$ ,  $Y$ , or  $H$  parameters.
4. In the RF testing laboratory, power measurement must be conducted under the impedance matching condition, whereas impedance measurement does not require such a condition.
5. The difference between  $S_{11}$  and  $\Gamma_{\text{in}}$  and  $S_{22}$  and  $\Gamma_{\text{out}}$  is

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

$$\Gamma_{\text{out}} = S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1 - S_{11}\Gamma_S}$$

$S_{11}$  is the input voltage reflection coefficient  $\Gamma_{\text{in}}$  when  $\Gamma_L = 0$ ,

$S_{22}$  is the output voltage reflection coefficient  $\Gamma_{\text{out}}$  when  $\Gamma_S = 0$ .

This difference means that in the impedance measurement the condition must be satisfied through a correct calibration of network analyzer.

6. The modulated carrier is transported or manipulated in the RF circuitry. Its performance can be easily characterized and clearly displayed by the frequency spectrum but not by the waveform. On the other hand, the demodulated signal is transported or manipulated in the digital circuitry. Its performance can be easily characterized and clearly displayed by the waveform but not by the frequency spectrum. Therefore, testing RF circuitry in the frequency domain is preferred while testing digital circuitry in the time domain is preferred.

7. Smith Chart is a complex plane of the voltage reflection coefficient. It presents the relationship between the voltage reflection coefficient with impedance and other parameters such as, admittance, VSWR (voltage standing wave ratio), and so on simultaneously. It is a powerful tool in the RF circuit design because using the chart,
- Many parameters can be directly converted from each other.
  - In the construction of an impedance matching network, the part can be correctly selected on the basis of the location of the impedance displayed in Smith Chart.
8. The principle of impedance testing by means of network analyzer is as follows:
- Measure the incident and reflected voltage,  $v_i$  and  $v_r$ , first;
  - Calculate the ratio of the incident and reflected voltage,  $\Gamma$ , second:

$$\Gamma = \frac{v_r}{v_i}.$$

- Calculate the impedance finally:

$$Z = \frac{1 + \Gamma}{1 - \Gamma}.$$

# GROUNDING

## 7.1 IMPLICATION OF GROUNDING

Grounding is a common concern in all circuit design for circuit blocks and systems, including RF, digital, and analog circuit designs. As long as the circuit operates in AC mode, grounding must be done properly no matter how low or high the operating frequencies are. In the RF frequency range, grounding becomes a more serious matter because RF frequencies are usually higher than other AC frequencies. Many problems or “bugs” arise from inappropriate grounding. As a matter of fact, grounding is one of the core technologies and an indispensable skill in RF circuit design. In high-speed digital circuit designs, grounding is as important as in RF circuit design.

In a circuit block or a system, there are many ground points or nodes marked in a schematic. Figure 7.1 shows a typical way to draw a schematic of circuitry, in which the positive terminal of the DC power supply point is marked with  $V_{cc}$  and the negative terminal of the DC power supply is marked with GND or by an up-side-down symbol “ $\Delta$ .” The positive terminal of the DC power supply point,  $V_{cc}$ , is connected to many subpoints of DC power supply,  $V_{cc1}$  to  $V_{cc8}$ , and the negative terminal of the DC power supply point, GND, is connected to many subpoints of ground points,  $G_1$  to  $G_8$ .

First, a question is raised: which grounded point is the reference ground point in a circuit block or system if these ground points are not equipotential? The answer is that the reference ground point is the terminal or grounded point of the DC power supply.

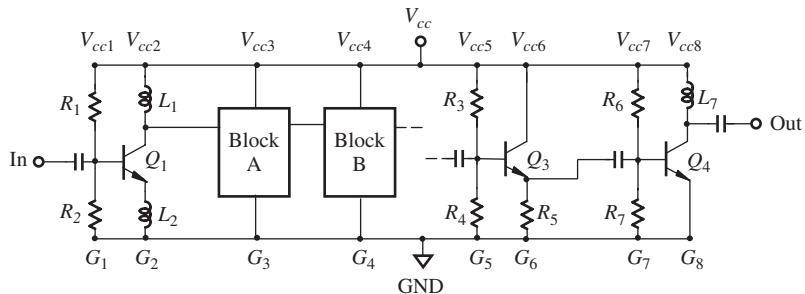


Figure 7.1. A typical way to draw a schematic of circuitry.

This is an irrefutable and unanimous assertion because all the currents flowing through all the branches of the circuitry must eventually return to the terminal or ground point of the DC power supply. A real ground point or node in a schematic must be equipotential to the ground point of the DC power supply, GND.

Second, the ground points or nodes,  $G_1$  to  $G_8$ , are connected together by a straight line segment as shown in the bottom of Figure 7.1. At these ground points or nodes, both the DC and AC grounding are expected simultaneously: that is, in respect to a reference ground point or node, GND, the following conditions of the AC and DC components of voltage and impedance must be satisfied, that is,

$$V_{AC} = 0, \quad V_{DC} = 0, \quad (7.1)$$

and

$$Z_{AC} = 0, \quad Z_{DC} = 0, \quad (7.2)$$

where the subscripts “AC” and “DC” denote the alternative and directive components of the parameters, respectively.

In other words, these ground points are expected to be equipotential with or short circuited to the reference ground point, GND. Consequently, these ground points,  $G_1$  to  $G_8$ , are named as fully ground points, in which both the AC and DC components of voltage and impedance are equal to zero in respect to the reference ground point, GND.

In RF circuit design with discrete parts, these kinds of ground points are usually embedded or “swallowed” by a ground surface. In the RFIC circuit design, the  $P+$  guard ring in an individual RF block functions as a ground surface.

Third, the points or nodes  $V_{cc1}$  to  $V_{cc8}$  are DC power supply or DC bias terminals. They are connected together by a straight line segment as shown on the top of Figure 7.1. At these ground points or nodes, AC grounding, but not DC grounding, is expected. That is, in respect to a reference ground point or node, GND, the following conditions for the AC and DC components of voltage and impedance must be satisfied, that is,

$$V_{AC} = 0, \quad V_{DC} \neq 0 \quad (7.3)$$

and

$$Z_{AC} = 0, \quad Z_{DC} \neq 0. \quad (7.4)$$

In other words, at all terminals of DC power supply or DC bias, both the AC impedance and AC voltage are expected to be equipotential with or short circuited to the reference ground point, GND, while both the DC impedance and DC voltages must be kept at an expected nonzero value. These ground points,  $V_{cc1}$  to  $V_{cc8}$ , are therefore named as “half ground points.”

Obviously, the grounding scheme for fully grounded points,  $G_1$  to  $G_8$ , is different from grounding scheme for half ground points,  $V_{cc1}$  to  $V_{cc8}$ , which will be introduced and discussed later.

## 7.2 POSSIBLE GROUNDING PROBLEMS HIDDEN IN A SCHEMATIC

In the textbook and in the classroom, the schematic of circuitry is always drawn as shown in Figure 7.1, in which many ambiguous grounding problems could hide in dark corners.

Figure 7.1 contains many full and half grounded points but specifies nothing about the following questions:

- How to ensure all the fully grounded points being AC and DC equipotential with the reference ground point or terminal?
- How to ensure all the half grounded points being AC but not DC short circuited to the reference ground point or terminal?
- Furthermore, if the conditions of equipotentiality, (7.1)–(7.4), are not satisfied, the problem of forward current and return current coupling will appear.

If a half grounded point is not equipotential with the reference ground point or terminal, its AC forward current would flow from the DC power supply  $V_{cc}$  to the half grounded point. (Usually, the internal resistance of a DC power supply is about a tenth of ohms, so that the positive and negative terminal is equipotential to the AC component of either voltage or current). The forward currents will be coupled from each other as they flow from the DC power supply,  $V_{cc}$ , to the half grounded points.

Similarly, if a full grounded point is not equipotential with the reference ground point or terminal, its AC return current would flow from the full grounded point to the reference grounded point, GND. The return currents would be coupled from each other in the way when they flow from the full grounded point to the reference grounded point, GND.

Figure 7.2 shows the forward and return currents, which are ignored in Figure 7.1, as is usual in a schematic drawing.

In the simulation design stage, the above questions are never raised. On the contrary, it is assumed that

- the fully grounded points are equipotential with the reference grounded point, GND, and conditions (7.1) and (7.2) are satisfied;
- the half ground points are equipotential with the reference ground point, GND, for the RF/AC signal but not for the DC voltage, and conditions (7.3) and (7.4) are satisfied;
- the forward and return current couplings are negligible.

This is, of course, far from reality. Should these hidden problems be erased, it is equal to say that “there are no grounding problems in the circuit design”!

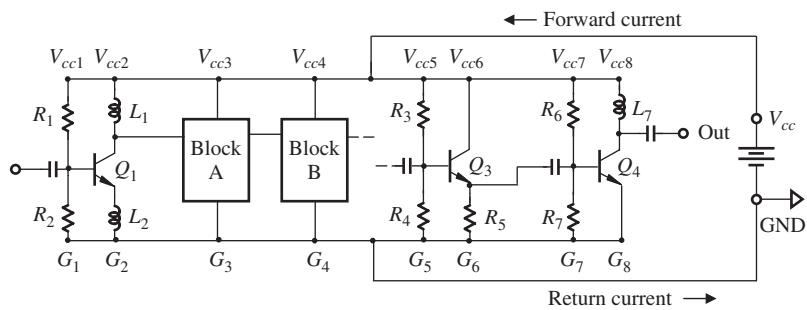


Figure 7.2. Forward and return currents added to the schematic as shown in Figure 7.1.

### 7.3 IMPERFECT OR INAPPROPRIATE GROUNDING EXAMPLES

On the basis of practical statistics of RF circuit design, most design problems do not exist in the simulation stage, but quite a high percentage of problems are due to imperfect or inappropriate RF/AC grounding. Imperfect or inappropriate RF/AC grounding leads to performance with unexpected erroneous functions. At the extreme end, it can put an RF circuit block or system out of work.

The following examples expose some typical problems in RF/AC grounding.

#### 7.3.1 Inappropriate Selection of Bypass Capacitor

**7.3.1.1 Underestimation or Ignoring of Bypass Capacitor.** Essentially, the AC bypass capacitor and the DC blocking capacitor are the same part in an RF circuit block. Their impedance approaches zero at the operating frequency or within the operating frequency bandwidth. They are commonly called *zero* capacitors, which is discussed in Section 7.4. In this section, only the bypass capacitor is mentioned because it is more often ignored than the DC blocking capacitor. The DC blocking capacitor is one of the parts in the circuit schematic and is taken care of in the simulation, whereas the bypass capacitor is very often not drawn in the circuit schematic and neglected in the simulation. Computers always assume that the DC power supply or DC bias is an ideal unit with short-circuited AC. It seems not necessary to have a bypass capacitor connected with it in parallel. The RF circuit designer does not need to worry about the bypass capacitor until he or she is going to test the circuit board. This might be one of reasons that the bypass capacitor is easily ignored.

Some think that the selection of the bypass capacitor is a tiny thing and not worthwhile to mention. In published electrical engineering books or in electrical engineering courses, it is hard to find any material dealing with the bypass capacitor.

Is the bypass capacitor very easy to deal with, and therefore, not worth mentioning? Statistics indicates that a very high percentage of failure rate in the testing of an RF circuit block is due to inappropriate selection of the bypass capacitor. Obviously, the task of selecting a bypass capacitor or “zero” capacitor is underestimated. In the testing stage, the bypass capacitor is an indispensable part not only for RF but also for digital circuit blocks. An improper bypass capacitor always leads to ridiculous test results, and sometimes, the tested circuit block is simply out of work even though the circuit block has demonstrated excellent performance in the simulation.

### 7.3.1.2 Blind Selection of Bypass Capacitor.

Here is a true story:

In a famous institute engaging in R&D for a wireless electronic product, an electrical engineer was about to test an RF block circuitry. He was aware that he had to attach a bypass capacitor at the DC power supply port. Then, he asked his supervisor, "Which value of capacitor must be selected from the engineering stock room?" Without any hesitation, his boss answered his question with brimming confidence: "0.01  $\mu\text{F}$ !" From then on, the majority of the electrical engineers would apply a 0.01  $\mu\text{F}$  capacitor as a bypass capacitor in whatever circuit they designed.

Some people had different opinions: "The bypass capacitor must be different for various circuits. You can choose the bypass capacitor when you are going to adjust your circuit block into an optimal state by testing. For example, if you are designing a low-noise amplifier LNA, you may swap out bypass capacitors one by one until you get a maximum of gain or a minimum of noise figure, or until you get both simultaneously."

This sounds a pretty good idea, but in reality, it starts to look like little more than a dream. The performance of the LNA might be unchanged even with many different values of bypass capacitors.

An easy but lazy way to solve the selection problem for a bypass capacitor is to put a huge crowd of capacitors on the DC power supply port, by which the RF/AC signal should be short circuited over a very wide frequency range. This method is very time effective and universal, dealing with all the half ground points in the circuit, regardless of the operating frequencies in the circuit block.

For example, Figure 7.3 shows a circuit block in which the main operating frequency is 2.4 GHz. A huge crowd of capacitors with values ranging from very low, 10 pF, to very high, 100  $\mu\text{F}$ , are connected between the DC power supply port and ground in parallel. The capacitors from  $C_1$  to  $C_6$  are chip capacitors, while the capacitors from  $C_7$  to  $C_{12}$  are electrolytic capacitors. The designer thought that, with such a setup, there should be no problem of reaching the goal of AC short circuiting and DC open circuiting at the DC power supply port.

Unfortunately, it is found that the huge crowd of capacitors is useless because the AC signal appearing on the DC power supply port is maintained at the same level whether

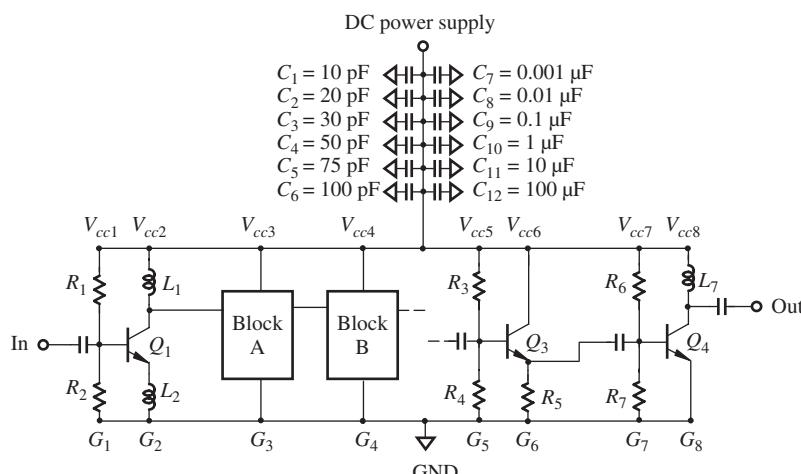


Figure 7.3. A huge crowd of capacitors are supposed to function as bypass capacitors.

they are connected to or removed from the DC power supply port. This inconceivable event frustrated the designer for a long time.

The answer to this problem can be found from Section 7.4 and 7.A.1, where the zero capacitor is discussed. The SRF (self-resonant frequency) of any capacitor in the huge crowd of capacitors is below the operating frequency of 2.4 GHz. From Table 7.A.1 it can be found that the correct value of the bypass capacitor must be 5.1 pF.

### 7.3.2 Imperfect Grounding

**7.3.2.1 The Long March.** Figure 7.4 shows a PCB (printed circuit board) test for an 800 MHz transceiver, in which the transceiver IC die has been packaged with 20 pins and soldered onto the PCB. The purpose of this test setup is to check the receiver sensitivity, since the IC die is bought from an outside company. The input is connected to a generator that provides a small modulated signal so as to simulate a received signal from the antenna. The output is connected to an audio distortion meter so that the sensitivity of the receiver can be measured by 12-dB SINAD (Ratio of Signal to Noise And Distortion) or 20-dB *Quieting*. For simplicity, only the parts and runners related to grounding are drawn in Figure 7.4.

The tested sensitivity is found to be much lower than what the manufacturer specified. The problem is due to an inappropriate PCB layout. One can see its extremely poor RF/AC grounding as shown in Figure 7.4. Let us take a look at how flawed the RF/AC grounding of the input SMA connector is. Step by step, the ground port of the input

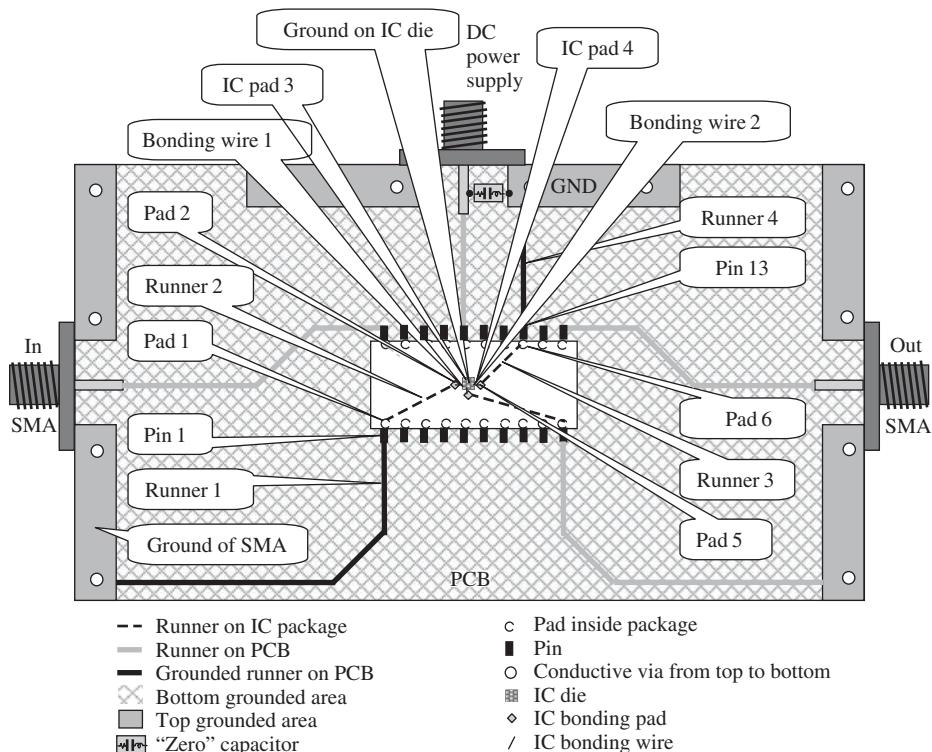


Figure 7.4. Long march from ground at SMA input to the GND at DC power supply.

SMA connector is eventually connected to the reference ground point of the DC power supply, GND, through a “Long March” of sorts, that is,

- ground of SMA connector at input;
- runner 1 on PCB;
- pin 1 of IC package;
- pad 1 inside IC package;
- runner 2 inside IC package;
- bonding pad 2 inside IC package;
- bonding wire 1;
- IC pad 3 on IC die;
- ground on IC die;
- IC pad 4 on IC die;
- bonding wire 2;
- bonding pad 5 inside IC package;
- runner 3 inside IC package;
- pad 6 inside package;
- pin 13 of IC package;
- runner 4 on PCB;
- destination: reference ground point, GND.

Should the thinness and tiny size of the bonding wires and ground wire inside an IC die come to mind, one would definitely feel sorry for such a grounding path, which mostly runs through microstrip lines with high-characteristic impedance. Through such a grounding path, the ground of the input SMA connector is definitely not equipotential with the reference ground point of the DC power supply, GND, but is better compared to a “balloon” floating in the air. Consequently, the testing of the sensitivity of the receiver is of course incorrect, and is usually greatly degraded.

It has been proved that the higher sensitivity of the receiver can be restored, in which it is enhanced by 6 dB after the test PCB is redesigned with good RF/AC grounding.

**7.3.2.2 Unequipotentiality on Grounded Surface.** A good grounded RF circuit block implies that equipotentiality is maintained over the entire grounded surface. In other words, the entire grounded surface is an electrically equipotential surface. Unequipotentiality on the grounded surface is one of the imperfect grounding problems. The tested results are questionable if the grounded surface is not an equipotential surface, because the tested parameter is not referenced to the same reference ground point, GND.

If the grounded surface is not in an equipotential state, the current will flow over the surface and result in so-called current coupling, which might be a problem in a dark corner and therefore puts the circuit block out of work!

It should be noted that equipotentiality is not directly related with high conductivity. For instance, a copper layer with gold plating has very high conductivity. However, its conductivity does not guarantee that its surface is electrically equipotential. On the contrary, the voltage is generally different from one point to another. If the shape of this copper plane is long and narrow on a substrate or PCB, it is a microstrip line. The voltage on a microstrip line varies from point to point or from line to line. Strictly speaking, equipotentiality on a metallic surface is never exactly true; however, it can

be well approached if its size or dimension is greatly less than the quarter wavelength, that is,

$$L_d \ll \lambda/4 \quad (7.5)$$

where  $L_d$  is the maximum dimension of the metallic surface.

In order to ensure the equipotentiality of a grounded surface, it is better to design the PCB or IC die with a dimension much less than the quarter wavelength. For complicated or large systems, in which the size of the PCB or IC die may be comparable with or larger than the quarter wavelength, a special means or scheme must be worked out so as to guarantee equipotentiality on the whole grounded surface. This will be discussed in the next chapter along with the current coupling problem.

### 7.3.3 Improper Connection

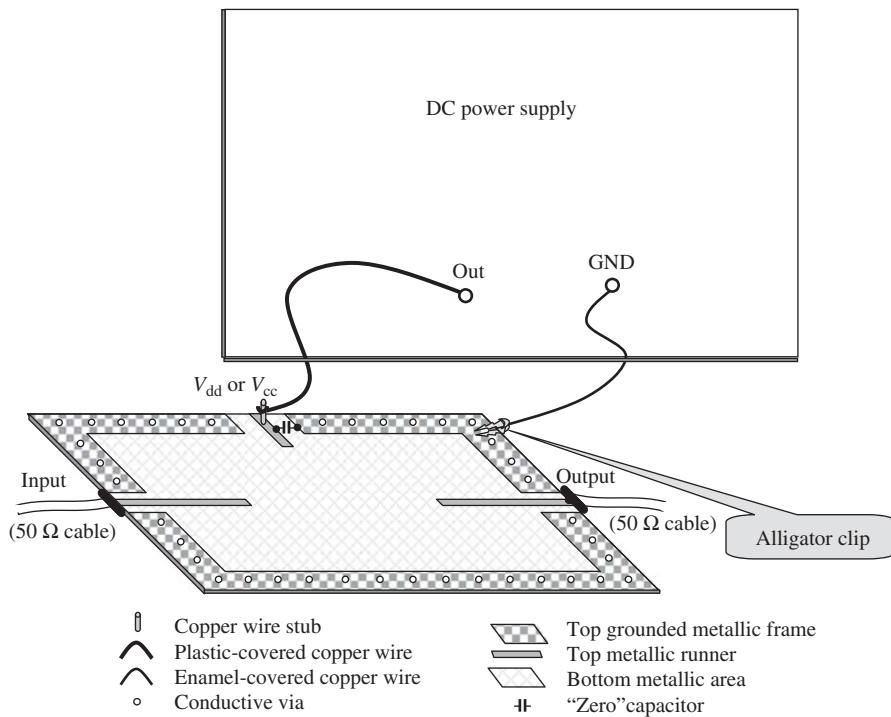
The following are two examples of improper connection about the ground points.

**7.3.3.1 Connection from Tested PCB to DC Power Supply.** Figure 7.5 shows a lazy approach to connecting the test PCB with the DC power supply. A short copper wire stub is erected and soldered at the terminal  $V_{dd}$  or  $V_{cc}$  of the test PCB. This stub is connected to the output of the DC power supply by a plastic-covered copper wire. The ground terminal of the DC power supply, GND, is connected to the test PCB by an enamel-covered copper wire. One end of the enamel-covered wire is connected to the ground terminal of the DC power supply. Another end of the enamel-covered wire is put on an alligator clip, which is placed in the top grounded metallic frame of the test PCB in a very convenient way.

Using an alligator clip, an enamel-covered copper wire, and a plastic-covered copper wire is really convenient. However, it brings about a lot of trouble:

- The short wire copper stub may function as a small antenna or an impedance compensator.
- The long plastic-covered copper wire may function like a big antenna with a heavy load, the output impedance of the DC power supply.
- The length of the enamel-covered copper wire may be comparable to or longer than the quarter wavelength corresponding to the operating frequency. Two ground terminals in the DC power supply and the tested PCB may be AC/RF disconnected.
- The “antennas” absorb whatever interference signals and noise they sense in the sky. The zero capacitor can AC short circuit those frequency spectrums around the operating frequency, but does nothing for those interferences whose spectrums are far from the operating frequency; especially, it does nothing to alleviate white noise.
- Noise not only comes from the sky but also from the DC power supply, and so on.

Without hesitation, such a lax method must be strictly prohibited in RF circuit testing.

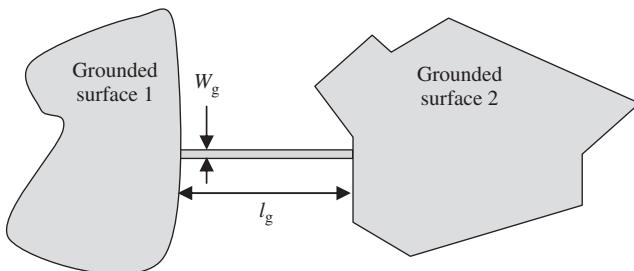


**Figure 7.5.** A lax way to connect tested PCB with DC power supply.

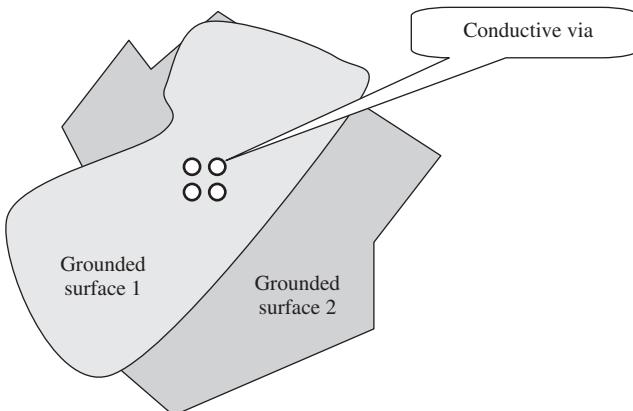
**7.3.3.2 Connection from Ground to Ground.** Very often there is more than one piece of grounded surface in an RF system or block. The connection between grounded surfaces may bring about trouble. Figures 7.6 and 7.7 show two types of connections between grounded surfaces.

In the circuit board of an RF module or RF system, two pieces of grounded surfaces are very often connected by a runner as shown in Figure 7.6. If the length of this runner is comparable to the quarter wavelength, the two pieces of the grounded surface are not equipotential. Consequently, grounding problems arise.

In an RFIC chip, two pieces of grounded surfaces are very often connected by a bunch of conductive vias as shown in Figure 7.7. If the diameter of the conductive via is too small, the two grounded surfaces may not be equipotential. Consequently, grounding problems may arise.



**Figure 7.6.** Two pieces of grounded surface is connected by a runner.



**Figure 7.7.** Two pieces of grounded surface is connected by conductive vias.

## 7.4 ‘ZERO’ CAPACITOR

In this section, our discussion about the ‘zero’ capacitor applied for RF/AC grounding is not for integrated circuits, but for circuits built with discrete parts. However, the discussion about the quarter wavelength for RF/AC grounding is suitable to both integrated circuits and circuits built by discrete parts. Among the discrete parts applied for RF/AC grounding, zero capacitors and microstrip lines are the two main parts widely applied in RF blocks and systems.

### 7.4.1 What is a Zero Capacitor

The term *zero capacitor* is a “professional” slang or terminology. It indicates that at the specified operating frequency the impedance of said capacitor approaches zero.

A zero capacitor can therefore function as a bypass capacitor if it is connected between the port to be RF/AC grounded and the reference ground point, GND, in parallel. At the port to be grounded, the DC voltage is kept unchanged, but the RF/AC signal is short circuited to the ground. On the other hand, a zero capacitor can function as a DC blocking capacitor if it is connected with the port in series, in which the DC voltage is blocked but the RF/AC signal crosses over without any attenuation. Both applications, as AC bypass capacitor and as DC blocking capacitor, are shown in Figure 7.8. At the nodes of  $V_{cc}$  and Bias, the zero capacitors are connected in parallel and thus function as AC bypass capacitors; at the nodes of input and output, the zero capacitors are connected in series and function as DC blocking capacitors.

In short, either the AC bypass capacitor or the DC blocking capacitor is the zero capacitor.

### 7.4.2 Selection of a Zero Capacitor

It is well known that the impedance for an ideal capacitor,  $Z_c$ , is

$$Z_c = \frac{1}{jC\omega}, \quad (7.6)$$

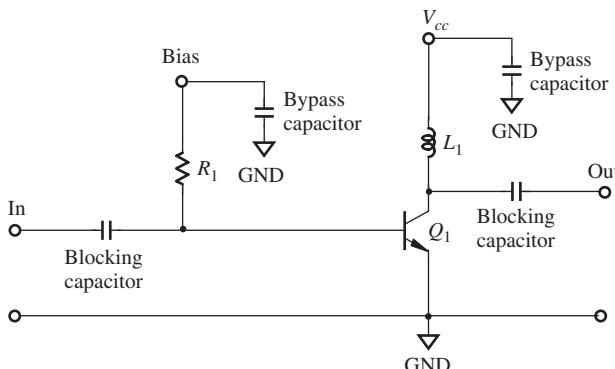


Figure 7.8. Zero capacitor functions as a bypass or a blocking capacitor.

where

$C$  = capacitance of the capacitor, and

$\omega$  = operating angular frequency.

The subscript 'c' denotes the capacitor's parameter.

It can be seen that the impedance of a capacitor,  $Z_c$ , is infinite to the DC current or voltage when  $\omega = 0$  and that the impedance of a capacitor is decreased as the value of the capacitor increases when  $\omega \neq 0$ .

Theoretically, at the RF frequency range, the impedance of an ideal capacitor could approach zero by infinitely increasing the capacitor's value  $C$  when  $\omega \neq 0$ . In other words, the zero capacitor is an ideal capacitor with infinite capacitance when  $\omega \neq 0$ . RF/AC grounding could be accomplished by connecting an ideal capacitor with infinite capacitance between the desired RF grounded terminal and a real grounded point.

In reality, an ideal capacitor does not exist, and applying infinite capacitance is not realistic or practical. In the RF frequency range, an actual capacitor is never close to being an ideal capacitor, so the ideal formula for the capacitor impedance is even less viable, although in the low-frequency range, an actual capacitor is pretty close to being ideal and can be described by expression (7.6).

There are many kinds of capacitors available in the market. In today's RF circuit design implemented either by discrete parts or by integrated circuits, most zero capacitors are chip capacitors because of their advantage of small size, low cost, and reliable performance.

In the RF frequency range, a chip capacitor has additional inductance and additional resistance and can be modeled as shown in Figure 7.9, in which a chip capacitor is represented by an ideal capacitor  $C$ , an ideal inductor  $L_S$ , and an ideal resistor  $R_S$  in series. The additional inductance  $L_S$  brings about additional phase shift of the RF/AC signal, and the additional resistance  $R_S$  represents additional attenuation of the RF/AC signal or the  $Q$  value of the actual capacitor.

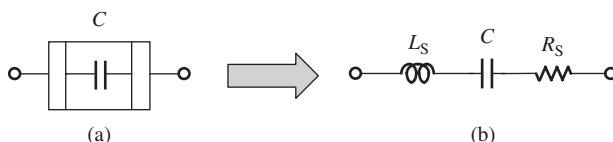


Figure 7.9. An actual chip capacitor (a) and its equivalent (b).

Instead of infinitely increasing the capacitor's value  $C$ , the impedance of an actual chip capacitor,  $Z_c$ , approaches  $R_S$  when its additional inductor is resonant with the capacitor itself at its SRF. The SRF of an actual chip capacitor can be expressed as

$$\text{SRF}_c = \frac{1}{2\pi\sqrt{L_S C}}, \quad (7.7)$$

where the subscript "c" denotes the capacitor's parameter.

Should the SRF be identical to the operating frequency, the capacitor is a zero capacitor since the value of the resistor  $R_S$  shown in Figure 7.9 is a tiny value or approaches zero at the operating frequency. In other words, the zero capacitor is re-recognized as a bypass or DC blocking capacitor with "zero impedance" when the operating frequency is equal to its SRF.

The SRF of a chip capacitor depends on the material, art, and technology adapted by the manufacturer. For the chip capacitor manufactured by MuRata, the SRF can be formulated as

$$\text{SRF}_c = \frac{5400}{\sqrt{C_{\text{specified}}}}, \quad (7.8)$$

where

$\text{SRF}_c$  = self-resonant frequency in MHz (megahertz), and  
 $C_{\text{specified}}$  = specified capacitance by the manufacturer in pF (picofarad).

Formula (7.8) is available within the range of specified capacitance

$$C_{\text{specified}} = 0.5\text{--}18,000 \text{ pF},$$

and for both sizes

$$W \times L = 50 \times 80 \text{ mils and } 30 \times 40 \text{ mils}$$

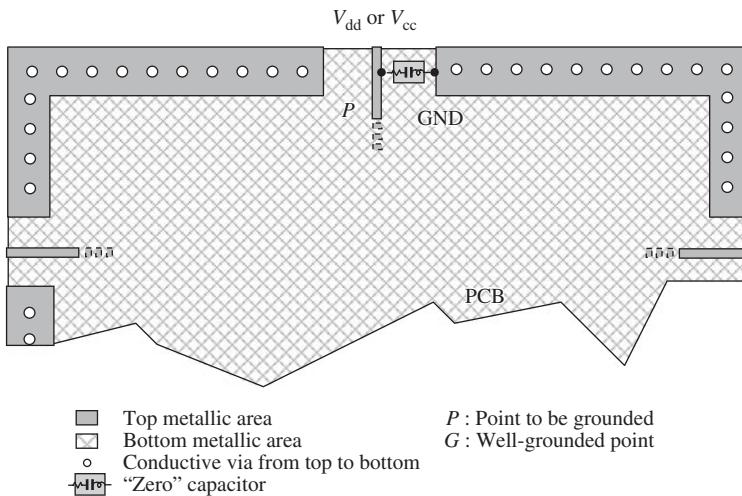
where  $W$  and  $L$  are the width and length of the chip capacitor, respectively.

In terms of formula (7.8), one can easily find out the zero capacitor for a specific operating frequency. In Appendix 7.A.1, we discuss how to find the additional inductance  $L_S$  and additional resistance  $R_S$  from the actual measurement of a chip capacitor, and finally how to obtain formula (7.8).

If the zero capacitor is not manufactured by MuRata, the formula may be slightly different. By means of the same testing procedures, an appropriate formula could be found, in which the coefficient "5400" in formula (7.8) would be different and replaced with some other value.

Figure 7.10 shows a formal PCB, which will be discussed in Chapter 9. The parts on the PCB could be discrete elements, RFIC chips, or both. For simplicity, the parts on the PCB are not drawn, since our concern is only focused on how to apply the zero capacitor for RF/AC grounding.

The terminals of the DC power supply and DC bias are half ground points, in which RF/AC signals must be grounded but the DC voltage must not be short circuited. A zero capacitor is therefore connected between the DC power supply terminal  $V_{dd}$  or  $V_{cc}$  and the point GND, which is the reference ground point on the PCB ground surface. Similarly, the zero capacitor can be applied to other half ground terminals, such as the DC bias, which are not drawn in Figure 7.10.



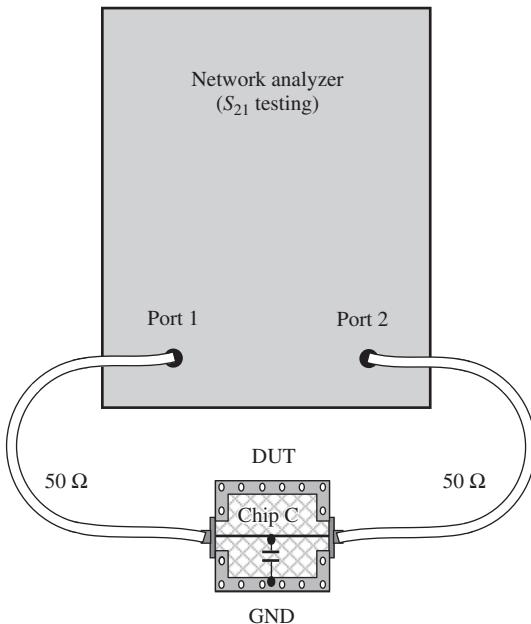
**Figure 7.10.** RF or AC grounded by a zero capacitor on a PCB (printed circuit board).

### 7.4.3 Bandwidth of a Zero Capacitor

The SRF of a chip capacitor can be found by means of  $S_{21}$  measurement.

Figure 7.11 shows an  $S_{21}$  testing setup for a chip capacitor. As a desired test unit (DTU), the tested chip capacitor is soldered between the  $50\ \Omega$  microstrip line and ground on a small PCB. The small PCB is then connected to ports 1 and 2 of the network analyzer by  $50\ \Omega$  RF cables. The tested chip capacitor is connected with port 1 and 2 of the network analyzer in parallel accordingly.

By referring to the equivalent model of a chip capacitor as shown in Figure 7.9 and the test setup as shown in Figure 7.11, the capacitor  $C$  is resonant with  $L_S$  in series at the SRF, so that the  $S_{21}$  is significantly lowered because of its connection with port 1 and 2



**Figure 7.11.**  $S_{21}$  testing for a chip capacitor.

TABLE 7.1. Bandwidth of Zero Capacitor,  $C = 15 \text{ pF}$ ,  
SRF = 1.394 GHz

Assigned Critical Value of $S_{21}$ (dB)	Bandwidth (MHz)
-10	370
-20	165
-30	70

of the network analyzer in parallel. Figure 7.12 shows a typical  $S_{21}$  frequency response of a tested chip capacitor,  $C = 15 \text{ pF}$ . It is found that

$$S_{21} = -45.1 \text{ dB},$$

at

$$\text{SRF} = 1.394 \text{ GHz}.$$

An actual zero capacitor is self-resonant at its SRF and approaches zero impedance within a frequency range around the SRF. In other words, the zero capacitor can enforce a node to approach zero impedance within a certain bandwidth as long as it is connected with this port to the reference ground in parallel. The bandwidth depends on which level of  $S_{21}$  is being assigned as the critical value of “approaching to zero impedance.” From Figure 7.12, the bandwidth of the zero capacitor can be found and is listed in Table 7.1.

The lower the assigned critical value of  $S_{21}$ , the better the approach to zero impedance. However, as shown in Figure 7.12 or Table 7.1, the lower the assigned critical value, the narrower the bandwidth to be covered is.

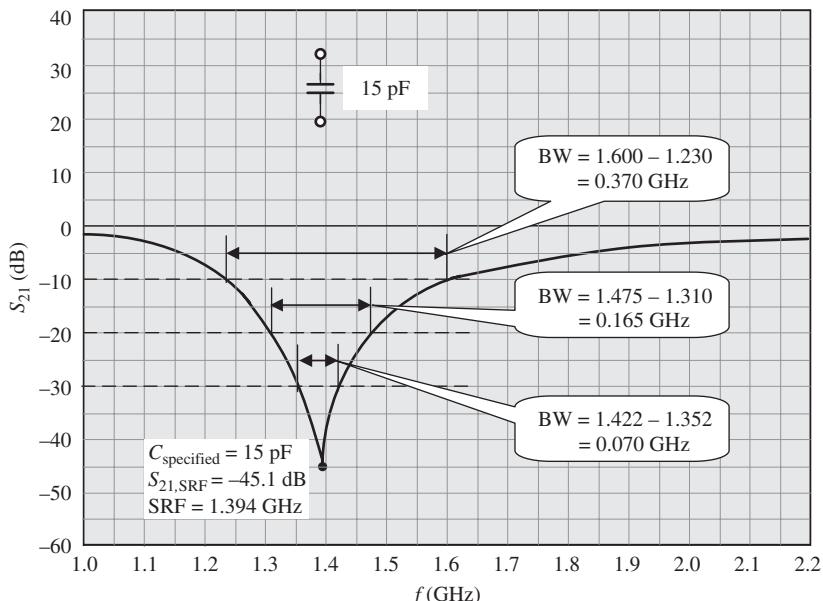
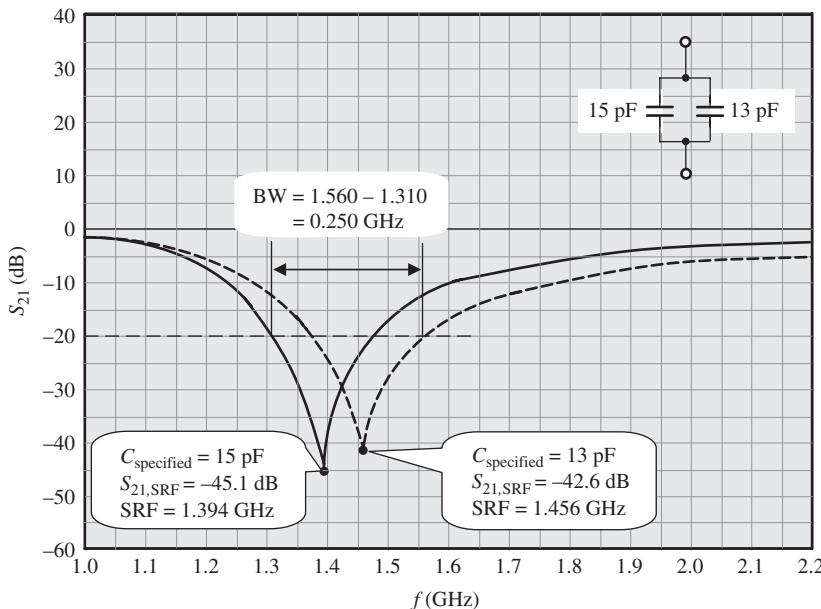


Figure 7.12.  $S_{21}$  frequency response of a chip capacitor tested with setup as shown in Figure 7.11.



**Figure 7.13.** Frequency bandwidth  $250 \text{ MHz} = 1.560 - 1.310 \text{ GHz}$  is covered by a zero capacitor combined by two capacitors  $C = 15 \text{ pF}$  and  $C = 13 \text{ pF}$  connected together in parallel.

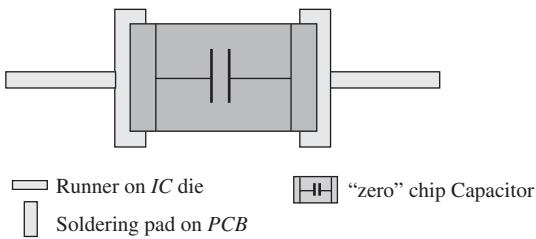
For a narrow band block or system, one zero capacitor is usually enough to cover its bandwidth. Figure 7.12 shows that the zero capacitor,  $C = 15 \text{ pF}$ , can cover a bandwidth of 165 MHz, or  $f = 1310 - 1475 \text{ MHz}$ , if the assigned critical value of  $S_{21}$  is  $-20 \text{ dB}$ .

However, for a wide band block or system, the zero capacitor can and must be formed by a combination of several capacitors in parallel, in which the SRF of the capacitors is shifted in a way so that zero impedance can be approached as  $S_{21}$  is lower than the assigned value over the expected bandwidth. For example, if the bandwidth to be covered is 250 MHz for  $f = 1.310 - 1.560 \text{ GHz}$  and the critical value of  $S_{21}$  is assigned as  $-20 \text{ dB}$ , the single zero capacitor,  $C = 15 \text{ pF}$ , is not enough to cover the bandwidth. However, the zero capacitor can be formed by a combination of two capacitors,  $C = 15 \text{ pF}$  and  $C = 13 \text{ pF}$ , connected together in parallel. The combined zero capacitor can cover a bandwidth even wider than the expected 250 MHz. Figure 7.13 shows the coverage of the expected bandwidth.

#### 7.4.4 Combined Effect of Multi-Zero Capacitors

Very often, multiple chip capacitors are applied to form a zero capacitor. This is not only for the sake of bandwidth coverage but also for the multibands existing in a block or system. For instance, there are three bands, the RF, LO, and intermediate frequency (IF) frequency bands, existing in a mixer block. The zero capacitor must be able to cover these three bands. This implies that there are at least three zero capacitors to be applied in a mixer block. If any of the three bands is wide band, which cannot be covered by only one zero capacitor, then the total number of capacitors needed to form a zero capacitor will be greater than 3.

An interesting question is then raised: does a resultant SRF due to the combination of all the individual zero capacitors exist, superseding the SRFs of the individual zero



**Figure 7.14.** Modified equivalent model of the “zero” chip capacitor obtained by addition of soldering pads and runners for RF modules.

capacitors, since these zero capacitors are connected between the positive terminal of DC power supply or DC bias and a real grounded point, GND, in parallel? According to empirical experimentation, a resultant SRF combined by individual zero capacitors may appear while the behavior of the individual zero capacitors is unchanged. The resultant SRF is usually located outside of the expected bandwidth.

Finally, it must be noted that in the simulation phase of circuit design, special care must be taken when dealing with the model of the zero chip capacitor, especially when the operating frequency is higher than gigahertz. In addition to what is shown in Figure 7.9, soldering pad and runners must be attached to the equivalent model as shown in Figure 7.14.

The modified equivalent model consists of the following five parts:

- A runner on the PCB;
- A soldering pad on the PCB;
- A zero chip capacitor;
- A soldering pad on the PCB;
- A runner on the PCB.

By design experience, the difference in SRF between the cases with and without the modification of the equivalent model of zero chip capacitor is around 5% in the gigahertz of frequency range.

#### 7.4.5 Chip Inductor is a Good Assistant

Since we select the chip capacitor as a zero capacitor, let us introduce the chip inductor, which is a good assistant to the zero capacitor in RF/AC grounding.

Theoretically, the impedance of an ideal inductor,  $Z_L$ , is

$$Z_L = jL\omega, \quad (7.9)$$

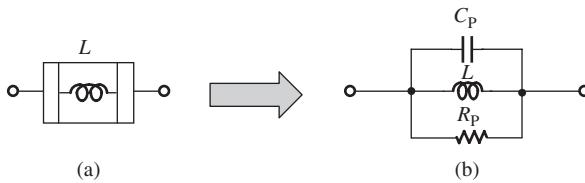
where

$L$  = inductance of the inductor, and

$\omega$  = operating angular frequency.

The subscript “L” denotes the inductor’s parameter.

It can be seen that the impedance of an ideal inductor,  $Z_L$ , is zero to the DC current or voltage when  $\omega = 0$  and that it increases as the value of the inductor increases when  $\omega \neq 0$ . The value of  $Z_L$  can be increased to a very high level, almost infinite, if the value of the inductor,  $L$ , is high enough for the RF signal.



**Figure 7.15.** An actual inductor (a) and its equivalent (b).

Just like the ideal capacitor, however, ideal inductors are never available, although a real inductor in the low-frequency range is pretty close to ideal and can be described by the expression (7.9). In the RF range, this ideal formula for the inductor impedance is not viable. Instead, a practical chip inductor in the RF range has additional capacitance and additional resistance in parallel and can be modeled as shown in Figure 7.15.

The impedance of an actual chip inductor approaches  $R_P$  when its additional capacitor is resonant with the inductance itself in parallel at an operating frequency. It is called an *infinite inductor* because ideally its impedance approaches infinity at its SRF, that is,

$$\text{SRF}_L = \frac{1}{2\pi\sqrt{LC_P}}, \quad (7.10)$$

where the subscript 'L' denotes the inductor's parameter.

Should the SRF be identical to the operating frequency, the inductor is then an inductor with "infinite" impedance, since the value of the resistor  $R_P$  shown in Figure 7.14 is a high value, or say, approaches infinity at operating frequency. In other words, the infinite inductor is re-recognized as an AC blocking inductor with infinite impedance when the operating frequency is equal to its SRF. As a matter of fact, the alias of the infinite inductor is the RF choke, which most people are familiar with.

The  $\text{SRF}_L$  expression (7.10) is similar to the the  $\text{SRF}_c$  expression (7.8). However, one must bear in mind that the  $\text{SRF}_L$  is a resonant frequency when  $L$  and  $C_P$  are connected in parallel and its impedance approaches infinity at  $\text{SRF}_L$ , while  $\text{SRF}_c$  is a resonant frequency when  $C$  and  $L_S$  are connected in series and its impedance approaches zero at  $\text{SRF}_c$ .

The SRF of a chip inductor depends on the material, art, and technology adapted by the manufacturer. For the chip inductor manufactured by MuRata, the SRF can be formulized as

$$\text{SRF}_L = \frac{8920}{\sqrt{L_{\text{specified}}}}, \quad (7.11)$$

where

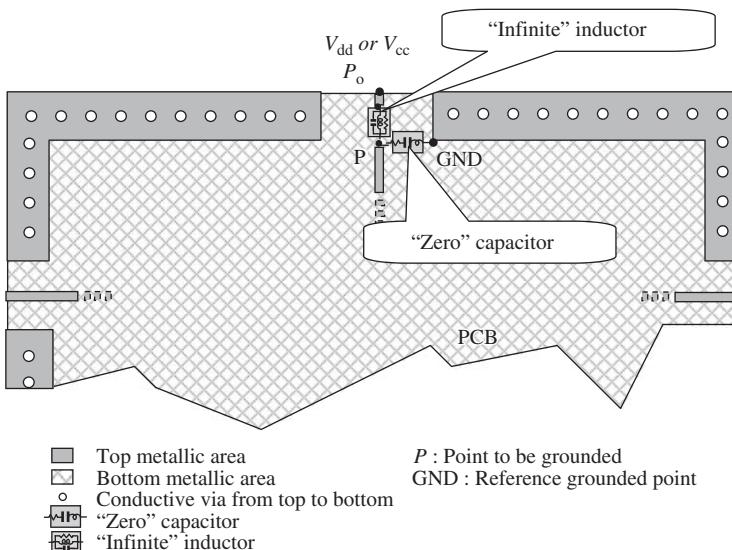
$\text{SRF}_L$  = self-resonant frequency in MHz, and

$L_{\text{specified}}$  = specified inductance by the manufacturer in nH (nanohenry).

Formula (7.11) is available within the range of specified inductance,

$$L_{\text{specified}} = 22-1800 \text{ nH}.$$

If the infinite inductor is not selected from the chip inductor manufactured by MuRata, the formula may be different. By means of the same testing procedures, an appropriate



**Figure 7.16.** RF or AC grounded by a zero capacitor with assistance of an infinite inductor on a PCB.

formula could be found, in which the coefficient “8920” in formula (7.11) would be different and replaced with other value.

In terms of formula (7.11), one can easily find out the infinite inductor for a specific operating frequency.

In Appendix 7.A.2 we discuss how to find the additional capacitance and additional resistance from the actual measurement of a chip inductor, and finally how to obtain the formula (7.11).

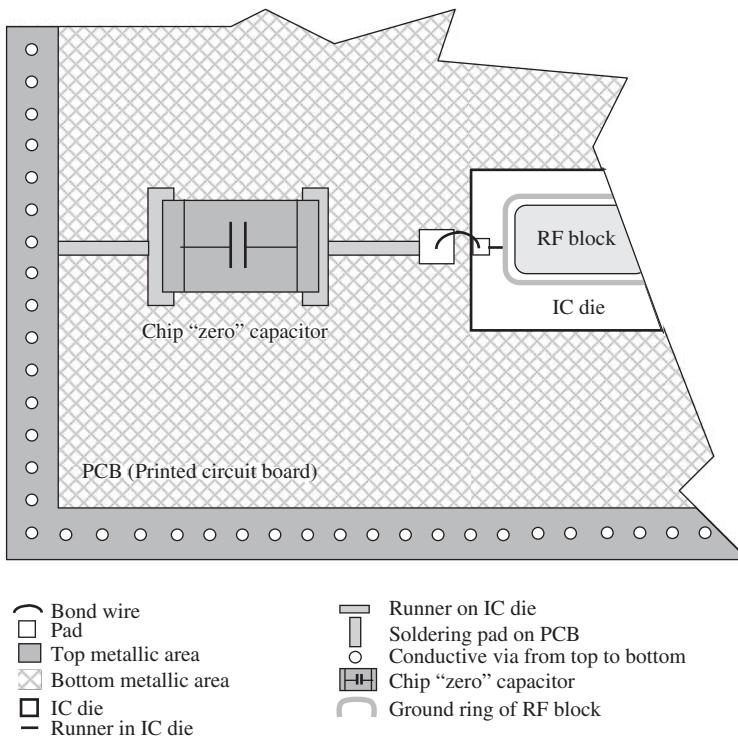
Figure 7.16 shows how an infinite inductor functions as an assistant of zero capacitor in the RF/AC grounding.

For simplicity, Figure 7.16 duplicates everything from Figure 7.10 except the portion containing the point to be RF/AC grounded, P. In Figure 7.16, an infinite inductor is inserted between point  $P_0$  and point P, where  $P_0$  is the outside adjacent point on the same runner before the infinite inductor is inserted. This infinite inductor blocks the RF signal from the outside point  $P_0$  to point P. At point P, the voltage or power of the RF signal from point  $P_0$  is significantly reduced to below a significant level, but the DC voltage at point P is kept the same as that at point  $P_0$ , since the impedance of the infinite inductor approaches infinity. Point P is in fact very close to being RF/AC grounded even if the zero capacitor is imperfect, as long as point  $P_0$  is the unique RF source.

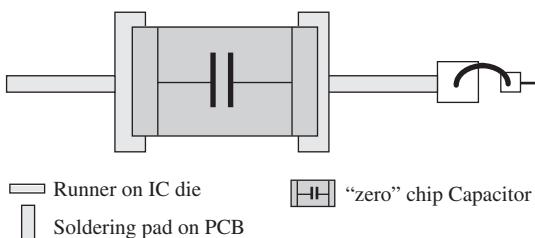
#### 7.4.6 Zero Capacitor in RFIC Design

Today, chip capacitors are widely applied in RF/AC grounding as zero capacitors for a test PCB or RF module. However, there does not seem to be a way that they can be directly applied in an RFIC die. The designer cannot but treat the zero capacitor as an off-chip discrete part. Even so, the chip capacitor is still a good part in RF/AC grounding for RFIC.

Figure 7.17 illustrates how the chip capacitor works together with other parts between the ground ring in the IC die and a ground frame on the PCB. It is assumed that the ground frame on PCB is equipotential with the reference ground point at the DC power supply terminal. The ground ring of an individual RF block in the IC die is a  $P+$  guard ring. It must be forced to be equipotential with the reference ground point at the DC power supply terminal. As shown in Figure 7.17, it is connected to the ground frame on



**Figure 7.17.** Zero capacitor is connected between the  $P+$  guard ring in IC die and the ground frame on PCB.



**Figure 7.18.** Modified equivalent model of zero chip capacitor obtained by addition of soldering pads and runners for RFIC.

the PCB through the following nine parts, which form the modified equivalent model of a zero capacitor for an RFIC and is redrawn in Figure 7.18.

- A short runner in the IC die from the ground ring to a tiny pad;
- A tiny bond pad in the IC die;
- A bond wire in the air between the tiny pad in the IC die and the bonding pad on the PCB;
- A bonding pad on the PCB;
- A runner on the PCB from the bonding pad on the PCB to the soldering pad on the PCB;
- Soldering pad on the PCB;
- A zero chip capacitor;
- Soldering pad on the PCB;
- A runner on the PCB from the zero chip capacitor to the ground frame on the PCB.

The equivalent model of the zero capacitor must be correspondingly modified from Figure 7.14 to Figure 7.18.

There are four differences between the modified equivalent models of the zero capacitor shown in Figure 7.18 and Figure 7.14:

- A short runner in the IC die from the ground ring to a tiny pad;
- A tiny bond pad in the IC die;
- A bond wire in the air between the tiny pad in the IC die and the bonding pad on the PCB;
- A bonding pad on the PCB.

By design experience, the difference of the zero chip capacitor's SRF because of these four additional items is significant. On the other hand, the existing models for these four additional parts are complicated and inaccurate. Especially in the model of the bonding wire between pads, the inaccuracy of the model arises not only from the inconsistent length of the bonding wire in production lines but also from the variation of its inclined angle with respect to the vertical jump. We are therefore aware of the importance and need to develop a zero capacitor directly in the IC die, which may be a remarkable project in the coming years.

## 7.5 QUARTER WAVELENGTH OF MICROSTRIP LINE

### 7.5.1 A Runner is a Part in RF Circuitry

A runner between parts in a circuitry can be a copper line with gold plating on the PCB or a gold line segment on an IC chip. Very often, the circuit designer just ignores its existence, only focusing on the lump parts, such as the inductor, capacitor, resistor, transistor, and so on. Very often, the test results of performance are quite far from agreement with the simulation results of performance. This may be partially due to the inaccuracies of the lump parts applied in the simulation and partially due to the disregarding of the existence of runners. In the RF frequency range, a runner is a part just like other lump parts. Sometimes it may even be more important than a lump part. All the runners in the circuitry must be taken care of in a good simulation. Otherwise, a considerable discrepancy between simulation and testing is inevitable, and sometimes this can put a circuit block out of work.

Essentially, a runner is a microstrip line in the RF frequency range. Its existence equivalents to an additional impedance existing between the two parts connected together by the runner.

The impedance of a microstrip line, based on transmission line theory, can be depicted as Figure 7.19 and can be expressed as

$$Z_P = Z_0 \frac{Z_L \cosh \gamma_t l + Z_0 \sinh \gamma_t l}{Z_0 \cosh \gamma_t l + Z_L \sinh \gamma_t l}, \quad (7.12)$$

where

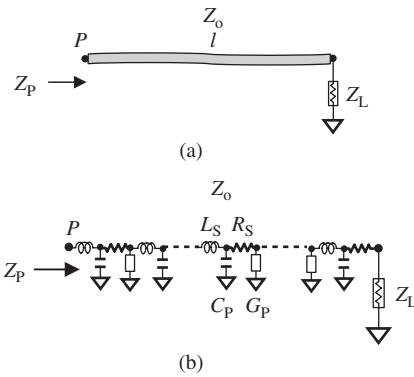
$Z_P$  = impedance at point P, looking into load with a distance  $l$ ,

$Z_0$  = characteristic impedance of the microstrip line,

$Z_L$  = load impedance of the microstrip line,

$l$  = length of the microstrip line, and

$\gamma_t$  = transmission coefficient.



**Figure 7.19.** A microstrip line with its characteristic and load impedances,  $Z_o$ ,  $Z_L$  and distribution parameters,  $R_{ms}$ ,  $L_{ms}$ ,  $G_{mp}$ ,  $C_{mp}$ . (a) A microstrip line with its characteristic and load impedance,  $Z_o$ ,  $Z_L$ . (b) A microstrip line with its distribution parameters,  $R_{ms}$ ,  $L_{ms}$ ,  $G_{mp}$ ,  $C_{mp}$ .

The characteristic impedance  $Z_o$  and the transmission coefficient  $\gamma$  can be expressed as

$$Z_o = \sqrt{\frac{R_{ms} + j\omega L_{ms}}{G_{mp} + j\omega C_{mp}}} \approx \sqrt{\frac{L_{ms}}{C_{mp}}}, \quad (7.13)$$

$$\gamma_t = \sqrt{(R_{ms} + jL_{ms}\omega)(G_{mp} + jC_{mp}\omega)} \approx \alpha + j\beta, \quad (7.14)$$

if

$$R_{ms} \ll \omega L_{ms}, \quad (7.15)$$

$$G_{mp} \ll \omega C_{mp}, \quad (7.16)$$

where

- $R_{ms}$  = resistance per unit length of the microstrip line,
- $L_{ms}$  = inductance per unit length of the microstrip line,
- $G_{mp}$  = conductance per unit length of the microstrip line,
- $C_{mp}$  = capacitance per unit length of the microstrip line,
- $\alpha$  = attenuation per unit length of the microstrip line, and
- $\beta$  = phase shift per unit length of the microstrip line.

A closer approximation for  $\alpha$  and  $\beta$  can be obtained by rearranging expression (7.14) for  $\gamma_t$  and using the binomial expression. Thus,

$$\alpha \approx \frac{1}{2} \left( \frac{R_{ms}}{\sqrt{L_{ms}/C_{mp}}} + G \sqrt{L_{ms}/C_{mp}} \right), \quad (7.17)$$

$$\beta \approx \omega \sqrt{L_{ms} C_{mp}}. \quad (7.18)$$

Figure 7.19 shows a microstrip line with its characteristic and load impedances,  $Z_o$  and  $Z_L$ , and its distribution parameters,  $R_{ms}$ ,  $L_{ms}$ ,  $G_{mp}$ , and  $C_{mp}$ .

In the cases when the attenuation is negligible, that is,

$$\alpha \rightarrow 0, \quad (7.19)$$

and conditions (7.15) and (7.16) are satisfied, expression (7.12) becomes

$$Z_P = Z_o \frac{Z_L \cos \beta l + jZ_o \sin \beta l}{Z_o \cos \beta l + jZ_L \sin \beta l}, \quad (7.20)$$

or

$$Z_P = Z_o \frac{Z_L \cos \frac{2\pi l}{\lambda} + jZ_o \sin \frac{2\pi l}{\lambda}}{Z_o \cos \frac{2\pi l}{\lambda} + jZ_L \sin \frac{2\pi l}{\lambda}}, \quad (7.21)$$

where  $\lambda$  is the electrical wavelength, and

$$\beta = \frac{2\pi}{\lambda}. \quad (7.22)$$

If the runner is considered as a part in the circuitry, the special feature of this part is that its impedance depends on not only the self-parameters, the length of runner  $l$ , and the characteristic impedance  $Z_o$ , but also the load impedance,  $Z_L$ .

Now, let us examine two special cases:

1. When the load is short circuited, that is,

$$Z_L = 0, \quad (7.23)$$

then

$$Z_P|_{Z_L \rightarrow 0} = jZ_o \tan \beta l = jZ_o \tan \frac{2\pi}{\lambda} l. \quad (7.24)$$

2. When the load is open circuited, that is,

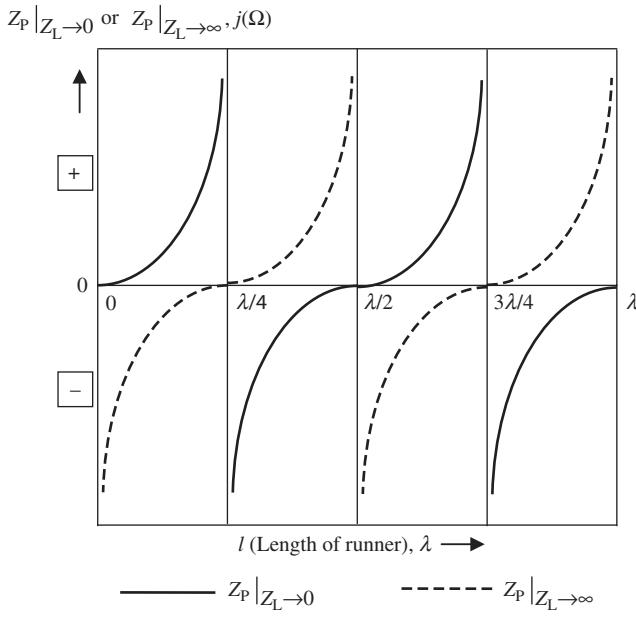
$$Z_L \rightarrow \infty, \quad (7.25)$$

then

$$Z_P|_{Z_L \rightarrow \infty} = \frac{1}{j} Z_o \cot \beta l = -jZ_o \cot \frac{2\pi}{\lambda} l. \quad (7.26)$$

Figure 7.20 plots the curves of  $Z_P|_{Z_L \rightarrow 0}$  and  $Z_P|_{Z_L \rightarrow \infty}$  versus  $l$ , by which one can find out the following:

1. If the length of the runner is varied from 0 to  $\lambda/4$  or from  $\lambda/2$  to  $3\lambda/4$ ,
  - the impedance of a runner  $Z_P$  is changed from 0 to  $\infty$  when the load is short circuited, that is,  $Z_L = 0$ ;
  - the impedance of a runner  $Z_P$  is changed from  $-\infty$  to 0 when the load is open circuited, that is,  $Z_L = \infty$ .
2. If the length of the runner is varied from  $\lambda/4$  to  $\lambda/2$  or from  $3\lambda/4$  to  $\lambda$ ,
  - the impedance of a runner  $Z_P$  is changed from  $-\infty$  to 0 when the load is short circuited, that is,  $Z_L = 0$ ;



**Figure 7.20.** Impedance of runner when  $Z_L = 0$  and  $Z_L = \infty$ .

- the impedance of a runner  $Z_P$  is changed from 0 to  $\infty$  when the load is open circuited, that is,  $Z_L = \infty$ .

Alternatively, in spite of the short-circuited load,  $Z_L = 0$ , or open-circuited load,  $Z_L \rightarrow \infty$ , the following can be concluded.

1. The variation of impedance due to the existence of runner must be ignored if

$$l \ll \lambda/4, \quad (7.27)$$

or

$$l \approx n\lambda/2 = 2n\lambda/4, \quad n = 1, 2, 3, 4, \dots \quad (7.28)$$

because under conditions (7.27) or (7.28), expressions (7.24) and (7.26) become

$$Z_P|_{Z_L \rightarrow 0} \rightarrow 0, \quad (7.29)$$

$$Z_P|_{Z_L \rightarrow \infty} \rightarrow \infty, \quad (7.30)$$

which implies that

$$Z_P \approx Z_L. \quad (7.31)$$

This means that the insertion of the runner does not introduce any conceivable change of impedance.

2. The variation of impedance due to the existence of runner must be

$$l \approx (2n + 1)\lambda/4, \quad (7.32)$$

because under this condition,

$$Z_P|_{Z_L \rightarrow o} \rightarrow \pm\infty, \quad (7.33)$$

$$Z_P|_{Z_L \rightarrow \infty} \rightarrow 0, \quad (7.34)$$

which implies that  $Z_P$  and  $Z_L$  are at extreme opposites, that is

$$Z_P \rightarrow \pm\infty, \quad \text{if } Z_L \rightarrow 0, \quad (7.35)$$

$$Z_P \rightarrow 0, \quad \text{if } Z_L \rightarrow \infty, \quad (7.36)$$

This indicates that the insertion of the runner brings about an earth-shaking change in impedance.

3. The variation of impedance due to the existence of runner must be taken care of if the length is neither as in the cases of (7.27) and (7.28) nor as in the case of (7.32), because in such a case, the variation of impedance due to the existence of runner is usually comparable with the impedance of other parts.

### 7.5.2 Why Quarter Wavelength is so Important

It should be noted that condition (7.28),  $l \approx n\lambda/2$ , is intentionally rewritten as  $l \approx n\lambda/2 = 2n\lambda/4$ . This is due to the recognition that all three conditions (7.27), (7.28), and (7.32) are related to the quarter wavelength  $\lambda/4$ .

Let us furthermore examine conditions (7.27), (7.28), and (7.32).

In the layout for an RF circuit block or an RF system, it is always sound to shorten the length of the runner between two parts as much as possible. The theoretical background is condition (7.27). It says that the length of the runner must be greatly shorter than the quarter wavelength. If so, the insertion of the runner does not introduce a conceivable change in impedance. In other words, the additional impedance due to the existence of the runner can be neglected or ignored.

If the length of a runner is about an even multiple of the quarter wavelength as shown in condition (7.28), a case similar to (7.27) exists. If so, the insertion of the runner does not introduce a conceivable change in impedance. In other words, the additional impedance due to the existence of the runner can be neglected or ignored.

However, if the length of a runner is about one quarter wavelength or an odd multiple of the quarter wavelength as shown in condition (7.32), the insertion of the runner introduces an earth-shaking change in impedance. The additional impedance due to the existence of the runner approaches infinity. The function of the runner in the circuitry dominates that of the other parts and, consequently, will put the circuitry out of work.

The three cases of the runner's length, (7.27), (7.28), and (7.32), are the criteria to judge the importance of a runner. These criteria become useful only if the quarter wavelength of the runner is well understood: this is why it is so important to understand the quarter wavelength.

Whether the circuit is built by discrete parts as a module or built on an IC chip, the first thing or first step in designing the layout of the circuit is to know how long

the quarter wavelength will be in the actual layout PCB or on the actual IC substrate. Otherwise, it is impossible to judge whether a runner is too short or too long.

The quarter wavelength can be calculated by formulas, but it is better obtained by actual testing.

### 7.5.3 Magic Open-Circuited Quarter Wavelength of Microstrip Line

If two parts are connected by a runner and the length of the runner is about one quarter wavelength or an odd multiple of the quarter wavelength as shown in condition (7.32), it equates to inserting a special part with infinite impedance between the two parts, which leaves the connection between these two parts broken or open circuited. This is, of course, an unexpected worst case.

Just like many things in the universe, the worst and best cases switch positions under certain circumstances. One or odd multiples of quarter wavelength runner is not good to make connection between two parts. It is, however, a magic part in the task of RF/AC grounding or isolation if one end of the runner is kept in the open-circuited state.

Figure 7.21 shows a runner or a microstrip line a quarter wavelength long, with an open-circuited load. From (7.25), (7.26), and (7.32), we have

$$Z_P|_{Z_L \rightarrow \infty} = jZ_0 \cot \frac{2\pi}{\lambda} l \Big|_{l=(2n+1)\lambda/4} = 0, \quad (7.37)$$

Expression (7.37) indicates that point P of the runner is grounded, since the values of  $Z_L$  and  $Z_P$  appearing in Figure 7.21 are the impedances with respect to the reference ground point.

It is therefore concluded that in order to ground point P, a simple connection by a conductive metallic runner will work well if one end of the runner is connected to point P and another end is open circuited and if the length of the runner is an odd multiple of the quarter wavelength as shown in equation (7.32). This technology is called *compulsory* or *enforced* or *coercing grounding*.

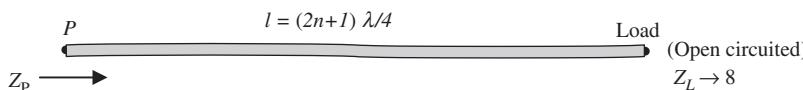
However, it should be noted that at the open-circuited load, the infinite impedance is just a theoretical approximation. Instead, some spray capacitance always exists, especially in a high RF frequency range. Consequently, in the practical product, the actual length of a quarter wavelength microstrip line for compulsory grounding purposes is a little bit shorter than the quarter wavelength.

From expression (7.32) it can be seen that the shortest length of a microstrip line from point P to the open-circuited load is

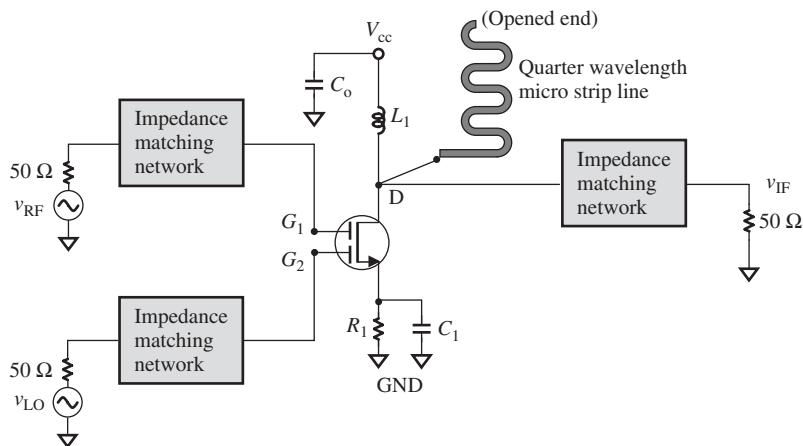
$$l = \frac{\lambda}{4}, \quad (7.38)$$

when

$$n = 0. \quad (7.39)$$



**Figure 7.21.** Point P is grounded by microstrip line if  $l = (2n + 1)\lambda/4$  with an open-circuited load, that is,  $Z_L \rightarrow \infty$ .



**Figure 7.22.** Isolation scheme by applying a quarter-wavelength microstrip line in a mixer.

A quarter wavelength microstrip line can be applied to improve isolation. The following example shows how the quarter wavelength microstrip line improves isolation in a mixer.

The device applied in the mixer shown in Figure 7.22 is a dual-gate metal–oxide–semiconductor field-effect transistor (MOSFET). One of the gates is injected with the LO injection and another gate is the input of RF signal. Owing to the application of the dual-gate MOSFET, the isolation between the input and output is poor. The LO injection at the LO port could leak to the RF or IF ports. Similarly, RF power from the RF port could leak to the LO or IF ports. The LO power leakage into the IF portion usually dominates over RF power leakage to the IF port because the LO injection has the highest power among these three ports.

By means of a quarter-wavelength microstrip line, a mixer can perform with excellent isolation between the LO, RF, and IF ports. As shown in Figure 7.22, a quarter-wavelength microstrip line corresponding to the LO frequency is attached to the drain of the dual-gate MOSFET at one end. Another end of this quarter wavelength microstrip line is in an opened-circuited state. Consequently, at the connected point, the drain of the MOSFET, the possible leakage site of the LO injection from gate 1 to point D is “absorbed” or “rejected” by the quarter wavelength microstrip line because at point D, any LO leaked injection is forced to be zero by the quarter-wavelength microstrip line. In other words, any voltage at the LO frequency is impossible to be established at point D, the drain of the MOSFET.

On the other hand, the RF frequency is usually close to the LO frequency. The microstrip line with an LO quarter wavelength can be very helpful to the isolation between the RF and IF ports at the same time, because the bandwidth of the attached quarter-wavelength microstrip line usually covers the RF frequency as well, so that the leaked RF signal from gate 2 to point D can likewise be absorbed or rejected.

In general, isolation can be enhanced by more than 20 dB with this technology.

Alternatively, the quarter-wavelength microstrip line corresponding to the LO frequency in Figure 7.22 can be replaced by a quarter-wavelength cable corresponding to the LO frequency. However, cables are too clumsy and thus are seldom applied in practical engineering.

A question is then raised: does the attached quarter-wavelength microstrip line radiate the power at the LO frequency since it looks like a small antenna? Yes, indeed! The

primary purpose of the attached quarter-wavelength microstrip line is to suck out the leakage of LO injection at point D, the drain of the dual-gate MOSFET. The absorbed LO leakage power therefore radiates to the space around the circuitry. However, the radiated power should be very low, negligible to the circuitry, and should not be a concern.

#### 7.5.4 Testing for Width of Microstrip Line with Specific Characteristic Impedance

In terms of the network analyzer, the RF circuit designer should be able to characterize a runner or microstrip line on a specific PCB. The first parameter to be studied is the width of the microstrip line or runner with a specific characteristic impedance.

Let us assume that the specific characteristic impedance of the runner is the standard reference impedance, that is,

$$Z_0 = 50 \Omega. \quad (7.40)$$

It is well known that the characteristic impedance  $Z_0$  of a runner or microstrip line is related to the width of runner  $W$ , which can be determined through impedance testing by a network analyzer.

If the load is set with the specific characteristic impedance, that is,

$$Z_L = Z_0 = 50 \Omega, \quad (7.41)$$

then from equation (7.21), we have

$$Z_P = Z_0, \quad (7.42)$$

regardless of how long the microstrip line is. This is the theoretical background of our test.

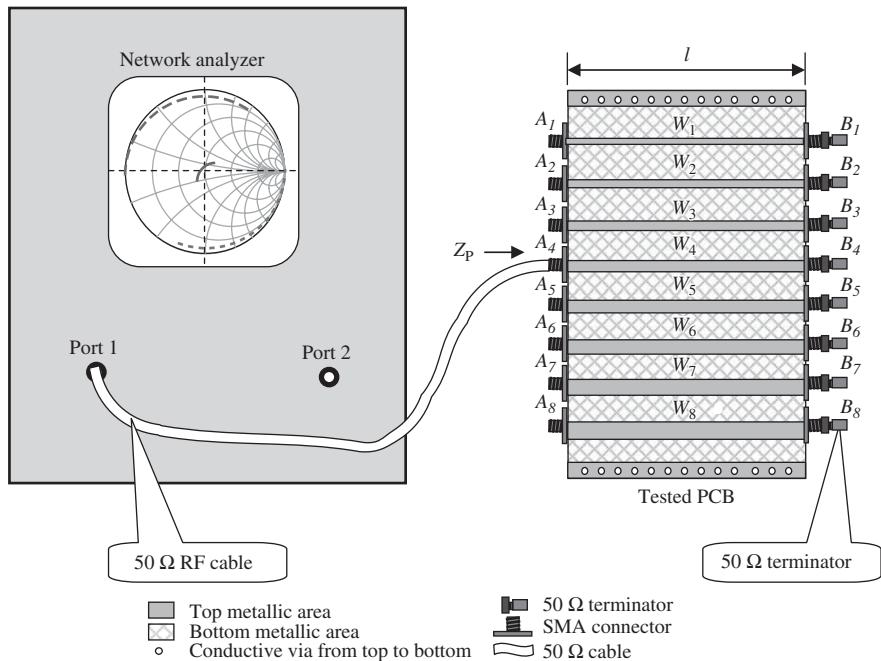
A PCB to be tested is shown on the right-hand side of Figure 7.23, in which many microstrip line or runners with different widths  $W_i$  but the same length are printed. Terminal  $A_i$  is connected to port 1 of the network analyzer, and terminal  $B_i$  is supposed to be connected with a  $50 \Omega$  terminator for all the microstrip lines. Testing can be started by connecting the  $50 \Omega$  RF cable to the terminal  $A_1$  and switching it to,  $A_2, A_3, A_4$ , and so on. (Owing to space restrictions, in Figure 7.23, only eight runners are drawn. The number of runners could be greater for higher accuracy of testing.) When the trace of  $S_{11}$  on the Smith Chart is moved to the center  $50 \Omega$  as a terminal  $A_i$  is connected, the microstrip line with terminal  $A_i$  is a microstrip line with characteristic impedance  $50 \Omega$ .

In RFIC design, the test PCB would be replaced by the test IC die on the wafer.

#### 7.5.5 Testing for Quarter Wavelength

The quarter wavelength can be obtained from the same setup as shown in Figure 7.23. Let us assume that the fourth runner with terminal  $A_4$  and  $B_4$  was selected by the testing described in Section 7.5.4. Its width  $W_4$  corresponds to the specific characteristic impedance  $Z_0 = 50 \Omega$ . We will focus our attention on this runner and ignore the others.

By simply replacing the  $50 \Omega$  terminator with either a “short” or “open” terminator, the quarter wavelength can be calculated from the readings of the impedance trace on the Smith Chart.



**Figure 7.23.** Impedance testing for determination of runner's width corresponding to 50  $\Omega$  characteristic impedance.

From equations (7.24) and (7.26) it can be seen that  $Z_P$  becomes a pure reactance in both cases when the 50  $\Omega$  terminator is replaced by either a short or open terminator, that is,

$$Z_P|_{Z_L \rightarrow 0} = jX_P|_{Z_L \rightarrow 0} = jZ_0 \tan \frac{2\pi}{\lambda} l, \quad (7.43)$$

$$Z_P|_{Z_L \rightarrow \infty} = jX_P|_{Z_L \rightarrow \infty} = -jZ_0 \cot \frac{2\pi}{\lambda} l, \quad (7.44)$$

Expressions (7.43) and (7.44) can be rewritten as

$$x|_{Z_L \rightarrow 0} = \tan \frac{2\pi}{\lambda} l, \quad (7.45)$$

$$x|_{Z_L \rightarrow \infty} = -\cot \frac{2\pi}{\lambda} l, \quad (7.46)$$

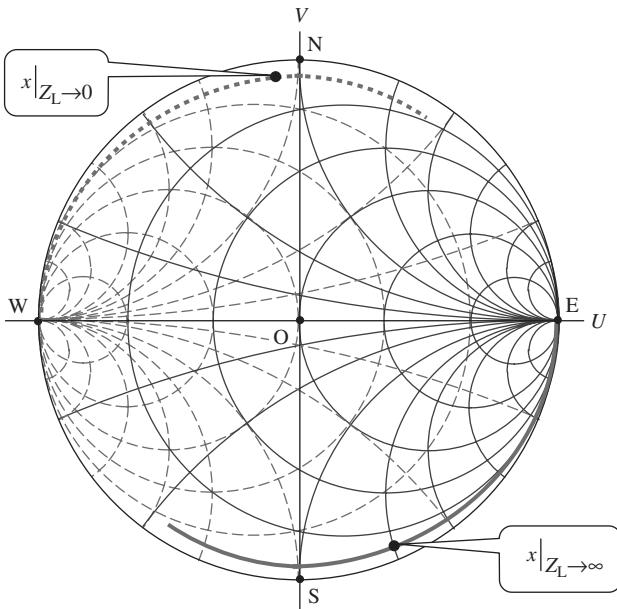
where

$$x|_{Z_L \rightarrow 0} = \frac{Z_P|_{Z_L \rightarrow 0}}{Z_0}, \quad (7.47)$$

and

$$x|_{Z_L \rightarrow \infty} = \frac{Z_P|_{Z_L \rightarrow \infty}}{Z_0}, \quad (7.48)$$

are the normalized reactances when  $Z_L$  is short circuited and open circuited, respectively. They are the readings from the impedance trace on the Smith Chart.



**Figure 7.24.** Two normalized reactances,  $x|_{Z_L \rightarrow 0}$  and  $x|_{Z_L \rightarrow \infty}$ , displayed on Smith Chart.

Figure 7.24 shows the two normalized reactances,  $x|_{Z_L \rightarrow 0}$  and  $x|_{Z_L \rightarrow \infty}$ , in the short-circuited and open-circuited cases, respectively. The two impedance traces somehow deviate from the big circle,  $r = 0$ , of the Smith Chart. This is because of the additional resistance that exists in a microstrip line or runner.

From expressions (7.45) and (7.46), the quarter wavelength can be calculated as

$$\frac{\lambda}{4} = \frac{\pi l}{2 \tan^{-1}(x|_{Z_L \rightarrow 0})}, \quad (7.49)$$

or

$$\frac{\lambda}{4} = -\frac{\pi l}{2 \cot^{-1}(x|_{Z_L \rightarrow \infty})}. \quad (7.50)$$

## APPENDICES

### 7.A.1 Characterizing of Chip Capacitor and Chip Inductor by Means of $S_{21}$ Testing

As IC technology developed in past decades, discrete parts have been replaced by IC chips on a large scale. However, the chip capacitor, chip inductor, and chip resistor are still indispensable parts in the implementation of a circuit module or system today. The reasons are as follows:

- They are the parts with advantages of small size, cost-efficiency, and reliability in the implementation of the circuit module or system.
- Even in a module or system replaced by IC chips, the zero chip capacitor must be applied in the testing.

- The chip inductor is very often applied as an off-chip part because the  $Q$  value of an on-chip inductor is too low.

This appendix is introduced for the characterization of the chip capacitor, chip inductor, and chip resistor, which are mainly abstracted from the author's design work in 1991.

The impedance on the Smith Chart is a complex number, and its real and imaginary parts can be changed from zero to infinity. Theoretically, the impedance of a chip part can be measured by single port testing, either  $S_{11}$  or  $S_{22}$ , using a network analyzer. Unfortunately, as pointed out in Section 6.4.1, at the very low- or high-impedance regions on the Smith Chart, the readings are quite inaccurate. On the other hand, the value of additional stray resistance  $R_S$  in a chip capacitor is very low as shown in Figure 7.9, whereas the value of additional stray resistance  $R_p$  in a chip inductor as shown in Figure 7.15 is very high. Consequently, it is inappropriate to characterize the chip capacitor or chip inductor by  $S_{11}$  or  $S_{22}$  testing.

Instead, the chip capacitor or chip inductor is characterized by  $S_{21}$  testing.

Figures 7.A.1 and 7.A.2 show the test setup for the characterization of chip capacitor and chip inductor, respectively.

Figure 7.A.1 shows the  $S_{21}$  testing setup for a chip capacitor. The tested chip capacitor is soldered between the  $50 \Omega$  microstrip line and ground in parallel on a small PCB as a DTU. Then the small PCB is connected with ports 1 and 2 of the network analyzer by  $50 \Omega$  RF cables. As a result, the tested chip capacitor is connected to ports 1 and 2 of the network analyzer in parallel accordingly.

Figure 7.A.2 shows the  $S_{21}$  testing setup for a chip inductor. The tested chip inductor is inserted and soldered between two short  $50\text{-}\Omega$  microstrip lines in series on a small PCB as a DTU. Then the small PCB is connected with ports 1 and 2 of the network analyzer by  $50\text{-}\Omega$  RF cables. As a result, the tested chip inductor is connected to ports 1 and 2 of the network analyzer in series accordingly.

By Ralph Carson's derivation (1975),  $S_{21}$  can be expressed as a function of the input and output parameters. Figure 7.A.3 shows the various parameters at the source and load.

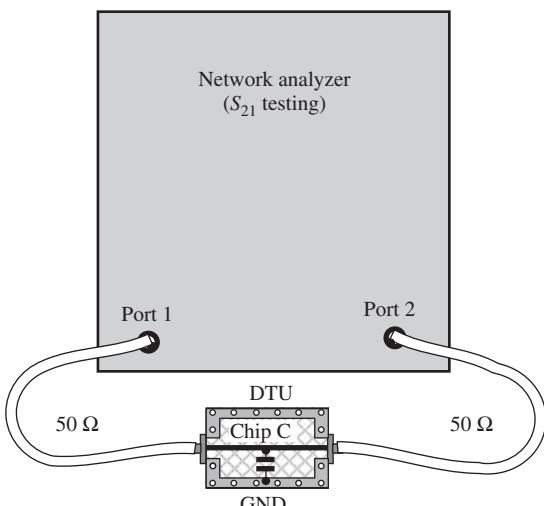


Figure 7.A.1.  $S_{21}$  testing for a chip capacitor.

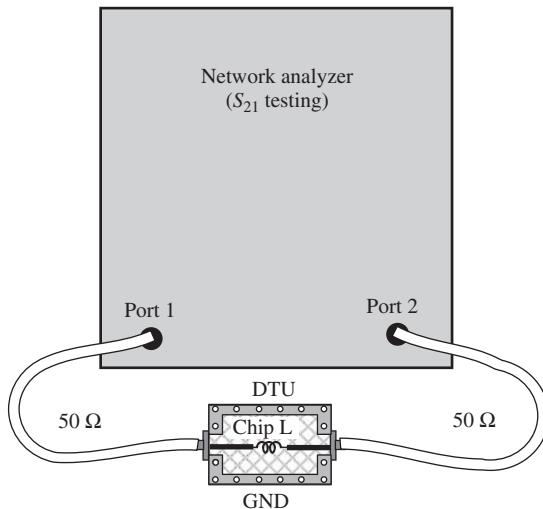


Figure 7.A.2.  $S_{21}$  testing for a chip inductor.

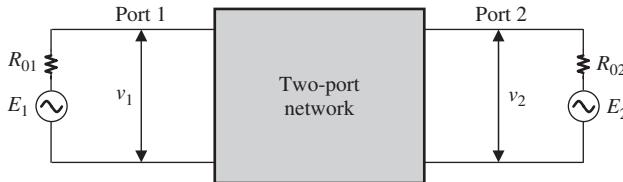


Figure 7.A.3. Source and load parameters in a two-port network.

The relationship between  $S_{21}$  and the various parameters can be expressed as

$$S_{21} = 2 \sqrt{\frac{R_{01}}{R_{02}}} \frac{v_2}{E_1}, \quad (7.A.1)$$

if

$$E_2 = 0, \quad (7.A.2)$$

where

- $v_1$  = input voltage,
- $v_2$  = output voltage,
- $E_1$  = voltage at source,
- $E_2$  = voltage at load,
- $R_{01}$  = source resistance, and
- $R_{02}$  = load resistance.

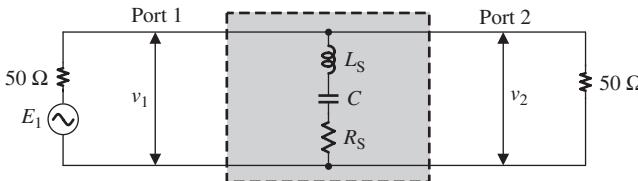
And if

$$R_{01} = R_{02} = 50 \Omega, \quad (7.A.3)$$

then, equation (7.A.1) becomes

$$S_{21} = 2 \frac{v_2}{E_1}. \quad (7.A.4)$$

The ratio  $v_2/E_1$  is determined by the DTU. It can be expressed by the parts of the DTU.



**Figure 7.A.4.** Replacement of “two-port network” in Figure 7.A.3 by chip capacitor’s equivalent model.

Let us replace the “two-port network” in Figure 7.A.3 by the equivalent model of the chip capacitor as shown in Figure 7.9; then, Figure 7.A.3 becomes Figure 7.A.4.

The ratio  $v_2/E_1$  can be found through a simple but tedious mathematic derivation

$$\frac{v_2}{E_1} = \frac{\sqrt{\left[R_s(50+2R_s) + \left(L_s\omega - \frac{1}{C\omega}\right)^2\right]^2 + \left[(50+2R_s)\left(L_s\omega - \frac{1}{C\omega}\right) - 2R_s\left(L_s\omega - \frac{1}{C\omega}\right)\right]^2}}{(50+2R_s)^2 + 4\left(L_s\omega - \frac{1}{C\omega}\right)^2}, \quad (7.A.5)$$

At the SRF,

$$\omega = \omega_{\text{SRF}} = \frac{1}{\sqrt{L_s C}}, \quad (7.A.6)$$

$$\frac{v_2}{E_1} = \frac{R_s}{50 + 2R_s}, \quad (7.A.7)$$

then,

$$S_{21,\text{SRF}} = 20 \log \left( 2 \frac{v_2}{E_1} \right) = 20 \log \left( 2 \frac{R_s}{50 + 2R_s} \right), \quad (7.A.8)$$

or

$$R_s = \frac{50}{2 \times \left( 10^{-\frac{S_{21,\text{SRF}}}{20}} - 1 \right)}, \quad (7.A.9)$$

$$L_s = \frac{1}{\omega_{\text{SRF}}^2 C}, \quad (7.A.10)$$

$$C = C_{\text{specified}}. \quad (7.A.11)$$

where

$S_{21,\text{SRF}} = S_{21}$  at the SRF,

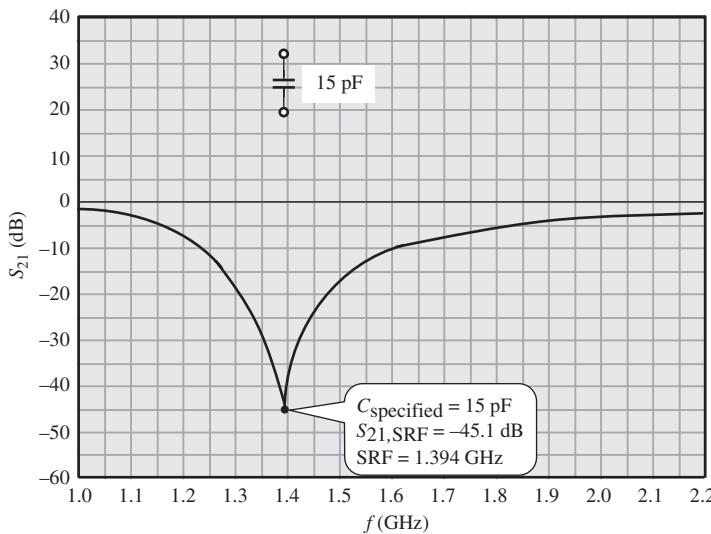
$\omega_{\text{SRF}}$  = angular SRF, and

$C_{\text{specified}}$  = specified value of capacitance by manufacturer.

Figure 7.A.5 shows the  $S_{21}$  frequency response of chip capacitor  $C = C_{\text{specified}} = 15 \text{ pF}$  with the test setup as shown in Appendix 7.A.1.

This is a test example for a chip capacitor with

$$C_{\text{specified}} = 15 \text{ pF},$$



**Figure 7.A.5.**  $S_{21}$  frequency response of chip capacitor  $C = C_{\text{specified}} = 15 \text{ pF}$  with test setup as shown in Figure 7.A.1.

and it is found that

$$\omega_{\text{SRF}} = 2\pi \times 1.394 \text{ GHz},$$

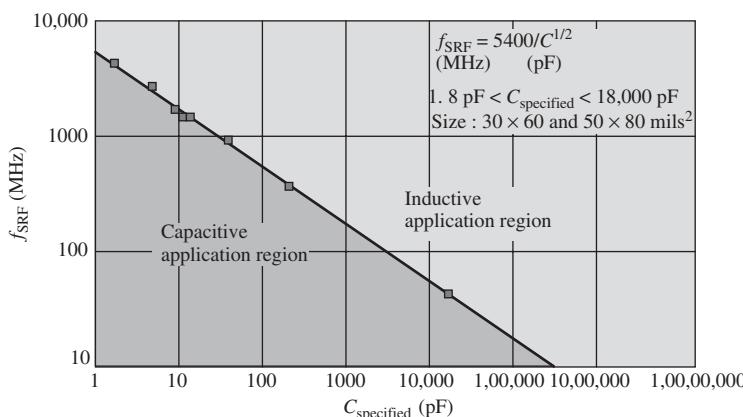
$$S_{21,\text{SRF}} = -45.1 \text{ dB}.$$

From expressions (7.A.9), (7.A.10), and (7.A.11), it can be seen that the values of  $R_S$  and  $L_S$  can be obtained from the reading of  $S_{21,\text{SRF}}$  at the SRF,  $\omega_{\text{SRF}}$ .

$$L_S = 0.83 \text{ nH},$$

$$R_S = 0.14 \Omega.$$

Figure 7.A.6 shows the SRFs of various MuRata chip capacitors. The tested range of capacitance is from 1.8 to 18,000 pF with two different sizes:  $30 \times 60$  and  $50 \times 80 \text{ mil}^2$ .



**Figure 7.A.6.** Plot of the SRF  $f_{\text{SRF}}$  versus specified value of chip capacitor  $C_{\text{specified}}$ , extracted from testing of MuRata chip capacitors.

TABLE 7.A.1. Some Important  $SRF_C$  (Self-Resonant Frequency) Values of Chip Capacitors

$SRF_C$ , MHz	Value of Chip Capacitor, pF
40	18,000
150	1296
450	144
500	117
800	46
900	36
1000	29
1500	13
2400	5.1
5400	1.0
6235	0.75

In the logarithmic coordinates, all the tested points are crowded on a straight line. Consequently, the SRFs can be formulized by a simple equation, that is,

$$SRF_C = \frac{5400}{\sqrt{C}}, \quad (7.A.12)$$

where the unit of  $C = C_{\text{specified}}$  is pF, and the unit of  $SRF_C$  is MHz.

Equation (7.A.12) proves extremely convenient to designers in the calculation of the SRF of a chip capacitor. However, instead of equation (7.A.12), an experienced engineer should be familiar with the values of the special chip capacitors listed in Table 7.A.1.

Figure 7.A.7 shows the variation of the in-series equivalent parasitic inductance  $L_S$  as  $C_{\text{specified}}$  is varied, which is simply calculated from the SRF via equations (7.A.10) and (7.A.11) for MuRata chip capacitors.

It is found that the in-series parasitic inductance  $L_S$  is kept almost constant from the low-capacitance parts to high-capacitance parts. Its average value is about 0.86 nH.

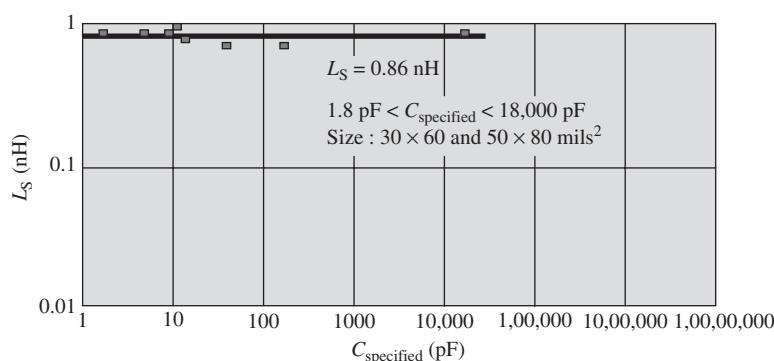
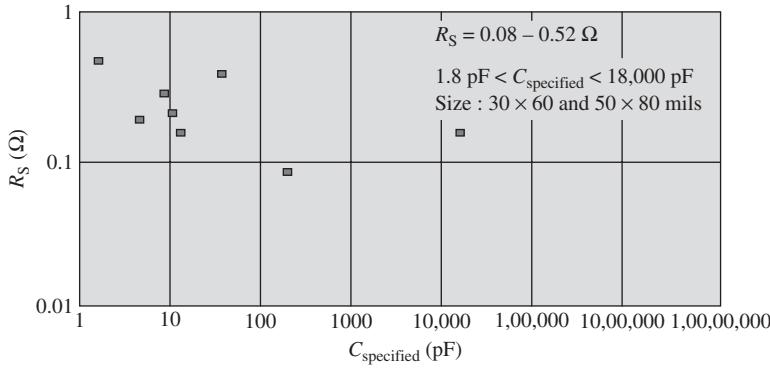
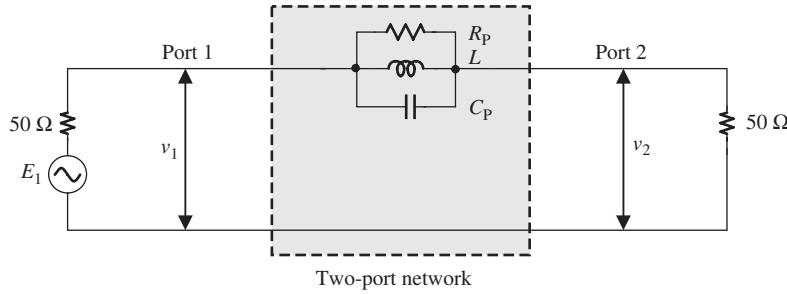


Figure 7.A.7. Plot of in-series parasitic inductance  $L_S$  versus specified capacitance  $C_{\text{specified}}$ , extracted from testing of MuRata chip capacitors.



**Figure 7.A.8.** Plot of in-series resistance  $R_S$  versus specified capacitance  $C_{\text{specified}}$ , extracted from testing of MuRata chip capacitors.



**Figure 7.A.9.** Replacement of two port network in Figure 7.A.3 by chip inductor's equivalent model.

Finally from the reading of  $S_{21,\text{SRF}}$ , we can calculate the parasitic resistance in series,  $R_S$ , though the equation (7.A.9). Figure 7.A.8 plots its result. The tested result displays that it is scattered between 0.08 and 0.52 Ω when  $C_{\text{specified}}$  is between 1.8 and 18,000 pF.

Similarly, let us replace the two port network in Figure 7.A.3 by the equivalent model of chip inductor as shown in Figure 7.15: then, Figure 7.A.3 becomes 7.A.9.

By a simple but tedious mathematic derivation, it is easy to find

$$\frac{v_2}{E_1} = \frac{50}{100 + R_p \left[ 1 + (R_p C_p \omega)^2 \left( 1 - \frac{1}{LC_p \omega^2} \right)^2 \right]^{1/2}}, \quad (7.A.13)$$

we have

$$S_{21} = 20 \log \left( 2 \frac{v_2}{E_1} \right) = 20 \log \frac{100}{100 + R_p \left[ 1 + (R_p C_p \omega)^2 \left( 1 - \frac{1}{LC_p \omega^2} \right)^2 \right]^{1/2}}. \quad (7.A.14)$$

At the SRF,

$$\omega = \omega_{\text{SRF}} = \frac{1}{(LC_p)^{1/2}}, \quad (7.A.15)$$

$$S_{21,\text{SRF}} = 20 \log \frac{100}{100 + R_p}, \quad (7.\text{A}.16)$$

or,

$$R_p = 100 \left[ 10^{-\frac{S_{21,\text{SRF}}}{20}} - 1 \right]. \quad (7.\text{A}.17)$$

$$C_p = \frac{1}{\omega_{\text{SRF}}^2 L}. \quad (7.\text{A}.18)$$

$$L = L_{\text{specified}}, \quad (7.\text{A}.19)$$

where

$S_{21,\text{SRF}}$  =  $S_{21}$  at the SRF,

$\omega_{\text{REF}}$  = angular SRF, and

$L_{\text{specified}}$  = specified value of inductance by manufacturer.

From expressions (7.A.17), (7.A.18), and (7.A.19), it can be seen that the values of  $R_p$  and  $C_p$  can be obtained from the reading of  $S_{21,\text{REF}}$  at SRF,  $\omega_{\text{SRF}}$ .

Figure 7.A.10 shows the  $S_{21}$  frequency response of chip inductor  $L = L_{\text{specified}}$  with test setup as shown in Appendix 7.A.2.

This is a test example for a chip inductor with

$$L_{\text{specified}} = 60 \text{ nH},$$

and it is found that

$$\omega_{\text{SRF}} = 2\pi \times 1.580 \text{ GHz},$$

$$S_{21,\text{SRF}} = -45.9 \text{ dB}.$$

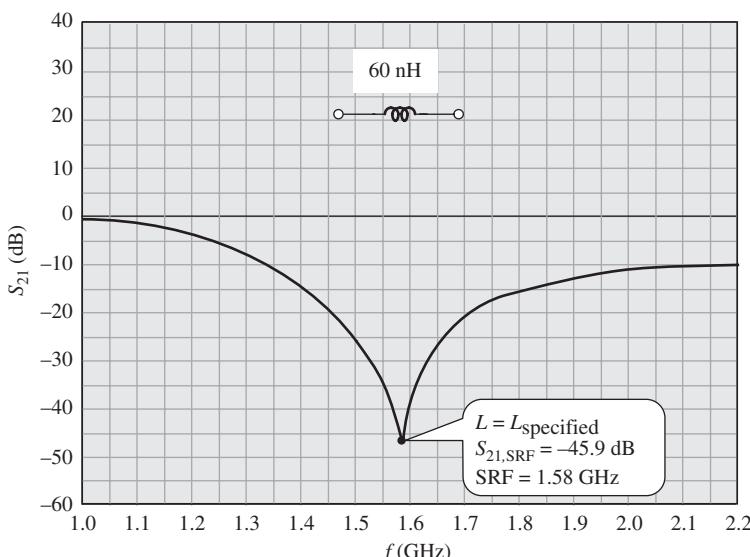


Figure 7.A.10.  $S_{21}$  frequency response of chip capacitor  $L = 60 \text{ nH}$  with test setup as shown in Figure 7.A.2.

From expressions (7.A.17), (7.A.18), and (7.A.19), it can be seen that the values of  $R_P$  and  $C_P$  can be obtained from the reading of  $S_{21,SRF}$  at the SRF,  $\omega_{SRF}$ .

$$C_P = 0.17 \text{ pF},$$

$$R_P = 19,624 \Omega.$$

Figure 7.A.11 shows the  $SRF_L$  of some MuRata chip inductors. The testing has been conducted for one size:  $80 \times 120$  mils<sup>2</sup>. It is a sampling test from  $L_{\text{specified}} = 22$ – $1800$  nH. Figure 7.A.11 illustrates the characteristics of the LC resonant circuit in parallel for a chip inductor, which appears as an inductor by itself when it is operating below its SRF. The bottom left area of Figure 7.A.11 below the  $f_{SRF}$  line is therefore called the *inductive application region*. On the contrary, it appears like a capacitor when it is operating above its SRF. The top right area of Figure 7.A.11 above the  $f_{SRF}$  line is thus called the *capacitive application region*.

In the logarithmic coordinates, all the tested points are crowded on a straight line. Consequently, the SRF can be formulized by a simple equation, that is,

$$SRF_L = \frac{8920}{\sqrt{L}}, \quad (7.A.20)$$

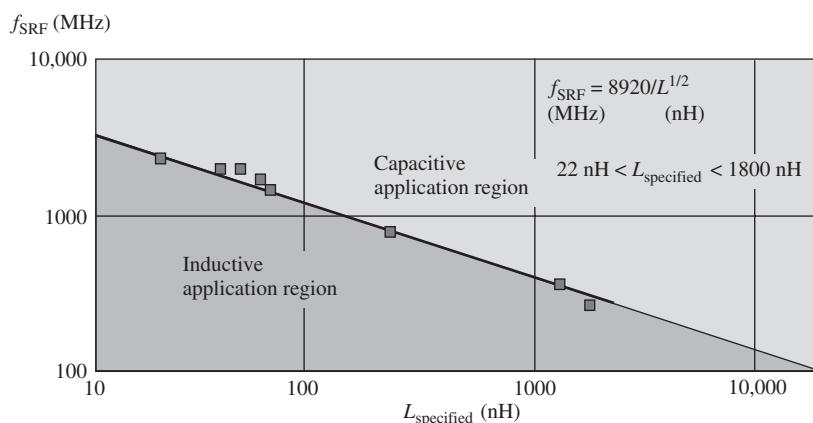
where the unit of  $L = L_{\text{specified}}$  is nH and the unit of  $SRF_L$  is MHz.

Equation (7.A.20) proves extremely convenient to designers in the calculation of the SRF of a chip inductor. However, instead of equation (7.A.20), an experienced engineer should be familiar with the values of the special chip inductors listed in Table 7.A.2.

Figure 7.A.12 shows the variation in the in-parallel equivalent parasitic capacitance  $C_P$  as  $L_{\text{specified}}$  is varied, which is simply calculated from the SRF via equations (7.A.18) and (7.A.19) for MuRata chip inductors.

It is found that the in-parallel parasitic capacitance  $C_P$  is kept almost constant from low-inductance parts to high-inductance parts. Its average value is about 0.2 pF.

Finally, from the reading of  $S_{21,SRF}$ , we can calculate the in-parallel parasitic resistance  $R_P$  through equation (7.A.17). Figure 7.A.13 plots its result. The tested result shows



**Figure 7.A.11.** Plot of the self-resonant frequency  $f_{SRF}$  versus specified value of chip inductor  $L_{\text{specified}}$ , extracted from testing of MuRata chip inductors.

TABLE 7.A.2. Some Important  $SRF_L$  (Self-Resonant Frequency) Values of Chip Inductors

$SRF_L$ , MHz	Value of Chip Inductor, nH
210	1800
300	884
450	393
500	318
800	124
900	98
1000	79.6
1500	35.4
2400	13.8
5400	2.7

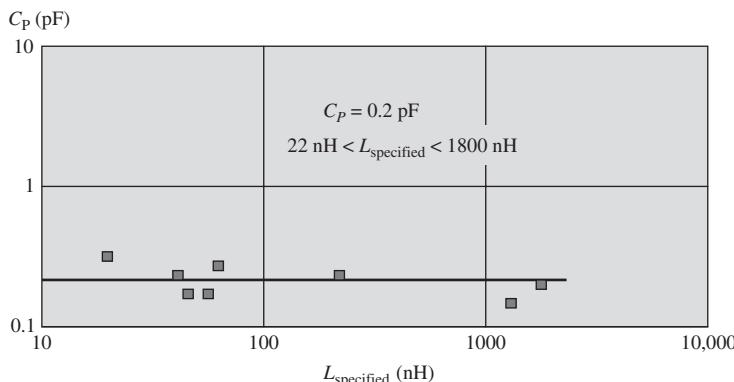


Figure 7.A.12. Plot of in-parallel capacitance  $C_P$  versus specified inductance  $L_{\text{specified}}$ , extracted from testing of MuRata chip inductors.

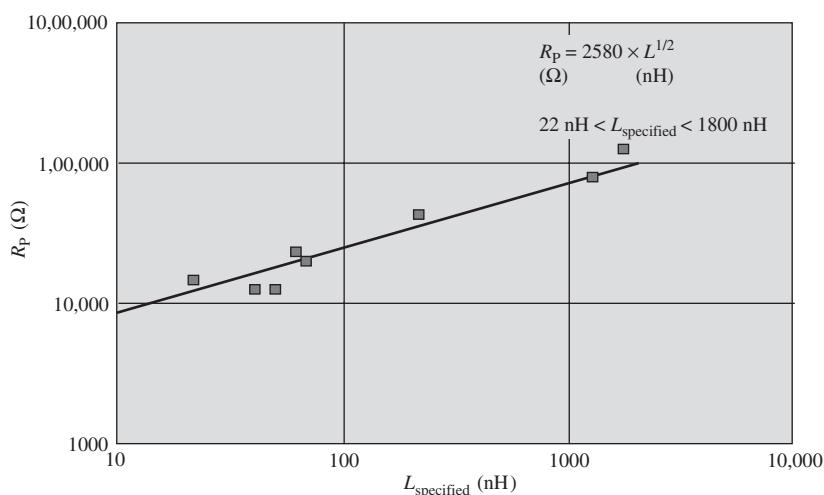


Figure 7.A.13. Plot of in-parallel resistance  $R_P$  versus specified inductance  $L_{\text{specified}}$ , extracted from testing of MuRata chip inductors.

that the in-parallel parasitic resistance  $R_P$  increases as the  $L_{\text{specified}}$  is increased. In the logarithmic coordinates, all tested points are crowded on a straight line. Consequently, the in-parallel parasitic resistance  $R_P$  can be summarized by a simple equation, that is,

$$R_P = \frac{2580}{\sqrt{L_{\text{specified}}}}, \quad (7.A.21)$$

where the unit of  $L_{\text{specified}}$  is nH, and the unit of  $R_P$  is  $\Omega$  (omega).

Equation (7.A.21) provides extreme convenience to the designers in the calculation of the in-parallel parasitic resistance  $R_P$  of a chip inductor.

### 7.A.2 Characterizing of Chip Resistor by Means of $S_{11}$ or $S_{22}$ Testing

The impedance of an ideal resistor is

$$Z_R = R. \quad (7.A.22)$$

Just like other chip parts, ideal resistors are never available, although a real resistor in the low-frequency range is pretty close to ideal and can be described by expression (7.A.22). In the RF range, this ideal formula for the resistor impedance is not viable. Instead, a practical chip resistor in the RF range can be modeled as shown in Figure 7.A.14.

The equivalent consists of the resistor itself,  $R = R_P$ , and an additional capacitor in parallel,  $C_P$ , or an additional inductor in parallel,  $L_P$ , depending on the value of  $R_{\text{specified}}$ , which is specified by the manufacturer. It has been found that in RF frequency range, usually,

$$R = R_P \neq R_{\text{specified}}. \quad (7.A.23)$$

If the resistance value of a chip resistor,  $R$ , is neither too low nor too high, it can be characterized by means of  $S_{11}$  or  $S_{22}$  testing as shown in Figure 7.A.15.

The reading of  $S_{11}$  or  $S_{22}$  corresponds to a complex impedance  $Z_S$ , that is,

$$Z_S = R_S + jX_S = \frac{1 + S_{11}}{1 - S_{11}}. \quad (7.A.24)$$

It should be noted that  $R_S$  and  $X_S$  are in series. In order to obtain  $R_P$  and  $C_P$  or  $L_P$  in parallel, the following conversion must be conducted:

$$Q = \frac{X_S}{R_S}, \quad (7.A.25)$$

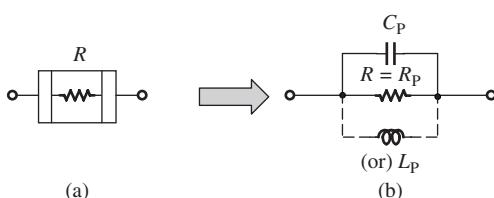
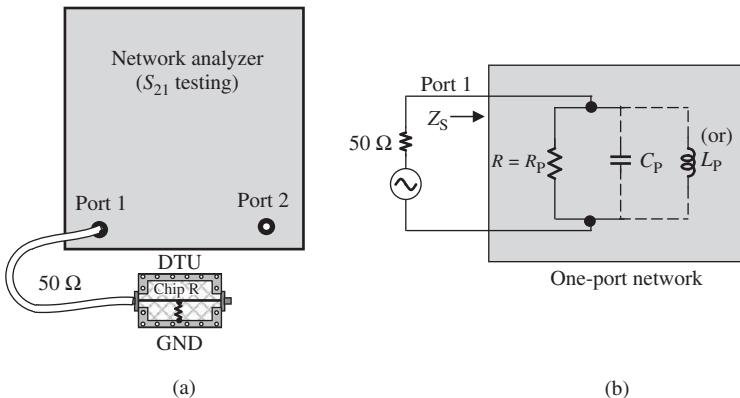


Figure 7.A.14. An actual chip resistor (a) and its equivalent (b).



**Figure 7.A.15.** Single-port  $S$  parameter measurement of a chip resistor (a)  $S_{11}$  testing setup for a chip resistor. (b) Replacement of 'One port Network' by chip resistor's equivalent model.

$$X_P = X_S \left( 1 + \frac{1}{Q_S^2} \right), \quad (7.A.26)$$

$$R_P = R_S (1 + Q_S^2), \quad (7.A.27)$$

$$C_P = -\frac{1}{\omega X_P}, \quad (7.A.28)$$

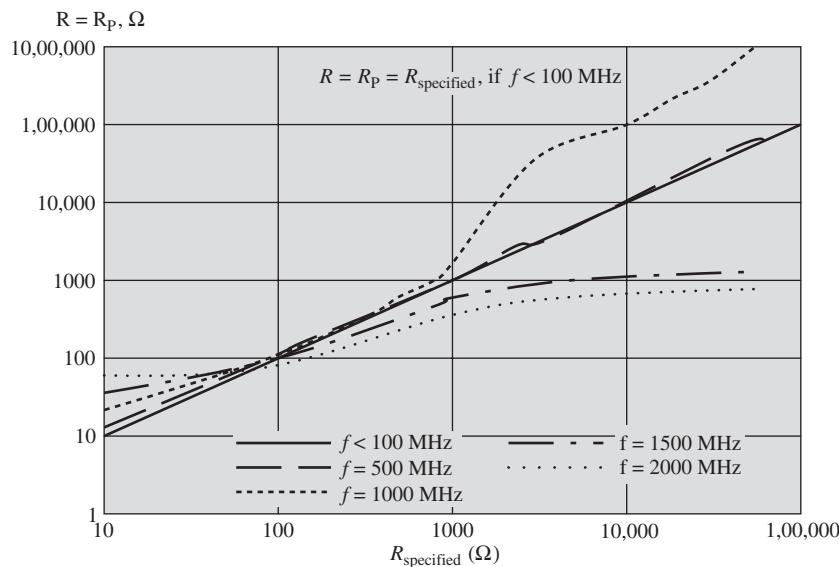
and

$$L_P = -\frac{X_P}{\omega}. \quad (7.A.29)$$

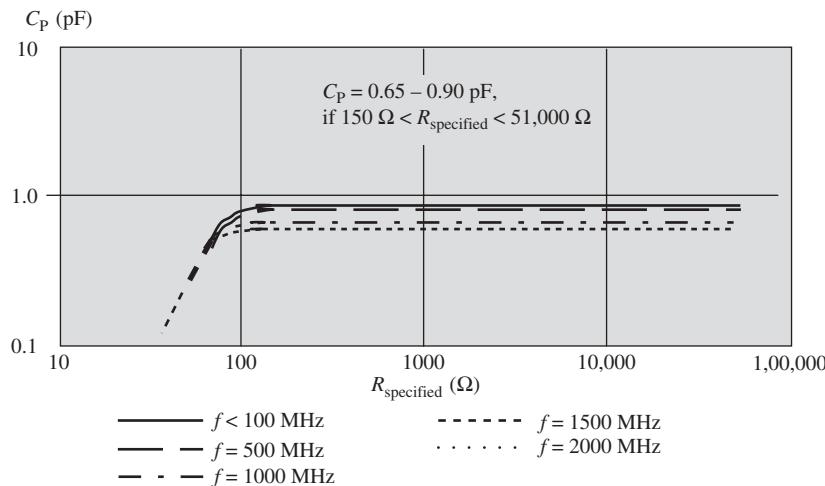
Figure 7.A.16 plots the chip resistor's resistance in parallel,  $R = R_P$ , versus the specified resistance,  $R_{\text{specified}}$ .

Compared with a chip inductor or a chip capacitor, a special feature of a chip resistor is that  $R_P$  is a function of the operating frequency. As shown in Figure 7.A.16,  $R_P$  is almost of the same value as  $R_{\text{specified}}$  when the operating frequency is less than about 100 MHz. As the operating frequency increases,  $R_P$  deviates from the  $R_{\text{specified}}$  more and more. At about 1000 MHz, the deviation approaches a maximum, that is, a chip resistor with  $R_{\text{specified}} = 10 \text{ k}\Omega$  becomes a resistor with  $R_P = 100 \text{ k}\Omega$ . The deviation is about 10 times. When the operating frequency is continuously increased, say, when the operating frequency is greater than about 1500 MHz, the deviation between  $R_P$  and  $R_{\text{specified}}$  depends on the value of  $R_{\text{specified}}$ . If  $R_{\text{specified}} > 80 \Omega$ , the deviation is not continuously increased but decreased. The value of  $R_P$  is lower than that of  $R_{\text{specified}}$ . On the contrary, if  $R_{\text{specified}} < 80 \Omega$ , the deviation between  $R$  and  $R_{\text{specified}}$  is increased so that the value of  $R$  is higher than that of  $R_{\text{specified}}$ . In the vicinity of  $R_{\text{specified}} = 80 \Omega$ , the deviation is negligible and  $R$  is almost independent from the operating frequency. A chip resistor with  $R_{\text{specified}} = 80 \Omega$  seems a good part to be applied in the practical circuit because it is neither sensitive to operating frequency nor deviated from its specified value.

Another special feature of a chip resistor is that its equivalent reactance is either  $C_P$  or  $L_P$ , depending on the value of  $R_{\text{specified}}$ . As shown in Figure 7.A.17, the reactance of the chip resistor is capacitive and the value of  $C_P$  is in the range of 0.65–0.90 pF if  $R_{\text{specified}}$  is greater than 100  $\Omega$ . However, the reactance of the chip resistor is actually inductance if  $R_{\text{specified}}$  is lower than about 6  $\Omega$ . This is quite an extraordinary feature.



**Figure 7.A.16.** Plot of in-parallel resistance  $R = R_P$  versus specified resistance  $R_{\text{specified}}$ , extracted from testing of MuRata chip resistors.



**Figure 7.A.17.** Plot of in-parallel capacitance  $C_P$  versus specified resistance  $R_{\text{specified}}$ , extracted from testing of MuRata chip resistors.

From Figure 7.A.17 it is also found that in the frequency range from 100 to about 2000 MHz, the value of  $C_P$  is not too sensitive to the variation of the operating frequency, although it is conceivable.

## REFERENCE

Carson RS. High-frequency amplifiers. New York: John Wiley & Sons, Inc.; 1975.

## FURTHER READING

- Gray PR, Hurst PJ, Lewis SH, Meyer RG. Analysis and design of analog integrated circuits. 4th ed. New York: John Wiley & Sons, Inc.; 2001.
- Jordan EC. Electromagnetic waves and radiating systems. 2nd ed. Englewood Cliffs (NJ): Prentice-Hall Inc.; 1968.
- Lafferty R. Measuring the self-resonant frequency of capacitors. IEEE Trans Compon Hybrids Manuf Technol [see also IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part A, B, C] 1982;5(4):528–530.
- Li R. Key issues in RF/RFIC circuit design. Beijing: Higher Education Press; 2005.
- Manka W. Alternative methods for determining chip inductor parameters. IEEE Trans Parts Hybrids Pack 1977;13(4):378–385.
- Tenhuunen H. CMOS interconnects. ESDlab (Electronic System Design Laboratory), Department of Electronics, KTH (Kungl Tekniska Hogskolan) Royal Institute of Technology, Kungl Tekniska Hogskolan; Stockholm, Sweden 2000.

## EXERCISES

1. What does it mean to be well grounded?
2. What do the terms zero capacitor and infinite inductor indicate?
3. What is the reason one can select a zero capacitor among the chip capacitors?
4. What is the reason one can select an infinite inductor among the chip inductors?
5. What is the relationship between the capacitance of a zero capacitor and its operating frequency?
6. What is the relationship between the inductance of an infinite inductor and its operating frequency?
7. The actual resistance of a chip resistor depends on the operating frequency. It can be 10 times higher or lower than the specified value. In which resistance range are the actual resistance values insensitive to the operating frequency?
8. A chip resistor has either capacitive or inductive impedance. Specify the ranges of resistance corresponding to capacitive and inductive impedance.
9. Using equations, when a microstrip line with characteristic impedance  $Z_0 = 50 \Omega$  and length  $l = 5 \text{ cm}$ , and the measured values of  $x|_{Z_L \rightarrow 0}$  are 0.577, 1.000, 1.732, respectively, calculate their corresponding quarter wavelengths.
10. How can we create a  $50\text{-}\Omega$  inductor using a microstrip line connected with a load in parallel? How can we create a  $50\text{-}\Omega$  capacitor using a microstrip line connected with a load in parallel?
11. You are looking into one end of a quarter-wavelength microstrip line with  $50\text{-}\Omega$  characteristic impedance. Its another end is short circuited. What impedance do you find?
12. What impedance do you find if
  - (a) the characteristic impedance is  $50 \Omega$ ,
  - (b) one end of the microstrip line is terminated by a  $50\text{-}\Omega$  terminator, and
  - (c) you look at another end of a microstrip line?
13. You are looking into one end of a quarter-wavelength microstrip line with  $50 \Omega$  characteristic impedance. Its other end is open circuited. What impedance do you find?

14. A mixer is built on a PCB. Its RF is 2.4 GHz, and its LO injection frequency is 1.50 GHz. How many zero capacitors are needed, and what are their values? If you are going to apply the infinite inductors to the circuitry, what are their values?
15. Point P is going to be grounded. A runner with  $50\ \Omega$  characteristic impedance and  $1/32$  wavelength is applied to connect point P and another point that is a real ground point. Point P is not grounded indeed. What is the actual impedance existing at point P?

## ANSWERS

1. “Well grounded” means that with respect to a reference ground point,
- (a) the following conditions are satisfied for all the fully grounding points:

$$\begin{aligned} V_{AC} &= 0 & V_{DC} &= 0 \\ Z_{AC} &= 0 & Z_{DC} &= 0. \end{aligned}$$

- (b) The following conditions are satisfied for all the fully grounding points:

$$\begin{aligned} V_{AC} &= 0 & V_{DC} &\neq 0 \\ Z_{AC} &= 0 & Z_{DC} &\neq 0. \end{aligned}$$

2. A zero capacitor is a capacitor whose impedance approaches zero around the operating frequency. The infinite inductor is an inductor whose impedance approaches infinite around the operating frequency.
3. The equivalent circuit of a chip capacitor is an LCR circuit in series. Its impedance approaches zero when the operating frequency is equal to its SRF. Therefore, a chip capacitor can be selected as a zero capacitor as long as its SRF is equal to the operating frequency.
4. The equivalent circuit of a chip inductor is an LCR circuit in parallel. Its impedance approaches infinite when the operating frequency is equal to its SRF. Therefore, a chip inductor can be selected as an infinite inductor as long as its SRF is equal to the operating frequency.
5. The relationship between the capacitance of a ‘zero’ capacitor and its operating frequency is

$$f = SRF_c = \frac{1}{2\pi\sqrt{L_S C}}.$$

6. The relationship between the inductance of an ‘infinite’ inductor and its operating frequency is

$$f = SRF_L = \frac{1}{2\pi\sqrt{LC_P}}.$$

7. In the resistance range around  $70, the actual resistor’s values of a chip resistor are insensitive to the operating frequency.$

8. On the basis of actual testing, the impedance of a chip resistor is inductive if its resistance value is greater than  $7 \Omega$ . The impedance of a chip resistor is capacitive if its resistance value is less than  $5 \Omega$ .
9. When a microstrip line has characteristic impedance  $Z = 50 \Omega$  and length  $l = 5 \text{ cm}$ , and the measured values of  $x|_{Z_L \rightarrow 0}$  are 0.577, 1.000, and 1.732, their corresponding quarter wavelength is 15, 10, and 7.5 cm, respectively.
10. A  $\lambda/8$  microstrip line with a short-circuited end is equivalent to a  $50\text{-}\Omega$  inductor, while a  $\lambda/8$  microstrip line with an open-circuited end is equivalent to a  $50\text{-}\Omega$  capacitor.
11. The impedance is infinite.
12.  $50 \Omega$ .
13.  $0 \Omega$  (short circuited).
14. Three zero capacitors:  $C_1 = 5.1 \text{ pF}$  for RF;  $C_2 = 13 \text{ pF}$  for LO;  $C_3 = 36 \text{ pF}$  for IF.  
Three infinite inductors:  $L_1 = 14 \text{ nH}$  for RF;  $L_2 = 35 \text{ nH}$  for LO;  $L_3 = 98 \text{ nH}$  for IF.
15. Point P is going to be grounded. A runner with  $50 \Omega$  characteristic impedance and  $1/32$  wavelength is applied to connect point P and another point that is a real ground point. Point P is not grounded indeed. Its actual impedance at point P is  $j9.95 \Omega$ .

# EQUIPOTENTIALITY AND CURRENT COUPLING ON THE GROUND SURFACE

## 8.1 EQUIPOTENTIALITY ON THE GROUND SURFACE

### 8.1.1 Equipotentiality on the Grounded Surface of an RF Cable

Let us start our discussion of the equipotentiality on grounded surface of a regular RF cable. Simply speaking, a regular RF cable is a conductive wire encircled by a conductive cylinder (Fig. 8.1). The outer cylinder at the two ends is usually connected as the ground terminal.

In the RF range, the following questions must be answered:

- Are the ground points on the outside cylinder  $G_a$  and  $G_b$  equipotential? Is

$$v_{G_a} = v_{G_b} ? \quad (8.1)$$

- Are the ground points of an RF cable at two ends,  $G_1$  and  $G_2$ , equipotential? Is

$$v_{G_1} = v_{G_2} ? \quad (8.2)$$

In the DC or low-frequency range, these points are, approximately, equipotential.

$$v_{G_a} = v_{G_b} = v_{G_1} = v_{G_2} \quad (8.3)$$

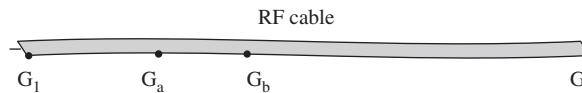


Figure 8.1. Regular RF cable.

In the RF range, they are in most cases not equipotential, that is,

$$v_{G_a} \neq v_{G_b}, \quad (8.4)$$

$$v_{G_1} \neq v_{G_2}, \quad (8.5)$$

unless the distances between a and b, or 1 and 2, are equal to a multiple of the half wavelength.

An RF cable can be simply considered as a waveguide tube. The transport of the RF power or energy is conducted and propagated along the central conductive wire. In most cases, two points on the cable are not equipotential not only on the outside cylinder but also on the central conductive wire unless the distance between these two points is equal to a multiple of the half wavelength.

### 8.1.2 Equipotentiality on the Grounded Surface of a PCB

The same question may be asked of the grounded surface of a PCB (printed circuit board): are the ground points on the grounded surface of a PCB equipotential, that is, as shown in Figure 8.2, is

$$v_A = v_B = v_C = v_D = v_E = v_F? \quad (8.6)$$

The answer is

- Yes, it is correct approximately if the size of the PCB is small, that is, the maximum dimension of PCB is much shorter than the quarter wavelength of the operating frequency.
- No, if the size of PCB is large, that is, the maximum dimension of PCB is comparable to or longer than the quarter wavelength of the operating frequency.

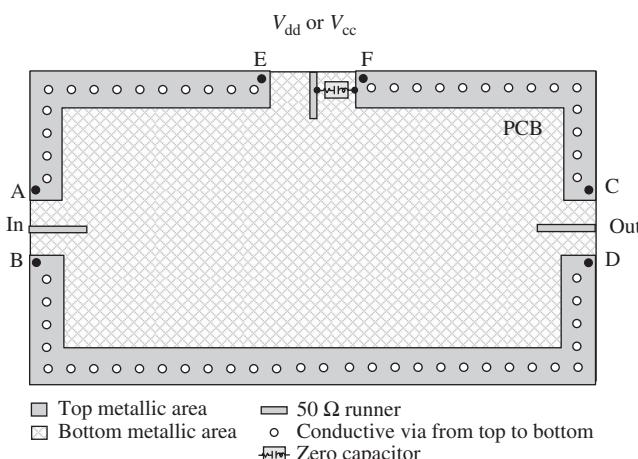


Figure 8.2. Discussion of equipotentiality between points A, B, C, D, E, F (the printed circuit on the PCB is neglected).

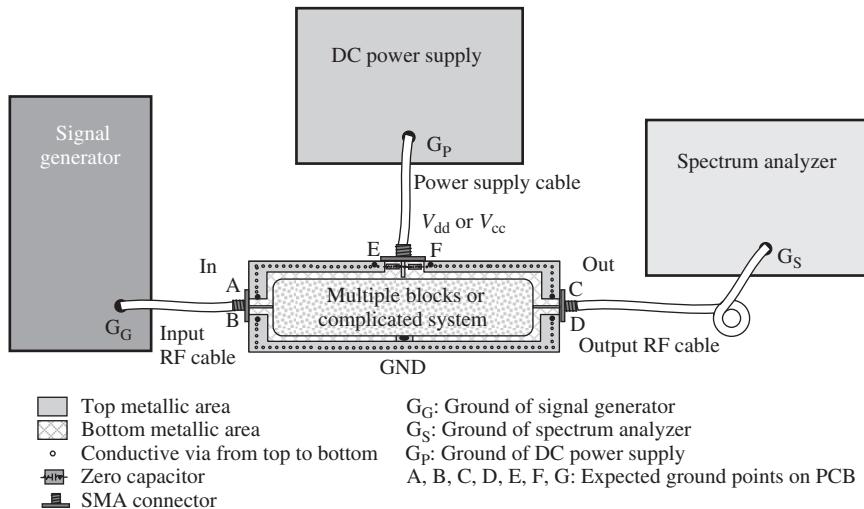


Figure 8.3. Example of test setup with a large PCB.

### 8.1.3 Possible Problems of a Large Test PCB

Let us examine a test setup when the test PCB is large. Figure 8.3 shows a large test PCB connected with a signal generator, a DC power supply, and a spectrum analyzer.

In the large PCB, points A and B are ground points of the input; points C and D are ground points of the output; and points E and F are ground points of the DC power supply. All these ground points are expected to be equipotential so that test results really represent the circuit's performance.

The grounded surface on a small PCB is equipotential approximately so that expression (8.6) is approximately correct. Consequently, testing is done under the same grounded reference point so that the test results do represent the circuit performance.

On the contrary, on a large PCB, the grounded points, as shown in Figure 8.3, are not equipotential, that is,

$$v_A = v_B \neq v_C = v_D \neq v_E = v_F \quad (8.7)$$

although the three pairs of points  $v_A$  and  $v_B$ ,  $v_C$  and  $v_D$ , and  $v_E$  and  $v_F$ , might each be equipotential since the distance between each pair of points is not greater than the size of an SMA (SubMiniature version A) connector, which is much less than a quarter wavelength corresponding to the operating frequency. According to the inequality (8.7), all the circuit branches on the tested PCB are tested under different grounded potentials; the tested results are skewed and do not well represent the circuit's performance.

Let us consider the grounded points connected by RF cables in Figure 8.3. They are the ground point of the signal generator  $G_G$ , the ground point of the DC power supply  $G_P$ , and the ground point of the spectrum analyzer  $G_S$ . As discussed in Section 8.1.1, the two ends of an RF cable are usually not equipotential, that is,

$$v_{G_G} \neq v_A \text{ (or } v_B\text{)}, \quad (8.8)$$

$$v_{G_S} \neq v_C \text{ (or } v_D\text{)}, \quad (8.9)$$

$$v_{G_P} \neq v_E \text{ (or } v_F\text{)}. \quad (8.10)$$

Also, usually

$$v_{G_G} \neq v_{G_S} \neq v_{G_D} \quad (8.11)$$

because  $v_{G_G}$ ,  $v_{G_S}$ , and  $v_{G_D}$  are the ground points of individual equipment so that they are not necessarily equal. Consequently, the expected ground points A, B, C, D, E, F, and G on the PCB are usually not equipotential through the connections between the PCB and the equipment by means of three cables.

On the other hand, it can be seen that the expected ground points A, B, C, D, E, F, and G on the PCB are connected together by the rectangular metallic frame. Does it ensure that these ground points are equipotential? In most cases, it does not because this is a large PCB. Two points on ground surface of a large PCB are usually not equipotential.

Consequently, if expressions (8.7) and (8.11) hold true, the unequipotentiality will make the testing of the circuitry on the PCB meaningless.

In order to make the testing meaningful, the design task of a test PCB is to force the expected ground points on PCB to be equipotential, that is,

$$v_A = v_C = v_E. \quad (8.12)$$

The equipotentiality condition (8.12) ensures that testing is conducted under “common grounding” conditions for the input, output, and other terminals so that correct test results can be expected although expressions (8.8) to (8.11) are still true. In RF circuit testing, conditions (8.8) to (8.11) are allowed as long as the power can be well transported by a cable from one end to the other.

### 8.1.4 Coercing Grounding

There are many methods to force the expected ground points on the PCB so as to achieve equipotentiality. The following three ways to force the ground surface to equipotentiality are recommended in AC/RF grounding for a large PCB.

#### 8.1.4.1 To Force Ground Surface Equipotential by “Zero” Chip Capacitors.

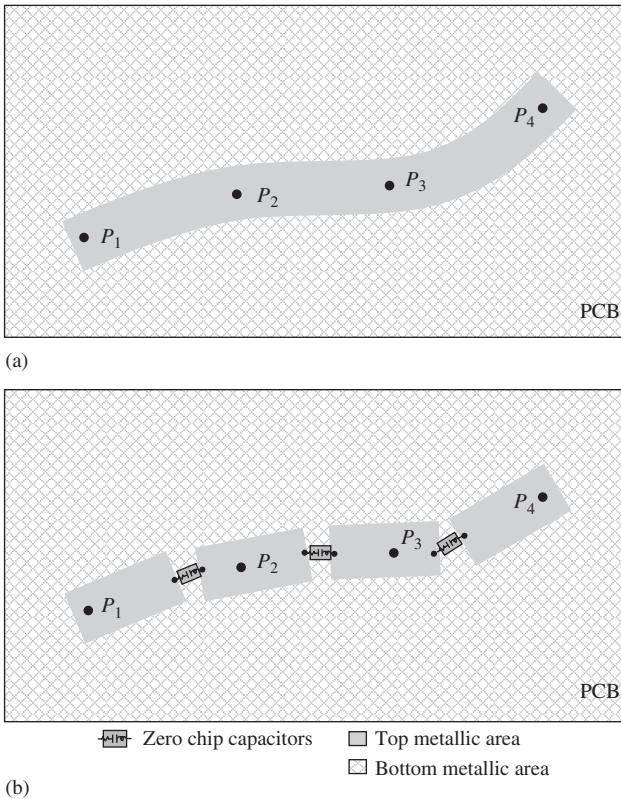
**Capacitors.** When the dimensions of the test PCB are comparable to or larger than the quarter wavelength at the operating frequency, a long runner or a large ground surface is usually in an unequipotentiality status.

Figure 8.4(a) shows a long runner on a large PCB. Usually, the AC/RF voltages on points  $P_1$ ,  $P_2$ ,  $P_3$ , and  $P_4$  are not equal, that is,

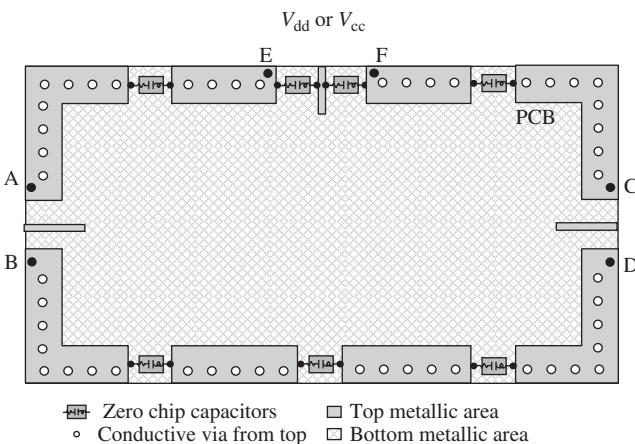
$$v_{P_1} \neq v_{P_2} \neq v_{P_3} \neq v_{P_4}. \quad (8.13)$$

However, as shown in Figure 8.4(b), if this long runner is cut into four or more small pieces such that the length of each small line segment is much less than quarter wavelength of the operating frequency and then they are connected by the zero capacitors, the AC/RF voltages at points  $P_1$ ,  $P_2$ ,  $P_3$ , and  $P_4$ , become equal approximately, that is,

$$v_{P_1} \approx v_{P_2} \approx v_{P_3} \approx v_{P_4}. \quad (8.14)$$



**Figure 8.4.** Unequipotentiality of a long runner changed to equipotentiality by means of zero chip capacitors. (a) Unequipotentiality on a long runner,  $v_{P_1} \neq v_{P_4}$ . (b) Equipotentiality on a long runner resumed by zero capacitors,  $v_{P_1} = v_{P_4}$ .



**Figure 8.5.** Unequipotentiality of a large ground surface changed to equipotentiality by means of zero chip capacitors (the printed circuit on the PCB is neglected).

Figure 8.5 shows a large PCB. Usually, the AC/RF voltages on points A, C, and E are unequal though those on individual pairs of points A and B, C and D, and E and F are equal, because they are quite close to each other, that is,

$$v_A \approx v_B \neq v_C \approx v_D \neq v_E \approx v_F. \quad (8.15)$$

However, as shown in Figure 8.5, if this large ground surface is cut into small pieces such that the maximum dimension of each small ground area is much less than quarter

wavelength of the operating frequency and then they are connected by the zero capacitors, the AC/RF voltage at points A, B, C, D, E, and F become equal approximately, that is,

$$v_A \approx v_B \approx v_C \approx v_D \approx v_E \approx v_F. \quad (8.16)$$

It should be noted that if the operating frequency covers a wide band or consists of more than one frequency band, a single zero capacitor between two small pieces of metal in Figures 8.4 and 8.5 is insufficient and must be replaced by multiple zero capacitors.

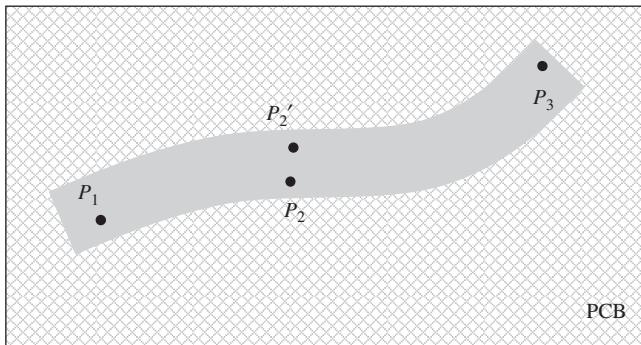
This technique is easier and more direct than the others to be introduced below; however, many zero capacitors must be used.

#### 8.1.4.2 To Force the Ground Surface Equipotential by Half Wavelength of Microstrip line.

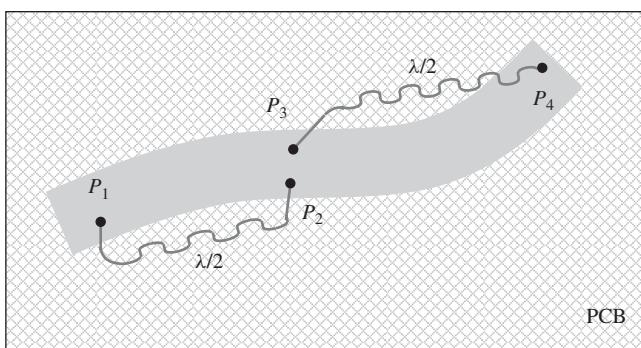
Instead of applying zero capacitors to force the ground surface into an equipotential state as shown in Figures 8.4 and 8.5, half-wavelength runners as shown in Figure 8.6 can be used because the voltage is the same at both ends of a half-wavelength runner.

Figure 8.6(a) shows a long runner on a PCB. Usually, the AC/RF voltage at points  $P_1$ ,  $P_2$  or  $P_3$ , and  $P_4$  are unequal, that is,

$$v_{P_1} \neq v_{P_2} \approx v_{P_3} \neq v_{P_4}, \quad (8.17)$$



(a)



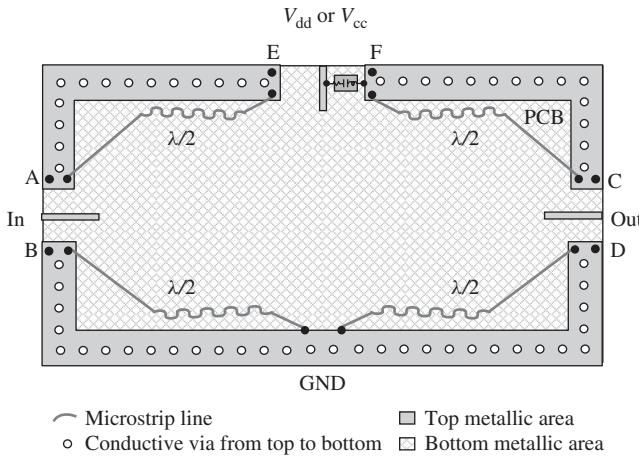
(b)

**Figure 8.6.** Unequipotentiality of a large ground surface changed to equipotentiality by means of a  $\lambda/2$  microstrip line.

(a) Unequipotentiality on a long runner,

$v_{P_1} \neq v_{P_2} \neq v_{P_3} \neq v_{P_4}$ . (b) Equipotentiality on a long runner resumed by  $\lambda/2$  microstrip line,

$$v_{P_1} \approx v_{P_2} \approx v'_{P_2} \approx v_{P_3}.$$



**Figure 8.7.** Unequipotentiality of a large ground surface changed to equipotentiality by means of  $\lambda/2$  microstrip line.

where  $v_{P_2}$  is approximately close to  $v_{P_3}$  because the points  $P_2$  and  $P_3$  are quite close to each other.

Now, as shown in Figure 8.6(b), if two microstrip lines with half wavelength are connected between  $P_1$  and  $P_2$  and between  $P_3$  and  $P_4$  on this long runner, the AC/RF voltage at points  $P_1$ ,  $P_2$ ,  $P_3$ , and  $P_4$  become equal approximately, that is,

$$v_{P_1} = v_{P_2} \approx v_{P_3} = v_{P_4}. \quad (8.18)$$

Figure 8.7 shows a large PCB. Usually, the AC/RF voltages at points A, C, and E are unequal though the AC/RF voltage on individual pair of points A and B, C and D, and E and F are equal, respectively, because they are quite close to each other, that is,

$$v_A \approx v_B \neq v_C \approx v_D \neq v_E \approx v_F. \quad (8.19)$$

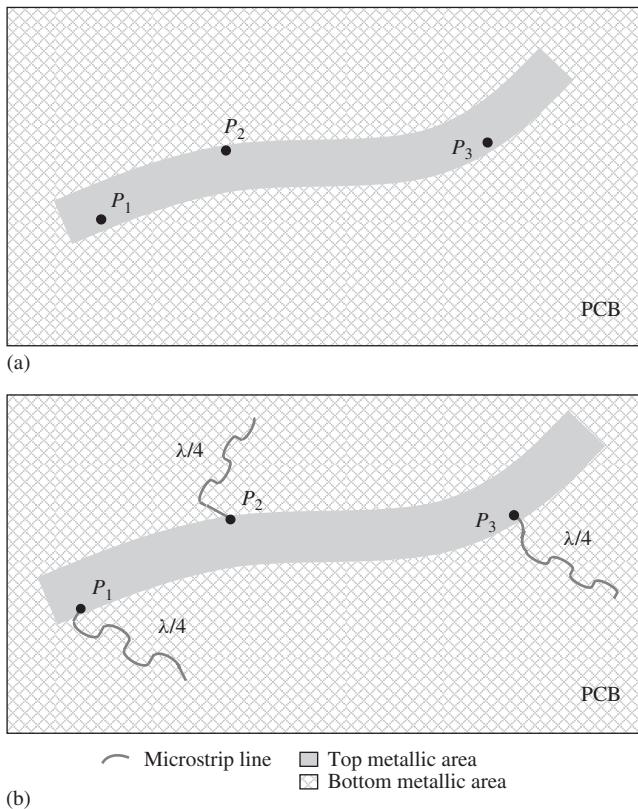
However, as shown in Figure 8.7, if four microstrip lines with half wavelength are connected between A and E, F and C, B and GND, and GND and D on this large ground surface, then the AC/RF voltages at points A, B, C, D, E, and F become equal approximately, that is,

$$v_A \approx v_B \approx v_C \approx v_D \approx v_E \approx v_F. \quad (8.20)$$

**8.1.4.3 To Force the Ground Surface Equipotential by Quarter Wavelength of Microstrip Line.** Instead of applying half-wavelength runners as shown in Figure 8.6, the quarter-wavelength runners as shown in Figure 8.8 can be used to force a long runner into an equipotential state because the voltage must be zero at one end if the other end of the microstrip line with quarter wavelength is open-circuited.

Figure 8.8(a) shows a long runner on a PCB. Usually, the AC/RF voltages at points  $P_1$ ,  $P_2$ , and  $P_3$  are unequal, that is,

$$v_{P_1} \neq v_{P_2} \neq v_{P_3}, \quad (8.21)$$



**Figure 8.8.** Unequipotentiality of a large grounded surface changed to equipotentiality by means of a  $\lambda/4$  microstrip line.

(a) Unequipotentiality on a long runner,  $v_{P_1} \neq v_{P_4}$ . (b) Equipotentiality on a long runner resumed by  $\lambda/4$  microstrip line,  $v_{P_1} = v_{P_4}$ :

However, if at points  $P_1$ ,  $P_2$ , and  $P_3$  a microstrip line with quarter wavelength is connected as shown in Figure 8.8(b), then the AC/RF voltages at points  $P_1$ ,  $P_2$ , and  $P_3$  become equal, that is,

$$v_{P_1} = v_{P_2} = v_{P_3} = 0. \quad (8.22)$$

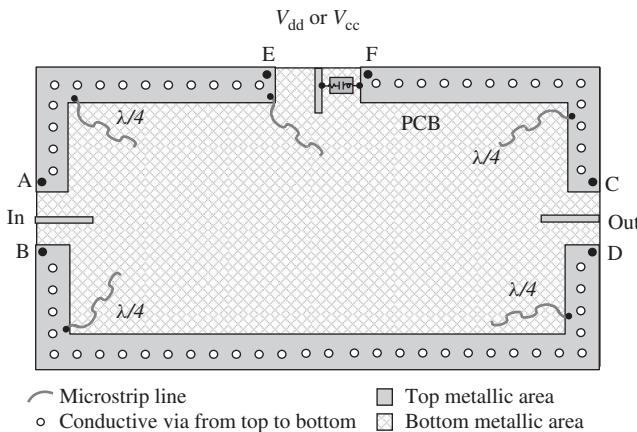
Figure 8.9 shows a large PCB. Usually, the AC/RF voltages on points A, C, and E are unequal though the AC/RF voltage on individual pair of points A and B, C and D, E and F are equipotential, respectively, because they are quite close to each other, that is,

$$v_A \approx v_B \neq v_C \approx v_D \neq v_E \approx v_F. \quad (8.23)$$

However, as shown in Figure 8.9, if five or more microstrip lines with quarter wavelength are connected to the PCB on the ground surface near the DC power supply, four corners and other points, then the AC/RF voltage at points A, B, C, D, E and F become equipotential approximately, that is,

$$v_A \approx v_B \approx v_C \approx v_D \approx v_E \approx v_F. \quad (8.24)$$

A half-wavelength runner is somewhat lengthy. The length of a quarter-wavelength microstrip line is 50% reduced down from that of a half wavelength so that it is a very effective part in the RF circuit grounding.



**Figure 8.9.** Unequipotentiality of a large ground surface changed to equipotentiality by means of  $\lambda/4$  microstrip line.

Besides microstrip line, the half/quarter-wavelength cables could be applied in the same manner as half/quarter-wavelength runners; however, the cables are too clumsy and thus are seldom applied in practical engineering.

### 8.1.5 Testing for Equipotentiality

We have talked a lot about equipotentiality on the ground surface. It is desirable to know how to test the equipotentiality on a large test PCB. Unfortunately, there is no formal equipment available for this task to date.

Figure 8.10 shows a tentative setup for the testing the equipotentiality of a large test PCB. The basic idea of the setup is to discover unequipotentiality on the ground surface by testing  $S_{11}$  or  $S_{22}$  by a network analyzer.

First, three RF cables must be prepared for the calibration. The first cable is termed the *calibrated cable* as shown in Figure 8.10. One end of this cable is connected to port 1 or port 2 on the network analyzer. The other end is denoted by the symbol C and must be calibrated in a “pick-tail” manner.

Calibration by the pick-tail is a special method in which the standard calibration kits’ “short,” “open,” “ $50 \Omega$ ,” and “through” are replaced by a set of cables (called *tails*) with one end in short, open,  $50 \Omega$ , and through and the other end connected to the network analyzer. The length of these cables is equal to the desired length of the first cable, the calibrated cable. A set of calibrated cables is shown in Figure 8.11.

The second cable is termed the *extended cable*, as shown in Figure 8.10. Its length must be a multiple of the half wavelength of the operating frequency, which can be determined by a network analyzer through careful measurement. This cable must be long enough so that the probe at its end can reach anywhere on the PCB. The expected minimum length is the half wavelength, of course.

The purpose of the second cable is just to extend the first cable to the third cable. In the testing of equipotentiality, the third cable, not the first and second cables, is actually connected in the test setup.

The third cable has a length equal to the sum of the lengths of the first and second cables. It is the actual cable used for testing as shown in Figure 8.10. One end is connected to the network analyzer and the other end is connected to a probe. Its intermediate point C is soldered to the reference ground point at the DC power supply terminal. The length between point C and the end connected to the network analyzer must be equal to the

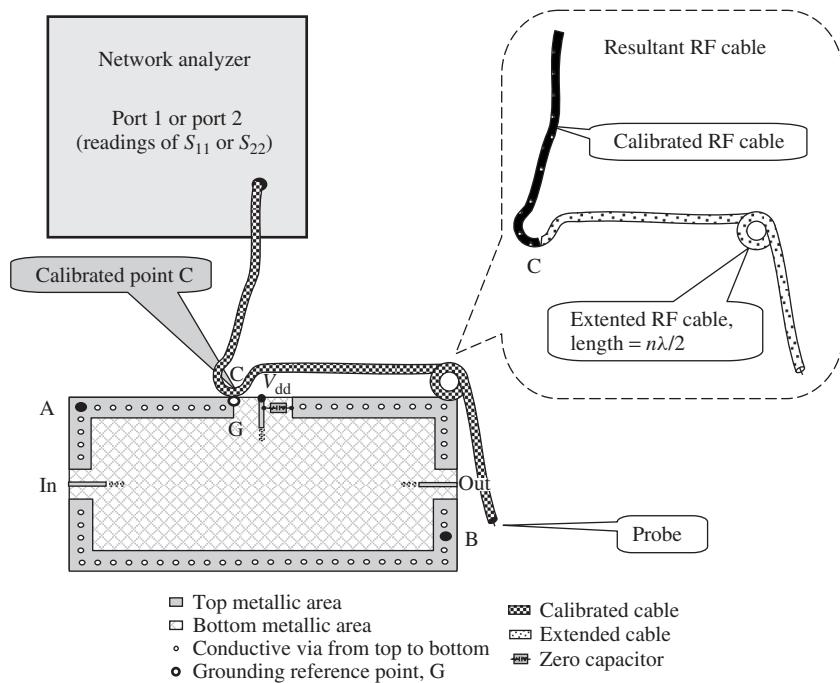
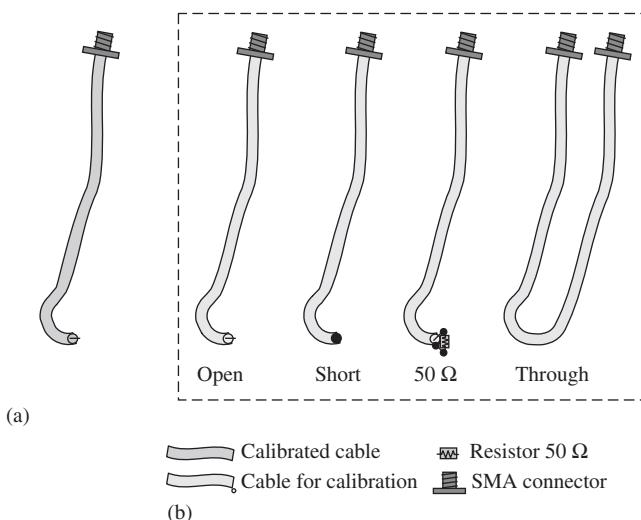


Figure 8.10. Testing of equipotentiality of a PCB.

Figure 8.11. Calibrated cable and calibration kits for testing the equipotentiality of a PCB.  
(a) Calibrated cable. (b) Calibration kit.

length of the first cable. Owing to the fact that point C has been calibrated and that the length between point C and the probe is a multiple of the half wavelength, the potential that the probe senses is equal to the potential at point C.

After calibration is completed and the third cable is fixed as shown in Figure 8.10, testing can begin. By moving the probe around on the PCB,  $S_{11}$  or  $S_{22}$  at different points on the PCB can be compared. If the ground surface and ground points on PCB are equipotential, the readings of  $S_{11}$  or  $S_{22}$  should remain unchanged as the probe moves

around on the PCB. On the other hand, if the readings of  $S_{11}$  or  $S_{22}$  fluctuate when the probe moves from point A to point B, it is an indication that the ground point A is not equipotential with point B. The more the variation of the readings, the more the unequipotentiality between the points.

It should be noted that the testing is sensitive to the presence of the human body when the third cable is held by hand. It is, therefore, suggested that a protective isolation glove be used on the cable where it is held by hand; it is even better to have a static protective metal ring on the wrist or arm of the tester.

This testing scheme has been developed by the author in recent years. It is only tentative and not yet satisfactory because it is qualitative, not quantitative.

## 8.2 FORWARD AND RETURN CURRENT COUPLING

### 8.2.1 Indifferent Assumption and Great Ignore

In Chapter 7, it was pointed out that the typical schematic of a circuitry is always drawn with many “indifferent assumptions” which can cause possible AC/RF grounding problems. People allow these oversights because their attentions are focused on the circuitry itself or because they think that these indifferent assumptions can be just temporarily ignored and easily realized by the circuit designer.

Experienced engineers, however, say: “These are not indifferent assumptions, but great ignorances!”

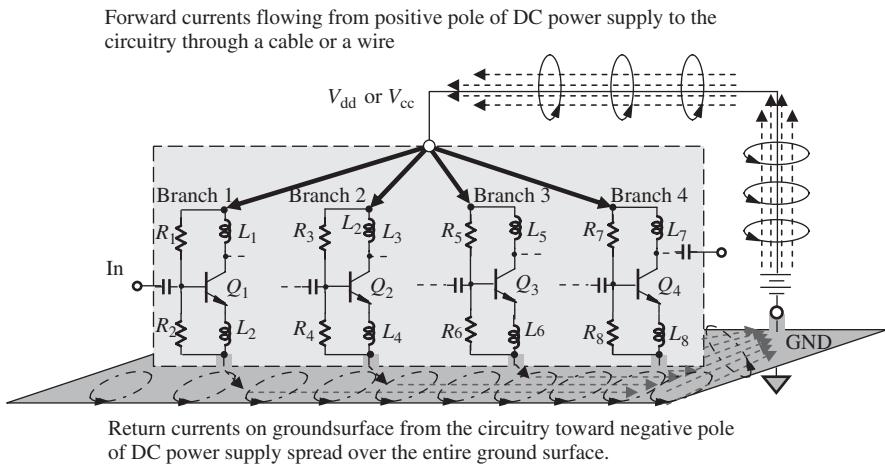
The first two problems about full and half ground points have been described in Chapter 7. Now let us discuss another problem about forward current and return current coupling.

When the half grounding points in a circuit block are not equipotential due to imperfect or inappropriate RF/AC grounding, the forward currents will be coupled magnetically on the path from the positive terminal of the DC power supply to every branch of the circuitry in a cable or a common runner. If each DC power supply node for each branch is made equipotential by means of zero capacitors, the forward current coupling would disappear. Similarly, when the full grounding points in a circuit block are not equipotential due to imperfect or inappropriate RF/AC grounding, the return currents will be coupled magnetically on the path from every branch to the negative terminal of the DC power supply. Usually, the return current has complicated paths because it flows over the entire grounded surface in a circuit block. If the entire grounded surface is equipotential, the return current coupling would disappear.

In short, there will be no current coupling problems if the grounded surface is perfectly equipotential. In reality, unequipotentiality exists more or less on the ground surface. Therefore, reduction or removal of forward and return current coupling becomes an important subject in RF circuit design.

Figure 8.12 creates images of the forward and return currents which resemble more to the actual performance. The forward current flows from the positive pole of the DC power supply to the circuitry through a cable or a wire. The return current flows from the circuitry toward the negative pole or ground of the DC power supply spread over the entire ground surface.

From Figure 8.12 it can be seen that the four forward subcurrents flowing into the four branches of circuitry are coupled together in a cable or a wire, while the four return subcurrents returning to the DC power supply are spread and coupled over the entire ground surface. In reality, the return currents flowing on the PCB form a complicated



**Figure 8.12.** Images of forward and return currents between the circuitry and the DC power supply.

pattern, depending on the placement of parts and the arrangement of the grounding areas.

The current coupling between adjacent branches is a kind of magnetic crosstalk. Signals from different branches coupling with each other in either the forward current flow path or the return current flow path are equivalent to the addition of a transformer into the circuitry. The transformer has complicated inputs and outputs from the four branches. It is particularly hard to figure out the return current coupling pattern on the ground surface. As a result, the circuit performance could become ridiculously wrong if the ground surface is unequipotential. This is the problem of forward and return current coupling.

### 8.2.2 Reduction of Current Coupling on a PCB

Figure 8.13 shows an example of forward current coupling on a PCB. For simplicity, the actual layout is replaced by the corresponding schematic in this section. On the PCB, it can be seen that there are three branches or three stages of this RF block. A DC power supply  $V_{dd}$  is provided to point P. Obviously, the currents for three branches  $I_1, I_2$ , and  $I_3$  are actually magnetically coupled along the runner PQ even though there are two zero capacitors connected from point P to the ground points E and F. This is called *forward current coupling*. The equivalent of this current coupling is a common transformer between these three branches. The equivalent transformer mainly depends on the length of the runner PQ. The length of PQ would be extended to include the length of the cable that connects the DC power supply to point P if there are no zero capacitors connected from the point P to the ground.

In the low-frequency range, the equivalent transformer plays a only tiny role in the circuitry and is negligible. At RF, however, the equivalent transformer becomes important and is sometimes the main source of trouble screwing up circuit function. Unfortunately, to new designers, this source of trouble is hidden in a dark corner.

A very straightforward method of reducing forward current coupling is to provide a DC power supply to each circuit branch separately. This method is also called “to provide a DC power supply to each circuit branch or block in parallel.” Figure 8.14 illustrates such a scheme. The DC currents for each branch,  $I_1, I_2$ , and  $I_3$ , are provided

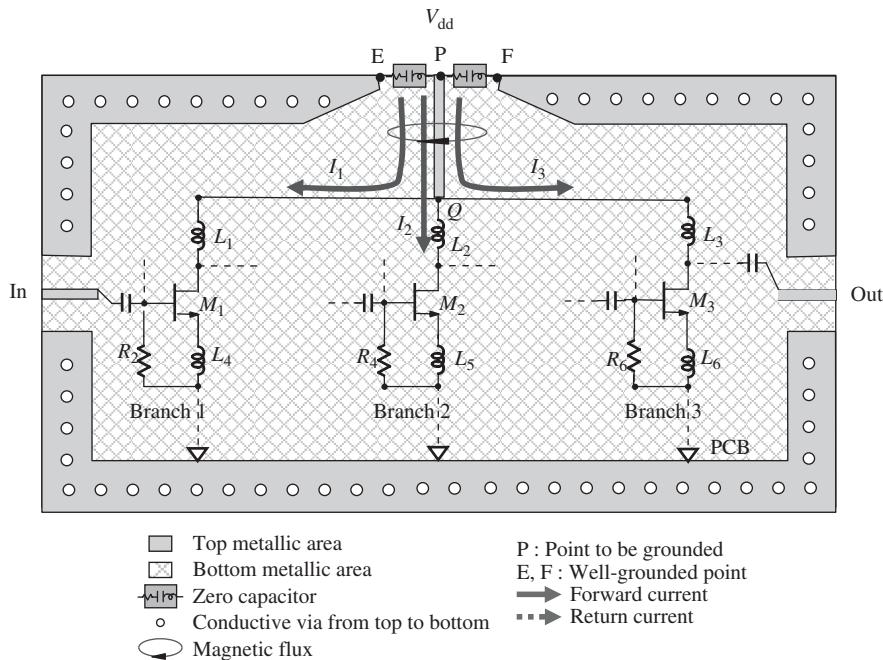


Figure 8.13. Forward current coupling due to the common runner PQ after zero capacitors.

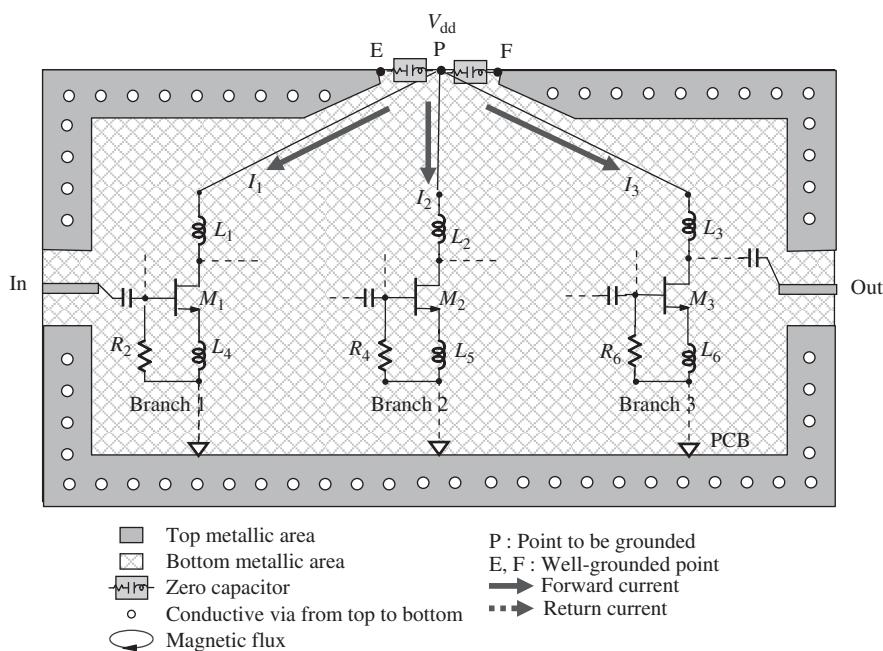


Figure 8.14. Reduction of forward current coupling by providing a DC power supply to each branch.

separately from point P to each branch without a common runner. Magnetic coupling still exists between the three paths for  $I_1$ ,  $I_2$ , and  $I_3$ , but the coupling is greatly reduced to a minuscule level.

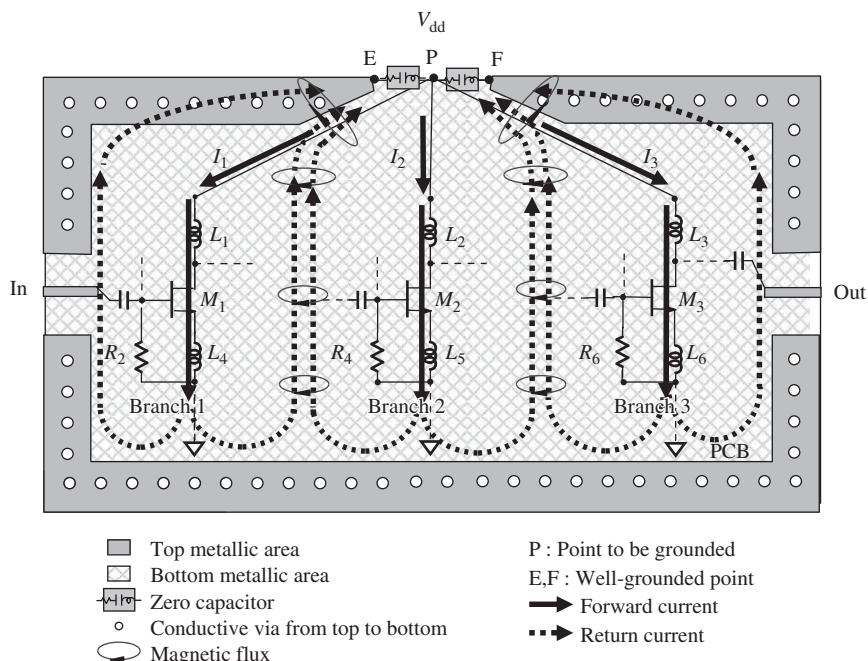
Now, let us consider the return current coupling. Unlike the forward current flowing along a runner, the return current actually flows over the entire grounded surface and then returns to the area around reference ground points E or F of the battery or the DC power supply.

Figure 8.15 roughly depicts both forward and return currents. The solid bold arrow represents the direction of the forward current, and the dashed bold arrow represents the direction of the return current. It can be seen that the return currents are strongly coupled in the area between the two branches: this is equal to having a transformer between the two branches in the circuitry. All equivalent transformers would seriously disturb the performance of the circuitry. If a designer ignores return current coupling, he might never know what is going on when his RF block does not work well or simply refuses to work altogether.

In order to reduce return current coupling, as shown in Figure 8.16, two slots are cut on the bottom metal plane so that the ground areas of adjacent branches of the circuitry are isolated from each other. Each return current from the individual branches returns along its own ground area to the point E or F. Consequently, the return current coupling is significantly reduced though it is still exists around points E, F, or P.

### 8.2.3 Reduction of Current Coupling in an IC Die

If a single RF block is built in an IC die, the scheme discussed above can be adapted; however, some modification is necessary. Figure 8.17 shows a single RF block built on an IC die for the reduction of forward and return current coupling.



**Figure 8.15.** Return current coupling due to current coupling on the ground surface.

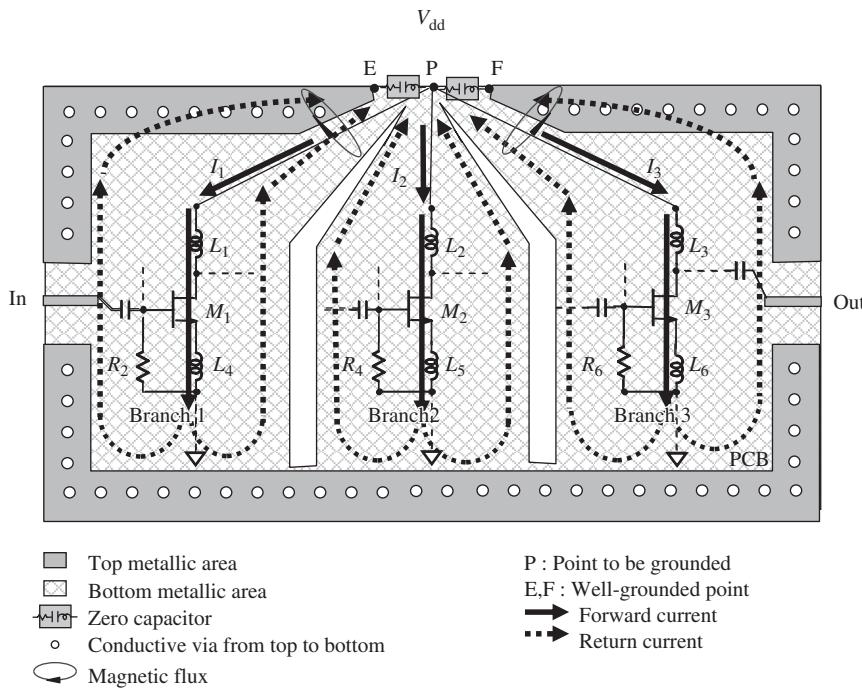


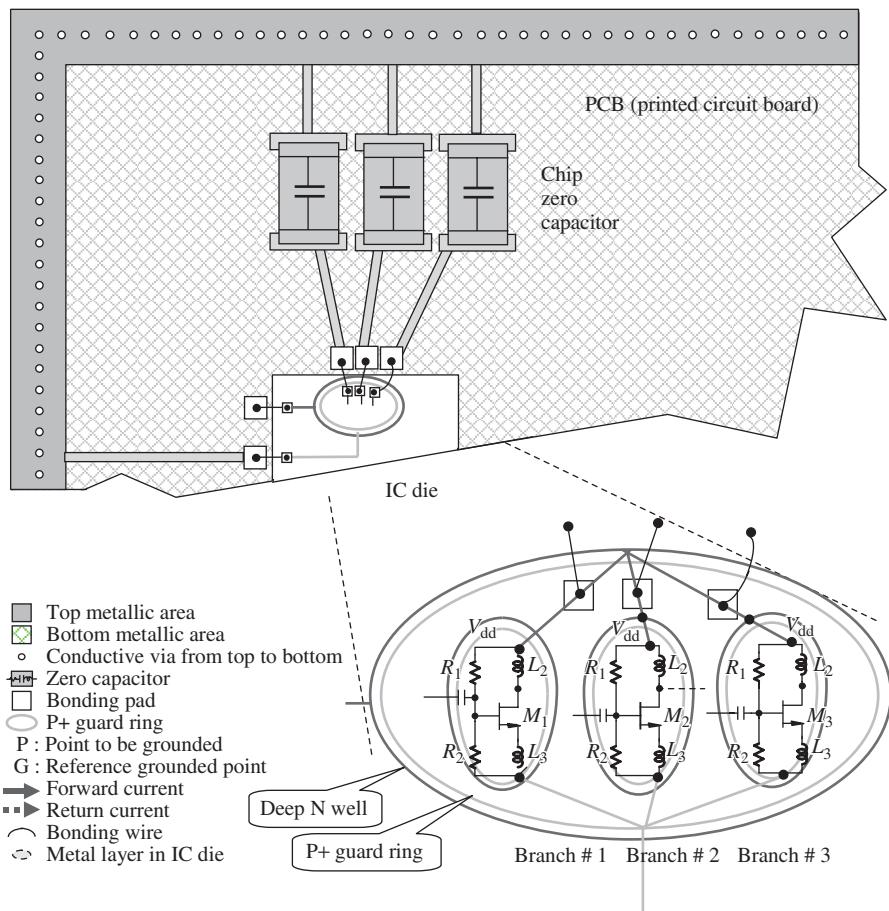
Figure 8.16. Reduction of return current coupling by means of slot-cutting.

In order to reduce forward current coupling in an RFIC die, the same scheme shown in Figure 8.16 can be applied to the RFIC die containing an RF block. The three branches are provided with separate DC power supplies to greatly reduce forward current coupling. However, three separate zero capacitors are chip capacitors and must be allocated outside the IC die as shown in Figure 8.17 so that three bonding wires and six pads are needed for the chip capacitors outside the IC die. For a simple RFIC block, this may not be a problem; however, if an IC die contains multiple RF blocks, the number of bonding wires and pads quickly becomes unacceptable. We are therefore aware of the importance and need to develop a zero capacitor directly in the IC die in the coming years.

In order to reduce return current coupling in an RFIC die, the ground metal in the IC die is divided into three pieces so that the three branches of circuitry have their individual ground areas. The individual ground areas are connected to the P+ guard ring, which is the ground ring encircling the RF block. The P+ guard ring then is connected to a pad and extruded out of the die by a bonding wire so as to connect to the metal frame on the PCB, which is equipotential with the reference ground point. The return currents from the three branches of circuitry are therefore isolated from each other, and the return current coupling is significantly reduced.

In Figure 8.17, the metal layer in the IC die is etched in a pattern with three oval areas. This beautiful and neat pattern is drawn just for illustration of the principle. In reality, the areas can take on any shape as long as the return current in the three branches of circuitry can be isolated. Instead of an oval shape, a pattern with etched long slots in the IC die appears more often in engineering design.

Finally, it should be noted that when circuitry is discussed, much attention is paid to the forward current which is drawn from the DC power supply. Very often, the return



**Figure 8.17.** Reduction of forward and return current coupling simultaneously.

currents, which originate from all the grounded points of the circuit to the grounded point of the DC power supply, are ignored. In an actual RF layout or a test board, return current coupling could completely screw up the performance of a circuit due to the coupling between branches through the return current. This problem is especially a dark corner for a new engineer who has little experience in layout work.

#### 8.2.4 Reduction of Current Coupling between Multiple RF Blocks

In a single RF block, the basic principles for reducing forward and return current coupling can be outlined as follows:

- providing DC power supply to branches of circuitry separately or in parallel;
- inserting long slots in the ground area so that each branch has its own ground area. All individual ground areas then are connected in parallel to the common ground surface, which is equipotential with the reference ground point at the DC power supply terminal.

The same principles can be applied to reduce forward and return current between multiple RF blocks. In other words, the above principles can be simply copied, replacing the words

- “Branch” by “Block,” and
- “Block” by “Multiple blocks.”

That is, in cases with multiple RF blocks, the basic principles for reducing forward and return current coupling can be outlined as follows:

- providing DC power supply to blocks of circuitry separately or in parallel;
- inserting long slots in the ground area so that each block has its individual ground area. All individual ground areas then are connected in parallel to the common ground surface, which is equipotential with the reference ground point at the power supply terminal.

In practical engineering designs, flexibility always exists for the separation of branches or blocks. In other words, a few branches could be combined as a big branch or a few blocks combined as a big block in the execution of the design principles mentioned above. These choices depend on the complexity of the circuitry, the current drain, and so on. As long as forward and return current coupling do not seriously impact the performance, the designer may try whatever works better to reach the goals.

### 8.2.5 A Plausible System Assembly

The “multiclosets” assembly type is a popular type of system assembly appearing in electronic products. Figure 8.18 shows its configuration, which can be described as follows:

- It is basically a metallic box with many closets. Each closet contains one block.
- There are 12 closets shown in Figure 8.18. The raw material of the metallic box is usually aluminum; in order to enhance its shielding function, the entire surface of the metallic box is gold-plated. As shown in Figure 8.18, a gold-plated aluminum

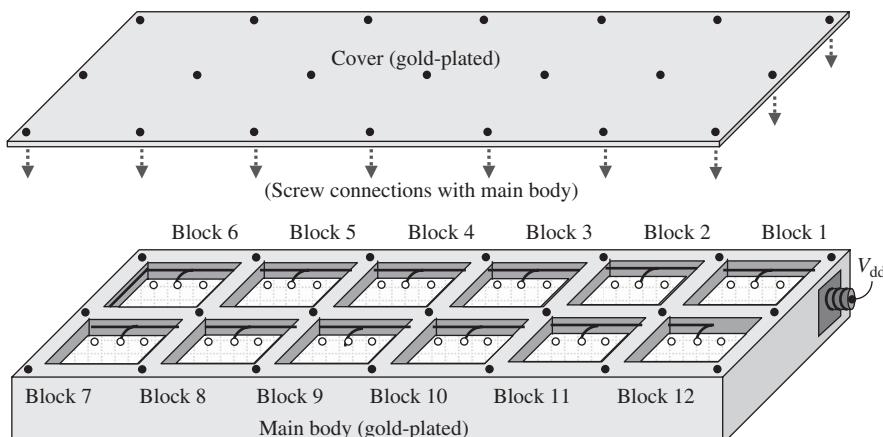


Figure 8.18. Configuration of a multicloset type of system assembly.

cover is connected to the main body of the metallic box by screws. (In Figure 8.18, the screw holes are marked by dark circles.) Consequently, the isolation between blocks or closets is, expectedly, very good.

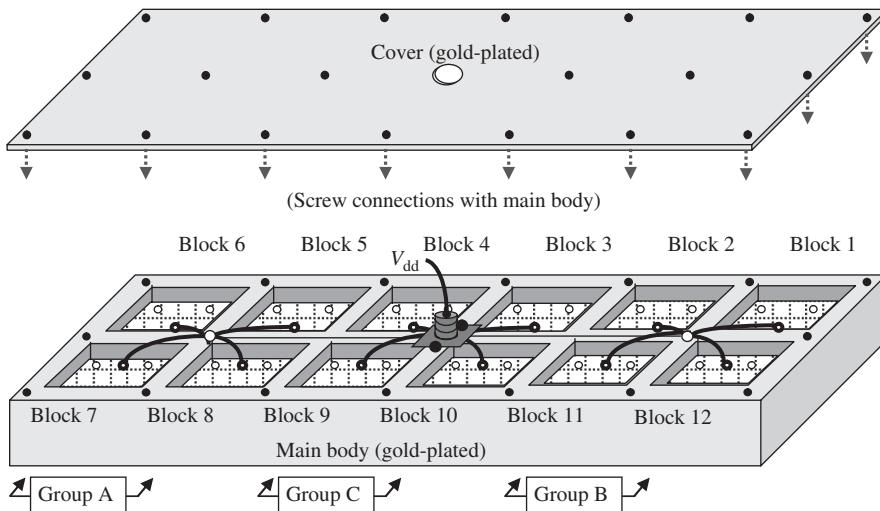
- A block in a closet consists of many electronic parts, including resistors, capacitors, inductors, IC dies or packages, and so on. All the electronic parts are assembled on a PCB or some other type of board.
- On the assembled PCB, a terminal marked with a bold black circle is connected to the DC power supply  $V_{dd}$ . The DC power supply goes from the side-connector to the terminals in block 1, block 2, block 3, ..., block 12 in numerical order. In Figure 8.18, the path of the DC power supply is shown with solid bold runners. Each individual block is fed by the DC power supply  $V_{dd}$  in series from the common connector.
- The ground surface on each PCB contains two holes marked with two white circles, which are electrically connected to the bottom side of the closet by conductive screws. Additional grounded points on each block's board can be added and electrically connected to the main body of the box wherever it is short and convenient.

The multicloset system assembly type looks nice and neat. There is not doubt that the shielding of the entire system assembly against outside interference is very good; however, its cost is high because the casting and gold-plating of the metallic box is expensive. Thus, the question is whether it is worth the cost to construct such an assembly or not, or alternatively, how valuable its performance would be.

The performance of this system assembly, as a matter of fact, can be characterized by the following three factors, in which problems are found correspondingly.

1. *Equipotentiality over the Entire Gold-Plated Surface of the Box.* The grounded potential may be different from block to block if the gold-plated surface of the box is not equipotential, especially when the size of the box is in the same order of the quarter wavelength or is greater than quarter wavelength. This may be a serious problem, though “the additional grounded point on each block board can be added and electrically connected to the main body of the box wherever it is short and convenient.”
2. *Forward Current Magnetic Coupling.* If the DC power supply nodes on each block are not equipotential, forward current magnetic coupling could greatly degrade the performance of the circuitry. Obviously, this is a problem in the multicloset type of system assembly shown in Figure 8.18 because “each individual block is fed with DC power supply from the common connector,  $V_{dd}$ , in series.”
3. *Return Current Magnetic Coupling.* If the voltage over all the gold-plated surface is not equipotential, return current magnetic coupling could also greatly degrade the performance of circuitry, and is likewise an obvious problem in the multicloset type of system assembly as shown in Figure 8.18 because the return currents from all individual blocks will flow over all the gold-plated surface with a complicated pattern and couple to each other.

**8.2.5.1 An Improved Path of DC power supply.** In Figure 8.18, the DC power supply  $V_{dd}$  from the common connector to the individual blocks is connected by



**Figure 8.19.** Improved connection path for DC power supply.

a path in series. This inevitably results in magnetic coupling in the forward current path. Figure 8.19 shows an alternative DC power supply path by which magnetic coupling of the forward current could be greatly reduced.

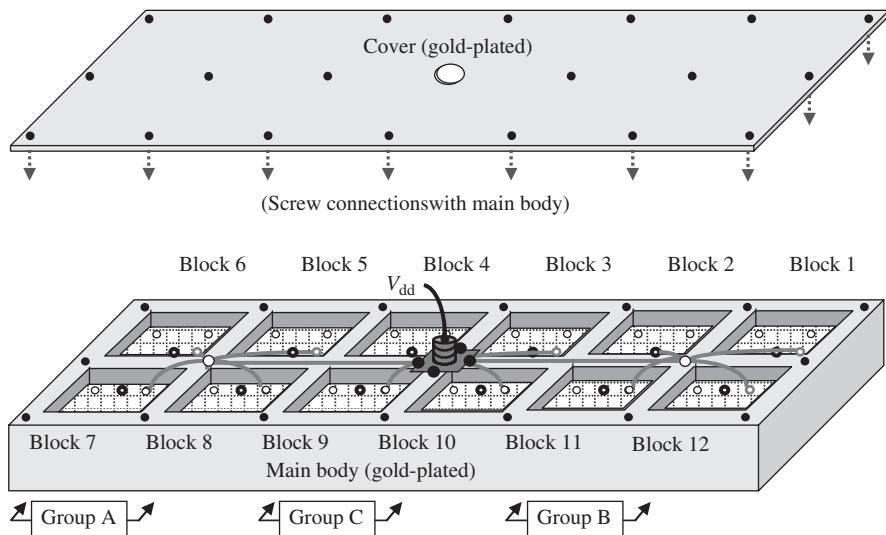
This differs from the path shown in Figure 8.18 in the following points:

- The terminal connector of the DC power supply is allocated at the center of the top side of the box, instead of the lateral side of the box.
- Ideally, the DC power supply must be provided from this DC terminal separately. This implies that 12 wires must go to their individual blocks respectively from the common DC power supply at the center of the box. In order to alleviate the crowding caused by 12 wires, the DC power supply can be divided into three groups of 4 wires, in which groups A and B are connected to the DC terminal along with group C, as shown in Figure 8.19.

**8.2.5.2 An Improved Path of Grounding.** In Figure 8.18, the entire surface of the gold-plated box functions as the ground surface for the individual blocks. The ground point for each block is simply selected on the surface of gold-plated box wherever it is short and convenient. This inevitably results magnetic coupling in the return current path. Figure 8.22 shows an improved grounding path by which the magnetic coupling of the return current can be greatly reduced.

This differs from the path shown in Figure 8.18 in the following points:

- The terminal connector of the common grounding is allocated at the center of the top side of the box, instead of at the lateral side of the box.
- Ideally, grounding in individual blocks must be provided from the common grounding terminal separately. This implies that 12 wires must go to their individual blocks respectively from the common grounding terminal at the center of the top side of the box. In order to alleviate the crowding caused by 12 wires, the grounding points can be divided into three groups of 4 wires, in which groups



**Figure 8.20.** Improved connection path for grounding.

A and B are connected to the center of the top side along with group C, as shown in Figure 8.20.

- It should be noted that only three grounding points corresponding to the three groups, and not the grounding points from individual blocks, are directly connected to the main body of the metallic box. This is the significant difference between Figures 8.20 and 8.18.

### 8.3 PCB OR IC CHIP WITH MULTIMETALLIC LAYERS

A multilayered PCB is often adapted in a complicated system. In parallel with multi-layer technology, there are multimetallic layers available in the IC circuit design today. Undoubtedly, multilayer technology is an important progression in the field of electrical engineering. It is very helpful to lay out a complicated system.

For instance, if the PCB of IC chip contains only one metallic layer, some runners must be replaced by jumping wires or “bridges” in the air where two runners must cross over each other. If the PCB contains a top and bottom metallic layer or if an IC chip contains two metallic layers, a runner can be moved up and down between the two metallic layers so that the jumping wires or bridges can be avoided.

A PCB with more than three metallic layers not only allows the runners laid on the PCB to be free of jumping wires or bridges, but also provides more flexibility to the runners’ layout. The following are some points to be noted as examples:

- In order to get better isolation between the RF and digital signals, one intermediate metallic layer can be assigned to function as a standard ground surface while the rest of the layers are assigned to contain all the runners. For example, all the runners involved with RF signals are collected in the top layers above the ground layer, while all the runners involved with digital signals are collected in the bottom layers below the ground layer.

- In order to get better isolation between RF and low-frequency signals, those runners involved with RF frequencies are collected in one metallic layer, while those runners involved with low frequencies are collected in another metallic layer.
- In order to get better isolation of a special part, that part can be “sandwiched” between two large grounded metallic portions in the metallic layers.

All these seem to be good ideas on the surface. However, for some reasons, the actual performance is often far from expectations. The expected good isolation seldom comes true although the expense for multiple metallic layers has been paid.

Therefore, it seems worthwhile to study and then answer the following questions:

- First, “Is the entire metallic layer an equipotential grounded surface?” If it is not equipotential, then the parts in the circuit connected to this metallic layer are referenced at different ground potential levels, and the performance, of course, would then be greatly different from that if the parts were referenced to an equipotential ground surface. The designer must verify the equipotentiality of the entire metallic layer before it is applied to the layout of the circuit. A demarcation line for equipotentiality is to see whether the maximum dimension of the metallic layer is much lower than the quarter wavelength of the highest operating frequency or not.
- If the entire metallic layer is verified to be an equipotential grounded surface, then the second question is, “Is it equipotential with the main ground reference point, the negative pole of the DC power supply?” If not, then forward and return currents will flow. Especially, return current will flow over the entire metallic layer with a very complicated pattern, making the case worse than when the entire metallic layer did not exist. In addition, if the metallic layer is not equipotential with the main ground reference point, not only RF return currents but also digital return currents would flow over the entire metallic layer and couple with the RF return currents. It would result in a big mess!
- Third, “Is the entire metallic layer equipotential with those grounded points not on the entire metallic layer?” and “By what method are they connected together?”
- Now, let us consider another potential big trouble source, the conductive vias from one layer to another. Conductive vias are the main tools to connect runners jumping from one layer to another. One should be aware that a conductive via is by no means a zero impedance part. Instead, it is a part with parasitic inductance, capacitance, and resistance, even though its entire surface is gold-plated. In either PCB or IC design, the equivalent model of a via should be developed and put into the circuit simulation. Parasitic inductance and resistance can screw up normal performance in a very bad way. For example, a conductive via with 10 mil diameter on an FR4 plastic PCB with 25 mil thickness may attenuate a 10-GHz signal by 5–10 dB.
- As for the sandwich configuration, one must worry about the additional capacitance between the top and bottom surface of sandwich, which could be considerable or, in extreme cases, form a resonant cavity.
- In practical design for PCBs or IC chips with multiple metallic layers, many more problems and questions can arise. For instance, forward current coupling may be due to providing an inappropriate DC power supply mode to an individual RF or digital block. Similarly, return current coupling may be due to an inappropriate grounding mode from the reference ground point to the individual ground point of an RF or digital block.

To summarize, a good design for a PCB or IC chip with multimetalllic layers must have the following conditions:

- The ground surface is equipotential with the reference ground point, that is, the negative pole of DC power supply.
- Forward and return current coupling are reduced to an insignificant level.
- The RF signal with the highest operating frequency, which flows through the conductive vias, does not have considerable attenuation.
- The additional capacitance, inductance, and resistance due to the multimetalllic configuration are insignificant.
- Additional interference between digital and digital, RF and RF, and RF and digital areas due to the multimetalllic configuration are negligible.

## FURTHER READING

- Jordan EC. Electromagnetic waves and radiating systems. 2nd ed. Prentice-Hall Inc; Englewood Cliffs, New Jersey 1968.
- Li R. Key issues in RF/RFIC circuit design. Beijing: Higher Education Press; 2005.
- Shim HW, Zeeff TM, Hubing TH. Decoupling strategies for printed circuit boards without power planes. Volume 1, IEEE International Symposium on Electromagnetic Compatibility; Rolla, Missouri 2000 Aug 19–23; 2002. pp. 258–261.
- Tenhunen H. CMOS Interconnects. (Lecture) Electronic System Laboratory. Kungl Tekniska Hogskolan; Stockholm, Sweden 2000.

## EXERCISES

1. What does forward current coupling mean? Usually, how can it be avoided?
2. What does return current coupling mean? Under what conditions does it happen? How can it be avoided?
3. By means of an open-circuited microstrip line with quarter wavelength to coerce a large PCB approaching to equipotentiality, which locations should they be put on the PCB to be most economical and effective?
4. The return current of input branch can flow back to the ground reference point over the ground surface. Alternatively, it can flow back to the ground point of the generator along the RF input cable. Which path is dominant?
5. A PCB with three metallic layers is designed for a system which contains both RF and digital circuitry. In order to enhance the isolation between the RF and the digital circuits, the second metallic layer is maintained as a common ground surface. The first metallic layer is the layout for the runners of all the RF circuits and the third metallic layer is the layout for the runners of all the digital circuits. Comment on such a PCB arrangement.
6. A PCB with two metallic layers is designed for a system which contains both RF and digital circuitry. What is a reasonable way to arrange its ground surface?

**ANSWERS**

1. On a PCB, some or all of the expected half grounded points are not equipotential to the reference grounding point so that some or all the currents flowing through the corresponding circuit branches would be magnetically coupled together in the current forward path from the DC power supply. The forward current coupling can be reduced or removed by using a bypass capacitor.
2. On the ground surface of a PCB, some or all of the expected fully grounded points are not equipotential to the reference grounding point so that the currents of some or all the currents flowing through the corresponding circuit branches would be magnetically coupled together in the current return path on the ground surface. The return current coupling can be reduced or removed by using a zero capacitor, or a  $\lambda/2$  microstrip line, or open-circuited  $\lambda/4$  microstrip line.
3. The microstrip lines should be placed on the four corners of the PCB.
4. The return current of input branch flowing back to the ground reference point over the ground surface dominates the return current flowing back to the ground point of the generator along the RF input cable.
5. It is a good PCB arrangement if the second metallic layer is equipotential to both RF and digital signals. There is no return current coupling on the second metallic layer. On the other hand, the second metallic layer plays the shielding role. On the contrary, it is a bad PCB arrangement if the second metallic layer is not equipotential to both RF and digital signals. The return current coupling will happen on the second metallic layer, which is the common ground surface. How to keep equipotentiality on the second metallic layer is quite a challenging subject to the PCB designer.
6. Ground surface must be separated into two portions, one for RF and the other for digital circuits.



# LAYOUT

## 9.1 DIFFERENCE IN LAYOUT BETWEEN AN INDIVIDUAL BLOCK AND A SYSTEM

The difference in layout between an individual block and a system can probably be summarized as follows:

First, the size of the PCB (printed circuit board) or the IC substrate for an individual block is usually smaller than that for a system. Therefore, the ground surface for an individual block is usually equipotential if the maximum dimension of the PCB or IC substrate is much shorter than the quarter wavelength corresponding to the operating frequency. Consequently, fewer grounding problems arise. On the contrary, the large ground surface of a system is usually not equipotential, so forward and return current coupling may take place. Coercive grounding is absolutely necessary and must be taken care of seriously.

Second, the number of metallic layers of the PCB or the IC substrate is different. Very often, a PCB or an IC substrate with two metallic layers is applied to lay out an individual block, whereas a PCB or an IC substrate with multiple metallic layers is applied to lay out a system. The conductive via is applied in both layouts. Via study is more important in the layout for a system than for a block because more vias with more complicated configurations are applied in the layout of a system.

Another remarkable difference is the number of DC power supplies. Multiple DC power supplies are usually required in the layout for a system, whereas a single power supply is usually sufficient in the layout for an individual block. Consequently, more

“zero” capacitors must be incorporated in the layout for a system than for an individual block.

## 9.2 PRIMARY CONSIDERATIONS OF A PCB

### 9.2.1 Types of PCBs

There are many types of PCBs existing in current electronic products, such as plastic, ceramic, glass, and so on. Nowadays, plastic and ceramic PCBs are the most popular types of board applied in the layout of both individual block and system. Table 9.1 compares the main features between plastic and ceramic PCBs.

The advantages of the plastic PCB in the circuit design are as follows:

- *Shorter processing times.* For example, only a couple of hours are needed to process a plastic PCB with two metallic layers. Today, advanced technology is available to process a plastic PCB with more than 10 metallic layers, though its processing time is longer and its cost is higher.
- *Robust mechanical property.* Compared to other types of PCB, a plastic PCB is mechanically more robust and sturdy. Its probability of suffering damage or being broken during packaging and transportation is low.

The disadvantage of using plastic PCBs in circuit design is the longer time needed to solder the parts on it manually. In the production line, the soldering process is accomplished by the “machine hand,” which costs much more than hand-soldering though it can shorten the soldering time.

As to the cost of using a plastic PCB in the circuit design, the cost of a PCB available for operation in the lower frequency range is lower than that of other types of PCB. However, the cost of a PCB available to operate in the high-frequency range, say, more than 2 GHz, is higher than that of other types of PCB.

On the contrary, the advantages and disadvantages of a ceramic PCB are just the opposite.

The advantages of a ceramic PCB in circuit design are the following:

- *Shorter soldering times.* For example, only a couple minutes is needed to solder the parts on a ceramic PCB by heating over an electric stove.

TABLE 9.1. Comparison of the Main Features between Plastic and Ceramic PCBs

	Plastic PCB	Ceramic PCB
Processing time	Couple of hours of processing for two metal layers	Couple of days of processing for two metal layers
Soldering of parts	Couple of hours By soldering iron or machine	Couple of minutes By heating over an electric stove
Size	Large	Small
Mechanical property	Robust and sturdy	Fragile, easily broken
Cost (Low frequency)	Low	High
(High frequency)	High	Low

- *Small size.* Because of its higher dielectric constant, its size can be smaller than that of a plastic PCB. The higher the dielectric constant, the smaller the size.

The disadvantages of using a ceramic PCB in circuit design are the following:

- *Longer processing times.* Because of the via penetration on the ceramic PCB by laser technology, its processing time is much longer than that of plastic PCB.
- *Poor mechanical property.* Compared to the plastic PCB, the mechanical property of a ceramic PCB is poor. Its probability of suffering damage or being broken during packaging and transportation is high.

As to the cost using a ceramic PCB in circuit design, the cost of a ceramic PCB available for operation in the lower frequency range is higher than that of a plastic PCB. However, the ceramic PCB can be applied in the higher frequency range, say, more than 2 GHz, with a cost lower than that of the plastic PCB.

### 9.2.2 Main Electromagnetic Parameters

- *Upper Limit of Operating Frequency.* The primary factor in choosing the raw PCB is the upper frequency limit of the PCB. If the operating frequency of the circuit is higher than the upper frequency limit of the PCB, the signal will suffer, with significant attenuation or distortion. For example, the upper frequency limit of an FR4 type plastic PCB is about 2.5 GHz. In general, the upper frequency limit of a PCB must be guaranteed to be at least 1.5 times higher than the highest operating frequency.
- *Dielectric Constant of PCB Material.* The dielectric constant of an FR4 type plastic PCB is around 4.3, while that of a ceramic PCB varies from less than 10 to 100s. The selection of the PCB dielectric constant is mainly determined by the considerations of PCB size, layout tolerance, and cost.
- *Attenuation Factor or Loss Angle.* The attenuation of signal due to imperfect PCB material is one of the important parameters of a PCB. This is specified by its loss angle. Usually, attenuation increases with the operating frequency. Consequently, the loss angle determines the upper limit of the operating frequency of the PCB.
- *Variation of Temperature Coefficient*, and so on.

The parameters mentioned above are the required information at the simulation stage.

### 9.2.3 Size

If the dimension of a PCB—either the length or the width—is much shorter than the quarter wavelength of a runner with  $50\text{-}\Omega$  characteristic impedance, then

$$L \ll \lambda/4, \quad (9.1)$$

where  $L$  is the maximum dimension of a PCB and  $\lambda$  is the wavelength corresponding to the operating frequency.

Expression (9.1) is the definition of a small-sized PCB.

On the contrary, if the dimension of a PCB—again, either the length or the width—is comparable to or greater than the quarter wavelength of a runner with a  $50\text{-}\Omega$  characteristic impedance, then

$$L \sim \text{or} > \lambda/4. \quad (9.2)$$

Expression (9.2) is the definition of a large-sized PCB.

A PCB has essentially different characteristics in these two sizes. In the case of a small-sized PCB, the ground surface is very close to being equipotential, whereas in the case of a large-sized PCB, equipotentiality is definitely not available on the ground surface. Therefore, a small-sized PCB is preferable for testing and packaging. A large-sized PCB requires special treatment, namely, coercive grounding on the ground surface.

#### 9.2.4 Number of Metallic Layers

How many metal layers should the PCB have? It depends on the complexity of the circuit to be built on the PCB.

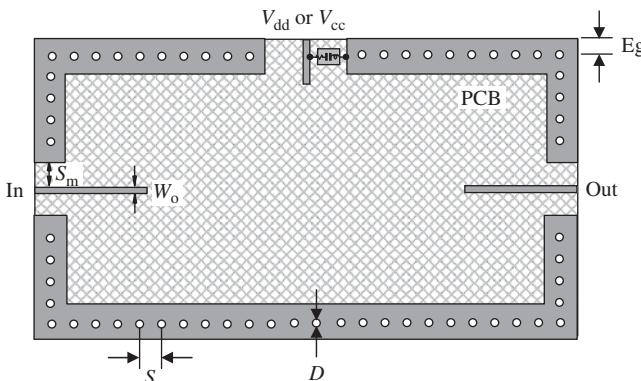
For an individual RF block, a PCB with two metal layers is preferred. The runners run on the top metallic layer, while most of the bottom metallic layer functions as the ground surface. Occasionally, a small area of the bottom layer may function as a jumping wire through two vias from the top to the bottom layer when a runner on the top layer is required to crossover to another one. A PCB with two metallic layers has the advantages of the simplest configuration and lowest cost. Layout problems on a PCB with two metallic layers can be easily found by manual inspection. On the contrary, layout problems in the intermediate layers of a PCB with more than two metallic layers are difficult and time consuming to detect. It is therefore preferable to use a PCB with two metallic layers in circuit implementation.

However, the entire circuitry of a complicated system may be too much for a PCB with only two metallic layers, in which case a PCB with multiple metallic layers is needed. Circuits on different metallic layers are connected by means of conductive vias. The distribution and arrangement of the circuitry and the conductive vias are important subjects in the layout of a system circuitry construction, and is discussed further in the following sections.

### 9.3 LAYOUT OF A PCB FOR TESTING

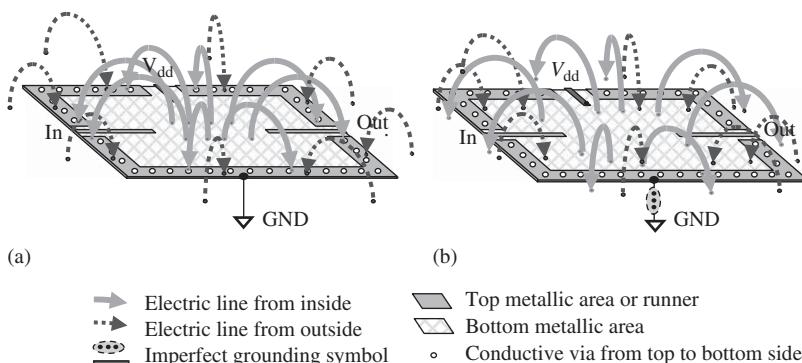
Figure 9.1 shows a preferred layout of a PCB for testing, in which the printed circuit on the PCB has been omitted. This is a PCB with only two metallic layers, one on the top and the other at the bottom. Usually, the entire metallic bottom layer remains as the main ground surface.

With the exception of three openings for input, output, and DC power supply entry, a rectangular ground frame, which is connected with the bottom metallic layer by many vias, is the main portion of the top metallic layer. The purpose of the rectangular ground frame on the top is the convenience of grounding. On the other hand, it also plays an appreciable shielding role. Figure 9.2(a) shows that most of the electric field lines radiating inside the PCB would bend down and terminate on the rectangular metallic frame; meanwhile, most of the electric field lines radiating outside the PCB would bend down and terminate on the rectangular metallic frame as well, as long as this metallic



- Top metallic area
- Bottom metallic
- 50  $\Omega$  runner
- Conductive via from top to bottom
- ▨ “Zero” capacitor

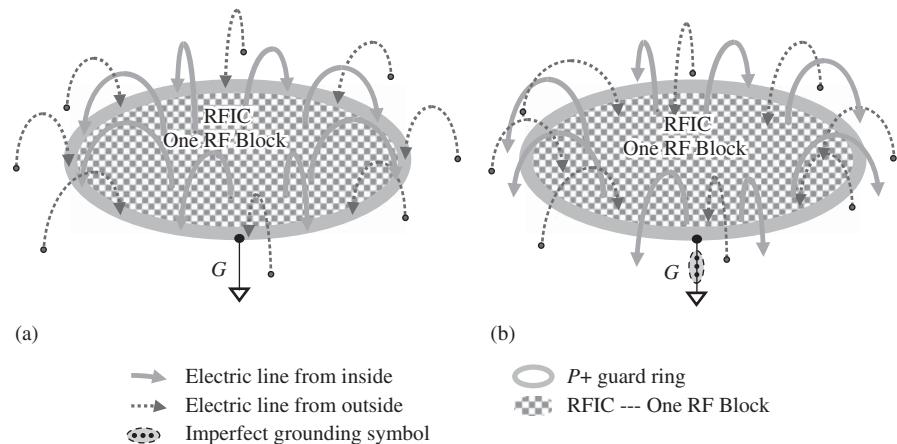
**Figure 9.1.** Preferred layout of a test PCB (the printed circuit on the PCB is not shown).



**Figure 9.2.** Electric field lines radiating either from the test circuit inside the PCB or from any interference source outside the PCB would be terminated on the rectangular grounded frame. (a) Perfect RF/AC grounding. (b) Imperfect RF/AC grounding.

frame is well grounded. (We say “most of electric field lines” but not “all electric field lines” because some of them would cross over the barrier of the rectangular grounded frame since the environment of the PCB is somewhat more complicated than a free space.) However, if this metallic frame is not well grounded as shown in Figure 9.2(b), the shielding function of this metallic frame would be absent, in which case most of the electric lines radiating inside the PCB would not bend down and terminate on the rectangular metallic frame and most of the field lines radiating outside the PCB would not bend down and terminate on the rectangular metallic frame as well.

In RFIC design, the P+ guard ring encircling the RF block plays a similar shielding role as the rectangular metallic frame on a PCB if the P+ guard ring is well grounded. Figure 9.3(a) shows such a function, in which most of the electric lines radiating inside the guard ring are bent down and terminated on the P+ guard ring and most of the electric lines radiating outside the guard ring are bent down and terminated on the P+ guard ring as well, if this P+ guard ring is well grounded. However, if this P+ guard ring is not well grounded as shown in Figure 9.3(b), the shielding function of guard ring



**Figure 9.3.** Shielding effect of P+ guard ring. (a) Perfect RF/AC grounding. (b) imperfect RF/AC grounding.

would disappear, in which case most of the electric lines radiating inside the guard ring would not be bent down and terminate on the P+ guard ring and most of the electric lines radiating outside the guard ring would not be bent down and terminate on the P+ guard ring as well.

In the DC or low-frequency range, the rectangular metallic frame on the top is undoubtedly well connected to the bottom metal layer. The rectangular metallic frame on the top is then at the same potential as the bottom metal layer. However, in the RF frequency range, they may not be equipotential to each other. In other words, the rectangular metallic frame might not be grounded well even though it is connected to the bottom ground by the conductive vias. The grounding state of the rectangular metallic frame mainly depends on the diameter of the conductive vias and the spacing between the vias. From practical engineering design experience, a good grounding state of the rectangular metallic frame is approached if

$$D > 10 \text{ mil}, \quad (9.3)$$

$$S \approx 4D \text{ to } 10D, \quad (9.4)$$

where

$D$  = the diameter of the conductive vias, and

$S$  = the spacing between the vias.

A conductive via is equivalent to a combination of an inductor, a capacitor, and a resistor. The values of these equivalent components mainly depend on the diameter of the via and other parameters related to the PCB's material and configuration. Usually, the equivalent inductance and resistance increase as the diameter of via is reduced. It is therefore desirable to make the vias as large as possible. Empirically, the diameter of via should be greater than 10 mils in the RF frequency range. On the other hand, in order to connect the rectangular metallic frame with the bottom ground well, it is desirable to punch more vias in the PCB. Empirically, the spacing between vias can be taken as  $4D$  to  $10D$ , depending on the operating frequency. The spacing  $S = 4D$  is for high RF-operating frequencies, whereas the spacing  $S = 10D$  is for low RF-operating frequencies.

Another two parameters are shown in Figure 9.1. One of them is the spacing between the input or output runner and the adjacent ground edge, that is,  $S_m$ , and the other is the spacing between the via and the edge of the rectangular metallic frame on the top, that is,  $Eg$ .

At the input, output, and DC power supply ports, one might like to paste more ground area beside the runners. However, the spacing  $S_m$  between the input or output runner and the adjacent ground edge must be wide enough so that the edge stray capacitance between the input or output runner and the edge of the ground area on both sides can be neglected. Empirically, the following condition must be satisfied:

$$S_m > 3W_o, \quad (9.5)$$

where

$S_m$  = the spacing between input or output runner and the adjacent ground edge  
and

$W_o$  = the width of input or output runner.

Usually, the spacing between the via and the edge of the rectangular metallic frame on the top is equal to the diameter of the vias, that is,

$$Eg \approx D, \quad (9.6)$$

where  $Eg$  is the spacing between the via and the edge of the rectangular metallic frame on the top.

## 9.4 VIA MODELING

In PCB design, the empirical formulas (9.3) to (9.6) for via have been arrived at from practical engineering experience. However, more concise modeling for vias must be developed since the via is one of the important parts of a multilayer PCB and plays an important role in circuit performance in the design of RF or digital circuits. On the other hand, vias constitute about 30–40% of total PCB cost. As a matter of fact, active study of vias is in progress nowadays.

In IC design, there are three types of via as shown in Figure 9.4:

1. *Blind via*. Located from the top metallic layer to an inner metallic layer between the top and bottom. Usually its depth is less than the diameter of the hole.
2. *Buried via*. Located between two inner metallic layers. It cannot be seen on either the top or bottom metallic layers.
3. *Through via*. A via from the top layer to the bottom layer. In other words, it a via going through all the layers.

### 9.4.1 Single Via

A via consists of two parts: one is the drilled hole and the other is the pad around the drill hole. Smaller vias are always preferred by the designers so that they able to drill more vias on the PCB; in addition, the parasitic capacitance of a via would be reduced as the size of via is decreased. However, decreasing the size of the via increases

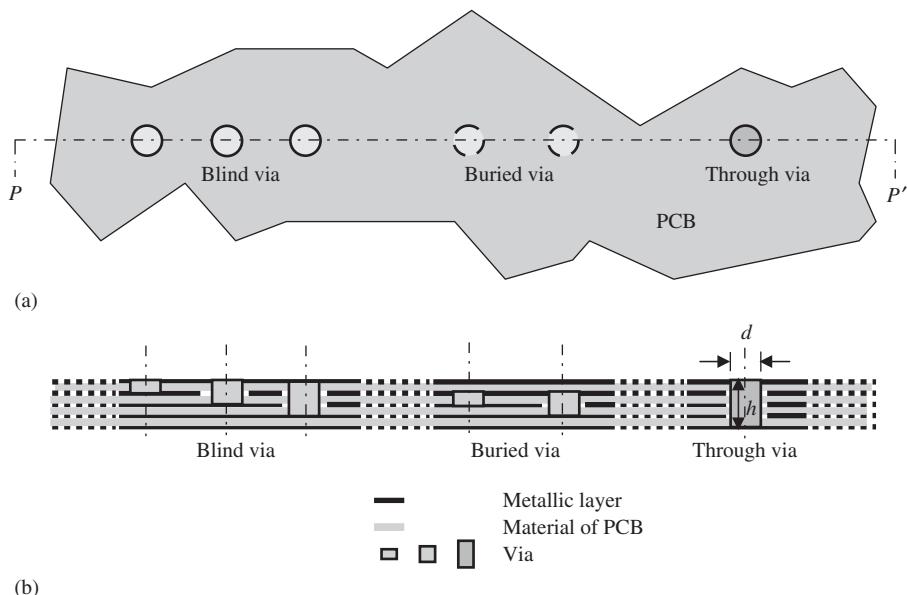


Figure 9.4. Three types of via on PCB. (a) Top view. (b) P–P' profile.

the cost. On the other hand, the size of via cannot be reduced indefinitely because of considerations of drilling and plating the via. The drilling time would be long and the drilled via would deviate from its original central position as the size of via is decreased. When the height or depth of the via becomes 6 times longer than the diameter of the via, a homogeneous copper plating on the wall of the drilled hole becomes impossible. For example, the minimum diameter of via for a PCB with six metallic layers and 50 mils thickness would be greater than 8 mils.

Figure 9.5 shows a single through via on a PCB or IC substrate. The parasitic capacitance  $C$  of a via can be approximately expressed as

$$C = 1.41 \cdot \varepsilon \cdot h \cdot \frac{d_1}{d_2 - d_1}, \quad (9.7)$$

where

$C$  = the parasitic capacitance of a via in pF,

$\varepsilon$  = the dielectric constant of the PCB material,

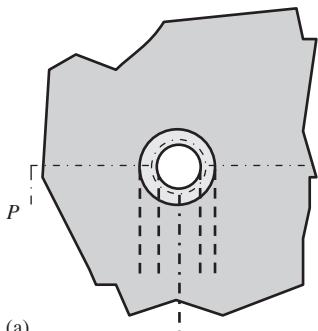
$h$  = the height of via or thickness of PCB in inches,

$d_1$  = the diameter of the drilled via in inches, and

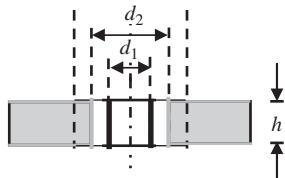
$d_2$  = the diameter of the isolated via of ground surface around the drilled via in inches.

Expression (9.7) is similar to that of capacitance for an ideal plane capacitor. The empirical constant “1.41” in expression (9.7) is due to the edge effect of the electrical field lines between the drilled via and the isolated via of ground surface around the drilled via.

For example, if  $\varepsilon = 4.4$ ,  $h = 50$  mils,  $d_1 = 20$  mils, and  $d_2 = 32$  mils, then  $C = 0.517$  pF.



(a)



(b)

**Figure 9.5.** Simple single via. (a) Top view. (b) P–P' profile.

Based on a cavity model analysis, the parasitic inductance  $L_{\text{via}}$  of the via can be approximately expressed as

$$L_{\text{via}} = 5.08 \cdot h \cdot \ln \left[ \frac{4h}{d_1} + 1 \right], \quad (9.8)$$

where

$L_{\text{via}}$  = the parasitic inductance of a via in nanohenries (nH),

$h$  = the height of via in inches, and

$d_1$  = the diameter of the drilled via in inches.

For example, if  $h = 50$  mils and  $d_1 = 10$  mils, then  $L = 1.015$  nH.

In the design for an RF circuit or a digital circuit to operate at data rates, the effect of parasitic inductance of a via on the circuit performance usually dominates over the effect of parasitic capacitance. The parasitic inductance is more dependent on the height than on the diameter, although the smaller the diameter, the higher the parasitic inductance.

In the design of an RF circuit or a digital circuit for operation at high data rates, any small and simple via could have a significant impact on the circuit performance. Consequently, the following are important points to a circuit designer:

1. Try to lay out the circuitry with as few vias as possible;
2. Try to use as thin a PCB as possible;
3. Try to make the diameter of the via as large as possible.

For a microstrip via or a circular-pad via, a more accurate expression for the parasitic inductance of a via was derived by Goldfarb and Pucel (1991):

$$L_{\text{via}} = \frac{\mu_0}{2\pi} \left[ h \cdot \ln \left( \frac{h + \sqrt{r^2 + h^2}}{r} \right) + \frac{3}{2} \left( r - \sqrt{r^2 + h^2} \right) \right], \quad (9.9)$$

where

$$\mu_0 = \text{the permeability of free space} = 4\pi \times 10^{-7} \text{ H/m},$$

$r = d_1/2 = \text{the radius of via in meters, and}$

$h = \text{the height of via in meters.}$

Let us introduce a via model on PZT (lead zirconate titanate) thin film (Bakar et al., 2008). The proposed equivalent circuit model derived for the via hole is shown in Figure 9.6. The via grounding is modeled by a series inductance ( $L_o$ ) and a shunt capacitance ( $C_{via}$ ) representing via hole pad; shunt inductance ( $L_{via}$ ) and resistance ( $r_{via}$ ) representing via metal plate; and another shunt resistance ( $r_o$ ) associated with dielectric loss. All the extracted equivalent circuit parameter values for a  $50 \times 50 \mu\text{m}^2$  via and cylindrical vias from 10 to  $50 \mu\text{m}$  diameter are summarized in Table 9.2.

So far, we have discussed only a single through via. The study of a single blind via or a single buried via is in progress and is still far from comprehensive understanding. Further works are expected in the near future.

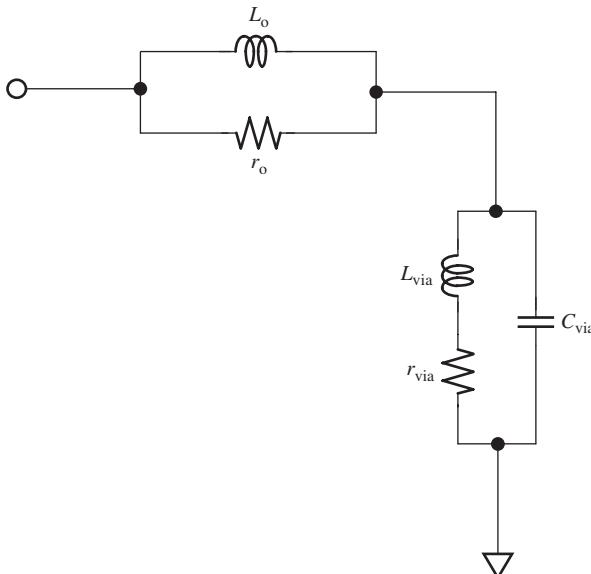


Figure 9.6. Proposed via hole equivalent circuit.

TABLE 9.2. Extracted Equivalent Circuit Parameter Values for a Via Hole on PZT (Lead Zirconate Titanate) Thin Films with Thickness =  $0.5 \mu\text{m}$ ;  $\epsilon = 300$

Via hole	$L_{via}$ , pH	$r_{via}$ , $\Omega$	$C_{via}$ , pF	$L_o$ , pH	$r_o$ , $\Omega$
$d = 10 \mu\text{m}$	2.550	0.040	2.985	1.904	0.284
$d = 20 \mu\text{m}$	2.405	0.034	3.256	1.904	0.245
$d = 30 \mu\text{m}$	2.314	0.023	3.590	1.727	0.174
$d = 40 \mu\text{m}$	2.204	0.019	4.363	1.566	0.118
$d = 50 \mu\text{m}$	1.904	0.016	6.446	1.060	0.102
$50 \times 50 \mu\text{m}^2$	1.813	0.014	7.462	0.962	0.097
$50 \times 50 \mu\text{m}^2$	1.813	0.014	7.462	0.962	0.097

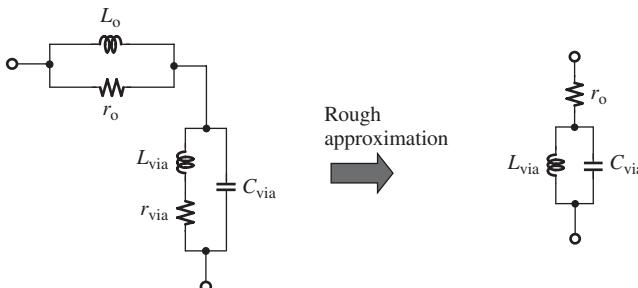
### 9.4.2 Multivias

The equivalent of a single via is a complicated combination of two inductors  $L_o$  and  $L_{via}$ , two resistors  $r_o$  and  $r_{via}$ , and one capacitor  $C_{via}$ , as shown in Figure 9.6. It can be roughly simplified to a simple combination of one inductor  $L_{via}$ , one resistor  $r_{via}$ , and one capacitor  $C_{via}$ , as shown in Figure 9.7 because it can be found from Table 9.2 that

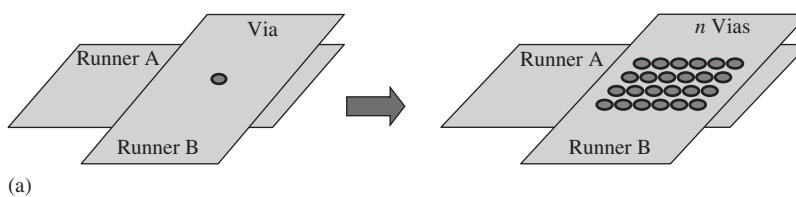
$$L_o < L_{via}, \quad (9.10)$$

$$r_o > r_{via}. \quad (9.11)$$

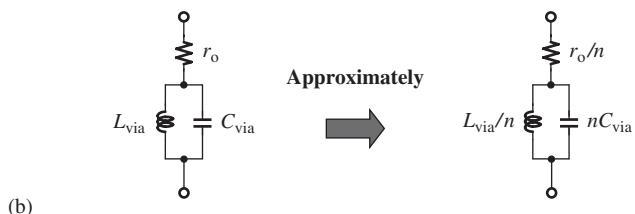
These vias are very often used for connecting two runners in an IC die. As shown in Figure 9.8(a), runner A can be connected to runner B by a single conductive via. Of course, it is not a perfect connection because the impedance of a via is not zero, but a complicated equivalent, as shown in Figure 9.7. In order to approach a good connection, the values of  $r_o$  and  $L_{via}$  must be reduced, whereas the value of  $C_{via}$  must be enhanced. As mentioned above, this can be done by increasing the via diameter. However, in an IC design, the via diameter cannot be changed but is fixed by the foundry. An alternative is to increase the number of vias as much as possible, say, from 1 to  $n$ , as shown in Figure 9.8(b). This results in an  $n$ -fold reduction of the values of  $r_o$  and  $L_{via}$  and approximately an  $n$ -fold increase in the value of  $C_{via}$  so as to approach a good connection.



**Figure 9.7.** Rough approximation of a single via's equivalent.



(a)



(b)

**Figure 9.8.** Changed equivalent as one via connection is replaced by the multivia connection. (a) One via connection is changed to multivia connection. (b) Equivalent when one via is changed to multivias.

## 9.5 RUNNER

The study of runners is extremely important in the circuit design because of the following reasons:

- A runner is a microstrip line. Therefore, a runner is just like any other part in the circuitry.
- In general, the number of runners is greater than the number of parts in the circuitry. Therefore, taking care of runners is very important in circuit design.

Let us examine the variation of impedance when a runner is inserted into the circuitry.

### 9.5.1 When a Runner is Connected with the Load in Series

Most runners are connected with the load in series, as shown in Figure 9.9.

We therefore have

$$Z_P = Z_o \frac{Z_L \cos \frac{2\pi l}{\lambda} + jZ_o \sin \frac{2\pi l}{\lambda}}{Z_o \cos \frac{2\pi l}{\lambda} + jZ_L \sin \frac{2\pi l}{\lambda}}. \quad (9.11)$$

There are three special cases for the length of a runner:

1. When

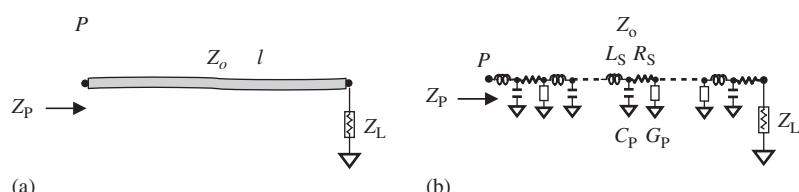
$$l \ll \frac{\lambda}{4}, \quad (9.12)$$

$$Z_P \approx Z_L. \quad (9.13)$$

2. When

$$l = n \frac{\lambda}{2} \quad (n = 1, 2, 3 \dots), \quad (9.14)$$

$$Z_P = Z_L. \quad (9.15)$$



**Figure 9.9.** Microstrip line with its characteristic and load impedance  $Z_o$  and  $Z_L$  and distribution parameters  $R_s$ ,  $L_s$ ,  $G_p$ ,  $C_p$ . (a) Microstrip line with its characteristic and load impedance  $Z_o$ ,  $Z_L$ . (b) Microstrip line with its distribution parameters  $R_s$ ,  $L_s$ ,  $G_p$ ,  $C_p$ .

3. When

$$l = n \frac{\lambda}{4} \quad (n = 1, 3, 5 \dots), \quad (9.16)$$

$$Z_P = \frac{Z_o^2}{Z_L}. \quad (9.17)$$

This is an impedance transformer case.

In the general case,

$$Z_P = Z_o \frac{Z_L Z_o}{Z_o^2 \cos^2 \beta l + Z_L^2 \sin^2 \beta l} + \frac{j(Z_o^2 - Z_L^2) \cos \beta l \sin \beta l}{Z_o^2 \cos^2 \beta l + Z_L^2 \sin^2 \beta l}. \quad (9.18)$$

There are three special cases for  $Z_L$ :

1. When

$$Z_L = Z_o, \quad (9.19)$$

$$Z_P = Z_o = Z_L. \quad (9.20)$$

This is an impedance-matched case.  $Z_P$  is equal to  $Z_L$  and is independent of the length of the runner.

2. When

$$Z_L = 0, \quad (9.21)$$

$$Z_P|_{Z_L \rightarrow 0} = jZ_o \tan \beta l = jZ_o \tan \frac{2\pi}{\lambda} l. \quad (9.22)$$

This is a short-circuited load case and the relationship between the impedance and the length is shown in Figure 9.10 by the solid line.

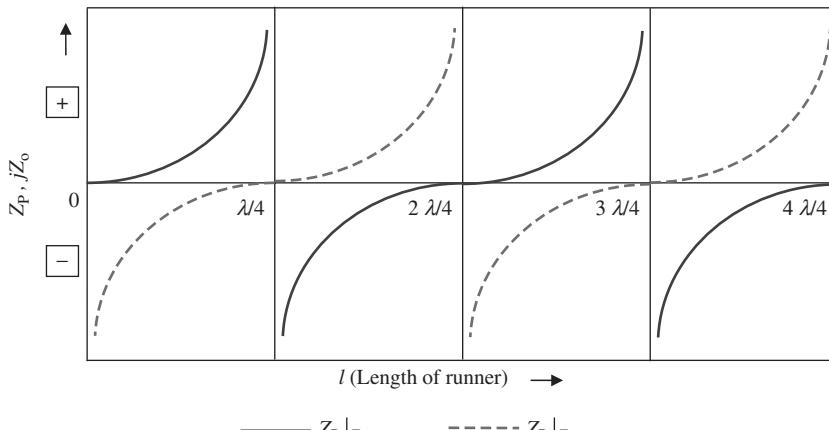


Figure 9.10. Impedance of runner in one wavelength when  $Z_L = 0$  and  $Z_L = \infty$ .

### 3. When

$$Z_L \rightarrow \infty, \quad (9.23)$$

$$Z_P|_{Z_L \rightarrow \infty} = \frac{1}{j} Z_0 \cot \beta l = -j Z_0 \cot \frac{2\pi}{\lambda} l. \quad (9.24)$$

This is an open-circuited load case and the relationship between the impedance and the length is shown in Figure 9.10 by the dashed line.

In general therefore, the following conclusions can be made:

1. The runner could be ignored, or in other words the variation of impedance from  $Z_L$  to  $Z_P$  approaches zero, that is,

$$Z_P \approx Z_L, \quad (9.25)$$

if

$$l \ll \frac{\lambda}{4}. \quad (9.26)$$

It is therefore encouraged to keep runners much shorter than their quarter wavelength in a circuit layout so that condition (9.25) is satisfied. If this is the case, we need not worry about the existence of runners.

2. The runner can be completely ignored, or in other words the variation of impedance from  $Z_L$  to  $Z_P$  is exactly equal to zero, that is,

$$Z_P \approx Z_L, \quad (9.27)$$

if

$$l = n\lambda/2 \quad (n = 1, 2, 3 \dots), \quad (9.28)$$

or

$$Z_0 = Z_L. \quad (9.29)$$

In circuit layout work, a runner with half wavelength or a multiple of half length is usually too long, so that condition (9.28) is meaningful in the coercing of grounding but is meaningless in the practical layout.

In circuit layout work, a runner with a characteristic impedance equal to its load impedance as shown in condition (9.29) is encouraged so that we do not need to worry about the existence of the runner if  $Z_L$  is a real number. The characteristic impedance of a runner is directly related to its width. It is better to select the width of the runner so that its characteristic impedance is equal to the load impedance. In the input or output nodes of an RF block or an RF system, the width of the runner is selected to correspond to  $50\text{-}\Omega$  impedance so as to be able to match the impedance of the test equipment, which is usually  $50 \Omega$ . At other nodes of circuitry, following condition (9.29), the width of a runner is preferred to correspond to the characteristic impedance  $Z_L$  of the part which is connected by that runner. However, condition (9.29) is usually not practical if  $Z_L$  is a complex number, whereas the  $Z_0$  is a real number.

### 9.5.2 When a Runner is Connected to the Load in Parallel

Figure 9.11 shows a runner with one end grounded and the other end connected to the load  $Z_L$  in parallel.

$$Z_1 = jZ_o \tan \beta l, \quad (9.30)$$

$$Z_P = Z_1 // Z_L = j \frac{jZ_o Z_L \tan \beta l}{Z_L + jZ_o \tan \beta l}, \quad (9.31)$$

$$Z_P = Z_o \frac{\frac{Z_L}{Z_o} \tan^2 \beta l}{\left(\frac{Z_L}{Z_o}\right)^2 + \tan^2 \beta l} + jZ_o \frac{j \left(\frac{Z_L}{Z_o}\right)^2 \tan^2 \beta l}{\left(\frac{Z_L}{Z_o}\right)^2 + \tan^2 \beta l}. \quad (9.32)$$

Figure 9.12 shows a runner with one end open and the other end connected to the load  $Z_L$  in parallel. A pad attached to the load corresponds to such a case.

$$Z_1 = -jZ_o \cot \beta l, \quad (9.33)$$

$$Z_P = Z_1 // Z_L = -j \frac{jZ_o Z_L \cot \beta l}{Z_L - jZ_o \cot \beta l}, \quad (9.34)$$

$$Z_P = Z_o \frac{\frac{Z_L}{Z_o} \cot^2 \beta l}{\left(\frac{Z_L}{Z_o}\right)^2 + \cot^2 \beta l} - jZ_o \frac{j \left(\frac{Z_L}{Z_o}\right)^2 \cot^2 \beta l}{\left(\frac{Z_L}{Z_o}\right)^2 + \cot^2 \beta l}. \quad (9.35)$$

A runner shorter than  $\lambda/4$  could be equivalent to an inductor if one of its terminations is grounded, or to a capacitor if one of its terminations is open (Table 9.3; Fig. 9.13).

### 9.5.3 Style of Runner

**9.5.3.1 Multiple Runners or Curves in Parallel.** In order to reduce the area of the IC die, it is a common practice to place multiple runners in parallel or in curves in a digital IC layout, as shown in Figure 9.14.

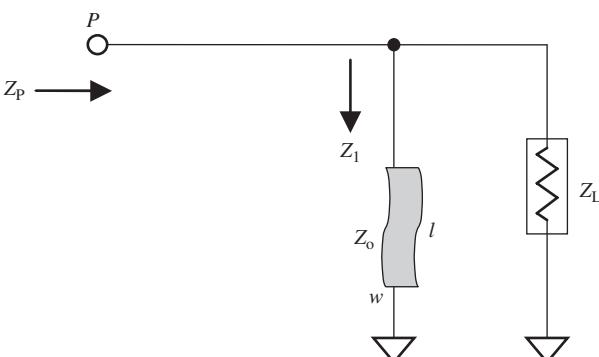
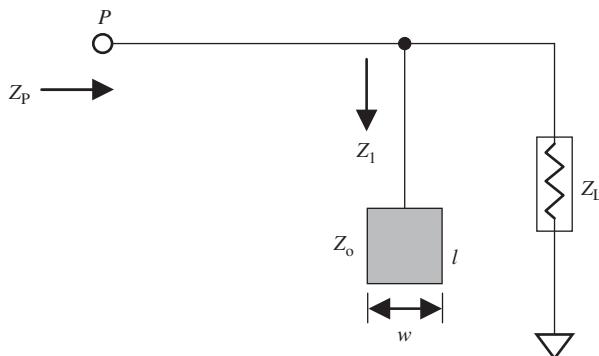
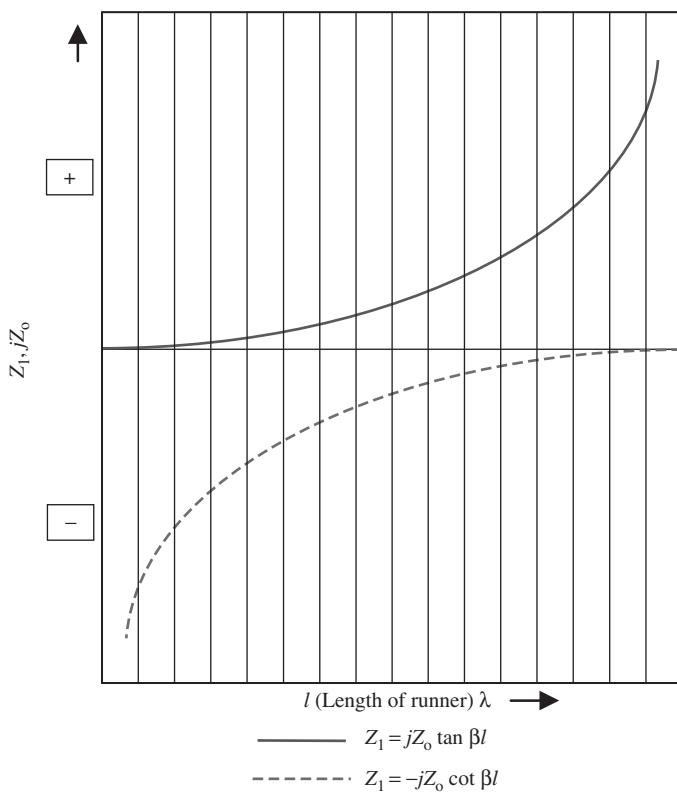


Figure 9.11. Microstrip line connected with load in parallel.



**Figure 9.12.** Pad connected with the load in parallel.



**Figure 9.13.** Impedance of runner in one quarter wavelength when  $Z_L = 0$  and  $Z_L = \infty$ .

However, this style of layout is absolutely not allowed in an RF circuit layout, because it brings about coplanar capacitance, mutual inductance, and so on, between the runners, and hence brings about a strong coupling between runners, eventually leading to malfunctioning of the circuitry.

**9.5.3.2 Runner with “nice looking” Style.** In layouts for digital circuitry, people like layout with nice looking styles. The runner is therefore aligned with an “East–West, South–North” orientation, appearing like streets and boulevards in a city’s traffic map, as shown in Figure 9.15. This is not the expected style in the layout for RF circuitry. Instead, the preferred style of the runner is such that it is “as short as possible.”

TABLE 9.3. Relationship between the Length of MSL and  $Z_1$  when  $Z_L = 0$  and  $Z_L = \infty$ 

$l$ (Length of runner)	$Z_1 = jZ_o \tan \beta l$	$Z_1 = -jZ_o \cot \beta l$
( $\lambda$ )	(Degree)	( $j\Omega$ )
0	0.00	0.00
1/64	5.63	4.92
2/64	11.25	9.95
3/64	16.88	15.17
4/64	22.50	20.71
5/64	28.13	26.73
6/64	33.75	33.41
7/64	39.38	41.03
<b>8/64</b>	<b>45.00</b>	<b>50.00</b>
9/64	50.63	60.93
10/64	56.25	74.83
11/64	61.88	93.54
12/64	67.50	120.71
13/64	73.13	164.83
14/64	78.75	251.37
15/64	84.38	507.68
1/4	90.00	$\infty$
		0.00

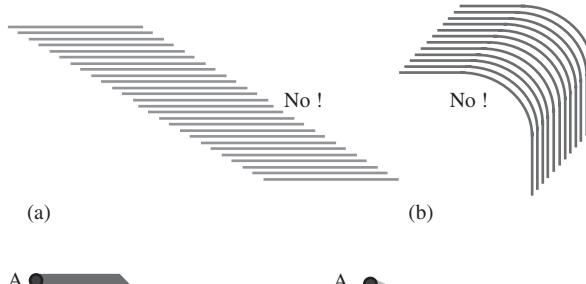


Figure 9.14. Multiple runners or multiple curves in parallel are not welcome. (a) Multiple runners in parallel. (b) Multiple curves in parallel.

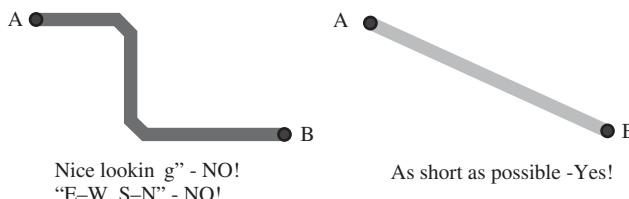
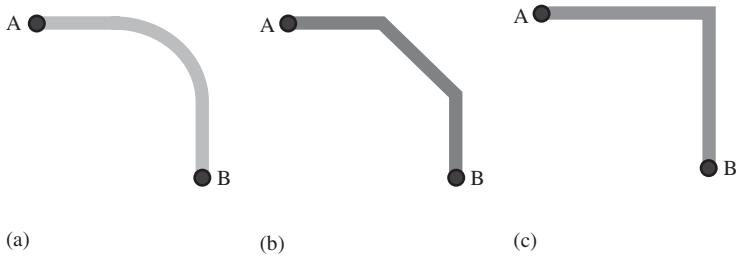


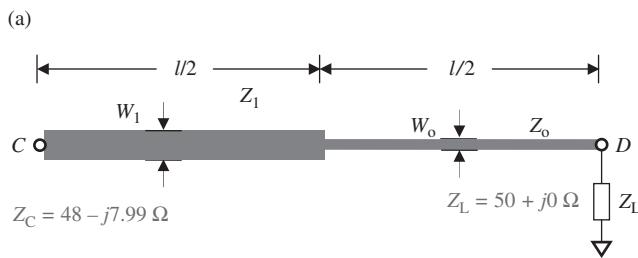
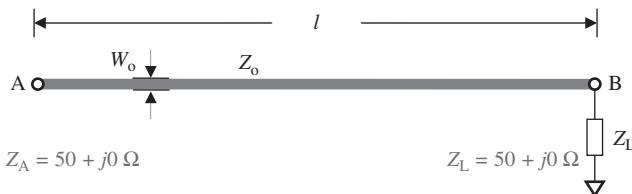
Figure 9.15. Two runner styles from A to B.

**9.5.3.3 Corner of Runner.** As shown in Figure 9.16, sometimes a runner from node A to node B must be bent at an approximately  $90^\circ$  angle. The layout shown in Figure 9.16(c) is not good because a strong electromagnetic field appears in the adjacent area of a runner with a sharp angle when an AC/RF signal is going through the runner. The radiation of the strong electromagnetic field consumes the power of the signal and also disturbs the signals in the circuitry. The layout shown in Figure 9.16(b) is an improvement. Finally, the layout shown in Figure 9.16(a) is the best. In short, one of the basic layout principles is to make the corner of the runner as smooth as possible.

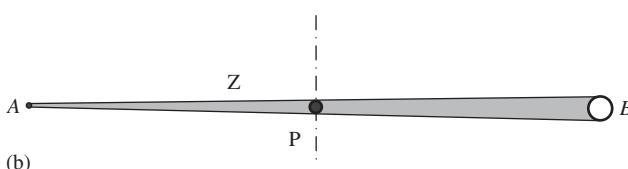
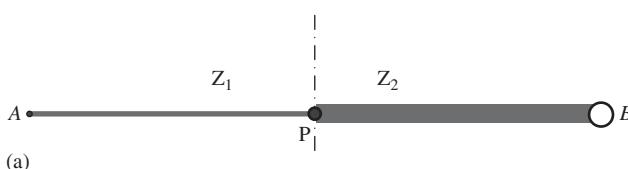
**9.5.3.4 Smoothness of the Runner.** The characteristic impedance of a runner is directly related to its width. Therefore, the width of a runner must not be changed



**Figure 9.16.** Three different corners of runner from A to B. (a) Circular—the best. (b)  $45^\circ$ —ok. (c) Rectangular—lousy.



Additional capacitor in series:  $-j7.99 \Omega = 1.99 \text{ pF}$   
 $f = 10.00 \text{ GHz}$



**Figure 9.17.** Comparison of impedance between even and uneven runner ( $f = 10 \text{ GHz}$ ).  
 (a) An even runner:  $W_o = 6 \mu\text{m}$ ,  $Z_o = 50.17 \Omega$ ,  $l = 100 \mu\text{m}$ . (b) An uneven runner:  $W_1 = 20 \mu\text{m}$ ,  $Z_1 = 26.65 \Omega$ ,  $l/2 = 50 \mu\text{m}$ ;  $W_o = 6 \mu\text{m}$ ,  $Z_o = 50.17 \Omega$ ,  $l/2 = 50 \mu\text{m}$ .

suddenly so as to avoid a sudden change of the characteristic impedance. In other words, the connection between two nodes must be an even runner as shown in Figure 9.17(a). If the width of a runner is suddenly changed, as shown in Figure 9.17(b), it becomes equivalent to an impedance transformer.

The width of a runner should be changed gradually as shown in Figure 9.18(b) if variation of a runner's width is necessary.

**9.5.3.5 Placement of Runners.** Runners running parallel to each other brings about coplanar capacitance, mutual inductance, and so on, between the runners, as mentioned above. Therefore, it is encouraged to lay out the runners not parallelly but perpendicularly as shown in Figure 9.19 if a PCB or an IC die has more than two metallic layers.

**9.5.3.6 Runners in Parallel.** If the runners must be parallel for some special purpose, the spacing between two runners must be wider than 3 times than the runner's width as shown in Figure 9.20 so that the electromagnetic interference between two adjacent runners can be reduced down below a conceivable level, that is,

$$S > 3W. \quad (9.36)$$

**9.5.3.7 Spacing between a Runner and the Ground Surface.** Similarly, as shown in Figure 9.21, the spacing between a runner and its adjacent ground surface in parallel must be greater than 3 times of the runner's width, that is,

$$S_{rg} > 3W. \quad (9.37)$$

If a runner is running between two adjacent ground surfaces as shown in Figure 9.22, the spacing between a runner and its two adjacent ground surfaces in parallel must be

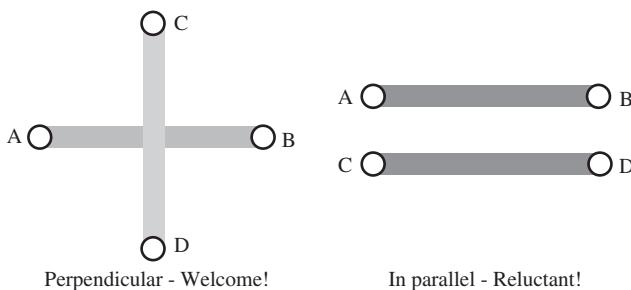


Figure 9.19. Two adjacent runners in perpendicular, not in parallel.

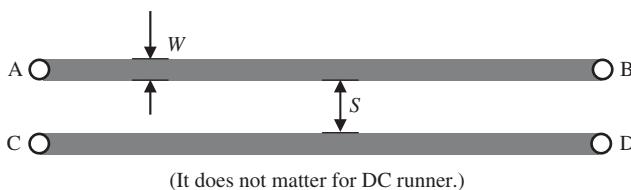


Figure 9.20. Spacing between two runners in parallel.

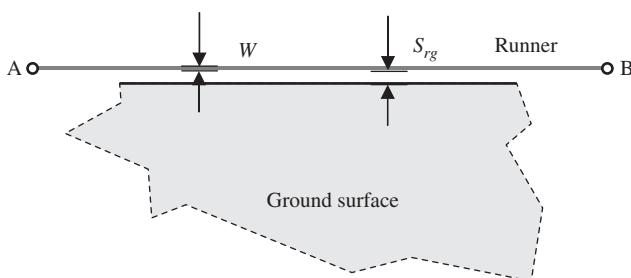
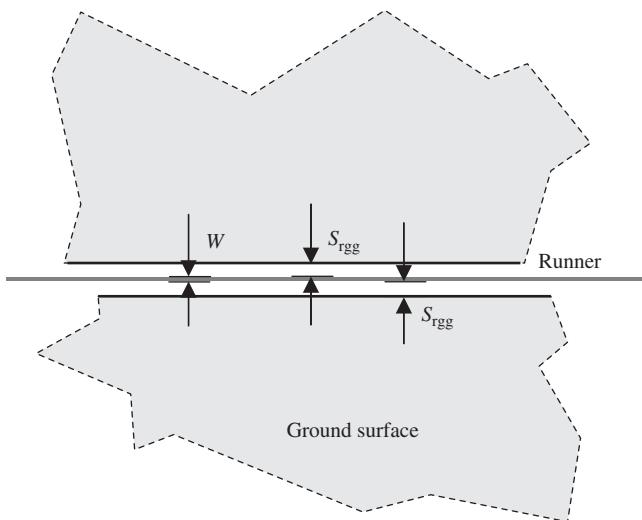
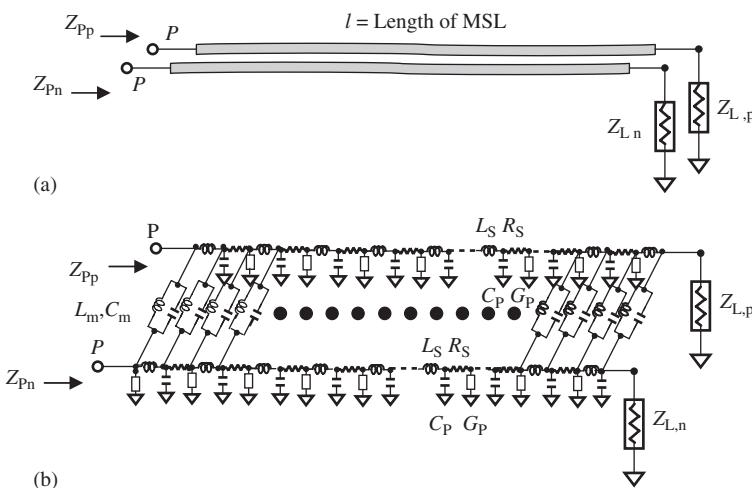


Figure 9.21. Spacing between the runner and the ground surface in parallel.



**Figure 9.22.** Spacing between the runner and two ground surfaces in parallel.



**Figure 9.23.** Runner (MSL) for a differential pair with its characteristic and load impedance  $Z_o$  and  $Z_L$ , and distribution parameters  $R_S$ ,  $L_S$ ,  $G_p$ ,  $C_p$ ,  $C_m$ , and  $L_m$ . (a) Microstrip lines for differential pair with its characteristic and load impedance  $Z_o$  and  $Z_L$ . (b) Equivalent circuitry of microstrip lines for a differential pair with its distribution parameters  $R_S$ ,  $L_S$ ,  $G_p$ ,  $C_p$ ,  $C_m$ , and  $L_m$ .

greater than 3 times of the runner's width, that is,

$$S_{rgg} > 3W. \quad (9.38)$$

**9.5.3.8 Special Runners for a Differential Pair.** As shown in Figure 9.23, a differential signal can be transported by two MSL lines in parallel, which are very close together so that they are strongly coupled with each other.

The equivalent model for two microstrip lines close together in parallel, as shown in Figure 9.23(a), is complicated and can be represented by Figure 9.23(b). The additional distribution parameters  $L_m$  and  $C_m$  represent the coupling between the two MSL lines. The configuration with two MSL lines with a narrow spacing is allowed and is beneficial to saving space. However, impedance must be matched in either input or output terminals.

Finally, the expected layout technology for the runners can be outlined in terms of three rules:

1. As short as possible
2. As smooth as possible
3. As perpendicular to each other as possible.

## 9.6 PARTS

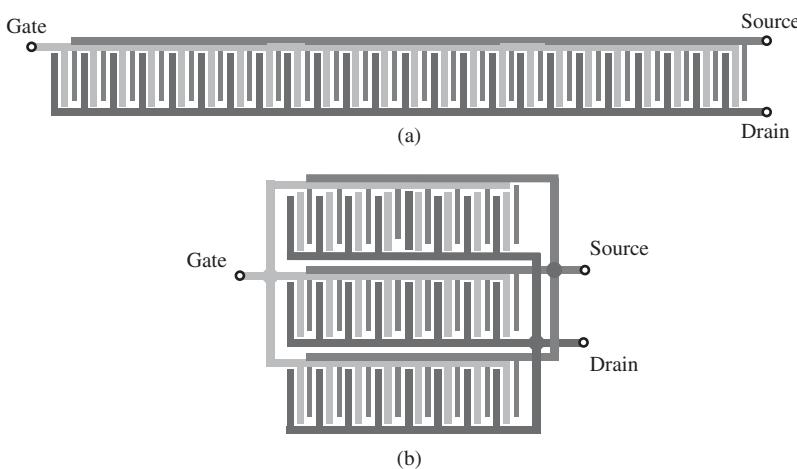
Considerations of how to handle the parts is another important objective in the layout of a circuitry.

### 9.6.1 Device

There is no choice in laying out a device if it is an individual part and is purchased from market or directly from a manufacturer. However, it is a very sensitive issue to lay out a device if the device is built on an IC chip. The size of the device can be adjusted by the designer by varying the number of its fingers. A device with only a few fingers is a small device, whereas a device with many fingers is a large device. There are no special considerations about the layout of a small device. On the contrary, when dealing with the layout for a large device with many fingers in an IC chip, all the fingers may align together and form a long “dragon.” A delay appears when AC/RF signals go along the “dragon” from one end to the other. Consequently, a phase shift appears from finger to finger, causing AC/RF current cancellation against each other when the currents from different fingers reach the terminal of the device. It is therefore preferable to arrange all the fingers of the device in a “square” shape instead of a “dragon” as shown in Figure 9.24.

### 9.6.2 Inductor

Inductors in an IC die occupy the largest area. Unfortunately, even more area must be allocated for a circuit block containing more than two inductors. As shown in Figure 9.25, a guard ring is needed for each inductor and the distance between inductors must



**Figure 9.24.** Devices must be arranged in a “square” but not a “dragon” shape. (a) Device laid out as a “dragon.” (b) Device laid out in a “square” shape. The latter is preferred.

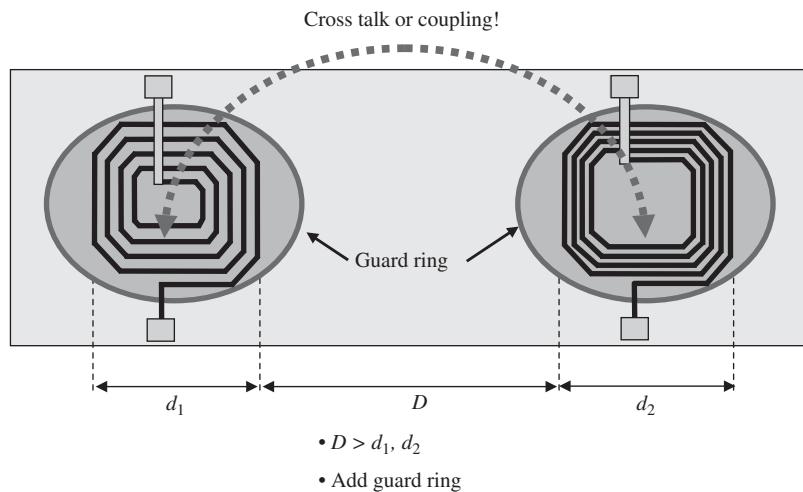


Figure 9.25. Cross talk between two inductors.

be larger than the highest diameter of the two spiral inductors. Otherwise, strong coupling between the inductors will take place.

### 9.6.3 Resistor

Compared to the other parts, a resistor in an IC chip has the highest tolerance. In order to reduce its tolerance, a desired resistor is always replaced by a combination of resistors with a higher resistance. The tolerance of the combination of resistors is the same as that of the individual resistor. However, a resistor with a higher resistance value usually has a lower relative tolerance. Consequently, the tolerance can be lowered by replacing a resistor by a combination of resistors with higher resistance. An example is shown in Figure 9.26, where an expected resistor with  $1\text{ k}\Omega$  is replaced by a combination of 10 resistors each with  $10\text{ k}\Omega$ .

### 9.6.4 Capacitor

As shown in Figure 9.27, one of the important R&D projects in IC circuit design at present is the reduction of the area of the capacitor, since the area of a capacitor in an IC chip

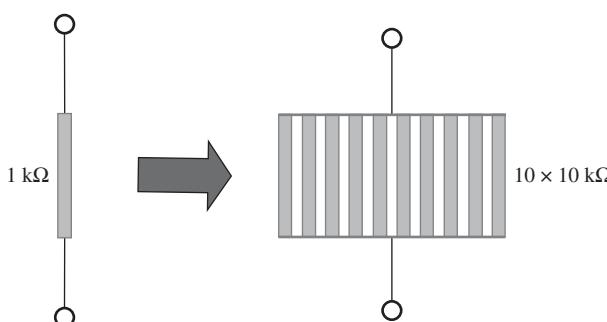
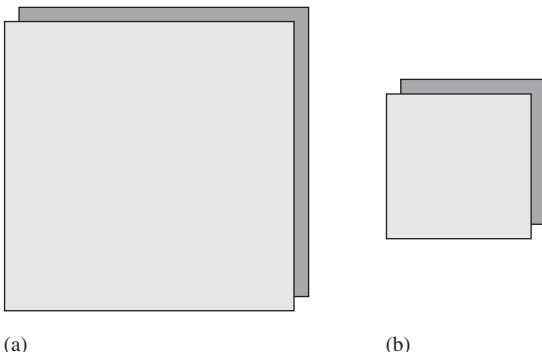


Figure 9.26. Desirable  $1\text{-k}\Omega$  resistor replaced by 10 resistors in parallel, the value of each resistor being  $10\text{ k}\Omega$ .



**Figure 9.27.** Area of capacitor is one of the important R&D projects at present. (a) Capacitor in RFIC limited by its area. It is rated less than about 20 pF. (b) Study of capacitor with high capacitance/area.

is large and the cost of an IC chip is proportional to its area. The objective of this R&D project is to search for an appropriate processing technique for capacitor implementation by which a capacitor with high capacitance per unit area can be obtained. Based on existing IC design technology levels, in order to avoid the high cost of IC design, a capacitor is not allowed in an IC chip if its value is greater than 20 pF. In other words, only those capacitors with a value less than 20 pF are allowed to be implemented on an IC chip.

## 9.7 FREE SPACE

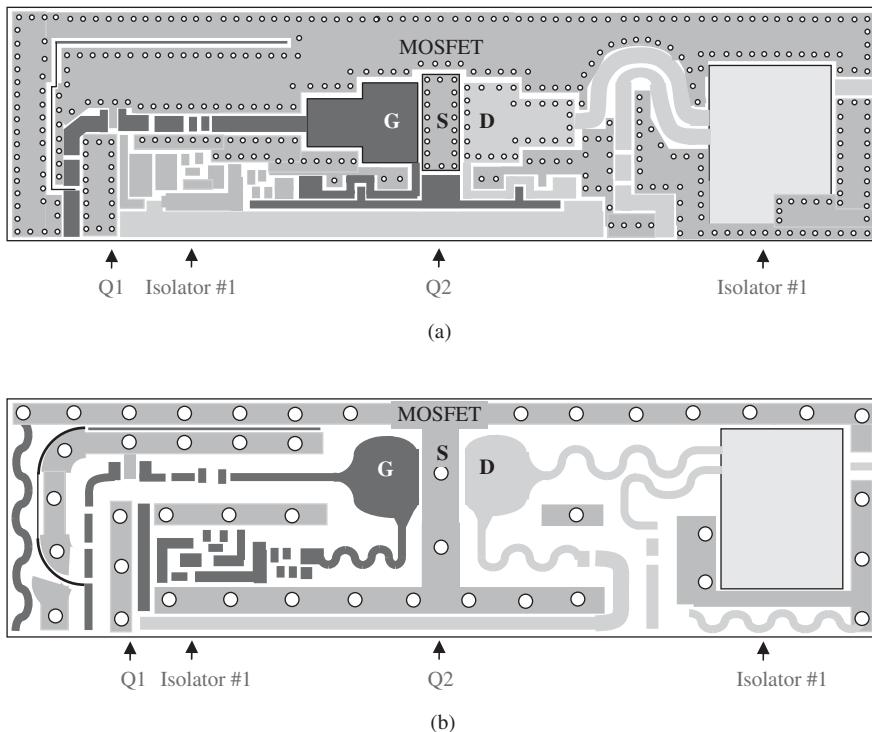
There is always free space in the layout of a PCB. Many designers like to fill up this space as partial ground surfaces. They might imagine that it would be beneficial for the isolation between circuit blocks or in shielding the special parts.

Figure 9.28(a) shows the original layout of a power amplifier in which the free space is filled with as much ground surface as possible.

The designer may celebrate his/her great achievement of no wasted space. Unfortunately, the performance of the power amplifier turns out to be far from expectations and is simply out of the range. After a study and deliberation of the problem for a while, it is found that the problem came from the wrong idea of filling free space with as much grounding as possible.

As a matter of fact, the original layout as shown in Figure 9.28(a) brings about the following:

- *Considerable coplanar capacitance.* Considerable coplanar capacitance is produced between the runners and between the runner and the adjacent ground surface.
- *Difference between standard MSL and coplanar MSL.* The designer may simulate the MSL with standard model, whereas in the layout it is replaced by a coplanar MSL. The difference between standard MSL and coplanar MSL becomes significant as the free space is filled.
- *Additional radiation due to the metallic pieces with sharp angles.* It is well known that a metallic piece with sharp angles forms a strong electromagnetic field around the sharp angle and consequently radiates radio waves and therefore increases attenuation of the signal. In particular, these problems exist in the layout portions for the gate and drain of a MOSFET device. They are formed by two metallic pieces of rectangular shape and then connected to a runner with a narrower runner.



**Figure 9.28.** Comparison of original and suggested layouts of a power amplifier (two stages, output power = 50 W). (a) Original layout. (b) Suggested layout.

This forms two impedance transformers due to the width of the metallic pieces suddenly changing twice. Consequently, it results in not only additional radiation, but also an unmatched impedance problem.

The suggested layout as shown in Figure 9.28(b) is able to improve the performance of the power amplifier considerably.

- *Reduction of coplanar capacitance.* Coplanar capacitance can be reduced by increasing the spacing between the runners and between the runner and the adjacent ground surface. The ratio of the spacing and the width of runner is kept greater than 3 : 1.
- *Elimination of coplanar MSL.* Owing to the increase of the spacing between the runners and between the runner and the adjacent ground surface, most coplanar MSL lines are eliminated. The designer simulating the MSL with the standard model is basically in agreement with its practical performance.
- *Reduction of the additional radiation due to the metallic pieces with sharp angles.* All the runners run in the circuitry smoothly without sharp angles. The additional radiation due to the metallic piece with sharp angle is therefore reduced significantly. The outstanding improvement is that the width of the metallic pieces in the layout portions for the gate and drain of the MOSFET varies gradually.

This results not only in reduction of additional radiation but also in the alleviation of the unmatched impedance problem.

- *Execution of equipotentiality on the entire ground surface.* In Figure 9.28(b), there are two microstrip lines with quarter wavelength added at the very west side and the east-bottom corner, which coerces the entire ground surface into being equipotential.

## REFERENCES

- Bakar RA, Awang Z, Sulaiman S, Lazim NFM, Khan ZI. Via hole modeling of PZT thin films for MMIC applications. In: 2008 IEEE International RF and Microwave Conference Proceedings; Kuala Lumpur, Malaysia; 2008 Dec 2–4. pp. 222–225.
- Goldfarb ME, Puce1 RA. Modeling via hole grounds in microstrip. IEEE Microw Guid Wave Lett 1991;1(6):135–137.

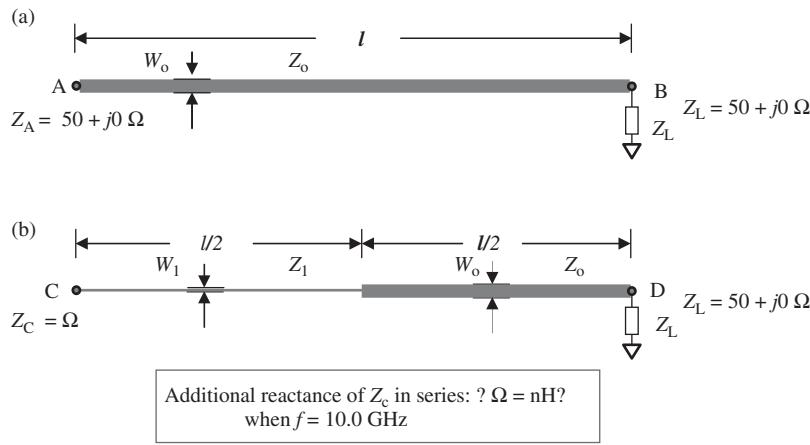
## FURTHER READING

- Cao Y, Simonovich L, Zhang Q-J. A broadband and parametric model of differential via holes using space-mapping neural network. IEEE Microw Wireless Compon Lett 2009;19(9):533–535.
- Gu Q, Yang YE, Tassoudji MA. Modeling and analysis of vias in multilayered integrated circuits. IEEE Trans Microw Theory Tech 1993;41(2):206.
- Kok P, Zutter DD. Capacitance of a circular symmetric model of a via hole including finite ground plane thickness. IEEE Trans Microw Theory Tech 1991;39(7):1229–1234.
- Kouzaev GA, Nikolova NK, Deen MJ. Circular-Pad via model based on cavity field analysis. IEEE Microw Wireless Compon Lett 2003;13(11):481–483.
- Wu T-L, Wang C-C, Kuo C-C, Hsieh J-S. A novel time-domain method for synthesizing broadband macro- $\pi$  models of differential via. IEEE Microw Wireless Compon Lett 2005;15(5):378–380.

## EXERCISES

1. What are the upper frequency limit and the relative dielectric constant of a FR4 PCB approximately?
2. Why does a guard ring in an RFIC with good grounding or an edge rectangular metallic frame on PCB with good grounding partially plays a shielding role?
3. A runner is inserted between two parts in the circuitry. What are the lengths of the runners for which the additional impedance due to the runner can be ignored?
4. A runner is inserted between two parts in the circuitry. What are the lengths of the runner such that the additional impedance due to the runner must be seriously taken care of?
5. Why are bunch of runners in parallel layout on a PCB or in the IC die very harmful to the circuit performance?
6. Why is a runner with a sharp corner a bad layout type in RF circuit design?
7. What is the minimum spacing that must be retained between two runners in parallel, or between a runner and the ground surface?

8. Refer to Figure 9.P.1, calculate the value of  $Z_C$  and the inductance when the operating frequency is 10 GHz.



**Figure 9.P.1.** Calculation of additional reactance due to the uneven layout. (a) An even runner:  $W_o = 6 \mu\text{m}$ ,  $Z_o = 50.2 \Omega$ ,  $l = 100 \mu\text{m}$ . (b) An uneven runner:  $W_1 = 2 \mu\text{m}$ ,  $Z_1 = 82.42 \Omega$ ,  $l/2 = 50 \mu\text{m}$ ;  $W_o = 6 \mu\text{m}$ ,  $Z_o = 50.2 \Omega$ ,  $l/2 = 50 \mu\text{m}$ .

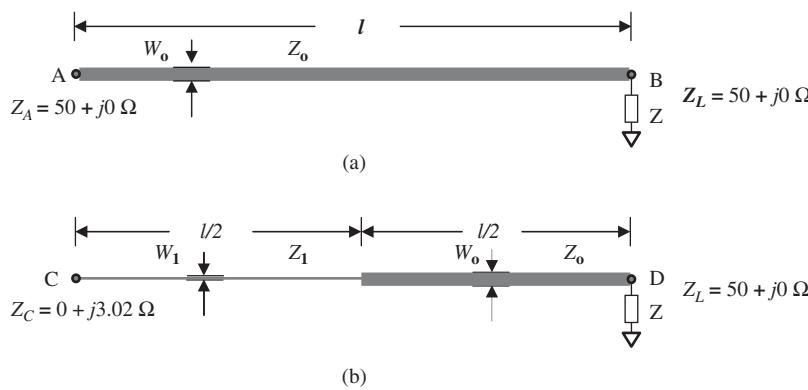
9. What is the equivalent circuit of a via?
10. How is the via's impedance reduced in RFIC design?
11. How is the via's impedance reduced in a PCB design?
12. In an RFIC layout, what is the important item to be taken care of for the layout of a large device with hundreds of fingers?
13. In an RFIC layout, what is the important item to be taken care of for the layout of a resistor?
14. Is it acceptable to fill all free space to function as part of ground surface on a layout?

## ANSWERS

1. Approximately the upper frequency limit and the relative dielectric constant of a FR4 PCB are 2.5–3.0 GHz and 4.3–4.6, respectively.
2. A guard ring in RFIC with good grounding or an edge rectangular metallic frame on PCB with good grounding plays the shielding role partially, because the electric field lines radiating either from the test circuit inside the PCB or from any interference source outside the PCB would be terminated on the guard ring in an RFIC with good grounding or on the rectangular grounded frame.
3. A runner inserted between two parts in the circuitry can be ignored if the length of the runner is
  - (a)  $l \ll \lambda/4$ , or
  - (b)  $l \approx n \lambda/2$  ( $n = 0, 1, 2, 3, 4 \dots$ ).
4. A runner inserted between two parts must be seriously taken care of if the length of the runner is

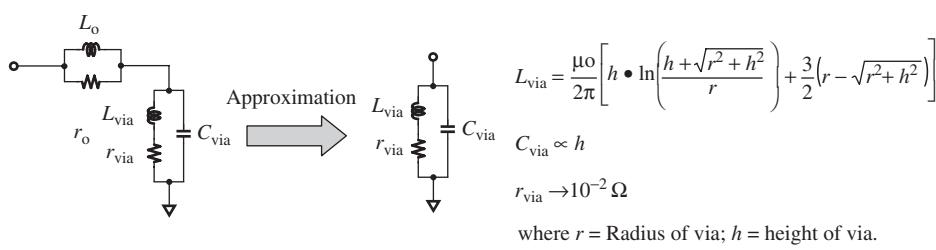
$$l \approx (2n + 1)\lambda/4, \quad (n = 0, 1, 2, 3, 4 \dots).$$

5. A bunch of runners in parallel layout on a PCB or in the IC die is very harmful to the circuit performance because it would bring appreciable coplanar capacitance, mutual inductance, and mutual capacitance.
6. A runner with sharp corners is a bad layout type in the RF circuit design because a strong electric field will be formed around the sharp corner, which would radiate the electromagnetic wave into the surroundings. Consequently, a sharp corner of a runner becomes a source of interference and attenuation of the signal.
7. A minimum spacing must be maintained between two runners in parallel, or between a runner and the ground surface, in such a way that the ratio of the spacing to the width of runner must be 3 or above.
8. The calculated value of  $Z_C$  and the inductance when the operating frequency is 10 GHz is shown in Figure 9.P.2.



**Figure 9.P.2.** Calculation of additional reactance due to the uneven layout ( $f = 10 \text{ GHz}$ ). (a) An even runner:  $W_0 = 6 \mu\text{m}$ ,  $Z_0 = 50.2 \Omega$ ,  $l = 100 \mu\text{m}$ . (b) An uneven runner:  $W_1 = 2 \mu\text{m}$ ,  $Z_1 = 82.42 \Omega$ ,  $l/2 = 50 \mu\text{m}$ ;  $W_0 = 6 \mu\text{m}$ ,  $Z_0 = 50.17 \Omega$ ,  $l/2 = 50 \mu\text{m}$ .

9. The equivalent circuit of a via is shown in Figure 9.P.3:



**Figure 9.P.3.** Proposed via hole equivalent circuit.

10. In order to reduce the via's impedance in an RFIC design, a single via should be replaced by multiple vias.

If the multiple =  $n$ , then

$$L_{\text{via}} \rightarrow L_{\text{via}}/n;$$

$$r_{\text{via}} \rightarrow r_{\text{via}}/n;$$

$$C_{\text{via}} \rightarrow nC_{\text{via}}.$$

11. In order to reduce the via's impedance down in a PCB design, one can enlarge the radius and reduce the height of the via.
12. In RFIC layout, one of the important items is to layout a large device not in a "dragon" shape but in a square shape if the device has tens or hundreds of fingers.
13. In RFIC layout, in order to reduce the tolerance down, a resistor must be replaced by a bunch of resistors with a higher value because the tolerance of a resistor is very high, which could be  $\pm 20\%$  at the present IC technology level.
14. It is not advisable to fill all the free space to function as part of ground surface on a layout. It may increase the coplanar capacitance significantly.

# MANUFACTURABILITY OF PRODUCT DESIGN

## 10.1 INTRODUCTION

The development process of a product can probably be divided into two stages: R&D and production. Let us take a look what happens when a product is developed in this way.

- In the R&D stage, the main design steps, including simulation, layout, and implementation of module or IC tape-out, are completed based on the product specifications.
- Some samples based on the prototype design are fabricated by hand.
- Testing of these samples is carried out.
- Patents are awarded and papers published.
- Product is put on the production line for mass production.
- However, the yield rate of the product is very poor, say, less than 40%. More than 60% of the funds for the direct and indirect material costs, including manpower and the manufacturing facility, is wasted.
- People from R&D and manufactory scream, shout, and blame each other for the product failure.
- Result: the company goes bankrupt.

What is going on?

The failure of this product is due to the lack of tolerance analysis prior to being put on the production line. To ensure the manufacturability of a product, tolerance analysis in the simulation stage must be carried out. Obviously, this is a very important step to guarantee the success of a new product.

The unique criterion to measure the success or failure of product development is the manufacturability of a product. What is the manufacturability of a product? Manufacturability of a product means that the product can be manufactured in a massive production line and retain the features of

1. High yield rate, including
  - meeting the specifications,
  - good repetition or identity of product,
  - high reliability;
2. Low cost, including the cost of
  - materials and parts,
  - manpower,
  - factory maintenance and equipment.

Therefore, the answer is simple: the yield rate of the product in the production line is higher than the expected goal. The definition of the expected goal, of course, is somewhat artificial but by summing up all the features of manufacturability there is a bottom line: the product must be profitable.

In the disastrous example above, not only is the tolerance analysis missing, but also are some other important segments of the design process, such as ALT (accelerated life testing), TOP (technical operating production), and pilot production. The typical design procedures for a new product are shown in Figure 10.1.

Instead of the two stages in the example, Figure 10.1 shows that three stages are required for a new product design to reach an acceptable manufacturability, which are as follows:

First stage: R&D

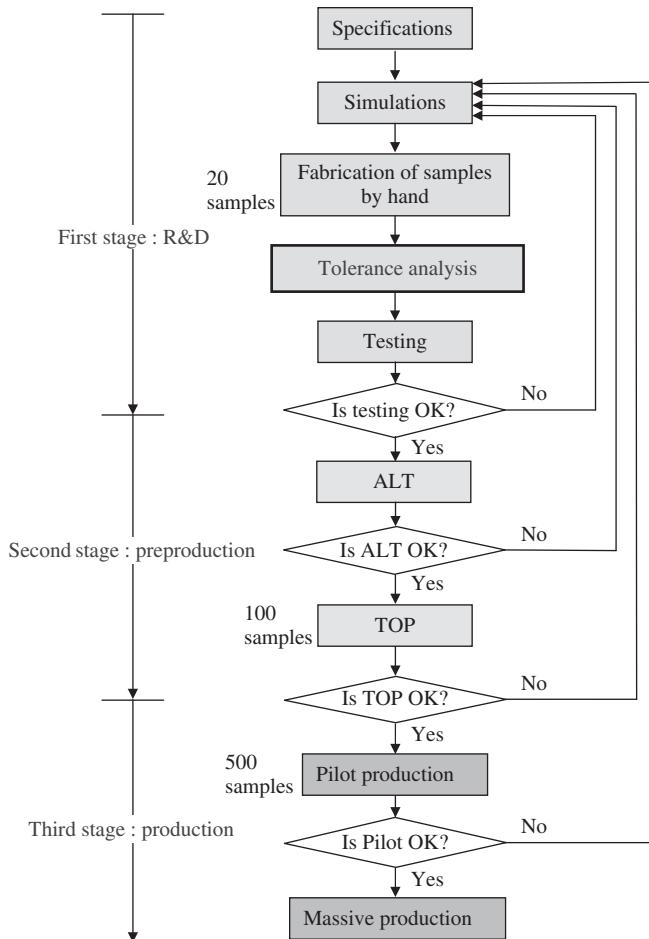
Second stage: preproduction

Third stage: production.

In order to ensure the manufacturability of a new product, the designer must be a qualified  $6\sigma$  engineer. Tolerance analysis is the key component in  $6\sigma$  design.

At present, in the simulation stage, tolerance analysis is usually conducted through the so-called corner analysis, in which tolerance analysis is done only for the devices but not for other parts such as resistors, capacitors, inductors, transformers, and so on. This may be fine in digital circuit design because in digital circuits the devices may play the decisive role in the contribution to tolerance. However, in RF circuit design, the devices and the other parts contribute to the tolerance with the same weight. Sometimes the tolerance contributed by other parts may even dominate over those of the devices. Conducting tolerance analysis for devices alone is not enough in a  $6\sigma$  design.

One of the main reasons why the yield rate in a production line is poor is that a high percentage of RF circuit designs do not employ tolerance analysis, or they confine tolerance analysis to the devices only. Often, tolerance analysis is ignored until the low yield rate shows up. In electrical engineering courses, tolerance analysis is often set aside or ignored so that new engineers fresh from school are somewhat green on it.



**Figure 10.1.** Typical design procedures of a new product with acceptable manufacturability.

## 10.2 IMPLICATION OF $6\sigma$ DESIGN

### 10.2.1 $6\sigma$ and Yield Rate

Appendix 10.A.1 introduces random variables and the fundamentals of random processes. For a random variable  $C$ , the ratio of relative deviation from the average  $m$  with respect to the standard deviation  $\sigma$  is defined as

$$z = \frac{C - m}{\sigma}, \quad (10.1)$$

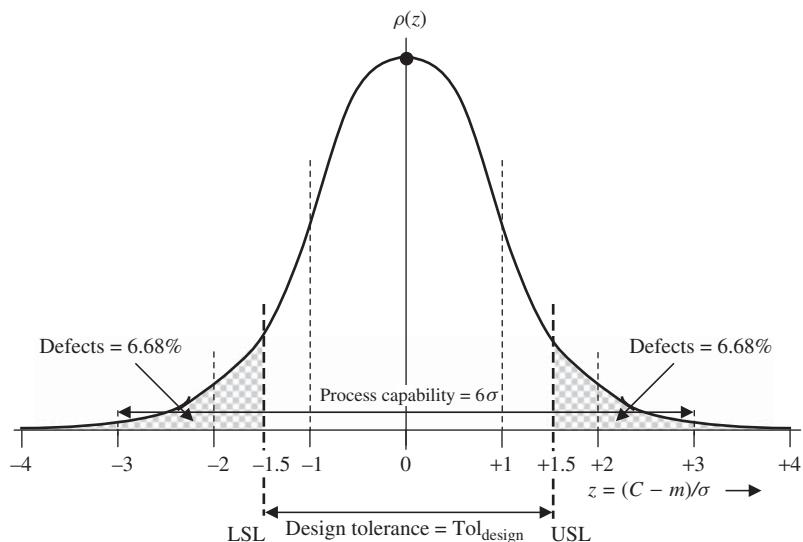
where

$z$  = the ratio of relative deviation from the average  $m$  with respect to the standard deviation  $\sigma$ ,

$C$  = a random variable,

$m$  = the average value of random variable, and

$\sigma$  = the variance or standard deviation or average square root of deviation.



**Figure 10.2.** Definition of design tolerance  $Tol_{\text{design}}$  and process capability  $6\sigma$ .

Figure 10.2 plots a random variable with a Gaussian distribution. The ordinate is the probability density  $\rho(z)$  and the abscissas is  $z$ , the ratio of the relative deviation from the average  $m$  with respect to the standard deviation  $\sigma$ .

The design tolerance  $Tol_{\text{design}}$  is defined as

$$Tol_{\text{design}} = USL - LSL, \quad (10.2)$$

where

$USL$  = the upper specification limit,  $\sigma$ ,

$LSL$  = the lower specification limit,  $\sigma$ .

The magnitudes of  $USL$  and  $LSL$  are the same but  $LSL$  is negative, whereas  $USL$  is positive because the Gaussian distribution is symmetrical with respect to its average value  $m$ . The design tolerance  $Tol_{\text{design}}$  represents the design capability of a circuit. In Figure 10.2,

$$USL = 1.5\sigma, \quad (10.3)$$

and

$$LSL = -1.5\sigma, \quad (10.4)$$

so that

$$Tol_{\text{design}} = 3\sigma. \quad (10.5)$$

The upper and lower specification limits,  $USL$  and  $LSL$ , and hence the design tolerance  $Tol_{\text{design}}$ , are specified by the designer. The shaded area in Figure 10.2 represents

the defects of the product. In the case shown in Figure 10.2,

$$\text{Defects} = 1 - f(z = \text{USL}) - f(z = \text{LSL}) = 1 - 43.32\% - 43.32\% = 13.36\%, \quad (10.6)$$

because from Figure 10.A.4,

$$f(z = \text{USL}) = f(z = 1.5) = 19.15\% + 14.98\% + 9.19\% = 43.32\%, \quad (10.7)$$

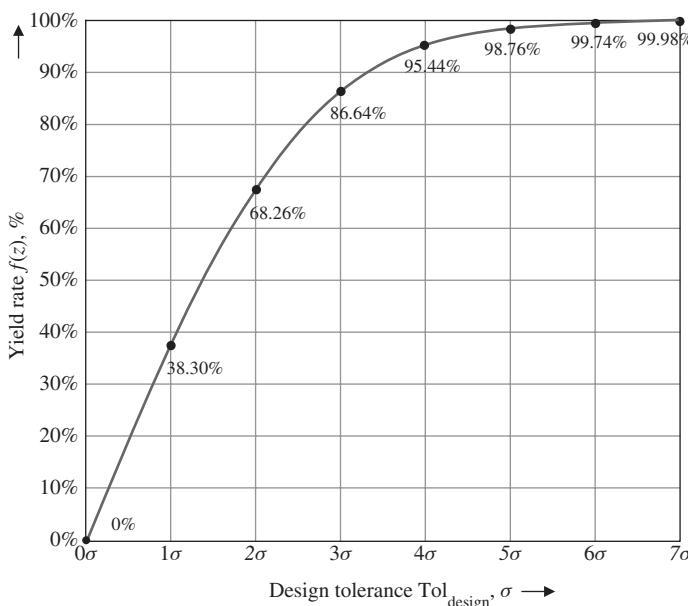
$$f(z = \text{LSL}) = f(z = -1.5) = 19.15\% + 14.98\% + 9.19\% = 43.32\%. \quad (10.8)$$

Obviously, if the design tolerance  $\text{Tol}_{\text{design}}$  is high, the defects will be low. Therefore, the design tolerance  $\text{Tol}_{\text{design}}$  is a measure of the design capability of a designer. The expected high value or goal of  $\text{Tol}_{\text{design}}$  is called the *process capability*, that is,

$$\text{Process capability} = 6\sigma. \quad (10.9)$$

Based on Figure 10.A.4, Figure 10.3 plots the relationship between the yield rate  $f(z)$  and the design tolerance  $\text{Tol}_{\text{design}}$ , by which the yield rate can be directly found from the value of design tolerance  $\text{Tol}_{\text{design}}$ . The design tolerance  $\text{Tol}_{\text{design}}$  is usually measured in units of  $\sigma$ . This is why people talk about a product with “how many  $\sigma$  design.” As a matter of fact, the relation between the yield rate  $f(z)$  and the design tolerance  $\text{Tol}_{\text{design}}$  in Figure 10.3 can also be listed as given in Table 10.1.

Figure 10.3 or Table 10.1 shows the increase of the yield rate as the design tolerance  $\text{Tol}_{\text{design}}$  is increased. The yield rate is quite sensitive to the design tolerance  $\text{Tol}_{\text{design}}$  when the latter is much less than  $6\sigma$ . A higher yield rate can be obtained if a higher  $\text{Tol}_{\text{design}}$  is allowed. The yield rate reaches **99.74%** if the design tolerance  $\text{Tol}_{\text{design}}$  is increased to  $6\sigma$ . In such a case, only 0.26% of the products are defective. However, if further effort is put on the design so that the design tolerance  $\text{Tol}_{\text{design}}$  is continuously



**Figure 10.3.** Relationship between yield rate  $f(z)$  and design tolerance  $\text{Tol}_{\text{design}}$ .

TABLE 10.1. Relationship between the Yield Rate  $f(z)$  and the Design Tolerance  $Tol_{\text{design}}$

Design tolerance $Tol_{\text{design}}, \sigma$	Yield rate $f(z), \%$
0	0
1	38.30
2	68.26
3	86.64
4	95.44
5	98.76
6	99.74
7	99.98

increased from  $6\sigma$  up to  $7\sigma$  as seen in Figure 10.3, the increase of the yield rate is very little. This is why the  $6\sigma$  design tolerance  $Tol_{\text{design}}$  is ruled as an expected high design goal but not  $7\sigma$ ,  $8\sigma$ ,  $9\sigma$ , or higher.

Also, it can be seen that the standard deviation of product  $\sigma$  is a key parameter in the  $6\sigma$  circuit design.

For example, a hypothetical customer is going to buy a capacitor specified as  $1 \pm 0.1$  pF. Two designers are in charge of the design works in the factory. Designer A designs the 1-pF capacitor with  $\sigma = 0.1$  pF, whereas designer B designs the 1-pF capacitor with  $\sigma = 0.2$  pF. The specification from the customer indicates that he wants to buy  $2\sigma$  products designed by designer A and  $1\sigma$  products designed by designer B. Consequently, the manufacturer can provide the customer with 68.26% of the mass-produced products designed by designer A and only 38.30% of the products designed by designer B. Obviously, designer A makes more profit for the factory.

Who is the  $6\sigma$  designer in respect to this customer's specification? A designer who can design the 1-pF capacitor with  $\sigma = 0.033$  pF is the  $6\sigma$  designer, because his  $6\sigma$  range  $-3\sigma = -0.1\text{pF} < z\sigma < +3\sigma = +0.1$  pF is what the customer asks for. The manufacturer can provide the customer with 99.74% of the mass-produced products designed by this designer.

### 10.2.2 $6\sigma$ Design for a Circuit Block

In Section 10.2.1, the discussion was confined to a single part. Now, we expand the discussion to a circuit block design.

A  $6\sigma$  design for a circuit block is such that the yield rate of this circuit block in mass production reaches 99.74% when  $6\sigma$  tolerance of all the parts applied in the circuit block is allowed.

Actually, the expected yield rate is 100% (actually 99.74% but close to 100%).

Similarly, referring to Figure 10.3, we may conclude the following:

- A  $5\sigma$  circuit design is such that the yield rate of the circuit block in mass production reaches 98.76% when  $5\sigma$  tolerance of all the parts applied in the circuit block is allowed.
- A  $4\sigma$  circuit design is such that the yield rate of the circuit block in mass production reaches 95.44% when  $4\sigma$  tolerance of all the parts applied in the circuit block is allowed.

- A  $3\sigma$  circuit design is such that the yield rate of the circuit block in mass production reaches 86.64% when  $3\sigma$  tolerance of all the parts applied in the circuit block is allowed.
- A  $2\sigma$  circuit design is such that the yield rate of the circuit block in mass production reaches 68.26% when  $2\sigma$  tolerance of all the parts applied in the circuit block is allowed.
- A  $1\sigma$  circuit design is such that the yield rate of the circuit block in mass production reaches 38.30% when  $1\sigma$  tolerance of all the parts applied in the circuit block is allowed.

The value of  $\sigma$  in the  $6\sigma$  design for a circuit block is contributed by the individual  $\sigma$  value of all the parts in the circuit block. It depends not only on the individual  $\sigma$  values but also on the weight of individual parts in the circuit block. Usually, it is hard to calculate this by hand. Monte Carlo analysis in most circuit simulation programs provides a powerful tool to obtain the value of  $\sigma$  and other parameters for a circuit block, which will be introduced below.

### 10.2.3 $6\sigma$ Design for a Circuit System

A  $6\sigma$  design for a circuit system is such that the yield rate of this circuit system in mass production reaches 99.74% when  $6\sigma$  tolerance of all the parts applied in the circuit system is allowed.

The other sentences used in Section 10.2.2 are also applicable to this section as long as the word “block” is replaced by the word “system.”

## 10.3 APPROACHING $6\sigma$ DESIGN

### 10.3.1 By Changing of Parts' $\sigma$ Value

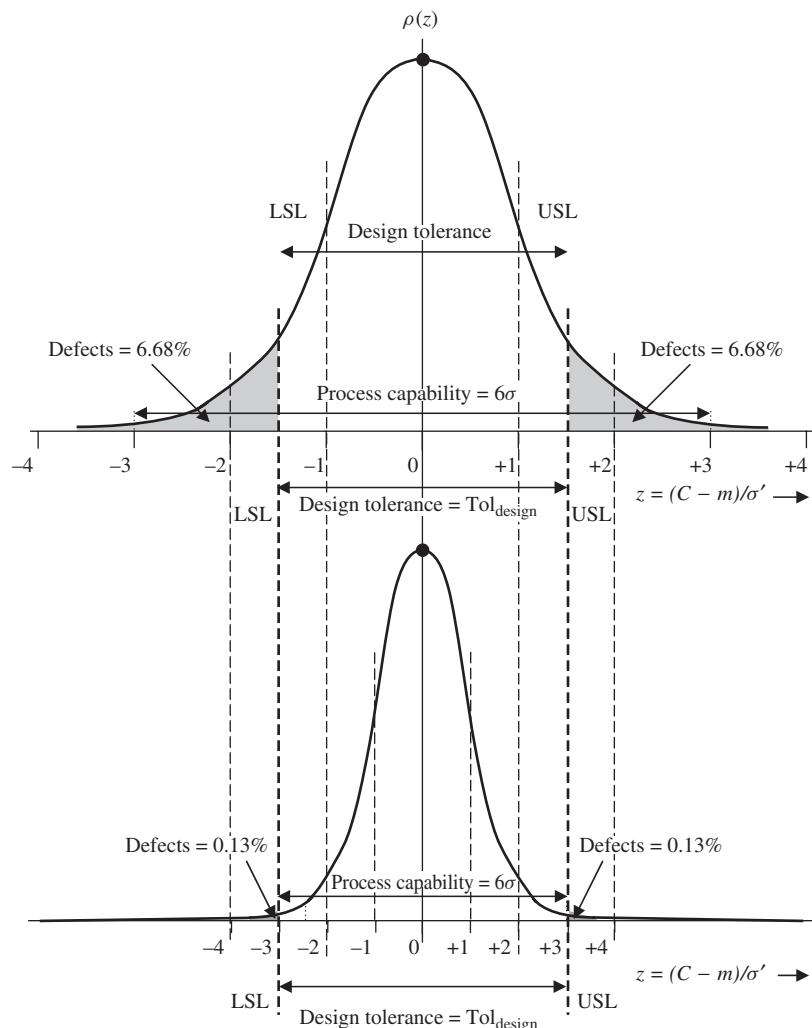
The  $6\sigma$  design is a tough task for a circuit design engineer because the design

- must ensure that the circuit block is in normal performance or that all the goals of performance are satisfied;
- must allow a  $6\sigma$  deviation for all the parts of the circuit block, such as resistors, capacitors, inductors, transistors, and so on.

It is possible to approach a  $6\sigma$  design by changing the part's  $\sigma$  value. Figure 10.4 illustrates such an approach.

The original design is depicted in the upper portion of Figure 10.4. By applying the values of  $USL = 1.5\sigma$  and  $LSL = -1.5\sigma$  as shown in the upper portion of Figure 10.4 into expression (10.2), the design tolerance  $Tol_{design}$  of  $3\sigma$  is found so that the original design is a  $3\sigma$  design. The defects of the product are  $2 \times 6.68\% = 13.36\%$ , and the yield rate is 86.64%.

With the same specified values of  $USL$  and  $LSL$ , the original  $3\sigma$  design in the upper portion of Figure 10.4 can be improved and converted to  $6\sigma$  design as shown in the bottom portion of Figure 10.4. Such a conversion can be realized by replacing the original part by a new part if the average value  $m$  of the part is kept unchanged



**Figure 10.4.** Conversion of  $3\sigma$  design to  $6\sigma$  design by 50% reduction of standard deviation from  $\sigma$  to  $\sigma'$ , that is,  $\sigma' = \sigma/2$ .

and if the standard deviation of the new part is lower than that of the original part, that is,

$$\sigma' = \frac{\sigma}{2}, \quad (10.10)$$

where

$\sigma'$  = the standard deviation of new part and  
 $\sigma$  = the standard deviation of original part.

Consequently, the original defect rate of 13.36% would be reduced to 0.26%, and the yield rate increased from 86.64% to 99.74%.

The cost of this increased yield rate is price, because parts with low standard deviation are more expensive than parts with high standard deviation. Expression (10.10)

indicates that the standard deviation of the new parts is lower than that of the original parts.

As a matter of fact, the relation (10.10) is specific to the conversion from  $3\sigma$  to  $6\sigma$  circuit design or for the conversion from the yield rate 86.64% to 99.74%. In general, by changing the standard deviation of the parts, a circuit design can be converted from any yield rate to another. If the conversion is made from a low to a high yield rate, the cost of the higher yield rate is the price.

### 10.3.2 By Replacing Single Part with Multiple Parts

In the real parts, such as the resistor, capacitor, and inductor, the  $\sigma$  value decreases as the value of parts increases if they are produced with the same processing steps. By utilizing this attribute, we have another way to approach the  $6\sigma$  design either for a discrete RF module or for an RFIC circuit. That is, we replace the original part with multiple parts whose value adds up to the higher part's value.

Let us illustrate this idea by an example. Assume that a circuit block consists of three parts. They are listed in the second column of Table 10.2.

In order to convert the circuit design from a low-sigma circuit design to a high-sigma circuit design, or from a low yield rate to a high yield rate, without changing the parts' standard deviation as discussed in Section 10.3.1, the original parts are replaced by multiple parts whose values add up to the higher part's value. The resultant value of multiple parts is kept the same as the original part's value through a reasonable combined configuration either in parallel or in series as shown in Figure 10.5. The resultant standard deviation of the multiple parts is lower than that of the original part because the multiple parts have the higher part's value and therefore have lower standard deviation. The multiple parts are listed in the third column of Table 10.2.

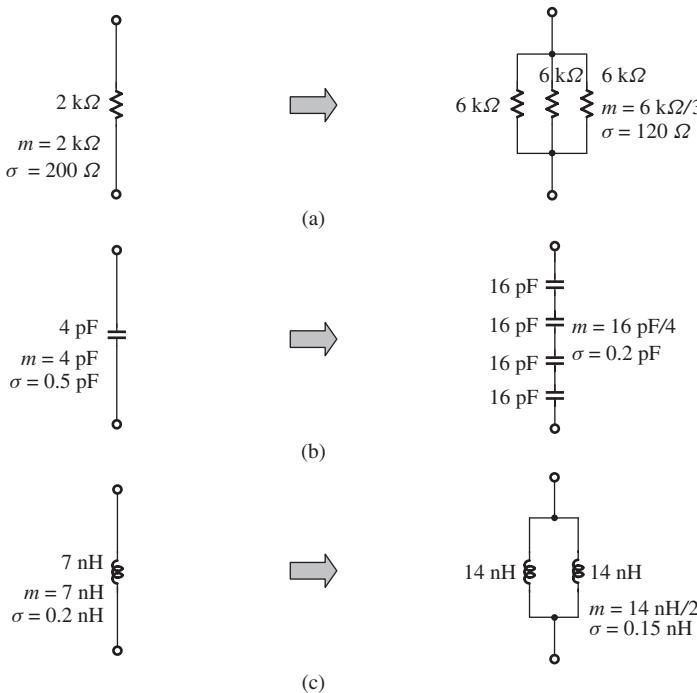
In Table 10.2,

- the original 2-k $\Omega$  resistor with  $\sigma = 200 \Omega$  is replaced by three 6-k $\Omega$  resistors with  $\sigma' = 120 \Omega$  in parallel;
- the original 4-pF capacitor with  $\sigma = 0.5 \text{ pF}$  is replaced by four 16-pF capacitors with  $\sigma' = 0.2 \text{ pF}$  in series;
- the original 7-nH inductor with  $\sigma = 0.2 \text{ nH}$  is replaced by two 14-nH inductors with  $\sigma' = 0.15 \text{ nH}$  in parallel.

The yield rate in the circuit block built either by the original parts or by multiple parts depends on the topology of the circuit block and all the parts in the circuit block. Unfortunately, the data listed in Table 10.2 is not enough to calculate the yield rate in the

TABLE 10.2. Replacement of Original One Resistor, One Capacitor, and One Inductor by Three Resistors, Four Capacitors, and Two Inductors

	Original part		Multiparts	
	$m$	$\sigma$	$m$	$\sigma'$
Resistor	2 k $\Omega$ (one resistor)	200 $\Omega$	6 k $\Omega/3$ (three resistors in parallel)	120 $\Omega$
Capacitor	4 pF (one capacitor)	0.5 pF	16 pF/4 (four capacitors in series)	0.2 pF
Inductor	7 nH (one inductor)	0.2 nH	14 nH/2 (two inductors in parallel)	0.15 nH



**Figure 10.5.** Replacement of one part by multiple parts. (a) One resistor of  $2\text{ k}\Omega$  is replaced by three  $6\text{ k}\Omega$  resistors in parallel. (b) One capacitor of  $4\text{ pF}$  is replaced by four  $16\text{ pF}$  capacitors in series. (c) One inductor of  $7\text{ nH}$  is replaced by two  $14\text{ nH}$  inductors in parallel.

circuit block built either by the original parts or by multiple parts. However, the yield rate in the circuit block built by multiple parts with the higher part's value is undoubtedly higher than that of the circuit block built by the original parts.

## 10.4 MONTE CARLO ANALYSIS

The tolerance evaluation of a circuit block can be conducted by a Monte Carlo analysis. Let us take a BPF (band-pass filter) as an example to illustrate the key features of this analysis.

### 10.4.1 A Band-Pass Filter

Figure 10.6 shows the schematic of a practical BPF in the frequency range of 403–470 MHz. The parts, including capacitors and inductors, are the chip parts manufactured by MuRata. The corresponding models for capacitors and inductors are applied in all the simulations.

Figure 10.7 shows the frequency response of the BPF in the frequency range from 200 to 700 MHz. Its insertion loss, IL, is

$$\text{IL} = -1.05\ \text{dB}, \quad (10.11)$$

when

$$403 < f < 470\ \text{MHz}, \quad (10.12)$$

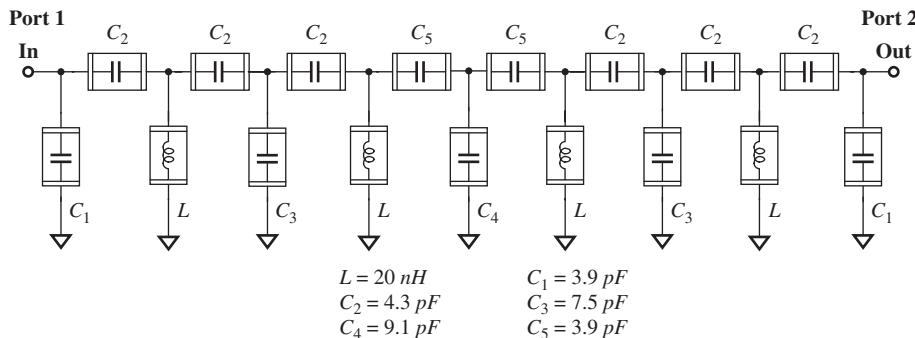


Figure 10.6. Band-pass filter, 403–470 MHz.

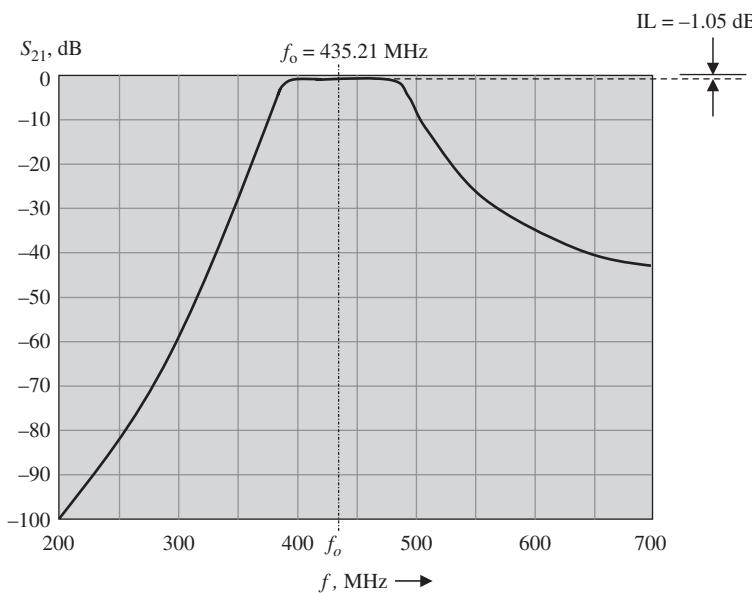


Figure 10.7. Frequency response of a band-pass filter, 403–470 MHz.

and

$$f_0 = \sqrt{403 \times 470} = 435.21 \text{ MHz.} \quad (10.13)$$

Figure 10.8 shows the return loss  $S_{11}$  on the Smith Chart, which indicates that the input impedance of the BPF is well matched to  $50 \Omega$  in the frequency range from 403 to 470 MHz.

### 10.4.2 Simulation with Monte Carlo Analysis

Figure 10.9 shows a simulation page for the Monte Carlo analysis.

A schematic of the BPF as shown in Figure 10.6 is represented by the circuit block labeled “BPF (band-pass filter), 403–470 MHz.”

The purpose of the Monte Carlo analysis is to find the values of the parts’ tolerance for approaching  $6\sigma$  design. In Figure 10.6, there are six different parts: one type of inductor and five kinds of capacitors; they are all involved in the tolerance analysis.

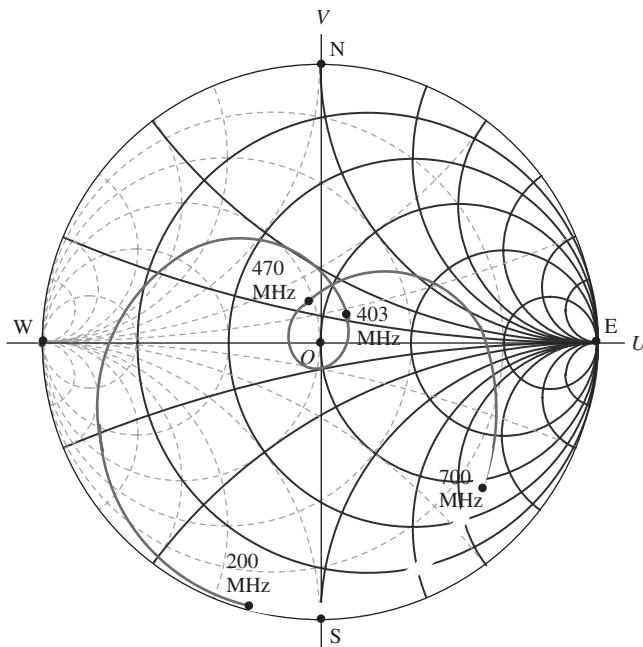
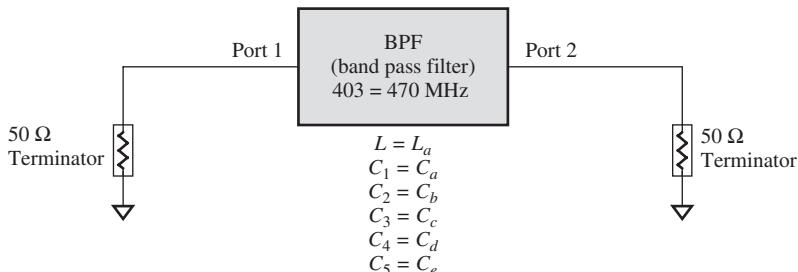


Figure 10.8.  $S_{11}$  of the band-pass filter,  
403–470 MHz.



Equation  $L_a$  = Randvar Gaussian, 20 nH  $\pm$  7%

Equation  $C_a$  = Randvar Gaussian, 3.9 pF  $\pm$  5%

Equation  $C_b$  = Randvar Gaussian, 4.3 pF  $\pm$  5%

Equation  $C_c$  = Randvar Gaussian, 7.5 pF  $\pm$  5%

Equation  $C_d$  = Randvar Gaussian, 9.1 pF  $\pm$  5%

Equation  $C_e$  = Randvar Gaussian, 3.9 pF  $\pm$  5%

Figure 10.9. Simulation page for Monte Carlo analysis.

The first objective in the simulation with Monte Carlo analysis is to specify

- all the values of different parts : random variables,
- distribution type of the variables : Gaussian distribution,
- their nominal value or mean  $m$ , and
- tolerance of each part: Tol (by percentage).

The nominal value or mean,  $m$ , and tolerance by percentage, Tol, of the parts are provided by the manufacturers.

In the Monte Carlo analysis, the second objective is to ask for the iteration number in sampling the random variables. The iteration number must be set large enough so that the simulated results can reflect all the possible variations due to the parts' tolerance. However, too many iterations may increase simulation time to an unacceptable level. In a practical simulation, the number of iterations is usually set to

$$\text{Iterations} \Rightarrow 50 - 100. \quad (10.14)$$

Figures 10.10 and 10.11 show the frequency response and the input return loss of the BPF, respectively. For clarity, Figure 10.10 contains only 10 iterations, that is, it contains only 10 curves of the frequency response.

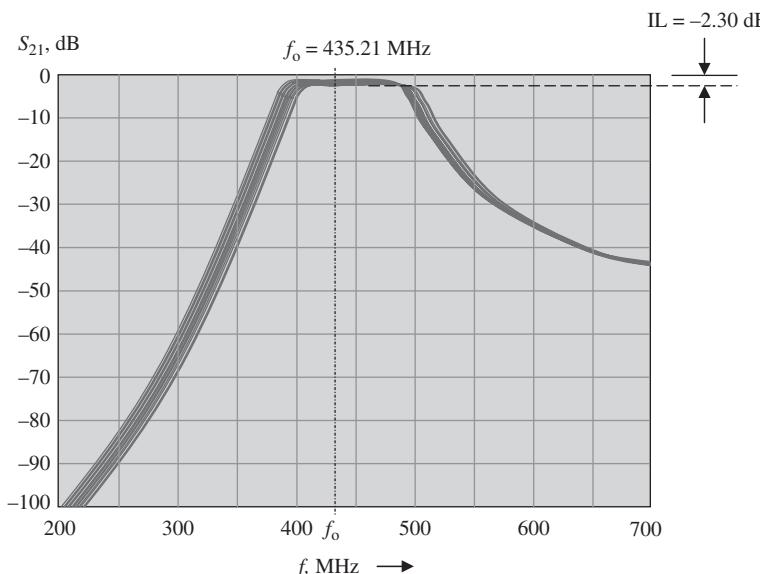
From Figures 10.10 and 10.11, it can be seen that the curve of the frequency response is no longer a clear-cut curve as shown in Figures 10.7 and 10.8. Instead, the curve spreads out over a small area in the direction of frequency or  $S_{21}$  due to the parts' tolerance.

As a BPF, the insertion loss is one of the important parameters to be concerned with. By comparing Figures 10.7 and 10.10, it is found that the insertion loss is degraded from  $-1.05$  to  $-2.30$  dB as a result of the parts' tolerance. Obviously, in order to ensure 100% yield rate for this BPF in the production line, the insertion loss must be announced as

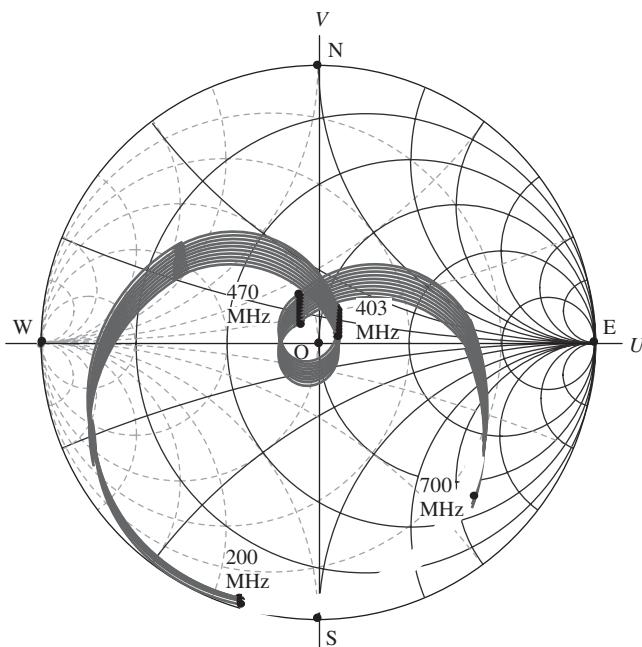
$$|IL| < 2.5 \text{ dB}, \quad (10.15)$$

where 2.5 dB is higher than 2.3 dB as read from Figure 10.10. The 0.2 dB difference is a buffer for inaccuracies in reading or testing.

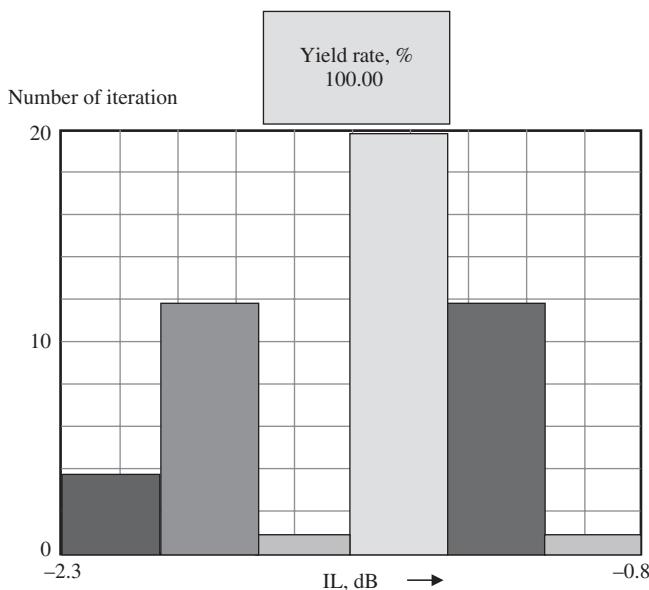
On the other hand, the simulation with Monte Carlo analysis can provide the yield rate and histogram display as shown in Figure 10.12. On the top, it shows that the yield rate is 100% if the goal of insertion loss is consistent with expression (10.15). The total number of iterations in the simulation is 50. The histogram shows the distribution of the iteration number and is listed in Table 10.3.



**Figure 10.10.** Frequency response of band-pass filter, 403–470 MHz.  
Number of iterations = 50.



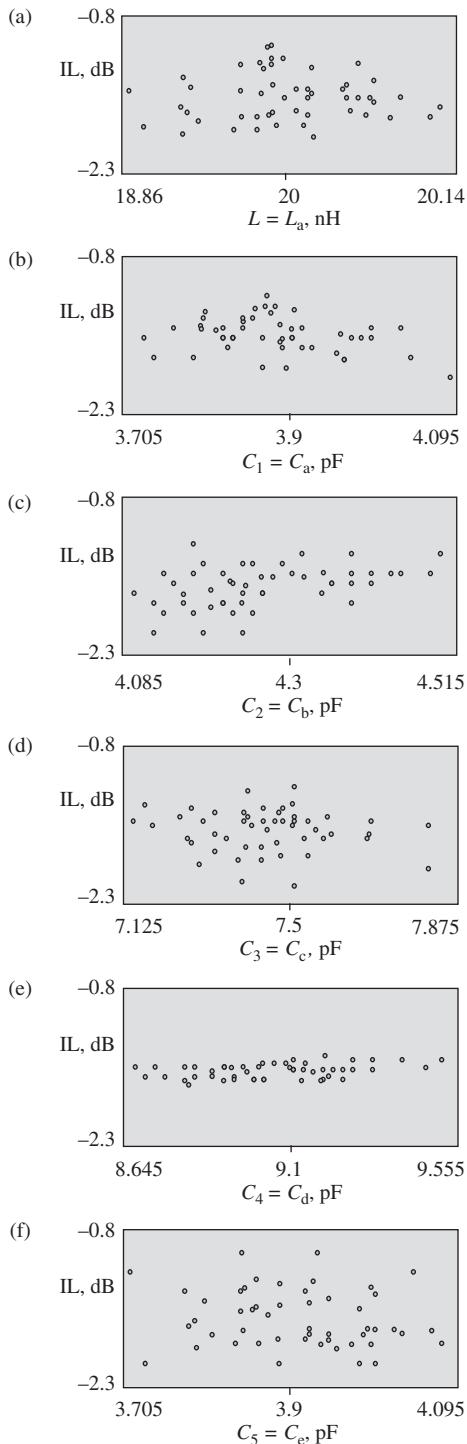
**Figure 10.11.**  $S_{11}$  of band-pass filter, 403–470 MHz. Number of iterations = 50.



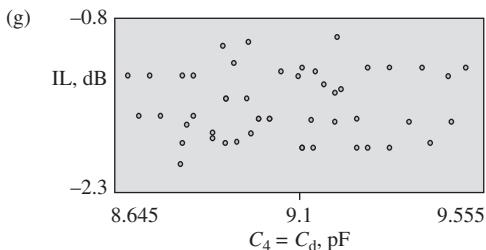
**Figure 10.12.** Display of insertion loss histogram and yield rate for  $|IL| < 2.5$  dB (iteration number = 50).

**TABLE 10.3.** Distribution of Iteration Number versus Insertion Loss

Insertion loss, dB	-2.30 to -2.05	-2.05 to -1.80	-1.80 to -1.55	-1.55 to -1.30	-1.30 to -1.05	-1.05 to -0.8
Number in iteration	4	12	1	20	12	1



**Figure 10.13.** Sensitivity of individual part's value on the insertion loss. (a) IL versus  $L$ : tolerance of  $L$  is 7%. (b) IL versus  $C_1$ : tolerance of  $C_1$  is 5%. (c) IL versus  $C_2$ : tolerance of  $C_2$  is 5%. (d) IL versus  $C_3$ : tolerance of  $C_3$  is 5%. (e) IL versus  $C_4$ : tolerance of  $C_4$  is 5%. (f) IL versus  $C_5$ : tolerance of  $C_5$  is 5%.



**Figure 10.14.** Sensitivity of  $C_4$  on the insertion loss after its tolerance of 5% is replaced by 7%.

### 10.4.3 Sensitivity of Parts on the Parameter of Performance

Another special function of the Monte Carlo analysis is to examine the sensitivity of each part on the parameters of performance. Figure 10.13(a)–(f) shows the sensitivity of the six individual parts,  $L$ ,  $C_1$ ,  $C_2$ ,  $C_3$ ,  $C_4$ , and  $C_5$ , on the insertion loss, respectively. The insertion loss is less than 2.5 dB. Fifty iterations are conducted in the analysis so that there are 50 points on the plot. Each point therefore represents one random value of the part in each iteration.

In the plots of Figure 10.13(a)–(d) and (f) for IL versus  $L$ ,  $C_1$ ,  $C_2$ ,  $C_3$ , and  $C_5$ , the tolerance of these five parts looks quite adequate. All the points are scattered within the range of the IL in an approximately homogeneous pattern. However, in the plot of Figure 10.13(e) for IL versus  $C_4$ , the tolerance of  $C_4$  seems somewhat smaller than what is needed because all the points are crowded in the middle area within the IL range. In other words, the tolerance of  $C_4$  can be relaxed somewhat, say, from the original 5% as shown in Figure 10.13 to 7%. After a resimulation with Monte Carlo analysis, replacing the 5% tolerance by 7% for  $C_4$ , Figure 10.13(e) is changed as shown in Figure 10.14.

As seen in Figure 10.14, all the points are now scattered within the range of the IL in an approximately homogeneous pattern. This implies that the replacement of 5% tolerance by 7% tolerance is adequate. This replacement is valuable because the  $C_4$  capacitors with 7% tolerance are much cheaper than those with 5% tolerance. The cost therefore can be reduced by examining the sensitivity of the parts.

Similarly, in addition to the insertion loss, the Monte Carlo analysis can be conducted for other parameters so as to examine all the parts' sensitivity and finally to ensure 100% yield rate. Therefore, Monte Carlo analysis is a powerful tool to realize the  $6\sigma$  circuit design for the mass production of the product.

The BPF example introduced above is a simple example. Based on the same principle, readers can expand the Monte Carlo analysis to more complicated circuit blocks, including transistors, microstrip lines, and so on.

## APPENDICES

### 10.A.1 Fundamentals of Random Process

Both regular processes and random processes exist in the world. They also exist in circuit design. The fabrication of the prototype samples in the first design stage is a regular process, while the mass production in the third design stage is a random process. For example, if the designed product is a 1-pF capacitor, then in the first design stage, after the simulation, a couple of capacitor samples are built and tested. The capacitance of each capacitor sample is under control and nonrandom. However, once this product is

put on the production line for large-scale production, the value of individual capacitors becomes a random variable. Some of the values are greater than 1 pF, while some are less than 1 pF. Their average value, of course, is 1 pF, that is,

$$m = \overline{\sum_i C_i} = 1 \text{ pF}, \quad (10.A.1)$$

where

- $C_i$  = the capacitance of an individual capacitor and
- $m$  = the average value of all the capacitors.

The number of capacitors with values close to the average value is greater than the number of capacitors with values much higher or lower than the average value. However, the sum of all the individual deviations from the average value would be zero because the chance of the capacitance of individual capacitor being “higher or lower than the average value” is the same, that is,

$$D_i = \Delta C_i = m - C_i, \quad (10.A.2)$$

$$\sum_i D_i = \sum_i \Delta C_i = \sum_i (m - C_i) = 0, \quad (10.A.3)$$

where  $D_i$  is the deviation of individual capacitor's value from the average value.

In order to characterize the deviation status of the random variable, a new parameter called *variance* or *standard deviation* is defined as the average value of the square root of deviation, that is,

$$\sigma = \sqrt{\overline{\sum_i \Delta C_i^2}} \text{ pF}, \quad (10.A.4)$$

where  $\sigma$  is the variance or standard deviation of the capacitor.

The standard deviation is always a positive value because every individual square root of deviation is positive.

The tolerance is the ratio of the standard deviation to the average value, that is,

$$\text{Tol} = \frac{\sqrt{\overline{\sum_i \Delta C_i^2}}}{m} = \frac{\sigma}{m}, \quad (10.A.5)$$

where Tol is relative tolerance of the capacitance.

If all the capacitors are tested for their values and are grouped with a small constant interval of capacitance, a plot of the values versus the number of capacitors can be shown as Figure 10.A.1. This is a histogram of the relative number of capacitors versus their values. It is symmetrical with respect to the point  $m$ .

Based on statistical theory, the histogram as shown in Figure 10.A.1 approaches a normal probability distribution with a probability density  $\rho(z)$ :

$$\rho(z) = \frac{e^{-\frac{z^2}{2\sigma^2}}}{\sqrt{2\pi}}, \quad (10.A.6)$$

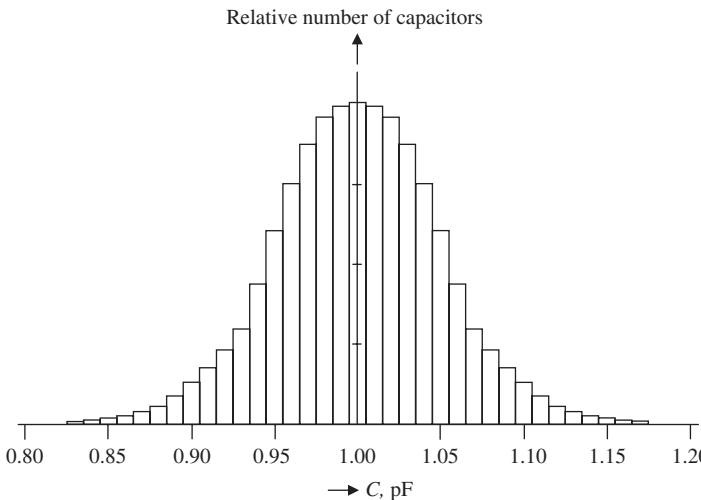


Figure 10.A.1. Histogram of relative number of capacitors versus their values.

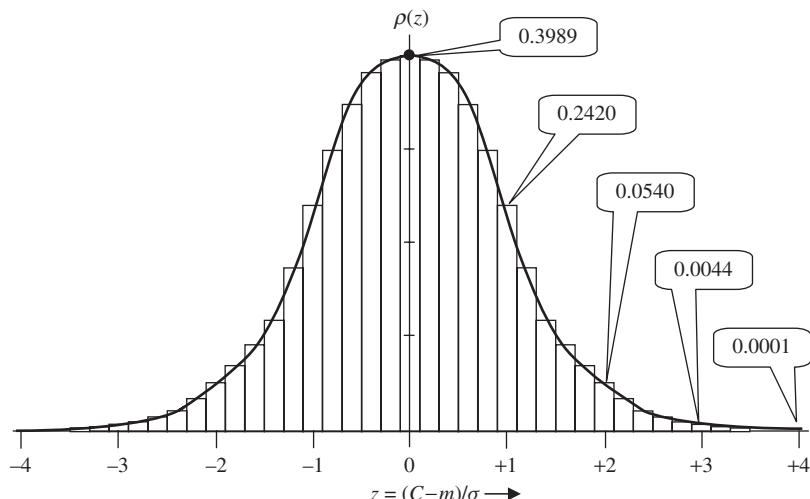


Figure 10.A.2. Distribution of the random variable  $C$  as a normal probability function.

where  $z$  is the ratio of deviation from the average to the standard deviation, say, a dimensionless capacitance, which is relative to the average value of the capacitor and is normalized by the variance  $\sigma$ .

$$z = \frac{C - m}{\sigma} \quad (10.A.7)$$

Figure 10.A.2 shows the normal probability density superimposed on the histogram of the capacitor as shown in Figure 10.A.1; the two are well matched. This verifies the assertion that the random variable  $C$  is a Gaussian or normal probability function. The abscissa scale  $z$  is a dimensionless and normalized scale, which corresponds to its normal Gaussian distribution.

A Gaussian distribution is a function symmetrical to  $z = 0$ , and it becomes a normal distribution when

$$m = 0, \quad (10.A.8)$$

and

$$\sigma = 1. \quad (10.A.9)$$

It is interesting to see how many capacitors there would be when  $z$  is varied from 0 to  $z$ , or  $C$  is varied from  $m$  to  $m + z\sigma$ . This is an integral of the function  $\rho(z)$ , integrated from  $z = 0$  to  $z = z$ , or from  $C = m$  to  $C = m + z\sigma$ , that is,

$$f(z) = \int_0^z \rho(x) dx = \int_0^z \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi}} dx. \quad (10.A.10)$$

The integral  $f(z)$  is the shadowed area shown in Figure 10.A.3.

The normal probability function is related to the error function  $\text{erf}(x)$  such that

$$f(z) = \int_0^z \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx = \frac{1}{2} \text{erf}\left(\frac{z}{\sqrt{2}}\right), \quad (10.A.11)$$

and

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy. \quad (10.A.12)$$

The value of  $f(z)$ , or the shadowed area in Figure 10.A.3, depends on the integrated interval along the abscissa. Figure 10.A.4 shows the values of  $f(z)$  in each interval of  $\sigma$ .

It can be seen that

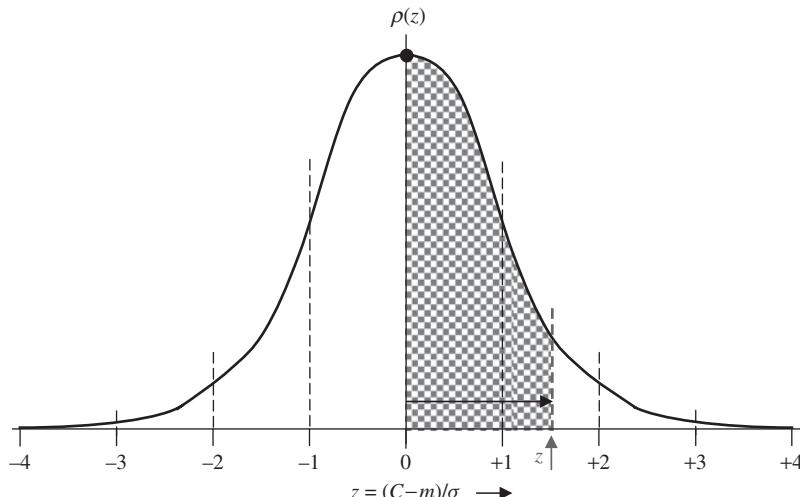
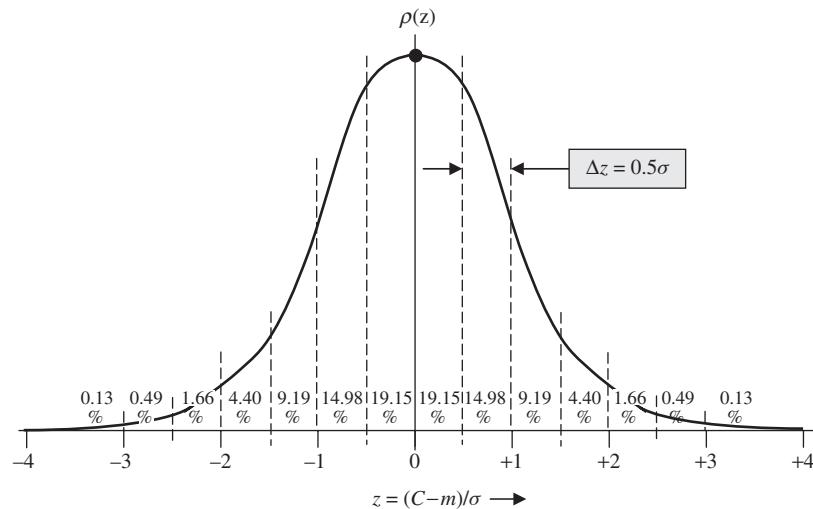
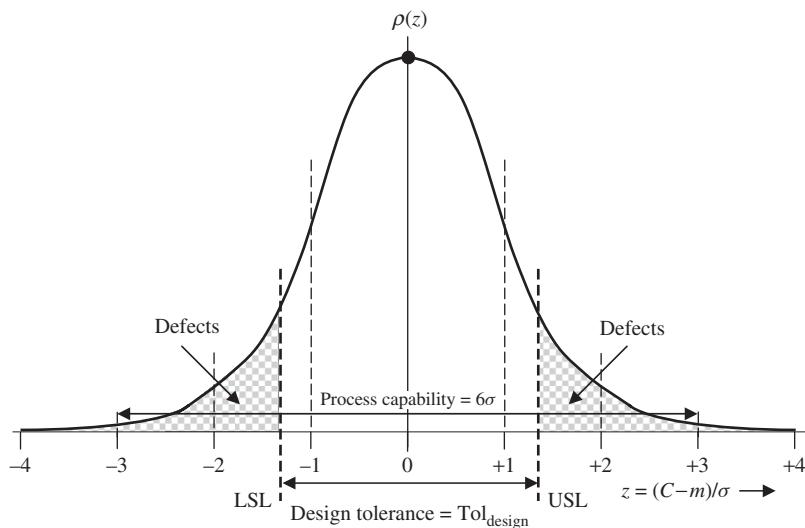


Figure 10.A.3. Integral of  $\rho(z)$  from 0 to  $z$  (shadowed area  $f(z)$ ).



**Figure 10.A.4.** Yield rate  $f(z)$ , the percentage of a random variable with a normal distribution, at each interval  $\Delta z = 0.5$ .



**Figure 10.A.5.** Definition of design tolerance  $Tol_{design}$  and process capability  $6\sigma$ .

when the interval is  $-0.0\sigma < z\sigma = (C - m) < +0.0\sigma$ , then the area is  $f(z) = 0.00\%$ ;

when the interval is  $-0.5\sigma < z\sigma = (C - m) < +0.5\sigma$ , then the area is  $f(z) = 38.30\%$ ;

when the interval is  $-1\sigma < z\sigma = (C - m) < +1\sigma$ , then the area is  $f(z) = 68.26\%$ ;

when the interval is  $-1.5\sigma < z\sigma = (C - m) < +1.5\sigma$ , then the area is  $f(z) = 86.64\%$ ;

TABLE 10.A.1. Table of the Normal Distribution

Z	0	1	2	3	4	5	6	7	8	9
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0754
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2258	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2518	0.2549
0.7	0.2580	0.2612	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2996	0.3023	0.3051	0.3078	0.3016	0.3133
0.9	0.3135	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4778	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4925	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

when the interval is  $-2\sigma < z\sigma = (C - m) < +2\sigma$ , then the area is  $f(z) = 95.44\%$ ;  
 when the interval is  $-2.5\sigma < z\sigma = (C - m) < +2.5\sigma$ , then the area is  $f(z) = 98.75\%$ ;

when the interval is  $-3\sigma < z\sigma = (C - m) < +3\sigma$ , then the area is  $f(z) = 99.74\%$ ;

when the interval is  $-3.5\sigma < z\sigma = (C - m) < +3.5\sigma$ , then the area is  $f(z) = 99.98\%$ ;

when the interval is  $z\sigma < -3\sigma$ , and  $z\sigma > +3\sigma$ , then the area is  $f(z) = 0.26\%$ .

Figure 10.A.5 shows the various terminologies of a process with normal distribution.

## 10.A.2 Index $C_p$ and $C_{pk}$ Applied in $6\sigma$ Design

**10.A.2.1 The Capability Index  $C_p$ .** In order to quantitatively describe the distance of the design tolerance from the expected high process capability  $6\sigma$ , the capability index is defined as

$$C_p = \frac{\text{Design\_Tolerance}}{\text{Process\_Capability}} \approx \frac{\text{USL} - \text{LSL}}{6\sigma}, \quad (10.A.13)$$

where

USL = the upper specification limit  $\sigma$  and

LSL = the lower specification limit  $\sigma$ .

**10.A.2.2 Adjusted Capability Index  $C_{pk}$ .** In order to quantitatively describe the deviation of the actual process mean from the nominal mean, another index, the adjusted capability index  $C_{pk}$ , is defined as

$$C_{pk} = C_p(1 - K), \quad (10.A.14)$$

where

$$K = \frac{m - m'}{(\text{USL} - \text{LSL})/2}, \quad (10.A.15)$$

where

$m$  = the nominal process mean  $\sigma$  and

$m'$  = the actual process mean  $\sigma$ .

## 10.A.3 Table of the Normal Distribution

$$f(z) = \int_0^z \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx = \frac{1}{2} \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right) \quad (10.A.16)$$

## FURTHER READING

Brayton RK, Director SW, Hachtel GD. Yield maximization and worst-case design with arbitrary statistical distributions. IEEE Trans Circ Syst 1980;27:756–764.

Juran JM. Quality control handbook. 3rd ed. New York: McGraw-Hill; 1979.

Kane VE. Process capability indices. J Qual Technol 1986;18(1):41–52.

Kim J, Plouchart J-O, Zamdmmer N. Design and manufacturability aspect of SOI CMOS RFICs. In: IEEE 2004 Custom Integrated Circuits Conference; Orlando, Florida; 2004. pp. 541–548.

- Li RC. Key issues in RF/RFIC circuit design. Beijing: Higher Education Press; 2005.
- Raghu NK. Off-Line quality control, parameter design, and the taguchi method. *J Qual Technol* 1985;17(4):176–188.
- Taguchi G. Volume 1, Experimental designs. 3rd ed. Tokyo, Japan: Maruzen Publishing Company; 1976 (In Japanese).
- Taguchi G. Volume 2, Experimental designs. 3rd ed. Tokyo, Japan: Maruzen Publishing Company; 1977 (In Japanese).
- Taguchi G. Off-Line and On-Line quality control system. In: Proceedings of International Conference on Quality Control; Tokyo, Japan; 1978.

## EXERCISES

1. What is the meaning of “manufacturability of a product”?
2. What is  $Tol_{\text{relative}}$ , the relative tolerance of a part?
3. What is  $6\sigma$  design?
4. Why do we pursue  $6\sigma$  design, but not  $7\sigma$  or “more than  $6\sigma$ ” design?
5. A company may hire a  $4\sigma$  Engineer Peter and a  $6\sigma$  Engineer Robert with \$80k and \$90k annual salary, respectively. (Note: The yield rate for  $4\sigma$  design is 86.64%, whereas the yield rate for  $6\sigma$  design is 99.74%.) Assume 1000 qualified products are sold annually, and the total cost for each product (not including Engineer’s salary) is \$100. From the company’s viewpoint, which Engineer should be hired with higher priority?
6. Yield rate of a product is increased from  $A\%$  to  $B\%$  ( $B > A$ , of course) due to the average tolerance of parts being reduced from  $m\%$  to  $n\%$  ( $m > n$ , of course). The purchasing department reports that the cost of each product set is increased from  $P$  to  $Q$  ( $Q > P$ , of course) due to tolerance reduction of the part. Assume that the total number of qualified products manufactured annually is  $S$  sets and the average selling price of each product is  $R$ . Establish an equation to calculate the change of the annual product’s profit,  $\Delta APP$ .
7. Give an example to show how to approach  $6\sigma$  design by changing of part’s  $\sigma$  value.
8. Give an example to show how to approach  $6\sigma$  design by replacing a single part with multiple parts.
9. What are the main objectives in the Monte Carlo analysis?

## ANSWERS

1. The meaning of manufacturability of a product is
  - (a) high yield rate, including
    - i. satisfaction of specifications,
    - ii. good repetition or identity of product,
    - iii. high reliability;
  - (b) Low cost, including the cost of
    - i. material and parts,
    - ii. manpower,
    - iii. manufactory maintenance and equipment.

2. The relative tolerance of a part  $Tol_{\text{relative}}$  is defined as

$$Tol_{\text{Relative}} = \frac{\sqrt{\sum_i \Delta R_i^2}}{m} = \frac{\sigma}{m} (\%)$$

where

$R_i$  = the sample value,

$m$  = the average value, and

$\sigma$  = the variance or standard deviation.

3. A  $6\sigma$  circuit design is a design such that the yield rate of this circuit block in mass production reaches 99.74% when  $6\sigma$  tolerance of all the parts applied in the circuit block is allowed. Actually, the expected yield rate is 100%, although 99.74% is close to 100%.
4. We pursue no further than  $6\sigma$  design because the yield rate for  $6\sigma$  design is already 99.74%, whereas the yield rate for  $7\sigma$  is 99.98%, which is a tiny increase of 0.14%. It is therefore considered unnecessary to pursue  $7\sigma$  or “more than  $6\sigma$ ” design.
5. If company hires Engineer Peter, the sum of his salary and the cost is

$$1000 \times \$100 / 86.64\% + \$80k = \$195.420k.$$

If company hires Engineer Robert, the sum of his salary and the cost is

$$1000 \times \$100 / 99.74\% + \$90k = \$190.261k.$$

Obviously, the company should consider hiring Engineer Robert with higher priority.

6. The equation to calculate the change of the annual product's profit,  $\Delta APP$ , is

$$\Delta APP = SR - SQ/B - (SR - SP/A) = S(P/A - Q/B).$$

7. If original work is a  $3\sigma$  design with the part's tolerance  $Tol_{\text{relative}} = \sigma/m$ , it will become a  $6\sigma$  design by changing of part's tolerance value from  $Tol_{\text{relative}} = \sigma'/m$  to  $Tol_{\text{relative}} = \sigma'/m$ , where  $\sigma' = \sigma/2$ .
8. A part with higher value usually has lower tolerance than a part with lower value. If the original work is a  $3\sigma$  design with the part's tolerance  $Tol_{\text{relative}} = \sigma/m$ , it can be approached by a  $6\sigma$  design by replacing a single part with multiple parts so that its tolerance can be changed from  $Tol_{\text{relative}} = \sigma/m$  to  $Tol'_{\text{relative}} = \sigma'/m$ , where  $\sigma' = \sigma/2$ . For example, one 1-kΩ resistor with  $\sigma/m = 10\%$  can be replaced by ten 10-kΩ resistors with  $\sigma'/m = 5\%$  in parallel, and any of other part is similarly replaced with the multiple parts.
9. The main objectives in Monte Carlo analysis are
- (a) to find the  $Tol_{\text{relative}}$  for each part applied in the circuit block;
  - (b) to find the sensitivity for each part to every goal.

# RFIC (RADIO FREQUENCY INTEGRATED CIRCUIT)

## 11.1 INTERFERENCE AND ISOLATION

### 11.1.1 Existence of Interference in Circuitry

An electrical circuit consists of basic parts such as resistors, capacitors, inductors, transistors, and so on. Each of these has its specific function in the circuitry. One should be aware of the fact that the parts' performance is different in DC and AC operational situations. When a circuit operates with a DC current flowing through it or DC voltage across it, every part appears as a resistor and is ruled by Ohm's law. When a circuit operates with AC or RF signals, every part appears as a small antenna since there is AC current flowing through it or AC voltage across it. All these alternating electromagnetic fields radiate from the part to its whole surroundings, which would disturb the desired signal at any node in the circuitry.

According to the theory of electromagnetic radiation, the interference thus produced in the low-frequency range is negligible and usually can be ignored. However, in the high-frequency range, such as the RF range, the interference due to the radiation is considerable and it does interfere with the circuit at every node in the circuitry.

It must be noted that these electromagnetic fields disturb the circuitry along two paths. One is that of the electromagnetic field radiated into the surrounding space and then returned to the circuitry. This can be categorized as an "air path." Another path is that of the electromagnetic field moving along the PCB in the RF module or along the IC substrate in the IC chip. It can be categorized as an "underground" path.

In general, as long as the circuit block is in AC or RF operation, the desired signal at any node in the circuitry is more or less disturbed by the interference from all the parts in the circuitry. Besides, interference may come from external sources, that is, outside the block. The interference might be especially strong if the adjacent block is a power amplifier or a powerful oscillator without any means of shielding.

### 11.1.2 Definition and Measurement of Isolation

Interference inevitably exists when a circuit is operated in AC or RF frequency. It is expected to be attenuated as much as possible. For instance, if the interference can be attenuated by 100 dB from its origin to the observed node in an RF module or RFIC (radiofrequency integrated circuit) die, then it becomes negligible and can usually be ignored. If the interference can be attenuated infinitely from its origin to the observed node, then the node is perfectly isolated from the origin. Therefore, isolation can be simply defined as the attenuation of the interference in magnitude, but with the opposite sign, that is,

$$\text{Isolation} = -\text{Attenuation of interference.} \quad (11.1)$$

Quantitatively, isolation between points A and B can be measured by the interference attenuation from node A to B as shown in Figure 11.1.

Assume that the power of the interference at node A is  $P_A$  and the power at node B is  $P_B$ , and both of them are expressed in decibels. The attenuation and isolation between A and B can be expressed as

$$\text{Attenuation} = 10 \log \frac{P_B}{P_A} \text{ (dB)}, \quad (11.2)$$

and

$$\text{Isolation} = 10 \log \frac{P_A}{P_B} \text{ (dB)}. \quad (11.3)$$

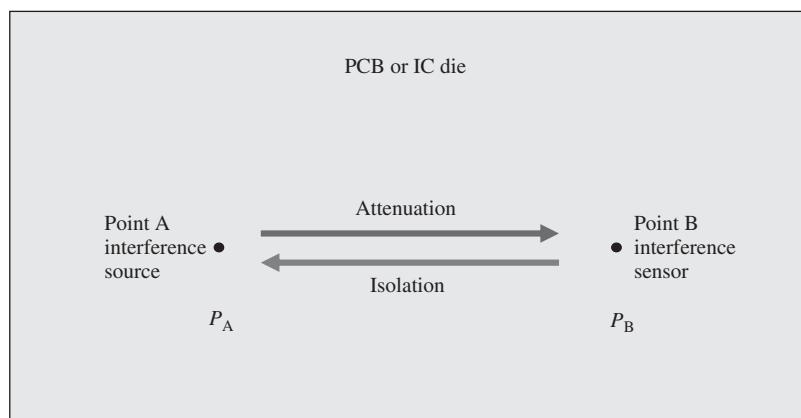


Figure 11.1. Definition of isolation between nodes A and B.

For example, if the interference is  $-10$  dB at node A and  $-30$  dB at node B, then the attenuation of interference is  $-20$  dB from node A to node B, and the isolation is  $20$  dB at node B with respect to node A.

Perfect isolation occurs when

$$P_B = 0, \quad (11.4)$$

then,

$$\text{Isolation} \rightarrow \infty. \quad (11.5)$$

### 11.1.3 Main Path of Interference in a RF Module

In RF module, the interference moving along the ‘underground’ path is insignificant because the plastic or ceramic PCB is usually a good insulator in the RF frequency range below a few gigahertz.

The main path of interference in a RF module is from the space because the dimension of the PCB is more than the dimension of a part, or is sometimes comparable with the quarter wavelength of the operating frequency. The electric lines bending into the RF module from the space could seriously disturb the performance of the circuitry. These electric lines consist of two portions. One is made up of the electric field lines radiating from local parts of the RF block. The other is made up of the field lines radiating from outside blocks.

Nothing can be done about the electric field lines radiating from the local RF block; however, those radiating from outside blocks can be shielded against. In most cases, the electric lines radiating from outside blocks are the main interference source rather than those radiating from the local block. This implies that a good shielding can usually ensure good isolation of an RF block from the outside blocks and thus ensure the performance without interference from outside RF blocks. In the era before the RFIC was developed, shielding equipment would always be a part of the RF module or RF block implemented by discrete parts. Shielding using a metallic shielding box will be introduced in Section 11.2.

### 11.1.4 Main Path of Interference in an IC Die

In RFIC, things are just the opposite. Interference from the space is minor because the substrate area for the RF block is small. Not too many electric lines bend into the RF block from the space to disturb the performance of the circuitry.

On the other hand, the interference through the substrate path is significant because the distance from the interference source to the sensor is very short so that attenuation is minimal; in addition, the permittivity of the substrate is much higher than that of air.

In order to explore the possibility of implementing an RFIC, one must focus on the study of the interference moving along the path of the substrate.

## 11.2 SHIELDING FOR AN RF MODULE BY A METALLIC SHIELDING BOX

At high or radio frequencies, a considerable electromagnetic field radiates from each part in the circuit, including runners. Each part starts to look like a small antenna.

The electromagnetic waves radiate into the surrounding sky and could mix together as a feedback signal returning to the circuitry so as to cause disorder in the circuit performance.

This problem was recognized as the key barrier in the early stages of RF circuit designs, which were fabricated with discrete parts before the 1970s.

In order to avoid the radiation of electromagnetic waves from the RF block to the sky, each individual RF block is shielded by a small metallic box. Figure 11.2 shows a typical shielding scheme. The top portion is a metallic shielding box. The hole at the top is for receiving the short stub on the PCB of the RF block so that the latter can be fixed after it is slid into the metallic shielding box. The PCB of RF block is depicted as the middle portion in Figure 11.2, in which the parts and runners are neglected. The PCB is slid into the metallic shielding box through the slot in the box and its position is fixed by the slot and by the short stub inserted into the hole on the top of the metallic shielding box. The input, output, bias, and DC power supply are assigned to six pins; these are inserted into the pin receivers on the main PCB as the two ears on the metallic shielding box are inserted into two slots on the main PCB simultaneously. After that, the two ears are bent toward the bottom side of the main PCB so that the PCB of the RF block is fixed on the main PCB along with the metallic shielding box. The runners on the main PCB are at the bottom side of PCB, so that they are not short circuited by the metallic shielding box when the box touches the main PCB.

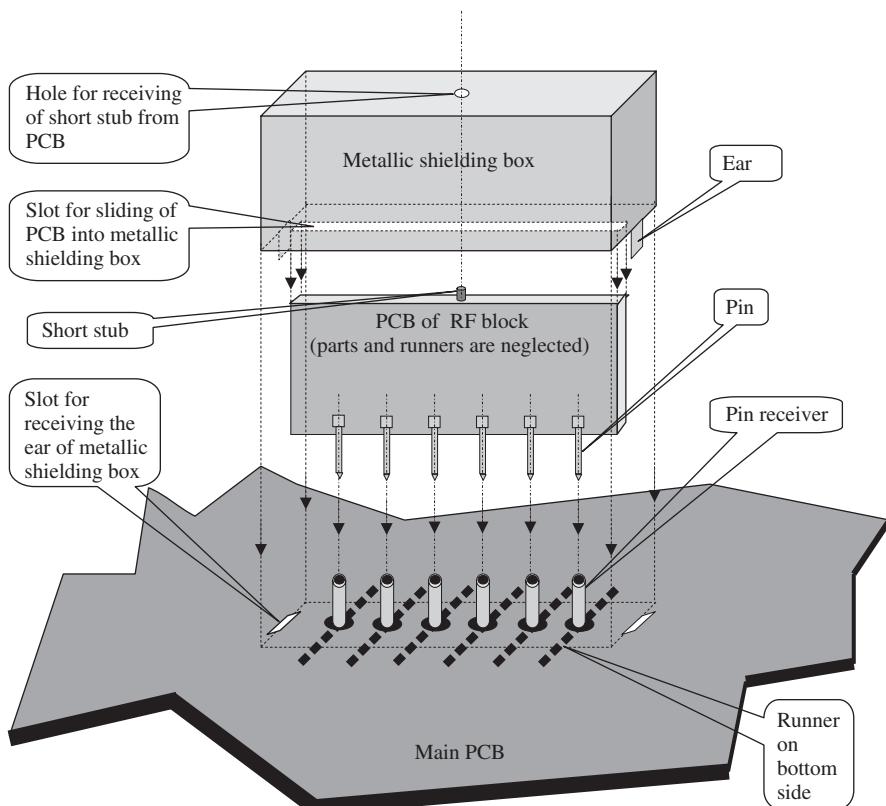


Figure 11.2. Shielding for an RF module.

The addition of the metallic shielding box is not an easy task.

First, the cost of adding the metallic shielding box is usually much higher than the direct material cost of the parts in the RF block.

Second, the insertion configuration of the pin and pin receivers is designed with really high-level technology. They require not only gold plating but also elastic contact with both compression and elongation. The most outstanding feature of the pin and pin receivers is that their characteristic impedances are kept at  $50 \Omega$  over all the operating frequency ranges. Much R&D effort has gone into this sophisticated mechanical-electrical design.

In the 1980s, a portable radio had approximately 10–15 RF modules with metallic shielding boxes. More than 100 pins must have been present in the main PCB. The reliability of all these pins is another subject of concern in the production line.

### 11.3 STRONG DESIRABILITY TO DEVELOP RFIC

In the 1960s, the appearance of the IC brought about a great revolution in electronic circuitry. Compared with the electronic products implemented with discrete parts, the great advantages of the IC are the following:

- greatly reduced cost (at least 10 times lower)
- greatly reduced size (more than 1000 times smaller)
- greatly enhanced reliability of product (at least 100 times more reliable).

The digital circuit was partially fabricated on IC chips in the 1960s; the digital IC became quite popular in circuit designs in the 1970s, though there were a few still fabricated by discrete parts. Digital circuit design experienced a real revolution from the 1960s and the 1970s. The cost and size were greatly reduced and reliability was significantly enhanced from design transitions from discrete parts to IC technology.

In this time period, however, the RF circuit in an electronic product was still fully fabricated by discrete parts. The technological progress in RF circuit design was obviously much slower than that of the digital circuit design. RF circuit design became the main bottleneck in the development of electronic products.

From the viewpoint of commercial competition, it became desirable to apply IC technology to RF circuit design since IC technology had so many advantages compared to the fabrication by discrete parts. These strong incentives existed in all fields of the electronic industry, because it was well known that the unique issues to become a product winner in a serious commercially competitive environment were lowering the cost, reducing the size, increasing the reliability, and enhancing performance.

On the other hand, from the engineering design viewpoint, it seemed hopeless to apply IC technology to RF circuit design because electromagnetic radiation would still exist even if very careful shielding was applied to the RF blocks. It was absolutely impossible to put metallic shielding boxes as shown in Figure 11.1 into an IC die. Consequently, at that time, very few people thought that it would be possible to fabricate an RF circuit block on an IC chip.

Was it really impossible or hopeless to apply IC technology into RF circuit design? In other words, was it impossible to have RFIC appearing in the market?

## 11.4 INTERFERENCE GOING ALONG IC SUBSTRATE PATH

As mentioned above, the main interference in an RF module is from the sky. In the RFIC, things are just the opposite. The interference from the sky is minor but interference through the substrate is significant. In order to explore the possibility of implementing an RFIC, we must focus on the study of interference through the substrate.

### 11.4.1 Experiment

Special experiments have been conducted to study the interference path along the IC substrate. Figure 11.3 shows the vertical profile of an IC die sample for such an experiment. The purpose of this experiment is to provide an interference source to a specific IC substrate and to study its attenuation as it moves along the substrate path.

Point A is an interference source and point B is the sensor for receiving the interference signal moving along the substrate path from A to B.

The source at point A looks like an n-channel MOSFET on the left-hand side, which will be connected to the network analyzer port 1. The network analyzer at port 1 provides an RF signal to this IC die sample as an interference source. The sensor at point B is a P+ buried pad with a metallic contact on the right-hand side, which will be connected to port 2 of the network analyzer and will function as an interference sensor. If the substrate is not a perfect insulator at the interference frequency, the interference signal from port 1 of the network analyzer will move along the substrate path from point A to B and will be received by the sensor. The attenuation of the interference signal from A to B can be found from the reading of  $S_{21}$  in the network analyzer.

If the substrate were an ideal insulator, the attenuation would be infinite and the interference would completely disappear. Usually, the substrate is not an ideal insulator and interference will appear with a certain amount of attenuation. As a matter of fact, the

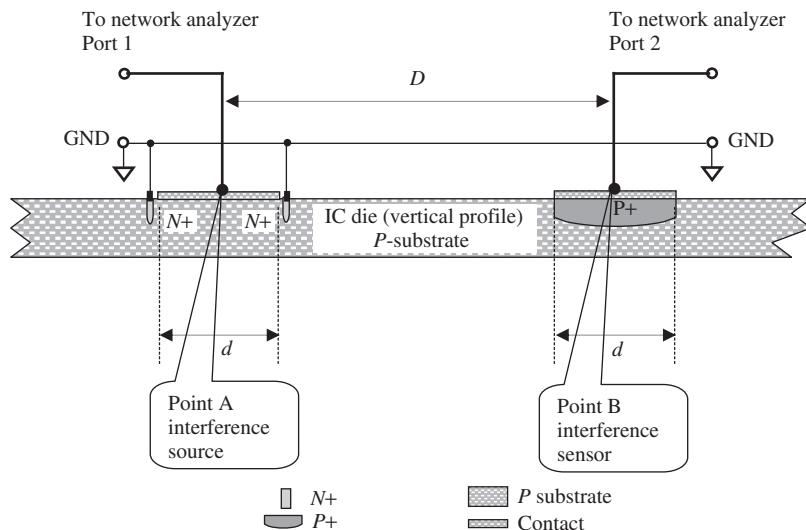


Figure 11.3. IC die sample for studying the interference from the substrate.

attenuation from source A to sensor B is a measure of the isolation between the sensor and the source.

Quantitatively, the value of  $S_{21}$  depends on the distance  $D$  between the sensor and source, the size of the source contact area and the sensor contact area  $d$ , and other parameters of the IC die, such as the thickness of the substrate, the characteristics of the substrate, and so on. As an example, Figure 11.4 shows the measured attenuation of the interference signal from source A to sensor B. Its negative value represents the value of the corresponding isolation. Both ordinates of attenuation and isolation are depicted in Figure 11.4. In this measurement, the distance between the source and sensor is 150  $\mu\text{m}$  and the size of the source or sensor contact area is 50  $\mu\text{m}$ . The experimental results are shown in Figure 11.4 and their approximate values are listed in Table 11.1.

Obviously, the interference moving along the substrate path is significant over the entire RF frequency range. At the high RF range, say,  $f > 1000$  MHz, the attenuation from source to sensor, or the isolation from sensor to source, is only 20 dB! The interference moving along the substrate path will considerably disturb the performance of the RF circuit. We conclude that the RF circuitry implemented by IC technology is impossible if no special IC scheme is developed. Even in the low RF range, say,  $f \approx 10$  MHz, the attenuation from the source A to sensor B, or the isolation from sensor B to source A, is only 40 dB. This is better than that in the high RF range, but is still not good enough to implement the circuit by IC technology.

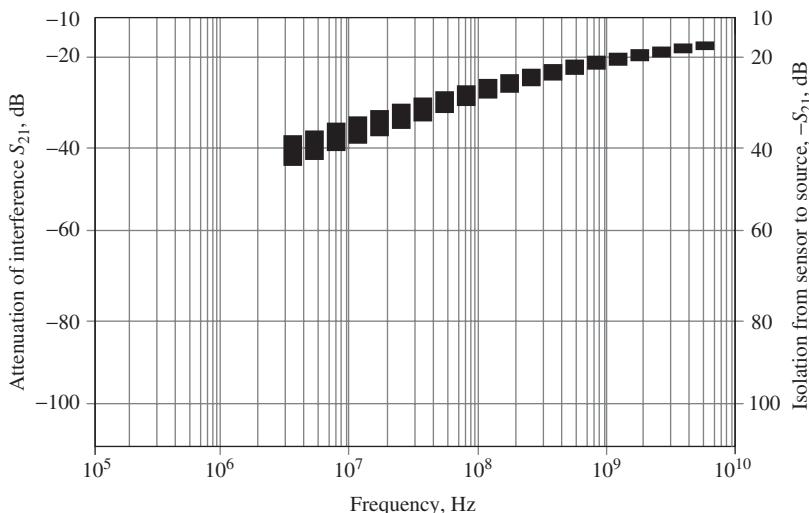


Figure 11.4. Measured attenuation of interference from source A to sensor B or isolation from sensor B to source A when  $D \approx 150 \mu\text{m}$ ,  $d \approx 50 \mu\text{m}$ .

TABLE 11.1. Interference Attenuation or Isolation When Interference Signal Going along IC Substrate Path

---

$S_{21} \approx -40 \text{ dB}$ , when $f = 10 \text{ MHz}$
$S_{21} \approx -30 \text{ dB}$ , when $f = 100 \text{ MHz}$
$S_{21} \approx -20 \text{ dB}$ , when $f = 1000 \text{ MHz}$

---

### 11.4.2 Trench

The experimental results shown in Figure 11.4 or Table 11.1 are very disappointing. However, after carefully reexamining the problem, one may find the solution by some other means.

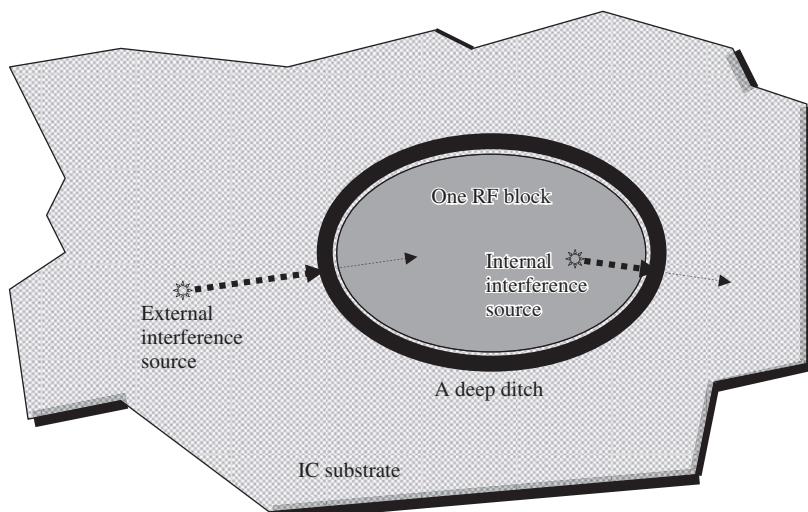
The simplest solution to attenuate or remove the interference moving along the substrate path is to remove the substrate partially so as to block the path by which the interference moves along the source to the sensor. This is called *trenching of an RF block*.

As a concrete example, let us make a trench on an RF block. That is, we dig a deep ditch around the RF block on the IC substrate as shown in Figure 11.5. For simplicity, the area of the RF block and the deep ditch are simply represented by an ellipse. In a practical IC layout, the shape of RF block area and the ditch are irregular and, in addition, the ditch is broken into several segments because the RF block must be connected to the outside blocks.

As shown in Figure 11.5, owing to the existence of the encompassing ditch, the external interference source is considerably blocked by the deep ditch in the path that the interference crosses over the ditch inward. Similarly, any internal interference source is considerably blocked by the deep ditch in the path that the interference must cross over.

Experiments indicate that trenching of an RF block is an effective means to suppress interference moving along the IC substrate path.

Of course, the attenuation of the interference depends on the width and depth of the ditch. The wider and deeper the ditch, the more will be the attenuation of the interference. In practical layouts, the area of the IC die and its cost increase as the width of the ditch increases. If a ditch is too wide, the cost might escalate to an unaffordable level. Additionally, the mechanical rigidity of the IC die will be affected by the depth of the ditch. If a ditch is too deep, the IC die may be damaged easily. Besides, this trenching increases the complexity of IC processing. Nevertheless, in the early stages of RFIC development, much effort was placed on researching the topic of ditch width and depth.



**Figure 11.5.** Trenching of a RF block by digging a deep ditch encompassing the RF block.

### 11.4.3 Guard Ring

By means of trenching around a RF block, the fabrication of the RFIC becomes possible. However, considering the many drawbacks and complexity of trenching, we cannot but ask: is it possible to replace trenching by any other technology or scheme? The answer is yes.

Based on the experimental setup as shown in Figure 11.3, Figure 11.6 shows another experimental setup, in which a P+ guard ring is added to encompass the interference source at point A. The purpose of this guard ring is to block the interference moving along the substrate path from the interference source at point A to the interference sensor at point B. Hopefully, this guard ring will function similar to a trenched ditch as mentioned above.

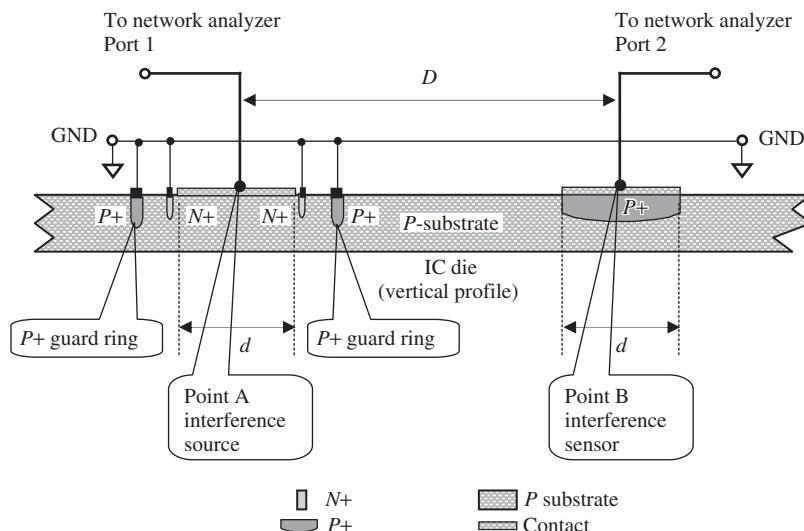
The experimental results with the P+ guard ring shown in Figure 11.7 are positive and encouraging. For comparison, the experimental results without the P+ guard ring shown in Figure 11.4 are also plotted in Figure 11.7.

It can be seen that the attenuation of the interference or isolation is significantly improved from the case without the P+ guard ring to the case with the P+guard ring. Table 11.2 lists the approximate values of  $S_{21}$  at the frequencies 10, 100, and 1000 MHz. The improvement of  $S_{21}$  is 30–40 dB at 10 MHz, 25–30 dB at 100 MHz, and 20 dB at 1000 MHz.

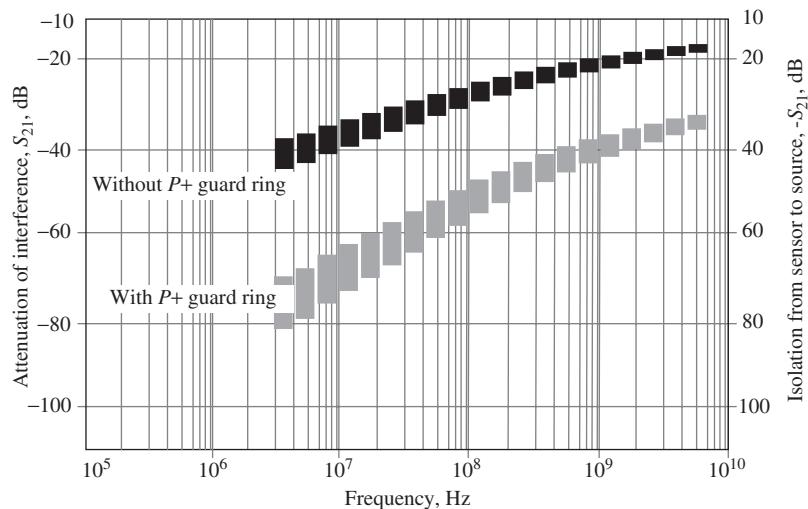
In the past decade, much effort has been put into research on the guard ring, including its type, shape, width, and spacing, as well as multiple guard rings and the distance between multiple rings.

An IC die with a guard ring has better mechanical rigidity than an IC die with a trench. Also, the width and depth of a guard ring are easier to control than those of a trenching ditch in the IC processing.

At present, most trench processing has been replaced by P+ and a deep N-well guard ring. Figure 11.8(a) and (b) shows the RF block encircled by a trenching ditch and two



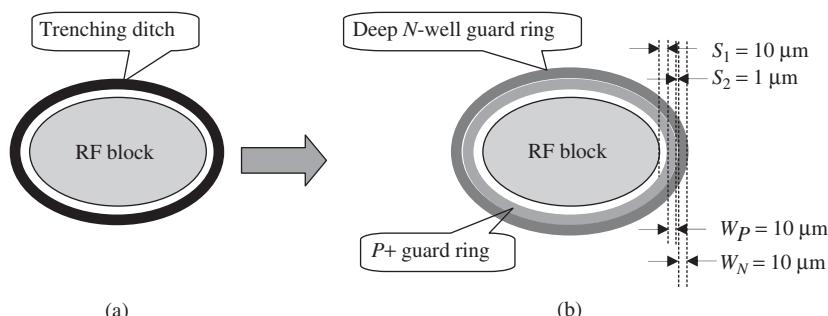
**Figure 11.6.** IC die sample for experiment of interference from substrate interference source circled by a P+guard ring.



**Figure 11.7.** Measured attenuation of interference from source A to sensor B or isolation from sensor B to source A when  $D \approx 150 \mu\text{m}$ ,  $d \approx 50 \mu\text{m}$ .

**TABLE 11.2.** Comparison of Interference Attenuation or Isolation between the Cases without and with P+ Guard Ring

Without P+guard ring, dB	With P+guard ring, dB	Frequency, MHz
$S_{21} \approx -40$	$\approx -80$ to $-70$	10
$S_{21} \approx -30$	$\approx -60$ to $-55$	100
$S_{21} \approx -20$	$\approx -40$	1000



**Figure 11.8.** Trenching ditch replaced by P+ and deep N-well guard ring. (a) Trenching ditch. (b) P+ and deep N-well guard ring.

guard rings (P+ and deep N-well), respectively. Experiments indicate that the attenuation of the interference, or isolation, is increased if the width of the guard ring or the spacing between the guard rings is increased. Circuit designers can select the appropriate width and spacing for guard rings based on the specific requirements of the attenuation of the interference or the isolation.

TABLE 11.3. Typical Width of Guard Ring and Spacing between Guard Rings in a RFIC Layout

Item	Value, $\mu\text{m}$
Spacing between RF block and P+guard ring	$S_1 = 10$
Width of P+guard ring	$W_P = 10$
Spacing between P+guard ring and deep N-well	$S_2 = 1$
Width of deep N-well guard ring	$W_N = 10$

In order to keep the size of the IC die small so as to keep the cost low, the width and spacing of guard rings are confined to less than 15  $\mu\text{m}$ . A set of typical values is marked in Figure 11.8 and listed in Table 11.3, which is applied in RFIC layout currently.

## 11.5 SOLUTION FOR INTERFERENCE COMING FROM SKY

As mentioned above, the interference from the sky in the RFIC die is minor because the substrate area for the RF block is small. Not many electric lines bend into the RF block from the sky which can disturb the performance of the circuitry. Nevertheless, it should still be reduced or removed as much as possible.

In the discussion on PCB in Chapter 9, it was pointed out that the rectangular metallic grounded frame on a PCB not only functions as a grounded surface but also partially plays a role as a shield if this metallic frame is well grounded. It was also pointed out that the P+ guard ring in an IC die plays the similar shielding role as the rectangular metallic frame on a PCB if the P+ guard ring is well grounded. As a matter of fact, not only the P+ guard ring but also the deep N-well guard ring plays the same shielding function. In other words, Figure 9.3 should be modified as Figure 11.9.

Most of the electric lines that are produced and then radiated from the circuit inside the RF block would be terminated on two guard rings. On the other hand, most of the electric lines from any interference source outside the RF block would be stopped at two

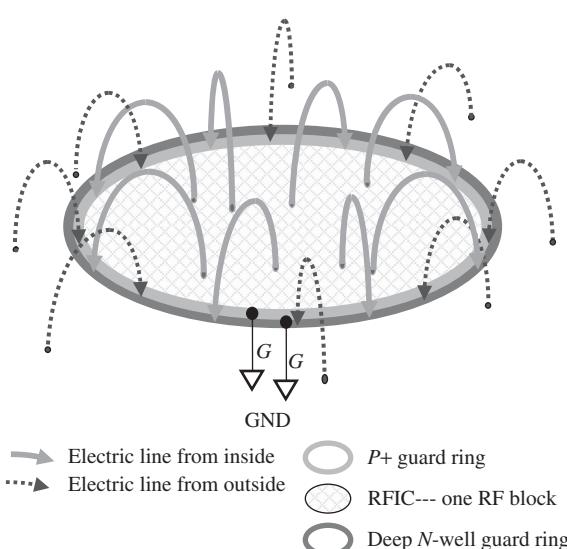
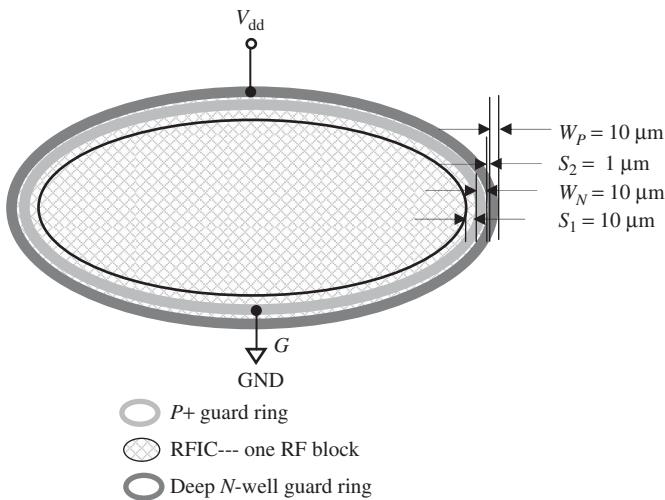


Figure 11.9. Electric lines either radiating from the RFIC die internally or radiated to the RFIC die externally terminated on the grounded guard ring.



**Figure 11.10.** P+ guard ring connected to GND and deep N-well guard ring connected to  $V_{dd}$  or  $V_{cc}$ .

guard rings as well. The isolation between the circuits inside and outside the two guard rings is greatly improved because of the existence of the rings. This is a very simple but very powerful idea, which is the main clue to solving the isolation problem in the RFIC fabrication.

In the actual RFIC layout, the P+ guard ring is connected to the ground (GND) whereas the deep N-well guard ring is connected to  $V_{dd}$  or  $V_{cc}$ . Figure 11.10 shows the connection for these two guard rings. The voltage drop between the P+ and N-well guard rings furthermore improves the attenuation of the interference, or isolation.

Through the use of P+ and N-well guard rings, the problem of interference or isolation between the RF blocks is basically solved. Thus, RFIC fabrication becomes realistic. This was a major milestone in the history of electronic circuitry.

## 11.6 COMMON GROUNDING RULES FOR RF MODULE AND RFIC DESIGN

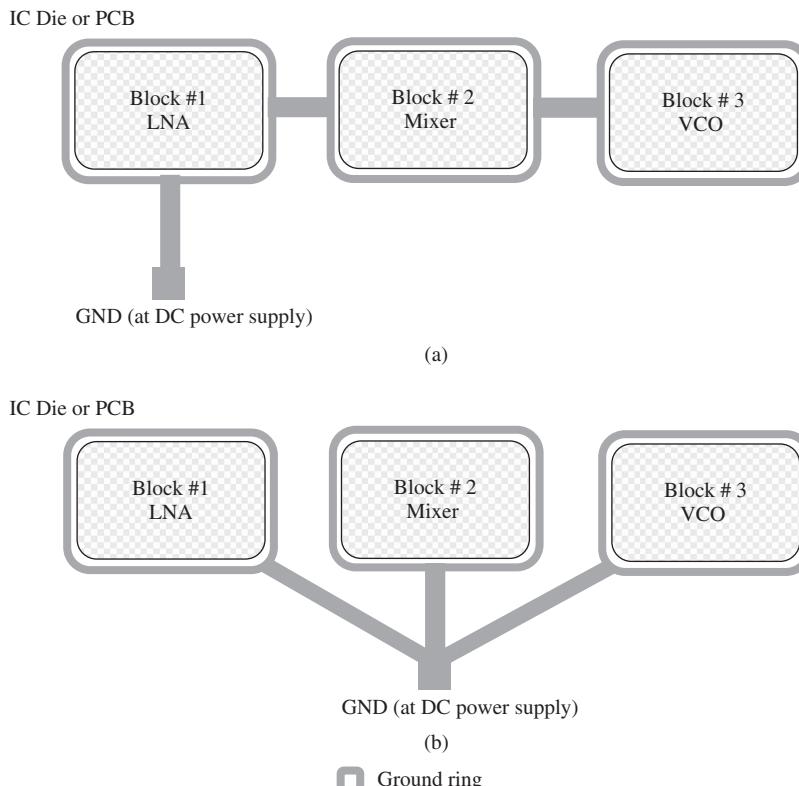
In addition to solving the problem of interference or isolation between RF blocks, some common grounding rules must be followed for both RF modules and RFICs.

### 11.6.1 Grounding of Circuit Branches or Blocks in Parallel

In Chapter 8, grounding of a PCB was discussed. In order to reduce or avoid return current coupling between multiple circuit branches or blocks, some long slots were inserted between the circuit branches or blocks so that the grounding surfaces of individual circuit branches or individual blocks were separated from one another.

In RFIC design, the grounding of circuit branches or blocks must also be separated so as to reduce or avoid return current coupling between multiple circuit branches or blocks.

As matter of fact, “grounding separately” is equivalent to “grounding in parallel.” Its opposite would be “grounding stacked together” or “grounding in series.” In an RFIC chip or a PCB containing multiple circuit branches or multiple blocks, grounding of



**Figure 11.11.** Grounding connection for multiple circuit branches or blocks. (a) Incorrect grounding connection in series or stacked together. (b) Correct grounding connection in parallel or separately.

circuit branches or blocks must be connected separately but not stacked together, that is, in parallel but not in series. Figure 11.11 shows correct and incorrect grounding methods.

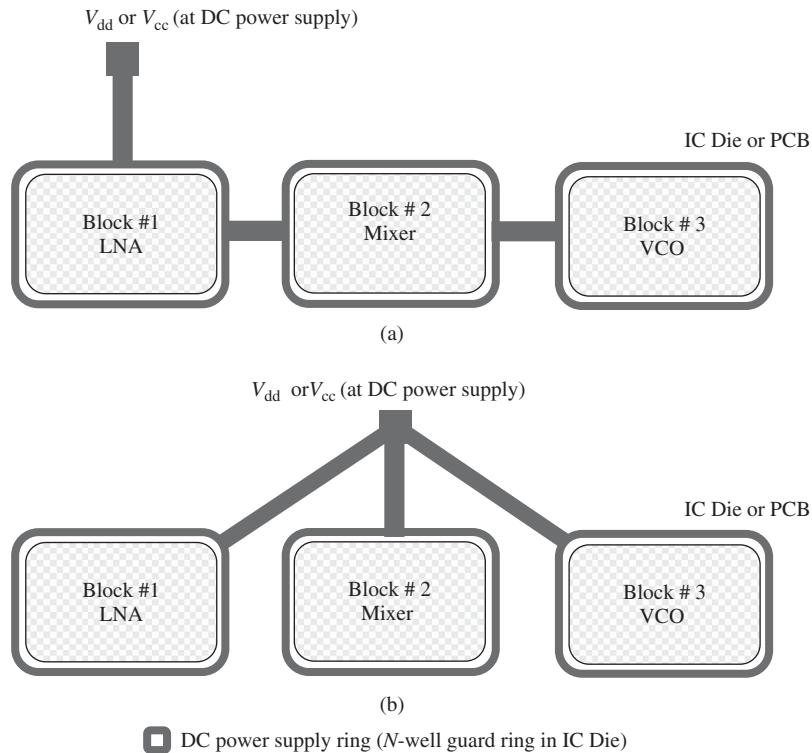
### 11.6.2 DC Power Supply to Circuit Branches or Blocks in Parallel

As mentioned in Chapter 7, the DC power supply point is a half ground point.

Similarly, the DC power supply to a PCB was discussed in Chapter 8. In order to reduce or avoid forward current coupling between multiple circuit branches or blocks, the DC power supply must be provided separately or in parallel.

In RFIC design, the DC power supply to circuit branches or blocks must also be separated or provided in parallel so as to reduce or avoid forward current coupling between multiple circuit branches or blocks.

As matter of fact, the statement “DC power supply provided separately” is equivalent to “DC power supply provided in parallel.” Its opposite is “DC power supply provided stacked together” or “DC power supply provided in series.” In an RFIC chip or a PCB containing multiple circuit branches or multiple blocks, the DC power supply to the circuit branches or blocks must be connected separately but not stacked together, that is, in parallel but not in series. Figures 11.12 shows correct and incorrect methods of providing the DC power supply.



**Figure 11.12.** DC power supply connection for multiple circuit branches or blocks. (a) Incorrect DC power supply connection in series or stacked together. (b) Correct DC power supply connection in parallel or separately.

## 11.7 BOTTLENECKS IN RFIC DESIGN

### 11.7.1 Low-Q Inductor and Possible Solution

In RFIC design, an inductor is built on the IC substrate. On a two-dimensional plane, it can take only a spiral configuration. Figure 11.13 shows a typical spiral inductor on the IC substrate and Figure 11.14 its equivalent model.

There are three problems affecting a spiral inductor:

1. *Extremely Low Q Value.* With respect to that of  $L_S$ , the value of  $R_S$  is usually high, so that the  $Q$  value of the spiral inductor is low.

In the RF range, the  $Q$  value of the spiral inductor is about 10 or so at present in most foundries. In an RF circuit design with discrete parts, the  $Q$  value of inductors is generally over 100. For example, the  $Q$  value of a chip inductor is around 120.

This is the biggest bottleneck in RFIC design.

2. *Large Area.* The spiral inductor's size is usually more than  $100 \mu\text{m}^2$ . This area is 10 or 100 times that of a resistor. The large area impacts the cost of the IC chip directly since the cost is directly proportional to its die area.

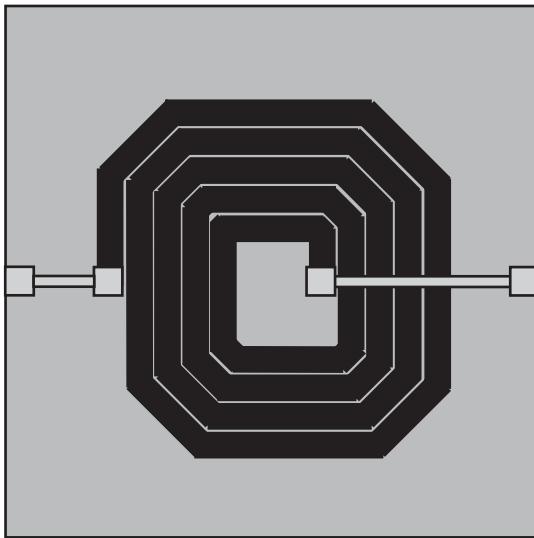


Figure 11.13. Spiral configuration of an inductor on an IC chip.

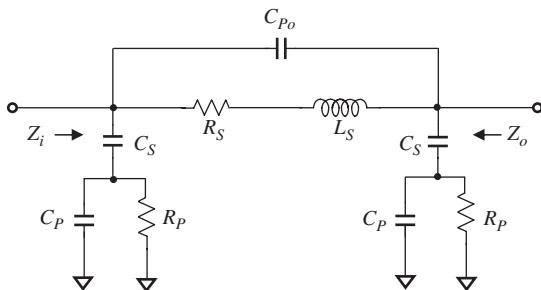


Figure 11.14. Model of a spiral inductor.

3. Other interference sources easily interfere with it and, additionally, it is an interference source to disturb other parts or blocks because its size is large and its magnetic flux or electric lines are open wide to the sky.

It seems that we cannot do too much about the last two problems, except that the interference could be somewhat alleviated by a reasonable layout. We will focus on the low- $Q$  problem, which leads many problems in RFIC design such as the following:

- It is impossible to build a filter directly on an IC chip. In the fabrication of a filter, the  $Q$  value of the inductor must be around 100 or higher. Otherwise, the insertion loss of the filter would approach an unacceptable level.
- Besides the filter, some other RF circuit blocks cannot be built on the RFIC chip due to the low  $Q$  value of the inductor. For example, the input impedance matching network of the LNA (low-noise amplifier) may not be built on the IC chip because the low  $Q$  value of the inductor could change an LNA to an HNA (high-noise amplifier).
- An off-chip inductor must be applied and connected to the RFIC chip if a high  $Q$  value of inductor is required in the circuitry. A pair of input and output bonding wires, pads, and pins for an off-chip inductor must be provided. The bonding wires

and pads bring about serious uncertainty and inconsistency in the production line. An IC package cannot afford too many additional bonding wires and pins. For instance, if 10 off-chip inductors were needed, the additional number of bonding wires and pins would be greater than 20.

In the last decade, some scientists and engineers have attempted to avoid the use of inductors in certain RF circuit designs such as filters. For instance, the log-domain filtering scheme is one of the techniques in the construction of a filter by only transistors, capacitors, and resistors. Unfortunately, this is realistic only when the operating frequency is low, say, below 1 GHz. Second, this filter is an active circuit, so the cost of DC current or power consumption must be paid. Similar to the log-domain filtering technique, much effort has been put on the development of the equivalent inductor circuit, the gyrator. The restrictions or drawbacks of the gyrator are the same as those of the log-domain filtering technique.

The inductor is one of the indispensable parts of an RF circuit design because it is a phase-shift part. Avoiding the use of inductors in an RFIC is almost impossible. People have been putting a lot of effort to enhance the  $Q$  value of the spiral inductor. Unfortunately, it looks like not too much exciting progress has been made to date.

Why is the  $Q$  value of a spiral inductor so low? Many hypotheses have been presented and many experiments conducted. They can be outlined below.

**SKIN EFFECT.** Some people believe that the low  $Q$  value of the spiral inductor is due to the skin effect on the spiral metallic loop. The depth of the skin effect  $\delta$  for copper can be evaluated and is found approximately

$$\delta \approx 0.66 \text{ } \mu\text{m}, \quad \text{when frequency} = 10 \text{ GHz}, \quad (11.6)$$

and

$$\delta \approx 6.6 \text{ } \mu\text{m}, \quad \text{when frequency} = 100 \text{ MHz}. \quad (11.7)$$

If the metal is copper and its relative electric permittivity

$$\varepsilon_r = 8.854 \times 10^{-12} \approx \frac{1}{36\pi \times 10^9},$$

its relative magnetic permeability

$$\mu_r \approx 4\pi \times 10^{-7} \text{ (H/m);}$$

its conductivity

$$\sigma \approx 5.8 \times 10^7 (\Omega^{-1}/\text{m}).$$

On the other hand, the thickness of the metal layer in IC is on the order of

$$T \sim 0.1 \text{ } \mu\text{m}. \quad (11.8)$$

One might conclude that the low  $Q$  value of the spiral inductor on an IC chip is due to the fact that the metal layer is too thin when the operating frequency is below 10 GHz.

On the basis of such a judgment, many experiments in which the wire thickness of the spiral inductor is increased have been conducted in order to enhance the  $Q$  value of IC inductors. Unfortunately, the  $Q$  value of IC inductors does not increase considerably even though the thickness of the metallic wire is pushed up to several micrometers.

This indicates that the thickness of the metal layer is not the main reason that brings about the low  $Q$  value of the IC spiral inductor.

**ATTENUATION DUE TO THE EXISTENCE OF THE SUBSTRATE.** The metallic winding of an IC spiral inductor is laid on a substrate. The eddy current may appear on the substrate, which leads to the attenuation of the RF signal. On the other hand, the electric permittivity  $\epsilon_s$  of the substrate is generally much higher than the permittivity of the free space, so that the electromagnetic field as well as its power in the substrate is much stronger than in the free space. The electromagnetic field can be stored, propagated, and consequently attenuated in the substrate.

Therefore, the substrate might be the main source causing the low value of the IC spiral inductor.

An outstanding experiment was conducted at the University of California, Los Angeles (UCLA), USA, in which the substrate beneath the spiral inductor is dug away. This is a difficult and highly technical task. Professors, students, scientists, and engineers at UCLA had been working hard and finally completed this special treatment successfully. The  $Q$  value of this spiral inductor was indeed increased, but unfortunately, it was still not high enough to satisfy most RFIC designs.

In recent years, the micromachined RF inductor was developed and reported in RFIC circuit design. The idea is basically the same as UCLA one. In addition to the highly technical works involving in the IC processing, the high cost is another problem.

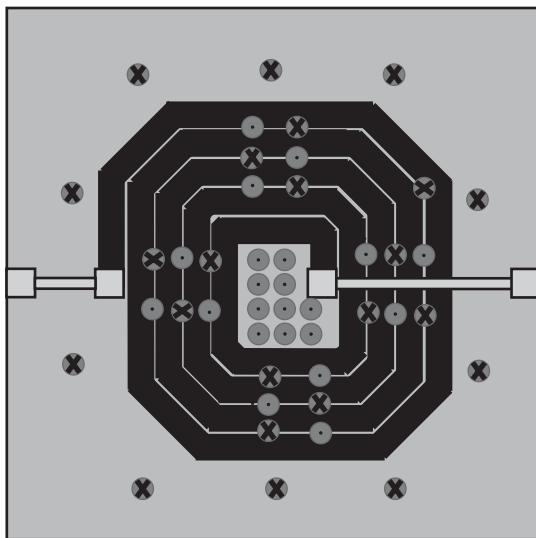
**FLUX LEAKAGE.** Flux can leak from the gap between windings. Flux leak implies energy loss or attenuation of signal power. If the gap between windings is reduced or narrowed, the flux leak would be reduced and the  $Q$  value of the spiral inductor should be increased. Unfortunately, the increase of the  $Q$  value through this method is very small.

Flux can be emitted from the gap between the windings to the sky. The flux escape also implies energy loss or attenuation of signal power. A “sandwich” experiment was conducted in which the spiral inductor was enclosed by a magnetic-shielding material. The purpose of this experiment was to enhance the  $Q$  value of the spiral inductor by stopping the flux escaping to the sky. Unfortunately, the  $Q$  value of the spiral inductor did not show a large difference before and after the spiral inductor was “sandwiched.”

Therefore, we again conclude that the low  $Q$  value of the spiral inductor is not mainly due to the flux leakage or escape from the gap between the windings to the sky.

**FLUX CANCELLATION.** Owing to the inherent drawback of the spiral configuration, the cancellation of flux between the windings is significant. As shown in Figure 11.15, in the spacing between two windings, the flux produced by the inner winding is cancelled by the flux produced by the outer winding. The flux cancellation seems to be more important than the flux leakage to the reduction of the  $Q$  value.

So far, many theories have been presented and many experiments have been conducted. There may be multiple reasons for the low  $Q$  value of spiral inductors. However, flux cancellation seems to be the most important reason. If so, then the low- $Q$  problem inherently exists because of the spiral configuration.



✖ Flux into paper  
● Flux out from paper

Figure 11.15. Cancellation of flux between two windings.

If the spiral configuration of the inductor is kept unchanged, a possible solution for the low- $Q$  problem is to compensate for its  $Q$  value by adding a negative resistance to the spiral inductor.

It is well known that the  $Q$  value of an inductor  $L$  can be expressed by its reactance  $L\omega$  divided by an equivalent resistance  $r$  in series, that is,

$$Q = \frac{L\omega}{r}. \quad (11.9)$$

If a resistor with negative resistance  $r'$  is connected to this inductor  $L$  in series, the  $Q$  value will be changed to

$$Q = \frac{L\omega}{r - r'} = \frac{L\omega}{\Delta r}, \quad (11.10)$$

where

$$\Delta r = r - r'. \quad (11.11)$$

Theoretically,  $Q$  can be enhanced to any value as needed as long as the resistance  $r$  can be compensated for by the negative resistance  $r'$  so that  $\Delta r$  stays below a small positive value. Figure 11.16 shows the change of an inductor's  $Q$  value by the addition

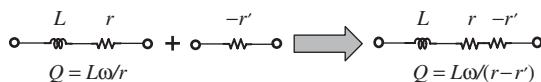


Figure 11.16. Change of an inductor's  $Q$  value by adding a negative resistance to the inductor in series.

of a negative resistance to the inductor in series. In the extreme case, the  $Q$  value can be compensated to infinity if the negative resistance  $-r'$  is equal to the original resistance  $r$  of the inductor, in magnitude.

However, it is not so simple in an actual engineering design. The difficult points are the following:

- generating a negative resistance;
- ensuring that there is not negative resistance outside the expected bandwidth;
- keeping the remaining negative resistance inside the bandwidth below a small positive value;
- reducing current consumption of the generated negative resistance, which is usually done by an active device;
- handling the noise generated as a result of the existence of the active device.

### 11.7.2 “Zero” Capacitor

The second bottleneck in RFIC design is the development of a “zero” capacitor directly in the IC chip.

In RF module design with PCB testing, zero capacitors can be selected from chip capacitors based on their self-resonant principle. However, the chip zero capacitor cannot be directly applied to the IC chip because of its large size.

As mentioned in Section 7.4.6, when a zero capacitor is needed in an RF block fabricated by RFIC, the circuit on the IC chip must come out of the IC chip to connect to the zero capacitor by bonding wires. This brings about much inaccuracy because of the addition of the bonding wires, pads, and the additional runners. Second, if many off-chip zero capacitors are needed and the IC chip is packaged with multiple pins, the number of pins required may be unacceptably high. Finally, if many off-chip zero capacitors are needed, the cost will be raised considerably.

Therefore, it is strongly desirable to develop a zero capacitor directly in the IC chip. Hopefully, progress of such a development will appear in the near future.

### 11.7.3 Bonding Wire

The bonding wire may create other problems to the RFIC designer.

Two ends of a bonding wire are connected to the pads, which are the points of impedance discontinuity, or to the terminals where the impedance is unmatched. In order to ensure impedance matching or smooth transition, an accurate model of the bonding wire is demanded. Unfortunately, the current models of bonding wires provided by foundries are very inaccurate.

The bonding wire is a very unstable part. The inaccuracy of the model is due to many uncertain factors: for example, bonding wires with the same length can have different heights or different tilted angles in the air. When they are soldered on the pad, the uncertainty of the soldered point may cause the inaccuracy of the model as well.

### 11.7.4 Via

Modeling of a via has been discussed in Section 9.4. The complication of via modeling comes from three factors:

1. There are three types of via: blind via, buried via, and through via;
2. The number of metallic layers in a PCB or in an IC substrate can be greater than 2;
3. The model of multi vias would be more complicated than that of a single one.

The study and research of via is still far from the engineering design needs, although it has been paid much attention to in recent years.

## 11.8 CALCULATING OF QUARTER WAVELENGTH

Hannu (2000) summarized a set of equations for the calculation of quarter wavelength in CMOS (complementary metal–oxide–semiconductorOxide–Semiconductor) processing. Figure 11.17 shows the various parameters of a microstrip line fabricated on silicon substrate in CMOS processing. It can be described by the distribution parameters, the capacitance per unit length in respect to the substrate  $C_{\text{msl}}$ , and the self-inductance per unit length along the runner  $L_{\text{msl}}$ .

$$C_{\text{msl}} = \epsilon_{\text{ox}} \left[ 2.42 + \frac{W}{X_{\text{int}}} - 0.44 \frac{X_{\text{int}}}{W} + \left( 1 - \frac{X_{\text{int}}}{W} \right)^6 \right] \text{ F/cm}, \quad (11.12)$$

$$L_{\text{msl}} = 2 \ln \left[ \frac{8X}{W} + \frac{W}{4X} \right] (\text{nH/cm}), \quad (11.13)$$

$$\epsilon_{\text{ox}} = 3.45 \times 10^{-13} \text{ F/cm}, \quad (11.14)$$

$$X = X_{\text{int}} + X_{\text{si}}, \quad (11.15)$$

where

$C_{\text{msl}}$  = the capacitance per unit length in respect to the substrate,

$W$  = the width if microstrip line,

$X_{\text{int}}$  = the thickness of oxide layer,

$\epsilon_{\text{ox}}$  = the electric permittivity of the silicon oxide layer,

$L_{\text{msl}}$  = the self-inductance per unit length along the runner, and

$X_{\text{si}}$  = the thickness of silicon substrate.

The characteristic impedance is

$$Z_0 = \sqrt{\frac{L_{\text{msl}}}{C_{\text{msl}}}}. \quad (11.16)$$

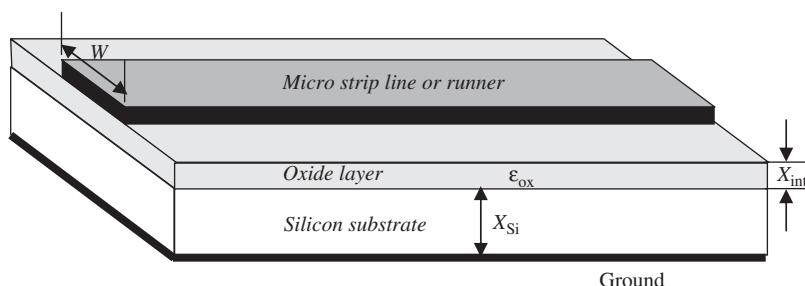


Figure 11.17. Various parameters of a microstrip line on a silicon substrate.

The quarter wavelength is

$$\frac{\lambda}{4} = \frac{2\pi}{4} \sqrt{C_{\text{msl}} L_{\text{msl}}} . \quad (11.17)$$

As an example, assuming that  $X_{\text{int}} = 0.5 \mu\text{m}$ ,  $X_{\text{si}} = 500 \mu\text{m}$ , and  $\epsilon_{\text{ox}} = 3.45 \times 10^{-13} \text{ F/cm}$ , the quarter wavelength with different frequencies and the width of the microstrip line are as listed in Tables 11.4 and 11.5.

TABLE 11.4. Example of Calculated Quarter Wavelength in CMOS Processing

$W, \mu\text{m}$	$C, \text{pF/cm}$	$L, \text{nH/cm}$	$Z_o, \Omega$	$f, \text{GHz}$	$\lambda/4, \text{mm}$	$\lambda/360, \mu\text{m}$	$f, \text{GHz}$	$\lambda/4, \text{mm}$	$\lambda/360, \mu\text{m}$
1	1.45	16.59	106.8	1.0	16.09	178.83	2.4	6.71	74.51
2	2.24	15.20	82.4	1.0	13.55	150.58	2.4	5.65	62.74
3	3.00	14.39	69.3	1.0	12.04	133.79	2.4	5.02	55.74
4	3.73	13.82	60.9	1.0	11.01	122.34	2.4	4.59	50.98
5	4.45	13.37	54.8	1.0	10.25	113.84	2.4	4.27	47.43
6	5.17	13.01	50.2	1.0	9.64	107.15	2.4	4.02	44.65
7	5.88	12.70	46.5	1.0	9.15	101.70	2.4	3.81	42.37
8	6.58	12.43	43.5	1.0	8.74	97.13	2.4	3.64	40.47
9	7.28	12.20	40.9	1.0	8.39	93.22	2.4	3.50	38.84
10	7.98	11.98	38.8	1.0	8.08	89.82	2.4	3.37	37.42
11	8.68	11.79	36.9	1.0	7.81	86.82	2.4	3.26	36.18
12	9.38	11.62	35.2	1.0	7.57	84.16	2.4	3.16	35.06
13	10.07	11.46	33.7	1.0	7.36	81.76	2.4	3.07	34.07
14	10.77	11.31	32.4	1.0	7.16	79.59	2.4	2.98	33.16
15	11.46	11.17	31.2	1.0	6.99	77.62	2.4	2.91	32.34
16	12.16	11.04	30.1	1.0	6.82	75.81	2.4	2.84	31.59
17	12.85	10.92	29.2	1.0	6.67	74.14	2.4	2.78	30.89
18	13.54	10.81	28.3	1.0	6.53	72.60	2.4	2.72	30.25
19	14.23	10.70	27.4	1.0	6.41	71.17	2.4	2.67	29.65
20	14.93	10.60	26.6	1.0	6.29	69.84	2.4	2.62	29.10
25	18.39	10.15	23.5	1.0	5.79	64.29	2.4	2.41	26.79
30	21.84	9.79	21.2	1.0	5.41	60.07	2.4	2.25	25.03
35	25.30	9.48	19.4	1.0	5.10	56.72	2.4	2.13	23.63
40	28.75	9.21	17.9	1.0	4.86	53.97	2.4	2.02	22.49
45	32.21	8.98	16.7	1.0	4.65	51.66	2.4	1.94	21.53
50	35.66	8.77	15.7	1.0	4.47	49.68	2.4	1.86	20.70
55	39.11	8.58	14.8	1.0	4.32	47.96	2.4	1.80	19.98
60	42.56	8.40	14.1	1.0	4.18	46.45	2.4	1.74	19.35
65	46.01	8.24	13.4	1.0	4.06	45.11	2.4	1.69	18.79
70	49.46	8.09	12.8	1.0	3.95	43.90	2.4	1.65	18.29
75	52.92	7.96	12.3	1.0	3.85	42.81	2.4	1.61	17.84
80	56.37	7.83	11.8	1.0	3.76	41.82	2.4	1.57	17.42
85	59.82	7.71	11.4	1.0	3.68	40.91	2.4	1.53	17.05
90	63.27	7.59	11.0	1.0	3.61	40.08	2.4	1.50	16.70
95	66.72	7.48	10.6	1.0	3.54	39.31	2.4	1.47	16.38
100	70.17	7.38	10.3	1.0	3.47	38.59	2.4	1.45	16.08
200	139.17	6.00	6.6	1.0	2.74	30.39	2.4	1.14	12.66
500	346.18	4.22	3.5	1.0	2.07	22.98	2.4	0.86	9.57

TABLE 11.5. Example of Calculated Quarter Wavelength in CMOS Processing

$W$ , $\mu\text{m}$	$C$ , $\text{pF/cm}$	$L$ , $\text{NH/cm}$	$Z_o$ , $\Omega$	$f$ , $\text{GHz}$	$\lambda/4$ , $\text{mm}$	$\lambda/360$ , $\mu\text{m}$	$f$ , $\text{GHz}$	$\lambda/4$ , $\text{mm}$	$\lambda/360$ , $\mu\text{m}$
1	1.45	16.59	106.8	5.8	2.77	30.83	10.0	1.61	17.88
2	2.24	15.20	82.4	5.8	2.34	25.96	10.0	1.36	15.06
3	3.00	14.39	69.3	5.8	2.08	23.07	10.0	1.20	13.38
4	3.73	13.82	60.9	5.8	1.90	21.09	10.0	1.10	12.23
5	4.45	13.37	54.8	5.8	1.77	19.63	10.0	1.02	11.38
6	5.17	13.01	50.2	5.8	1.66	18.47	10.0	0.96	10.72
7	5.88	12.70	46.5	5.8	1.58	17.53	10.0	0.92	10.17
8	6.58	12.43	43.5	5.8	1.51	16.75	10.0	0.87	9.71
9	7.28	12.20	40.9	5.8	1.45	16.07	10.0	0.84	9.32
10	7.98	11.98	38.8	5.8	1.39	15.49	10.0	0.81	8.98
11	8.68	11.79	36.9	5.8	1.35	14.97	10.0	0.78	8.68
12	9.38	11.62	35.2	5.8	1.31	14.51	10.0	0.76	8.42
13	10.07	11.46	33.7	5.8	1.27	14.10	10.0	0.74	8.18
14	10.77	11.31	32.4	5.8	1.24	13.72	10.0	0.72	7.96
15	11.46	11.17	31.2	5.8	1.20	13.38	10.0	0.70	7.76
16	12.16	11.04	30.1	5.8	1.18	13.07	10.0	0.68	7.58
17	12.85	10.92	29.2	5.8	1.15	12.78	10.0	0.67	7.41
18	13.54	10.81	28.3	5.8	1.13	12.52	10.0	0.65	7.26
19	14.23	10.70	27.4	5.8	1.10	12.27	10.0	0.64	7.12
20	14.93	10.60	26.6	5.8	1.08	12.04	10.0	0.63	6.98
25	18.39	10.15	23.5	5.8	1.00	11.08	10.0	0.58	6.43
30	21.84	9.79	21.2	5.8	0.93	10.36	10.0	0.54	6.01
35	25.30	9.48	19.4	5.8	0.88	9.78	10.0	0.51	5.67
40	28.75	9.21	17.9	5.8	0.84	9.31	10.0	0.49	5.40
45	32.21	8.98	16.7	5.8	0.80	8.91	10.0	0.46	5.17
50	35.66	8.77	15.7	5.8	0.77	8.57	10.0	0.45	4.97
55	39.11	8.58	14.8	5.8	0.74	8.27	10.0	0.43	4.80
60	42.56	8.40	14.1	5.8	0.72	8.01	10.0	0.42	4.65
65	46.01	8.24	13.4	5.8	0.70	7.78	10.0	0.41	4.51
70	49.46	8.09	12.8	5.8	0.68	7.57	10.0	0.40	4.39
75	52.92	7.96	12.3	5.8	0.66	7.38	10.0	0.39	4.28
80	56.37	7.83	11.8	5.8	0.65	7.21	10.0	0.38	4.18
85	59.82	7.71	11.4	5.8	0.63	7.05	10.0	0.37	4.09
90	63.27	7.59	11.0	5.8	0.62	6.91	10.0	0.36	4.01
95	66.72	7.48	10.6	5.8	0.61	6.78	10.0	0.35	3.93
100	70.17	7.38	10.3	5.8	0.60	6.65	10.0	0.35	3.86
200	139.17	6.00	6.6	5.8	0.47	5.24	10.0	0.27	3.04
500	346.18	4.22	3.5	5.8	0.36	3.96	10.0	0.21	2.30

In the practical RFIC design, the width of the runner is typically taken in the range

$$100 \text{ } \mu\text{m} \geq W \geq 5 \text{ } \mu\text{m}. \quad (11.18)$$

$$X_{\text{int}} = 0.5 \text{ } \mu\text{m}, X_{\text{si}} = 500 \text{ } \mu\text{m}, \text{ and } \varepsilon_{\text{ox}} = 3.45 \times 10^{-13} \text{ F/cm}.$$

## REFERENCE

Hannu T. CMOS interconnects. Kungl Tekniska Hogskolan: Electronic System Laboratory; 2000.

## FURTHER READING

- Ashby KB, Finley WC, Bastek JJ, Moinian S, Koullias IA. 70 High Q inductors for wireless applications in a complementary silicon bipolar process. In: Proceedings of the 1994 Bipolar/BiCMOS Circuits and Technology Meeting; Minneapolis, MN, 1994 Oct 10–11. 1994. pp. 179–182.
- Ashby KB, Koullias IA, Finley WC, Bastek JJ, Moinian S. High Q inductors for wireless applications in a complementary silicon bipolar process. *IEEE J Solid-State Circ* 1996;31(1):4–9.
- Behzad R. Design of analog CMOS integrated circuits. McGraw-Hill Companies; New York, NY, 2001.
- David J-B, Musalem F-X, Albert P. RF high-Q spiral inductor design. In: Proceedings of the 1st ISA/IEEE Conference, Sensor for Industry; Rosemont, IL, Nov 5–7. 2001. pp. 78–82.
- Erben U, Abele P, Behammer D, Ergraber H, Schmacher H. High-Q inductors on silicon using a quasi thin film microstrip technique. In: Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems; 1998 Sep 17–18, Digest of Papers. 1998; pp. 155–159.
- Frei MR, Belk NR, Dennis DC, Carroll MS, Lin W, Pinto MR, Archer VD, Ivanov TG, Moinian S, Ng KK, Chu J. Integration of high-Q inductors in a latch-up resistant CMOS technology. In: International Electron Devices Meeting; Washington, DC, Dec 1999; IEDM Technical Digest. 1999. pp. 757–760.
- Groves R, Malinowski J, Volant R, Jadus D. High Q inductors in a SiGe BiMOS process utilizing a thick metal process add-on module. In: Proceedings of the 1999 Bipolar/BiCMOS Circuits and Technology Meeting; Minneapolis, MN, 1999 Sep 26–28. 1999. pp. 149–152.
- Hafizi M, Shen F, Taoing F, Schulze K, Ruth R, Schwab R, Karlsen P, Simmonds D, Qizheng G. RF front-end of direct conversion receiver RFIC for CDMA-2000. *IEEE J Solid-State Circ* 2004;39(10):1622–1632.
- Han F-Y, Wu J-M, Horng T-S, Tu C-C. A rigorous study of package and PCB effects on W-CDMA Upconverter RFICs. *IEEE Trans Microw Theory Tech* 2006;54(10):3793–3803.
- Jenei S, Decoutere S, Van Huylenbroeck S, Vanhorebeek G, Nauwelaers B. High Q inductors and capacitors on Si substrate. In: 2001 Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems; 2001 Sep 12–14, Digest of Papers. 2001. pp. 64–70.
- Kaczman D, Dozier C, Godambe N, Shah M, Guimaraes H, Rachedine M, Alam M, Han L, Shepherd W, Cashen D, Ganger J, Couglar K, Getka B, Brotkowski E, Wong D, Hayes D. A tri-band (2100/1900/800MHz) single-chip cellular transceiver for WCDMA/HSDPA. (Invited Paper), 2005 IEEE Radio Frequency Integrated Circuits Symposium. 2005. pp. 281–284.
- Kamogawa K, Nishikawa K, Tokumitsu T, Tanaka M. A novel high-Q inductor based on Si 3D MMIC technology and its application. In: 1999 IEEE Radio Frequency Integrated Circuits (RFIC) Symposium; Kyoto, Japan, 1999 June 13–15. pp. 185–188.
- Kang S-M (Steve), Leblebici Y. CMOS digital integrated circuits. WCB/McGraw-Hill Companies, Inc.; 1999.
- Kim S-J, Lee Y-G, Yun S-K, Lee H-Y. Realization of high-Q inductors using wirebonding technology. In: The 1st IEEE Asia Pacific Conference on ASICs, AP-ASIC '99; Seoul, South Korea, 1999 Aug 23–25. pp. 13–16.
- Klemmer N, Hartung J. High Q inductors for MCM-Si technology. In: 1997 IEEE Multi-Chip Module Conference, Santa Cruz, California 1997. MCMC '97; 1997 Feb 4–5. 1997. pp. 33–37.

- Lakdawala H, Zhu X, Luo H, Santhanam S, Carley LR, Fedder GK. Micromachined high-Q inductors in a 0.18 –  $\mu\text{m}$  copper interconnect low-k dielectric CMOS process. *IEEE J Solid-State Circ* 2002;37(3):394–403.
- Lee C-Y, Chen T-S, Deng JD-S, Kao CH. A simple systematic spiral inductor design with perfected Q improvement for CMOS RFIC application. *IEEE Trans Microw Theory Tech* 2005;53(2):523–528.
- Lin JW, Chen CC, Cheng YT. A robust high-Q micromachined RF inductor for RFIC applications. *IEEE Trans Electron Devices* 2005;52(7):1489–1496.
- Lin J-W, Chen CC, Huang JK, Cheng YT. An optimum design of the micromachined RF inductor. In: *IEEE Radio Frequency Integrated Circuits Symposium*, Hsinchu, Taiwan 2004. pp. 639–642.
- Lin Y-S, Liang H-B, Chen C-C, Wang T, Lu S-S. A high quality factor and low power loss micromachined RF bifilar transformer for UWB RFIC applications. *IEEE Electron Device Lett* 2006;27(8):684–688.
- Nelson C, Sigliano R, Makihara C. Built-in passive components in multilayer ceramics for wireless applications. In: *Proceedings of 3rd International Symposium on Advanced Packaging Materials*; Braselton, Georgia, 1997 Mar 9–12. 1997. pp. 75–80.
- Rahn DG, Cavin MS, Dai FF, Fong NHW, Griffith R, Macedo J, Moore D, Rogers JWM, Toner M. A fully integrated multiband MIMO WLAN transceiver RFIC. *IEEE J Solid-State Circ* 2005;40(8):1629–1641.
- Rotella F, Howard D, Racanelli M, Zampardi P. Characterizing and optimizing high Q inductors for RFIC design in silicon processes. In: *2003 IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*; Irvine, CA. Jun 8–10. 2003. pp. 339–342.
- Samavati H, Hajimini A, Shahani AR, Nasserbakht GN, and Lee TH. Fractal capacitors. *IEEE J Solid-State Circ* 1998;33(12):2035–2041.
- Sia CB, Ong BH, Chan KW, Yeo KS, Ma J-G, Do MA. Physical layout design optimization of integrated spiral inductors for silicon-based RFIC applications. *IEEE Trans Electron Devices* 2005;52(12):2559–2567.
- Theerachet S, Simon Wong S. A 0-dB IL 2140+/-30MHz bandpass filter utilizing Q-enhanced spiral inductors in standard CMOS. *IEEE J Solid-State Circ* 2002;37(5):579–586.
- Van Schuylenbergh K, Chua CL, Fork DK, Lu J-P, Griffiths B. On-chip out-of-plane high-Q inductors. In: *Proceedings of IEEE Lester Eastman Conference on High Performance Devices*; Palo Alto, CA. Aug 6–8. 2002. pp. 364–373.
- Wang T, Lin Y-S, Li S-S. An ultralow-loss and broadband micromachined RF inductor for RFIC input-matching applications. *IEEE Trans Electron Device* 2006;53(3):568–570.
- Yodprasit U, Ngarmnil J. Q-enhancing technique for RF CMOS active inductor. In: *ISCAS 2000, IEEE International Symposium on Circuits and Systems*; Geneva, Switzerland; 2000 May 28–31. 2000.
- Yodprasit U, Ngarmnil J. Q-enhancing technique for RF CMOS active inductor. In: *ISCAS 2000, IEEE International Symposium on Circuits and Systems*; Geneva, Switzerland; 2000 May 28–31. 2000.

## EXERCISES

1. What is definition of “isolation between two points” on a PCB or in an IC die?
2. What is definition of “attenuation of the interference between two points” on a PCB or in an IC die?
3. What are the main advantages of the technology moving forward from RF module (built by discrete parts) to RFIC (radio frequency integrated circuit)?

4. What is the greatest barrier in RFIC development?
5. What is the essential difference in the interference path between an RF module and an RFIC?
6. What is the main function of the P+ and deep N-well guard ring in RFIC?
7. More than 70 dB isolation between RF blocks in an RFIC chip has been reached at present. How many decibels of isolation would be satisfactory as the final goal to be pursued?
8. What are the common grounding rules in RFIC?
9. What are the main bottlenecks in RFIC development at present?

## ANSWERS

1. If two points on a PCB or in an IC die are A and B, the isolation between these two points is defined as

$$\text{Isolation} = 10 \log \frac{P_A}{P_B} (\text{dB}).$$

2. If two points on a PCB or in an IC die are A and B, the attenuation of the interference between these two points is defined as

$$\text{Attenuation} = 10 \log \frac{P_B}{P_A} (\text{dB}).$$

3. There are three main advantages of the technology moving forward from RF module (built by discrete parts) to RFIC (radio frequency integrated circuit):

- (a) lower cost;
- (b) smaller size; and
- (c) higher reliability.

4. The greatest barrier in the development of RFIC is the low  $Q$  value of the spiral inductor in an RFIC die.

5. The interference path for an RF module is from the “sky,” whereas the interference path for a RFIC die is from the “ground” (substrate).

6. The main function of the P+ and deep N-well guard ring in RFIC is isolation.

7. Around 120 dB or above.

8. The common grounding rules in RFIC are: be separate, don’t combine! Be in parallel, don’t be in series!

9. The main bottlenecks in RFIC development at present are

- (a) low  $Q$  value of spiral inductor
- (b) on-chip zero capacitor
- (c) modeling of bonding wire and pad, and
- (d) further improvement of isolation.



---

# PART 2

---

---

## RF SYSTEM

---



# MAIN PARAMETERS AND SYSTEM ANALYSIS IN RF CIRCUIT DESIGN

## 12.1 INTRODUCTION

With the exceptions of the oscillator or the VCO (voltage-controlled oscillator), the main task of RF circuit design is to ensure good power transport or power manipulation in most electronic products or systems, such as wireless communication systems, navigation systems, and electronic control systems. Consequently, the main parameters to be seriously concerned with in RF circuit design are as follows:

### 1. Power gain

In RF circuit design, it is very often necessary to intensify a weak signal to an appropriate power level when the signal is passed over an RF block. In other words, the designed RF block must have enough power gain. For instance, in the receiver of a communication system, the signal sensed by the antenna is a very weak carrier modulated with the desired digital data or other signals. The demodulator will be able to detect the desired signal from the carrier only if the power of the weak carrier is intensified to reach or exceed a “conceivable” level. The “conceivable” level depends on the incoming noise power, the type of demodulator, and so on. In a transmitter, the power of the modulated signal must be intensified to become a strong signal so that it can be transmitted by the antenna and then propagated to the receiver, which is located a long distance from the transmitter.

It should be noted that power gain is not equal to voltage gain. It is well known that power is equal to the product of voltage and current, that is,

$$P = v \cdot i = \frac{v^2}{z} = i^2 z, \quad (12.1)$$

where

$P$  = the power of AC or RF signal,

$v$  = the voltage of AC or RF signal,

$i$  = the current of AC or RF signal,

$z$  = the impedance over which the AC or RF signal voltage is dropped or the AC or RF current of signal is flowing through.

Power gain is emphasized in RF circuit design, while voltage gain is emphasized in the digital circuit design. An RF block with a high voltage gain may have very low power gain if its input impedance is very low and its output impedance is very high, or vice versa.

### 2. Noise

Whether the power of an RF signal is strong or weak is relative to the noise power “attached” to the RF signal. If the noise power is higher than the signal power, the signal will be “swallowed” by the noise and become undetectable. For instance, the carrier-to-noise ratio (CNR) at the demodulator input must higher than a threshold value, say, 10 dB. Otherwise, the signal cannot be demodulated and separated from the carrier. Therefore, noise is another important parameter in the RF circuit design.

### 3. Nonlinearity

The third important parameter in RF circuit design concerns the distortion of the weak signal. In order to ensure the fidelity of power transport and manipulation, distortion of the desired signal must be avoided or reduced as much as possible.

Distortion is mainly caused by the nonlinearity of the device and all the interferences in the circuit block. The characteristic of a practical device contains linear and nonlinear portions. Its nonlinear terms produce many harmonics and spurious products which contribute to the distortion of the signal.

Just like noise, the harmonics, spurious products, and interferences in the circuit block are not welcome in an RF circuit design. However, they are essentially different from noise. Noise is a random phenomenon. The frequency spectrum of a “white” noise spreads from minus infinity to plus infinity so that it cannot be effectively filtered out entirely by a practical filter. On the contrary, the spurious products and interferences are definite disturbances with definite frequencies, so that they can be reduced or removed by means of special circuit design schemes.

The effect of noise and spurious products on a receiver is different. Noise is directly related to the sensitivity of the receiver. The sensitivity of a receiver is high if its noise figure is low and vice versa. The spurious products do cause distortion of the signal, but they are not directly related to the sensitivity of a receiver. For instance, a receiver with low noise has high sensitivity. However, it can have either high or low distortion of a voice signal. In other words, a receiver can sense the very weak signal at the antenna, but the purity of the signal after demodulation mainly depends on the distortion. In cases where the distortion is high, a girl’s voice can become a boy’s sound even if the sensitivity of the receiver is high. In cases where the distortion is low, the fidelity of the voice can be kept even if the sensitivity of the receiver is low.

In addition to the three main parameters mentioned above, other important parameters in RF circuit design are as follows:

4. Stability
5. DC power supply
6. Current drain
7. Part count
8. Special parameters in some special RF circuit blocks such as
  - Load-pulling effect to a VCO
  - Phase noise of a VCO
  - PAE (power added efficiency) of a power amplifier
  - Sensitivity of the receiver
  - Output power of the transmitter, and so on.

## 12.2 POWER GAIN

Generally, there are three well-known gains in an electronic circuit block: the power gain, voltage gain, and current gain. The meaning of voltage and current gain is quite straight forward. It is simply the ratio of the input and the output. However, there are different power gains with different meanings.

Power gains can be explained by the signal flow graph introduced in Appendix 12.A.3. Instead of the signal flow graph, we are going to explain power gains through a new concept, namely, reflection power gain. The derivation of reflection power gain seems more intuitive than the derivation by means of the signal flow graph.

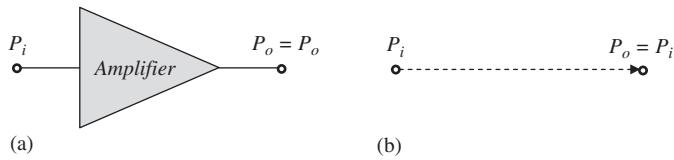
### 12.2.1 Basic Concept of Reflection Power Gain

It is easy to understand that amplification power gain can be obtained from the input to the output of an amplifier due to amplification. However, it may be a little bit hard to understand that reflection power gain can be obtained from the power reflection at a terminal or at a load. Let us introduce the concept of reflection power gain through the concept of amplification power gain.

For an amplifier, the power gain due to amplification is defined as the ratio of the output power and the input power. Usually, this definition is related to two nodes, the input and output nodes of the amplifier, when the amplifier exists. However, this definition of power gain due to amplification can be replaced by a new concept, in which the definition is not related to both input and output nodes but to the output node only. Now, the power gain due to amplification is defined as the ratio of the output power when the amplifier exists to the output power when the amplifier does not exist and the output node is directly connected to the input node. Figure 12.1 illustrates this new definition.

The expression of power gain is kept the same before and after the definition of power gain is replaced by such a new concept, that is,

$$G_{\text{amp}} = \frac{P_o}{P_i}, \text{ or } G_{\text{amp,dB}} = P_o - P_i, \text{ dB} \quad (12.2)$$



**Figure 12.1.** Definition of power gain based on the readings only at output node in two cases when amplifier does or does not exist,  $G = P_o/P_i$ . (a) At the output of the amplifier,  $P_o = P_o$  when amplifier exists. (b) At the output of amplifier,  $P_o = P_i$  when amplifier does not exist.

where

$G_{\text{amp}}$  = the power gain due to amplification, measured in watts (W) or milliwatts (mW), and

$G_{\text{amp,dB}}$  = the power gain due to amplification, measured in decibels (dB).

In terms of this concept, power gain due to reflection can be well defined. Figure 12.2 illustrates the definition of power gain due to reflection based on the readings at the load only.

Similar to the definition of the amplification power gain in expression (12.2), the reflection power gain is defined as the ratio of the delivered power from the source to the load when reflections  $\Gamma_S$  and  $\Gamma_L$  exist to the delivered power from the source to the load when reflections  $\Gamma_S$  and  $\Gamma_L$  do not exist, that is,

$$G_{\text{ref}} = \frac{P_{R_L}|_{\text{with reflection}}}{P_{R_L}|_{\text{without reflection}}}, \quad (12.1)$$

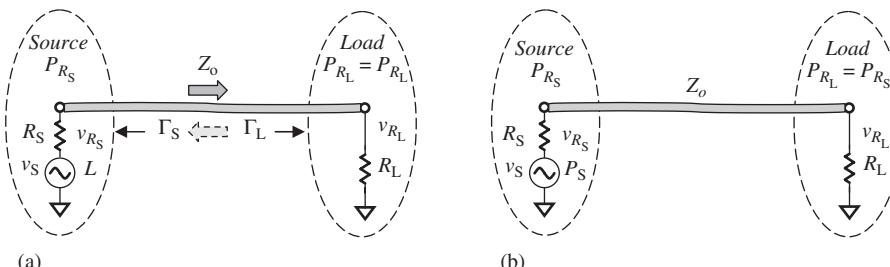
where

$G_{\text{ref}}$  = the power gain due to reflection,

$P_{R_L}|_{\text{with reflections}}$  = the delivered power from the source to the load when reflections exist, and

$P_{R_L}|_{\text{without reflections}}$  = the delivered power from the source to the load when reflections do not exist.

By comparing power gain due to amplification as shown in Figure 12.1 and power gain due to reflection as shown in Figure 12.2, the corresponding relations of these two power gains are listed in Table 12.1.



**Figure 12.2.** Definition of power gain based on the readings at the load in two cases when reflections do or do not exist. (a) At the load,  $P_{R_L} = P_{R_L}$  when reflections exist. (b) At the load,  $P_{R_L} = P_{R_S}$  when reflection do not exist.

TABLE 12.1. Corresponding Relations between Power Gain Due to Amplification and Power Gain Due to Reflection

Power Gain Due to Amplification	Power Gain Due to Reflection
Amplification	Reflection
Input	Source
Output	Load
$P_i$	$P_{R_S}$
$P_o$	$P_{R_L}$

Now let us derive the expressions for  $P_{R_L}|_{\text{with reflections}}$  and  $P_{R_L}|_{\text{without reflections}}$  so that the expression of power gain due to reflection can be described as a function of the reflection coefficients  $\Gamma_S$  and  $\Gamma_L$ .

In cases where reflections exist, recall expressions (2.7), (2.9), (2.11) and (2.14),

$$v_S = v_{S_0} e^{j\omega t}, \quad (2.7)$$

$$v_{R_L}|_{t=T_d} = v_{S_0} \frac{R_L}{Z_S - z_L} (1 - \Gamma_L) \sum_{n=0}^{\infty} e^{j\omega[t-(2n+1)T_d]} (\Gamma_S \Gamma_L)^n, \quad (2.9)$$

If

$$T_d \rightarrow 0, \quad (2.11)$$

then

$$v_{R_L}|_{t=T_d} = v_{S_0} e^{j\omega t} \frac{R_L}{Z_S + Z_L} \frac{1 - \Gamma_L}{1 - \Gamma_S \Gamma_L}, \quad (2.14)$$

$$v_{R_L}|_{t=T_d} = v_{S_0} e^{j\omega t} \frac{R_L}{Z_S + Z_L} (1 - \Gamma_L) + v_{S_0} e^{j\omega t} \frac{R_L}{Z_S + Z_L} (1 - \Gamma_L) \sum_{n=1}^{\infty} (\Gamma_S \Gamma_L)^n. \quad (2.15)$$

where  $T_d$  is the delay while the delivered signal travels from the source to the load.

Without loss of generality, for simplicity let us assume that the impedances of both the source and the load are pure resistance. Then, Figure 2.1 can be redrawn as Figure 12.2(a) and expression (2.15) can be copied as (12.4)

$$v_{R_L}|_{t=T_d} = v_{S_0} e^{j\omega t} \frac{R_L}{R_S + R_L} \frac{1}{1 - \Gamma_S \Gamma_L} - v_{S_0} e^{j\omega t} \frac{R_L}{R_S + R_L} \frac{\Gamma_L}{1 - \Gamma_S \Gamma_L}. \quad (12.2)$$

It should be noted that the delivered voltage from the source to the load  $v_{R_L}$  in expression (12.4) contains two terms. The first term is the incident voltage and the second is the reflected voltage. Their corresponding incident and reflected powers are

$$\begin{aligned} P_{R_L}|_{\text{with-reflection}} &= v_{S_0}^2 e^{j2\omega t} \frac{R_L}{(R_S + R_L)^2} \frac{1}{|1 - \Gamma_S \Gamma_L|^2} \\ &- v_{S_0}^2 e^{j2\omega t} \frac{R_L}{(R_S + R_L)^2} \frac{|\Gamma_L|^2}{|1 - \Gamma_S \Gamma_L|^2}, \end{aligned} \quad (12.3)$$

In expression (12.5), the first term is the incident power to the load and the second term is the reflected power from the load. They can be combined together as

$$P_{R_L} \mid_{\text{with-reflection}} = v^2 S_0 e^{j2\omega t} \frac{R_L}{(R_S + R_L)^2} \frac{1 - |\Gamma_L|^2}{|1 - \Gamma_S \Gamma_L|^2}. \quad (12.4)$$

Obviously, for the cases when the reflections do not exist, that is,  $\Gamma_S = \Gamma_L = 0$ , then

$$P_{R_L} \mid_{\text{without-reflection}} = v^2 S_0 e^{j2\omega t} \frac{R_L}{(R_S + R_L)^2}. \quad (12.5)$$

Substituting (12.6) and (12.7) into (12.3), we have

$$G_{\text{ref}} = \frac{1 - |\Gamma_L|^2}{|1 - \Gamma_S \Gamma_L|^2}. \quad (12.6)$$

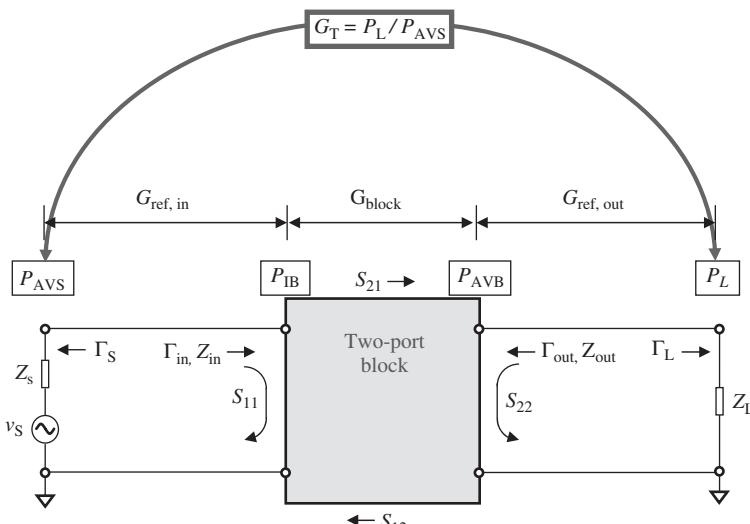
From expression (12.8) it can be seen that the reflection power gain can be positive if both  $\Gamma_L$  and  $\Gamma_S$  lie between 0 and 1. This is not surprising because the power on the load in the case with reflection can be higher than that in the case without reflection at the load but with reflection at the source.

### 12.2.2 Transducer Power Gain

Figure 12.3 shows the various powers, power gains, voltage reflection coefficients, and S-parameters in a two-port block.

In Figure 12.3, there are four powers:

1.  $P_{IB}$  = the power input to the block,
2.  $P_{AVS}$  = the power available from the source,
3.  $P_L$  = the power delivered to the load, and
4.  $P_{AVB}$  = the power available from the block.



**Figure 12.3.** Various powers, power gains, voltage reflection coefficients, and S-parameters in a two-port block.

The transducer power gain is the power gain from the source to the load in cases where neither the input impedance nor the output impedance is matched. As shown in Figure 12.3, the transducer power gain is defined as the ratio of the power delivered to the load  $P_L$  and power available from the source  $P_{AVS}$ . It consists of three portions, that is,

$$G_T = \frac{P_L}{P_{AVS}} = G_{\text{ref,in}} \cdot G_{\text{block}} \cdot G_{\text{ref,out}}. \quad (12.7)$$

According to the discussion in Section 12.2.1, the reflection power gains at the input and output portions are

$$G_{\text{ref,in}} = \frac{1 - |\Gamma_{in}|^2}{|1 - \Gamma_S \Gamma_{in}|^2}, \quad (12.8)$$

and

$$G_{\text{ref,out}} = \frac{1 - |\Gamma_L|^2}{|1 - S_{22} \Gamma_L|^2}, \quad (12.9)$$

respectively.

Instead of  $\Gamma_{out}$ ,  $S_{22}$  in expression (12.11) represents the reflection at the output of the block when the incident power is coming from the load.  $S_{22}$  in expression (12.11) looks just like  $\Gamma_S$  in expression (12.10).

The power gain  $G_{\text{block}}$  is obviously

$$G_{\text{block}} = |S_{21}|^2. \quad (12.10)$$

Substituting expressions (12.10) to (12.12) into (12.9), we have

$$G_T = \frac{P_L}{P_{AVS}} = \frac{1 - |\Gamma_S|^2}{|1 - \Gamma_{in} \Gamma_S|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22} \Gamma_L|^2}, \quad (12.11)$$

or

$$G_T = \frac{P_L}{P_{AVS}} = \frac{1 - |\Gamma_S|^2}{|1 - S_{11} \Gamma_S|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - \Gamma_{out} \Gamma_L|^2} \quad (12.12)$$

because

$$\frac{|1 - S_{11} \Gamma_S|^2}{|1 - S_{22} \Gamma_L|^2} = \frac{|1 - \Gamma_{in} \Gamma_S|^2}{|1 - \Gamma_{out} \Gamma_L|^2}. \quad (12.13)$$

In addition to the transducer power gain  $G_T$ , the operating power gain  $G_P$  is defined as a ratio of the power delivered to the load  $P_L$  and the power input to the block  $P_{IB}$ , that is,

$$G_P = \frac{P_L}{P_{IB}} = G_T|_{\Gamma_{in}=\Gamma_S^*} = \frac{1}{1 - |\Gamma_{in}|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22} \Gamma_L|^2}, \quad (12.14)$$

because

$$P_{IB} = P_{AVS}|_{\Gamma_S^*=\Gamma_{in}}. \quad (12.15)$$

when

$$\Gamma_S^* = \Gamma_{in}. \quad (12.16)$$

The third power gain, the available power gain  $G_A$ , is defined as the ratio of the power available from the block,  $P_{AVB}$ , to the power available from the source,  $P_{AVS}$ , that is,

$$G_A = \frac{P_{AVB}}{P_{AVS}} = G_T|_{\Gamma_{out}=\Gamma_L^*} = \frac{1 - |\Gamma_S|^2}{|1 - S_{11}\Gamma_S|^2} |S_{21}|^2 \frac{1}{1 - |\Gamma_{out}|^2}. \quad (12.17)$$

Because

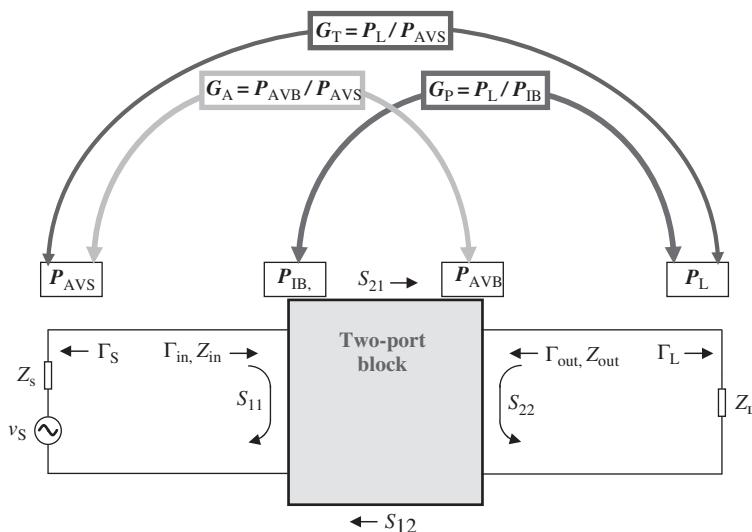
$$P_{AVB} = P_L|_{\Gamma_{out}=\Gamma_L^*}, \quad (12.18)$$

when

$$\Gamma_{out} = \Gamma_L^*. \quad (12.19)$$

Figure 12.4 shows the relationships between the four powers,  $P_{IB}$ ,  $P_{AVS}$ ,  $P_L$ , and  $P_{AVB}$ , and the three power gains  $G_T$ ,  $G_P$ , and  $G_A$ .

The transducer power gain  $G_T$  describes the power gain of a two-port block in a general form. The operating power gain  $G_P$  is a special case of the transducer power gain  $G_T$  when the input impedance of the block is well matched. The available power gain  $G_A$  is a special case of the transducer power gain  $G_T$  when the output impedance of the block is well matched. In general, the transducer power gain  $G_T$  is lower than  $G_P$  or  $G_A$ .



**Figure 12.4.** Transducer power gain, operating power gain, and available power gain in a two-port block.

### 12.2.3 Power Gain in a Unilateral Case

In general cases, the relations of the various voltage reflection coefficients  $\Gamma_S$ ,  $\Gamma_L$ ,  $\Gamma_{in}$ , and  $\Gamma_{out}$  in a two-port block are

$$\Gamma_{in} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L}, \quad (12.20)$$

$$\Gamma_{out} = S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1 - S_{11}\Gamma_S}. \quad (12.21)$$

When voltage reflections at the source and the load of a two-port block exist, that is,

$$\Gamma_S \neq 0, \quad (12.22)$$

and

$$\Gamma_L \neq 0; \quad (12.23)$$

then

$$\Gamma_{in} \neq S_{11}, \quad (12.24)$$

and

$$\Gamma_L \neq S_{22}, \quad (12.25)$$

where

$\Gamma_S$  = the voltage reflection coefficient at the source,

$\Gamma_L$  = the voltage reflection coefficient at the load,

$\Gamma_{in}$  = the voltage reflection coefficient at the input of the two-port block, and

$\Gamma_{out}$  = the voltage reflection coefficient at the output of the two-port block.

In the special case when the block becomes *unilateral*, that is, when

$$S_{12} = 0; \quad (12.26)$$

then

$$\Gamma_{in} = S_{11}, \quad (12.27)$$

$$\Gamma_{out} = S_{22}, \quad (12.28)$$

and then expressions (12.13) and (12.14) become

$$G_T = G_{T,S_{12}=0} = G_S G_o G_L = \frac{1 - |\Gamma_S|^2}{|1 - S_{11}\Gamma_S|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2}, \quad (12.29)$$

where

$$G_S = \frac{1 - |\Gamma_S|^2}{|1 - S_{11}\Gamma_S|^2}, \quad (12.30)$$

$$G_o = |S_{21}|^2, \quad (12.31)$$

$$G_L = \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2}. \quad (12.32)$$

It can be seen that the transducer power gain  $G_T$  is composed of three different independent power gains when  $S_{12} = 0$ .

In the unilateral case, the operating power gain and available power gain become

$$G_P = G_{P,S_{12}=0} = \frac{1}{1 - |S_{11}|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2}, \quad (12.33)$$

and

$$G_A = G_{A,S_{12}=0} = \frac{1 - |\Gamma_S|^2}{|1 - S_{11}\Gamma_S|^2} |S_{21}|^2 \frac{1}{1 - |S_{22}|^2}, \quad (12.34)$$

respectively.

#### 12.2.4 Power Gain in a Unilateral and Impedance-Matched Case

Furthermore, let us examine an ideal case in which the block is not only unilateral but also reflection-free at both the source and the load, that is,

$$S_{12} = 0, \quad (12.35)$$

and

$$\Gamma_S = \Gamma_L = 0, \quad (12.36)$$

which are the conditions of impedance matching. Under these conditions, we have

$$G_S = G_{S,S_{12}=0,\Gamma_S=\Gamma_L=0} = 1, \quad (12.37)$$

$$G_o = G_{o,S_{12}=0,\Gamma_S=\Gamma_L=0} = |S_{21}|^2, \quad (12.38)$$

$$G_L = G_{L,S_{12}=0,\Gamma_S=\Gamma_L=0} = 1, \quad (12.39)$$

$$G_P = G_{P,S_{12}=0,\Gamma_S=\Gamma_L=0} = \frac{1}{1 - |S_{11}|^2} |S_{21}|^2, \quad (12.40)$$

$$G_A = G_{A,S_{12}=0,\Gamma_S=\Gamma_L=0} = \frac{1}{1 - |S_{22}|^2} |S_{21}|^2, \quad (12.41)$$

$$G_T = G_{T,S_{12}=0,\Gamma_S=\Gamma_L=0} = |S_{21}|^2. \quad (12.42)$$

Design engineers usually present their simulation results of  $|S_{21}|^2$  as power gain. This implies that it is a transducer power gain  $G_o$  not only in the unilateral case  $S_{12} = 0$  but also in cases without reflection at both the source and the load, that is,  $\Gamma_S = \Gamma_L = 0$ .

### 12.2.5 Power Gain and Voltage Gain

In a digital circuit design, the main task is to realize the digital status transportation. Engineers always try to reduce power consumption as much as possible. Power gain is not important in digital circuit design. Instead, the voltage gain is the important parameter, by which the status transportation is executed.

On the other hand, in an RF/RFIC circuit design, the main task is to realize power transport. Consequently, power gain is of constant concern. The voltage gain is meaningless until the corresponding impedance is specified.

However, as long as the input and output impedances are given, one can convert the voltage gain to power gain through the well-known relation

$$G_P = \frac{\frac{V_{\text{out}}^2}{Z_{\text{out}}}}{\frac{V_{\text{in}}^2}{Z_{\text{in}}}} = G_v^2 \frac{Z_{\text{in}}}{Z_{\text{out}}}, \quad (12.43)$$

where  $G_P$  and  $G_v$  are the power and voltage gains, respectively.

### 12.2.6 Cascaded Equations of Power Gain

In the system analysis for multiple RF blocks or an RF subsystem, the system power gain is contributed by the gain of individual RF blocks, which can be calculated in terms of so-called cascaded equations of power gain.

Let us derive the cascaded equation for power gain of a system containing only two blocks as shown in Figure 12.5.

The power gain  $G_k$  shown in Figure 12.5 represents the transducer power gain  $G_{T,k}$ , that is,

$$G_k = G_{T,k} = \frac{P_{L,k}}{P_{S,k}}, \quad (12.44)$$

$$G_{k+1} = G_{T,k+1} = \frac{P_{L,k+1}}{P_{S,k+1}}. \quad (12.45)$$

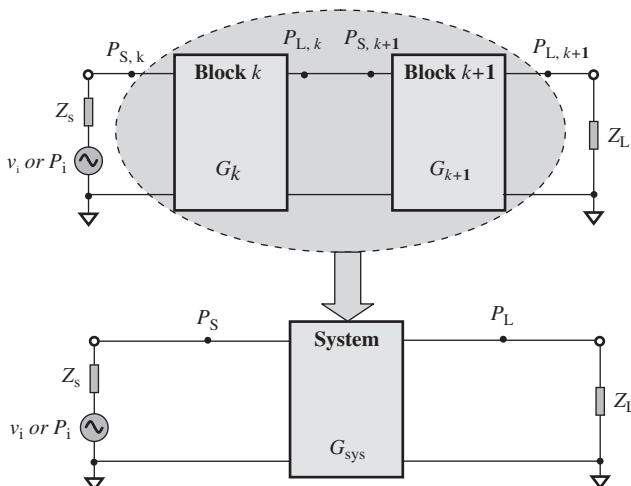


Figure 12.5. Power gain of a system cascaded by two block  $k$  and  $k + 1$ .

Assuming that the input and output impedances between blocks are well matched, then the system power gain  $G_{\text{SYS}}$  should be

$$G_{\text{SYS}} = \frac{P_{L,k+1}}{P_{S,k}} = \frac{P_{L,k} P_{L,k+1}}{P_{S,k} P_{L,k}} = G_k G_{k+1}, \quad (12.46)$$

where

- $P_{L,k}$  = the power delivered to the load of the  $k$ th block,
- $P_{L,k+1}$  = the power delivered to the load of the  $(k + 1)$ th block,
- $P_{S,k}$  = the power available from the source of the  $k$ th block,
- $P_{S,k+1}$  = the power available from the source of the  $(k + 1)$ th block,
- $G_k$  = the power gain of the  $k$ th block,
- $G_{T,k}$  = the transducer power gain of the  $k$ th block,
- $G_{k+1}$  = the power gain of the  $(k + 1)$ th block,
- $G_{T,k+1}$  = the transducer power gain of the  $(k + 1)$ th block, and
- $G_{\text{sys}}$  = the power gain of the system in the numeric scale.

In the logarithmic form or in units of dB, equation (12.48) becomes

$$G_{\text{SYS,dB}} = G_k + G_{k+1}. \quad (12.47)$$

The system gain  $G_{\text{SYS}}$  is a simple product of the individual gains  $G_k$  and  $G_{k+1}$  in the numeric scale. In other words, the system gain  $G_{\text{SYS,dB}}$  is a simple sum of the individual gains  $G_k$  and  $G_{k+1}$  in the logarithmic scale.

Figure 12.6 shows a general case of a system containing  $n$  blocks; equations (12.48) and (12.49) can, in this case, be extended to

$$G_{\text{SYS}} = G_1 G_2 G_3 \dots G_n, \quad (12.48)$$

$$G_{\text{SYS,dB}} = G_1 + G_2 + G_3 + \dots + G_n, \quad (12.49)$$

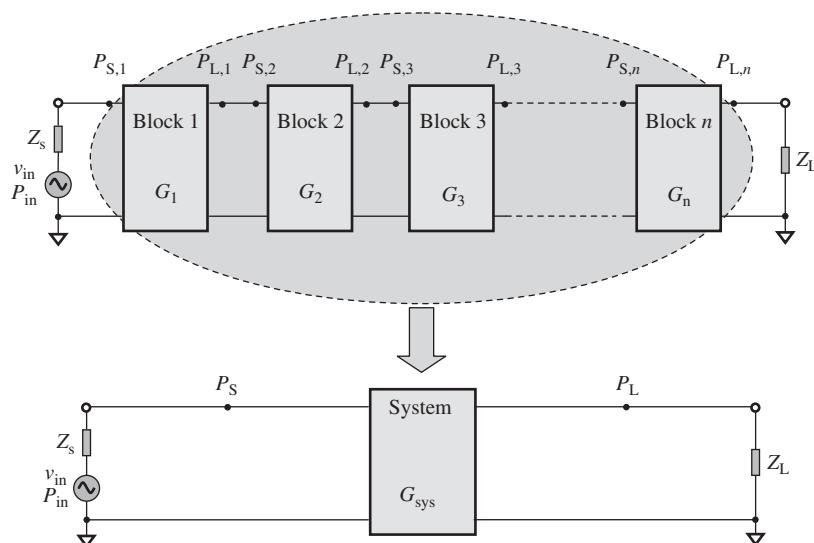


Figure 12.6. Power gain of a system cascaded by  $n$  blocks.

or

$$G_{\text{SYS}} = \prod_1^n G_k, \quad (12.50)$$

$$G_{\text{SYS,dB}} = \sum_1^n G_k. \quad (12.51)$$

Equations (12.48) to (12.53) are called the *cascaded equations* of power gain.

## 12.3 NOISE

Noise inherently exists in the active or passive parts of an RF circuit. In the microelectrical world, the charge amount of the main electric carriers, that is, electrons and the positively charged holes, is not a continuous quantity, and the motion of the carriers is a chaotic or random process in the time domain. The macro parameter that describes the random process always contains two portions. One part represents its average value, while the other represents its random fluctuation from its average value. The random fluctuation can be described by its root-mean-square value or variance. For example, an electric current always contains two portions. One is the current with its average value, which we are familiar with. The other is the noise current, which is always neglected in fields where such a tiny fluctuation of current is not important. In the electronic circuit design field, the random fluctuation of the charge carriers is one of the main sources of noise and should never be neglected. Noise current, noise voltage, or noise power is one of the most important parameters in the performance of a block or a system.

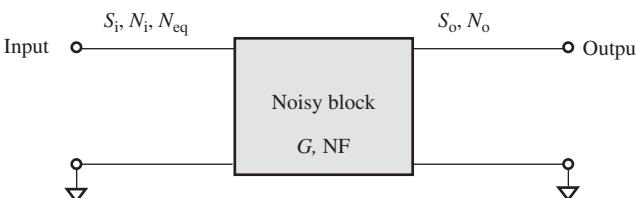
In a circuit or a system, another source of noise lies outside its environment. When a signal is transported or operated in a circuit block or system, it is inevitably interfered with or imposed on by noise that radiates from other blocks or systems. This topic is outside the limited topics covered in this book and will not be discussed here; readers who are interested in this subject can refer to other specific books or papers.

Noise degrades the performance of the signal. One of the main tasks in circuit design is to reduce or suppress the noise produced by the circuit itself or from the environment outside so as to ensure a good performance of the signal.

### 12.3.1 Significance of Noise Figure

Figure 12.7 shows various parameters in the input and output of a practical RF circuit block.

In reality, a noise-free circuit block does not exist. All practical blocks are noisy blocks: a practical circuit block always produces an additional part of noise in addition to



**Figure 12.7.** Various parameters related to the definition of noise figure.

the signal. Therefore, its total output noise power is greater than the output noise power due to input noise, that is,

$$N_o = GN_i + \Delta N, \quad (12.52)$$

where

$\Delta N$  = the additional noise power produced by the circuit block, and  
 $N_i, N_o$  = the noise power at the input and output, respectively.

The additional noise means that the increase of noise is positive:

$$\Delta N > 0. \quad (12.53)$$

On the other hand, the input signal is intensified by the RF circuit block to the output signal by the power gain factor  $G$ , which is transducer power gain  $G_T$  usually, that is,

$$S_o = GS_i, \quad (12.54)$$

where

$S_i, S_o$  = the signal power at input and output respectively, and  
 $G = G_T$  = the transducer power gain of the circuit block.

From expressions (12.54) and (12.56), the ratio of the output signal to the output noise is

$$\frac{S_o}{N_o} = \frac{GS_i}{GN_i + \Delta N} = \frac{S_i}{N_i + \frac{\Delta N}{G}} < \frac{S_i}{N_i}. \quad (12.55)$$

It can be seen that the additional noise therefore reduces the ratio of the signal to noise from the the input to the output.

At any point of a circuit block, the signal can be detected only if the power of the signal is greater than the power of the noise by a certain amount. The power ratio of the signal to noise is a measure of the possibility of detecting a signal in a practical circuit.

Instead of  $\Delta N$ , a new parameter called the NF (noise figure) is defined as

$$NF = \frac{\frac{S_o}{N_o}}{\frac{S_o}{N_o} - 1}. \quad (12.56)$$

By the substitution of relation (12.56) into (12.58), we have

$$NF = \frac{N_o}{GN_i} = \frac{N_o}{N_{o,i}}, \quad (12.57)$$

where

$N_{o,i} = GN_i$  = the output noise power portion due to the input noise.

From expressions (12.54) and (12.59),

$$NF = \frac{GN_i + \Delta N}{GN_i}. \quad (12.58)$$

In other words,

$$NF = \frac{\text{Total output noise power}}{\text{Output noise power due to the input noise}}. \quad (12.59)$$

The expressions from (12.58) to (12.61) have the same meaning: Noise figure of an RF circuit block represents the additional noise produced from the RF circuit block itself.

In an ideal case when the circuit block is noise free, that is, the additional noise is zero, then from the definition of the noise figure from (12.58) to (12.61),

$$\frac{S_o}{N_o} = \frac{S_i}{N_i}, \quad (12.60)$$

$$N_o = G N_i = N_{o,i}, \quad (12.61)$$

or

$$NF = 1 \text{ (or } 0 \text{ dB).} \quad (12.62)$$

An NF of 1 indicates that the input ratio of the signal to noise is equal to the output ratio of the signal to noise. In other words, the total output noise is equal to only the output noise caused by the input noise power. In short, there is no additional noise existing in the circuit block if the noise figure of the circuit block is 1, or 0 dB.

An ideal or noise-free circuit block never exists in practical RF circuit blocks. Therefore, in a practical RF circuit block, the ratio of signal to noise (S/N), the noise, and the noise figure are always

$$\frac{S_o}{N_o} < \frac{S_i}{N_i}, \quad (12.63)$$

$$N_o > G N_i = N_{o,i}, \quad (12.64)$$

or

$$NF > 1 \text{ (or } 0 \text{ dB).} \quad (12.65)$$

### 12.3.2 Noise Figure in a Noisy Two-Port RF Block

Based on Haus' theory (Haus et al., 1960), the noise figure of a noisy block can be expressed by

$$NF = NF_{\min} + \frac{R_n}{G_S} [(G_S - G_{S,\text{opt}})^2 + (B_S - B_{S,\text{opt}})^2], \quad (12.66)$$

where

$$Y_S = G_S + jB_S, \quad (12.67)$$

$$Y_{S,\text{opt}} = G_{S,\text{opt}} + jB_{S,\text{opt}}, \quad (12.68)$$

and

- $NF$  = the noise figure of the noisy block,
- $NF_{min}$  = the minimum noise figure of the noisy block,
- $R_n$  = the equivalent noise resistance,
- $Y_S$  = the admittance of the input source,
- $G_S$  = the conductance of the input source,
- $B_S$  = the susceptance of the input source,
- $Y_{S,opt}$  = the optimum admittance of the input source,
- $G_{S,opt}$  = the optimum conductance of the input source, and
- $B_{S,opt}$  = the optimum susceptance of the input source.

The profound significance of this equation is that there is an optimum admittance of the input source existing in a noisy RF block when

$$Y_S = Y_{S,opt}, \quad (12.69)$$

or when

$$G_S = G_{S,opt}, \quad (12.70)$$

$$B_S = B_{S,opt}. \quad (12.71)$$

Under the optimum condition (12.71), the noisy two-port block reaches a minimum of noise figure, that is,

$$NF = NF_{min}. \quad (12.72)$$

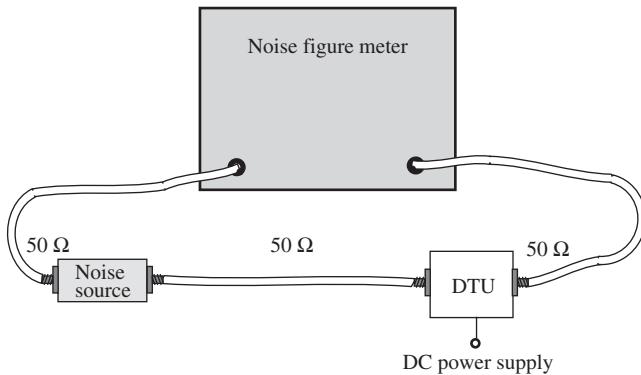
The expressions of optimum admittance  $Y_{S,opt}$  or  $G_{S,opt}$ ,  $B_{S,opt}$ , the equivalent noise resistance  $R_n$ , and the minimum of noise figure  $NF_{min}$  have been derived and can be found in the references (Gray et al., 2001; Lee, 1998).

### 12.3.3 Notes on Noise Figure Testing

In the low RF frequency range, say, below the ultrahigh frequency (UHF) range, the noise figure of an RF circuit block is directly measured by the noise figure meter as shown in Figure 12.8. Special attention must be paid to the setup adjustments, such as given below:

- Tests must be conducted in a shielded room to shield the circuit from the RF signal.
- All the electric lights and DC power supplies, except for the DC power supplies of the DTU (desired test unit) and the noise figure meter, must be turned off, as they produce a lot of noise at low as well as high frequencies.
- All the cables must be placed in appropriate positions so that the reading of the noise figure in the noise figure meter is at a minimum.

In the high RF frequency range, a noise figure meter is not available. An alternative method is to measure the noise figure by a network analyzer. This is not as simple as measurement by a noise figure meter, but the test is not impacted seriously by the environment as long as the network analyzer is well calibrated and the noise floor is taken care of.



**Figure 12.8.** Noise figure of an RF block directly measured by a noise figure meter.

### 12.3.4 An Experimental Method to Obtain Noise Parameters

Through experiments, one can obtain the noise parameters of an RF block.

Let us recall expression (12.68):

$$NF = NF_{min} + \frac{R_n}{G_s} [(G_s - G_{s,opt})^2 + (B_s - B_{s,opt})^2]. \quad (12.68)$$

There is a remarkable significance in this expression of the noise figure. It indicates that for any noisy two-port circuit block, a minimum of the noise figure can be obtained as long as the source admittance  $Y_s = G_s + jB_s$  is adjusted to its optimum values  $Y_{s,opt} = G_{s,opt} + jB_{s,opt}$ .

Mathematically, this equation can be considered as the noise figure  $NF$  being a function of the variables  $G_s$  and  $B_s$ , with four noise parameters,  $NF_{min}$ ,  $R_n$ ,  $G_{s,opt}$ , and  $B_{s,opt}$ . By means of testing the noise figure  $NF$  four times with four different source admittances  $G_s$  and  $B_s$ , four equations can be established on the basis of expression (12.68) and the four unknown parameters  $NF_{min}$ ,  $R_n$ ,  $G_{s,opt}$ , and  $B_{s,opt}$  can be solved for. This is called the *four-point experimental method* of noise figure testing.

$$NF_1 = NF_{min} + \frac{R_n}{G_{s1}} [(G_{s1} - G_{s,opt})^2 + (B_{s1} - B_{s,opt})^2], \quad (12.73)$$

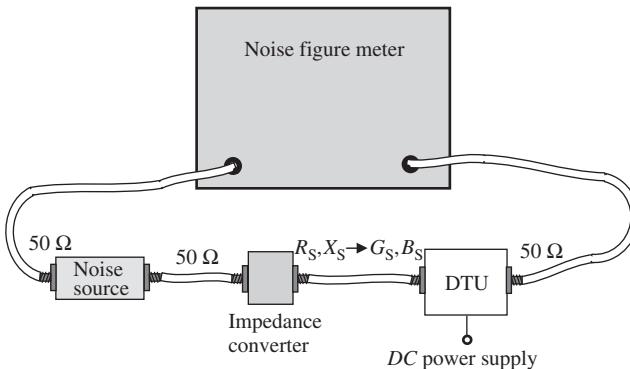
$$NF_2 = NF_{min} + \frac{R_n}{G_{s2}} [(G_{s2} - G_{s,opt})^2 + (B_{s2} - B_{s,opt})^2], \quad (12.74)$$

$$NF_3 = NF_{min} + \frac{R_n}{G_{s3}} [(G_{s3} - G_{s,opt})^2 + (B_{s3} - B_{s,opt})^2], \quad (12.75)$$

$$NF_4 = NF_{min} + \frac{R_n}{G_{s4}} [(G_{s4} - G_{s,opt})^2 + (B_{s4} - B_{s,opt})^2], \quad (12.76)$$

where the subscripts 1, 2, 3, and 4 denote the four sets of tested  $NF$  and the variables  $G_s$  and  $B_s$  in each test, respectively.

The test setup is shown in Figure 12.9. The output impedance of the noise source is converted to  $Z_S = R_S + jX_S$  via an impedance converter. In order to calculate the four



**Figure 12.9.** Setup for four-point experimental method of noise figure testing.

noise parameters via equations (12.75)–(12.78), the impedance  $Z_s$  must be converted to admittance  $Y_s$  through the following relations:

$$Y_s = G_s + jB_s = \frac{1}{Z_s} = \frac{1}{R_s + jX_s} = \frac{R_s - jX_s}{R_s^2 + X_s^2}, \quad (12.77)$$

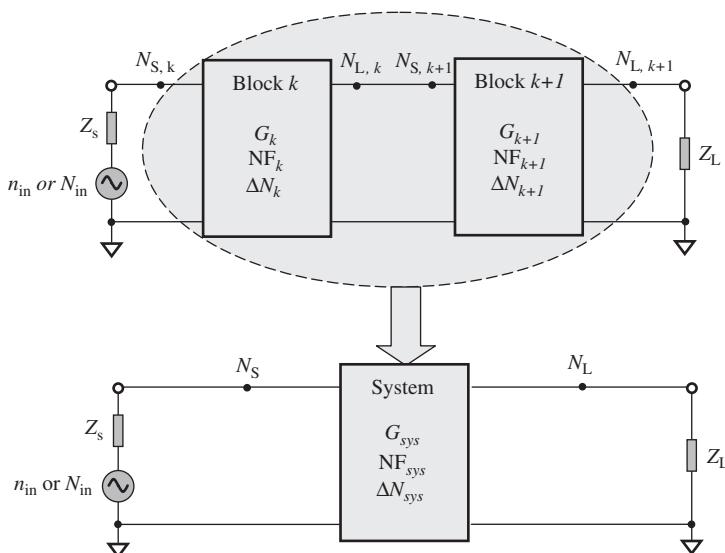
$$G_s = \frac{R_s}{R_s^2 + X_s^2}, \quad (12.78)$$

$$B_s = -\frac{X_s}{R_s^2 + X_s^2}. \quad (12.79)$$

Once these four parameters, that is,  $NF_{min}$ ,  $R_n$ ,  $G_{s,opt}$ , and  $B_{s,opt}$ , are known, we can calculate the noise figure for the case with any source admittance.

### 12.3.5 Cascaded Equations of Noise Figure

Let us derive the cascaded equation for noise figure of a system containing only two blocks as shown in Figure 12.10, where  $G_k$  and  $G_{k+1}$  are the power gains of blocks  $k$  and  $k+1$ , respectively, and  $NF_k$  and  $NF_{k+1}$  are their noise figures, respectively.



**Figure 12.10.** Noise figure of a system cascaded by two blocks  $k$  and  $k+1$ .

By the definition of the noise figure, the resulting noise figure is the ratio of the total output power to the output power due to the input noise power, that is,

$$\text{NF}_k = \frac{G_k N_i + \Delta N_k}{G_k N_i} = 1 + \frac{\Delta N_k}{G_k N_i}, \quad (12.80)$$

$$\text{NF}_{k+1} = \frac{G_{k+1} N_i + \Delta N_{k+1}}{G_{k+1} N_i} = 1 + \frac{\Delta N_{k+1}}{G_{k+1} N_i}. \quad (12.81)$$

Then,

$$\text{NF}_{\text{SYS}} = \frac{G_{k+1}(G_k N_i + \Delta N_k) + \Delta N_{k+1}}{G_k G_{k+1} N_i} = 1 + \frac{G_{k+1} \Delta N_k + \Delta N_{k+1}}{G_k G_{k+1} N_i}, \quad (12.82)$$

where

- $N_i$  = the noise source power delivered to the  $k$ th block,
- $\Delta N_k$  = the additional noise power of the  $k$ th block,
- $\Delta N_{k+1}$  = the additional noise power of the  $(k+1)$ th block,
- $\Delta N_{\text{SYS}}$  = the additional noise power of the system,
- $G_k$  = the power gain of the  $k$ th block,
- $G_{k+1}$  = the power gain of the  $(k+1)$ th block,
- $\text{NF}_k$  = the noise figure of the  $k$ th block,
- $\text{NF}_{k+1}$  = the noise figure of the  $(k+1)$ th block, and
- $\text{NF}_{\text{SYS}}$  = the noise figure of system in the numeric scale.

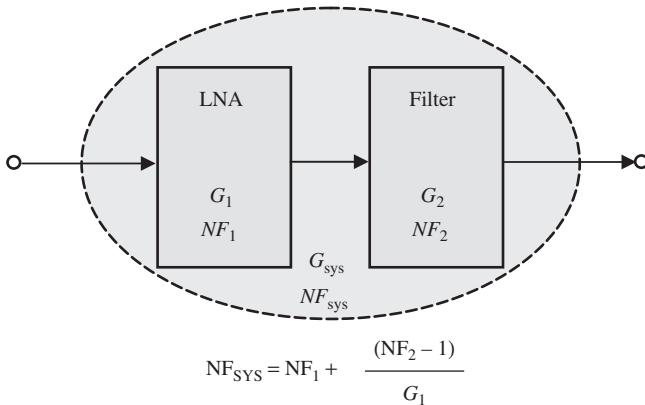
Substituting (12.82) and (12.83) into (12.84), we have

$$\text{NF}_{\text{SYS}} = 1 + \frac{G_{k+1}(\text{NF}_k - 1)G_k N_i + (\text{NF}_{k+1} - 1)G_{k+1} N_i}{G_k G_{k+1} N_i}, \quad (12.83)$$

$$\text{NF}_{\text{SYS}} = \text{NF}_k + \frac{(\text{NF}_{k+1} - 1)}{G_k}. \quad (12.84)$$

An important feature of the noise figure can be found from equation (12.86). The system noise figure or resulting noise figure  $\text{NF}_{\text{SYS}}$  consists of two terms. The first term  $\text{NF}_k$  is contributed by the first block  $k$ . The second term  $(\text{NF}_{k+1} - 1)/G_k$ , is contributed by the second block  $k+1$ , but it is reduced by a factor  $G_k$ . This means that, from the viewpoint of the system, the noise figure of the first block plays a more important role than that of the second one, and that the noise figure of the second block can be reduced by the gain of first block if the gain of the first block is positive and appreciable. In a receiver, in order to minimize the noise figure of the system or to enhance the sensitivity of a receiver, the first objective is to design the first block LNA (low-noise amplifier) with a low noise figure but high gain. This is the main goal for the LNA design in a receiver because the LNA is always located in the front end of the receiver.

For example, there are two RF blocks comprising the LNA and a FLT (Filter) cascaded together as shown in Figure 12.11. Table 12.2 lists design goals in the two examples. The noise figure of first block LNA is the same (2.5 dB), while the insertion loss is different in both examples. The 7-dB insertion loss of the filter in the second example is higher than the 3.5-dB insertion loss of the filter in first example. The result is that the system noise figure is the same (2.67 dB) in both examples. This is not surprising because the gain of the LNA block in the second example (17.4 dB) is 4.9 dB higher than that in first example (12.5 dB).



**Figure 12.11.** System noise figure if a system is cascaded by block 1 and 2.

**TABLE 12.2.** System Noise Figure Contributed from Two Blocks with Different Noise Figures and Gain

	LNA (Block 1)	FLT (Block 2)	System
Example 1			
$NF_{sys}, \text{dB}$	2.5	3.5	2.67
$G_{sys}, \text{dB}$	12.5	-3.5	9.00
Example 2	LNA (Block 1)	FLT(Block 2)	System
$NF_{sys}, \text{dB}$	2.5	7	2.67
$G_{sys}, \text{dB}$	17.4	-7	10.40

In the general case, if a system contains  $n$  blocks as shown in Figure 12.12, equation (12.91) can be extended to

$$NF_{sys,watt} = NF_1 + \frac{(NF_2 - 1)}{G_1} + \frac{(NF_3 - 1)}{G_1 G_2} + \frac{(NF_4 - 1)}{G_1 G_2 G_3} + \cdots + \frac{(NF_n - 1)}{G_1 G_2 \dots G_{n-1}}, \quad (12.85)$$

or

$$NF_{sys,watt} = NF_1 + \sum_2^n \frac{(NF_k - 1)}{\prod_1^{k-1} G_j}. \quad (12.86)$$

Expressions (12.87) or (12.88) are the cascaded equations of the noise figure.

### 12.3.6 Sensitivity of a Receiver

Figure 12.13 shows the basic parameters applied in the sensitivity testing of a receiver.

There is an equivalent noise source  $\sqrt{e_{nt}^2}$  and a signal source  $E_i$  in Figure 12.13. The input equivalent noise source is the thermal noise of the resistor  $R_i$ . Referring to Appendix 12.A.4, the average square of its voltage is

$$\overline{e_{nt}^2} = 4kTR_i\Delta f, \quad (12.87)$$

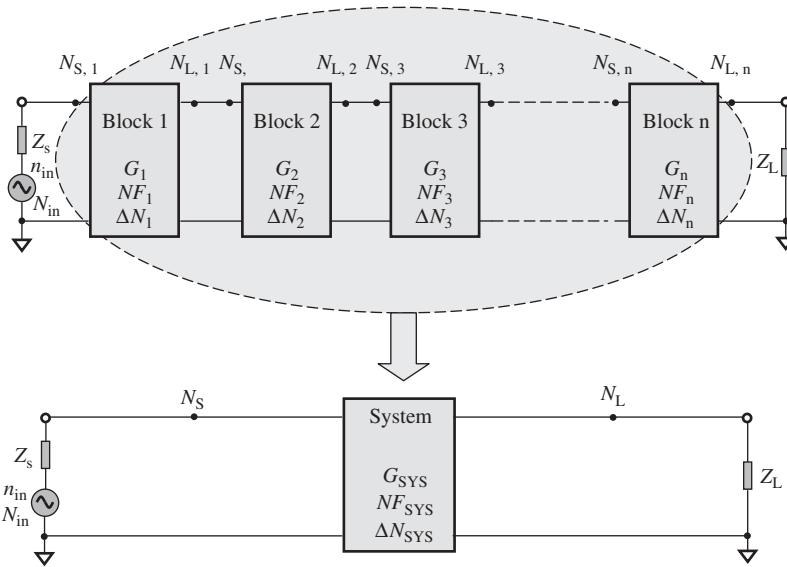


Figure 12.12. Noise figure of a system cascaded by  $n$  blocks.

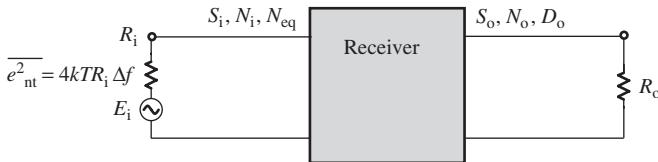


Figure 12.13. Basic parameters applied to the sensitivity testing of a receiver.

where

$e_{nt}$  = the random voltage of the equivalent noise source,

$k$  = the Boltzmann constant,  $1.38 \times 10^{-23}$  J/K,

$T$  = the absolute temperature,

$R_i$  = the internal resistance of input signal source, and

$\Delta f$  = the operating frequency bandwidth.

The input signal source  $E_i$  is usually expressed by the voltage in units of  $\mu\text{V}$  (microvolts), or sometimes by the power with units of  $\text{dB}_m$ .

If these two sources are impedance-matched with the input impedance of the receiver, their powers at the input of the receiver are

$$N_i = \frac{\overline{e_{nt}^2}}{4R_i} = kT \Delta f, \quad (12.88)$$

$$S_i = \frac{E_i^2}{4R_i}. \quad (12.89)$$

A new parameter  $N_{eq}$  is defined as an equivalent input power, which becomes the output noise power after it is intensified by  $G$  times, that is,

$$N_{eq} = \frac{N_o}{G}. \quad (12.90)$$

Substituting  $N_o$  from (12.59) into (12.92), we have

$$N_{eq} = NF \cdot N_i. \quad (12.91)$$

At the output of the receiver, test equipment such as the audio distortion meter are able to separate the signal power from the combined power the distortion and the noise. It is therefore an appropriate parameter defined as SINAD, which is a ratio of the output signal power to the output combined power of noise and distortion, that is,

$$\text{SINAD} = \frac{S_o}{N_o + D_o}. \quad (12.92)$$

In the input portion, another ratio, called RISE, is the ratio of the sum of the input signal power and the input equivalent noise power ( $S_i + N_{eq}$ ) to the input equivalent noise power  $N_{eq}$ , that is,

$$\text{RISE} = \frac{S_i + N_{eq}}{N_{eq}} = \frac{S_i}{N_{eq}} + 1. \quad (12.93)$$

It indicates the strength of the input signal power over the input equivalent noise power.

The sensitivity of a receiver is defined as a critical point of the input signal at which the input signal power approaches the input equivalent noise power, that is,

$$S_i = N_{eq}, \quad (12.94)$$

and from expression (12.95), we have

$$\text{RISE} = 2. \quad (12.95)$$

In general cases, from (12.95), we have

$$\frac{S_i}{N_{eq}} = \text{RISE} - 1. \quad (12.96)$$

Then, from (12.91) and (12.98), we have

$$E_i = 2\sqrt{S_i R_i} = 2\sqrt{N_{eq}(\text{RISE} - 1)R_i}. \quad (12.97)$$

By substituting (12.90) and (12.93) into (12.99), we have

$$E_i = 2\sqrt{\text{NF}(\text{RISE} - 1)R_i kT \Delta f}, \quad (12.98)$$

or

$$S_i = \frac{E_i^2}{4R_i} = \text{NF}(\text{RISE} - 1)kT \Delta f, \quad (12.99)$$

where

$S_i, S_o$  = the signal power at input and output, respectively (dB<sub>w</sub> or dB<sub>m</sub>),  
 $N_i, N_o$  = the noise power at input and output, respectively (dB),

- $D_o$  = the distortion power at the output (dB),  
 $R_i$  = the input resistance of the generator ( $\Omega$ ),  
 $N_{eq}$  = the equivalent input noise power of the output noise, and  
 $E_i$  = the sensitivity looking from input ( $\mu V$ ).

If condition (12.96) or (12.97) is satisfied, then the expressions for the sensitivity of the receiver (12.100) and (12.101) become

$$E_i = 2\sqrt{NF \cdot R_i kT \Delta f}, \quad (12.100)$$

or

$$S_i = \frac{E_i^2}{4R_i} = NF \cdot kT \Delta f. \quad (12.101)$$

Obviously, the sensitivity of a receiver consists of three portions:

1. Noise figure NF of the entire receiver,
2. Noise floor (noise spectrum density),  $kT$ , in the environment of receiver, and
3. Frequency bandwidth,  $\Delta f$ , of receiver.

Then

$$S_{i,dB} = 10 \log_{10} S_i = 10 \log_{10} NF + 10 \log_{10}(kT) + 10 \log_{10}(\Delta f). \quad (12.102)$$

If

$$T = 290K, \quad (12.103)$$

the noise floor (noise spectrum density)  $kT$  is

$$(kT)_{Watt/Hz} = 1.38 \cdot 10^{-23} \cdot 290 = 4.002 \cdot 10^{-21} \text{ Watt/Hz}, \quad (12.104)$$

$$10 \log_{10}(kT)_{dBW/Hz} = 10 \log_{10}(4.002 \cdot 10^{-21}) = -204 \text{ dBW/Hz}, \quad (12.105)$$

$$10 \log_{10}(kT)_{dBm/Hz} = 10 \log_{10}(4.002 \cdot 10^{-21} / 10^{-3}) = -174 \text{ dBm/Hz}. \quad (12.106)$$

For example, if

$$NF = 6 \text{ dB, and} \quad (12.107)$$

$$\Delta f = 1 \text{ MHz}, \quad (12.108)$$

then

$$10 \log_{10}(\Delta f) = 60 \text{ dB,} \quad (12.109)$$

$$S_{i,dB} = 6 - 174 + 60 = -108 \text{ dBm.} \quad (12.110)$$

If

$$R_i = 50 \Omega, \quad (12.111)$$

$$E_i = 10^{[-108-30+10\log(4Ri)]/20} = 10^{-5.75} = 1.78 \mu V. \quad (12.112)$$

Typically, there are two ways to test the sensitivity of the receiver:

1. **12-dB SINAD.** Figure 12.14 shows the setup and the input signal from the generator for the testing of the 12-dB SINAD sensitivity of receiver.

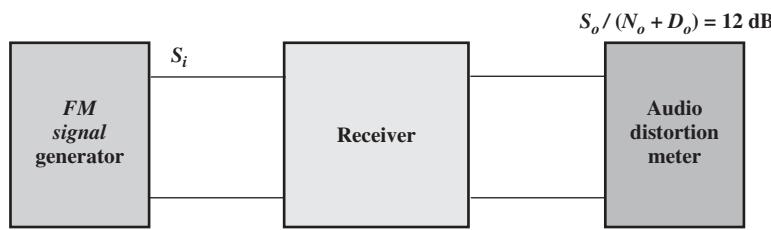
The input signal power  $S_i$  is provided by an RF signal generator. It is a frequency-modulated signal, in which the carrier of the desired frequency in the operating frequency range is modulated with a 1-kHz sinusoidal signal. At the beginning, the output power of the generator is turned off. It is then increased in very small increments slowly. The SINAD value at the audio distortion meter will increase accordingly. When the SINAD value increases up to 12 dB as the input signal power from generator is increased, the corresponding value of the input signal power  $S_i$  is the sensitivity  $S_{\text{sen}12}$ , that is,

$$S_{i,\text{dB}}|_{\text{SINAD}=12 \text{ dB}} = S_{\text{sen}12}, \quad (12.113)$$

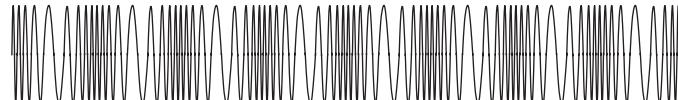
where  $S_{i,\text{sen}12}$  = the 12-dB SINAD sensitivity.

2. **20-dB Quieting.** Figure 12.15 shows the setup and the input signal from generator for the testing the 20-dB quieting sensitivity of a receiver.

The input signal power  $S_i$  is provided by an RF signal generator. It is a pure carrier signal. At the beginning, the output power of the generator is turned off so that only noise is detected by the audio distortion meter. The output power of the generator is then increased in very small increments slowly. The noise at the audio distortion meter gradually quiets down and the ratio of signal to noise  $S_o/(N_o + D_o)$  increases accordingly as the input signal power is increased. Finally, the ratio  $S_o/(N_o + D_o)$  is increased by up to 20 dB as the input signal power from the

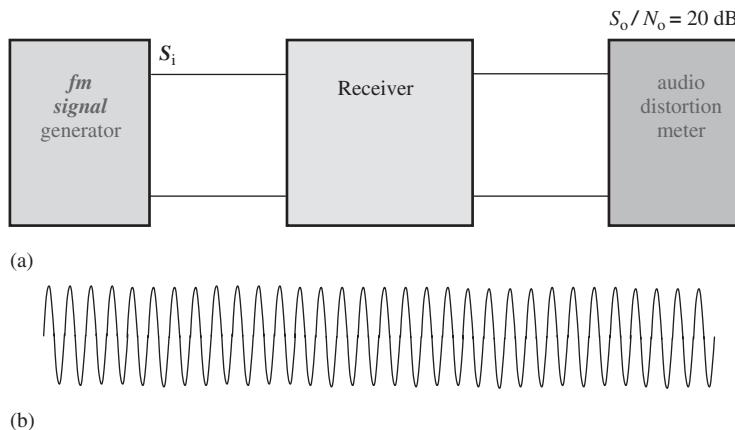


(a) Setup for testing of 12 dB SINAD sensitivity of a receiver.



(b)  $S_i$  = Carrier  $f_o$  modulated with  $f_D = 1$  kHz audio signal.

**Figure 12.14.** 12-dB SINAD sensitivity testing for a receiver. (a) Setup for testing of 12-dB SINAD sensitivity of a receiver. (b)  $S_i$  = Carrier  $f_o$  modulated with  $f_D = 1$  kHz audio signal.



**Figure 12.15.** 20-dB quieting sensitivity testing for a receiver. (a) Setup for testing the 20 dB quieting sensitivity of a receiver. (b)  $S_i$  = Carrier signal with frequency  $f_o$ .

generator is increased. The corresponding value of the input signal power  $S_i$  when the ratio  $S_o/N_o$  reaches 20 dB is the sensitivity  $S_{sen20}$ , that is,

$$S_{i,dB}|_{S_o/N_o=20 \text{ dB}} = S_{sen.20} \quad (12.114)$$

where  $S_{sen20}$  = the 20-dB quieting sensitivity.

As a matter of fact, experiments indicate that either the 12-dB SINAD or the 20-dB quieting is close to the condition of (12.96) or (12.97), so that their corresponding sensitivities  $S_{i,sen12}$  and  $S_{sen20}$  are supposed to be approximately the same. Their difference due to the different test methodology is normal and acceptable.

## 12.4 NONLINEARITY

### 12.4.1 Nonlinearity of a Device

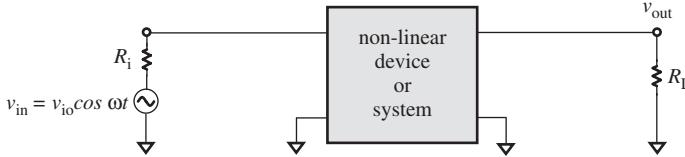
Nonlinearity of a system mainly arises from the nonlinearity of the device. Nonlinearity exists more or less in any practical device. In reality, an ideal linear device or system does not exist, and any “linear” device or system is just a reasonable approximation.

Nonlinearity of a device or system produces spurious products. All the spurious products are sources of interference with the signal and eventually cause distortion of the signal. There are two kinds of spurious products: harmonics and complicated spurious products.

#### 12.4.1.1 Harmonics When Input Is a Signal with Only One Frequency.

Assuming that the input voltage of the signal is a pure sinusoidal wave as shown in Figure 12.16, that is,

$$v_{in} = v_{io} \cos \omega t, \quad (12.115)$$



**Figure 12.16.** Harmonics produced from a nonlinear device or system even if the input is a signal with only one frequency.

the nonlinearity of a device can be expressed in the general form,

$$v_{out} = a_0 + a_1 v_{in} + a_2 v_{in}^2 + a_3 v_{in}^3 + a_4 v_{in}^4 + \dots, \quad (12.116)$$

where

$a_i$  = the  $i$ th-order nonlinear coefficients,  
 $i = 0, 1, 2, 3, 4 \dots$

Substituting (12.117) into (12.118), we have

$$v_{out} = V_{dc} + \sum_{n=1}^{\infty} A_n \cos n\omega t, \quad (12.117)$$

where

$V_{dc}$  = the DC component of output voltage, and  
 $A_n$  = the amplitude of  $n$ th harmonic.

They are

$$V_{dc} = a_0 + \frac{1}{2} a_2 v_{10}^2 + \frac{3}{8} a_4 v_{10}^4 + \dots \quad (12.118)$$

$$A_1 = a_1 v_{10} + \frac{4}{3} a_3 v_{10}^3 + \dots \quad (12.119)$$

$$A_2 = \frac{1}{2} a_2 v_{10}^2 + \frac{1}{2} a_4 v_{10}^4 + \dots \quad (12.120)$$

$$A_3 = \frac{1}{4} a_3 v_{10}^3 + \dots \quad (12.121)$$

$$A_4 = \frac{1}{8} a_4 v_{10}^4 + \dots \quad (12.122)$$

...

The output voltage is a complicated entity with infinite harmonics.

These coefficients  $A_n$  may be dependent on or independent of the current flowing through the device or system or of the voltage drop across the device or system. As a general trend, these coefficients are reduced as the order increases. However, sometimes there are extraordinary cases in which the nonlinearity coefficients of higher orders have higher values than those of lower orders. The infinite number of harmonics, the terms containing the factors  $\cos \omega t, \cos 2\omega t, \cos 3\omega t, \cos 4\omega t, \dots$ , are produced by the nonlinearity of the device or system. The new resulting coefficient of each harmonic  $A_n$  is another infinite series containing the original nonlinearity coefficients  $a_i$ . These new coefficients look very clumsy at first glance. In a practical engineering design, it is sufficient to keep only the first three or four terms containing the coefficients  $a_i$  with low orders.

A remarkable feature of these new coefficients can be discovered from equations (12.120) to (12.124): that is, the new coefficients with odd harmonics  $A_{2n+1}$  contain only old coefficients with odd orders  $a_{2i+1}$ , while the new coefficients with even harmonics  $A_{2n}$  contain only even coefficients with even orders  $a_{2i}$ . This is important to engineers who are asked for the REDUCTION of spurious products. In RF or RFIC circuits where a high linearity is required, such as the LNA or PA (power amplifier), devices with a higher value of  $a_1$  but lower values of the other coefficients should be chosen. In the circuit design of a mixer, which operates on the basis of the second-order of nonlinearity, devices with higher values of  $a_2$  but lower values of the other coefficients should be chosen.

By means of a spectrum analyzer, the harmonics can be seen on the screen as shown in Figure 12.17. This is an important technique to measure the nonlinearity of a device.

For an expected linear block or system, the power differences between the first and second harmonics and between the second and third harmonics are expected to be as large as possible. For instance, in an LNA design, they are expected to be

$$\Delta_1 = P_1 - P_2 > 20 \text{ dB}, \quad (12.123)$$

and

$$\Delta_2 = P_2 - P_3 > 10 \text{ dB}. \quad (12.124)$$

**12.4.1.2 Spurious Products When the Input Is a Signal with Two Frequencies.** If the input voltage signal consists of two frequencies as shown in Figure 12.18, that is,

$$v_{\text{in}} = v_{\text{io}} \cos \omega_1 t + v_{\text{io}} \cos \omega_2 t. \quad (12.125)$$

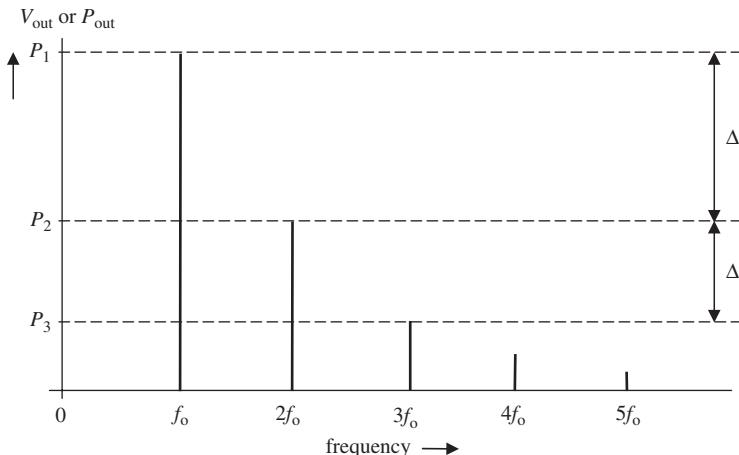


Figure 12.17. The frequency spectrum of a device when the input is a signal with only one frequency.

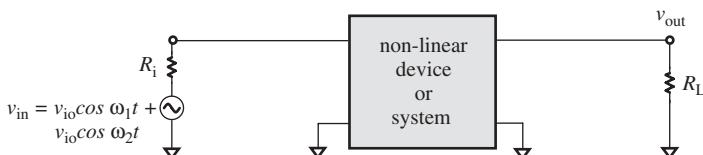


Figure 12.18. Spurious products produced from a nonlinear device or system if the input is a signal with two frequencies.

Substituting (12.127) into (12.118), the output voltage becomes a complicated entity with infinite harmonics and infinite spurious products with the composed frequencies.

$$v_{\text{out}} = \sum_{m=0, n=0}^{\infty} A_{mn} \cos(m\omega_1 \pm n\omega_2)t, \quad (12.126)$$

where  $A_{mn}$  is the amplitude of the harmonic with angular frequency  $= \omega_1 \pm n\omega_2$ ; they are

$$A_{00} = a_o + a_2 v_{\text{io}}^2 + \frac{9}{4} a_4 v_{\text{io}}^4 + \dots \quad (12.127)$$

$$A_{10} = A_{01} = a_1 v_{\text{io}} + \frac{9}{4} a_3 v_{\text{io}}^3 + \dots \quad (12.128)$$

$$A_{20} = A_{02} = \frac{1}{2} a_2 v_{\text{io}}^2 + 2 a_4 v_{\text{io}}^4 \dots \quad (12.129)$$

$$A_{11} = a_2 v_{\text{io}}^2 + 3 a_4 v_{\text{io}}^4 + \dots \quad (12.130)$$

$$A_{30} = A_{03} = a_3 v_{\text{io}}^3 + \dots \quad (12.131)$$

$$A_{21} = A_{12} = \frac{3}{4} a_4 v_{\text{io}}^4 + \dots \quad (12.132)$$

$$A_{40} = A_{04} = \frac{1}{8} a_4 v_{\text{io}}^4 + \dots \quad (12.133)$$

$$A_{31} = A_{13} = \frac{1}{2} a_4 v_{\text{io}}^4 + \dots \quad (12.134)$$

$$A_{22} = \frac{3}{4} a_4 v_{\text{io}}^4 + \dots \quad (12.135)$$

...

A remarkable feature of these new coefficients is that the coefficients of the odd spurious products or harmonics contain only odd coefficients with odd orders, while the coefficients of the even spurious products or harmonics contain only even coefficients with even orders.

From expression (12.128), it can be seen that the spectrum becomes quite complicated, in which the frequencies of the spurious products can be summarized as

$$\omega_{\text{spur}} = |m\omega_1 \pm n\omega_2|. \quad (12.136)$$

It should be noted that  $A_{00}$  is the DC component of output voltage when both  $m$  and  $n$  are equal to 0. When either  $m$  or  $n$  is zero but the other is nonzero, the output contains the desired signal and spurious products with harmonics. When both  $m$  and  $n$  are nonzero but positive integers, the output contains spurious products with composed frequencies. For example, in an LNA design, the third-order intermodulation is due to the spurious product with

$$m = 2, \text{ and } n = 1, \quad (12.137)$$

or

$$m = 1, \text{ and } n = 2, \quad (12.138)$$

and the second-order intermodulation is due to the spurious product with

$$m = 2, \text{ and } n = 2, \quad (12.139)$$

which are called *half IF spurs* or the *Able–Baker spurs*.

The resulting output contains not only the harmonics of the two frequencies, but also the other combined spurious products. The order of the spurious products is the number of frequencies involved in the terms, which is listed in Table 12.3.

As examples, Table 12.4 lists three sets of spurious products when the input signals are two frequencies.

$$\omega_1 = 850 \text{ MHz}, \omega_2 = 828.6 \text{ MHz}, \text{ IF} = \omega_1 - \omega_2 = 21.4 \text{ MHz}, \quad (12.140)$$

$$\omega_1 = 850 \text{ MHz}, \omega_2 = 805.0 \text{ MHz}, \text{ IF} = \omega_1 - \omega_2 = 45.0 \text{ MHz}, \quad (12.141)$$

$$\omega_1 = 850 \text{ MHz}, \omega_2 = 776.65 \text{ MHz}, \text{ IF} = \omega_1 - \omega_2 = 73.35 \text{ MHz}. \quad (12.142)$$

It should be noted that one of the two frequencies is kept the same at 850 MHz, that is,

$$\omega_1 = 850 \text{ MHz}. \quad (12.143)$$

And, the difference of the two frequencies is denoted by IF, that is,

$$\text{IF} = \omega_1 - \omega_2. \quad (12.144)$$

The three sets of spurious products listed in Table 12.4 are relisted and sorted from low to high frequency in Table 12.5.

It is very interesting to compare these three sets of spurious products. Figure 12.19(a)–(c) plots the first nine rows for the three sets of spurious products listed in Table 12.4, respectively, which are the low-frequency spurious products below 1000 MHz. The order of the spurious products is marked beside the spectrum accordingly.

By comparing Figure 12.19(a), (b) and (c), the following can be found:

- All of them have low-frequency spurious products of IF, 2IF, and 3IF with second, fourth, and sixth orders of nonlinearity in low-frequency end. These three frequency components are squeezed toward the low-frequency end more as the IF is decreased.
- All of them have some spurious products around  $\omega_1$ . They scatter from  $\omega_1$  to further away as the IF is increased.

Consequently, the first set of spurious products, IF = 21.4 MHz, may have the least opportunity to interfere with other circuit blocks or systems because there are no spurious products in a wide frequency range from about from 65 to 780 MHz. In other words, if there are other circuit blocks or systems operating in the frequency range

TABLE 12.3. Order of the Spurious Products and Their Frequencies If the Input Is a Signal with Two Frequencies

Order of the Spurious Products	Frequency
1	$\omega_1$
1	$\omega_2$
2	$2\omega_1$
2	$2\omega_2$
2	$\omega_1 - \omega_2$
2	$\omega_1 + \omega_2$
3	$3\omega_1$
3	$3\omega_2$
3	$2\omega_1 - \omega_2$
3	$2\omega_1 + \omega_2$
3	$2\omega_2 - \omega_1$
3	$2\omega_2 + \omega_1$
4	$4\omega_1$
4	$4\omega_2$
4	$3\omega_2 - \omega_1$
4	$3\omega_2 + \omega_1$
4	$3\omega_1 - \omega_2$
4	$3\omega_1 + \omega_2$
4	$2(\omega_1 - \omega_2)$
4	$2(\omega_1 + \omega_2)$
5	$5\omega_1$
5	$5\omega_2$
5	$4\omega_2 - \omega_1$
5	$4\omega_2 + \omega_1$
5	$4\omega_1 - \omega_2$
5	$4\omega_1 + \omega_2$
5	$3\omega_1 - 2\omega_2$
5	$3\omega_1 + 2\omega_2$
5	$3\omega_2 - 2\omega_1$
5	$3\omega_2 + 2\omega_1$
6	$6\omega_1$
6	$6\omega_2$
6	$5\omega_2 - \omega_1$
6	$5\omega_2 + \omega_1$
6	$5\omega_1 - \omega_2$
6	$5\omega_1 + \omega_2$
6	$4\omega_1 - 2\omega_2$
6	$4\omega_1 + 2\omega_2$
6	$4\omega_2 - 2\omega_1$
6	$4\omega_2 + 2\omega_1$
6	$3(\omega_1 - \omega_2)$
6	$3(\omega_1 + \omega_2)$

from 65 to 780 MHz, the first set of spurious products is no longer an interference source.

This feature is applied in communication system designs where the selection of the LO (local oscillator) frequency, and hence the IF frequency for a given RF frequency,

TABLE 12.4. Examples of the Spurious Products and Their Frequencies If the Input Is a Signal with Two Frequencies

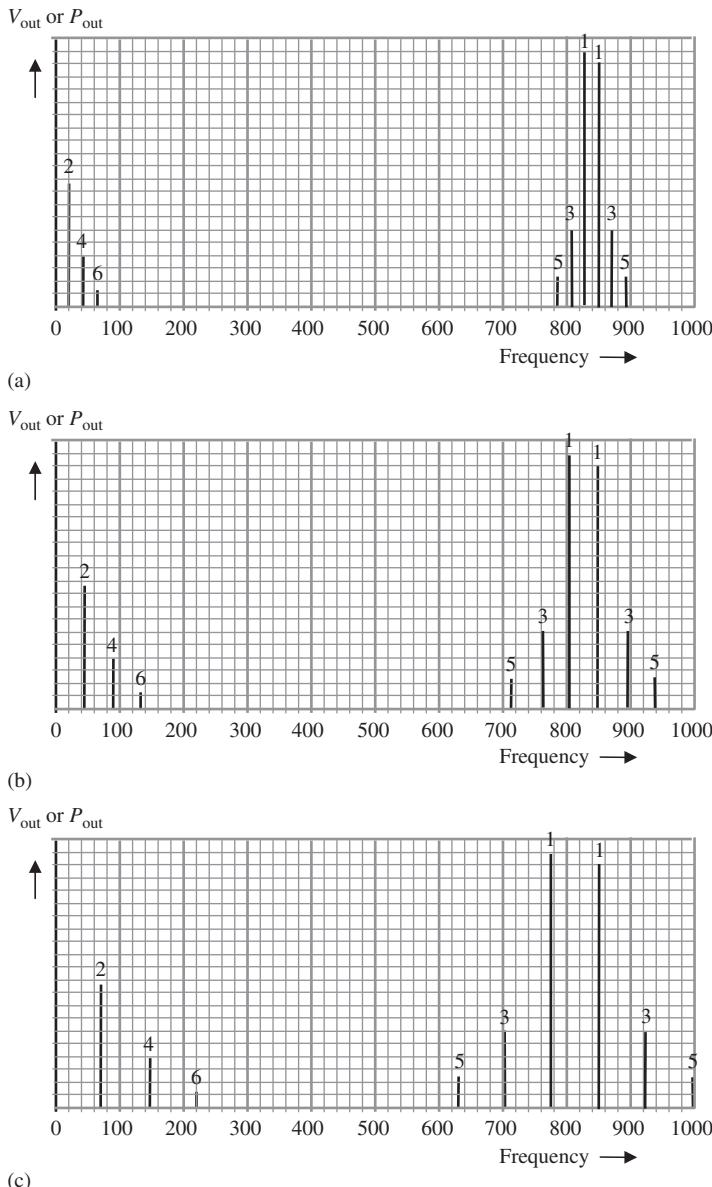
Order	Frequency	(1) Spurious, MHz	(2) Spurious, MHz	(3) Spurious, MHz
1	$\omega_1$	850.0	850	850.00
1	$\omega_2$	828.6	805	776.65
2	$2\omega_1$	1700.0	1700	1700.00
2	$2\omega_2$	1657.2	1610	1553.30
2	$\omega_1 - \omega_2$	21.4	45	73.35
2	$\omega_1 + \omega_2$	1678.6	1655	1626.65
3	$3\omega_1$	2550.0	2550	2550.00
3	$3\omega_2$	2485.8	2415	2329.95
3	$2\omega_1 - \omega_2$	871.4	895	923.35
3	$2\omega_1 + \omega_2$	2528.6	2505	2476.65
3	$2\omega_2 - \omega_1$	807.2	760	703.30
3	$2\omega_2 + \omega_1$	2507.2	2460	2403.30
4	$4\omega_1$	3400.0	3400	3400.00
4	$4\omega_2$	3314.4	3220	3106.60
4	$3\omega_1 - \omega_2$	1721.4	1745	1773.35
4	$3\omega_1 + \omega_2$	3378.6	3355	3326.65
4	$3\omega_2 - \omega_1$	1635.8	1565	1479.95
4	$3\omega_2 + \omega_1$	3335.8	3265	3179.95
4	$2(\omega_1 - \omega_2)$	42.8	90	146.70
4	$2(\omega_1 + \omega_2)$	3357.2	3310	3253.30
5	$5\omega_1$	4250.0	4250	4250.00
5	$5\omega_2$	4143.0	4025	3883.25
5	$4\omega_1 - \omega_2$	2571.4	2595	2623.35
5	$4\omega_1 + \omega_2$	4228.6	4205	4176.65
5	$4\omega_2 - \omega_1$	2464.4	2370	2256.60
5	$4\omega_2 + \omega_1$	4164.4	4070	3956.60
5	$3\omega_1 - 2\omega_2$	892.8	940	996.70
5	$3\omega_1 + 2\omega_2$	4207.2	4160	4103.30
5	$3\omega_2 - 2\omega_1$	785.8	715	629.95
5	$3\omega_2 + 2\omega_1$	4185.8	4115	4029.95
6	$6\omega_1$	5100.0	5100	5100.00
6	$6\omega_2$	4971.6	4830	4659.90
6	$5\omega_1 - \omega_2$	3421.4	3445	3473.35
6	$5\omega_1 + \omega_2$	5078.6	5055	5026.65
6	$5\omega_2 - \omega_1$	3293.0	3175	3033.25
6	$5\omega_2 + \omega_1$	4993.0	4875	4733.25
6	$4\omega_1 - 2\omega_2$	1742.8	1790	1846.70
6	$4\omega_1 + 2\omega_2$	5057.2	5010	4953.30
6	$4\omega_2 - 2\omega_1$	1614.4	1520	1406.60
6	$4\omega_2 + 2\omega_1$	5014.4	4920	4806.60
6	$3(\omega_1 - \omega_2)$	64.2	135	220.05
6	$3(\omega_1 + \omega_2)$	5035.8	4965	4879.95

is very important. If  $\omega_1 = 850$  MHz, as in our example above, is selected as the RF frequency in the mixer design, then the selection of  $\omega_2 = 828.6$  MHz or IF = 21.4 MHz is better than the selections of  $\omega_2 = 805$  MHz or IF = 45 MHz and  $\omega_2 = 776.65$  MHz or IF = 73.35 MHz.

TABLE 12.5. Examples of the Spurious Products and Their Sorted Frequencies If the Input Is a Signal with Two Frequencies

Order	Frequency	(1) Spurious, MHz	(2) Spurious, MHz	(3) Spurious, MHz
2	$\omega_1 - \omega_2$	21.4	45	73.35
4	$2(\omega_1 - \omega_2)$	42.8	90	146.70
6	$3(\omega_1 - \omega_2)$	64.2	135	220.05
5	$3\omega_2 - 2\omega_1$	785.8	715	629.95
3	$2\omega_2 - \omega_1$	807.2	760	703.30
1	$\omega_2$	828.6	805	776.65
1	$\omega_1$	850.0	850	850.00
3	$2\omega_1 - \omega_2$	871.4	895	923.35
5	$3\omega_1 - 2\omega_2$	892.8	940	996.70
6	$4\omega_2 - 2\omega_1$	1614.4	1520	1406.60
4	$3\omega_2 - \omega_1$	1635.8	1565	1479.95
2	$2\omega_2$	1657.2	1610	1553.30
2	$\omega_1 + \omega_2$	1678.6	1655	1626.65
2	$2\omega_1$	1700.0	1700	1700.00
4	$3\omega_1 - \omega_2$	1721.4	1745	1773.35
6	$4\omega_1 - 2\omega_2$	1742.8	1790	1846.70
5	$4\omega_2 - \omega_1$	2464.4	2370	2256.60
3	$3\omega_2$	2485.8	2415	2329.95
3	$2\omega_2 + \omega_1$	2507.2	2460	2403.30
3	$2\omega_1 + \omega_2$	2528.6	2505	2476.65
3	$3\omega_1$	2550.0	2550	2550.00
5	$4\omega_1 - \omega_2$	2571.4	2595	2623.35
6	$5\omega_2 - \omega_1$	3293.0	3175	3033.25
4	$4\omega_2$	3314.4	3220	3106.60
4	$3\omega_2 + \omega_1$	3335.8	3265	3179.95
4	$2(\omega_1 + \omega_2)$	3357.2	3310	3253.30
4	$3\omega_1 + \omega_2$	3378.6	3355	3326.65
4	$4\omega_1$	3400.0	3400	3400.00
6	$5\omega_1 - \omega_2$	3421.4	3445	3473.35
5	$5\omega_2$	4143.0	4025	3883.25
5	$4\omega_2 + \omega_1$	4164.4	4070	3956.60
5	$3\omega_2 + 2\omega_1$	4185.8	4115	4029.95
5	$3\omega_1 + 2\omega_2$	4207.2	4160	4103.30
5	$4\omega_1 + \omega_2$	4228.6	4205	4176.65
5	$5\omega_1$	4250.0	4250	4250.00
6	$6\omega_2$	4971.6	4830	4659.90
6	$5\omega_2 + \omega_1$	4993.0	4875	4733.25
6	$4\omega_2 + 2\omega_1$	5014.4	4920	4806.60
6	$3(\omega_1 + \omega_2)$	5035.8	4965	4879.95
6	$4\omega_1 + 2\omega_2$	5057.2	5010	4953.30
6	$5\omega_1 + \omega_2$	5078.6	5055	5026.65
6	$6\omega_1$	5100.0	5100	5100.00

As a matter of fact, in the selection of LO and hence IF frequency for a communication system, not only low-frequency products but also high-frequency ones must be analyzed. In other words, the analysis of only the first nine rows listed in Table 12.4 must be extended to cover higher orders of nonlinearity of the device, and the plot of Figure 12.19 must be extended to higher frequencies. This is tedious but important work in the system design.



**Figure 12.19.** Frequency spectrum of spurious products when the input is a signal with two frequencies. (a) IF = 21.4 MHz. (b) IF = 45.0 MHz. (c) IF = 73.35 MHz.

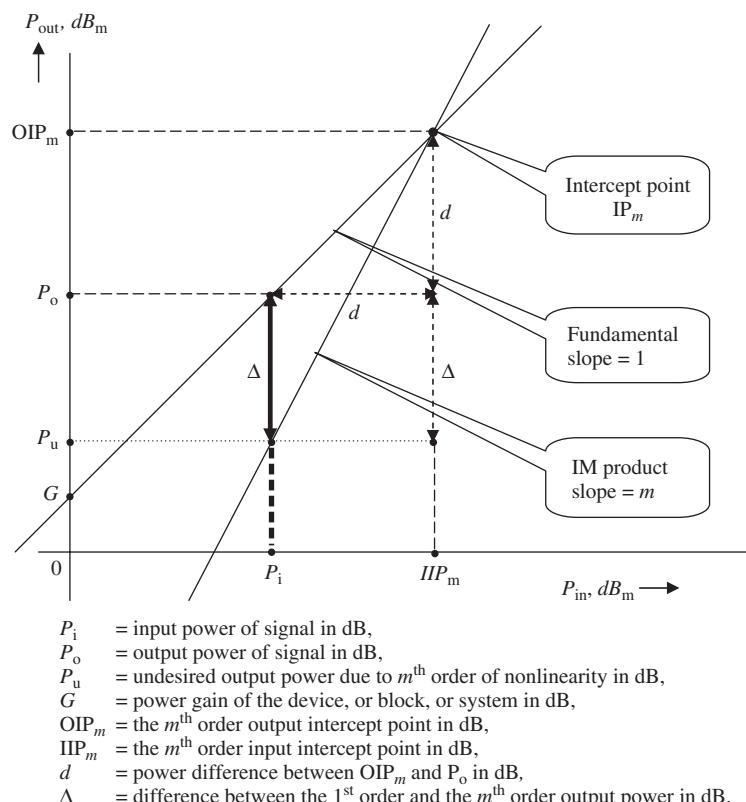
### 12.4.2 IP (Intercept Point) and IMR (Intermodulation Rejection)

Intermodulation of a RF block mainly comes from the nonlinearity of the device or the block. It therefore depends on the device characteristic itself. On the other hand, it also depends on the circuit topology and the characteristics of the other parts. This implies that the intermodulation would be different for different devices, different topologies, or different parts applied in the circuit block. Conversely, different RF blocks have different levels of capability to reject intermodulation. IMR (intermodulation rejection) therefore

becomes an important specification to measure the nonlinearity or linearity of a RF block.

The intercept point is just an intermediate parameter by which the IMR can be evaluated. Intercept points are categorized on the basis of their order of nonlinearity. For instance, the  $m$ th-order intercept point is defined as the intercept point of the input/output power curves of first-order signal and the  $m$ th-order IM (intermodulation) product for a given frequency. Figure 12.20 shows the  $m$ th-order intercept point  $\text{IP}_m$ .

The straight line with a slope of 1 depicts the relation between the input and the output signal power. The output power is  $G \text{ dB}_m$  if the input power is  $0 \text{ dB}_m$ , where  $G$  is the power gain of the device or system. Another line with a slope of  $m$  represents the  $m$ th-order IM product or spurious product. When the input signal power  $P_i$  is low, the output power of the  $m$ th-order IM product  $P_u$  is still much lower than the output power of the first-order output signal  $P_o$ . Let  $\Delta$  denote the output power difference between the first order and the  $m$ th order when the input signal power is  $P_i$ . The  $m$ th-order IM product output increases much faster than the first-order output signal power as the input signal power is increased from  $P_i$ , and eventually the two lines intercept at one point as shown in Figure 12.20. This point is called the *intercept point* of the first-order signal and the  $m$ th-order IM product. The corresponding abscissa and ordinate coordinates are called the  $m$ th-order output intercept point,  $\text{OIP}_m$  and the  $m$ th-order input intercept point,  $\text{IIP}_m$ , respectively.



**Figure 12.20.** Plot of the first-order signal and the  $m$ th-order IM product on the input–output power plane.

Let us discuss how to find out the  $m$ th-order IMR from the  $m$ th-order intercept point  $\text{IP}_m$  since the intercept point is just an intermediate parameter.

First, let us find out the relationship between the input intercept point  $\text{IIP}_m$  and the input signal power  $P_i$ . By simple geometric derivation,

$$\Delta = m(\text{IIP}_m - P_i) - 1(\text{IIP}_m - P_i) = (m - 1)(\text{IIP}_m - P_i), \quad (12.145)$$

$$\text{IIP}_m = \frac{\Delta}{(m - 1)} + P_i, \quad (12.146)$$

$$\text{OIP}_m = G + \text{IIP}_m. \quad (12.147)$$

By means of equations (12.148) and (12.149), either the input intercept point  $\text{IIP}_m$  or output intercept point  $\text{OIP}_m$  can be obtained by the measurement of  $\Delta$  at an appropriate value of  $P_i$  and the power gain  $G$ .

Now, let us calculate the  $m$ th-order intermodulation rejection,  $\text{IMR}_m$ , from the  $m$ th-order input intercept point,  $\text{IIP}_m$ . By simple geometry as shown in Figure 12.21, we have

$$\text{IMR}_m = (\text{IIP}_m - P_{i,s}) - \frac{(\text{IIP}_m - P_{i,s})}{m}, \quad (12.148)$$

$$\text{IMR}_m = \alpha(\text{IIP}_m - P_{i,s}), \quad (12.149)$$

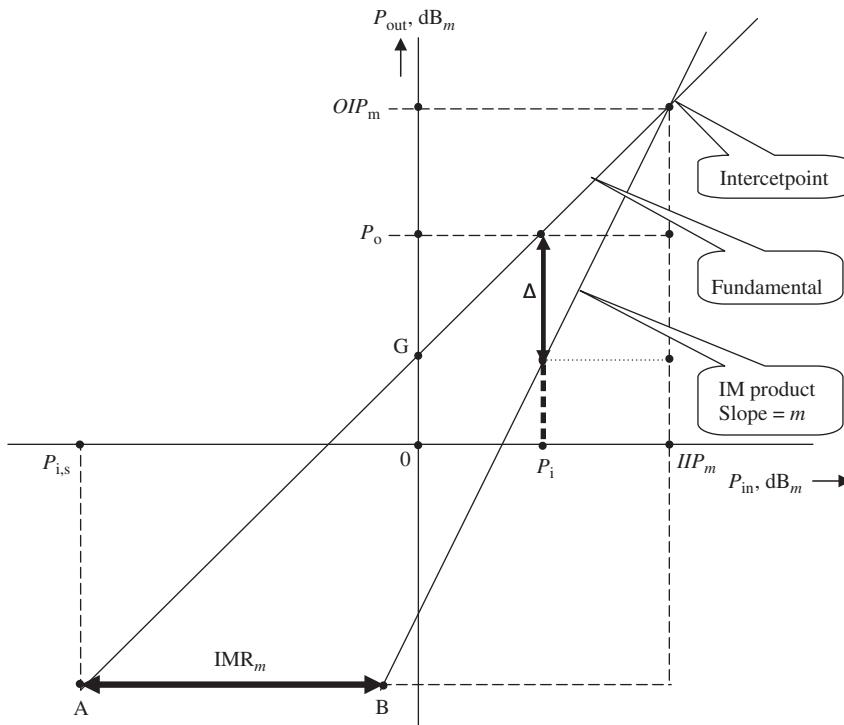


Figure 12.21. Calculation of  $\text{IMR}_m$  from intercept point  $\text{IIP}_m$ .

where

$$\alpha = \frac{(m - 1)}{m}. \quad (12.150)$$

Figure 12.21 illustrates the meaning of the  $m$ th-order intermodulation rejection  $IMR_m$ , which is expressed by the length of AB. In Figure 12.21,  $P_{i,s}$  denotes the input power at the sensitivity point, the minimum of a detectable input signal level.

- Point A represents the input signal power at the sensitivity point and its corresponding output signal power.
- Point B is on the  $m$ th-order IM product line with the same output power as point A.

The input power at point B is  $(IMR_m)$  dB higher than point A. In other words, when the input signal is at the sensitivity point, the  $m$ th-order IM product in the input is  $(IMR_m)$  dB stronger than the input signal power. As both input powers are increased toward  $IIP_m$ , they have the same output at the intercept point. Therefore, it implies that this system has a capability to suppress or to reject the  $m$ th-order IM product down  $(IMR_m)$  dB.

**12.4.2.1 Second-Order Intercept Point,  $IP_2$ .** As mentioned earlier, the power or amplitude of an IM or spurious product is relatively reduced as its order is increased. Therefore, more attention is paid to the IM or spurious products with lower orders. The IM or spurious products with  $m = 2$  or 3 have been more focused on than the others since the IM or spurious products with  $m = 1$  is the desired signal.

In this section, let us discuss the second-order spurious product. When

$$m = 2, \quad (12.151)$$

the expressions (12.148), (12.149), (12.151), and (12.152) become

$$IIP_2 = \Delta + P_i, \quad (12.152)$$

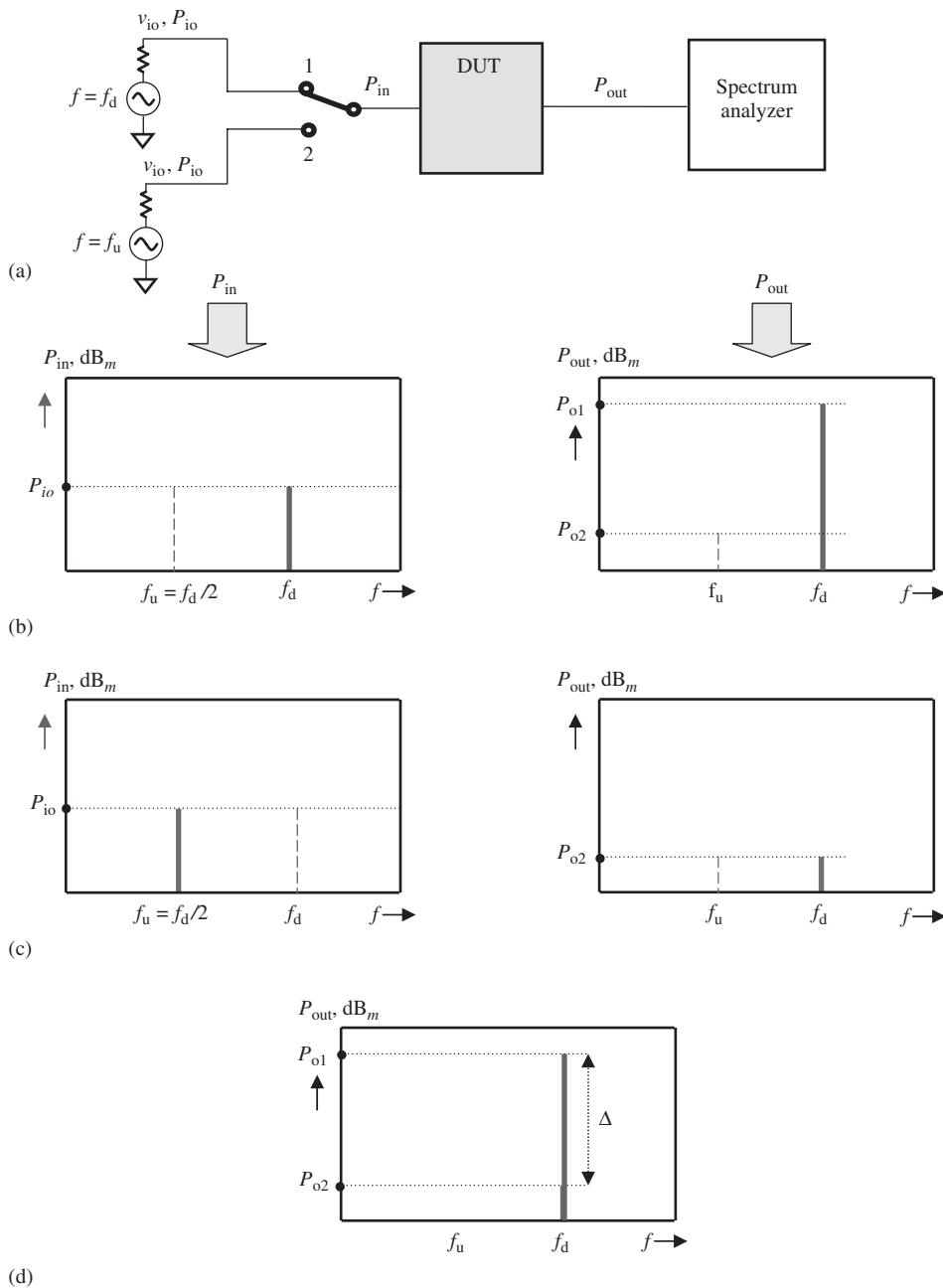
$$OIP_2 = G + IIP_2, \quad (12.153)$$

$$IMR_2 = \alpha(IIP_2 - P_{i,s}) = \frac{1}{2}(IIP_2 - P_{i,s}), \quad (12.154)$$

$$\alpha = \frac{1}{2}. \quad (12.155)$$

By directly measuring  $\Delta$  in the test laboratory, one can calculate  $IIP_2$ ,  $OIP_2$ , and  $IMR_2$  by means of equations (12.154) to (12.157). The setup and display of  $IP_2$  testing is shown in Figure 12.22. Figure 12.22(a) shows its setup. There are two generators generating the same voltage or power level,  $v_{io}$  or  $P_{in}$ , but with different frequencies  $f_d$  and  $f_u$ . Two input signals are toggled by a switch. The power output  $P_{out}$  from the desired test block or system is displayed on the screen of the spectrum analyzer. There are three experimental steps:

The first step is to test for the response from the input to output when the switch is thrown on position 1. The input signal is one with the desired frequency  $f_d$  and input voltage  $v_{io}$  or input power  $P_{io}$ . The output power  $P_{o1}$  is read at the desired frequency  $f_d$  from the screen of the spectrum analyzer. The display is shown in Figure 12.22(b).



**Figure 12.22.** Setup and display of IP<sub>2</sub> testing. (a) Setup of IP<sub>2</sub> testing. (b) Switch is thrown on position 1. (c) Switch is thrown on position 2. (d) Calculation of  $\Delta$ .

The second step is to test for the response from the input to output when the switch is thrown on position 2. The input signal is one with the undesired frequency  $f_u = f_d/2$  and with the same input voltage  $v_{io}$  or input power  $P_{io}$  as in step 1. The output power  $P_{o2}$  is read at the desired frequency  $f_d$  from the screen of the spectrum analyzer. The display is shown in Figure 12.22(c).

As shown in Figure 12.22(d), the third step is to calculate the difference of output powers at the desired frequency  $f_d$  based on the readings of output power in two previous test steps,  $P_{o1}$  and  $P_{o2}$ , that is,

$$\Delta = P_{o1} - P_{o2}. \quad (12.156)$$

Note that

$$P_i = P_{io}. \quad (12.157)$$

By substituting (12.158) and (12.159) into (12.154), (12.155), and (12.156), the values of  $IIP_2$ ,  $OIP_2$ , and  $IMR_2$  can be obtained accordingly.

This is one way to obtain the value of  $IMR_2$  by means of the test setup shown in Figure 12.22(a).

Another way to obtain the value  $IMR_2$  is to calculate it from the nonlinear coefficients of the nonlinear device or system if the transfer function as shown in expression (12.118) is known.

Recalling the expressions from (12.119) to (12.124), the transfer function (12.119) can be rewritten as

$$\begin{aligned} v_{out} = & [a_0 + (1/2)a_2 v_{io}^2 + (3/8)a_4 v_{io}^4 + \dots] + [a_1 v_{io} + (3/4)a_3 v_{io}^3 + \dots] \cos \omega t \\ & + [(1/2)a_2 v_{io}^2 + (1/2)a_4 v_{io}^4 + \dots] \cos 2\omega t + [(1/4)a_3 v_{io}^3 + \dots] \cos 3\omega t \\ & + [(1/8)a_4 v_{io}^4 + \dots] \cos 4\omega t + \dots. \end{aligned} \quad (12.158)$$

Assuming that  $\omega$  in equation (12.119) is the angular frequency of an undesired signal  $\omega_u$ , which is half of  $\omega_d$ , the angular frequency of desired signal, that is

$$\omega_u = \frac{\omega_d}{2}, \quad \text{or} \quad f_o = \frac{f_d}{2}, \quad (12.159)$$

where

$\omega_u$  or  $f_u$  = the undesired angular frequency or frequency, and  
 $\omega_d$  or  $f_d$  = desired angular frequency or frequency,

The IM product of the undesired signal with frequencies  $\omega_u$  in transfer function (12.160) containing the term with frequency  $2\omega_u = \omega_d$  is

$$v_{out,u} = \dots + [(1/2)a_2 v_{iu}^2 + (1/2)a_4 v_{iu}^4 + \dots] \cos 2\omega_u t + \dots, \quad (12.160)$$

where  $v_{iu}$  = the voltage amplitude of undesired signal.

According to the measurement of IM rejection, we raise this undesired signals' level  $v_{out,u}$  up to a reference level of desired signal  $v_{out,d}$ , which is

$$v_{out,d} = \dots + [a_1 v_{io} + \dots] \cos \omega_d t + \dots, \quad (12.161)$$

where  $v_{io}$  = the voltage amplitude of the desired signal.

This implies that the term with  $\cos 2\omega_u t$  in expression (12.162) for the transfer function of the undesired signal is equivalent to the term with  $\cos \omega_d t$  in expression (12.163) for the transfer function of the desired signal. Then, if the terms higher than

fourth-order nonlinearity in expression (12.162) and the terms higher than second-order nonlinearity in expression (12.163) are neglected, from the expressions (12.162) and (12.163) we have

$$a_1 v_{\text{io}} \cos \omega_d t \approx \frac{1}{2} a_2 v_{\text{iu}}^2 \cos 2\omega_u t. \quad (12.162)$$

Note that  $2\omega_u = \omega_d$ ,

$$v_{\text{iu}} \approx \left( 2 \frac{a_1}{a_2} v_{\text{io}} \right)^{1/2}. \quad (12.163)$$

Consequently,

$$\text{IMR}_{2,v} = \frac{v_{\text{iu}}}{v_{\text{io}}} \approx \left( 2 \frac{a_1}{a_2} \right)^{1/2} v_{\text{io}}^{-1/2}, \quad (12.164)$$

$$\text{IMR}_{2,\text{dB}} = 20 \log \frac{v_{\text{iu}}}{v_{\text{io}}} \approx 20 \log \left[ \left( 2 \frac{a_1}{a_2} \right)^{1/2} v_{\text{io}}^{-1/2} \right]. \quad (12.165)$$

The second-order intermodulation rejection  $\text{IMR}_2$ , is not only dependent on the ratio of the nonlinearity coefficients  $a_1$  and  $a_2$  but is also sensitive to the voltage amplitude of the input signal  $v_{\text{io}}$ . The higher the voltage amplitude of the input signal  $v_{\text{io}}$ , the lower the  $\text{IMR}_2$ .

#### 12.4.2.2 Third-Order Intercept Point, $\text{IP}_3$ .

When

$$m = 3, \quad (12.166)$$

then expressions (12.148), (12.149), (12.151), and (12.152) become

$$\text{IIP}_3 = \frac{\Delta}{2} + P_i, \quad (12.167)$$

$$\text{OIP}_3 = G + \text{IIP}_3, \quad (12.168)$$

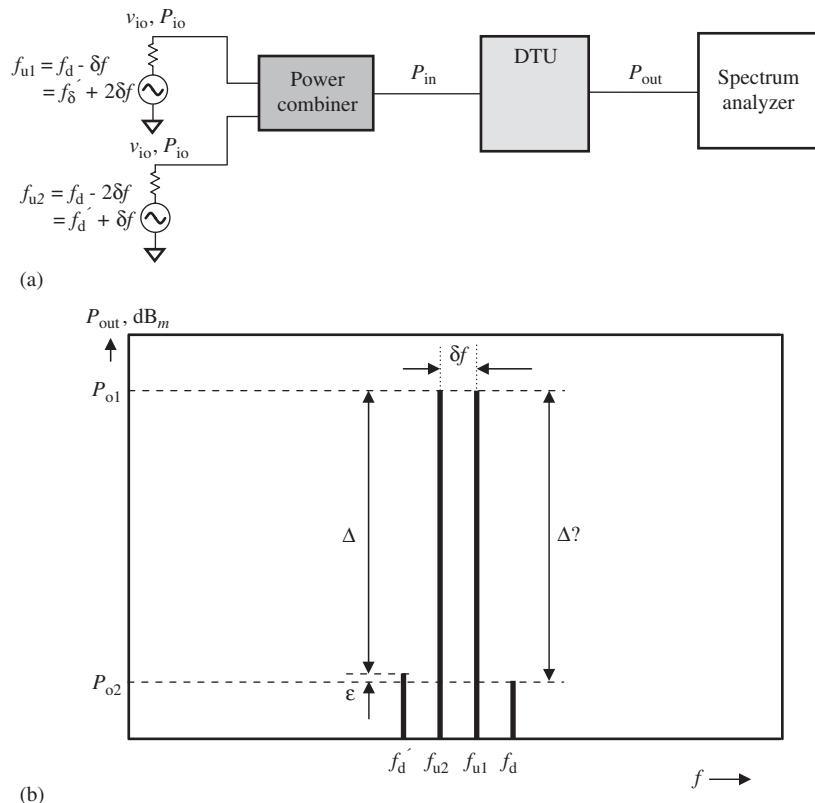
$$\text{IMR}_3 = \alpha(\text{IIP}_3 - P_{i,s}) = \frac{2}{3}(\text{IIP}_3 - P_{i,s}), \quad (12.169)$$

$$\alpha = \frac{2}{3}. \quad (12.170)$$

By directly measuring  $\Delta$  in the test laboratory, one can calculate  $\text{IIP}_3$  and then  $\text{IMR}_3$  by means of equations (12.169) and (12.171).

Figure 12.23(a) shows the setup of  $\text{IP}_3$  testing. There are two generators generating two undesired signals with the same power level  $v_{\text{io}}$  or  $P_{\text{io}}$  but with two different undesired frequencies  $f_{u1}$  and  $f_{u2}$ . These two input signals are combined together by a power combiner and then delivered to the DTU. The power output  $P_{\text{out}}$  from the DTU is displayed on the screen of the spectrum analyzer.

Figure 12.23(b) shows the display of  $\text{IP}_3$  testing on the screen of the spectrum analyzer. In addition to the two undesired signals at the frequencies  $f_{u1}$  and  $f_{u2}$ , two



**Figure 12.23.** Setup and display of IP<sub>3</sub> testing. (a) Setup of IP<sub>3</sub> testing. (b) Display of IP<sub>3</sub> testing.

spurious products of third order with frequency  $f_d$  and  $f'_d$  appear on the screen of the spectrum analyzer as well. The relations between  $f_{u1}$ ,  $f_{u2}$ ,  $f_d$ , and  $f'_d$  are

$$f_d = 2f_{o1} - f_{o2}, \quad \text{or} \quad \omega_d = 2\omega_{u1} - \omega_{u2}, \quad (12.171)$$

$$f'_d = 2f_{o2} - f_{o1}, \quad \text{or} \quad \omega'_d = 2\omega_{u2} - \omega_{u1}. \quad (12.172)$$

This is due to the two input undesired frequencies, which are

$$f_{o1} = f_d - \delta f = f'_d + 2\delta f, \quad \text{or} \quad \omega_{u1} = \omega_d - \delta\omega = \omega'_d + 2\delta\omega, \quad (12.173)$$

$$f_{o2} = f_d - 2\delta f = f'_d + \delta f, \quad \text{or} \quad \omega_{u2} = \omega_d - 2\delta\omega = \omega'_d + \delta\omega. \quad (12.174)$$

The difference  $\Delta$  between the desired and undesired output powers can be read from the screen of the spectrum analyzer, that is,

$$\Delta = P_{o1} - P_{o2}. \quad (12.175)$$

And note that

$$P_i = P_{io}. \quad (12.176)$$

By substituting (12.177) and (12.178) into (12.169), (12.170) and (12.171), the values of  $\text{IIP}_3$ ,  $\text{OIP}_3$ , and  $\text{IMR}_3$  can be obtained, respectively.

This is one way to obtain the value of  $\text{IMR}_3$  by means of the test setup shown in Figure 12.23(a).

There is a minor problem that occurs quite often in the actual testing. Sometimes, the two values of  $\Delta$  at  $f_d$  and  $f'_d$  appearing on the display screen are not the same, but have a difference  $\varepsilon$ . This is due to the large number of computation iterations in the computer system and results from the truncation of decimal numbers. The case is worse with the lesser  $\Delta$ , and should be taken into account in the calculation of  $\text{IP}_3$  or  $\text{IMR}_3$ .

Another way to obtain the value of  $\text{IMR}_3$  is to calculate it from the nonlinear coefficients of the nonlinear device or system if the transfer function as shown in expression (12.128) is known.

Recalling the expressions from (12.128) to (12.137), the transfer function (12.128) can be rewritten as

$$\begin{aligned}
 v_{\text{io}} = & [a_0 + a_2 v_{\text{io}}^2 + (9/4)a_4 v_{\text{io}}^4 + \dots] + [a_1 v_{\text{io}} + (9/4)a_3 v_{\text{io}}^3 + \dots] \cos \omega_1 t \\
 & + [a_1 v_{\text{io}} + (9/4)a_3 v_{\text{io}}^3 + \dots] \cos \omega_2 t + [(1/2)a_2 v_{\text{io}}^2 + 2a_4 v_{\text{io}}^4 + \dots] \cos 2\omega_1 t \\
 & + [(1/2)a_2 v_{\text{io}}^2 + 2a_4 v_{\text{io}}^4 + \dots] \cos 2\omega_2 t + [a_2 v_{\text{io}}^2 + 3a_4 v_{\text{io}}^4 + \dots] \cos(\omega_1 - \omega_2)t \\
 & + [a_2 v_{\text{io}}^2 + 3a_4 v_{\text{io}}^4 + \dots] \cos(\omega_1 + \omega_2)t + [a_3 v_{\text{io}}^3 + \dots] \cos 3\omega_1 t \\
 & + [a_3 v_{\text{io}}^3 + \dots] \cos 3\omega_2 t + [(3/4)a_3 v_{\text{io}}^3 + \dots] \cos(2\omega_2 - \omega_1)t \\
 & + [(3/4)a_3 v_{\text{io}}^3 + \dots] \cos(2\omega_2 + \omega_1)t + [(3/4)a_3 v_{\text{io}}^3 + \dots] \cos(2\omega_1 - \omega_2)t \\
 & + [(3/4)a_3 v_{\text{io}}^3 + \dots] \cos(2\omega_1 + \omega_2)t + [(1/8)a_4 v_{\text{io}}^4 + \dots] \cos 4\omega_1 t \\
 & + [(1/8)a_4 v_{\text{io}}^4 + \dots] \cos 4\omega_2 t + [(1/2)a_4 v_{\text{io}}^4 + \dots] \cos(3\omega_2 - \omega_1)t \\
 & + [(1/2)a_4 v_{\text{io}}^4 + \dots] \cos(3\omega_2 + \omega_1)t + [(1/2)a_4 v_{\text{io}}^4 + \dots] \cos(3\omega_1 - \omega_2)t \\
 & + [(1/2)a_4 v_{\text{io}}^4 + \dots] \cos(3\omega_1 + \omega_2)t + [(3/4)a_4 v_{\text{io}}^4 + \dots] \cos 2(\omega_1 - \omega_2)t \\
 & + [(3/4)a_4 v_{\text{io}}^4 + \dots] \cos 2(\omega_1 + \omega_2)t. \tag{12.177}
 \end{aligned}$$

Let us focus on the term containing the frequency  $(2\omega_2 - \omega_1)$  in the transfer function (12.179), that is,

$$v_{\text{out}} = \dots + [(3/4)a_3 v_{\text{io}}^3 + \dots] \cos(2\omega_1 - \omega_2)t + \dots \tag{12.178}$$

Now if

$$f_1 \rightarrow f_{\text{u}1}, \quad \text{or} \quad \omega_1 \rightarrow \omega_{\text{u}1}, \tag{12.179}$$

$$f_2 \rightarrow f_{\text{u}2}, \quad \text{or} \quad \omega_2 \rightarrow \omega_{\text{u}2}, \tag{12.180}$$

then from (12.173), we have

$$2\omega_1 - \omega_2 = 2\omega_{\text{u}1} - \omega_{\text{u}2} = \omega_{\text{d}}. \tag{12.181}$$

Equation (12.180) can be rewritten as

$$v_{\text{out},u} = \dots + [(3/4)a_3 v_{iu}^3 + \dots] \cos \omega_d t + \dots \quad (12.182)$$

where  $v_{iu}$  = the voltage amplitude of undesired signal.

This is the IM product of the two undesired signals with frequencies  $\omega_{u1}$  or  $\omega_{u2}$  and with the same amplitude  $v_{io}$ . The measurement of IM rejection is to raise this undesired signals' level  $v_{\text{out},u}$  up to a reference level of desired signal  $v_{\text{out},d}$ , which is

$$v_{\text{out},d} = \dots + [a_1 v_{io} + \dots] \cos \omega_d t + \dots, \quad (12.183)$$

where  $v_{io}$  = the voltage of the desired signal.

This implies that the term with  $\cos 2\omega_d t$  in expression (12.184) for the transfer function of the undesired signal corresponds to the term with  $\cos \omega_d t$  in expression (12.185) for the transfer function of the desired signal. Then, if the terms higher than fourth-order nonlinearity in expression (12.184) and the terms higher than second-order nonlinearity in expression (12.185) are neglected, from the expressions (12.184) and (12.185) we have

$$a_1 v_{iu} \cos \omega_d t \approx \frac{3}{4} a_3 v_{iu}^3 \cos(2\omega_{u1} - \omega_{u2})t. \quad (12.184)$$

Then

$$v_{iu} \approx \left( \frac{4}{3} \frac{a_1}{a_3} v_{io} \right)^{1/3}. \quad (12.185)$$

Consequently,

$$\text{IMR}_{3,v} = \frac{v_{iu}}{v_{io}} \approx \left( \frac{4}{3} \frac{a_1}{a_3} \right)^{1/3} v_{io}^{-2/3}, \quad (12.186)$$

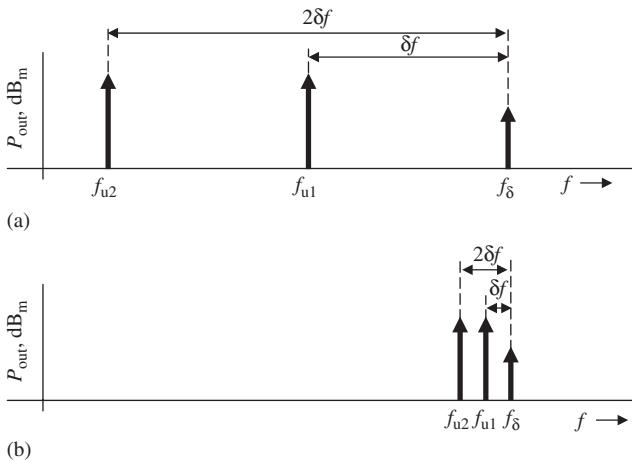
$$\text{IMR}_{3,dB} = 20 \log \frac{v_{iu}}{v_{io}} \approx 20 \log \left[ \left( \frac{4}{3} \frac{a_1}{a_3} \right)^{1/3} v_{io}^{-2/3} \right]. \quad (12.187)$$

The third-order IMR, that is  $\text{IMR}_3$ , is not only dependent on the ratio of the nonlinearity coefficients  $a_1$  and  $a_3$  but is also sensitive to the input signal voltage amplitude  $v_{io}$ . The higher the signal voltage amplitude, the less the  $\text{IMR}_3$ .

Finally, it must be pointed out that there is a unique feature of the third-order spurious product: the difference between the two undesired frequencies  $\delta f$  could be any value! There is no restriction on  $\delta f$  either in the  $\text{IP}_3$  testing setup as shown in Figure 12.23(a) or in expressions (12.175) and (12.176). The range of  $\delta f$  could be very wide, say, several thousand megahertz, or very narrow, say, a few hertz, as shown in Figure 12.24.

Consequently, in practical circuit or system designs, testing for the third-order spurious product must be conducted very carefully, in which the two undesired frequencies  $\delta f$  must cover a wide frequency range from a few kilohertz to couple hundred megahertz.

Theoretically, the nonlinearity of a device or a system could be described by any order of IM or spurious product. However, by means of the unique feature of the third-order spurious product, the third-order intercept point  $\text{IP}_3$  is chosen by the authorities as a criterion to judge the nonlinearity of a device or a system. Let us demonstrate this with the example of two cellular phones produced by Motorola and Nokia, respectively.



**Figure 12.24.** Undesired signal frequencies  $f_{u1}$  and  $f_{u2}$  and desired signal frequency  $f_d$ . (a)  $\delta f$  is wide. (b)  $\delta f$  is narrow.

Assume that both phones are operating in the same operating frequency range and have the same channel spacing or bandwidth, say, 12.5 kHz. In order to compare their nonlinearity performance without prejudice, the FCC (Federal Communication Committee) could choose a value of  $\delta f$  less than 6.25 KHz. Because of this choice, in the examination by the FCC, the two undesired signals are entirely located within the channel bandwidth and can never be filtered away by either the Motorola or the Nokia phone because a filter design with a few kilohertz of bandwidth in the RF frequency range is almost impossible. The examination by the FCC would do justice to both Motorola and Nokia phones without any prejudice. This is why the third-order spurious product, but not the other orders of spurious products, is taken as the criterion of nonlinearity in a communication system.

**12.4.2.3 The 1-dB Compression Point and  $IP_3$ .** The intercept point as shown in Figure 12.20 or 12.21 is a linear approximation of the actual measurement. The actual testing result for  $IP_3$  can be represented by Figure 12.25.

There are two curves shown on the plot. They are straight segments with slopes  $m = 1$  or  $m = 3$ , respectively, when the input power  $P_{\text{in}}$  is low, but start to bend downwards when the input power  $P_{\text{in}}$  is increased up to a certain level. As the input power continuously increases, the plots then deviate from straight lines more and more. The actual intercept point is marked as  $IP_{3,\text{act}}$  in Figure 12.25. The two straight segments can be extended in the increasing direction of the input power. The extended intercept point  $IP_{3,\text{ext}}$  of these two straight segments with slopes of  $m = 1$  and  $m = 3$  is the third-order intercept point that we are looking for. The value of  $IIP_3$  at the extended intercept point  $IP_{3,\text{ext}}$  is lower than that at the actual intercept point  $IP_{3,\text{act}}$ , whereas the value of  $OIP_3$  at the extended intercept point  $IP_{3,\text{ext}}$  is higher than that at the actual intercept point  $IP_{3,\text{act}}$ .

Testing for the third intercept point is somewhat more complicated than testing only for the curve containing the line with a slope of  $m = 1$ . One can easily get stuck in testing only the curve containing the line with a slope of  $m = 1$  if the third-order intercept point  $IP_{3,\text{ext}}$  can be evaluated from the bending point because the bending point is correlated with the third-order intercept point and is not too far from the extended intercept point  $IP_{3,\text{ext}}$ .

Instead of the bending point, the 1-dB compression point is utilized to evaluate the third-order intercept point  $IP_{3,\text{ext}}$ . As shown in Figure 12.25, the 1-dB compression point is a special input power  $P_{1\text{dB}}$  at which the output power deviates by 1 dB from

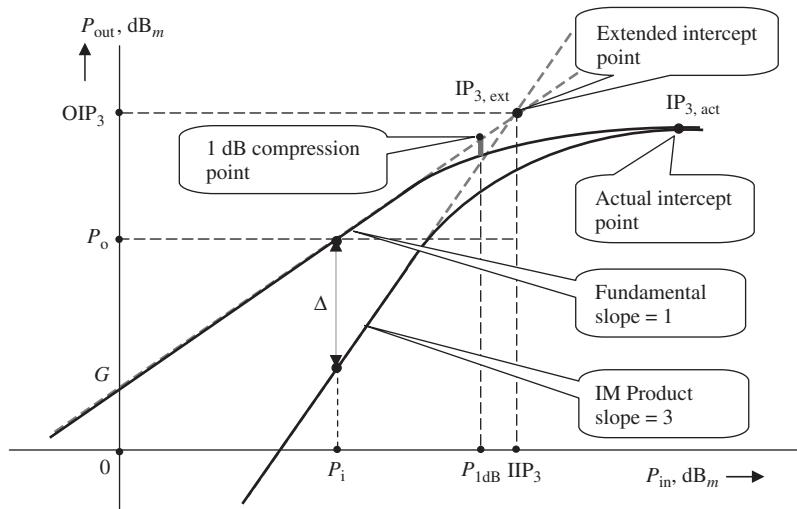


Figure 12.25. 1-dB compression point and  $IP_3$ .

the extended straight line with a slope of  $m = 1$ . The relationship between the 1-dB compression point,  $P_{1\text{dB}}$ , and the third-order input intercept point,  $IIP_3$ , can be evaluated as

$$IIP_3 \rightarrow P_{1\text{dB}} + (3-10) \text{ dB}. \quad (12.188)$$

The uncertain factor,  $(3-10)$  dB, depends on the type of device, circuit topology, current drain, DC voltage supply, and other parameters of the device or system.

In order to cover enough of the linear portion, the input power in the third-order intercept point testing should be

$$P_i < -30 \text{ dB}_m \quad (12.189)$$

so as to ensure that the plot can be extended from the linear portion to get the third-order intercept point  $IP_{3,\text{ext}}$ .

### 12.4.3 Cascaded Equations of Intercept Point

Let us derive the cascaded equation for the intercept point of a system containing only two blocks as shown in Figure 12.26, where  $G_k$  and  $G_{k+1}$  are the transducer power gains of blocks  $k$  and  $k + 1$ , respectively, and  $IP_{mk}$  and  $IP_{mk+1}$  are the  $m$ th-order input intercept point of blocks  $k$  and  $k + 1$ , respectively.

There are simple geometrical relations shown in Figure 12.20, which are

$$\Delta = d(m - 1), \quad (12.190)$$

$$\Delta = P_o - P_u, \quad (12.191)$$

$$d = OIP_m - P_o. \quad (12.192)$$

We have

$$P_u = P_o - \Delta = P_o - d(m-1), \quad (12.193)$$

$$P_u = P_o - (\text{OIP}_m - P_o)(m-1), \quad (12.194)$$

$$P_u = mP_o - (m-1)\text{OIP}_m. \quad (12.195)$$

In the numeric scale or in units of watts, equation (12.197) becomes

$$P_u = \frac{P_o^m}{\text{OIP}_m^{(m-1)}}. \quad (12.196)$$

Referring to the individual blocks  $k$  and  $k+1$  in Figure 12.26, equation (12.197) can be rewritten as

$$P_{o,k} = \frac{P_{o,k}^m}{\text{OIP}_{m,k}^{(m-1)}}, \quad (12.197)$$

$$P_{o,k+1} = \frac{P_{o,k+1}^m}{\text{OIP}_{m,k+1}^{(m-1)}}, \quad (12.198)$$

where

- $P_{o,k}$  = the desired output power from the  $k$ th block,
- $P_{o,k+1}$  = the desired output power from the  $(k+1)$ th block,
- $P_{u,k}$  = the undesired output power from the  $k$ th block,
- $P_{u,k+1}$  = the undesired output power from the  $(k+1)$ th block,
- $\text{OIP}_{m,k}$  = the output intercept point from the  $k$ th block, and
- $\text{OIP}_{m,k+1}$  = the output intercept point from the  $(k+1)$ th block.

Note that

$$P_{o,k} = \frac{P_{o,k+1}}{G_{k+1}}. \quad (12.199)$$

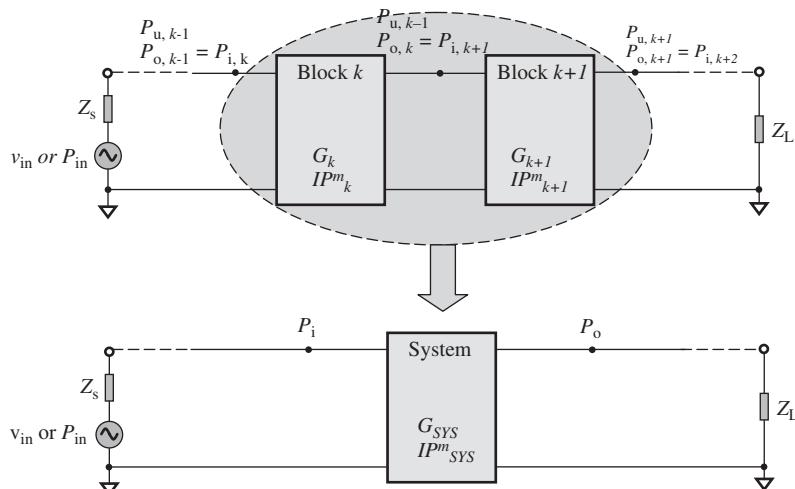


Figure 12.26. Intercept point of a system cascaded by two block  $k$  and  $k+1$ .

Then

$$P_{o,k}^m = \frac{P_{o,k+1}^m}{G_{k+1}^m}. \quad (12.200)$$

The undesired output power at  $Z_L$  from block  $k$  is

$$P_{o,k} G_{k+1} = \frac{P_{o,k}^m G_{k+1}}{\text{OIP}_{m,k}^{(m-1)}} = \frac{P_{o,k+1}^m G_{k+1}}{G_{k+1}^m \text{OIP}_{m,k}^{(m-1)}} = \frac{P_{o,k+1}^m}{G_{k+1}^{(m-1)} \text{OIP}_{m,k}^{(m-1)}}, \quad (12.201)$$

where the expression (12.199),  $P_{o,k} = \frac{P_{o,k}^m}{\text{OIP}_{m,k}^{(m-1)}}$ , and the relation,  $P_{o,k}^m = \frac{P_{o,k+1}^m}{G_{k+1}^m}$ , are applied in the expression (12.203).

From equation (12.203), the undesired voltage at  $Z_L$  from block  $k$  is

$$v_{u,k} = \sqrt{P_{u,k} G_{k+1} \cdot Z_L} = \left[ \frac{P_{o,k+1}^m Z_L}{G_{k+1}^{(m-1)} \text{OIP}_{m,k}^{(m-1)}} \right]^{1/2}. \quad (12.202)$$

From equation (12.200), the undesired voltage at  $Z_L$  from block  $k + 1$  is

$$V_{u,k+1} = \left[ \frac{P_{o,k+1}^m Z_L}{\text{OIP}_{m,k+1}^{(m-1)}} \right]^{1/2}. \quad (12.203)$$

Then, the total undesired voltage or the system undesired voltage across  $Z_L$  is

$$V_{u,sys} = V_{u,k} + V_{u,k+1} = \sqrt{P_{o,k+1}^m Z_L} \left[ \left( \frac{1}{G_{k+1} \text{OIP}_{m,k}} \right)^{\frac{(m-1)}{2}} + \left( \frac{1}{\text{OIP}_{m,k+1}} \right)^{\frac{(m-1)}{2}} \right]. \quad (12.204)$$

The total undesired output power or the system undesired output power across  $Z_L$  is

$$P_{u,sys} = \frac{V_{u,sys}^2}{Z_L} = P_{o,k+1}^m \left[ \left( \frac{1}{G_{k+1} \text{OIP}_{m,k}} \right)^{\frac{(m-1)}{2}} + \left( \frac{1}{\text{OIP}_{m,k+1}} \right)^{\frac{(m-1)}{2}} \right]^2, \quad (12.205)$$

$$\frac{P_{u,sys}}{P_{o,k+1}} = P_{o,k+1}^{(m-1)} \left[ \left( \frac{1}{G_{k+1} \text{OIP}_{m,k}} \right)^{\frac{(m-1)}{2}} + \left( \frac{1}{\text{OIP}_{m,k+1}} \right)^{\frac{(m-1)}{2}} \right]^2. \quad (12.206)$$

At the system intercept point,

$$P_{u,sys} = P_{o,k+1}, \quad (12.207)$$

$$P_{o,k+1} = \text{OIP}_{m,sys}, \quad (12.208)$$

and then equation (12.208) becomes

$$\left(\frac{1}{\text{OIP}_{m,\text{sys}}}\right)^{\frac{(m-1)}{2}} = \left(\frac{1}{G_{k+1} \text{OIP}_{m,k}}\right)^{\frac{(m-1)}{2}} + \left(\frac{1}{\text{OIP}_{m,k+1}}\right)^{\frac{(m-1)}{2}}. \quad (12.209)$$

Mathematically, equation (12.211) looks like a very nice form of the relationship between the total output intercept point,  $\text{OIP}_{m,\text{sys}}$ , and the individual output intercept point,  $\text{OIP}_{m,k}$  and  $\text{OIP}_{m,k+1}$ . The total output intercept point,  $\text{OIP}_{m,\text{sys}}$ , is not only a function of the individual output intercept point,  $\text{OIP}_{m,k}$  and  $\text{OIP}_{m,k+1}$ , but also is related to the gain of the second individual block,  $G_{k+1}$ .

Now let us transfer equation (12.211) from the output intercept point,  $\text{OIP}_m$ , to the input intercept point,  $\text{IIP}_m$ . Note that

$$\text{IIP}_{m,\text{sys}} = \frac{\text{OIP}_{m,\text{sys}}}{G_{\text{sys}}} = \frac{\text{OIP}_{m,\text{sys}}}{G_k G_{k+1}}. \quad (12.210)$$

Also,

$$\text{IIP}_{m,k} = \frac{\text{OIP}_{m,k}}{G_k}, \quad (12.211)$$

and

$$\text{IIP}_{m,k+1} = \frac{\text{OIP}_{m,k+1}}{G_{k+1}}. \quad (12.212)$$

Then we have

$$\left(\frac{1}{\text{IIP}_{m,\text{sys}}}\right)^{\frac{(m-1)}{2}} = \left(\frac{1}{\text{IIP}_{m,k}}\right)^{\frac{(m-1)}{2}} + \left(\frac{G_k}{\text{IIP}_{m,k+1}}\right)^{\frac{(m-1)}{2}}. \quad (12.213)$$

This can be rewritten as

$$\left(\frac{1}{\text{IIP}_{m,\text{sys}}}\right)^{\frac{(m-1)}{2}} = \left(\frac{1}{\text{IIP}_{m,k}}\right)^{\frac{(m-1)}{2}} + \left(\frac{1}{\frac{\text{IIP}_{m,k+1}}{G_k}}\right)^{\frac{(m-1)}{2}}. \quad (12.214)$$

A special feature of the intercept point can be found from equation (12.215) or (12.216). The system intercept point or the resulting intercept point  $\text{OIP}_{\text{sys}}$  or  $\text{IIP}_{\text{sys}}$  looks like the formula of resistance of two resistors connected together in parallel. From the cascaded equation (12.216), it can be seen that the intercept point of second block  $\text{IIP}_{k+1}$  contributes more to the system intercept point  $\text{IIP}_{\text{sys}}$  than the intercept point of first block  $\text{IIP}_k$  because  $\text{IIP}_k$  has same weight as  $\text{IIP}_{k+1}/G_k$  in the equation (12.216). In other words, the gain in first block  $G_k$  promotes the function of the intercept point in second block. This is the reverse of the feature in the cascaded equation of noise figure described in Section 12.3.5. High power gain in first block is harmful to the system intercept point but beneficial to the system noise figure.

In order to become further familiar with this feature of cascaded intercept points, Table 12.6 lists different system intercept points contributed from two blocks with different settlements of intercept point and gain.

TABLE 12.6. Different System Intercept Point Contributed from Two Blocks with Different Settlements of Intercept Point and Gain

<i>Example 1</i>	Block <i>k</i> LNA	Block <i>k</i> + 1 Mixer	System
IIP <sub>3</sub> , dB <sub>m</sub>	3	100	3.0
Gain, dB	10	5	15.0
<i>Example 2</i>	Block <i>k</i> LNA	Block <i>k</i> + 1 Mixer	System
IIP <sub>3</sub> , dB <sub>m</sub>	100	5	0.0
Gain, dB	5	5	10.0
<i>Example 3</i>	Block <i>k</i>	Block <i>k</i> + 1	System
IIP <sub>3</sub> , dB <sub>m</sub>	3	5	-5.6
Gain, dB	10	5	15.0
<i>Example 4</i>	Block <i>k</i> LNA	Block <i>k</i> + 1 Mixer	System
IIP <sub>3</sub> , dB <sub>m</sub>	3	5	-11.2
Gain, dB	16	5	21.0

In example 1, the system IIP<sub>3,sys</sub> is equal to the IIP<sub>3</sub> of block *k* because the IIP<sub>3,k+1</sub> of block *k* + 1 is 100 dB<sub>m</sub>, which is extreme high.

In example 2, the system IIP<sub>3,sys</sub> is contributed almost only by the IIP<sub>3,k+1</sub> of block *k* + 1 because the IIP<sub>3,k</sub> of block *k* is 100 dB<sub>m</sub> which is also extreme high. However, the system IIP<sub>3,sys</sub> is not the same as IIP<sub>3,k+1</sub> of block *k* + 1 because the gain *G<sub>k</sub>* of block *k* declines the contribution of IIP<sub>3,k+1</sub> to the system IIP<sub>3,sys</sub>.

In example 3, the system IIP<sub>3,sys</sub> is lower than that in example 1 because the IIP<sub>3,k+1</sub> of block *k* + 1 in example 3 is much lower than that in example 1.

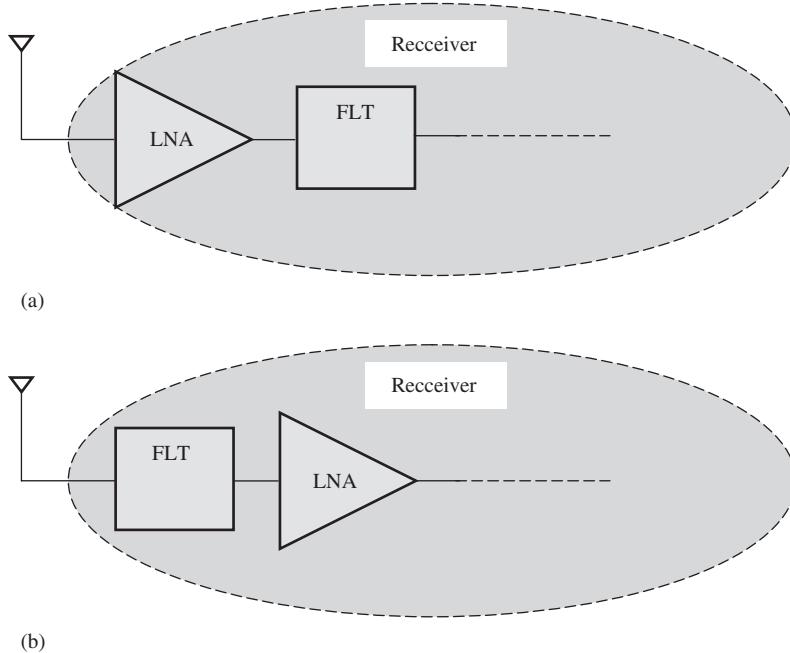
In example 4, the system IIP<sub>3,sys</sub> is lower than that in example 3 because the gain *G<sub>k</sub>* = 16 dB of block *k* in example 4 is higher than that in example 3.

The opposite feature of cascaded equations between the noise figure and the intercept point leads to different considerations in the front-end design of a receiver. Figure 12.27(a) is one type of front-end receiver design which is beneficial to the system noise figure, while Figure 12.27(b) is another type of front-end receiver design which is beneficial to the system intercept point. It should be noted that the sensitivity of a receiver is directly related to the system noise figure, while the linearity of a system is directly related to the system intercept point.

Now let us extend our discussion to a general case: a system consisting of *n* blocks as shown in Figure 12.28. Equations (12.211) and (12.216) can be extended as

$$\begin{aligned} \left( \frac{1}{\text{OIP}_{m,\text{SYS}}} \right)^{\frac{(m-1)}{2}} &= \left( \frac{1}{\text{OIP}_{m,1} G_2 G_3 G_4 \dots G_n} \right)^{\frac{(m-1)}{2}} + \left( \frac{1}{\text{OIP}_{m,2} G_3 G_4 \dots G_n} \right)^{\frac{(m-1)}{2}} + \dots \\ &\quad + \left( \frac{1}{\text{OIP}_{m,n-1} G_n} \right)^{\frac{(m-1)}{2}} + \left( \frac{1}{\text{OIP}_{m,n}} \right)^{\frac{(m-1)}{2}} \end{aligned} \quad (12.215)$$

$$\left( \frac{1}{\text{IIP}_{m,\text{SYS}}} \right)^{\frac{(m-1)}{2}} = \left( \frac{1}{\text{IIP}_{m,1}} \right)^{\frac{(m-1)}{2}} + \left( \frac{G_1}{\text{IIP}_{m,2}} \right)^{\frac{(m-1)}{2}} + \left( \frac{G_1 G_2}{\text{IIP}_{m,3}} \right)^{\frac{(m-1)}{2}} + \dots$$



**Figure 12.27.** Two types of receiver front-end design. (a) This receiver front-end design is beneficial to the system sensitivity. (b) This receiver front-end design is beneficial to the system linearity.

$$+ \left( \frac{G_1 G_2 G_3 \cdots G_{n-1}}{\text{IIP}_{m,n}} \right)^{\frac{(m-1)}{2}}. \quad (12.216)$$

Or,

$$\left( \frac{1}{\text{OIP}_{m,\text{SYS}}} \right)^{\frac{(m-1)}{2}} = \sum_{k=1}^{n-1} \left( \frac{1}{\text{OIP}_{m,k} \prod_{j=k+1}^n G_j} \right)^{\frac{(m-1)}{2}} + \left( \frac{1}{\text{OIP}_{m,n}} \right)^{\frac{(m-1)}{2}}, \quad (12.217)$$

$$\left( \frac{1}{\text{IIP}_{m,\text{SYS}}} \right)^{\frac{(m-1)}{2}} = \left( \frac{1}{\text{IIP}_{m,1}} \right)^{\frac{(m-1)}{2}} + \sum_{k=2}^n \left( \frac{\prod_{j=1}^{k-1} G_j}{\text{IIP}_{m,k}} \right)^{\frac{(m-1)}{2}}. \quad (12.218)$$

Equations (12.211), (12.215), (12.219), and (12.220), are cascaded equations of the intercept point.

It should be noted that in the derivations of the cascaded equations above, all the individual blocks are assumed to have flat frequency responses. This implies that the frequency response of a block to the input desired signal with frequency  $f_d$  and the input spurious component with frequency  $f_s$  has the same gain, while the undesired IM product with the frequency  $f_u$  is produced. This is not true if there is frequency selectivity in a block, where the gain is not the same for the input desired signal with frequency  $f_d$  and the input spurious component with frequency  $f_s$ . Sagers (1982) pointed out that the cascaded equations of intercept point must be modified as follows:

Assuming that block  $k$  in Figure 12.29 has an additional selectivity  $\text{SEL}_k$  dB at the input spurious frequency  $f_s$ , where the gain at the input spurious frequency  $f_s$  is  $\text{SEL}_k$  dB

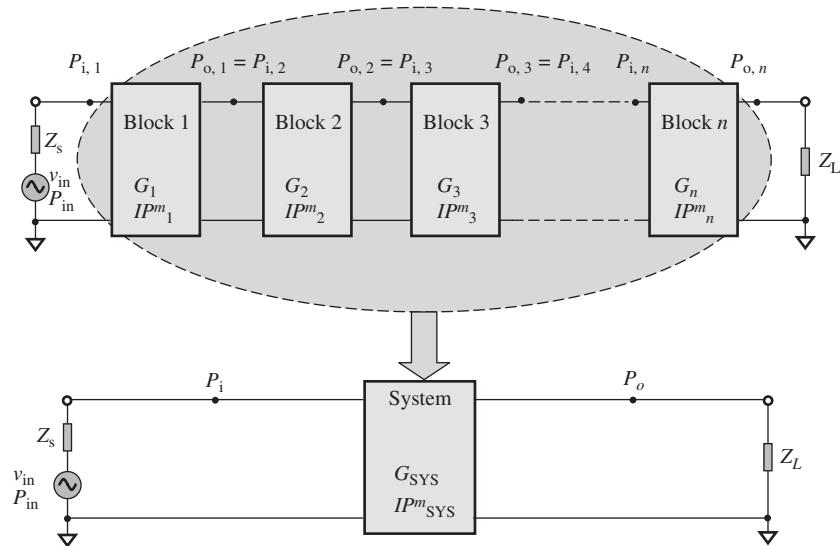


Figure 12.28. Intercept point of a system cascaded by  $n$  blocks.

lower than the gain at the input desired signal frequency  $f_d$ , the cascaded equations of intercept point (12.211), (12.215), (12.219), and (12.220), must be modified with three steps:

1. The block  $k$  having an additional selectivity should be combined with the sequential block  $k + 1$  to form a new block  $k'$  as shown in Figure 12.30.  
The power gain of the new block  $k'$  is

$$G'_k = G_k + G_{k+1}, \quad (12.219)$$

where

$G'_k$  = the combined power gain to the desired input signal in the new block  $k'$ ,

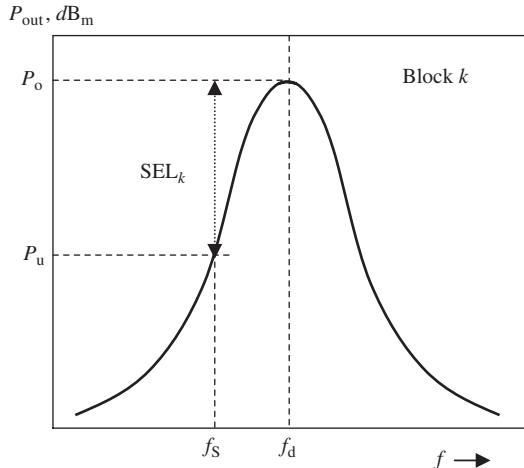
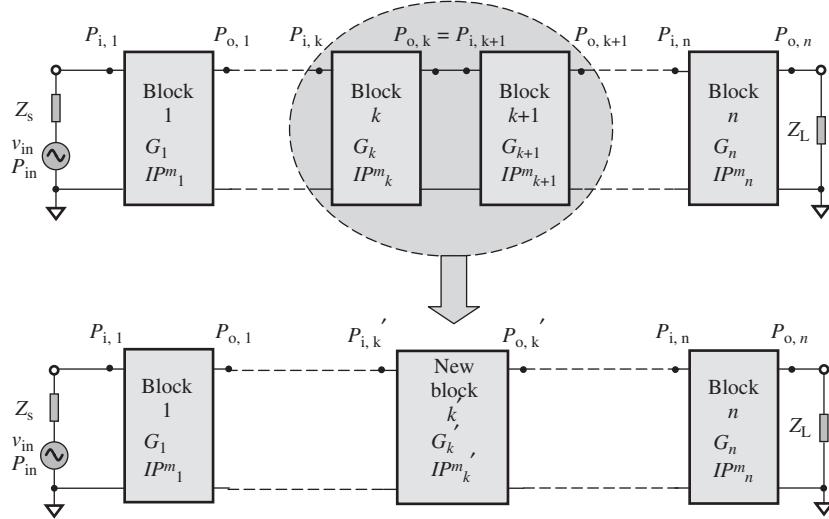


Figure 12.29. Selectivity at the input spurious frequency  $f_s$  in block  $k$ .



**Figure 12.30.** Combining of block  $k$  and  $k + 1$  as a new block  $k'$ .

$G_k$  = the power gain to the desired input signal in the block  $k$ , and  
 $G_{k+1}$  = the power gain to the desired input signal in the block  $k + 1$ .

2. The input intercept point,  $\text{IIP}_m$  of the new block  $k'$  is

$$\text{IIP}'_k = \text{IIP}_{m+1} (\text{SEL}_k)^{\frac{m}{m-1}} \quad (12.220)$$

where

$\text{IIP}'_k$  = the combined  $\text{IIP}_m$  to the desired input signal in the new block  $k'$ ,

$\text{IIP}_m$  = the  $\text{IIP}_m$  to the desired input signal in the block  $k$ ,

$\text{IIP}_{m+1}$  = the  $\text{IIP}_m$  to the desired input signal in the block  $k + 1$ , and

$\text{SEL}_k$  = the selectivity of block  $k$  at spurious frequency  $f_s$ .

3. The output intercept point  $\text{OIP}_m$  of the new block  $k'$  is

$$\text{OIP}'_k = G'_k \text{IIP}'_k = G'_k \text{IIP}_{m+1} \text{SEL}^{\frac{m}{m-1}}. \quad (12.221)$$

where

$\text{OIP}'_k$  = the combined  $\text{OIP}_m$  to the desired input signal in the new block  $k'$ .

#### 12.4.4 Nonlinearity and Distortion

If the original RF signal has an ideal sinusoidal waveform, a distorted RF signal implies that it is not a perfect sinusoidal waveform, but has an imperfect portion, in which either the amplitude or the phase deviates from its ideal state. The deviation from the ideal

state is called *distortion*, which is an intuitive parameter and is measured by percentage, or dB. It is defined as

$$D_{v,\%} = \frac{\Delta V}{V}, \quad (12.222)$$

$$D_{P,\text{dB}} = 20 \log \frac{\Delta V}{V} \approx 20 \log(D_{v,\%}), \quad (12.223)$$

where

$D_v, \%$  = the distortion in %, and

$D_{P,\text{dB}}$  = the distortion in dB.

In terms of relation (12.225), distortion by percent and by decibel can be interconverted. For example, 5% of distortion is equal to about  $-26.02$  dB:

$$(-26.02) \text{ dB} = 20 \log \left( \frac{5}{100} \right). \quad (12.224)$$

It should be noted that the mechanism of the distortion and the noise of an RF signal are essentially different. The noise spreads or swallows its waveform trace but does not bring about distortion, while all the harmonics and all the IM or spurious products contribute to distortion but not to noise. Their impacts on a communication system are essentially different. For instance, a receiver with low noise has a high sensitivity, so the antenna can sense a very weak input signal. However, a considerable distortion could be present in the voice signal. A girl's voice may end up sounding like a boy's voice. On the contrary, a receiver with high noise has low sensitivity, so the antenna can sense an input signal only when the input signal power exceeds a certain level. However, its distortion may be low so that the voice could be recognized easily. In short, both spurious products and noise are not welcome in a communication system, but their impacts on the system are different. Spurious products cause distortion, while noise reduces the sensitivity.

## 12.5 OTHER PARAMETERS

### 12.5.1 Power Supply Voltage and Current Drain

The DC power supply voltage applied into a circuit block or system is becoming lower and lower. The reasons are simple: to reduce the cost and shrink the size of the block or system.

For instance, in a wireless communication system, the DC power supply and current drain are not considered important in the design of base stations or repeaters, but they are quite important in the design of the handset. More battery cells must be provided, and thus higher costs result if a high DC power supply voltage is required.

At present, the DC power supply voltage in most handsets of cellular phones is 3 V. The competition between cellular phone manufacturers is centered on current drain. High current drain in the circuits means high current consumption from the battery. The current consumption of a handset is the product of the total current drain and the total operating hours per day. The units of current consumption are (mA·h) or (A·h).

Total current consumption is important to customers. It is acceptable if a handset can be normally operated all day and requires battery charging at night. However, the

customer will be upset if a handset can only be operated in the morning, and must be charged in the afternoon.

The total current drain  $i_{\text{tot}}$  or the current consumption  $i_{\text{tot}} \cdot H_{\text{tot}}$  of a handset can be evaluated as follows.

First, the total current drain  $i_{\text{tot}}$  can be categorized into three types of subcurrent drain:

1. *Standby Current Drain,  $i_{\text{standby}}$* . This is the current drain when the handset is operating in “waiting” status, that is, the handset is operating neither in transmit mode nor in receive mode, but is merely waiting for incoming signals.
2. *Receive Mode Current Drain,  $i_{\text{receive}}$* . This is the current drain when the handset is operating in receive mode, listening to the incoming signals.
3. *Transmit Mode Current Drain,  $i_{\text{transmit}}$* . This is the current drain when the handset is operating in transmit mode to deliver outbound signals.

Second is the corresponding operating periods during which the handset operates in the respective modes, that is,

- operating period when the handset is in standby mode,  $T_{\text{standby}}$ ,
- operating period when the handset is in receive mode,  $T_{\text{receive}}$ , and
- operating period when the handset is in transmit mode,  $T_{\text{transmit}}$ .

The ratio of these three operating periods differs between customers. For example, for a policeman, the period ratio of these three current drains might be approximately

$$T_{\text{standby}} : T_{\text{receive}} : T_{\text{transmit}} = 1 : 2 : 2.$$

For a student, the period ratio of these three current drains might be

$$T_{\text{standby}} : I_{\text{receive}} : T_{\text{transmit}} = 5 : 1 : 1,$$

and for a housekeeper, the period ratio of these three current drains could be

$$T_{\text{standby}} : I_{\text{receive}} : T_{\text{transmit}} = 10 : 1 : 1.$$

A third parameter is the total operating hours per day,  $T_{\text{tot}}$ .

$$T_{\text{tot}} = T_{\text{standby}} + T_{\text{receive}} + T_{\text{transmit}}$$

Consequently, the average current drain of a handset,  $i_{\text{avg}}$ , can be calculated by the following formula:

$$i_{\text{avg}} = i_{\text{standby}} \frac{T_{\text{standby}}}{T_{\text{tot}}} + i_{\text{receive}} \frac{T_{\text{receive}}}{T_{\text{tot}}} + i_{\text{transmit}} \frac{T_{\text{transmit}}}{T_{\text{tot}}}. \quad (12.225)$$

And the total current consumption can be calculated by the following formula:

$$i_{\text{avg}} T_{\text{tot}} = i_{\text{standby}} T_{\text{standby}} + i_{\text{receive}} T_{\text{receive}} + i_{\text{transmit}} T_{\text{transmit}}. \quad (12.226)$$

The battery of a handset is chosen by evaluating  $i_{\text{avg}} T_{\text{tot}}$  from (12.228).

### 12.5.2 Part Count

Part count is the total number of parts in a circuit block or a system. It is a simple statistical number, but is also an important parameter in a circuit block design or a system design. This is because the reliability of a product, either a circuit block or a system, significantly depends on its part count. In the early stages of developing cell phones, the part count of the entire handset was about 200–300; the reliability was quite low and its price was high. At present, the part count has been significantly reduced from a few hundred down to a few tens, resulting in increased reliability; therefore, its price has correspondingly come down. If the expected SOC (system-on-a-chip) goal is realized in the future, the part count will approach 1, theoretically increasing the reliability to infinity.

## 12.6 EXAMPLE OF RF SYSTEM ANALYSIS

Usually, an RF system consists of a number of blocks. The system parameters must be calculated from the corresponding parameters of all the individual blocks. The basic skill for this task is calculating the system parameters from the corresponding parameters of two individual blocks. As long as we master such a skill for two blocks, it can be extended to a system constructed with more than two blocks.

Today, computer simulation is a very convenient and sophisticated tool not only in circuit design but also in system performance analysis. However, applying cascaded equations on a Microsoft Excel worksheet enables engineers to do a system analysis in a fast and handy way, although such an analysis is confined to the primary stage. Table 12.7 shows such an example.

This is a system analysis table for the front-end of a receiver. It consists of eight blocks, which are as follows:

1. Harmonic filter/balun
2. LNA
3. Mixer
4. IF VGA (Voltage Gain-controlled Amplifier)
5. IF LPF (low-pass filter)
6. IF VGA
7. IF LPF
8. SP amplifier
9. Back end (base band portion).

The ninth block in last column is the back end of the receiver and is excluded and constructed in the baseband portion. The values of performance in the ninth block are provided by the base band designer.

TABLE 12.7. System Analysis of a Receiver Front End by the Cascaded Equations

<b>System Goals</b>	Lower frequency	: 2412	MHz	Temperature	: 300 K°				
	Upper frequency	: 2902.5	MHz	RISE at 12 dB SINAD	: 6.0 dB				
				RISE at 20 dB quiet	: 7.6 dB				
	System BW	: 13.5	kHz	Worst case factor	: 0.1				
<b>Performance</b>									
(Column)	1	2	3	4	5	6	7	8	9
	HF/Balun	LNA	Mixer	IF VGA	IF LPF	IF VGA	IFLPF	SP Amp	Bk.End
<b>Gain</b>	dB	-2.0	16	1	25	-2	25	-2	0
NF	dB	2	3	12	5	2	5	2	0
IP3	dB <sub>m</sub>	20	0	12	0	18	0	18	-13
IP2	dB <sub>m</sub>	100	80	40	100	100	100	100	1000
SEL at Δf	dB	0	0	0	10	0	10	0	0
SEL at Δ2f	dB	0	0	0	15	0	15	0	0
SEL at 1/2 IF	dB	0	10	0	0	0	0	0	0
SEL at 1/2 IF	dB	0	10	0	0	0	0	0	0
Gain(worst)	dB	-2.2	14.4	0.9	22.5	-2.2	22.5	-2.2	-2.2
NF(worst)	dB	2.2	3.3	13.2	5.5	2.2	5.5	2.2	2.2
IP3(worst)	dB <sub>m</sub>	20	0	12	0	18	0	18	-13
IP2,(worst)	dB <sub>m</sub>	100	80	40	100	100	100	100	1000
<b>Calculations</b>		H	G	F	E	D	C	B	A
NF <sub>SYS</sub>	dB	5.8	3.8	12.4	5.0	7.0	5.0	9.0	7.0
12 dB SINAD	μV	6.87	5.46	14.7	6.26	7.89	6.27	9.90	7.86
12 dB SINAD	dB <sub>m</sub>	-90.3	-92.3	-83.6	-91.1	-89.0	-91.0	-87.1	-89.1
20 dB Quiet	μV	8.67	6.89	18.6	7.91	9.97	7.92	12.5	9.93
20 dB Quiet	dB <sub>m</sub>	-88.2	-90.2	-81.6	-89.0	-87.0	-89.0	-85.0	-87.0
IP <sub>3,SYS</sub>	dB <sub>m</sub>	-39.3	-41.3	-25.3	-24.3	-16.8	-18.8	-18.2	-13.2
IMR <sub>3</sub>	dB	31.6	31.6	36.5	42.1	45.8	45.8	48.2	48.2

There are three portions in Table 12.7:

1. *System Goals.* These are general goals for all the blocks, which are not involved in the calculation.
2. *Performance.* The expected goals of performance for each individual block are listed, such as gain, NF, and IP<sub>3</sub>.

**3. Calculation.** Calculations are only conducted in this portion on the basis of the expected goals in the performance portion. In the process of calculation, the calculated results are registered in the corresponding cells in this portion.

Instead of simultaneously doing the calculations for all the blocks, the calculations are confined to only two blocks at a time. In other words, only cascaded equations for two blocks are applied in the calculations in Table 12.7.

In order to show the calculation path, let us take the calculation of the noise figure as an example. In Table 12.7, the NF values of every block in the “Performance” portion and the  $NF_{sys}$  values of every block in the “Calculation” portion are circled with an ellipse. The NF values of each block in the “Performance” portion are the respective expected goals of each individual block and the  $NF_{sys}$  values of each block in the “Calculation” portion is a systematic value of noise figure looking from the block towards all the following blocks on the right side. These values are grouped and named group A, B, C, D, E, F, G, and H, respectively. The calculation path for noise figure is also shown in Table 12.7.

Now calculation is started from group A in the last two blocks. The original noise figure of the back-end block,  $NF_{sys}$ , shown in the calculation portion is 7 dB, which is provided by the base band circuit designer. The noise figure of the SP amplifier, NF, is 0 dB. The new  $NF_{sys}$  is the original noise figure of the back-end block,  $NF_{sys}$ , cascaded with the noise figure of the SP amplifier, NF. In terms of the cascaded equation for two blocks, the new or resultant noise figure  $NF_{sys}$  can be calculated and is obtained as 7 dB, which is registered into the  $NF_{sys}$  cell of the SP amplifier located in the “Calculation” portion.

Following a similar calculation path, the calculation moves from group A to B, and then from B to C, from C to D, from D to E, and so on. At last, the final value of  $NF_{sys}$  appears in group H. This value is calculated and is obtained as 5.8 dB. This is the value of the system NF for the entire receiver.

By a similar calculation path as shown for the noise figure above, other parameters such as gain and intercept point can be calculated starting from the last block and working toward the first block. All the values located in the column of the first block in Table 12.7, the HF/Balun, are the entire system values for the receiver. In Table 12.7, not only the gain, noise figure, and intercept point, but also other parameters, such as the 12-dB SINAD, 20-dB quieting, and IMR are also presented.

It should be noted that in Table 12.7 there are two rows that present the worst case gain and noise figures, in which the worst case factor is  $0.1 = 10\%$ . The purpose of system analysis using the worst case is to retain the necessary “room” for the high-performance reliability of the system.

The calculations performed in Table 12.7 are quite flexible. If using a Microsoft Excel spread sheet, the calculated results are automatically corrected if any expected goal in the “Performance” portion in Table 12.7 is varied. This is very helpful to the system engineer in adjusting the design goals between the individual blocks. Table 12.7 shows a successful design since the final results would produce excellent performance:

- 12-dB, SINAD =  $-90.3 \text{ dB}_m$ ,
- 20-dB quite =  $-88.2 \text{ dB}_m$ ,
- IMR<sub>3</sub> = 31.6 dB.

The first two items represent the sensitivity of the receiver and the last item is related to the distortion of the receiver.

## APPENDICES

### 12.A.1 Conversion between Watts, Volts, and $\text{dB}_m$ in a System with $50 \Omega$ Input and Output Impedance

#### Equations

$$\text{Voltage} : V = \sqrt{50 \times 10^{\frac{P_{\text{dBm}} - 30}{10}}}, \quad (12.\text{A}.1)$$

$$\text{Power} : P = \frac{V^2}{50}, \quad (12.\text{A}.2)$$

$$\text{Power in dB} : P_{\text{dBm}} = 10 \log \left( \frac{V^2}{50} \right) + 30. \quad (12.\text{A}.3)$$

### 12.A.2 Relationship between voltage reflection coefficient, $\Gamma$ , and Transmission coefficients when the load $R_o$ is equal to the standard characteristic resistance, $50 \Omega$

( $R_o = 50 \Omega$ , standard characteristic resistance)

Voltage reflection coefficient:

$$\Gamma = \frac{R - R_o}{R + R_o}. \quad (12.\text{A}.4)$$

Voltage standing wave ratio:

$$\text{VSWR} = \frac{1 + \Gamma}{1 - \Gamma}. \quad (12.\text{A}.5)$$

Voltage standing wave ratio in dB:

$$\text{VSWR}_{\text{dB}} = 20 \log(\text{VSWR}). \quad (12.\text{A}.6)$$

Return loss in dB:

$$\text{RL}_{\text{dB}} = S_{11,\text{dB}} = 20 \log \Gamma. \quad (12.\text{A}.7)$$

Transmission loss in dB:

$$\text{TL}_{\text{dB}} = 10 \log(1 - \Gamma^2). \quad (12.\text{A}.8)$$

Power reflection coefficient:

$$\Gamma_P = \gamma = \Gamma^2. \quad (12.\text{A}.9)$$

Transmitted power in %:

$$P_{T,\%} = 100(1 - \Gamma^2). \quad (12.\text{A}.10)$$

Reflected power in %:

$$P_{R,\%} = 100\Gamma^2. \quad (12.\text{A}.11)$$

TABLE 12.A.1. Conversion between Watts, Volts, and  $\text{dB}_m$  in a System with  $50\Omega$  Input/Output Impedance

$P, \text{dB}_m$	Voltage, mV	Power, mW	$P, \text{dB}_m$	Voltage, $\mu\text{V}$	Power, nW
0	223.61	1.0000	-50	707.11	10.0000
-1	199.29	0.794328	-51	630.21	7.943282
-2	177.62	0.630957	-52	561.67	6.309573
-3	158.30	0.501187	-53	500.59	5.011872
-4	141.09	0.398107	-54	446.15	3.981072
-5	125.74	0.316228	-55	397.64	3.162278
-6	112.07	0.251189	-56	354.39	2.511886
-7	99.88	0.199526	-57	315.85	1.995262
-8	89.02	0.158489	-58	281.50	1.584893
-9	79.34	0.125893	-59	250.89	1.258925
-10	70.71	0.1000	-60	223.61	1.0000
-11	63.02	0.079433	-61	199.29	0.794328
-12	56.17	0.063096	-62	177.62	0.630957
-13	50.06	0.050119	-63	158.30	0.501187
-14	44.62	0.039811	-64	141.09	0.398107
-15	39.76	0.031623	-65	125.74	0.316228
-16	35.44	0.025119	-66	112.07	0.251189
-17	31.59	0.019953	-67	99.88	0.199526
-18	28.15	0.015849	-68	89.02	0.158489
-19	25.09	0.012589	-69	79.34	0.125893
-20	22.36	0.0100	-70	70.71	0.1000
-21	19.93	0.007943	-71	63.02	0.079433
-22	17.76	0.006310	-72	56.17	0.063096
-23	15.83	0.005012	-73	50.06	0.050119
-24	14.11	0.003981	-74	44.62	0.039811
-25	12.57	0.003162	-75	39.76	0.031623
-26	11.21	0.002512	-76	35.44	0.025119
-27	9.99	0.001995	-77	31.59	0.019953
-28	8.90	0.001585	-78	28.15	0.015849
-29	7.93	0.001259	-79	25.09	0.012589
-30	7.07	0.0010	-80	22.36	0.0100
-31	6.30	0.000794	-81	19.93	0.007943
-32	5.62	0.000631	-82	17.76	0.006310
-33	5.01	0.000501	-83	15.83	0.005012
-34	4.46	0.000398	-84	14.11	0.003981
-35	3.98	0.000316	-85	12.57	0.003162
-36	3.54	0.000251	-86	11.21	0.002512
-37	3.16	0.000200	-87	9.99	0.001995
-38	2.82	0.000158	-88	8.90	0.001585
-39	2.51	0.000126	-89	7.93	0.001259
-40	2.24	0.0001	-90	7.07	0.0010
-41	1.99	0.000079	-91	6.30	0.000794
-42	1.78	0.000063	-92	5.62	0.000631
-43	1.58	0.000050	-93	5.01	0.000501
-44	1.41	0.000040	-94	4.46	0.000398
-45	1.26	0.000032	-95	3.98	0.000316
-46	1.12	0.000025	-96	3.54	0.000251
-47	1.00	0.000020	-97	3.16	0.000200
-48	0.89	0.000016	-98	2.82	0.000158
-49	0.79	0.000013	-99	2.51	0.000126

(continued)

TABLE 12.A.1. (Continued)

$P$ , dB <sub>m</sub>	Voltage, mV	Power, mW	$P$ , dB <sub>m</sub>	Voltage, $\mu$ V	Power, nW
-50	707.11	10.0000	-100	2236.07	100.0000
-51	630.21	7.943282	-101	1992.90	79.432823
-52	561.67	6.309573	-102	1776.17	63.095734
-53	500.59	5.011872	-103	1583.01	50.118723
-54	446.15	3.981072	-104	1410.86	39.810717
-55	397.64	3.162278	-105	1257.43	31.622777
-56	354.39	2.511886	-106	1120.69	25.118864
-57	315.85	1.995262	-107	998.81	19.952623
-58	281.50	1.584893	-108	890.19	15.848932
-59	250.89	1.258925	-109	793.39	12.589254
-60	223.61	1.0000	-110	707.11	10.0000
-61	199.29	0.794328	-111	630.21	7.943282
-62	177.62	0.630957	-112	561.67	6.309573
-63	158.30	0.501187	-113	500.59	5.011872
-64	141.09	0.398107	-114	446.15	3.981072
-65	125.74	0.316228	-115	397.64	3.162278
-66	112.07	0.251189	-116	354.39	2.511886
-67	99.88	0.199526	-117	315.85	1.995262
-68	89.02	0.158489	-118	281.50	1.584893
-69	79.34	0.125893	-119	250.89	1.258925
-70	70.71	0.1000	-120	223.61	1.0000
-71	63.02	0.079433	-121	199.29	0.794328
-72	56.17	0.063096	-122	177.62	0.630957
-73	50.06	0.050119	-123	158.30	0.501187
-74	44.62	0.039811	-124	141.09	0.398107
-75	39.76	0.031623	-125	125.74	0.316228
-76	35.44	0.025119	-126	112.07	0.251189
-77	31.59	0.019953	-127	99.88	0.199526
-78	28.15	0.015849	-128	89.02	0.158489
-79	25.09	0.012589	-129	79.34	0.125893
-80	22.36	0.0100	-130	70.71	0.1000
-81	19.93	0.007943	-131	63.02	0.079433
-82	17.76	0.006310	-132	56.17	0.063096
-83	15.83	0.005012	-133	50.06	0.050119
-84	14.11	0.003981	-134	44.62	0.039811
-85	12.57	0.003162	-135	39.76	0.031623
-86	11.21	0.002512	-136	35.44	0.025119
-87	9.99	0.001995	-137	31.59	0.019953
-88	8.90	0.001585	-138	28.15	0.015849
-89	7.93	0.001259	-139	25.09	0.012589
-90	7.07	0.0010	-140	22.36	0.0100
-91	6.30	0.000794	-141	19.93	0.007943
-92	5.62	0.000631	-142	17.76	0.006310
-93	5.01	0.000501	-143	15.83	0.005012
-94	4.46	0.000398	-144	14.11	0.003981
-95	3.98	0.000316	-145	12.57	0.003162
-96	3.54	0.000251	-146	11.21	0.002512
-97	3.16	0.000200	-147	9.99	0.001995
-98	2.82	0.000158	-148	8.90	0.001585
-99	2.51	0.000126	-149	7.93	0.001259
—	—	—	-150	7.07	0.0010

TABLE 12.A.2. Conversion between VSWR and Other Reflection and Transmission Coefficients ( $R_0 = 50\Omega$ , Standard Characteristic Resistance)

$R_o$	$R$	$\Gamma$	VSWR	VSWR <sub>dB</sub>	RL <sub>dB</sub>	TL <sub>dB</sub>	$\Gamma_P$	$P_{T,\%}$	$P_{R,\%}$
50.00	2500.00	0.96	50.00	33.98	-0.35	7.69	0.92	7.69	92.31
50.00	1250.00	0.92	25.00	27.96	-0.70	14.79	0.85	14.79	85.21
50.00	1000.00	0.90	20.00	26.02	-0.87	18.14	0.82	18.14	81.86
50.00	750.00	0.88	15.00	23.52	-1.16	23.44	0.77	23.44	76.56
50.00	500.00	0.82	10.00	20.00	-1.74	33.06	0.67	33.06	66.94
50.00	400.00	0.78	8.00	18.06	-2.18	39.51	0.60	39.51	60.49
50.00	300.00	0.71	6.00	15.56	-2.92	48.98	0.51	48.98	51.02
50.00	200.00	0.60	4.00	12.04	-4.44	64.00	0.36	64.00	36.00
50.00	100.00	0.33	2.00	6.02	-9.54	88.89	0.11	88.89	11.11
50.00	90.00	0.29	1.80	5.11	-10.88	91.84	0.08	91.84	8.16
50.00	80.00	0.23	1.60	4.08	-12.74	94.67	0.05	94.67	5.33
50.00	70.00	0.17	1.40	2.92	-15.56	97.22	0.03	97.22	2.78
50.00	60.00	0.09	1.20	1.58	-20.83	99.17	0.01	99.17	0.83
50.00	50.00	0.00	1.00	0.00	$\infty$	100.00	0.00	100.00	0.00
50.00	41.50	-0.09	1.20	1.62	-20.64	99.14	0.01	99.14	0.86
50.00	35.60	-0.17	1.40	2.95	-15.48	97.17	0.03	97.17	2.83
50.00	31.20	-0.23	1.60	4.10	-12.71	94.64	0.05	94.64	5.36
50.00	27.80	-0.29	1.80	5.10	-10.89	91.86	0.08	91.86	8.14
50.00	25.00	-0.33	2.00	6.02	-9.54	88.89	0.11	88.89	11.11
50.00	12.50	-0.60	4.00	12.04	-4.44	64.00	0.36	64.00	36.00
50.00	8.33	-0.71	6.00	15.57	-2.92	48.97	0.51	48.97	51.03
50.00	6.25	-0.78	8.00	18.06	-2.18	39.51	0.60	39.51	60.49
50.00	5.00	-0.82	10.00	20.00	-1.74	33.06	0.67	33.06	66.94
50.00	3.33	-0.87	15.00	23.52	-1.16	23.44	0.77	23.44	76.56
50.00	2.50	-0.90	20.00	26.02	-0.87	18.14	0.82	18.14	81.86
50.00	2.00	-0.92	25.00	27.96	-0.70	14.79	0.85	14.79	85.21
50.00	1.00	-0.96	50.00	33.98	-0.35	7.69	0.92	7.69	92.31
50.00	0.00	-1.00	$\infty$	$\infty$	0.00	0.00	1.00	0.00	100.00

### 12.A.3 Definition of Powers in a Two-Port Block by Signal Flow Graph

Figure 12.A.1 shows the various powers, voltage reflection coefficients, and  $S$ -parameters in a two-port block and its signal flow graph.

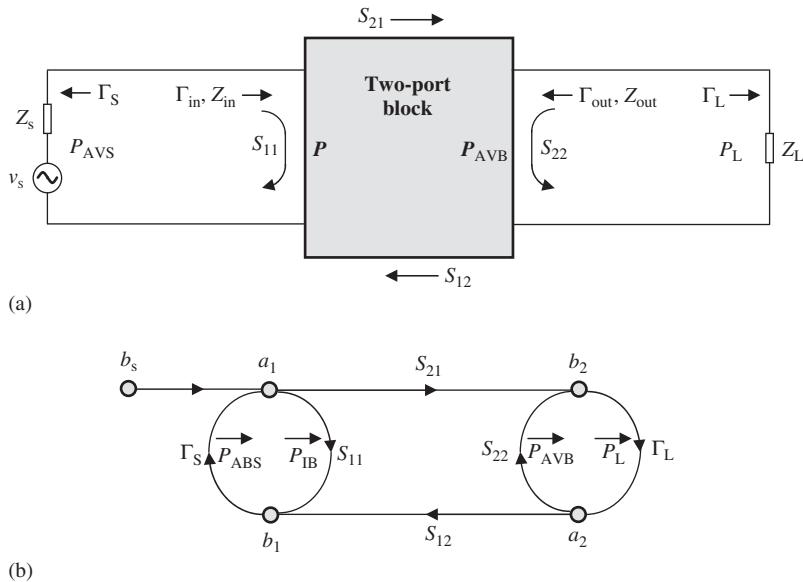
There are four powers in a two-port block. According to the signal flow graph as shown in Figure 12.A.1, they are as follows:

1. Power delivered to the load  $P_L$  :

$$P_L = \frac{|b_s|^2}{|1 - \Gamma_{in}\Gamma_S|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2}, \quad (12.A.12)$$

or

$$P_L = \frac{|b_s|^2}{|1 - S_{11}\Gamma_S|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - \Gamma_{out}\Gamma_L|^2}, \quad (12.A.13)$$



**Figure 12.A.1.** Various powers, voltage reflection coefficients, S-parameters and its signal flow graph in a two-port block. (a) Powers, voltage reflection coefficients, and S-parameters in a two-port block. (b) Signal flow graph of a two-port block.

because

$$\frac{|1 - S_{11}\Gamma_S|^2}{|1 - S_{22}\Gamma_L|^2} = \frac{|1 - \Gamma_{in}\Gamma_S|^2}{|1 - \Gamma_{out}\Gamma_L|^2}. \quad (12.A.14)$$

2. Power available from the block,  $P_{AVB}$ :

$$P_{AVB} = P_L|_{\Gamma_L=\Gamma_{out}^*} = \frac{|b_s|^2}{|1 - S_{11}\Gamma_S|^2} |S_{21}|^2 \frac{1}{1 - |\Gamma_{out}|^2}. \quad (12.A.15)$$

3. Power input to the block  $P_{IB}$ :

$$P_{IB} = \frac{|b_s|^2}{|1 - \Gamma_{in}\Gamma_s|^2}. \quad (12.A.16)$$

4. Power available from the source,  $P_{AVS}$ :

$$P_{AVS} = P_{IB}|_{\Gamma_{in}=\Gamma_S^*} = \frac{|b_s|^2}{1 - |\Gamma_s|^2}. \quad (12.A.17)$$

#### 12.A.4 Main Noise Sources

Any random process of current or voltage in a part or a circuit block results in noise. In other words, noise inevitably exists in a part or a circuit block because in the basic parts, such as the resistor, capacitor, inductor, transistor, and so on, there are many random processes existing in the motions of electrons and other charge carriers. Consequently,

it is found that there are many kinds of noise sources, such as shot noise, thermal noise, flicker noise, burst noise (popcorn noise), avalanche noise, and so on.

There are three main noise sources in a circuit block:

1. *Shot noise.* A current in a device, part, or runner is composed of a large number of moving carriers, either positive charges or electrons with random velocities. The product of the electric charge and velocity of a charge or an electron forms a current element. As a result of the different velocities, these current elements are different from each other. The sum of all the current elements at a junction of a semiconductor or a cross-section of a part looks like a large number of current pulses. The average value of these random current pulses is called *current* and its fluctuation around the average value is called *shot noise*. According to statistics, the mean-square value of its fluctuation is

$$\overline{i_n^2} = 2qI \Delta f, \quad (12.A.18)$$

where

- $i_n$  = the current fluctuation, a random variable,
- $I$  = the average value of current,
- $q$  = the charge of an electron, and
- $\Delta f$  = the bandwidth in which noise source is acting.

The special feature of thermal noise is that it is directly related to the DC current  $I$ . The higher the DC current flowing through the part, the higher the shot noise that the part has.

2. *Thermal noise.* Essentially, thermal noise is due to the fluctuation of resistance in a device, a part, or a runner. It is well known that resistance is directly associated with the collision between electrons and between electrons and other charge carriers or particles. Fluctuation of the resistance in parts or devices implies a fluctuation of collision, which is a random process.

The random thermal motion and collision of electrons and other charge particles produce thermal noise, whereas the random drift velocity of the moving charge carriers or current fluctuation produces shot noise. The velocity of the electron thermal motion is much higher than the drift velocity of the electron and other charge carriers. Therefore, thermal noise is produced by a completely different mechanism than shot noise. Thermal noise still exists even when shot noise is removed by a zero DC current.

According to statistics, the thermal noise of a resistor can be represented by either voltage or current noise source, that is,

$$\overline{e_n^2} = 4kTR\Delta f, \quad (12.A.19)$$

or

$$\overline{i_n^2} = 4kT \frac{1}{R} \Delta f, \quad (12.A.20)$$

where

- $e_n$  = the voltage fluctuation,
- $i_n$  = the current fluctuation,

- $k$  = the Boltzmann constant,  
 $T$  = the room temperature in kelvin,  
 $R$  = the resistance of the resistor, and  
 $\Delta f$  = the bandwidth in which noise source is acting.

The special feature of thermal noise is that its noise power spectrum density is independent of frequency, that is,

$$\frac{\overline{e_n^2}}{4R\Delta f} = kT, \quad (12.A.21)$$

or

$$\frac{\overline{i_n^2}R}{4\Delta f} = kT. \quad (12.A.22)$$

3. *Flicker noise.* Flicker noise is due to contamination and crystal defects. The traps capture and release carriers in a random fashion associated with a noise with energy concentrated at low frequencies. Just like the shot noise, the existence of flicker noise is always associated with a direct current. Therefore, a resistor that is not carrying a direct current does not have flicker noise. According to statistics, flicker noise can be represented by a noise current source as follows:

$$\overline{i_n^2} = kI^a \frac{\Delta f}{f}, \quad (12.A.23)$$

where

- $i_n$  = the current fluctuation,  
 $k$  = a constant for a particular device,  
 $I$  = the direct current,  $a = 0.5$  to  $2$ ,  
 $f$  = the operating frequency, and  
 $\Delta f$  = the operating bandwidth.

## REFERENCES

- Gray PR, Hurst PJ, Lewis SH, et al. Analysis and design of analog integrated circuits. 4th ed. New York: John Wiley & Sons, Inc.; 2001.
- Haus HA, Atkinson WR, Fonger WH, McLeod WW, Branch GM, Harris WA, Stodola EK, Davenport WB, Harrison SW, Telpey TE. Representation of noise in linear twoports. Proc IRE 1960;48:69–74.
- Lee TH. The design of CMOS radio-frequency integrated circuits. Cambridge University Press; 1998.
- Sagers RC. Intercept point and undesired responses. 32nd IEEE Vehicular Technology Conference; 1982 May 23–25; 1982.

## FURTHER READING

- Abidi AA. General relations between IP2, IP3, and offsets in differential circuits and the effects of feedback. IEEE Trans Microw Theory Tech 2003;51(5):1610–1612.

- Allen PE, Holberg DR. CMOS analog circuit design. 2nd ed. Oxford University Press, Inc; 2002.
- Bastani B, Bautista EE, Nagaraj G, et al. A quadrature down converter for direct conversion receivers with high 2nd and 3rd order intercept points. Proceedings of the 2000 3rd IEEE International Caracas Conference on Devices, Circuits and Systems, Cancun, Mexico; 2000 March 15–17; 2000. pp. C19–1–C19–4.
- Bautista EE, Bastani B, Heck J. A high IIP2 downconversion mixer using dynamic matching. *IEEE J Solid-State Circ* 2000;35(12):1934–1941.
- Bernard S. Digital communications. Englewood Cliffs (NJ): Prentice Hall; 1988.
- Bodway GE. Two port power flow analysis using generalized scattering parameters. *Microw J* 1967;10(6):61–69.
- Carson RS. High-frequency amplifiers. New York: John Wiley & Sons, Inc; 1975.
- Chu TS. Intermodulation in CDMA. Volume 2, 5th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Wireless Networks-Catching the Mobile Future, Holmdel, NJ; September 18–23; 1994. pp. 595–600.
- Collins CE, Dildine R, Pollard RD, et al.. A new method for determination of single sideband noise figure. Volume 3, IEEE MTT-S International Microwave Symposium Digest; 1994 May 23–27; 1994. pp. 1357–1360.
- Fong KL, Meyer RG. High-frequency nonlinearity analysis of common-emitter and differential-pair transconductance stages. *IEEE J Solid-State Circ* 1998;33(4):548–555.
- Fulton FF. Two-tone nonlinearity testing - the intercept point p/sub i/. *G-MTT Int Microw Symp Digest* 1973;73(1):112–112.
- Gagliardi RM. Satellite communications. Belmont (CA): Lifetime Learning Publications; 1984.
- Geens A, Rolain Y. Noise figure measurements on nonlinear devices. *IEEE Trans Instrum Meas* 2001;50(4):971–975.
- Gonzalez G. Microwave transistor amplifiers, analysis and design. Englewood Cliffs (NJ): Prentice-Hall, Inc; Englewood Cliffs, NJ, 1984.
- Ignea A, Kortvelyessy R, Stanescu O. A new approach on nonlinearity distortion distortion. Volume 1, 5th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Service, TELSIKS 2001, Nis Yugoslavia; September 19–21; 2001. pp. 301–304.
- Iversen CR, Kolding TE. Noise and intercept point calculation for modern radio receiver planning. *IEE Proc Commun* 2001;148(4):255–259.
- Kang S, Kim B. Second order nonlinearity analysis of Gilbert mixer. Volume 1, IEEE MTT-S International Microwave Symposium Digest; 2003 June 8–13; 2003a. pp. A45–A48.
- Kang S, Kim B. Second order nonlinearity analysis of Gilbert mixer. *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*; 2003 June 8–10; 2003b. pp. 559–562.
- Kang SM (Steve), Leblebici Y. CMOS digital integrated circuits. New York, NY: WCB/McGraw-Hill Companies, Inc; 1999.
- Kivekas K, Parssinen A, Halonen KAI. Characterization of IIP2 and DC-offsets in transconductance mixers. *IEEE Trans Circ Syst Analog Digital Signal Process (See also IEEE Trans Circuits Syst Express Briefs)* 2001;48(11):1028–1038.
- Kurokawa K. Power waves and the scattering matrix. *IEEE Trans Microw Theory Tech* 1965; 13.
- Larson LE. Modern digital and analog communication systems. 2nd ed. Holt, Rinehart and Winston, Inc; 1989.
- Lathi BP. The design of CMOS radio-frequency integrated circuits. Cambridge University Press; 1983.
- Lin M, Li Y, Chen H. A novel IP3 boosting technique using feedforward distortion distortion cancellation method for 5GHz CMOS LNA. Volume 1, IEEE MTT-S International, Microwave Symposium Digest; 2003 June 8–13; 2003. pp. A185–A188.
- Manstretta D, Brandolini M, Svelto F. Second-order intermodulation mechanisms in CMOS down-converters. *IEEE J Solid-State Circ* 2003;38(3):394–406.

- Meer DE. Noise figures [two-port linear transducers]. Volume 1, Proceedings of the IEEE National Aerospace and Electronics Conference; NAECON 1989, Dayton, OH; 1989 May 22–26; 1989. pp. 269–276.
- Millman J, Halkias CC. Integrated electronics: analog and digital circuits and systems. McGraw-Hill Book Company; New York, NY; 1972.
- Petrov AR. System approach for low 1/f noise, high IP2 dynamic range CMOS mixer design. Proceedings of the 15th Biennial, University/Government/Industry Microelectronics Symposium; 2003 June 30–July 2; 2003. pp. 74–77.
- Sagers RC. Intercept point and undesired responses. *IEEE Trans Veh Technol* 1983;VT-32(1):121–133.
- Smith J. Modern communication circuits. New York: McGraw-Hill Publishing Company; 1986.
- Taub H, Schilling D. Digital integrated electronics. McGraw-Hill, Inc; New York, NY; 1977.
- Vendelin GD, Pavio AM, Rohde UL. Microwave circuit design using linear and nonlinear techniques. New York: John Wiley & Sons, Inc; 1990.
- Wohlmuth HD, Simburger W. A high-IP3 RF receiver chip set for mobile radio base stations up to 2GHz. *IEEE J Solid-State Circ* 2001;36(7):1132–1137.
- Woods D. Reappraisal of the unconditional stability criteria for active 2-port networks in terms of S parameters. *IEEE Trans Circ Syst* 1976;23(2).
- Ziemer RE, Tranter WH. Principles of communications. New York: Houghton Mifflin Company; 1976.

## EXERCISES

1. List the main parameters of an RF block or a system.
2. What is reflection power gain?
3. What is transducer power gain, operating power gain, and available power gain?
4. In which cases is  $G_T = |S_{21}|^2$ ?
5. An RF system consists of two RF blocks in series. Write down the cascaded equation of power gain.
6. What is noise figure of a noisy RF block?
7. An RF system consists of block 1 and 2 in series; write down the cascaded equation of noise figure.
8. An RF system consists of block 1 and 2 in series with performance are shown as following table:

	Block 1	Block 2	System
NF, dB	LNA 2.5	Mixer 9.0	?
Gain, dB	12.5	4.5	?

Calculate the corresponding parameters for the system.

9. Highlight Haus' theory on the noise of a noisy two-port block.
10. Prove that the sensitivity of a receiver is

$$S_i = \frac{E_i^2}{4R_i} = NF \cdot kT \Delta f$$

11. Describe the test setup and procedures for a receiver's sensibility by 12-dB SINAD.
12. Describe the test setup and procedures for a receiver's sensibility by 20-dB quieting.
13. An RF system consists of two RF blocks in series. Write down the cascaded equation of IIP<sub>3</sub>.
14. A RF system consists of block 1 and 2, with performance are shown in the following table:

	Block 1	Block 2	System
	LNA	Mixer	
IIP <sub>3</sub> dB <sub>m</sub>	4	5	?
Gain, dB	12	-3	?

Calculate the corresponding parameters for the system.

15. When a transceiver is going to be registered officially, the FCC examines the nonlinearity of the receiver by means of IP<sub>3</sub> testing. Why is there no testing for IP<sub>2</sub>, IP<sub>4</sub>, or other intercept points?
16. In terms of simple geometrical derivation, derive the relationship between the input or output intercept point, IIP<sub>m</sub> or OIP<sub>m</sub> and the input power P<sub>i</sub>.
17. An RF system consists of two RF blocks in series. Write down the cascaded equation of IP<sub>2</sub>.
18. Does noise of a block or a system contribute to the distortion of signal?
19. The spurious product comes from the nonlinearity of the devices. Does spurious product impact on the sensitivity of a receiver?
20. A person has a handset. The current consumptions of his handset is

$$i_{\text{standby}} = 5 \text{ mA}; i_{\text{receive}} = 15 \text{ mA}; i_{\text{transmit}} = 50 \text{ mA}$$

and the operating time in different mode is

$$T_{\text{standby}} = 16 \text{ h}; T_{\text{receive}} = 4 \text{ h}; T_{\text{transmit}} = 1 \text{ h}.$$

What should be the milliampere hour (mAh) of the battery selected?

## ANSWERS

1. The main parameters of a RF block or a system are
  - power gain
  - DC power supply
  - noise
  - current drain
  - nonlinearity
  - part count

- stability
  - special parameters.
2. Reflection power gain is one in which readings are taken only at the output node:

$$G_{\text{ref}} = \frac{P_{R_L| \text{with reflection}}}{P_{R_L| \text{without reflection}}}$$

3. • Transducer power gain

$$G_T = \frac{P_L}{P_{\text{AVS}}} = \frac{1 - |\Gamma_S|^2}{|1 - \Gamma_{\text{in}}\Gamma_S|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2}$$

$$G_T = \frac{P_L}{P_{\text{AVS}}} = G_{\text{ref,in}} \cdot G_{\text{block}} \cdot G_{\text{ref,out}}$$

- Operating power gain

$$G_P = \frac{P_L}{P_{\text{IB}}} = G_T|_{\Gamma_S^* = \Gamma_{\text{in}}} = \Gamma_{\text{in}} = \frac{1}{1 - |\Gamma_{\text{in}}|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2}$$

$$P_{\text{IB}} = P_{\text{AVS}}|_{\Gamma_S^* = \Gamma_{\text{in}}} \quad \Gamma_S^* = \Gamma_{\text{in}}$$

- Available power gain

$$G_A = \frac{P_{\text{AVB}}}{P_{\text{AVS}}} = G_T|_{\Gamma_{\text{out}} = \Gamma_L^*} = \frac{1 - |\Gamma_S|^2}{|1 - S_{11}\Gamma_S|^2} |S_{21}|^2 \frac{1}{1 - |\Gamma_{\text{out}}|^2}$$

$$P_{\text{AVB}} = P_L|_{\Gamma_{\text{out}} = \Gamma_L^*} \quad \Gamma_{\text{out}} = \Gamma_L^*$$

4. In unilateral and impedance-matched cases, that is,  
if

$$S_{12} = 0, \quad \text{and} \quad \Gamma_S = \Gamma_L = 0,$$

then

$$G_T = |S_{21}|^2.$$

5. For an RF system consisting of two RF blocks in series, the cascaded equation of power gain is

$$G_{\text{SYS}} = G_1 \cdot G_2(\text{W}) \quad \text{or} \quad G_{\text{SYS}} = G_1 + G_2(\text{dB})$$

6. Noise figure of an RF circuit block represents the additional noise existing in the RF circuit block.

$$\text{NF} = \frac{\frac{S_i}{N_i}}{\frac{S_o}{N_o}}$$

or

$$NF = \frac{N_o}{GN_i} = \frac{N_o}{N_{o,i}}$$

or

$$NF = \frac{\text{Total output noise power}}{\text{Output noise power due to input noise}}$$

7. For an RF system consisting of block 1 and 2 in series, the cascaded equation of noise figure is

$$NF_{\text{SYS}} = NF_1 + \frac{NF_2 - 1}{G_1}$$

8. For an RF system consisting of block 1 and 2 in series, the performance parameters are shown in following table (columns 1 and 2):

	Block 1	Block 2	System
	LNA	Mixer	
NF. dB	2.5	9.0	3.36
G. dB	12.5	4.5	17.0

The corresponding parameters for the system have been calculated and are filled in column 3 of the table.

9. For a two-port noisy block,

$$NF = NF_{\min} + \frac{R_n}{G_s} [(G_s - G_{s,\text{opt}})^2 + (B_s - B_{s,\text{opt}})^2]$$

where

$R_n$  = the equivalent noise resistance of the RF block,  
 $Y_S$  = the Admittance of the RF block,

$$Y_S = G_S + jB_S$$

$$Y_{S,\text{opt}} = G_{S,\text{opt}} + jB_{S,\text{opt}}$$

if

$$Y_S = Y_{S,\text{opt}}$$

then

$$NF = NF_{\min}$$

10. The equation of receiver's sensitivity  $S_i$  can be derived as follows:

$$\sqrt{e^2_{\text{nt}}} = 4kTR_i \Delta f \quad N_i = \frac{\sqrt{e^2_{\text{nt}}}}{4R_i} = kT \Delta f$$

$$N_{\text{eq}} = \frac{N_o}{G} = \text{NF} \cdot N_i = \text{NF}kT \Delta f$$

Define

$$\text{SINAD} = \frac{S_o}{N_o + D_o} \quad \frac{S_i}{N_{\text{eq}}} = \text{RISE} - 1$$

$$\text{RISE} = \frac{S_i + N_{\text{eq}}}{N_{\text{eq}}} = \frac{S_i}{N_{\text{eq}}} + 1$$

$$E_i = 2\sqrt{S_i R_i} = 2\sqrt{N_{\text{eq}}(\text{RISE} - 1)R_i} = 2\sqrt{\text{NF}(\text{RISE} - 1)R_i kT \Delta f}$$

$$S_i = \frac{E_i^2}{4R_i} = \text{NF}(\text{RISE} - 1)kT \Delta f = \text{NF} \cdot kT \Delta f$$

when

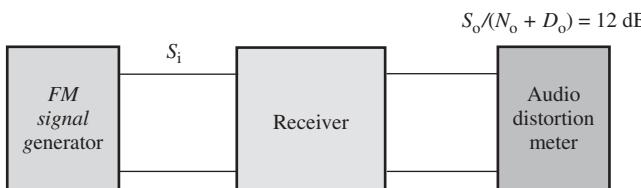
$$\text{RISE} = 2 \quad S_i = N_{\text{eq}}$$

$$E_i = 2\sqrt{\text{NF} \cdot R_i kT \Delta f}$$

$$S_i = \frac{E_i^2}{4R_i} = \text{NF} \cdot kT \Delta f$$

11. The 12-dB SINAD test setup for a receiver's sensibility is shown in Figure 12.P.1. Its test procedures are very simple, such as

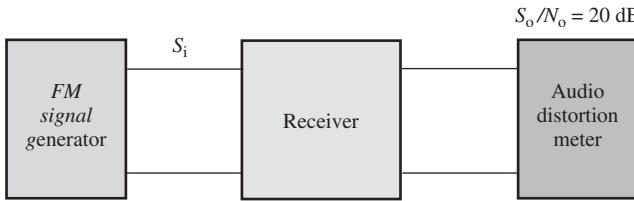
- Set  $S_i$  = Carrier  $f_o$  modulated with  $f_D = 1$  kHz audio signal.
- Gradually increase power  $S_i$  from 0 up to the point at which  $\text{SINAD} = 12$  dB.
- Write down the sensitivity  $E_i$  or  $S_i$  from the generator output.



**Figure 12.P.1.** 12-dB SINAD test setup for a receiver.

12. The test setup for a receiver's sensibility by 20-dB quieting is shown in Figure 12.P.2. Its test procedures are simple, such as

- Set carrier  $S_i$  at the frequency 1 kHz away from the desired RF carrier frequency.
- Gradually increase the power  $S_i$  from 0 up to the point at which the FM noise  $N_o$  has quietened down 20 dB from  $S_o$ .
- Write down the sensitivity  $E_i$  or  $S_i$  from the generator output.



**Figure 12.P.2.** 20-dB quieting test setup for a receiver.

13. For an RF system consisting of two RF blocks in series, the cascaded equation of  $\text{IIP}_3$  is as follows:

$$\frac{1}{\text{IIP}_{3,\text{sys}}} = \frac{1}{\text{IIP}_{3,1}} + \frac{G_1}{\text{IIP}_{3,2}}$$

14. For an RF system consisting of blocks 1 and 2, the performance parameters are shown as following table (columns 1 and 2):

	Block 1	Block 2	System
IIP <sub>3</sub> , dB <sub>m</sub>	LNA 4	Mixer 5	-7.33
Gain, dB	12	-3	9

The corresponding parameters for the system have been calculated and filled in the column 3 of the Table.

15. Instead of IP<sub>2</sub>, IP<sub>4</sub>, or other intercept points, the third-order intercept point, IP<sub>3</sub>, is chosen by the FCC as a criterion to judge the nonlinearity of a device or a system. This is due to the special feature of the IP<sub>3</sub> testing, that is, any value of  $\delta f$  in the IP<sub>3</sub> testing is possible to produce the third-order intermodulation product. If FCC examines the nonlinearity of a block or a system by means of IP<sub>3</sub> testing with a low value of  $\delta f$  within the channel to filter the IP<sub>3</sub> product out is almost impossible because a filter design with a few kilohertz of bandwidth in the RF frequency range is almost impossible. This justifies examining the nonlinearity of a block or a system because it is not possible to filter the IP<sub>3</sub> product out by the filtering technology.

16. As shown in Figure 12.21, the simple geometrical relations are

$$d + \Delta = md,$$

$$(m - 1)d = \Delta,$$

$$d = \text{IIP}_m - P_i,$$

then we have

$$\text{IIP}_m = \Delta / (m - 1) + P_i$$

$$\text{OIP}_m = G + \text{IIP}_m$$

17.

$$\left( \frac{1}{\text{IIP}_{2,\text{sys}}} \right)^{\frac{1}{2}} = \left( \frac{1}{\text{IIP}_{2,1}} \right)^{\frac{1}{2}} + \left( \frac{G_k}{\text{IIP}_{2,2}} \right)^{\frac{1}{2}}$$

18. The noise of a block or a system relates to its receiving sensitivity but does not contribute to the distortion of signal, which is mainly determined by its nonlinearity.
19. The spurious product comes from the nonlinearity of the devices and causes distortion but does not impact on the sensitivity of a receiver.
20. The selected battery must have an electrical capacity of 190 mAh at least, since

$$i_{\text{avg}} T_{\text{tot}} = i_{\text{standby}} T_{\text{standby}} + i_{\text{receive}} T_{\text{receive}} + i_{\text{transmit}} T_{\text{transmit}}$$

$$i_{\text{avg}} T_{\text{tot}} = 5 \times 16 + 15 \times 4 + 50 \times 1 = 190 \text{ mAh}$$



# SPECIALITY OF "ZERO IF" SYSTEM

## 13.1 WHY DIFFERENTIAL PAIR?

### 13.1.1 Superficial Difference between Single-Ended and Differential Pair

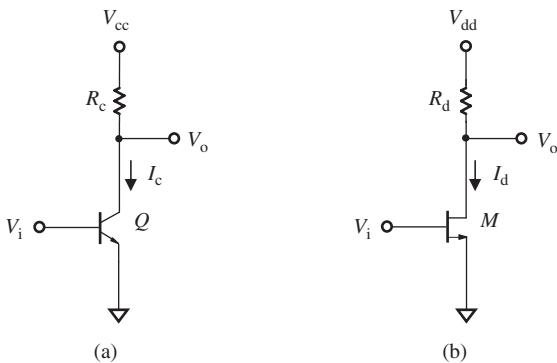
In the early years of electronic circuit development, only single-ended stages were studied and implemented. Gradually, it was found that some special features of performance could be achieved only by a differential pair instead of a single-ended stage or a double-balanced mixer. At present, both single-ended and differential pairs have been developed for most RF blocks, such as the LNA (low-noise amplifier), mixer, VCO (voltage-controlled oscillator), filter, and PA (power amplifier). In the recent years, a number of differential circuits have appeared in RF or RFIC designs, especially in "zero IF" or direct conversion communication systems.

Without the input and output impedance matching networks, a single-ended stage is shown in Figure 13.1, and a differential pair is shown in Figure 13.2.

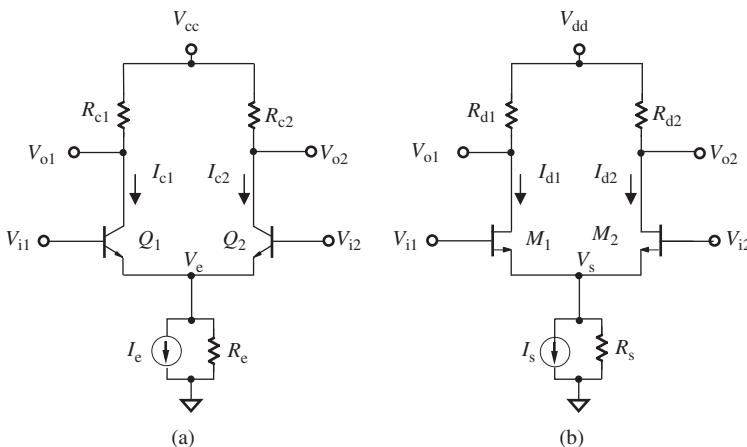
The differential pair appears as two single-ended blocks joined together with a common emitter or source resistor,  $R_e$  or  $R_s$ , in a bipolar or a MOSFET differential pair. There is a current source,  $I_e$  or  $I_s$ , connected to this common resistor in parallel. The pair of devices should have the same size and configuration, as well as identical collector or drain resistors, that is, for the bipolar differential pair,

$$R_{c1} = R_{c2}, \quad (13.1)$$

$$Q_1 = Q_2, \quad (13.2)$$



**Figure 13.1.** Typical single-ended block. (a) Bipolar single-ended stage. (b) MOSFET single-ended block.



**Figure 13.2.** Typical differential pair. (a) Bipolar differential pair. (b) MOSFET differential pair.

and for the MOSFET differential pair,

$$R_{d1} = R_{d2}, \quad (13.3)$$

$$M_1 = M_2. \quad (13.4)$$

In short, both differential branches should be symmetrical to each other, including their values, size, configuration, and layout.

The apparent differences are shown in Figures 13.1 and 13.2. The number of parts in a differential pair is nearly double that of a single-ended stage. For instance, as shown in Figure 13.1, the part count of each branch in a differential pair is equal to 2, which is the part count of a single-ended block. Actually, two more parts exist in the differential pair, the current sources \$I\_e\$ or \$I\_s\$ and the tail resistors \$R\_e\$ or \$R\_s\$.

Consequently, either in an RF circuit with discrete parts or in an RFIC design, the layout area for the differential pair is about double that of the single-ended stage. The cost in an RFIC design is proportional to the layout area on the IC chip; this means that the price for a differential pair is nearly double that of a single-ended stage. In practice, the part count and layout area would actually be more than double, as one of the parts in the differential pair is a current source, which is usually not a single part and is not present in a single-ended stage.

If a differential pair is built by two single-ended stages, its current drain is usually double that of each single-ended stage if each differential output has approximately the same voltage gain as that in the output of the single-ended stage.

In general, the noise figure of a differential pair is higher than that of a single-ended stage. One of the reasons for this is that, if the current flow through the device remains the same, shot noise is double that of a single stage. Another reason is that thermal noise is also approximately double that of a single-ended stage because there are twice as many resistors. At this point, different opinions do exist and some people disagree with the assertion that the noise figure is increased in the differential pair. On the contrary, they believe that “common mode” noise is eliminated or reduced substantially in a differential pair compared to a single-ended stage.

A differential input signal between two input ports is expected to be faithfully magnified or transported by a differential pair and presented to its differential output signal between two output ports. Assuming that the input of the differential pair is a pure differential signal, the output of the differential pair should thus be a pure differential signal as well. In order to approach this ideal differential performance, the symmetry of the layout must be maintained as much as possible. On the contrary, symmetry is meaningless for a single-ended stage.

Finally, a special part, called the *balun*, is needed for the differential pair. A balun either splits a single-ended signal to a pair of differential signals or combines a pair of differential signals into a single-ended signal. Chapter 15 introduces and discusses the balun.

From the comparison above, it is apparent that the differential pair is devoid of any merit compared to a single-ended stage. If so, what is use of the differential pair? The answer can be found in Sections 13.1.2 and 13.1.3.

Many negative comments have been leveled at the differential pair when comparing it with the single-ended stage. The “apparent differences between the single-ended stage and the differential pair” have been deduced solely from their schematics. Their essential differences, however, are discovered from their performance and other aspects. As a matter of fact, the remarkable difference between the single-ended stage and the differential pair is their performance in nonlinearity. Let us find out this difference.

### 13.1.2 Nonlinearity in Single-Ended Stage

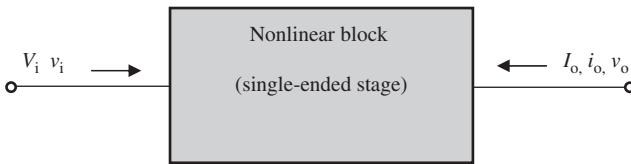
In reality, all transistors, bipolar or MOSFET, are nonlinear parts. Any circuit block containing any nonlinear part is called a *nonlinear block*. Therefore, a single-ended stage is a nonlinear block as shown in Figure 13.3. The transfer function of a single-ended stage can be expressed in a general form, that is,

$$I_o = I_{DC} + i_o = a_o + a_1 V_i + a_2 V_i^2 + a_3 V_i^3 + a_4 V_i^4 + \dots, \quad (13.5)$$

where

- $I_o$  = the resultant output current,
- $I_{DC}$  = the output DC current,
- $i_o$  = the output AC current,
- $V_i$  = the input voltage, and
- $a_i$  = the  $i$ th order of transconductance coefficient.

In general, the nonlinearity of the single-ended stage includes both odd and even orders from 0, 1, 2, 3, to infinity.



**Figure 13.3.** Nonlinear block containing a single-ended stage.

If the input signal is an AC sinusoidal voltage, that is,

$$V_i = v_i = v_{io} \cos \omega t, \quad (13.6)$$

where

- $v_i$  = the input AC sinusoidal signal,
- $v_{io}$  = the amplitude of the input signal, and
- $\omega$  = the angular frequency of the sinusoidal signal.

Then the transfer function (13.5) becomes

$$\begin{aligned} I_o &= a_o + a_1 v_{io} \cos \omega t + a_2 v_{io}^2 \cos^2 \omega t + a_3 v_{io}^3 \cos^3 \omega t + a_4 v_{io}^4 \cos^4 \omega t + \dots \\ &= a_o + a_1 v_{io} \cos \omega t + a_2 v_{io}^2 [1/2 + (1/2) \cos 2\omega t] + a_3 v_{io}^3 [(3/4) \cos \omega t \\ &\quad + (1/4) \cos 3\omega t] + a_4 v_{io}^4 [3/8 + (4/8) \cos 2\omega t + (1/8) \cos 4\omega t] + \dots \\ &= [a_o + (1/2)a_2 v_{io}^2 + (3/8)a_4 v_{io}^4 + \dots] + [a_1 v_{io} + (3/4)a_3 v_{io}^3 + \dots] \cos \omega t \\ &\quad + [(1/2)a_2 v_{io}^2 + (4/8)a_4 v_{io}^4 + \dots] \cos 2\omega t + [(1/4)a_3 v_{io}^3 + \dots] \cos 3\omega t \\ &\quad + [(1/8)a_4 v_{io}^4 + \dots] \cos 4\omega t + \dots \end{aligned} \quad (13.7)$$

It can be seen that, when the input is a sinusoidal signal, the output current of a single-ended stage contains both DC and AC currents, and can be rewritten as

$$I_o = I_{DC} + i_o, \quad (13.8)$$

$$I_{DC} = a_o + v_{io}^2 [(1/2)a_2 + (3/8)a_4 v_{io}^2 + \dots], \quad (13.9)$$

$$\begin{aligned} i_o &= [a_1 v_{io} + (3/4)a_3 v_{io}^3 + \dots] \cos \omega t + [(1/2)a_2 v_{io}^2 + (4/8)a_4 v_{io}^4 + \dots] \cos 2\omega t \\ &\quad + [(1/4)a_3 v_{io}^3 + \dots] \cos 3\omega t + [(1/8)a_4 v_{io}^4 + \dots] \cos 4\omega t + \dots \end{aligned} \quad (13.10)$$

From expression (13.9), it can be seen that the output DC current consists of two parts. The first part is the first term  $a_o$ , which is irrelevant to  $v_i$ . The second part is the second term containing the factor  $v_{io}^2$ , which is dependent on the input AC sinusoidal voltage  $v_i$  and is called the *DC offset*, that is,

$$\text{DC offset} = v_{io}^2 [(1/2)a_2 + (3/8)a_4 v_{io}^2 + \dots]. \quad (13.11)$$

Obviously, the DC offset relies directly on the input AC signal, so it is unstable and always fluctuates since the received signal varies from time to time and from place to place. A remarkable feature of the DC offset is that it is contributed by only the even orders but not the odd orders of its nonlinearity.

From expression (13.10), it can be seen that the output AC current consists of all the harmonics. The odd-order harmonics come from the odd-order nonlinearity terms, and the even-order harmonics come from the even-order nonlinearity terms.

### 13.1.3 Nonlinearity in a Differential Pair

The derivation of the nonlinearity of a nonlinear block containing a differential pair is similar to that for a nonlinear block containing a single-ended stage. Bear in mind the following relation:

$$v_{i1} = -v_{i2}. \quad (13.12)$$

If

$$v_{i1} = v_{io} \cos \omega t, \quad (13.13)$$

then,

$$v_{i2} = -v_{io} \cos \omega t, \quad (13.14)$$

where

$v_{i1}$  = the input AC sinusoidal signal at differential branch 1,

$v_{i2}$  = the input AC sinusoidal signal at differential branch 2, and

$v_{io}$  = the amplitude of each differential input signal.

Referring to Figure 13.4, the transfer function of each differential branch can be expressed in a general form, that is,

$$I_{o1} = I_{DC1} + i_{o1} = a_{o1} + a_{11}V_{i1} + a_{21}V_{i1}^2 + a_{31}V_{i1}^3 + a_{41}V_{i1}^4 + \dots, \quad (13.15)$$

$$I_{o2} = I_{DC2} + i_{o2} = a_{o2} + a_{12}V_{i2} + a_{22}V_{i2}^2 + a_{32}V_{i2}^3 + a_{42}V_{i2}^4 + \dots, \quad (13.16)$$

where the second subscript denotes the branch number of the differential pair. And, if it is an ideal differential pair, we have

$$a_{i1} = a_{i2} = a_i. \quad (13.17)$$

If the input signal is an AC differential sinusoidal voltage, that is,

$$V_{i1} = v_{i1} = v_{io} \cos \omega t, \quad (13.18)$$

$$V_{i2} = v_{i2} = -v_{io} \cos \omega t, \quad (13.19)$$

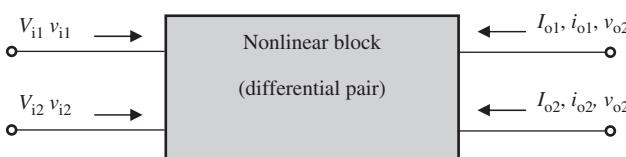


Figure 13.4. Nonlinear block containing a differential pair.

then expressions (13.15) and (13.16) become

$$\begin{aligned}
 I_{o1} &= a_o + a_1 v_{io} \cos \omega t + a_2 v_{io}^2 \cos^2 \omega t + a_3 v_{io}^3 \cos^3 \omega t + a_4 v_{io}^4 \cos^4 \omega t + \dots \\
 &= a_o + a_1 v_{io} \cos \omega t + a_2 v_{io}^2 [1/2 + (1/2) \cos 2\omega t] \\
 &\quad + a_3 v_{io}^3 [(3/4) \cos \omega t + (1/4) \cos 3\omega t] \\
 &\quad + a_4 v_{io}^4 [3/8 + (4/8) \cos 2\omega t + (1/8) \cos 4\omega t] + \dots \\
 &= [a_o + (1/2) a_2 v_{io}^2 + (3/8) a_4 v_{io}^4 + \dots] + [a_1 v_{io} + (3/4) a_3 v_{io}^3 + \dots] \cos \omega t \\
 &\quad + [(1/2) a_2 v_{io}^2 + (4/8) a_4 v_{io}^4 + \dots] \cos 2\omega t + [(1/4) a_3 v_{io}^3 + \dots] \cos 3\omega t \\
 &\quad + [(1/8) a_4 v_{io}^4 + \dots] \cos 4\omega t + \dots. \tag{13.20}
 \end{aligned}$$

$$\begin{aligned}
 I_{o2} &= a_o - a_1 v_{io} \cos \omega t + a_2 v_{io}^2 \cos^2 \omega t - a_3 v_{io}^3 \cos^3 \omega t + a_4 v_{io}^4 \cos^4 \omega t + \dots \\
 &= a_o - a_1 v_{io} \cos \omega t + a_2 v_{io}^2 [1/2 + (1/2) \cos 2\omega t] \\
 &\quad - a_3 v_{io}^3 [(3/4) \cos \omega t + (1/4) \cos 3\omega t] \\
 &\quad + a_4 v_{io}^4 [3/8 + (4/8) \cos 2\omega t + (1/8) \cos 4\omega t] + \dots \\
 &= [a_o + (1/2) a_2 v_{io}^2 + (3/8) a_4 v_{io}^4 + \dots] - [a_1 v_{io} + (3/4) a_3 v_{io}^3 + \dots] \cos \omega t \\
 &\quad + [(1/2) a_2 v_{io}^2 + (4/8) a_4 v_{io}^4 + \dots] \cos 2\omega t \\
 &\quad - [(1/4) a_3 v_{io}^3 + \dots] \cos 3\omega t + [(1/8) a_4 v_{io}^4 + \dots] \cos 4\omega t - \dots. \tag{13.21}
 \end{aligned}$$

The transfer function of the differential pair is

$$\begin{aligned}
 \Delta I &= I_{o1} - I_{o2} = i_o \\
 &= 2[a_1 v_{io} + (3/4) a_3 v_{io}^3 + \dots] \cos \omega t + 2[(1/4) a_3 v_{io}^3 + \dots] \cos 3\omega t + \dots. \tag{13.22}
 \end{aligned}$$

The output differential current contains only the input differential signal and its odd-order of harmonics, and thus the DC offset is obviously zero, that is,

$$\text{DC offset} = 0. \tag{13.23}$$

Summarizing, in an ideal differential pair,

- the even orders of nonlinearity are cancelled by each other;
- DC offset does not exist.

This is the answer for the question: what is the differential pair for?

If an RF block is built with an ideal differential configuration and if the input is perfectly differential, the DC offset and the even orders of nonlinearity disappear.

The next question is, of course: why do we pursue the goal of zero DC offset and cancel the even orders of nonlinearity. For the answer, let us move to the next section.

### 13.1.4 Importance of Differential Configuration in a Direct Conversion or Zero IF Communication System

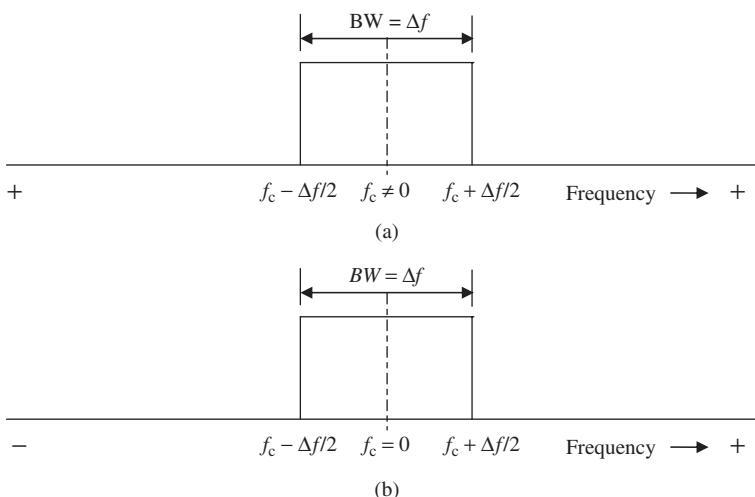
In a dual conversion communication system, the spectrum of the demodulated signals is allocated around the central frequency  $f_c$  as shown in Figure 13.5(a). The central frequency is not zero and its spectrum is spread over the bandwidth around the central frequency. The problem of DC offset does not exist because the spectrum of the demodulated signals does not contain the DC component.

In a direct conversion communication system, also called the *zero IF* communication system, the spectrum of the demodulated signals is allocated around the frequency  $f = 0$  as shown in Figure 13.5(b). The central frequency is zero and its spectrum is spread across the zero frequency point, which covers both negative and positive frequencies.

This implies that the DC or zero frequency is the main component of the signal's spectrum. Any conceivable interference at the DC or zero frequency might disturb the received signal and even put the communication system out of work completely.

The interference on the DC component in the spectrum can come from either outside or inside the circuit block. The interference from the outside depends on the environment, circuit configuration, and system design, which are outside the scope of this book. Internal interference from the circuit block is also a complicated product in which many factors are involved. The main issue, however, is that the DC offset is inevitably produced from the nonlinear devices of the circuitry. This is the main source of the DC offset problem in the zero IF communication system and becomes the key issue in the success or failure in the design of such a communication system.

Now, let us return to the discussion in Sections 13.1.2 and 13.1.3. In a circuit block with single-ended configuration, DC offset is inevitably produced by the nonlinearity of the device. However, in a circuit block with an ideal differential pair configuration, the produced DC offsets cancel each other. Furthermore, in a circuit block with an ideal differential pair configuration, all the even orders of nonlinearity cancel each other. The differential configuration of the circuit block is obviously a powerful means to remove or reduce the DC offset. This is why in a zero IF radio, most RF blocks, including the LNA, mixer, and modulator, are differential blocks.



**Figure 13.5.** Spectrum of demodulated signal in a communication system. (a) In a dual conversion communication system. (b) In a direct conversion or "zero IF" communication system.

However, it must be pointed out that this outstanding behavior of a differential pair is so far only a theoretical prediction. In order to realize such an ideal objective, the engineering of the circuit design must be strict and accurate. There are four essential tasks:

- To guarantee the symmetry of the circuit layout. Although perfect symmetry is impossible to achieve, it must be as close as possible;
- To strictly control the accuracy or tolerance of parts in either the discrete circuit or IC chip design;
- To avoid any interference;
- To have good RF grounding.

### 13.1.5 Why Direct Conversion or Zero IF?

Recalling all the disadvantages of the differential pair mentioned in Section 13.1, the advantage of building differential circuits for the zero IF communication system seems doubtful, unless the zero IF system confers some huge benefit.

From the engineering design viewpoint, designing a dual conversion communication system is easier than designing a direct conversion or zero IF communication system. From the viewpoint of system performance, a dual conversion system is better than a direct conversion or zero IF communication system. At present, dual conversion systems are still applied in the military or most advanced communication systems.

In spite of these various viewpoints, for a popular product, such as a cellular phone or WLAN (wireless local area network), cost is the first priority. A direct conversion or zero IF communication system is much cheaper than a dual conversion communication system. The price reduction is twofold:

- All the RF filters applied in the dual conversion communication system can be removed in a direct conversion communication system.
- It is possible to put a full direct conversion communication system on one IC chip since all the RF filters are gone.

This is the final answer to the question "Why differential pair?"

## 13.2 CAN DC OFFSET BE BLOCKED OUT BY A CAPACITOR?

In order to clarify this question, let us examine how a capacitor blocks DC voltage from the source to the load. Assume that a series of pulses are delivered from the source to the load. A capacitor  $C$  is inserted between the source and the load. The waveform at the source  $v_S$  is drawn in the upper left corner of Figures 13.6–13.8. The pulses are positive, with an average DC voltage  $V_o$  and a repetition period  $T$ . The waveform at the load  $R_L$  is drawn in the upper right corner of Figures 13.6–13.8, and can be categorized into three cases.

The first case is when

$$T \gg R_L C. \quad (13.24)$$

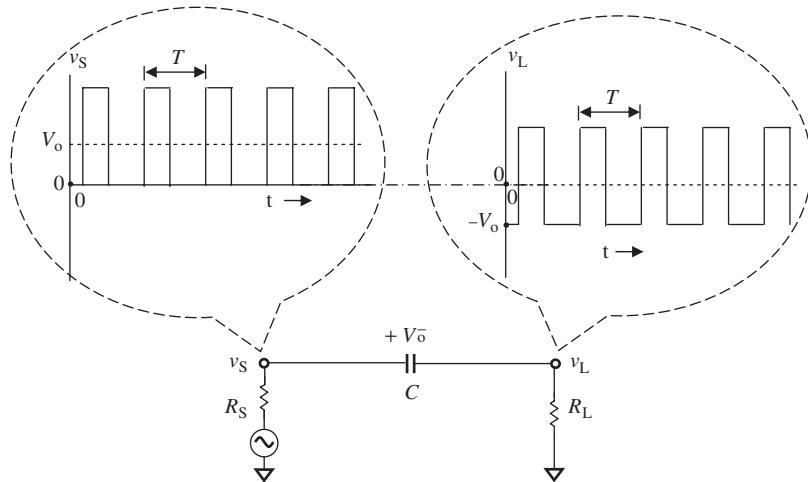


Figure 13.6. A capacitor blocks DC voltage  $V_o$  from the source  $v_S$  to  $v_L$  when  $T \gg R_L C$ .

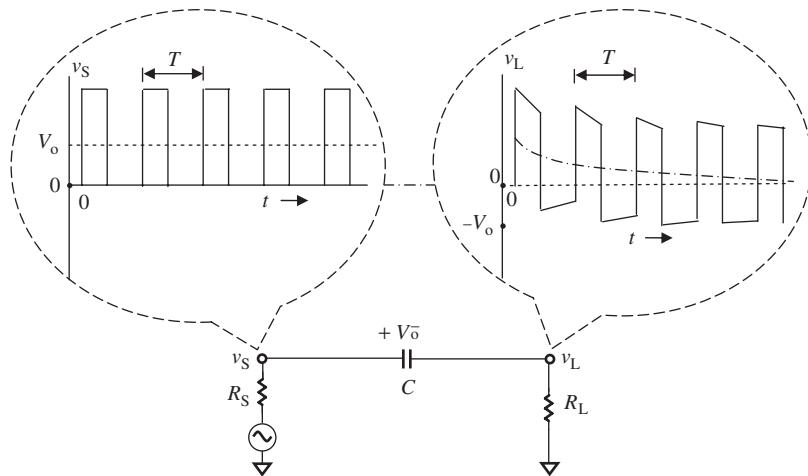


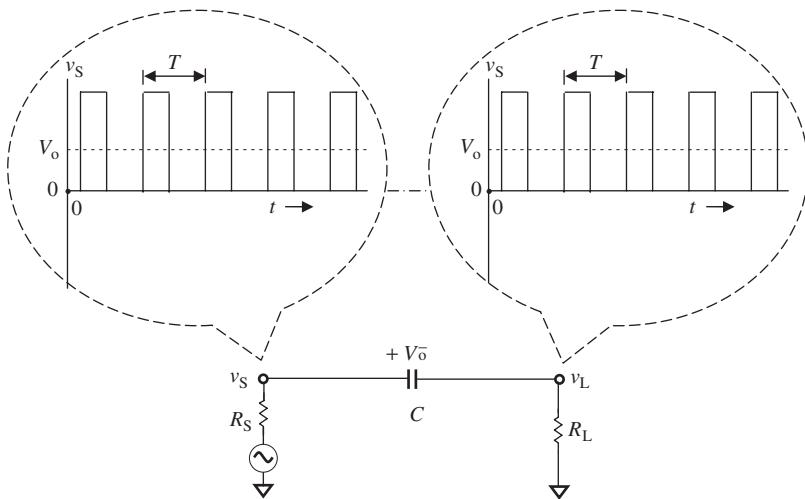
Figure 13.7. A capacitor blocking DC voltage  $V_o$  from the source  $v_S$  to  $v_L$  when  $T \sim R_L C$ .

Figure 13.6 shows the waveforms at the source  $R_S$  and at the load  $R_L$  when  $T \sim R_L C$ . In this case, the charge and discharge period of the capacitor is much shorter than the repetition period. This implies that the capacitor is able to follow the variation of the source voltage  $v_S$ . Consequently, the average DC voltage is built across the capacitor. The waveform at the load  $R_L$ , as shown in the upper right corner of Figure 13.6, is kept unchanged, but its average DC voltage has dropped from  $V_o$  to 0, the actual ground point. Therefore, in this case, the average DC voltage  $V_o$  existing in the source is blocked by the capacitor so that at the load the average DC voltage is 0.

The second case is when

$$T \sim R_L C. \quad (13.25)$$

Figure 13.6 shows the waveforms at the source  $R_S$  and at the load  $R_L$  when  $T \sim R_L C$ . Here, the charge and discharge period of the capacitor is around the same order as the



**Figure 13.8.** A capacitor blocking DC voltage  $V_o$  from the source  $v_s$  to  $v_L$  when  $T \ll R_L C$ .

repetition period. Hence, the capacitor is not able to follow the variation of the source voltage  $v_s$  well. Consequently, the average DC voltage built across the capacitor gradually increases from 0 to  $V_o$ . At the load  $R_L$ , the waveform changes and experiences the charged and discharged processes alternatively. The voltage across the capacitor, when the capacitor is charged, is

$$V = V_{oi}(1 - e^{-R_L C}), \quad (13.26)$$

where  $V_{oi}$  is the voltage to be charged to.

When the capacitor is discharged,

$$V = V_{oj} e^{-R_L C}, \quad (13.27)$$

where  $V_{oj}$  is the voltage to be discharged from.

Correspondingly, from the waveform at the load  $R_L$  as shown in the upper right corner of Figure 13.7, it can be seen that its average DC voltage is dropped from  $V_o$  to 0 with a time constant  $\tau = R_L C$ . In this case, the average DC voltage  $V_o$  existing in the source is blocked from the source to the load by the capacitor after a time of about  $\tau = R_L C$ .

The third case is when

$$T \ll R_L C. \quad (13.28)$$

Figure 13.8 shows the waveforms at the source  $R_S$  and at the load  $R_L$  when  $T \ll R_L C$ . In this case, the charge and discharge time of the capacitor is much longer than the repetition period. This implies that the capacitor is not able to follow the variation of the source voltage  $v_s$  well. The time for the average DC voltage across the capacitor from 0 to  $V_o$  is very long, so that the DC voltage across the capacitor is almost unchanged during the short repetition period  $T$ . At the load  $R_L$ , the waveform is as shown in the upper right corner. The average DC voltage  $V_o$  is still maintained at the same level as at the source.

From these three cases, it can be seen that a capacitor can function as a DC blocking capacitor only when condition (13.24) is satisfied. Let us convert this inequality (13.24) to an approximate equation, that is,

$$T = 10 R_L C_{\max}, \quad (13.29)$$

where  $C$  has been replaced by  $C_{\max}$ , which implies that the condition (13.24) is satisfied even in the worst case.

The repetition frequency of the pulses  $f$  corresponding to  $T$  in Figures 13.6–13.8 should be in the order of 100 MHz to 10 GHz; in terms of (13.29), the value of the DC blocking capacitor can be evaluated. In Table 13.1, the values of  $R_L$  are selected from 10 to 10,000  $\Omega$ . In actual circuits, this value is usually much higher than 50  $\Omega$  before impedance matching.  $R_L = 1\text{k}\Omega$  is taken as a reasonable typical value.

Table 13.1 lists the maximum values  $C_{\max}$  of a capacitor functioning as a DC blocking capacitor. Table 13.2 repeats the  $\text{SRF}_c$  (self-resonant frequency) values of chip capacitors from Table 7.A.1. The values of  $C_{\max}$  are much lower than those of the “zero” chip capacitor if the seventh column, when  $R_L = 1\text{k}\Omega$  in Table 13.1, is compared with Table 13.2. For instance,

- if the operating frequency is 800 MHz, it can be found from Table 13.1 that the capacitance of the DC blocking capacitor must be less than 0.125 pF if  $R_L = 1\text{k}\Omega$ , while from Table 13.2 it can be found that the capacitance of a zero chip capacitor is 46 pF;
- if the operating frequency is 2400 MHz, from Table 13.1 it can be found that the capacitance of DC blocking capacitor must be less than 0.042 pF if  $R_L = 1\text{k}\Omega$ , while from Table 13.2 it can be found that the capacitance of a zero chip capacitor is 5.1 pF.

It is well known that the signal is not attenuated if a zero capacitor is inserted at the input or output in series. However, it is impossible for a zero capacitor to function as a DC blocking capacitor because its value is too high. If the zero capacitor is replaced by a capacitor listed in Table 13.1, the low value of the capacitor would bring about a significant attenuation of the signal. More importantly, capacitors with values of less than 0.1 or 0.2 pF are meaningless in a practical circuit design, no matter whether the circuit is built by discrete parts or by an RFIC chip.

Consequently, one cannot expect the DC offset current to be blocked by a capacitor unless a special design scheme is developed. DC offset clearly exists in a single-ended stage, but disappears in an ideal differential pair. However, in reality, an ideal differential pair does not exist; DC offset persists in an imperfect differential pair.

The cancellation of DC offset thus has become a hot topic in the development of direct conversion or zero IF communication systems.

### 13.3 CHOPPING MIXER

DC offset cancellation must be executed in the receiver of a zero IF communication system. On the other hand, DC offset is mainly produced in the mixer block. Therefore, the so-called chopping mixer was developed for the DC offset cancellation.

Chopping technology was invented several decades ago during the development of the super-regeneration radio. By “chopping” the input and output signals of the mixer at

TABLE 13.1. Maximum Capacitance  $C_{\max}$  in pF as a DC Blocking Capacitor, pF

$f$ , MHz	$R_L = 10 \Omega$	$R_L = 30 \Omega$	$R_L = 50 \Omega$	$R_L = 100 \Omega$	$R_L = 500 \Omega$	$R_L f = 1 \text{ k}\Omega$	$R_L = 2 \text{ k}\Omega$	$R_L = 3 \text{ k}\Omega$	$R_L = 5 \text{ k}\Omega$	$R_L = 10 \text{ k}\Omega$
100	100.000	33.33	20.000	10.000	2.000	<b>1.000</b>	0.500	0.333	0.200	0.100
200	50.000	16.67	10.00	5.000	1.000	<b>0.500</b>	0.250	0.167	0.100	0.050
300	33.333	11.11	6.667	3.333	0.667	<b>0.333</b>	0.167	0.111	0.067	0.033
400	25.000	8.333	5.000	2.500	0.500	<b>0.250</b>	0.125	0.083	0.050	0.025
500	20.000	6.667	4.000	2.000	0.400	<b>0.200</b>	0.100	0.067	0.040	0.020
600	16.667	5.556	3.333	1.667	0.333	<b>0.167</b>	0.083	0.056	0.033	0.017
700	14.286	4.762	2.857	1.429	0.286	<b>0.143</b>	0.071	0.048	0.029	0.014
800	12.500	4.167	2.500	1.250	0.250	<b>0.125</b>	0.063	0.042	0.025	0.013
900	11.111	3.704	2.222	1.111	0.222	<b>0.111</b>	0.056	0.037	0.022	0.011
1000	10.000	3.333	2.000	1.000	0.200	<b>0.100</b>	0.050	0.033	0.020	0.010
1500	6.667	2.222	1.333	0.667	0.133	<b>0.067</b>	0.033	0.022	0.013	0.007
2400	4.167	1.389	0.833	0.417	0.083	<b>0.042</b>	0.021	0.014	0.008	0.004
5800	1.724	0.575	0.345	0.172	0.034	<b>0.017</b>	0.009	0.006	0.003	0.002
10,000	1.000	0.333	0.200	0.100	0.020	<b>0.010</b>	0.005	0.003	0.002	0.001

Note:  $R_L = 1 \text{ k}\Omega$  is taken as a typical value.

TABLE 13.2. Some  $SRF_c$  (Self-Resonant Frequency) Values of Chip Capacitors

$SRF_c$ , MHz	Value of Chip Capacitor, pF
40	18,000
50	11,664
100	2916
150	1296
450	144
500	117
800	46
900	36
1000	29
1500	13
2400	5.1
5400	1.0
6235	0.75

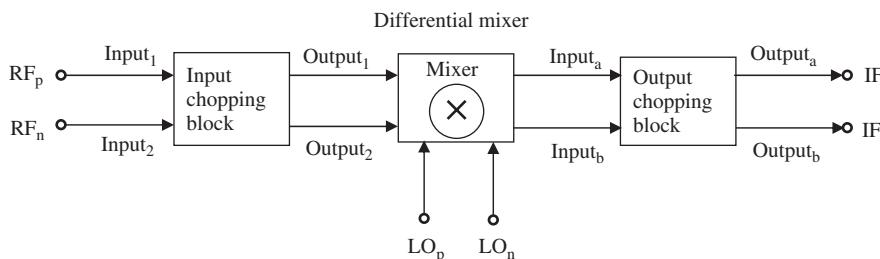


Figure 13.9. Block diagram of a “chopping” mixer.

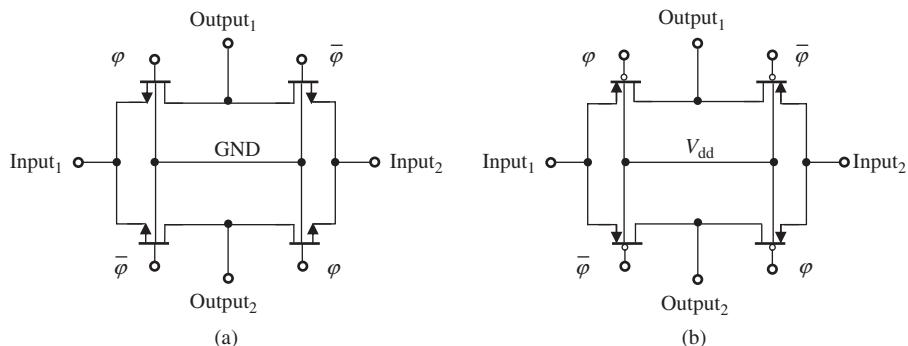
the RF input port and IF output port, respectively, the even orders of nonlinearity of the mixer or the DC offset can be significantly reduced.

Figure 13.9 shows the block diagram of a “chopping mixer.” It has three portions: an input chopping block, a differential mixer core, and an output chopping block. The core of the differential mixer is a mixer with a Gilbert cell.

The input chopping block is inserted between the RF input port and the input of the differential mixer. It chops the RF signal with a repetition frequency of  $f_{CH}$ . The output chopping block is inserted between the IF output of the differential mixer and the IF output ports. It chops the IF signal with the same repetition frequency  $f_{CH}$  as that of the input chopping block.

Figure 13.10(a) and (b) depicts the input and output chopping blocks, respectively. Both are simply cross-toggled switches. The input chopping block consists of four n-channel MOSFETs and its substrate is grounded, while the output chopping block consists of four p-channel MOSFETs and its substrate is connected to the DC power supply  $V_{dd}$ . These differences are due to the different positions of the chopping circuits. The input chopping circuit is working in the low DC voltage portion, while the output chopping circuit is working in the high DC voltage portion.

Two differential inputs are alternatively connected to two differential outputs through two gate controls  $\varphi$  and  $\bar{\varphi}$ , which are fed with pulse signals of an appropriate chopping



**Figure 13.10.** Input and output chopping circuit of a chopping mixer. (a) Input chopping circuit. (b) Output chopping circuit.

frequency  $f_{CH}$  and of the voltage amplitude with 0 and  $V_{dd}$  in an iterative manner. The phase of  $\varphi$  and  $\bar{\varphi}$  is kept at a  $180^\circ$  difference:  $\varphi$  is at 0 when  $\bar{\varphi}$  is at  $V_{dd}$ , and vice versa. The RF signal is therefore turned ON/OFF per the chopping frequency.

The chopping frequency is experimentally determined by simulation as well as benchwork. The criterion for determining this frequency is the maximization of  $IP_2$  or minimization of the DC offset for the entire chopping mixer. The value of  $IP_2$  is a good measure of the DC offset because it is the dominating term among the even-order nonlinearities which bring about the DC offset. Like the differential mixer itself, the layouts of the input and output chopping blocks must be kept symmetrical as much as possible.

Figure 13.11 shows the schematic of the entire chopping mixer, including the input and output chopping circuits. The Gilbert cell and the input and output chopping blocks are built by MOSFETs. The input RF signals are fed to the inputs of the input chopping circuit, and then the chopped RF signals from the outputs of the input chopping circuit are fed to the sources of the lower MOSFET pair. They are therefore called the *CG configuration* of the RF input. The special property of the CG configuration is that the RF input has a relatively wide bandwidth but low conversion gain compared to other configurations, such as the CS or CD configurations.

One possible way to connect the two RF input terminals  $RF_p$  and  $RF_n$  is to use an RF transformer. The RF input terminals  $RF_p$  and  $RF_n$ , as shown in Figure 13.11, are connected to the second winding of the transformer with its center tapped to the ground, while the primary winding of the transformer can be connected as either a differential or single-ended RF input.

It is desirable to understand why the chopping mixer is able to cancel the DC offset. Figure 13.12, in which the variations of the frequency spectrums of various signals are depicted step by step, explains its operating principle. The frequency spectrum of a signal in Figure 13.12 is drawn with a rectangular frame.

Figure 13.12(a) and (b) shows the frequency spectra of the RF signal and LO injection at the input, respectively. The spectra of the RF signal and LO injection are located in the same frequency range because they are designed for a direct conversion or zero IF system. Their central frequencies are marked with  $f_{RF}$  and  $f_{LO}$ .

The input chopping block chops the incoming RF signal ON and OFF with a repetition frequency of  $f_{CH}$ . Figure 13.12(c) shows the frequency spectrum after the input chopping block. The spectrum of the RF signal is split into two portions by chopping:

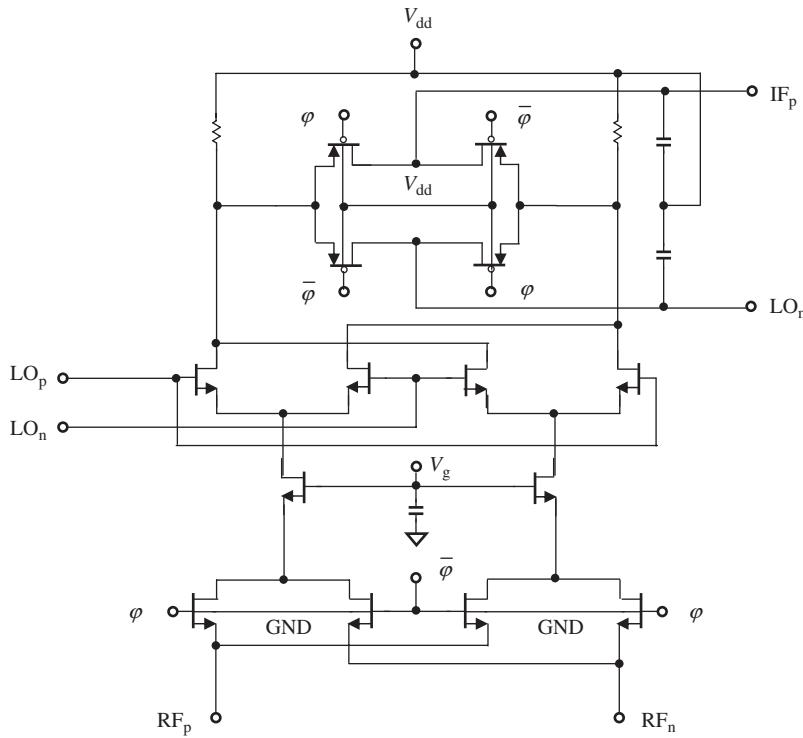


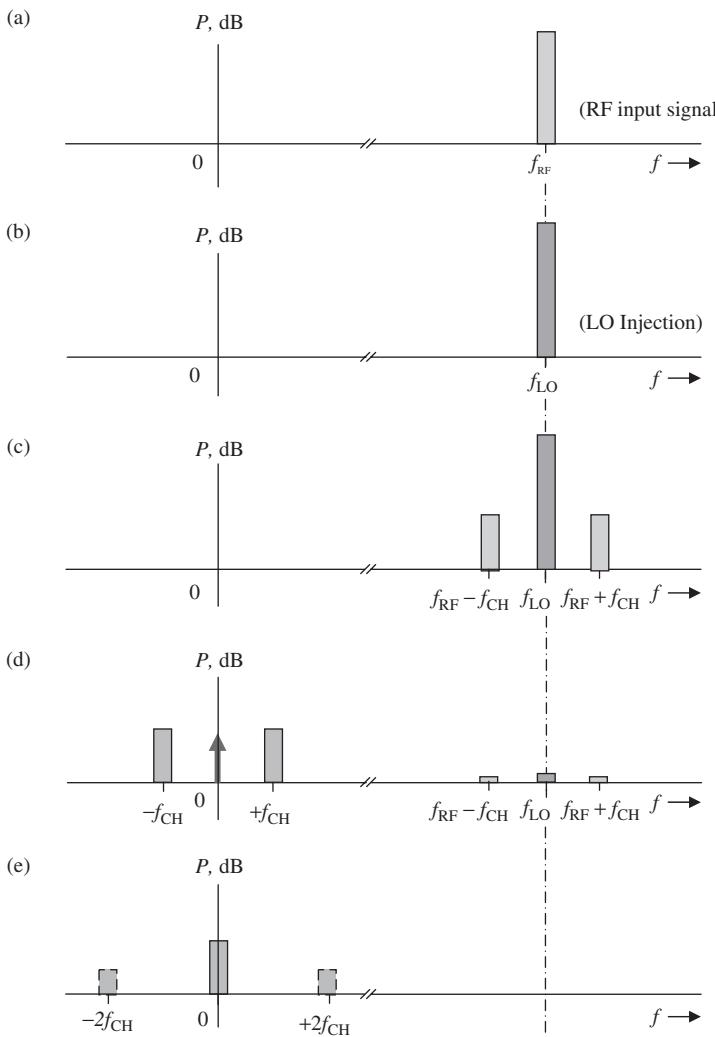
Figure 13.11. Topology of a chopping mixer.

one is the spectrum shifted from the original central location  $f_{RF}$  to  $f_{RF} - f_{CH}$ ; the other is the spectrum shifted from the original central location  $f_{RF}$  to  $f_{RF} + f_{CH}$ . The LO spectrum in Figure 13.12(c) is kept unchanged as shown in Figure 13.12(b).

Figure 13.12(d) shows the frequency spectrum after the Gilbert cell. The Gilbert cell mixes the RF signal and LO injection as shown in Figure 13.12(c) and produces two IF signals with spectrums located at  $-f_{CH}$  and  $f_{CH}$ , respectively as shown in Figure 13.12(d). Owing to imperfect isolation, a few percent of the power of the RF signal and LO injection are leaked from the RF and LO input ports to the IF output ports so that their spectra can be seen in Figure 13.12(d). Meanwhile, a DC offset is produced because of the imperfect differential structure. This is marked with an upward arrow in Figure 13.12(d).

The output chopping block chops the IF signals ON and OFF with the same repetition frequency  $f_{CH}$  as in the input chopping block. Figure 13.12(e) shows the spectrum of the IF signals after the output chopping block. Each spectrum of the IF signals is split into two portions. The IF signal with its spectrum located at  $-f_{CH}$  is split into two portions: one is the spectrum shifted from the original central location  $-f_{CH}$  to  $-2f_{CH}$ , while the other is the spectrum shifted from the original central location  $-f_{CH}$  to 0. The IF signal with its spectrum located at  $+f_{CH}$  is split into two portions as well: one is the spectrum shifted from the original central location  $+f_{CH}$  to  $+2f_{CH}$ , while the other is the spectrum shifted from the original central location  $+f_{CH}$  to 0. As a result, there are two spectra overlapping in the frequency range with a central frequency  $f = 0$ , which is the so-called zero IF.

In the output impedance matching network which is not shown in the circuit block diagram, the impedance is matched in the range of the zero IF, so that the two IF signals



**Figure 13.12.** Principle of DC offset reduction in the chopping mixer. (a) At the RF inputs. (b) At the LO inputs. (c) After input chopping but before Gilbert cell. (d) After Gilbert cell but before output chopping.

with their central frequencies located at  $-2f_{\text{CH}}$  and  $+2f_{\text{CH}}$ , which are drawn with dashed rectangular frames in Figure 13.12(e), are greatly reduced or removed.

### 13.4 DC OFFSET CANCELLATION BY CALIBRATION

Theoretically, zero DC offset does not exist in an ideal differential pair; however, it definitely exists in a real-life differential pair. In reality, an ideal differential pair does not exist. In engineering design, asymmetry between the two branches of the differential pair, such as nonidentical parts, inhomogeneous expansion, or constriction of the substrate or PCB, asymmetrical layout, etc., is inevitable. Consequently, a DC offset always exists in any circuit containing active devices, despite the differential configuration. In current communication systems implementing direct conversion or zero IF modulation technology, DC offset cancellation is one of most important tasks to be done in addition

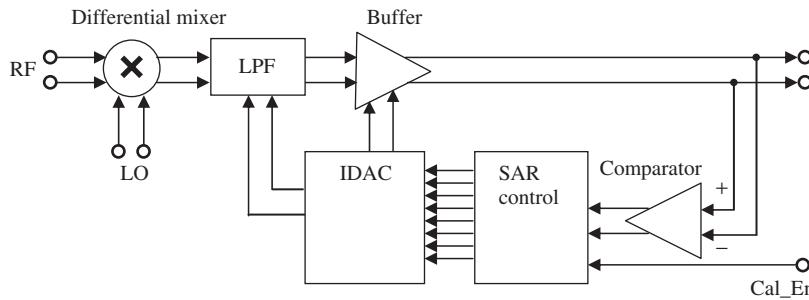


Figure 13.13. Simplified block diagram of DC offset calibration in Intersil's baseband IC, HFA3783.

to adapting the differential configuration in the circuitry. As long as DC offset cancellation is implemented well in a communication system, the DC offset will be kept below a threshold level, and the system will operate in a normal state.

There are various ways to implement DC offset cancellation. Figure 13.13 shows DC offset cancellation executed in terms of DC offset calibration, which has been developed by Intersil in their baseband IC, HFA3783, for WLAN products.

To keep the DC offset constant, all the DC offsets produced by the mixer, LPF (low-pass filter), and buffer are calibrated at an appropriate time during receiving, transmitting, or stand-by modes.

The comparator in Figure 13.13 senses the DC offset from the buffer output. Then, the output of the buffer is fed to a decision circuit called the SAR (successive approximation register) state control. The eight digital outputs of the SAR control the IDAC (current output of digital-to-analog conversion) which provides controls to the LPF and buffer so as to adjust and bring the DC offset down to a minimum.

The DC offset calibration is initialized by the *Cal\_En* (calibration enable) input either automatically or manually.

### 13.5 REMARK ON DC OFFSET CANCELLATION

It is difficult to measure the DC offset directly in the test laboratory because its absolute value is too low. DC offset can be indirectly measured by IP<sub>2</sub> since DC offset is contributed by the even orders of the nonlinearity of a device or block or by the even orders of the interference inside and outside the block. The second order of nonlinearity or interference is, of course, the most effective contributor. It is therefore reasonable to take IP<sub>2</sub> as a measure of the DC offset of a device, a block, or a system. Further study might be conducted on the equivalence between the DC offset and IP<sub>2</sub>.

Empirically, in a direct conversion or zero IF communication system, an RF block or system would not have any problem due to DC offset if its IIP<sub>2</sub> is greater than 80 dB<sub>m</sub>.

### FURTHER READING

Eynde FO, Vraninckx J, Gootschalckx P. A fully\_integrated zero-if DECT transceiver. ISSCC 2000/Session 8/Wireless Rx/Tx/Paper TA8.1, 2000 IEEE International Solid-State Circuits Conference, San Francisco, CA, 2000. pp. 138–139.

Faulkner M. IM2 removal in direct conversion direct conversion receivers. Volume 3, IEEE VTS 53rd Vehicular Technology Conference VTC, Rhodes; May 6–9; 2001. pp. 1897–1901.

- Faulkner M. DC offset DC offset and IM2 removal in direct conversion direct conversion receivers. IEE Proc Commun 2002; 149(3): 179–184.
- Huang PC, Chen YH, Wang CK. A 2-V CMOS 455-kHz FM/FSK demodulator using feedforward offset cancellation limiting amplifier. IEEE J Solid-State Circ 2001; 36(1): 135–138.
- Hull CD, Tham JL, Chu RR. A direct-conversion receiver for 999 MHz (ISM Band) spread-spectrum digital cordless telephone. IEEE J Solid-State Circ 1996; 31(12): 1955–1963.
- Matinpour B, Chakraborty S, Hamai M, et al. A novel dc-offset cancellation technique for even-harmonic direct conversion direct conversion receivers. Volume 2, IEEE MTT-S International Microwave Symposium Digest; 2000 June 11–16, Boston, MA. 2000. pp. 631–634.
- Matinpour B, Chakraborty S, Laskar J. Novel DC-offset cancellation techniques for even-harmonic direct conversion direct conversion receivers. IEEE Trans Microw Theory Tech 2000; 48(12): 2554–2559.
- Popov OV. Dynamic DC offset DC offset impact on the 802.11a receiver performance. 1st IEEE International Conference on Circuits and Systems for Communications Proceedings of ICCSC '02; Jun 26–28; 2002. pp. 250–253.
- Sevenhuijsen J, Verstraeten B, Taraborrelli S. Trends in silicon radio large scale integration: zero if receiver! zero I & Q transmitter! zero discrete passives! Topics in circuits for communications. 2000 IEEE, IEEE Communications Magazine, January 2000; 142–147.
- Shang S, Mirabbasi S, Saleh R. A technique for DC-offset removal and carrier phase error compensation in integrated wireless receivers. Volume 1, Proceedings of the 2003 International Symposium on Circuits and Systems; 2003 May 25–28. 2003. pp. I–173–I–176.
- Shoval A, Johns DA, Snelgrove WM. DC offset DC offset performance of four LMS adaptive algorithms. Volume 2, IEEE International Symposium on Circuits and Systems, ISCAS '94, London; May 30–June 2, 1994. pp. 409–412.
- Shoval A, Johns DA, Snelgrove WM. Comparison of DC offset DC offset effects in four LMS adaptive algorithms. IEEE Trans Circuits Syst II Analog Digital Signal Process (see also IEEE Trans Circuits Syst II Express Briefs) 1995; 42(3): 176–185.
- Sivonen P, Vilander A, Parssinen A. Cancellation of second-order intermodulation distortion and enhancement of IIP2 in common-source and common-emitter RF transconductors IEEE Trans Circuits Syst 2005; 52(2): 305–317.
- Stroet PM, Mohindra R, Hahn S, et al. A zero-if single-chip transceiver for up to 22Mb/s QPSK 802.11b Wireless LAN. ISSCC 2001/Session 13/Wireless LAN/13.5, 2001 IEEE International Solid-State Circuits Conference; 2001. pp. 204–206.
- Xinping H, Caron M. Gain/phase imbalance and DC offset DC offset compensation in quadrature modulators. Volume 4, IEEE International Symposium on Circuits and Systems, ISCAS 2002; 2002 May 26–29. 2002. pp. IV–811–IV–814.
- Yoshida H, Tsurumi H, Suzuki Y. DC offset DC offset canceller in a direct conversion direct conversion receiver for QPSK signal reception. Volume 3, 9th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Boston, MA; Sept 8–11, 1998. pp. 1314–1318.

## EXERCISES

1. In a zero IF receiver, does it mean that the demodulated signal is a DC voltage or DC current since zero IF means that  $f_{IF} = 0$ ?
2. How do you test and characterize the IF port of a mixer designed for a zero IF receiver?
3. What is the main advantage of a zero IF communication system?
4. What is the main feature of the DC offset?

5. What is the main advantage of applying the differential pair in the circuitry for a zero IF communication system?
6. What are the main disadvantages to apply the differential pair in the circuitry for a zero IF communication system?
7. A differential pair can cancel more noise than a single-ended stage does: is it true?
8. What is the main scheme of a chopping mixer to reduce the DC offset?
9. Why the blocking capacitor in a zero IF communication system cannot block the DC offset?
10. Assuming that a 10-pF blocking capacitor is connected with a 1-kΩ load, is it possible to block the DC offset existing in a 1-GHz RF signal?

## ANSWERS

1. In a zero IF receiver, it is true that the center point of the demodulated signal's spectrum has the component of a DC voltage or DC current because zero IF means that  $f_{IF} = 0$ . However, the spectrum of the demodulated signal covers a frequency bandwidth, from  $-\Delta f_{IF}/2$  to  $+\Delta f_{IF}/2$ , where  $\Delta f_{IF}$  is the bandwidth of IF. The DC voltage or DC current is only one of the components in its entire frequency bandwidth.
2. It is impossible to test and characterize the IF port of a mixer designed for a zero IF receiver when  $f_{IF} = 0$ , because all the RF testing through the  $S$  parameters pertains to AC cases and is meaningless in the DC case. Instead, the IF port of mixer must be tested and characterized in the cases when  $f_{IF} = \Delta f_{IF}/4$  and  $\Delta f_{IF}/2$ .
3. The low cost is the main advantage of a zero IF communication system. There are two factors by which the cost of a zero IF communication system can be greatly reduced:
  - (a) In a zero IF communication system, many filters, either an RF or an IF filter, can be removed.
  - (b) Consequently, the system can be implemented by IC technology since there are no filters in the system.
4. The main feature of DC offset is that this DC component is directly related and very sensitive to the input RF signal voltage or power. The DC offset from the nonlinearity of a single device is

$$\text{DC offset} = v_{io}^2 [(1/2) a_2 + (3/8) a_4 v_{io}^2 + \dots]$$

where  $a_2, a_4, \dots$  are the even-order nonlinearity coefficients of the device.

5. The main advantage of applying the differential pair in the circuitry for a zero IF communication system is
  - (a) Theoretically, the DC offset= 0.
  - (b) Also, all the odd-order nonlinearities of the device are cancelled from each other.
6. In a zero IF communication system, the main disadvantages in applying the differential pair rather than the single-ended stage are
  - (a) the number of parts in a differential pair is nearly double that of a single-ended stage;

- (b) the layout area for the differential pair is nearly double that of the single-ended stage;
  - (c) the cost for a differential pair is nearly double that of a single-ended stage;
  - (d) the part count and layout area would actually be more than double that of a single-ended stage;
  - (e) the current drain is usually double that of each single-ended stage;
  - (f) the symmetry of the layout must be maintained in a differential pair circuitry but not in a single-ended stage;
  - (g) a balun is needed for the conversion between the differential pair and the single-ended stage.
7. It is not true that a differential pair can cancel more noise than a single-ended stage does. On the contrary, the noise figure of a differential pair is higher than that of a single-ended stage generally.
8. The main scheme of a chopping mixer to reduce the DC offset is to shift the entire frequency spectrum of the RF signal back and forth, so that the DC offset can be moved and filtered away.
9. The blocking capacitor in a zero IF communication system can block the DC offset if the condition below is satisfied:

$$T \gg CR_L,$$

where

$T$  = the period of RF signal,

$C$  = the capacitance of the blocking capacitor, and

$R_L$  = resistance of load.

Unfortunately, the condition of  $T \gg CR_L$  usually is not satisfied in the RF circuitry within the RF frequency range, so that the blocking capacitor in a zero IF communication system cannot block the DC offset.

10. The time constant

$$CR_L = 10^{-8}\text{s}$$

if  $C = 10\text{ pF}$ , and  $R_L = 1\text{k}\Omega$ .

On the other hand, the period of RF signal

$$T = 1/f = 10^{-9}\text{s}.$$

Obviously, the condition of  $T \gg CR_L$  is not satisfied so that it is impossible to block the DC offset existing in a 1-GHz RF signal.

# DIFFERENTIAL PAIRS

## 14.1 FUNDAMENTALS OF DIFFERENTIAL PAIRS

### 14.1.1 Topology and Definition of a Differential Pair

For the integrity of description for the single-ended block and the differential pair, let us copy Figure 13.1, Figure 13.2, and equations (13.1) to (13.4) to this section and relabel them as Figure 14.1, Figure 14.2, and equations (14.1) to (14.4), respectively.

$$R_{c1} = R_{c2}, \quad (14.1)$$

$$Q_1 = Q_2, \quad (14.2)$$

$$R_{d1} = R_{d2}, \quad (14.3)$$

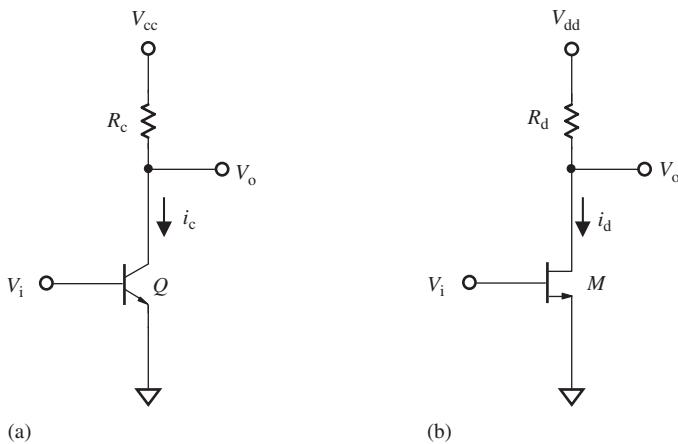
$$M_1 = M_2, \quad (14.4)$$

It should be reiterated that both differential branches are expected to be symmetrical to each other in all aspects, including their values, sizes, configurations, and layouts.

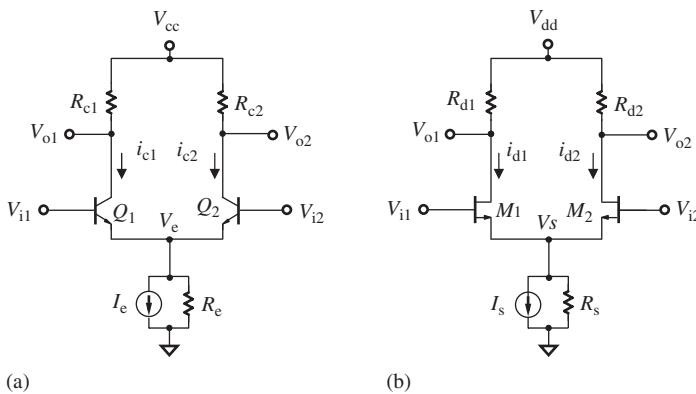
In order to clarify the DC and AC characters of the differential pair, it is better to denote the voltages and currents with DC and AC components, such as

$$V_{i1} = V_{II} + v_{i1}, \quad (14.5)$$

$$V_{i2} = V_{I2} + v_{i2}, \quad (14.6)$$



**Figure 14.1.** Typical single-ended block. (a) A bipolar single-ended block. (b) A MOSFET single-ended block.



**Figure 14.2.** Typical differential pair.  
(a) A bipolar differential pair.  
(b) A MOSFET differential pair.

$$V_{o1} = V_{O1} + v_{o1}, \quad (14.7)$$

$$V_{o2} = V_{O2} + v_{o2}, \quad (14.8)$$

$$I_{c1} = I_{C1} + i_{c1}, \quad (14.9)$$

$$I_{c2} = I_{C2} + i_{c2}, \quad (14.10)$$

$$I_{d1} = I_{D1} + i_{d1}, \quad (14.11)$$

$$I_{d2} = I_{D2} + i_{dc2}, \quad (14.12)$$

where

\$V\_{i1}\$ = sum of DC and AC voltage at input port 1,

\$V\_{i2}\$ = sum of DC and AC voltage at input port 2,

\$V\_{I1}\$ = DC voltage at input port 1,

\$V\_{I2}\$ = DC voltage at input port 2,

\$v\_{i1}\$ = AC voltage at input port 1,

\$v\_{i2}\$ = AC voltage at input port 2,

\$V\_{o1}\$ = sum of DC and AC voltage at output port 1,

\$V\_{o2}\$ = sum of DC and AC voltage at output port 2,

\$V\_{O1}\$ = DC voltage at output port 1,

- $V_{O2}$  = DC voltage at output port 2,  
 $v_{o1}$  = AC voltage at output port 1,  
 $v_{o2}$  = AC voltage at output port 2,  
 $I_{c1}$  = sum of DC and AC current into collector of  $Q_1$ ,  
 $I_{c2}$  = sum of DC and AC current into collector of  $Q_2$ ,  
 $I_{C1}$  = DC current into collector of  $Q_1$ ,  
 $I_{C2}$  = DC current into collector of  $Q_2$ ,  
 $i_{c1}$  = AC current into collector of  $Q_1$ ,  
 $i_{c2}$  = AC current into collector of  $Q_2$ ,  
 $I_{d1}$  = sum of DC and AC current into drain of  $M_1$ ,  
 $I_{d2}$  = sum of DC and AC current into drain of  $M_2$ ,  
 $I_{D1}$  = DC current into drain of  $M_1$ ,  
 $I_{D2}$  = DC current into drain of  $M_2$ ,  
 $i_{d1}$  = AC current into drain of  $M_1$ ,  
 $i_{d2}$  = AC current into drain of  $M_2$ ,

For a differential pair, the input and output signals can be expressed by voltage as shown below:

$$V_{I1} = V_{I2}, \quad (14.13)$$

$$v_{i1} = -v_{i2}, \quad (14.14)$$

or

$$|v_{i1}| = |v_{i2}|, \angle v_{i1} - \angle v_{i2} = 180^\circ \text{ or } -180^\circ, \quad (14.15)$$

and

$$V_{O1} = V_{O2}, \quad (14.16)$$

$$v_{o1} = -v_{o2}, \quad (14.17)$$

or

$$|v_{o1}| = |v_{o2}|, \angle v_{o1} - \angle v_{o2} = 180^\circ \text{ or } -180^\circ, \quad (14.18)$$

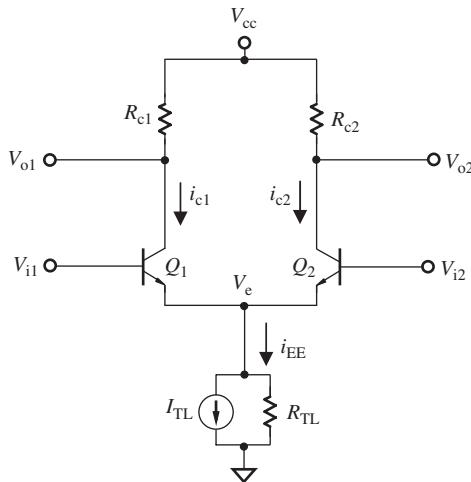
and the collector currents of a bipolar differential pair or drain currents of a MOSFET (metal–oxide–semiconductor field-effect transistor) differential pair are in a perfectly symmetrical state, that is,

$$I_{c1} = I_{c2}, \quad (14.19)$$

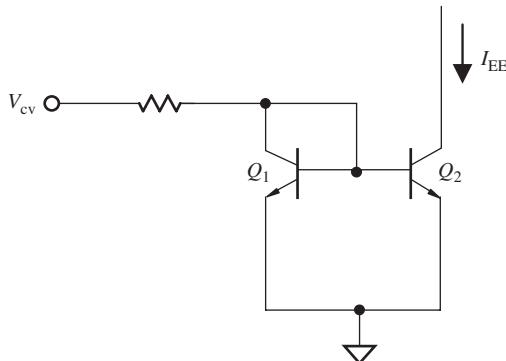
or

$$I_{d1} = I_{d2}. \quad (14.20)$$

Expressions (14.1) to (14.20) are tedious. However, they are important in distinguishing the DC and AC components in the expressions of voltages and currents. Some published books or papers discuss “DC transfer characteristics” of a device and then treat them as “AC transfer characteristics.” As matter of fact, the transfer characteristics of a device cover both DC and AC components.



**Figure 14.3.** Plot of bipolar differential pair for discussing its transfer characteristics.



**Figure 14.4.** Tail current  $I_{TL}$  and  $R_{TL}$  can be replaced by a current mirror and controlled by  $V_{cv}$ .

### 14.1.2 Transfer Characteristic of a Bipolar Differential Pair

Figure 14.3 shows a differential pair with bipolar devices. The emitter resistor  $R_E$  is renamed as the tail resistor,  $R_{TL}$ . The tail current  $I_{TL}$  and  $R_{TL}$  can be replaced by a current mirror and controlled by a control voltage  $V_{cv}$  as shown in Figure 14.4.

The transfer characteristics can be derived from the exponential relationship between the base voltage and the collector current.

$$I_{c1} = I_{S1} \left( 1 + \frac{V_{ce1}}{V_{A1}} \right) \exp \left( \frac{V_{be1}}{V_T} \right), \quad (14.21)$$

$$I_{c2} = I_{S2} \left( 1 + \frac{V_{ce2}}{V_{A2}} \right) \exp \left( \frac{V_{be2}}{V_T} \right), \quad (14.22)$$

where

$I_{c1}$  = collector current of  $Q_1$ ,

$I_{c2}$  = collector current of  $Q_2$ ,

$I_{S1}$  = saturated current of  $Q_1$ ,

$I_{S2}$  = saturated current of  $Q_2$ ,

$V_{ce1}$  = collector–emitter voltage drop of  $Q_1$ ,

$V_{ce2}$  = collector-emitter voltage drop of  $Q_2$ ,  
 $V_{be1}$  = base-emitter voltage drop of  $Q_1$ ,  
 $V_{be2}$  = base-emitter voltage drop of  $Q_2$ ,  
 $V_{A1}$  = early voltage of  $Q_1$ ,  
 $V_{A2}$  = early voltage of  $Q_2$ ,  
 $V_T$  = thermal voltage =  $kT/q = 26$  mV at 300 K,  
 $k$  = Boltzmann constant =  $8.617 \times 10^{-5}$  eV/deg,  
 $T$  = environment temperature, K, and  
 $q$  = charge of an electron =  $1.6 \times 10^{-19}$  C.

Assuming that the transistors and collector resistors are identical, that is,

$$I_{S1} = I_{S2} = I_S, \quad (14.23)$$

$$V_{ce1} = V_{ce2} = V_{ce}, \quad (14.24)$$

$$V_{A1} = V_{A2} = V_A, \quad (14.25)$$

$$\alpha_{F1} = \alpha_{F2} = \alpha_F, \quad (14.26)$$

$$R_{c1} = R_{c2} = R_c. \quad (14.27)$$

where

$\alpha_{F1}$  = ratio of collector to emitter current of  $Q_1$ ,

$\alpha_{F2}$  = ratio of collector to emitter current of  $Q_2$ ,

$R_{c1}$  = collector resistor of  $Q_1$ , and

$R_{c2}$  = collector resistor of  $Q_2$ .

Then, from equations (14.21) and (14.22) we have

$$V_{be1} = V_T \ln \frac{I_{c1}}{I_S \left(1 + \frac{V_{ce}}{V_A}\right)}, \quad (14.28)$$

$$V_{be2} = V_T \ln \frac{I_{c2}}{I_S \left(1 + \frac{V_{ce}}{V_A}\right)}. \quad (14.29)$$

Define the input and output differential voltages  $V_{id}$  and  $V_{od}$  as

$$V_{id} = V_{i1} - V_{i2}, \quad (14.30)$$

$$V_{od} = V_{o1} - V_{o2}. \quad (14.31)$$

In the loop of  $V_{i1} \rightarrow$  emitter of  $Q_1$  and  $Q_2 \rightarrow V_{i2}$ ,

$$V_{be1} - V_{be2} = V_{i1} - V_{i2}, \quad (14.32)$$

From expression (14.21) to (14.32) we have

$$\frac{I_{c1}}{I_{c2}} = \exp \left( \frac{V_{be1} - V_{be2}}{V_T} \right) = \exp \left( \frac{V_{i1} - V_{i2}}{V_T} \right) = \exp \left( \frac{V_{id}}{V_T} \right). \quad (14.33)$$

By introducing the tail current  $I_{TL}$  we have

$$I_{TL} = \frac{I_{c1} + I_{c2}}{\alpha_F}. \quad (14.34)$$

From expressions (14.33) and (14.34) we have

$$I_{c1} = \frac{\alpha_F I_{TL}}{1 + \exp\left(-\frac{V_{id}}{V_T}\right)}, \quad (14.35)$$

$$I_{c2} = \frac{\alpha_F I_{TL}}{1 + \exp\left(\frac{V_{id}}{V_T}\right)}, \quad (14.36)$$

$$\Delta I_c = I_{c1} - I_{c2} = -\alpha_F I_{TL} \tanh\left(\frac{V_{id}}{2V_T}\right). \quad (14.37)$$

This is the transfer characteristic of the bipolar differential pair and is plotted in Figure 14.5.

The output differential current in the transfer characteristic expression (14.37) can be converted to the output differential voltage.

Note that

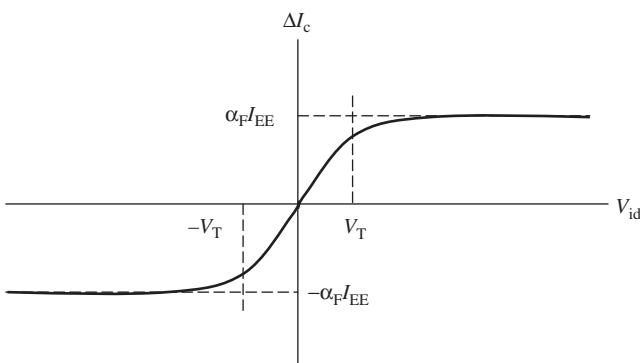
$$V_{o1} = V_{cc} - I_{c1}R_c, \quad (14.38)$$

$$V_{o2} = V_{cc} - I_{c2}R_c. \quad (14.39)$$

Finally,

$$V_{od} = V_{o1} - V_{o2} = \alpha_F I_{TL} R_c \tanh\left(\frac{-V_{id}}{2V_T}\right). \quad (14.40)$$

Expression (14.37) or (14.40) presents a relationship between the output differential current or voltage and the input differential voltage of a bipolar differential pair. Obviously, both of them are proportional to the tail current  $I_{TL}$ . On the other hand, the output differential current or voltage does not have a linear relationship with the input differential voltage. Instead, it is a tangent hyperbolic function of the input differential



**Figure 14.5.** Transfer characteristic of a bipolar differential pair with emitter coupling.

voltage. However, from Figure 14.5 it can be seen that they can be approximated as a linear relationship if

$$|V_{id}| < V_T. \quad (14.41)$$

### 14.1.3 Small Signal Approximation of a Bipolar Differential Pair

It should be noted that all the current and voltages in equation (14.40) contain both DC and AC components. From (14.5) to (14.8), (14.13) to (14.18), and (14.30) and (14.31) we have

$$\begin{aligned} V_{od} &= V_{o1} - V_{o2} = (V_{O1} + v_{o1}) - (V_{O2} + v_{o2}) \\ &= (V_{O1} - V_{O2}) + (v_{o1} - v_{o2}) = v_{o1} - v_{o2} = v_{od}, \end{aligned} \quad (14.42)$$

$$\begin{aligned} V_{id} &= V_{i1} - V_{i2} = (V_{I1} + v_{i1}) - (V_{I2} + v_{i2}) \\ &= (V_{I1} - V_{I2}) + (v_{i1} - v_{i2}) = v_{i1} - v_{i2} = v_{id}. \end{aligned} \quad (14.43)$$

Then, expression (14.40) can be changed into a relationship between the AC output differential voltage  $v_{od}$  and the AC input differential voltage  $v_{id}$ :

$$v_{od} = V_{o1} - V_{o2} = \alpha_F I_{TL} R_c \tanh\left(\frac{-v_{id}}{2V_T}\right). \quad (14.44)$$

If the AC input differential voltage  $v_{id}$  is a small signal, that is,

$$v_{id} \ll 2V_T. \quad (14.45)$$

It is well known that a tangent hyperbolic function in the expression can be extended to a Maclaurin series, such as  
if

$$|x| < \frac{\pi}{2} \quad (14.46)$$

then,

$$\tanh x = x - \frac{1}{3}x^3 + \frac{2}{15}x^5 - \frac{17}{315}x^7 + \frac{62}{2835}x^9 - \dots \quad (14.47)$$

Now,  
if

$$x = -\frac{v_{id}}{2V_T}, \quad (14.48)$$

then expression (14.44) becomes

$$\begin{aligned} v_{od} &= \alpha_F I_{TL} R_c \left[ \left( \frac{-v_{id}}{2V_T} \right) - \frac{1}{3} \left( \frac{-v_{id}}{2V_T} \right)^3 + \frac{2}{15} \left( \frac{-v_{id}}{2V_T} \right)^5 \right. \\ &\quad \left. - \frac{17}{315} \left( \frac{-v_{id}}{2V_T} \right)^7 + \frac{62}{2835} \left( \frac{-v_{id}}{2V_T} \right)^9 - \dots \right] \\ &= \alpha_F I_{TL} R_c \left[ -\frac{1}{2} \left( \frac{v_{id}}{V_T} \right) + \frac{1}{24} \left( \frac{v_{id}}{V_T} \right)^3 - \frac{1}{240} \left( \frac{v_{id}}{V_T} \right)^5 \right. \\ &\quad \left. + \frac{17}{40,320} \left( \frac{v_{id}}{V_T} \right)^7 - \frac{62}{1,451,520} \left( \frac{-v_{id}}{2V_T} \right)^9 - \dots \right], \end{aligned} \quad (14.49)$$

in which the terms with orders higher than 2 could be neglected. Then, approximately, we obtain

$$v_{od} \approx -\frac{1}{2} \alpha_F I_{TL} R_c \frac{v_{id}}{V_T}. \quad (14.50)$$

In actual RF circuit design, expressions (14.44) and (14.50), respectively, correspond to the cases of a modulator and a mixer design. In present designs, the Gilbert cell is the core of the designed block in both modulators and mixers. The big difference is that if the input is not a small signal in the modulator design, an arc-tangent-hyperbolic block must be added before the Gilbert cell so as to “neutralize” the tangent hyperbolic function as shown in (14.44) and result in a linear modulation.

#### 14.1.4 Transfer Characteristic of a MOSFET Differential Pair

Figure 14.6 shows a MOSFET differential pair. The source resistor  $R_s$  is renamed as the tail resistor,  $R_{TL}$ .

The transfer characteristics can be derived from the square relationship between the gate voltage and the drain current.

$$I_{d1} = \frac{\mu_n C_{ox}}{2} \frac{W_1}{L_1} (V_{gs1} - V_{th})^2 (1 + \lambda V_{ds1}), \quad (14.51)$$

$$I_{d2} = \frac{\mu_n C_{ox}}{2} \frac{W_2}{L_2} (V_{gs2} - V_{th})^2 (1 + \lambda V_{ds2}), \quad (14.52)$$

where

$I_{d1}$  = drain current of  $n$  channel MOSFET  $M_1$ ,

$I_{d2}$  = drain current of  $n$  channel MOSFET  $M_2$ ,

$W_1$  = width of  $n$  channel MOSFET  $M_1$ ,

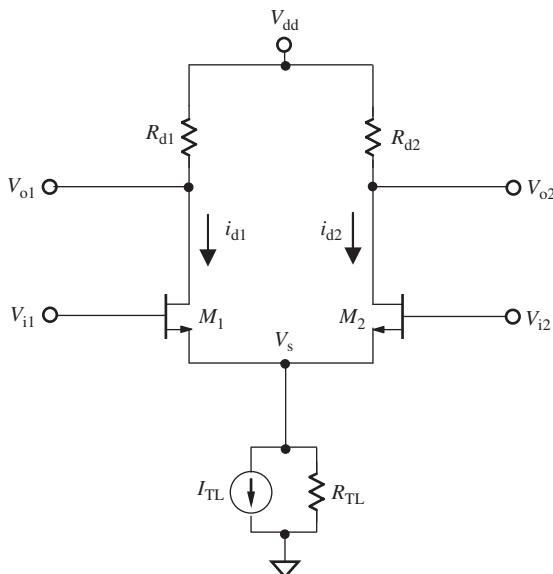
$W_2$  = width of  $n$  channel MOSFET  $M_2$ ,

$L_1$  = length of  $n$  channel MOSFET  $M_1$ ,

$L_2$  = length of  $n$  channel MOSFET  $M_2$ ,

$V_{gs1}$  = gate-source voltage for  $n$  channel MOSFET  $M_1$ ,

$V_{gs2}$  = gate-source voltage for  $n$  channel MOSFET  $M_2$ ,



**Figure 14.6.** Plot of MOSFET differential pair for discussing its transfer characteristics.

$V_{th}$  = threshold voltage for  $n$  channel MOSFET, the minimum gate-to-channel voltage needed for the carriers in the channel to exist,

$V_{ds1}$  = drain-source voltage for  $n$  channel MOSFET  $M_1$ ,

$V_{ds2}$  = drain-source voltage for  $n$  channel MOSFET  $M_2$ ,

and

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}, \quad (14.53)$$

where

$t_{ox}$  = thickness of the gate oxide layer, and

$\varepsilon_{ox}$  = permittivity of oxide layer.

From equations (14.51) and (14.52) we have

$$V_{gs1} - V_{th} = \sqrt{\frac{2I_{d1}}{\mu_n C_{ox} \frac{W_1}{L_1} (1 + \lambda V_{ds1})}}, \quad (14.54)$$

$$V_{gs2} - V_{th} = \sqrt{\frac{2I_{d2}}{\mu_n C_{ox} \frac{W_2}{L_2} (1 + \lambda V_{ds2})}}. \quad (14.55)$$

Assuming that the transistors are identical, that is,

$$W_1 = W_2 = W, \quad (14.56)$$

$$L_1 = L_2 = L, \quad (14.57)$$

$$V_{ds1} = V_{ds2} = V_{ds}. \quad (14.58)$$

From (14.54) to (14.58) we have

$$V_{id} = V_{i1} - V_{i2} = V_{gs1} - V_{gs2} = \frac{\sqrt{I_{d1}} - \sqrt{I_{d2}}}{\sqrt{\frac{\mu_n C_{ox}}{2} \frac{W}{L} (1 + \lambda V_{ds})}}. \quad (14.59)$$

On the other hand,

$$I_{d1} + I_{d2} = I_{TL}. \quad (14.60)$$

From (14.59) and (14.60),

$$I_{d1} = \frac{I_{TL}}{2} + \frac{\mu_n C_{ox}}{4} \frac{W}{L} (1 + \lambda V_{ds}) V_{id} \sqrt{\frac{4I_{TL}}{\frac{\mu_n C_{ox}}{2} \frac{W}{L} (1 + \lambda V_{ds})} - V_{id}^2}, \quad (14.61)$$

$$I_{d2} = \frac{I_{TL}}{2} - \frac{\mu_n C_{ox}}{4} \frac{W}{L} (1 + \lambda V_{ds}) V_{id} \sqrt{\frac{4I_{TL}}{\frac{\mu_n C_{ox}}{2} \frac{W}{L} (1 + \lambda V_{ds})} - V_{id}^2}. \quad (14.62)$$

Consequently,

$$\Delta I_d = I_{d1} - I_{d2} = \frac{\mu_n C_{ox}}{2} \frac{W}{L} (1 + \lambda V_{ds}) V_{id} \sqrt{\frac{4I_{TL}}{\frac{\mu_n C_{ox}}{2} \frac{W}{L} (1 + \lambda V_{ds})} - V_{id}^2}, \quad (14.63)$$

$$V_{od} = V_{o1} - V_{o2} = (V_{dd} - I_{d1} R_d) - (V_{dd} - I_{d2} R_d) = -\Delta I_d R_d, \quad (14.64)$$

$$V_{od} = -R_d \frac{\mu_n C_{ox}}{2} \frac{W}{L} (1 + \lambda V_{ds}) V_{id} \sqrt{\frac{4I_{TL}}{\frac{\mu_n C_{ox}}{2} \frac{W}{L} (1 + \lambda V_{ds})} - V_{id}^2}. \quad (14.65)$$

Expression (14.65) presents a relationship between the input and output differential voltages of a MOSFET differential pair. This is called the *V–V transfer characteristic* of a MOSFET differential pair with source coupling. Obviously, the transfer characteristic of the input and output differential voltages in a MOSFET differential pair as shown in expression (14.65) is not a linear relationship. The output differential voltage is a complicated function of the input differential voltage.

#### 14.1.5 Small Signal Approximation of a MOSFET Differential Pair

It should be noted that all the currents and voltages in equation (14.65) contain both DC and AC components. From (14.42) and (14.43), it is found that expression (14.65) can be changed to a relationship between the AC output differential voltage  $v_{od}$  and the AC input differential voltage  $v_{id}$ :

$$v_{od} = -R_d \frac{\mu_n C_{ox}}{2} \frac{W}{L} (1 + \lambda V_{ds}) v_{id} \sqrt{\frac{4I_{TL}}{\frac{\mu_n C_{ox}}{2} \frac{W}{L} (1 + \lambda V_{ds})} - v_{id}^2}. \quad (14.66)$$

The square-root term in expression (14.66) can be extended to a Maclaurin series form:

if

$$x < a^2, \quad (14.67)$$

$$\sqrt{a^2 - x^2} = a \left[ 1 - \frac{1}{2} \left( \frac{x}{a} \right)^2 - \frac{1}{8} \left( \frac{x}{a} \right)^4 - \frac{3}{48} \left( \frac{x}{a} \right)^6 - \dots \right]. \quad (14.68)$$

If the input differential voltage  $v_{id}$  is a small signal, expression (14.68) can be applied to expression (14.66). Then, expression (14.66) becomes

$$v_{od} = -R_d \sqrt{2\mu_n C_{ox} \frac{W}{L} (1 + \lambda V_{ds}) I_{TL}} \times \left[ v_{id} - \frac{\mu_n C_{ox} \frac{W}{L} (1 + \lambda V_{ds})}{4I_{TL}} v_{id}^3 - \frac{1}{32} \left\{ \frac{\mu_n C_{ox} \frac{W}{L} (1 + \lambda V_{ds})}{I_{TL}} \right\}^2 v_{id}^5 - \dots \right], \quad (14.69)$$

in which we can neglect the terms with orders higher than 2, approximately obtaining

$$v_{od} \approx - \left( \sqrt{2\mu_n C_{ox} \frac{W}{L} (1 + \lambda V_{ds}) I_{TL}} \right) R_d v_{id}. \quad (14.70)$$

In actual RF circuit design, expressions (14.66) and (14.70) correspond to the cases of a modulator and a mixer design, respectively. In present designs, the Gilbert cell is the core of the designed block in both modulator and mixer. The big difference is that if the input is not a small signal in the modulator design, an additional block must be added before the Gilbert cell so that the square-root term as shown in (14.66) can be neutralized to result in a linear modulation.

#### 14.1.6 What Happens If Input Signal Is Imperfect Differential

From Section 14.1.2–14.1.5, the transfer characteristic of a bipolar and MOSFET differential pair has been introduced. Through the study of the transfer characteristic, a remarkable special feature of the differential pair is found, that is, that there is zero DC offset in an ideal differential pair. However, it must be noted that this conclusion is obtained under two ideal conditions:

1. The differential pair is ideal, with a perfectly symmetrical configuration, perfectly identical parts, perfectly symmetrical layout, and perfectly symmetrical grounding and DC power supply provided.
2. The input signal is perfectly differential.

The first ideal condition is easily understood via the derivation of expressions (14.13) to (14.18). However, the second ideal condition is relatively hard to overlook.

The DC offset cannot be zero, and the even orders nonlinearity cannot be cancelled by each other if the input signal is not perfectly differential, for example,

$$v_{i1} = v_{ic} + \frac{v_{id}}{2}, \quad (14.71)$$

$$v_{i2} = v_{ic} - \frac{v_{id}}{2}, \quad (14.72)$$

where

- $v_{id}$  = differential portion of input signal, and
- $v_{ic}$  = common portion of input signal.

This is not a perfect differential signal because

$$v_{i1} \neq -v_{i2}, \quad (14.73)$$

Consequently, the output is not a perfectly differential signal and can be expressed in the same form as the input, that is,

$$v_{o1} = v_{oc} + \frac{v_{od}}{2}, \quad (14.74)$$

$$v_{o2} = v_{oc} - \frac{v_{od}}{2}, \quad (14.75)$$

where

- $v_{od}$  = differential portion of output signal, and
- $v_{oc}$  = common portion of output signal.

By replacing expressions (14.13) and (14.14) of differential inputs with expressions (14.71) and (14.72) of imperfect differential inputs in the derivation of the output differential current or voltage as shown in the expressions (13.20) to (13.21), it can be found that the DC offset cannot be zero and the even orders of nonlinearity of the device do not cancel each other.

In reality, the input signal is always imperfect. Since this is the case, merits of the differential pair, zero DC offset, and cancellation of even orders of nonlinearity, all disappear. It seems meaningless to apply the differential pair in RF circuits.

However, we need not be without hope when facing adverse situations. It is now necessary to understand how well the differential pair can promote the differential portion of the input signal, and, on the other hand, how strongly the differential pair can reject the common portion of the input signal. If it can greatly magnify the differential portion and effectively suppress the common portion of the input signal, the differential pair might become a powerful circuit element. If this capability exists, zero DC offset can be approached and the even orders of nonlinearity of the device can be reduced to an insignificant level. In short, we now desire to understand, with respect to the desired differential portion, what the capability of the rejection the common portion is in a differential pair.

In a concrete way, we are going to study the CMRR (common mode rejection ratio) of the differential pair. The general form of the transfer equations for differential mode and common mode operation are

$$v_{od} = A_{dm}v_{id} + A_{cm-dm}v_{ic}, \quad (14.76)$$

$$v_{oc} = A_{dm-cm}v_{id} + A_{cm}v_{ic}, \quad (14.77)$$

where

- $A_{dm}$  = differential mode voltage gain, that is,  $v_{od}/v_{id}$  when  $v_{ic} = 0$ ,
- $A_{cm-dm}$  = common mode to differential mode voltage gain, that is,  $v_{od}/v_{ic}$  when  $v_{id} = 0$ ,

$A_{dm-cm}$  = differential mode to common mode voltage gain, that is,  $v_{oc}/v_{id}$  when  $v_{ic} = 0$ , and  
 $A_{cm}$  = common mode voltage gain, that is,  $v_{oc}/v_{ic}$  when  $v_{id} = 0$ .

The calculation of  $A_{dm-cm}$  and  $A_{cm-dm}$  for the bipolar transistor and MOSFET is somewhat tedious. They are usually a couple orders smaller than  $A_{dm}$  and  $A_{cm}$ , and can be neglected. Let us confine our discussion to only  $A_{dm}$  and  $A_{cm}$  since we are interested in the CMRR, which is defined as

$$\text{CMRR} = \frac{A_{dm}}{A_{cm}}. \quad (14.78)$$

This is the ratio of differential mode voltage gain to common mode voltage gain. There are three cases:

1. If  $\text{CMRR} < 1$ , or  $\text{CMRR} < 0 \text{ dB}$ , it signifies that the differential pair magnifies the input common portion more than it magnifies the input differential portion. The common mode operation rejects the differential mode operation.
2. If  $\text{CMRR} = 1$ , or  $\text{CMRR} = 0 \text{ dB}$ , it signifies that the differential pair magnifies the differential portion with the same number of times as it magnifies the common portion. The differential pair just operates like a single-ended stage.
3. If  $\text{CMRR} > 1$ , or  $\text{CMRR} > 0 \text{ dB}$ , it signifies that the differential pair magnifies the input differential portion more than it magnifies the input common portion. The differential mode operation rejects the common mode operation.

The first two cases are not welcome. A differential pair must operate in the third case, that is,  $\text{CMRR} > 1$  or  $\text{CMRR} > 0 \text{ dB}$ . Otherwise, the application of the differential pair is meaningless. As matter of fact, the value of CMRR is expected to be as high as possible. The higher the CMRR, the lower the DC offset, and more the cancellation of even orders of nonlinearity of the device. Generally speaking, a good design of a differential pair must ensure a CMRR of over 20 dB.

## 14.2 CMRR (COMMON MODE REJECTION RATIO)

### 14.2.1 Expression of CMRR

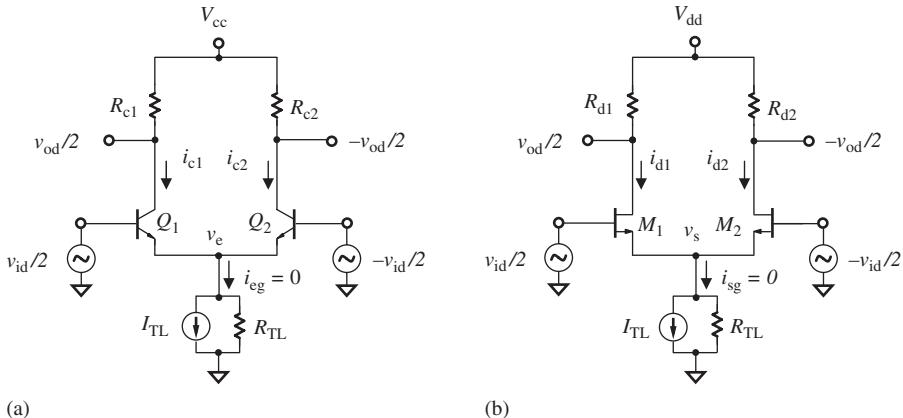
It is easy to imagine that the CMRR depends on the configuration of the differential pair, current drain, characteristics of transistors, and so on. In order to obtain the relationship between the CMRR and other circuit parameters in the differential pair, some basic assumptions are made about the differential pair:

1. The differential pair is ideal with a perfectly symmetrical configuration, perfectly identical parts, perfectly symmetrical layout, perfectly symmetrical grounding and DC power supply, and so on.
2. The input is an imperfect signal. It contains both differential and common mode portions.

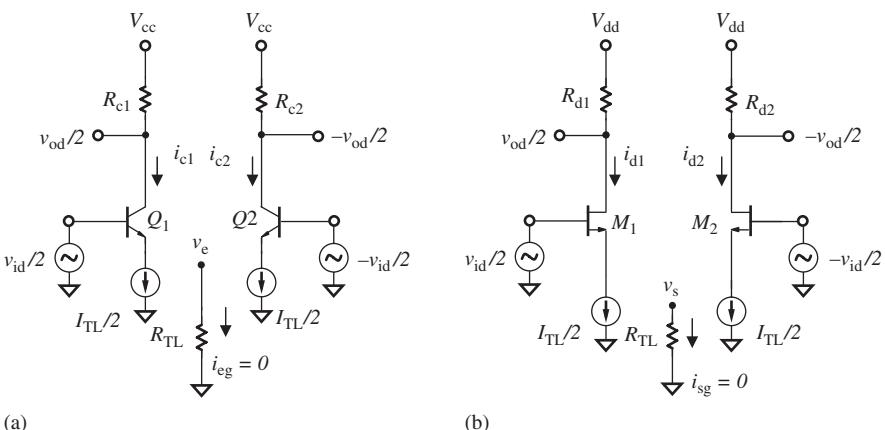
First, let us derive the expression of the voltage gain when the differential pair operates in the differential mode,  $A_{dm}$ , by which  $V_{i1}$ ,  $V_{i2}$ ,  $V_{o1}$ , and  $V_{o2}$  are replaced  $v_{id}/2$ ,  $-v_{id}/2$ ,  $v_{od}/2$ , and  $-v_{od}/2$ , respectively, Figure 14.2 is modified as Figure 14.7. In this figure, the tail current, either  $i_{eg}$  or  $i_{sg}$ , which flows from the node  $v_e$  or  $v_s$  to the ground, is zero, because the currents contributing to  $i_{eg}$  from both devices  $Q_1$  and  $Q_2$  or  $i_{sg}$  from both devices  $M_1$  and  $M_2$  are the same in magnitude but  $180^\circ$  different in phase.

This implies that the emitters or sources of devices in Figure 14.7 are virtually grounded and that the tail resistor  $R_{TL}$  is short circuited. In other words, in Figure 14.7(a) or (b), the left and right branches of the differential pair are two identical and independent half circuits. Figure 14.8 shows the two identical and independent half circuits after the left and right branches of the differential pair in Figure 14.3 are completely separated, in which the tail resistor  $R_{TL}$  becomes a dummy part.

The equivalents of the differential mode half circuit in the differential pair are sketched in Figure 14.9, which represents a low-frequency approximation. All the



**Figure 14.7.** Differential mode portion of differential pairs. (a) Bipolar differential pair. (b) MOSFET differential pair.



**Figure 14.8.** Differential mode portion of differential pairs. (a) Bipolar differential pair. (b) MOSFET differential pair.

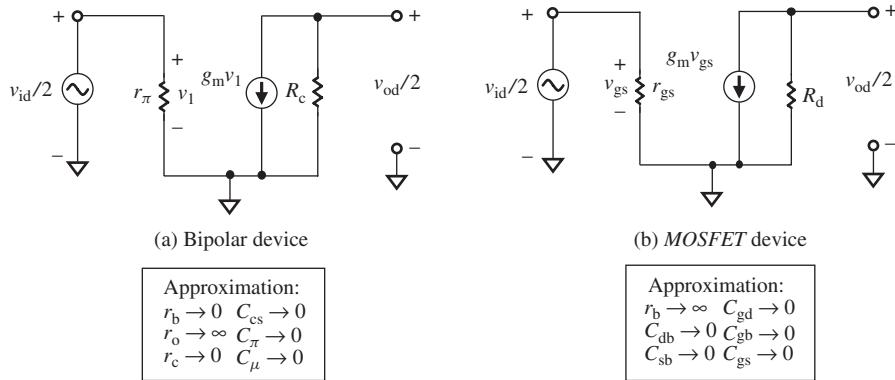


Figure 14.9. Equivalent of differential mode half circuit with low-frequency approximation.  
 (a) Bipolar device. (b) MOSFET device.

capacitors in the transistor's model are neglected. In the bipolar device model, only  $r_\pi$  and  $R_c$  are reserved and in the MOSFET device model only  $r_{gs}$  and  $R_d$  are reserved.

For the bipolar device, from Figure 14.9(a), we have

$$\frac{v_{od}}{2} = -g_m v_1 R_c = -g_m \frac{v_{id}}{2} R_c. \quad (14.79)$$

For the MOSFET device, from Figure 14.9(b), we have

$$\frac{v_{od}}{2} = -g_m v_{gs} R_d = -g_m \frac{v_{id}}{2} R_d. \quad (14.80)$$

Note that

$$R_c = R_{c1} = R_{c2}, \quad (14.81)$$

$$R_d = R_{d1} = R_{d2}, \quad (14.82)$$

$$v_1 = v_{gs} = \frac{v_{id}}{2}. \quad (14.83)$$

Let us merge the two equations (14.79) and (14.80) into one:

$$\frac{v_{od}}{2} = -g_m \frac{v_{id}}{2} R. \quad (14.84)$$

If  $R_c$  and  $R_d$  are commonly denoted by  $R$ , then

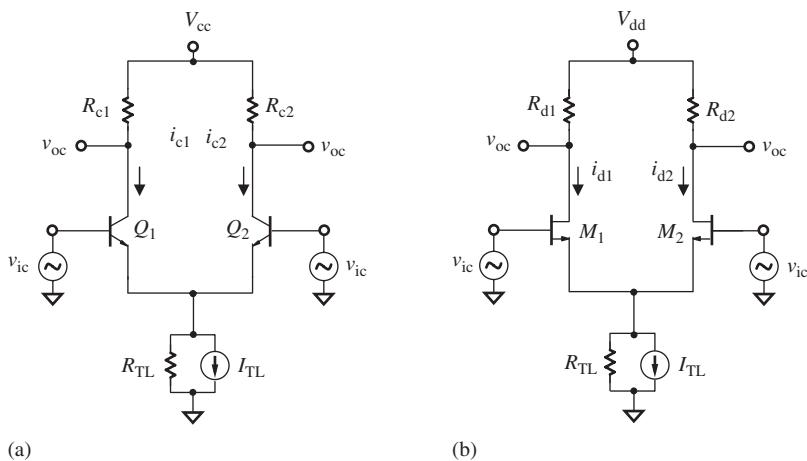
$$A_{dm} = \frac{v_{od}}{v_{id}} = -g_m R. \quad (14.85)$$

The differential mode voltage gain  $A_{dm}$  is proportional to  $g_m$  and  $R$ . The  $A_{dm}$  can be enhanced by increasing the value of  $R$  ( $R_c$  or  $R_d$ ). However, it might be detrimental to obtain a higher differential output power gain. Eventually, the effective means to enhance the  $A_{dm}$  is to increase the  $g_m$  of the devices.

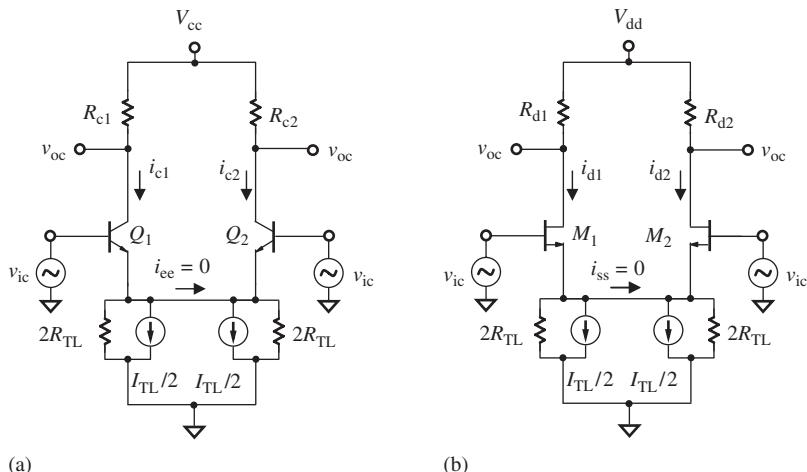
Next, let us derive the expression of the voltage gain when the differential pair operates in the common mode,  $A_{cm}$ . By replacing  $V_{i1}$ ,  $V_{i2}$ ,  $V_{o1}$ , and  $V_{o2}$  with  $v_{ic}$ ,  $v_{ic}$ ,  $v_{oc}$ , and  $v_{oc}$ , respectively, Figure 14.2 is modified to Figure 14.10.

The left and right branches of the differential pairs in Figure 14.10(a) or (b) are perfectly identical to each other. Since they are two identical, independent half circuits, Figure 14.10 can be redrawn as Figure 14.11, in which the current between the emitters  $i_{ee}$  or the current between the sources  $i_{ss}$ , is zero because each half circuit is driven by the same input voltage,  $v_{ic}$ . The connection between the emitters or sources of the device can be removed so that Figure 14.11 becomes Figure 14.12.

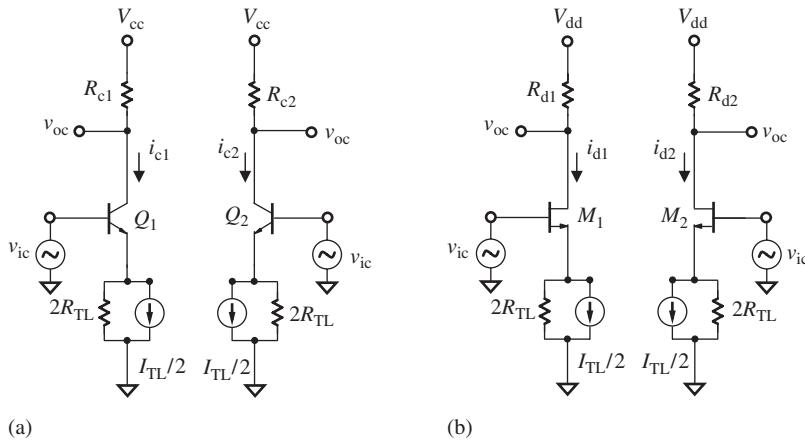
The equivalents of the common mode half circuit in the differential pair are sketched in Figure 14.13, which represents a low-frequency approximation. All the capacitors in the transistor's model are neglected. In the bipolar device model, only  $r_\pi$  and  $R_c$  are reserved, and in MOSFET device model, only  $r_{gs}$  and  $R_d$  are reserved.



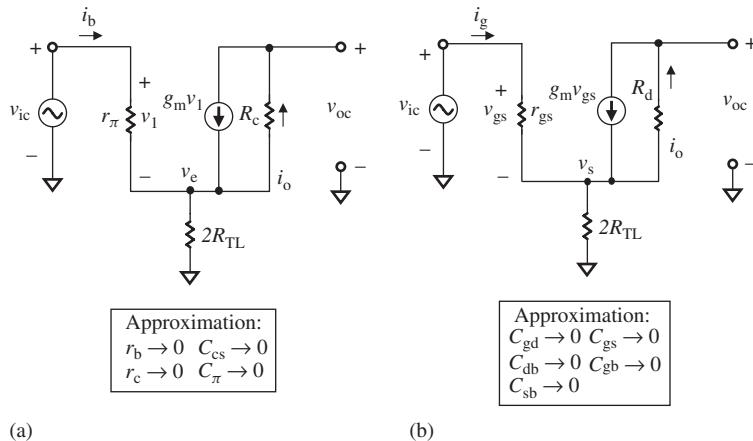
**Figure 14.10.** Common mode portion of differential pairs. (a) Bipolar differential pair. (b) MOSFET differential pair.



**Figure 14.11.** Common mode portion of differential pairs connected together. (a) Bipolar differential pair. (b) MOSFET differential pair.



**Figure 14.12.** Common mode portion of differential pairs disintegrated. (a) Bipolar differential pair. (b) MOSFET differential pair.



**Figure 14.13.** Equivalent of common mode half circuit. (a) Bipolar device. (b) MOSFET device.

As matter of fact, Figure 14.13 is a common-emitter or common-source stage with emitter or source degeneration. Its transconductance \$G\_m\$ can be calculated by means of simple KCL and KVL, although the calculation itself is somewhat tedious. For a detailed mathematical derivation, readers can refer to the book “Analysis and Design of Analog Integrated Circuits” (Gray et al., 2001).

For a differential pair built by bipolar devices,

$$G_m = \frac{i_o}{v_{ic}} = \frac{g_m}{1 + 2g_m R_{TL}(g_m r_{\pi} + 1)}, \quad (14.86)$$

For a differential pair built by MOSFETs,

$$G_m = \frac{i_o}{v_{ic}} = \frac{g_m}{1 + 2g_m R_{TL}(g_m r_{gs} + 1)}. \quad (14.87)$$

Again, let us merge the two equations (14.86) and (14.87) into one, obtaining

$$G_m = \frac{i_o}{v_{ic}} = \frac{g_m}{1 + 2g_m R_{TL} (g_m r_{\pi gs} + 1)}, \quad (14.88)$$

where  $r_\pi$  and  $r_{gs}$  are commonly denoted by  $r_{\pi gs}$ .

On the other hand, when

$$r_o \gg R_{TL} \quad \text{or} \quad r_d \gg R_{TL}. \quad (14.89)$$

we have

$$v_{oc} = -i_o R = -G_m R v_{ic}; \quad (14.90)$$

then

$$A_{cm} = \frac{v_{oc}}{v_{ic}} = -G_m R = -\frac{g_m R}{1 + 2g_m R_{TL} \left(1 + \frac{1}{g_m r_{\pi gs}}\right)}. \quad (14.91)$$

Common mode operation in a differential pair produces DC offset, which is very harmful to the RF signal in a direct conversion communication system. In a differential pair, the common mode voltage gain  $A_{cm}$  should be as low as possible. From equation (14.91) it can be seen that the tail resistor  $R_{TL}$  is a key part in suppressing the common mode voltage gain.

If  $R_{TL}$  is zero, equation (14.91) becomes (14.85). Consequently  $A_{cm}$  will be equal to  $A_{dm}$ , and  $CMRR = 1$ . In this case, the differential pair functions as a single-ended stage.

$CMRR$  is increased if the value of the tail resistor  $R_{TL}$  is higher. However, because of the limited DC power supply voltage, the value of  $R_{TL}$  cannot be too high.

By the expression (14.78),  $CMRR$  is defined as the ratio of the differential mode voltage gain to the common mode voltage gain. From equations (14.85) and (14.91), we have

$$CMRR = \frac{A_{dm}}{A_{cm}} = 1 + 2g_m R_{TL} \left(1 + \frac{1}{g_m r_{\pi gs}}\right). \quad (14.92)$$

If the second term in the parenthesis is much smaller than 1 and can be neglected, then expression (14.92) can be approximated as

$$CMRR = \frac{A_{dm}}{A_{cm}} \approx 1 + 2g_m R_{TL}. \quad (14.93)$$

This measures the rejection capacity opposing the common mode voltage and the magnifying capability to the differential mode voltage. A good differential pair must have a high  $A_{dm}$  but low  $A_{cm}$ . From equation (14.93) it can be seen that a higher  $g_m$  of the device and a higher tail resistance  $R_{TL}$  are preferred.

### 14.2.2 CMRR in a Single-Ended Stage

It is meaningless to talk about CMRR in a single-ended block because the “differential,” or  $180^\circ$  phase difference, does not exist. However, as shown in Figure 14.14, if the single-ended input node and the ground are taken as an input differential pair, and the single-ended output node and the ground are taken as an output differential pair, what would be the CMRR?

In Figure 14.14(a) or (b), the voltage gain is obviously the same, whether the input is either  $v_{id}$  or  $v_{ic}$ , that is,

$$A_{dm} = \frac{v_{od}}{v_{id}} = \frac{v_{oc}}{v_{ic}} = A_{cm}. \quad (14.94)$$

Then,

$$\text{CMRR} = \frac{A_{dm}}{A_{cm}} = 1. \quad (14.95)$$

This means that there is no capability to reject common mode input in a single-ended stage. In other words, a single-ended stage cannot distinguish between differential mode input and common mode input.

### 14.2.3 CMRR in a Pseudo-Differential Pair

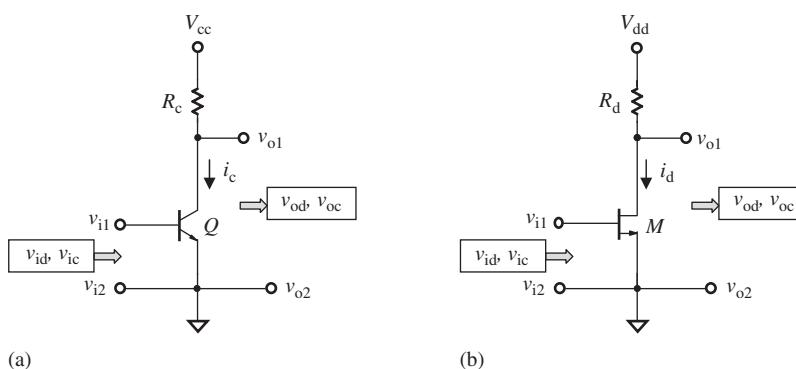
In Figure 14.15(a) or (b), the “differential pair” is a simple combination of two identical and independent single-ended stages. If their inputs are perfect differential signals, their outputs will undoubtedly be differential signals.

The two single-ended stages are identical so that their voltage gains  $A_v$  are the same, that is,

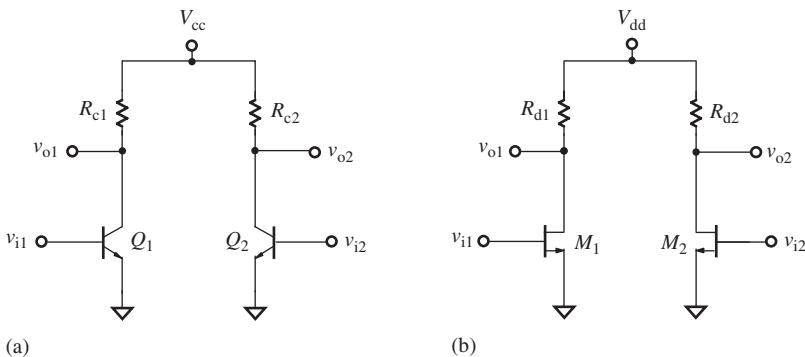
$$A_v = \frac{v_{o1}}{v_{i1}} = \frac{v_{o2}}{v_{i2}}. \quad (14.96)$$

Mathematically,

$$A_v = \frac{v_{o1}}{v_{i1}} = \frac{v_{o2}}{v_{i2}} = \frac{v_{o1} - v_{o2}}{v_{i1} - v_{i2}}. \quad (14.97)$$



**Figure 14.14.** Single-ended node and ground are hesitantly treated as a differential pair.  
(a) A bipolar single-ended block. (b) A MOSFET single-ended block.



**Figure 14.15.** A “differential pair” is a simple combination of two identical and independent single-ended stages. (a) Bipolar device. (b) MOSFET device.

If

$$v_{i1} - v_{i2} = v_{id}, \quad (14.98)$$

$$v_{o1} - v_{o2} = v_{od}, \quad (14.99)$$

then from relation (14.97) we have

$$A_v = \frac{v_{od}}{v_{id}} = A_{dm}. \quad (14.100)$$

If

$$v_{i1} - v_{i2} = v_{ic}, \quad (14.101)$$

$$v_{o1} - v_{o2} = v_{oc}, \quad (14.102)$$

then from relation (14.97) we have

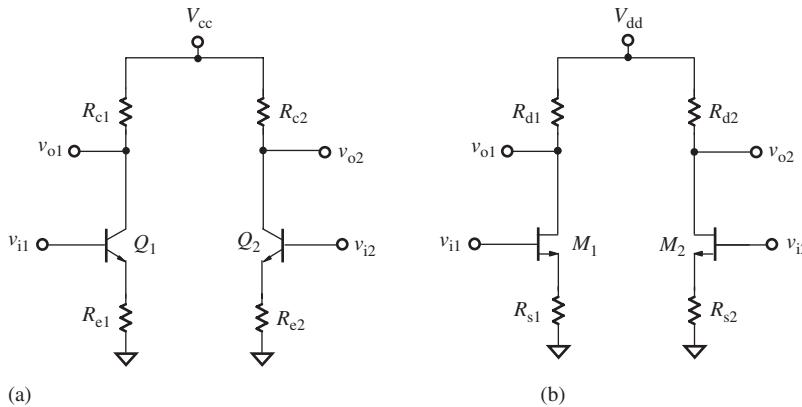
$$A_v = \frac{v_{oc}}{v_{ic}} = A_{cm}. \quad (14.103)$$

Finally, from expression (14.100) and (14.103) we have

$$\text{CMRR} = \frac{A_{dm}}{A_{cm}} = 1. \quad (14.104)$$

Just like the single-ended stage, the value of CMRR for a differential pair constructed by two identical and individual single-ended stages is the same,  $\text{CMRR} = 1$  or 0 dB. The differential pair as shown in Figure 14.15 has no capability of common mode rejection, so it is called a *pseudodifferential pair*. In both the single-ended stage and pseudodifferential pair, the common point is that there is no common coupling part, the resistor  $R_{TL}$ , which combines the two individual single-ended blocks together. This is the reason why  $\text{CMRR} = 1$  or 0 dB in expressions (14.95) and (14.104).

Another type of pseudodifferential pair is shown in Figure 14.16. There are resistors,  $R_{e1}$  and  $R_{e2}$ , connected from the emitter of the bipolar transistor to the ground in Figure 14.16(a), and  $R_{s1}$  and  $R_{s2}$ , connected from the source of the MOSFET transistor to



**Figure 14.16.** Another type of pseudodifferential pair. (a) Bipolar device. (b) MOSFET device.

the ground in Figure 14.16(b). However, there are no tail resistors. A real differential pair must have common tail resistor  $R_{TL}$  coupling the differential branches together. Without the tail resistor  $R_{TL}$ , the differential pair does not have the capability for common mode rejection, so that  $CMRR = 1$  or  $0 \text{ dB}$ .

#### 14.2.4 Enhancement of CMRR

It may be not an exaggeration to say that the CMRR is the paramount goal in the design of a differential pair. Most of the effort in the design is placed on how to enhance the value of the CMRR.

As mentioned earlier, the existence of the tail resistance  $R_{TL}$  determines whether the differential pair has the capability of common mode rejection or not. The CMRR is proportional to the value of tail resistance  $R_{TL}$ ; consequently, a direct way to enhance CMRR is to increase the value of  $R_{TL}$ . However, if the value of  $R_{TL}$  is too high, the DC voltage drop across the  $R_{TL}$  would be too great, and the DC voltage drop across the transistor would be too small. This would be harmful to the linearity of the transistor.

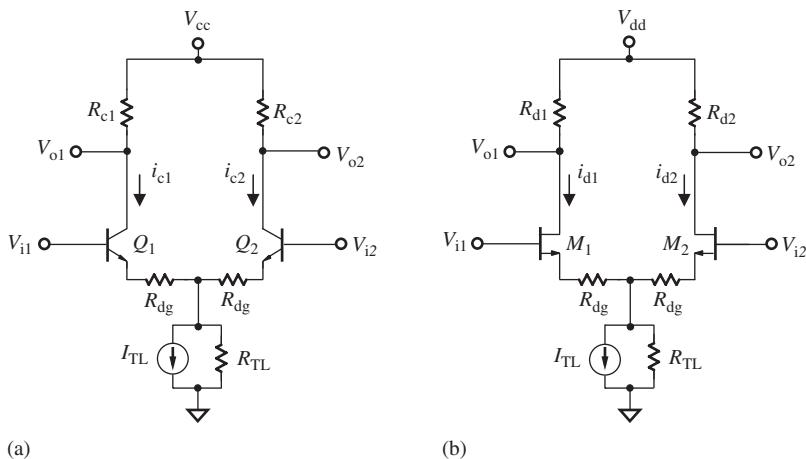
On the basis of equation (14.92) or (14.93), CMRR is proportional to the value of the transconductance of transistor  $g_m$ . Therefore, another way to enhance the CMRR is to select the transistor and adjust its operating state so that the value of  $g_m$  is increased as much as possible.

CMRR can be improved through modifying topology. For example, we can insert a degeneration resistor  $R_{dg}$  between each emitter or source and  $R_{TL}$  as shown in Figure 14.17. It can be verified that the equations (14.85), (14.91), and (14.92) are then modified to

$$A_{dm} = \frac{-g_m R}{1 + g_m R_{dg} \left( 1 + \frac{1}{g_m R_{\pi gs}} \right)}, \quad (14.105)$$

$$A_{cm} \approx \frac{-g_m R}{1 + 2g_m R_{TL} \left( 1 + \frac{1}{g_m R_{\pi gs}} \right) \left( 1 + \frac{R_{dg}}{2R_{TL}} \right)}, \quad (14.106)$$

$$CMRR \approx \frac{A_{dm}}{A_{cm}} = \frac{1 + 2g_m R_{TL} \left( 1 + \frac{1}{g_m R_{\pi gs}} \right) \left( 1 + \frac{R_{dg}}{2R_{TL}} \right)}{1 + g_m R_{dg} \left( 1 + \frac{1}{g_m R_{\pi gs}} \right)}. \quad (14.107)$$



**Figure 14.17.** Modified differential pairs by inserting  $R_{dg}$ . (a) Bipolar differential pair. (b) MOSFET differential pair.

It should be noted that in all the expressions of CMRR, the low-frequency approach has been applied so that all the capacitors in the transistors' model are neglected. In the actual engineering design, this approach must be abandoned if the operating frequency is high and the capacitors in the transistors' model cannot be neglected. However, all the expressions shown above still provide good references in RF circuit design.

## REFERENCE

Gray PR, Hurst PJ, Lewis SH, Meyer RG. Analysis and design of analog integrated circuits. 4th ed. New York, NY: John Wiley & Sons, Inc; 2001.

## FURTHER READING

- Giacoleto LJ. Differential amplifiers. New York: John Wiley & Sons, Inc; 1970.  
 Li RCH. Key issues in RF/RFIC circuit design. Beijing: Higher Education Press; 2005.  
 Middlebrook RD. Differential amplifiers. New York: John Wiley & Sons, Inc; 1963.  
 Sandoval-Ibarra F, Vargas-Calderon EJ, Herrera-Barragan JA, et al. Designing a CMOS differential pair. Proceedings of the 15th International Conference on Electronics, Communications, and Computers (CONIELECOMP 2005). Washington, DC: IEEE; 2005.  
 Shaker MO, Mahmoud SA, Soliman AM. New CMOS fully differential transconductor and its application. ISCAS 2006; Greece: IEEE; 2006. pp 77–80.

## EXERCISES

1. What is the basic topology and the definition of a differential pair?
2. What is the transfer characteristic of a bipolar differential pair?
3. What is the range of  $v_{id}$  such that the bipolar differential pair is approximately a linear device?

4. What are the transfer characteristics of a MOSFET differential pair?
5. What is the range of  $v_{id}$  such that the MOSFET differential pair is approximately a linear device?
6. What does CMRR mean? Discuss the cases with different CMRR values.
7. Prove  $\text{CMRR} = A_{\text{dm}}/A_{\text{cm}} = 1$  in a single-ended stage.
8. A pseudodifferential pair consists of two identical single-ended stages, which are simply tied together in parallel. Illustrate  $\text{CMRR} = A_{\text{dm}}/A_{\text{cm}} = 1$  in such a pseudodifferential pair.
9. Describe one or two ways to improve the CMRR of a differential pair.
10. With respect to the single-ended stage, does the differential pair have a higher capability to cancel some special interference?
11. With respect to the single-ended stage, does the differential pair have the capability to cancel “common mode noise”?
12. In the layout for differential circuitry, two long adjacent lines in parallel are very often provided in the input or output terminals. Please comment on their impact on circuit performance.

## ANSWERS

1. The basic topology of a differential pair can be sketched as Figure 14.P.1 and the definition of the differential pair can be outlined as follows:

$$\begin{aligned} V_{i1} &= V_{II} + v_{i1} \\ V_{i2} &= V_{I2} + v_{i2} \end{aligned}$$

$$\begin{aligned} V_{o1} &= V_{O1} + v_{o1} \\ V_{o2} &= V_{O2} + v_{o2} \end{aligned}$$

$$V_{II} = V_{I2}$$

$$V_{O1} = V_{O2}$$

$$v_{i1} = -v_{i2}$$

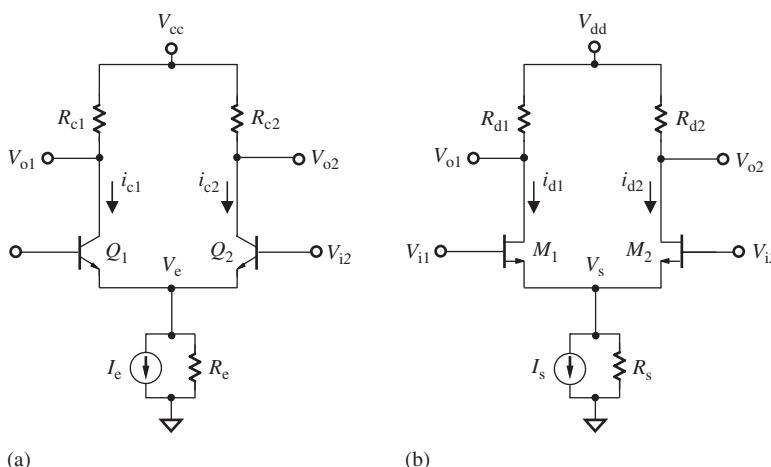
$$v_{o1} = -v_{o2}$$

$$|v_{i1}| = |v_{i2}|$$

$$|v_{o1}| = |v_{o2}|$$

$$\angle v_{i1} - \angle v_{i2} = 180^\circ \text{ or } -180^\circ$$

$$\angle v_{o1} - \angle v_{o2} = 180^\circ \text{ or } -180^\circ.$$



**Figure 14.P.1.** Typical differential pairs. (a) A bipolar differential pair. (b) A MOSFET differential pair

2. The transfer characteristic of a bipolar differential pair is

$$V_{\text{od}} = V_{\text{o1}} - V_{\text{o2}} = \alpha_F I_{\text{TL}} R_{\text{c}} \tanh \left( \frac{-V_{\text{id}}}{2V_{\text{T}}} \right).$$

3. The range of  $V_{\text{id}}$  that the bipolar differential pair is a linear device approximately is

$$-V_{\text{T}} < V_{\text{id}} < V_{\text{T}}.$$

4. The transfer characteristics of a MOSFET differential pair is

$$v_{\text{od}} = -R_{\text{d}} \frac{\mu_n C_{\text{ox}}}{2} \frac{W}{L} (1 + \lambda V_{\text{ds}}) v_{\text{id}} \sqrt{\frac{4I_{\text{TL}}}{\frac{\mu_n C_{\text{ox}}}{2} \frac{W}{L} (1 + \lambda V_{\text{ds}})} - v_{\text{id}}^2}$$

5. From the expression of  $v_{\text{od}}$ , it can be seen that the linear range of  $v_{\text{id}}$  for MOSFET approximately is

$$v_{\text{id}}^2 < \frac{4I_{\text{TL}}}{\frac{\mu_n C_{\text{ox}}}{2} \frac{W}{L} (1 + \lambda V_{\text{ds}})} = \frac{4I_{\text{TL}}}{I_{\text{d}}} (V_{\text{gs}} - V_{\text{th}})^2$$

or

$$v_{\text{id}} < 2 \sqrt{\frac{I_{\text{TL}}}{I_{\text{d}}} (V_{\text{gs}} - V_{\text{th}})}$$

because

$$I_{\text{d}} = \frac{\mu_n C_{\text{ox}}}{2} \frac{W}{L} (V_{\text{gs}} - V_{\text{th}})^2 (1 + \lambda V_{\text{ds}}).$$

6. CMRR is the ratio of differential mode gain to common mode gain.

$$\text{CMRR} = \frac{A_{\text{dm}}}{A_{\text{cm}}}$$

There are three cases:

- (a) If  $\text{CMRR} < 1$ , or  $\text{CMRR} < 0 \text{ dB}$ , it signifies that the differential pair magnifies the input common portion more than it magnifies the input differential portion. The common mode operation rejects the differential mode operation.
- (b) If  $\text{CMRR} = 1$ , or  $\text{CMRR} = 0 \text{ dB}$ , it signifies that the differential pair magnifies the differential portion with the same number of times as it magnifies the common portion. The differential pair operates just like a single-ended stage.
- (c) If  $\text{CMRR} > 1$ , or  $\text{CMRR} > 0 \text{ dB}$ , it signifies that the differential pair magnifies the input differential portion more than it magnifies the input common portion. The differential mode operation rejects the common mode operation.

7. To prove  $\text{CMRR} = A_{\text{dm}}/A_{\text{cm}} = 1$  in a single-ended stage:

$$v_{\text{i2}} = v_{\text{o2}} = 0$$

$$V_{\text{id}} = v_{\text{i1}}$$

$$v_{ic} = v_{i1}/2$$

$$v_{od} = v_{o1}$$

$$v_{oc} = v_{o1}/s$$

and

$$v_{o1} = A_{11}v_{i1}$$

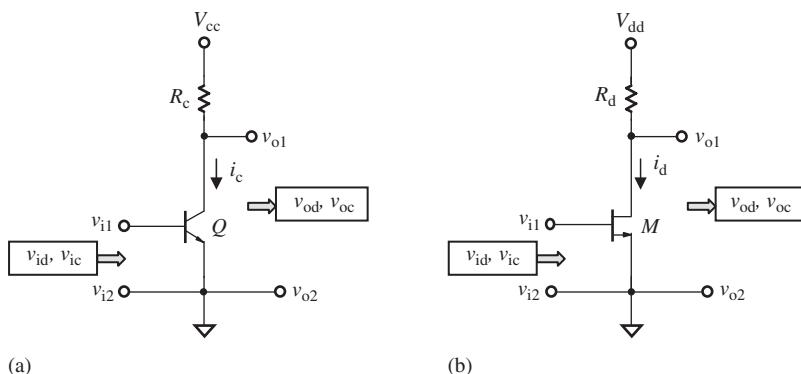
$$0 = A_{21}v_{i1}$$

or

$$v_{od} = A_{dm}v_{id} + A_{cm-dm}v_{id}/2$$

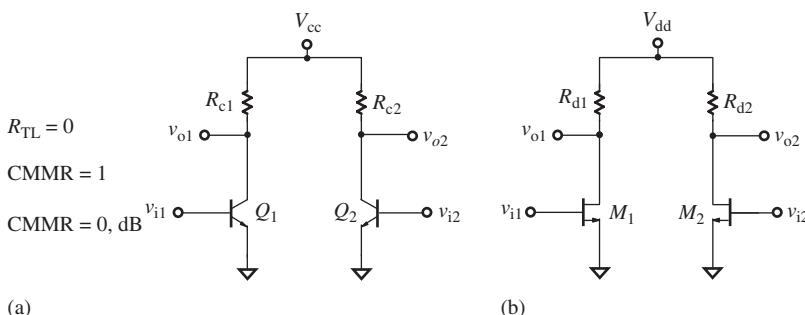
$$v_{od} = A_{dm-cm}v_{id} + A_{cm}v_{id}/2$$

$$\text{CMRR} = A_{dm}/A_{cm} = 1$$

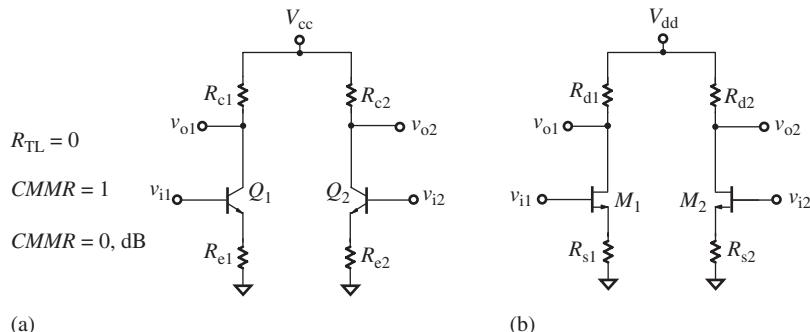


**Figure 14.P.2.** Single-ended node and ground are hesitantly treated as differential pairs. (a) A bipolar single-ended block. (b) A MOSFET single-ended block.

8. A pseudodifferential pair consists of two identical single-ended stages, which are simply tied together in parallel.  $\text{CMRR} = A_{dm}/A_{cm} = 1$  in such a pseudodifferential pair.



**Figure 14.P.3.** Pseudodifferential pair. (a) Bipolar device; \$R\_{TL} = 0, R\_{e1} = 0, R\_{e2} = 0\$. (b) MOSFET device; \$R\_{TL} = 0, R\_{s1} = 0, R\_{s2} = 0\$.



**Figure 14.P.4.** Another type of pseudodifferential pair. (a) Bipolar device;  $R_{TL} = 0$ ,  $R_{e1} \neq 0$ ,  $R_{e2} \neq 0$ . (b) MOSFET device;  $R_{TL} = 0$ ,  $R_{S1} \neq 0$ ,  $R_{S2} \neq 0$ .

9. One way to improve the CMRR of a differential pair is to add a small resistor between the source and the tail resistor; another way is to apply a good transformer between two sources and the ground. Its turn ratio is 1:1. This transformer would force two branches of the differential pair to operate in a good symmetrical way.
  10. In respect to the single-ended stage, the differential pair has higher capability to cancel some special interference, such as when the interference source is located at the central position which is a symmetrical axis of the differential pair.
  11. In respect to the single-ended stage, the differential pair is impossible to cancel so-called common mode noise. As a matter of fact, distinguishing the noise into common mode noise and differential mode noise is a misunderstanding of the noise.
  12. One can layout differential circuitry with two long adjacent lines in parallel at the input or output terminals. However, the impedance between input and output must be kept as identical as possible, and the attenuation due to the long line must be negligible.

# RF BALUN

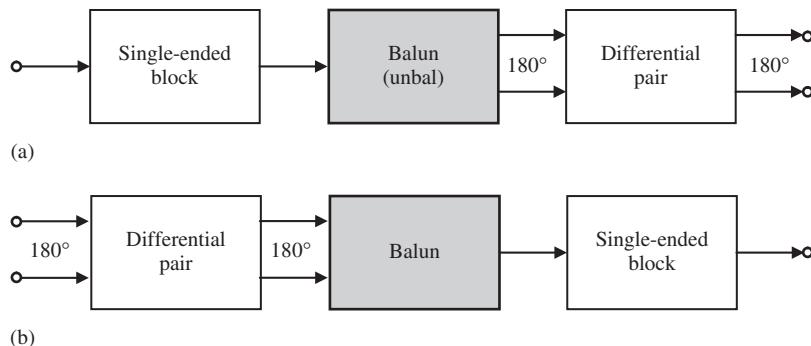
## 15.1 INTRODUCTION

A balun is a transformation block between a single-ended stage and a differential pair.

A balun transforming a signal from a single-ended stage to a differential pair splits a single-ended signal into a pair of differential signals with the same magnitude but  $180^\circ$  phase difference. It is called *splitter* as shown in Figure 15.1(a). A differential pair plays some special functions in the transport or manipulation of a signal. For instance, the capability of common mode rejection potentially exists in a differential pair but not in a single-ended stage.

A balun transforming a signal from a differential pair to a single-ended stage combines a pair of differential signals with the same magnitude but  $180^\circ$  phase difference as a single-ended signal. It is called *combiner* as shown in Figure 15.1(b). The conversion from a differential pair to a single-ended stage simplifies both simulation and testing of a block.

The word “balun” is a portmanteau formed from the words “balanced” and “unbalanced,” in which balanced implies a differential configuration and unbalanced represents a single-ended configuration. It would seem reasonable to rename the balun in Figure 15.1(a) as “Unbal.” Sometimes, people refer to the balun type in Figure 15.1(a) as a “splitter” and the balun type in Figure 15.1(b) as a “combiner.” The splitter is a reversed combiner, and vice versa. From now on, we will only focus on the splitter unless otherwise specified.



**Figure 15.1.** A balun, which is a block between a single-ended block and a differential pair.  
 (a) From single-ended block to differential pair. (b) From differential pair to single-ended block.

The balun is usually implemented by passive parts and, if so, can be called a *passive balun*. However, active baluns are also available, in which the block contains both active and passive parts. In this chapter, our discussion is confined to passive baluns.

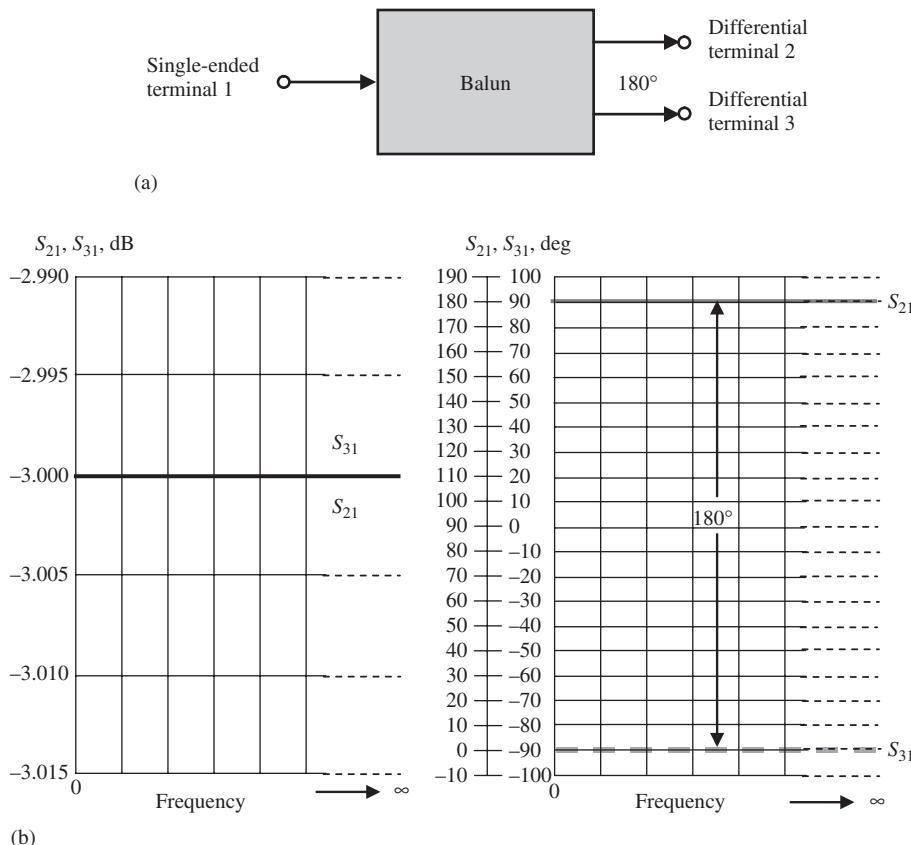
Figure 15.2 represents the main features of an ideal balun, which can be outlined as follows:

1. There are three terminals: a single-ended terminal and two differential terminals. Typically, the impedance looking into the terminal is  $50\ \Omega$  at all three terminals. However, other values are also possible, in which the impedances of the single-ended terminal and the two differential terminals differ; however, the two differential terminals' impedances must be kept equal.
2. At the two differential terminals, the magnitudes of the signal are equal but their phases are kept at  $180^\circ$  difference. The absolute value of the phase at the three terminals depends on the type of balun and the reference ground point, GND, in the circuit block. Two phase ordinates are provided in Figure 15.2(b) for different phase scales.
3. Total insertion loss from the single-ended terminal to the differential terminals is zero. This implies that the power of the signal at each differential terminal is 3 dB lower than that of the single-ended terminal because the signal power at the single-ended terminal is the sum of the signal powers at the two differential terminals. In reality, in a splitter the signal power at each differential terminal is more than 3 dB lower than the signal power at the single-ended terminal because the insertion loss is never zero in a practical balun.

A variety of RF baluns have been developed in past decades. In this chapter, three types of baluns are emphasized: the transformer, LC, and microstrip line.

In the simulation stage of RF circuit design, the ideal transformer balun is a good candidate for a transformation block between single-ended and differential pair blocks. The greatest advantage of the transformer balun is its frequency response with an almost infinite bandwidth. In addition, the insertion loss of the balun can be set to zero since it is an ideal balun.

In the testing stage, the LC and microstrip line baluns are recommended. The greatest advantage of an LC balun is its simplicity: it can be implemented in a laboratory quickly and easily. Surprisingly, instead of a narrowband response as might be expected, the



**Figure 15.2.** Insertion loss and phase shift of an ideal balun. (a) Terminals of a balun. (b) Insertion loss of an ideal balun.

LC balun behaves with a reasonably wideband frequency response; this behavior will be explained in the corresponding sections.

Compared with the LC balun, the frequency response in a microstrip line balun is wider. In addition, its outstanding advantage is its low cost, since its basic parts are microstrip lines.

As matter of fact, these three types of balun have become popular in RF circuit design, even though other types can also be found in practical applications. In Section 15.5, mixing type baluns will be introduced, while another two types of baluns specially designed for the PA (power amplifier) will be presented in Chapter 21.

## 15.2 TRANSFORMER BALUN

The distinguishing features of the transformer balun are as follows:

1. An ideal transformer balun has an infinitely wide frequency response and zero insertion loss. In reality, the ideal transformer balun is never realized; however, with respect to other types of baluns, a transformer balun has a wider frequency

range and a lower insertion loss if the operating frequency is lower than about 1 GHz.

2. The impedance matching can be adjusted by its turn ratio. If the impedances of the input source and the output load are purely resistive without any reactive portions, a transformer balun can function as an impedance matching network with a simple adjustment of its turn ratio. The task of impedance matching can therefore be avoided by the use of a transformer balun. This is a considerable benefit because, in RF circuit design, the task of impedance matching is usually a “must” and can sometimes be complicated.

If the impedance of the input source and the output load are not purely resistive, applying a transformer balun still eases the task of impedance matching. By adjusting its turn ratio, the real portion of the original impedance, at least, can be easily matched. The reactances of the impedances in either the single-ended or differential sides can be interpreted from each other. This makes the impedance matching work much easier. The interpretation of reactance between the single-ended and differential pair will be discussed in subsequent sections of this chapter.

### 15.2.1 Transformer Balun in RF Circuit Design with Discrete Parts

In past years, the ferrite transformer balun has been widely applied in communication systems, such as VHF or UHF radios, and even in radios with operating frequency ranges of around 800–900 MHz. Its cost is acceptable and its size has been reduced to the order of a few millimeters.

Unfortunately, the ferrite transformer balun is restricted by its upper frequency limit. If the operating frequency becomes higher than about 1 GHz, it simply goes out of order.

Instead, other types of baluns have been developed for application in higher RF ranges above 1 GHz of communication and other systems.

### 15.2.2 Transformer Balun in RFIC Design

In RFIC design, one must seriously consider which kind of transformer balun is to be applied in the circuit implementation. The key point is to focus on whether an off-chip or on-chip transformer balun is chosen.

There are many types of off-chip and on-chip transformer baluns. The ferrite transformer is popular as an off-chip transformer balun, while the spiral-coil transformer is popular as an on-chip transformer.

The merit of a ferrite transformer as an off-chip discrete part is the higher  $Q$  value of the coil. However, its demerits are very obvious: higher cost, larger size, and the upper limit of the operating frequency, which is about 1 GHz.

On the other hand, the merits of the spiral transformer on the RFIC chip are also very obvious: lower cost, smaller size, and a much higher upper limit of operating frequency. Its chief demerit is its extremely low  $Q$  value.

Among all the merits and demerits, cost is inevitably the first priority to be considered in a circuit design. Consequently, from the viewpoint of cost, it is preferable to build the transformer balun on the RFIC chip, despite its low  $Q$  value.

From the viewpoint of IC packaging, the off-chip transformer balun is likewise unwelcome, as one off-chip transformer balun requires at least three additional pins on the IC package. Neither is it welcome from the viewpoint of performance, because

the bonding wires and pads for a balun's inputs and outputs bring about uncertainty of performance and additional attenuation. These two considerations also encourage building the transformer on the RFIC chip despite its low  $Q$  value.

In summary, the following decision must be made:

- The on-chip transformer balun should be adapted in an RFIC design if its low  $Q$  value is acceptable.
- Otherwise, the off-chip transformer balun should be adapted in an RFIC design if a high  $Q$  value is required.

### 15.2.3 An Ideal Transformer Balun for Simulation

An ideal transformer balun for simulation is suggested as shown in Figure 15.3.

The basic configuration of the suggested transformer balun is to combine two ideal, identical transformers stacked together as the core of the resultant transformer balun. At the single-ended portion, the two primary windings of the ideal transformers are connected together in series; at the connected point, there is a resistor with a value of  $1 \text{ G } \Omega$ , which is meaningless in practical performance but is merely a dummy part to satisfy computer requirements for a DC grounding path in the simulation. At the differential portion, the two secondary windings of the two ideal transformers are also connected in series and their connected point is the central grounded point of the balun.

It should be specially noted that the turns ratio of each transformer is assigned as

$$\text{Turn ratio} = 1 : n = 1 : \sqrt{2}, \quad (15.1)$$

where

$n$  = the turns ratio of secondary winding of the transformer with respect to the primary winding.

It should also be noted that the turns ratio remains unchanged before and after the two transformers are stacked together to form a transformer balun. In the resultant transformer balun, the total impedance of the differential pair is twice that of the single-ended side, that is,

$$Z_{L,T} = 2Z_L, \quad (15.2)$$

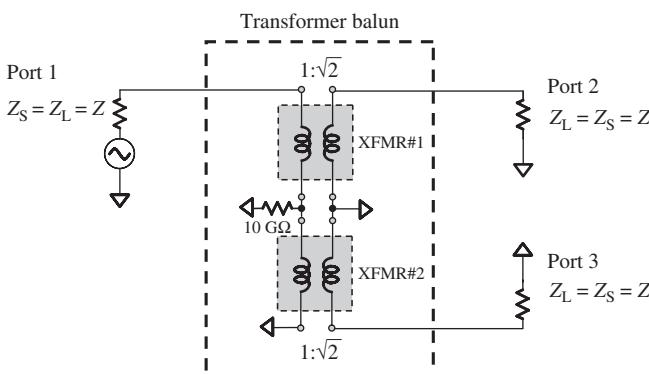


Figure 15.3. A transformer balun built by two ideal and identical transformers with turns ratio =  $1 : \sqrt{2}$  and  $Z_S = Z_L = Z$ .

where

$Z_{L,T}$  = the total load impedance from port 2 to port 3 in Figure 15.3 and  
 $Z_L$  = the individual load impedance at port 2 or port 3 in Figure 15.3.

It is well known that the relationship between the turns ratio and the impedances of source and load in the resultant transformer balun is

$$n = \sqrt{\frac{Z_{L,T}}{Z_S}} = \sqrt{\frac{2Z_L}{Z_S}} = \sqrt{2}. \quad (15.3)$$

Then we have

$$Z_L = Z_S = Z. \quad (15.4)$$

We remove the subscript “L” or “S” since  $Z_L = Z_S$ . This is why the impedance at each port in Figure 15.3 is symbolized with “ $Z_L = Z_S = Z$ .”

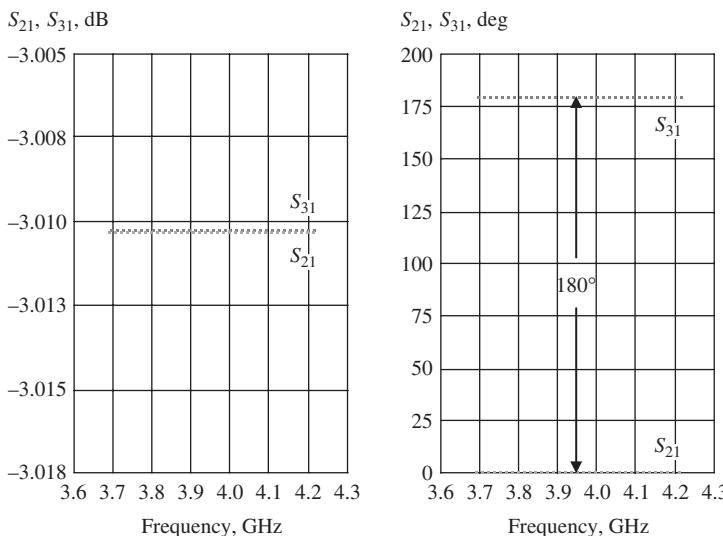
Now let us explore the features of this ideal transformer balun in terms of the following two simulations.

SIMULATION A. Simulation of the ideal transformer balun operating in the frequency range of a group 1 UWB system:  $f = 3696\text{--}3960\text{--}4224$  MHz,  $Z_S = Z_L = 50 \Omega$ .

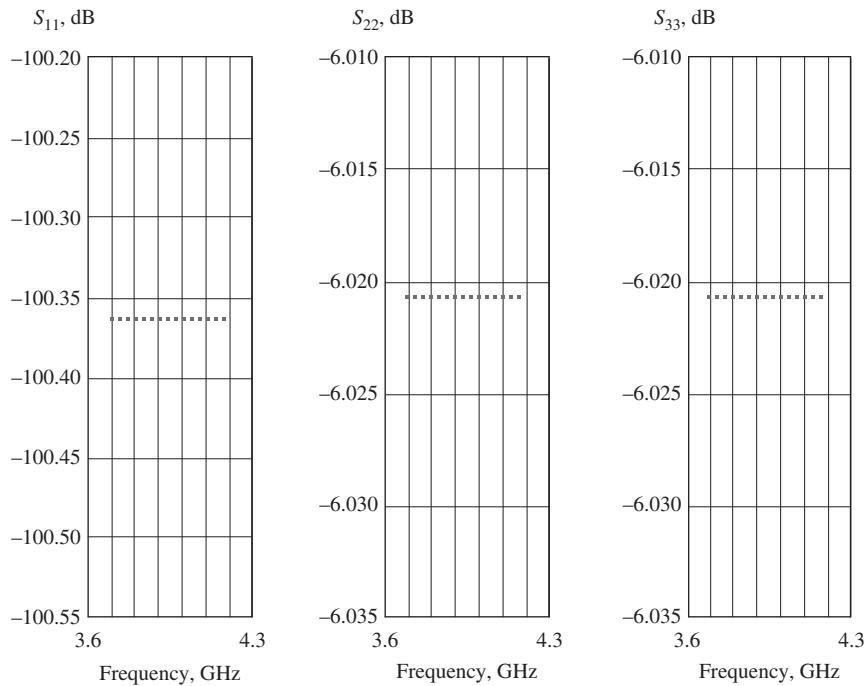
Figures 15.4 and 15.5 display the simulation results.

Figure 15.4 shows the insertion loss and phase shift while Figure 15.5 shows the return loss of the ideal transformer balun operating in the frequency range of a group 1 UWB system. Table 15.1 lists the values of the relevant  $S$ -parameters of the performance.

As expected, the frequency responses of the  $S$ -parameters are quite flat since the two stacked transformers are ideal. Unlike  $S_{11}$ , the magnitudes of the return loss  $S_{22}$  and  $S_{33}$  are not lowered because of the existence of three ports rather than two. However, this imperfection does not impact the simulation at all.



**Figure 15.4.** Insertion loss and phase shift of the ideal transformer balun operated in the frequency range of a group 1 UWB system:  
 $f = 3696\text{--}3960\text{--}4224$  MHz,  
 $Z_S = Z_L = 50 \Omega$ .



**Figure 15.5.** Return loss of the ideal transformer balun operated in the frequency range of a group 1 UWB system:  $f = 3696\text{--}3960\text{--}4224$  MHz,  $Z_S = Z_L = 50 \Omega$ .

**TABLE 15.1.** Values of Relevant  $S$ -Parameters of the Ideal Transformer Balun Operating in the Frequency Range of a group 1 UWB System:  $f = 3696\text{--}3960\text{--}4224$  MHz,  $Z_S = Z_L = 50 \Omega$

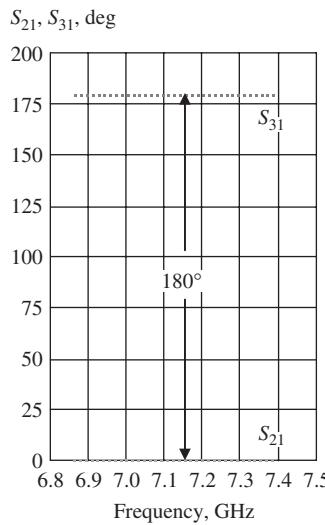
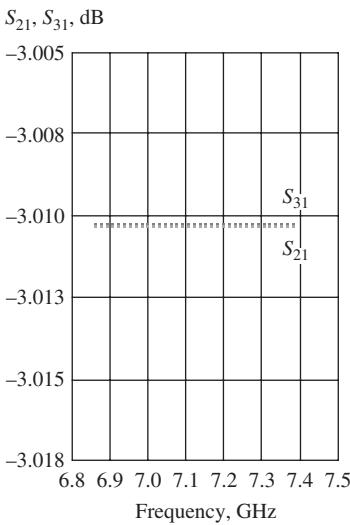
Frequencies	$f$	3696–3960–4224		MHz
Bandwidth	$\Delta f$		528	MHz
Insertion loss	$S_{21}$	Magnitude	$-3.01 \pm 0.01$	dB
		Phase	$0^\circ \pm 0.1^\circ$	—
	$S_{31}$	Magnitude	$-3.01 \pm 0.01$	dB
		Phase	$+180^\circ \pm 0.1^\circ$	—
Return loss	$S_{11}$	Magnitude	$-100.36 \pm 0.01$	dB
	$S_{22}$	Magnitude	$-6.02 \pm 0.01$	dB
	$S_{33}$	Magnitude	$-6.02 \pm 0.01$	dB

**SIMULATION B.** Simulation of the ideal transformer balun operated in the frequency range of a group 3 UWB system:  $f = 6864\text{--}7128\text{--}7392$  MHz,  $Z_S = Z_L = 50 \Omega$ .

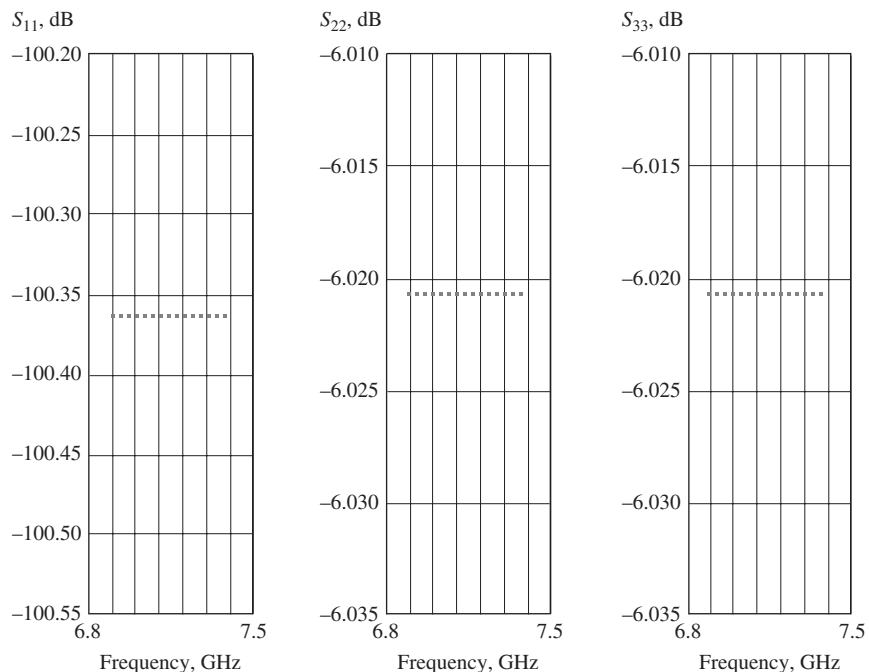
Figures 15.6 and 15.7 display the simulated results.

Figure 15.6 shows the insertion loss and phase shift, while Figure 15.7 shows the return loss of the ideal transformer balun operating in the frequency range of a group 3 UWB system. Table 15.2 lists the values of the relevant  $S$ -parameters of the performance.

Again, as expected, the frequency responses of the  $S$ -parameters are quite flat since the two stacked transformers are ideal. The magnitudes of the return loss  $S_{22}$  and  $S_{33}$  are



**Figure 15.6.** Insertion loss and phase shift of the ideal transformer balun operated in the frequency range of a group 3 UWB system:  $f = 6864\text{--}7128\text{--}7392$  MHz,  $Z_S = Z_L = 50 \Omega$ .



**Figure 15.7.** Return loss of the ideal transformer balun operated in the frequency range of a group 3 UWB system:  $f = 6864\text{--}7128\text{--}7392$  MHz,  $Z_S = Z_L = 50 \Omega$ .

similarly not lowered like  $S_{11}$  because of the existence of three ports rather than two. However, this imperfection does not impact the simulation at all.

In summary, a balun with the same impedance value at all its terminals, including the one single-ended and two differential terminals, is preferred because the impedance

TABLE 15.2. Values of Relevant S-Parameters of the Ideal Transformer Balun Operating in the Frequency Range of a Group 3 UWB System:  $f = 6864\text{--}7128\text{--}7392 \text{ MHz}$ ,  $Z_S = Z_L = 50 \Omega$

Frequencies	$f$	6864–7128–7392		MHz
Bandwidth	$\Delta f$		528	MHz
Insertion loss	$S_{21}$	Magnitude	$-3.01 \pm 0.01$	dB
		Phase	$0^\circ \pm 0.1^\circ$	—
	$S_{31}$	Magnitude	$-3.01 \pm 0.01$	dB
		Phase	$+180^\circ \pm 0.1^\circ$	—
Return loss	$S_{11}$	Magnitude	$-100.36 \pm 0.01$	dB
	$S_{22}$	Magnitude	$-6.02 \pm 0.01$	dB
	$S_{33}$	Magnitude	$-6.02 \pm 0.01$	dB

matching for the differential pair can be replaced by the impedance matching for a single-ended stage. In other words, this circuitry with a differential configuration can be treated as one with a single-ended stage. This, of course, simplifies the simulation task significantly.

An ideal transformer balun is never found in reality. However, this is nevertheless a powerful tool in the simulation stage because of its huge advantage in the simulation work. After the simulation is done, the ideal transformer balun, of course, is replaced by a practical balun in the product.

#### 15.2.4 Equivalence of Parts between Single-Ended and Differential Pair in Respect to an Ideal Transformer Balun

The ideal transformer balun built by two ideal and identical transformers as shown in Figure 15.3 has very special features. One feature is the same impedance appearing in the single-ended port and two differential pair ports, that is,  $Z_L = Z_S = Z$ , if the turns ratio of each transformer is  $1 : \sqrt{2}$ . Another special feature is that the equivalence of parts between the single-ended and differential pair enables these parts to be interpreted by each other.

What does “interpretation” of parts between the single-ended and the differential pair mean? If part A is added to the single-ended port and part B is added to each port of the differential pair simultaneously, then at the operating frequency  $\omega_0$  or  $f_0$  and the S-parameters or impedances at all three ports can be kept unchanged before and after these parts are added, as long as part B is “interpreted” from part A or part A is interpreted from part B based on the following special rules.

- A capacitor in parallel at the single-ended port is interpreted as an inductor in parallel at each port of differential pair, or vice versa.
- A capacitor in series at the single-ended port is interpreted as an inductor in series at each port of differential pair, or vice versa.
- An inductor in parallel at the single-ended port is interpreted as a capacitor in parallel at each port of differential pair, or vice versa.
- An inductor in series at the single-ended port is interpreted as a capacitor in series at each port of differential pair, or vice versa.

- The reactance of the inductor must be equal to the negative reactance of capacitor, that is

$$L\omega_0 = \frac{1}{C\omega_0}, \quad (15.5)$$

or

$$LC = \frac{1}{\omega_0^2}, \quad (15.6)$$

where  $\omega_0$  = the operating angular frequency.

It should be noted that the inductor and capacitor mentioned in these special rules are added to different sides of the transformer balun. In other words, if the capacitor is added to the single-ended side, then the inductor is added to the differential pair side, or vice versa.

Figure 15.8 shows four cases when a part is added to the single-ended port and the interpreted part is added to each port of the differential pair, respectively. Formula (15.5) is the relationship between the added parts, by which the added parts can be interpreted from each other. Simulations do prove that the  $S$ -parameters or impedances at all three ports are unchanged at the operating frequency  $\omega_0$  or  $f_0$ , and are nearly unchanged or only slightly changed within certain bandwidths around the operating frequency, before and after the parts are added to the transformer balun.

This is a very useful regulation! In terms of this regulation, simulation for a circuit with a differential pair configuration could be replaced by simulation for a circuit with a single-ended configuration!

In order to verify the interpretation regulation and to check the equivalence of impedances looking into and outward from the transformer balun, let us conduct another seven experimental simulations. The first five simulations are conducted for the verification of the interpretation regulation. The added part number will be incremented from zero to four in steps so that they are not the same as shown in Figure 15.8, where only one part was added to either the single-ended port or each port of differential pair. The last two simulations are conducted to check the equivalence of the impedances looking into and outward from the transformer balun. In all the simulations, the operating frequency is  $f_0 = 7128$  MHz, which is the central frequency of a group 3 UWB system. As matter of fact, we are also interested in seeing what would happen in the entire frequency bandwidth of group 3, UWB system,  $f = 6864$ – $7392$  MHz, with a central frequency of  $f_0 = 7128$  MHz.

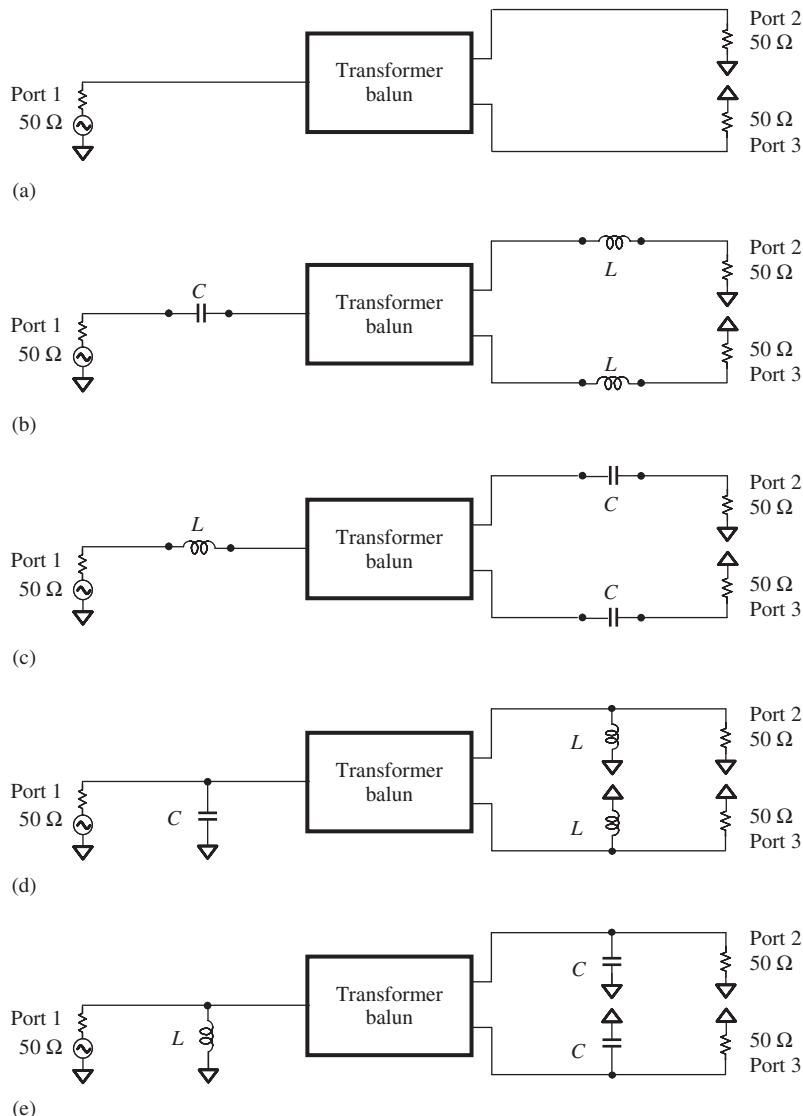
Simulation 1 is conducted for the original transformer balun without any parts added.

The setup is shown in Figure 15.3 and its simulation results are presented with Figures 15.6 and 15.7. However, for the integrity of description, it will be partially repeated in Figures 15.9–15.11.

Figure 15.9 is the simulation setup with the transformer balun as shown in Figure 15.3. Figures 15.10 and 15.11 show the insertion loss, phase shift, and return loss respectively.

The following can be found at the operating frequency  $f_0 = 7128$  MHz.

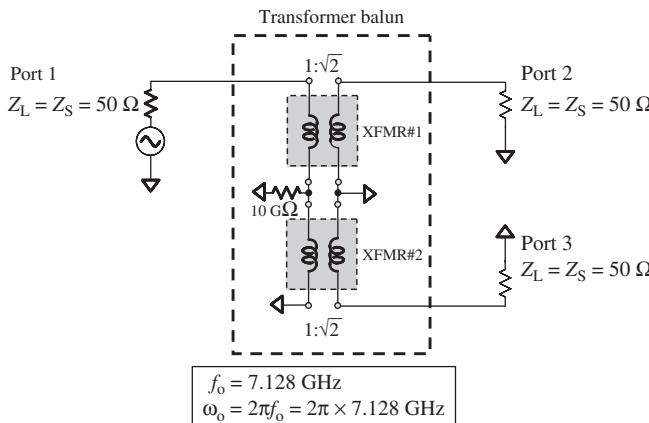
- Instead of the expected  $-3$  dB, the insertion losses  $S_{21}$  and  $S_{31}$  are both  $-3.01$  dB. This indicates that the attenuation of the transformer balun is  $0.01$  dB, which



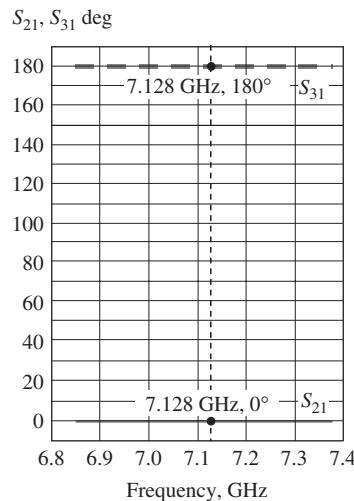
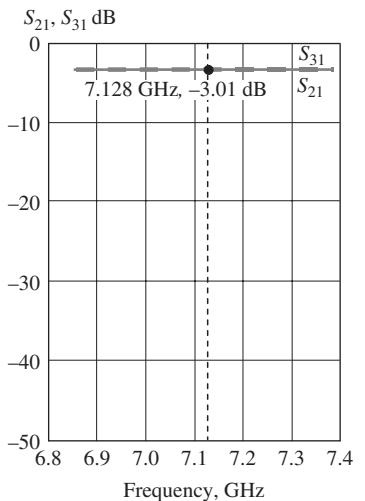
**Figure 15.8** Four ways to add parts to a transformer balun, in which the  $S$ -parameters or impedances at all three ports as shown in this figure are kept unchanged before and after these parts are added (note that  $LC = 1/\omega_0^2$ ). (a) Original transformer balun. (b) A capacitor is added to the single-ended side in series. (c) An inductor is added to the single-ended side in series. (d) A capacitor is added to the single-ended side in parallel. (e) An inductor is added to the single-ended side in parallel.

in an ideal transformer balun should be 0 dB. The difference of 0.01 dB is due to the error of truncated decimal numbers in the digital computation.

- The phase of  $S_{21}$  is  $0^\circ$  and the phase of  $S_{31}$  is  $180^\circ$ , so that the phase difference between  $S_{21}$  and  $S_{31}$  is  $180^\circ$ . This indicates that ports 2 and 3 are a real differential pair.



**Figure 15.9.** Simulation 1: transformer balun without any added part.



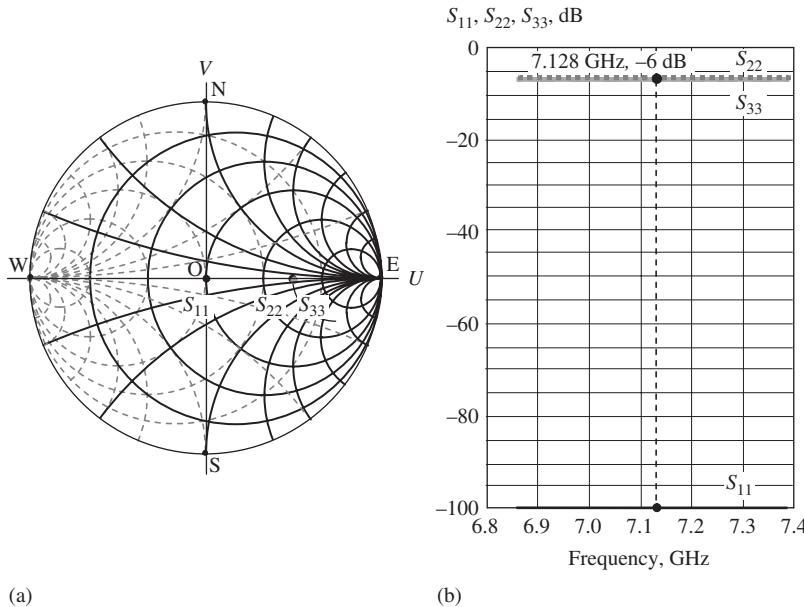
**Figure 15.10.** Insertion loss and phase shift of transformer balun for a group 3 UWB system ( $S_{21}, S_{31}$ ) in simulation 1: transformer balun without any added part.

- The magnitude of the return loss  $S_{11}$  at port 1 is dropped down to  $-100$  dB, which in an ideal transformer balun should be  $-\infty$  [dB]. Again, the difference is due to the error of truncated decimal numbers in the digital computation. Nevertheless, this indicates that port 1 is well matched to  $50 \Omega$ . At ports 2 and 3,  $S_{22} = S_{33} = -6$  dB. Because of the existence of three rather than two ports, they are not dropped down as is  $S_{11}$ . Regardless, they can be ignored because they do not play any negative role to the simulated parameters.

At other frequencies within the frequency bandwidth  $f = 6864$ – $7392$  MHz, the following observations can be made:

- The frequency responses of all the  $S$ -parameters are quite flat within the frequency bandwidth. This is due to the infinite bandwidth response resulting from the ideal transformers. Consequently, the performance in all frequencies within the bandwidth is almost the same as that at the operating frequency  $f_o = 7128$  MHz.

Simulation 2 is conducted for the transformer balun with one part added.



**Figure 15.11.** Return loss of transformer balun for group 3 UWB system ( $S_{11}, S_{22}, S_{33}$ ) in simulation 1: transformer balun without any added part. (a) On Smith Chart. (b) By rectangular coordination [dB].

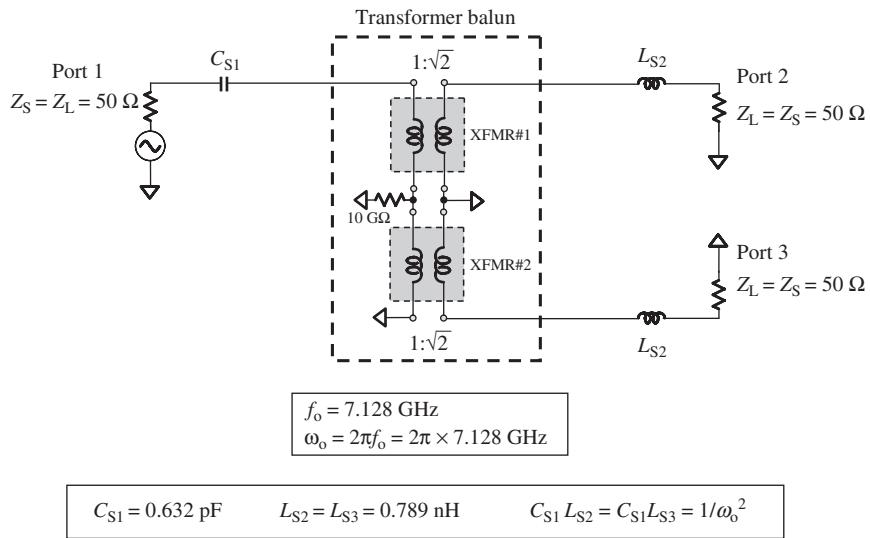
Figure 15.12 is the simulation setup of the transformer balun. Figures 15.13 and 15.14 show the insertion loss, phase shift, and return loss, respectively.

It can be found that, at the operating frequency  $f_o = 7128$  MHz, there is no large difference in performance between simulations 1 and 2, except that

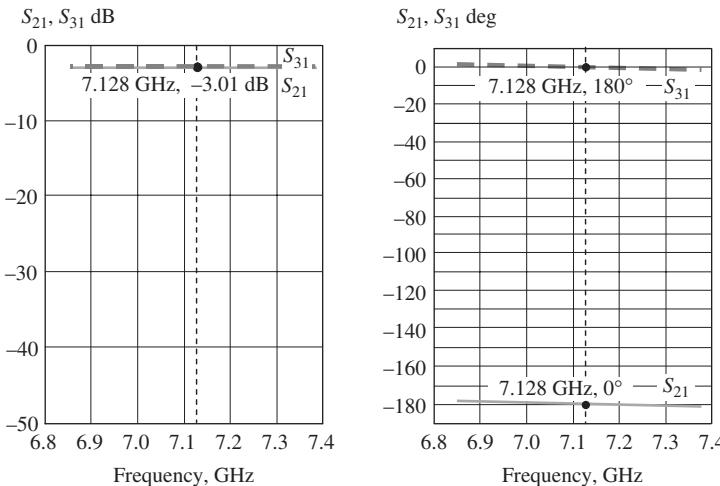
- the phase of  $S_{21}$  is  $-180^\circ$  and the phase of  $S_{31}$  is  $0^\circ$ , but the phase difference between  $S_{21}$  and  $S_{31}$  is still  $180^\circ$ . This indicates that ports 2 and 3 are a real differential pair;
- the magnitude of the return loss  $S_{11}$  at port 1 is dropped down to  $-90$  dB, which should be  $-\infty$  in an ideal transformer balun. Again, this difference is due to the error of truncated decimal numbers in the digital computation. Nevertheless, it indicates that port 1 is well matched to  $50\ \Omega$ .

At other frequencies within the frequency bandwidth  $f = 6864$ – $7392$  MHz

- the frequency responses of  $S_{21}$  and  $S_{31}$  are not entirely flat, but tilted a little bit, since the capacitor  $C_{S1}$  or the inductor  $L_{S2}$  is dependent on the frequency. The tilted magnitude is less than  $0.5$  dB and the phase shift is less than  $5^\circ$ . Interestingly, the tilted directions of the phase of  $S_{21}$  and  $S_{31}$  are the same so that the phase difference of the two differential ports is kept at around  $180^\circ$ ;
- the magnitude of the return loss at port 1,  $S_{11}$ , varies considerably within the frequency band. At low and high frequencies, the values of  $S_{11}$  are around  $-40$  dB, while at the central frequency the value of  $S_{11}$  is dropped down to less than  $-96$  dB. The reason for this is the same as that of the previous variation, that is, the impedance of the one additional part, either  $C_{S1}$  or  $L_{S2}$ , or  $L_{S3}$ , is dependent



**Figure 15.12.** Simulation 2: transformer balun with one added part.



**Figure 15.13.** Insertion loss and phase shift of transformer balun for a group 3 UWB system ( $S_{21}, S_{31}$ ) in simulation 2: transformer balun with one added part.

on the frequency. At ports 2 and 3,  $S_{22} = S_{33} = -6 \text{ dB}$ . These values are not dropped down as is  $S_{11}$  due to the existence of three ports rather than two ports. Regardless, they can be ignored because they do not play any negative role to the simulated parameters.

Simulation 3 is conducted for the transformer balun with two parts added.

Figure 15.15 is the simulation setup with the transformer balun. Figures 15.16 and 15.17 show the insertion loss, phase shift, and return loss, respectively.

It can be found that, at the operating frequency  $f_0 = 7128 \text{ MHz}$ , there is no large difference in performance between simulations 2 and 3, except that

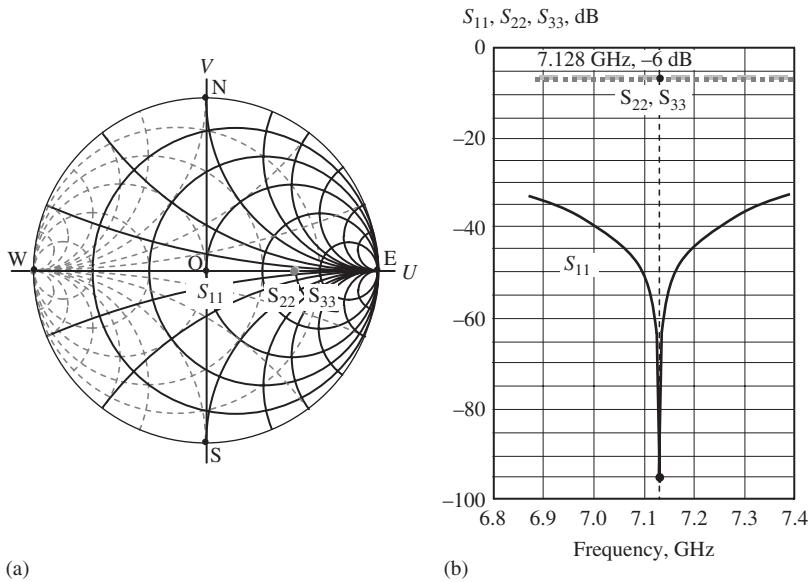


Figure 15.14. Return loss of transformer balun for a group 3 UWB system ( $S_{11}, S_{22}, S_{33}$ ) in simulation 2: transformer balun with one added part. (a) On Smith chart. (b) By rectangular coordination [dB].

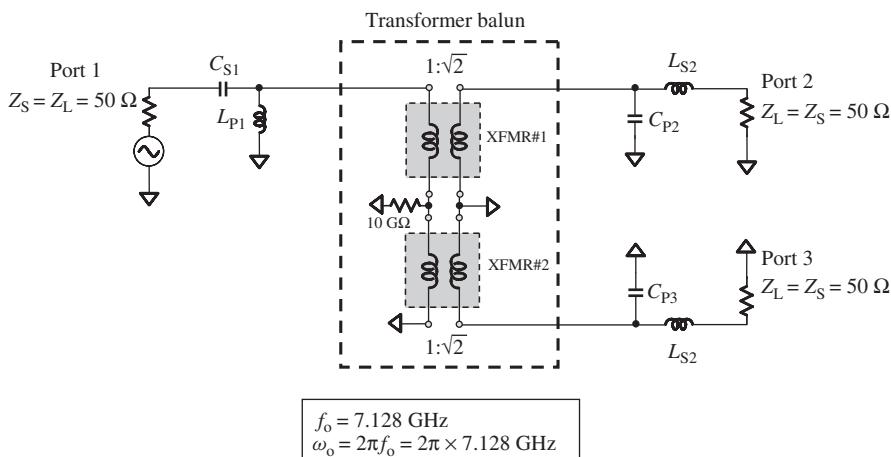
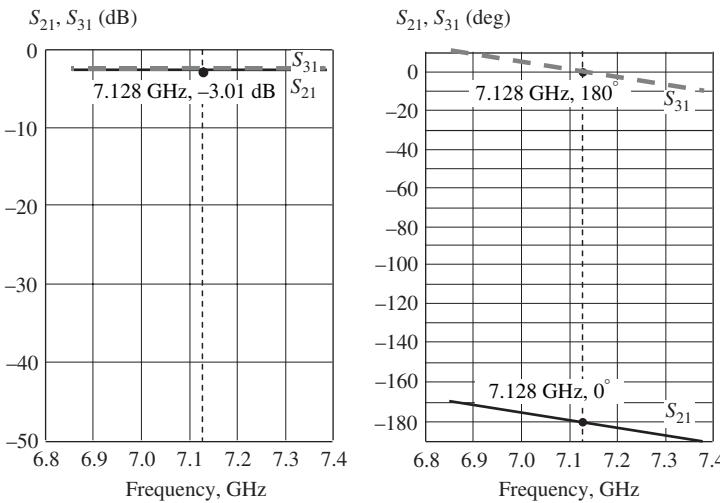


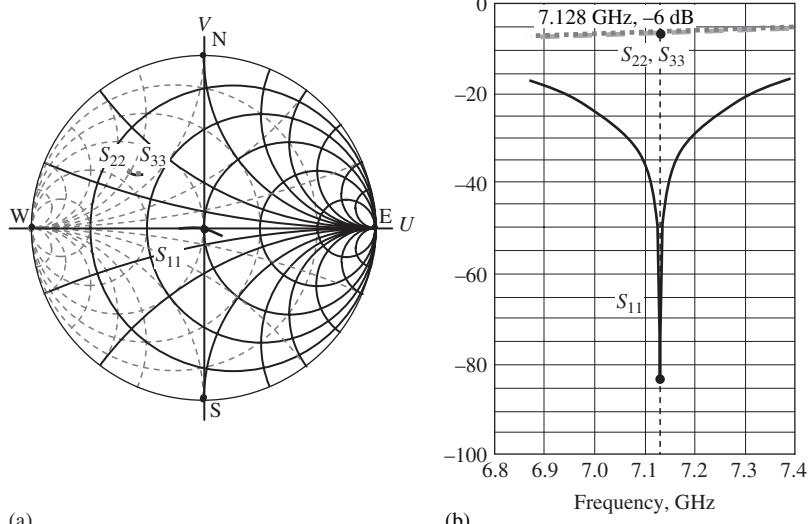
Figure 15.15. Simulation 3: transformer balun with two added parts.

- the magnitude of the return loss  $S_{11}$  at port 1 is dropped down to  $-83$  dB, which should be  $-\infty$  dB in an ideal transformer balun. Again, this difference is due to the error of truncated decimal numbers in the digital computation. Nevertheless, it indicates that port 1 is well matched to  $50 \Omega$ .

At other frequencies within the frequency bandwidth  $f = 6864\text{--}7392$  MHz,



**Figure 15.16.** Insertion loss and phase shift of transformer balun for a group 3 UWB system ( $S_{21}, S_{31}$ ) in simulation 3: transformer balun with two added parts.



**Figure 15.17.** Return loss of transformer balun for a group 3 UWB system ( $S_{11}, S_{22}, S_{33}$ ) in simulation 3: transformer balun with two added parts. (a) On Smith chart. (b) By rectangular coordination [dB].

- the frequency responses of  $S_{21}$  and  $S_{31}$  are not entirely flat; they are tilted more than in simulation 2, since there are two capacitors and two inductors,  $C_{S1}, L_{P1}$  and  $L_{S2}$ , and  $C_{P2}$ , which are dependent on the frequency. The tilted magnitude is less than 1.0 dB and the phase shift is less than 15°. Interestingly, the tilted directions of the phase of  $S_{21}$  and  $S_{31}$  are the same, so that the phase difference of the two differential ports is kept at around 180°;
- the magnitude of the return loss at port 1,  $S_{11}$ , varies considerably within the frequency band. At low and high frequencies, the values of  $S_{11}$  are around -20 to -40 dB, whereas at the central frequency the value of  $S_{11}$  is dropped down to

less than  $-83$  dB. The reason for this is the same as that of the previous variation, that is, the impedances of the additional parts,  $C_{S1}, L_{P1}$  and  $L_{S1}$ , and  $C_{P2}$ , are dependent on the frequency. At ports 2 and 3,  $S_{22}$  and  $S_{33}$  are changed from  $-7$  to  $-5$  dB. These values are not dropped as is  $S_{11}$  due to the existence of three ports rather than two ports. Regardless, they can be ignored because they do not play any negative role to the simulated parameters.

Simulation 4 is conducted for the transformer balun with three parts added.

Figure 15.18 is the simulation setup with the transformer. Figures 15.19 and 15.20 show the insertion loss, phase shift, and return loss, respectively.

It can be found that, at the operating frequency  $f_0 = 7128$  MHz, there is no large difference in performance between simulations 3 and 4, except

- the magnitude of the return loss  $S_{11}$  at port 1 is dropped down to  $-77$  dB, which should be  $-\infty$  dB in an ideal transformer balun. Again, this difference is due to the error of truncated decimal numbers in the digital computation. Nevertheless, it indicates that the port 1 is still well matched to  $50 \Omega$ .

At other frequencies within the frequency bandwidth  $f = 6864$ – $7392$  MHz,

- the frequency responses of  $S_{21}$  and  $S_{31}$  are not entirely flat, but tilted more than in simulation 3, since there are three capacitors and three inductors,  $C_{S1}, L_{P1}, L_{S1}$  and  $L_{S2}, C_{P2}, C_{S2}$ , which are dependent on the frequency. The tilted magnitude is less than  $1.5$  dB and the phase shift is less than  $25^\circ$ . Interestingly, the tilted directions of the phase of  $S_{21}$  and  $S_{31}$  are the same, so that the phase difference of the two differential ports is kept at around  $180^\circ$ ;

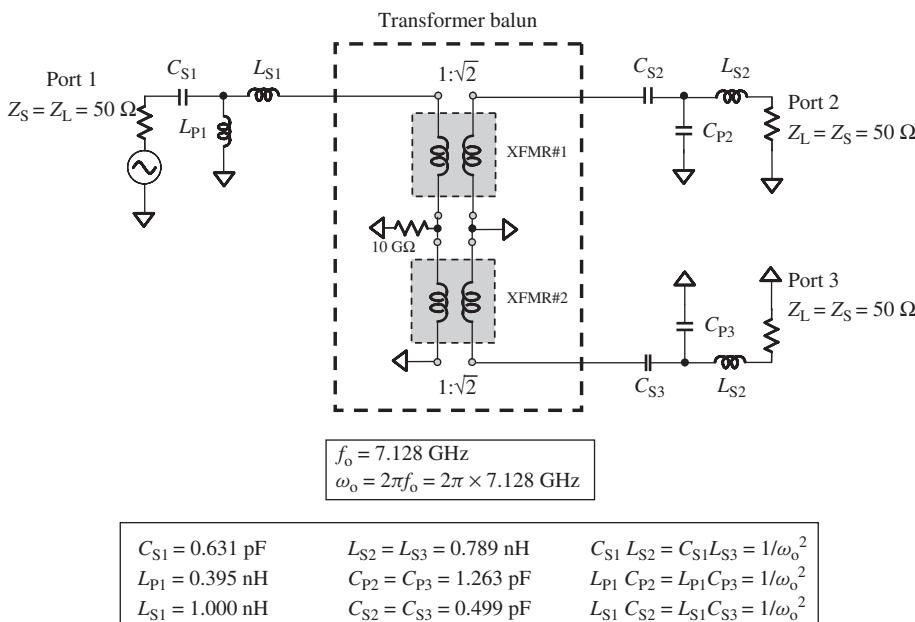
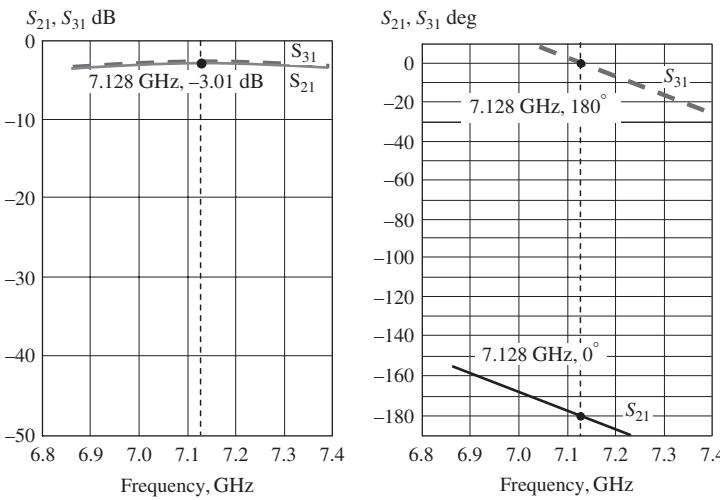
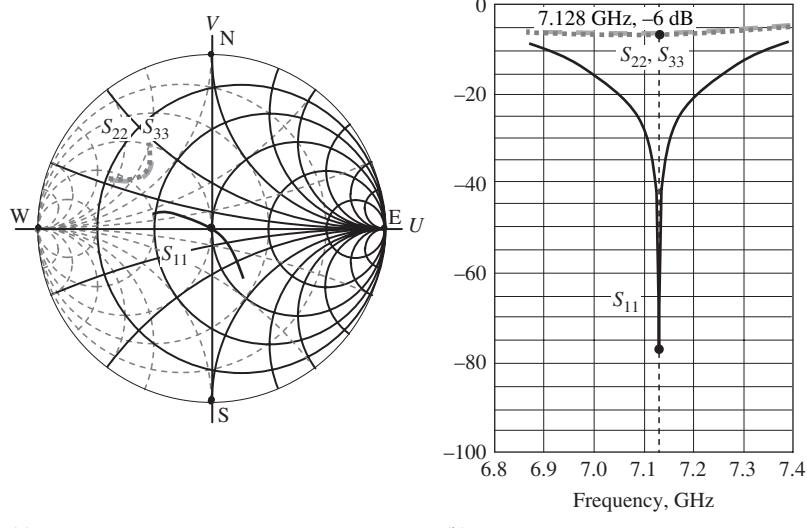


Figure 15.18. Simulation 4: transformer balun with three added parts.



**Figure 15.19.** Insertion loss and phase shift of transformer balun for a group 3 UWB system ( $S_{21}, S_{31}$ ) in simulation 4: transformer balun with three added parts.



**Figure 15.20.** Return loss of transformer balun for a group 3 UWB system ( $S_{11}, S_{22}, S_{33}$ ) in simulation 4: transformer balun with three added parts. (a) On the Smith chart. (b) By rectangular coordination [dB].

- the magnitude of the return loss at port 1,  $S_{11}$ , varies considerably within the frequency band. At low and high frequencies, the values of  $S_{11}$  are around -10 to -30 dB, while at the central frequency the value of  $S_{11}$  is dropped down to less than -77 dB. The reason for this is the same as that of the previous variation, that is, the impedance of the additional parts,  $C_{S1}, L_{P1}, L_{S1}$ , and  $L_{S2}, C_{P2}, C_{S2}$ , are dependent on the frequency. At ports 2 and 3,  $S_{22}$  and  $S_{33}$  are changed from -6 to -4 dB. These values are not dropped down as is  $S_{11}$  due to the existence of three ports rather than two ports. Regardless, they can be ignored because they do not play any negative role to the simulated parameters.

Simulation 5 is conducted for the transformer balun with four parts added.

Figure 15.21 is the simulation setup with the transformer balun. Figures 15.22 and 15.23 show the insertion loss, phase shift, and return loss, respectively.

It can be found that at the operating frequency,  $f_o = 7128, there is no large difference in performance between simulations 4 and 5, except that$

- the magnitude of the return loss  $S_{11}$  at port 1 is dropped down to  $-81$  dB, which should be  $-\infty$  dB in an ideal transformer balun. Again, this difference is due to

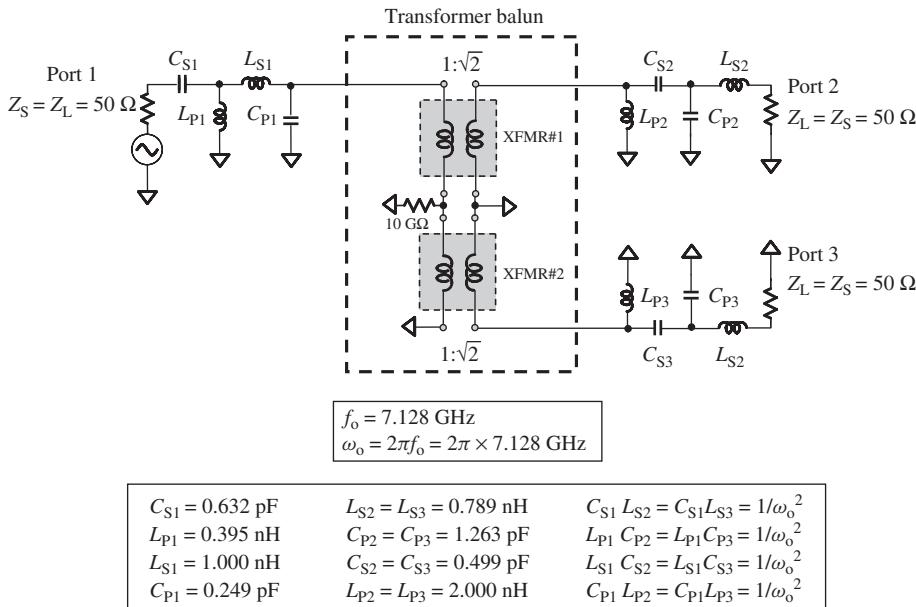


Figure 15.21. Simulation 5: transformer balun with four added parts.

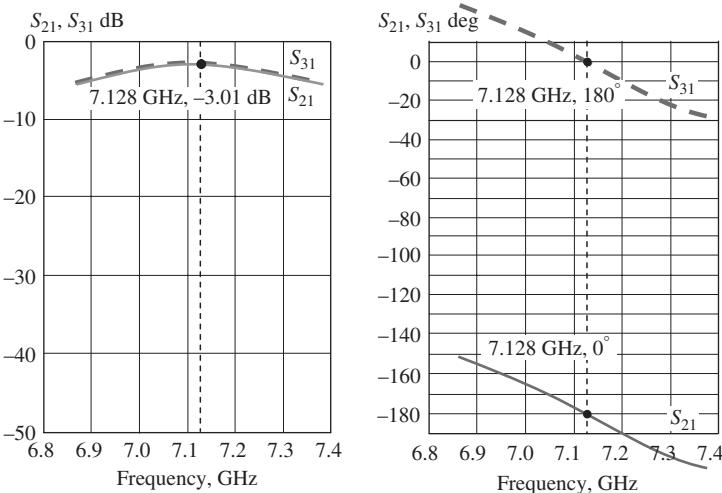
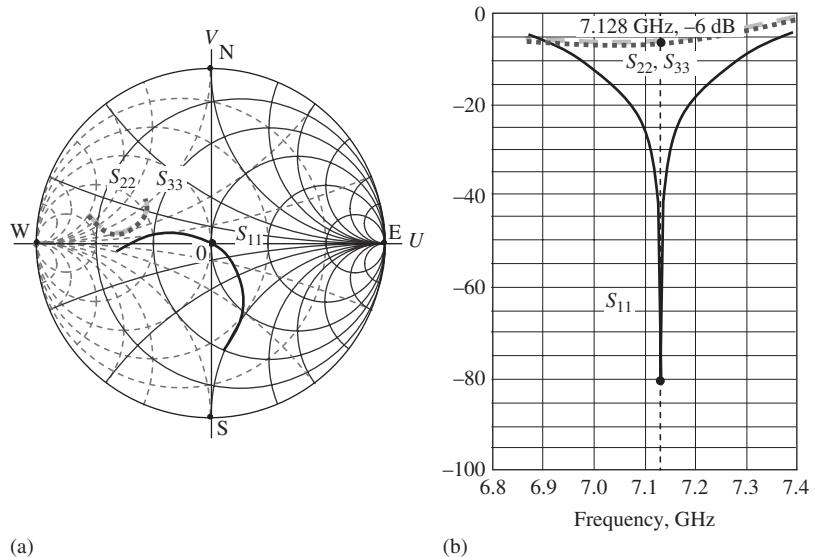


Figure 15.22. Insertion loss and phase shift of transformer balun for a group 3 UWB system ( $S_{21}, S_{31}$ ) in simulation 5: transformer balun with four added parts.



**Figure 15.23.** Return loss of transformer balun for group 3 UWB system ( $S_{11}, S_{22}, S_{33}$ ) in simulation 5: transformer balun with four added parts. (a) On Smith chart. (b) By rectangular coordination [dB].

the error of truncated decimal numbers in the digital computation. Nevertheless, it indicates that port 1 is still well matched to  $50 \Omega$ .

At other frequencies within the frequency bandwidth  $f = 6864\text{--}7392 \text{ MHz}$ ,

- the frequency responses of  $S_{21}$  and  $S_{31}$  are not entirely flat; they are tilted more than in simulation 4, since there are four capacitors and four inductors,  $C_{S1}, L_{P1}, L_{S1}, C_{P1}$ , and  $L_{S2}, C_{P2}, C_{S2}, L_{P2}$ , which are dependent on the frequency. The tilted magnitude is less than 3.0 dB and the phase shift is less than  $30^\circ$ . Interestingly, the tilted directions of the phase of  $S_{21}$  and  $S_{31}$  are the same, so that the phase difference of the two differential ports is kept at around  $180^\circ$ ;
- the magnitude of the return loss at port 1,  $S_{11}$ , varies considerably within the frequency band. At low and high frequencies, the values of  $S_{11}$  are around  $-5$  to  $-20 \text{ dB}$ , whereas at the central frequency, the value of  $S_{11}$  is dropped down to less than  $-81 \text{ dB}$ . The reason for this is the same as that of the previous variation, that is, the impedances of the additional parts,  $C_{S1}, L_{P1}, L_{S1}, C_{P1}$ , and  $L_{S2}, C_{P2}, C_{S2}, L_{P2}$ , are dependent on the frequency. At ports 2 and 3,  $S_{22}$  and  $S_{33}$  are changed from  $-7$  to  $-2 \text{ dB}$ . These values are not dropped down as is  $S_{11}$  due to the existence of three ports rather than two ports. Regardless, they can be ignored because they do not play any negative role to the simulated parameters.

In summary, from the five aforementioned simulations, it can be seen that the interpretation regulation in respect to the transformer balun is well verified because of the following:

1. At the assigned operating frequency  $f_0 = 7128 \text{ MHz}$ ,

- in all five simulations,  $S_{11}$  is always kept at very low level, that is,  $S_{11} < -70$  dB. This means that at port 1, the impedance is always kept at a value very close to  $50 \Omega$ ;
  - in all five simulations, the magnitudes of  $S_{21}$  and  $S_{31}$  are kept at  $-3.01$  dB and the phase between  $S_{21}$  and  $S_{31}$  is kept at  $180^\circ$ . This means that ports 2 and 3 are a real differential pair with only  $0.01$  dB attenuation from port 1 and that the impedance at both ports is kept at a value very close to  $50 \Omega$ ;
  - these results are due to the fact that in every simulation the added parts are interpreted from each other by the formula (15.5). Consequently, it is concluded that the  $S$ -parameters or impedances at all three ports are unchanged at the operating frequency  $\omega_o$  or  $f_o$  if the added parts at the single-ended side and at the differential pair side are interpreted from each other by the formula (15.5).
2. At the frequency range around the operating frequency  $f = 6864\text{--}7392$  MHz,
- in all five simulations,  $S_{11}$  is always kept at low levels, although not as low as that at  $f_o = 7128$  MHz. This means that at port 1, the impedance is always kept at a value around  $50 \Omega$ ;
  - in all five simulations, the magnitudes of  $S_{21}$  and  $S_{31}$  are less than  $-3$  dB but not less than  $-5$  dB and, interestingly, the phase between  $S_{21}$  and  $S_{31}$  is kept around  $180^\circ$ . It means that ports 2 and 3 can still be considered as a differential pair, but with an additional attenuation of  $1\text{--}2$  dB;
  - these results are due to the fact that in every simulation the added parts are interpreted from each other by the formula (15.5). Consequently, it is concluded that the  $S$ -parameters or impedances at all three ports are approximately unchanged in the frequency range of  $f = 6864\text{--}7392$  MHz if the added parts at the single-ended side and at differential pair side are interpreted from each other by the formula (15.5).

Now let us return to simulations 6 and 7.

Simulation 6 is conducted to check the impedance  $Z_1$  looking into the transformer balun from the single-ended side if the added four parts  $C_{S1}, L_{P1}, L_{S1}$ , and  $C_{P1}$  at the single-ended side as shown in Figure 15.21 are removed. Figure 15.24 shows its simulation setup.

Simulation 7 is conducted to check the impedance  $Z_2$  at one of differential branches immediately after the transformer balun, looking into the differential port 2. Figure 15.25 shows its simulation setup.

Tested results from simulations 6 and 7 indicates that

$$Z_1 = Z_2 = Z, \quad (15.7)$$

and

$$\begin{aligned} \text{when } f = 7.1 \text{ GHz, } Z_1 &= Z_2 = Z = 10.16 + j56.7 \Omega; \\ \text{when } f = 7.2 \text{ GHz, } Z_1 &= Z_2 = Z = 10.76 + j59.1 \Omega. \end{aligned}$$

From simulations 6 and 7, it is concluded that the impedance looking from the single end before the transformer balun is the same as that looking into one of the differential branches after the transformer balun. In other words, the impedance of one differential

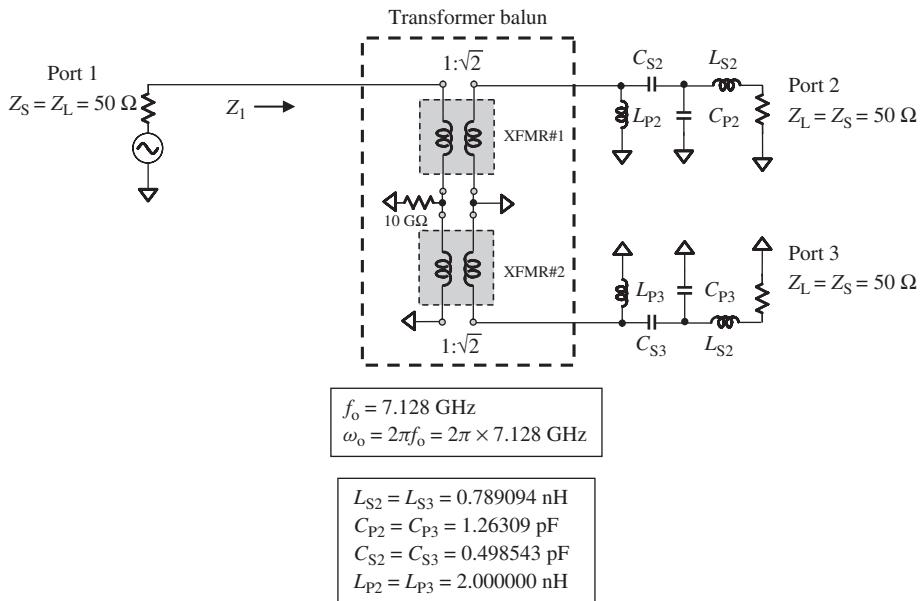


Figure 15.24. Simulation setup for testing of impedance  $Z_1$ .

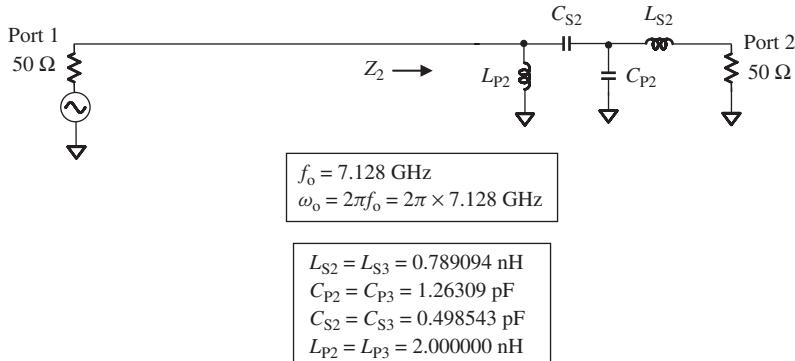
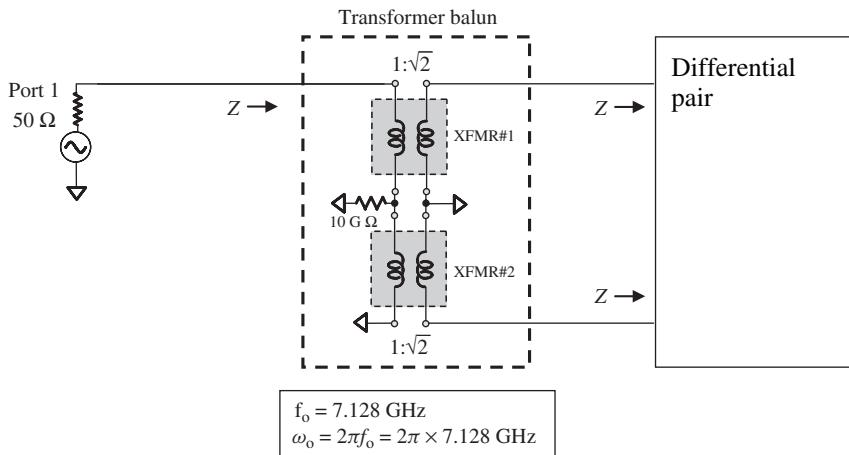


Figure 15.25. Simulation setup for testing of impedance  $Z_2$ .

branch  $Z_2$  can be found at the single-ended port as  $Z_1$ . This transformer balun looks like a monster-revealing mirror. You stand at the front of this magic mirror, you can find out the impedance behind it.

### 15.2.5 Impedance Matching for Differential Pair by means of Transformer Balun

Now let us illustrate how to replace the impedance matching for a differential pair by impedance matching for a single-ended stage, so that a circuitry with a differential configuration can be treated as circuitry with a single-ended stage.



**Figure 15.26.** Differential pair is connected with a transformer balun built by two stacked transformers with their turns ratio =  $1 : \sqrt{2}$ .

Assume that the impedance of each differential branch is  $Z$ . If it must be impedance-matched to  $50 \Omega$ , the impedance matching can be conducted as follows:

1. As shown in Figure 15.26, we connect the differential terminals of the transformer balun, which is built by two stacked transformers with their turn ratio =  $1 : \sqrt{2}$ , to the terminals of the differential pair. As mentioned in simulations 6 and 7, the impedance looking from the single-ended stage before the transformer balun is the same as that looking into a differential branch after the transformer balun. All of them have same value  $Z$  as shown in Figure 15.26.
2. At the single-ended port, we carry out impedance matching between port 1 with the  $50 - \Omega$  source to the load  $Z$  and form the single-ended impedance matching network as shown in Figure 15.27, in which the value of the parts are listed in the bottom-left corner.
3. We interpret the single-ended impedance matching network to the differential impedance matching network in terms of the rules described in Section 15.2.4. The results are shown in Figure 15.28.
4. Removing port 1, the single-ended impedance matching network, and the transformer balun, and realigning the differential impedance matching network and the differential pair, we have the result shown in Figure 15.29.

The four-step procedure above demonstrates how to replace the impedance matching for a differential pair by impedance matching for a single-ended stage. It should be taken into account that this is exactly true only at the operating frequency, and is only approximately true in the frequency range around the operating frequency within a certain frequency bandwidth. In cases of narrow bandwidth, with a relative bandwidth lower than 15%, the replacement of differential impedance matching by single-ended impedance matching should not be problematic in most cases. In cases of wide bandwidth, with a relative bandwidth higher than 15%, the replacement is not problematic at the assigned operating frequency but may have considerable discrepancies at other frequencies.

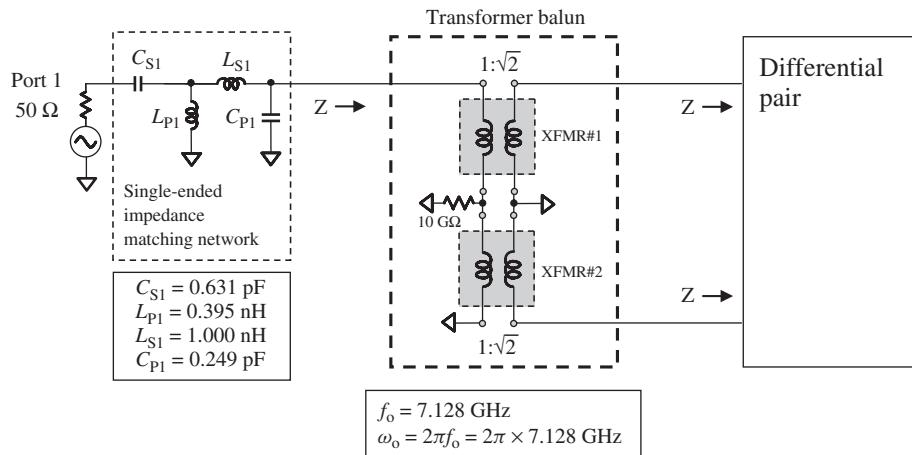


Figure 15.27. Impedance matching at single-ended side.

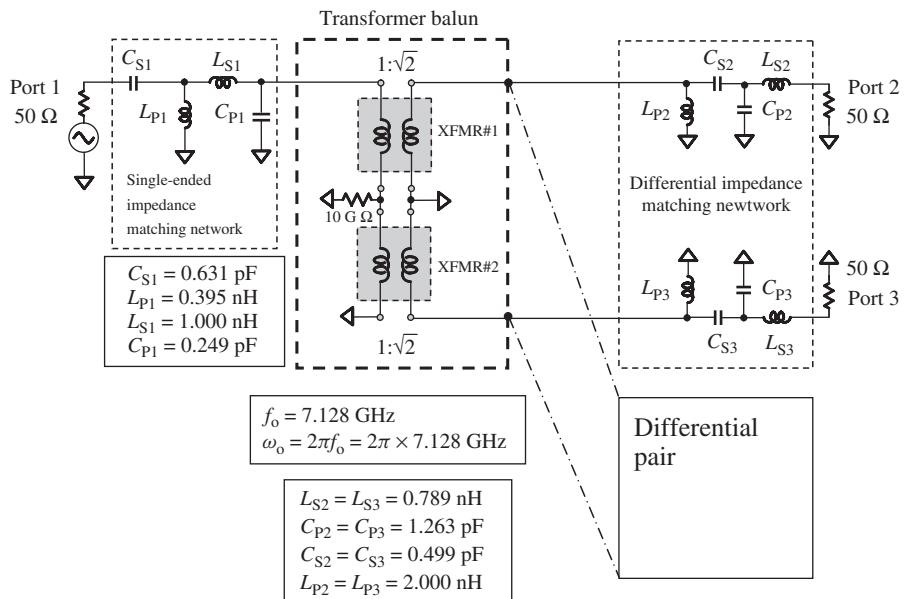


Figure 15.28. Interpretation of single-ended impedance matching network to differential impedance matching network (refer to Figure 15.21).

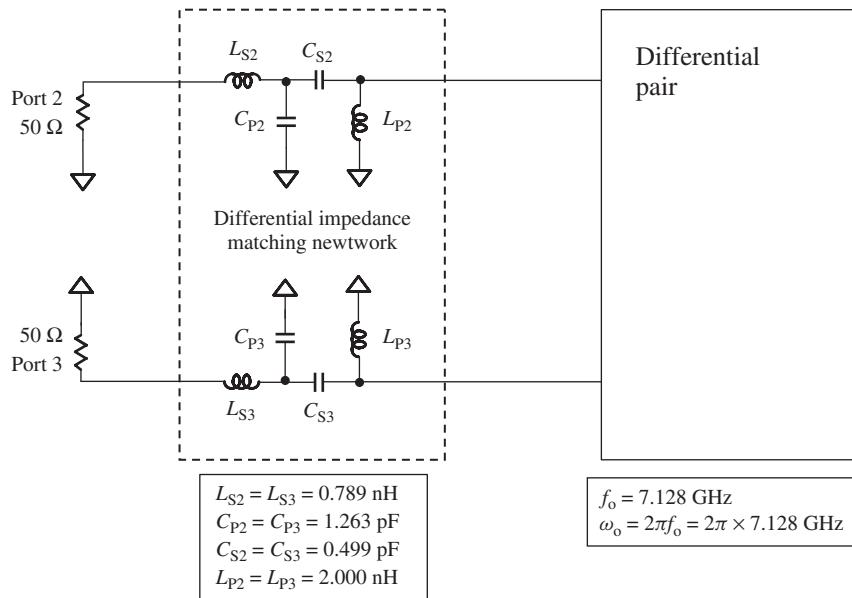


Figure 15.29. Final simulation result for differential impedance matching network.

### 15.3 LC BALUN

Figure 15.30 depicts a simple LC balun. It looks like a bridge and can serve as a splitter or a combiner. We will focus on the splitter since the combiner is simply a reversed splitter.

The distinguishing features of the LC balun are as follows:

- *Simplest Configuration.* The LC balun shown in Figure 15.30 is built by only two identical capacitors and two identical inductors. It is the simplest balun among all the baluns that have been developed so far.
- *Cost-Effective design.* It is a balun with the lowest cost among all the baluns that have been developed so far, due to its aforementioned simplicity.
- *Direct Functioning as an Impedance Matching Network.* The values of  $L$  and  $C$  in an LC balun are calculated on the basis of the impedances of the single-ended stage and differential pair. Therefore, the LC balun itself is the impedance matching network between the single-ended stage and the differential pair.

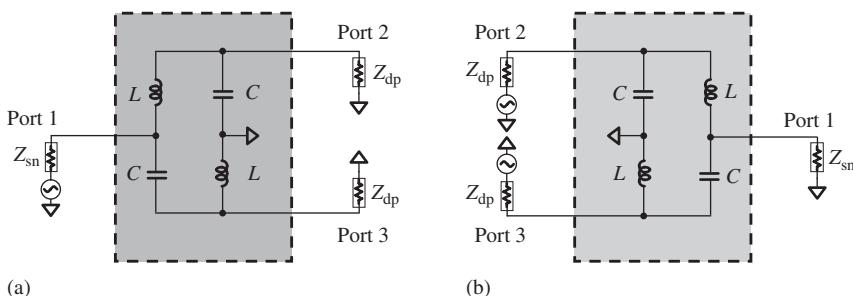


Figure 15.30. LC balun functioning as a splitter or a combiner. (a) LC splitter. (b) LC combiner.

- *Existence of Equivalence of Parts between Single-Ended Stage and Differential Pair.* It enables these parts to be interpreted from each other. Just like the transformer balun discussed in Section 15.2, the equivalence of parts between the single-ended and differential pair allows the simulation of the differential pair circuitry to be replaced by a simulation of single-ended circuitry.
- *Relatively Narrow Bandwidth of the LC Balun.* By comparing with other baluns, the bandwidth of LC balun is relatively narrow because the body of the LC balun itself, two inductors and two capacitors, is dependent on the frequency. This is the unique drawback of the LC balun.

### 15.3.1 Simplicity of LC Balun Design

An analysis of the simple balun shown in Figure 15.30(a) is given in Appendix 15.A.2. Designing a simple LC balun is tantamount to calculating the values of its inductors and capacitors. By the derivation in Appendix 15.A.2,

$$L\omega_0 = \sqrt{2Z_{dp}Z_{sn}^*}, \quad (15.8)$$

and

$$C\omega_0 = \frac{1}{\sqrt{2Z_{dp}Z_{sn}^*}}, \quad (15.9)$$

where

$L$  = the inductance of the inductor in the simple LC balun,

$C$  = the capacitance of the capacitor in the simple LC balun,

$Z_{sn}$  = the source impedance at the single-ended port,

$Z_{dp}$  = the load impedance at each port of the differential pair,

$\omega_0$  = the operating angular frequency,

and

$$Z_{sn} = R_{sn} + jX_{sn}, \quad (15.10)$$

$$Z_{dp} = R_{dp} + jX_{dp}. \quad (15.11)$$

In terms of formulas (15.8) and (15.9), a simple LC balun is easily designed.

### 15.3.2 Performance of a Simple LC Balun

Now let us explore the features of this simple LC balun in terms of the following two simulations.

**SIMULATION A.** Simulation of simple LC balun operated in the frequency range of a group 1 UWB system:  $f = 3696\text{--}3960\text{--}4224$  MHz,  $f_0 = 3960$  MHz,  $Z_S = Z_L = 50 \Omega$ .

Figure 15.31 shows setup of the simulation for a group 1 UWB system.

Figure 15.32 shows the insertion loss and phase shift, whereas Figure 15.33 shows the return loss of the simple LC balun operating in the frequency range of a group

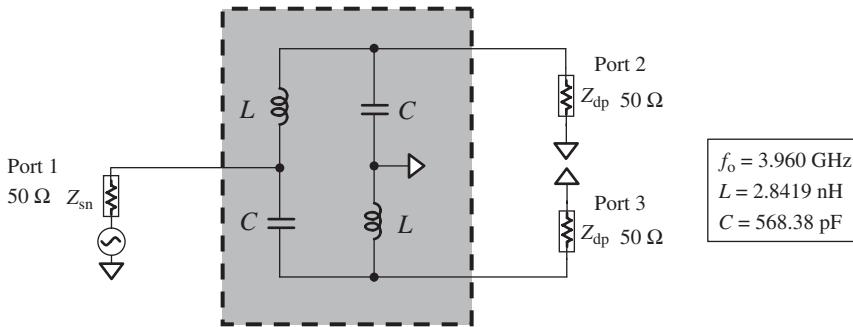


Figure 15.31. Setup of simulation of LC balun for group 1, UWB system.

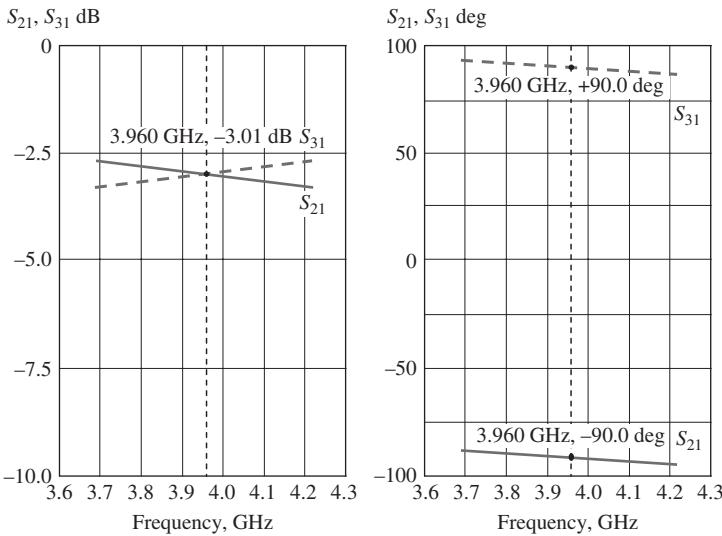
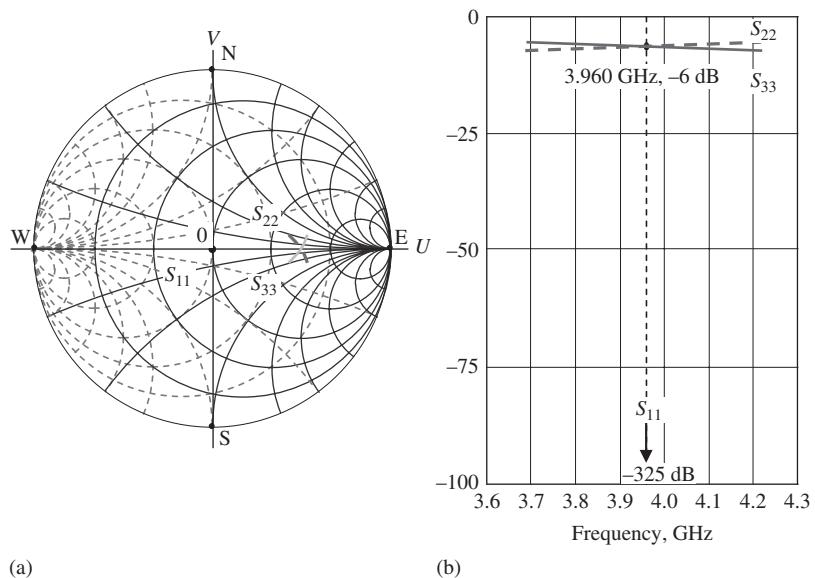


Figure 15.32. Insertion loss and phase shift of LC balun for a group 1 UWB system.

1 UWB system. Table 15.3 lists the values of the relevant  $S$ -parameters and their performance.

At the operating frequency  $f_0 = 3960 \text{ MHz}$ ,  $S_{21} = S_{31} = 3.01 \text{ dB}$ , which means that at each port of the differential pair the impedance is matched with the LC balun well. In addition, the phase at one port of the differential is  $90^\circ$  and the phase at the other is  $-90^\circ$ , so that the phase difference between the two ports is  $180^\circ$ . The magnitude of the return loss at port 1,  $S_{11}$ , is dropped down to  $-325 \text{ dB}$ , so that it is not shown on the plot. At ports 2 and 3,  $S_{22} = S_{33} = -6 \text{ dB}$ . These values are not lowered as is  $S_{11}$ , because of the existence of three ports rather than two ports. However, this imperfection does not impact the simulation at all.

At other frequencies within the frequency bandwidth  $f = 3696\text{--}4224 \text{ MHz}$ , the frequency responses of  $S_{21}$  and  $S_{31}$  are not flat, but tilted, since the capacitor  $C$  and inductor  $L$  are dependent on frequency. However, the tilted magnitude is less than  $0.5 \text{ dB}$  and the phase shift is less than  $5^\circ$ . Interestingly, the tilt directions of the phases of  $S_{21}$  and  $S_{31}$  are the same, so that the phase difference of the two differential ports is kept around  $180^\circ$ . The magnitude of the return loss at port 1,  $S_{11}$ , is dropped down to about  $-325 \text{ dB}$  so that it is not shown on the plot. At ports 2 and 3,  $S_{22}$  and  $S_{33}$  are approximately  $-6 \text{ dB}$ ,



**Figure 15.33.** Return loss of LC balun for group 1 UWB system ( $S_{11}, S_{22}, S_{33}$ ). (a) On Smith chart. (b) By rectangular coordination [dB].

**TABLE 15.3.** Values of Relevant  $S$ -Parameters of the Ideal Transformer Balun Operating in the Frequency Range of a Group 1 UWB system:  $f = 3696\text{--}3960\text{--}4224$  MHz,  $Z_S = Z_L = 50 \Omega$

Frequencies	$f$	3696–3960–4224		MHz
Bandwidth	$\Delta f$		528	MHz
Insertion loss	$S_{21}$	Magnitude	$-3.01 \pm 0.5$ (maximum)	dB
		Phase	$-90^\circ \pm 5^\circ$ (maximum)	—
	$S_{31}$	Magnitude	$-3.01 \pm 0.5$ (maximum)	dB
		Phase	$+90^\circ \pm 5^\circ$ (maximum)	—
Return loss	$S_{11}$	Magnitude	$-325 \pm \dots$	dB
	$S_{22}$	Magnitude	$-6.00 \pm 0.75$	dB
	$S_{33}$	Magnitude	$-6.00 \pm 0.75$	dB

with a variation of less than 0.75 dB. These values are not lowered as is  $S_{11}$ , because of the existence of three ports rather than two ports. This imperfection, just like that at the operating frequency  $f_0 = 3960$  MHz, does not impact the simulation at all.

**SIMULATION B.** Simulation of the ideal transformer balun operated in the frequency range of a group 3 UWB system:  $f = 6864\text{--}7128\text{--}7392$  MHz,  $f_0 = 7128$  MHz,  $Z_S = Z_L = 50 \Omega$ .

Figure 15.34 shows the setup of simulation for a group 3 UWB system.

Figure 15.35 shows the insertion loss and phase shift, while Figure 15.36 shows the return loss of the ideal transformer balun operating in the frequency range of a group 3 UWB system. Table 15.4 lists the values of interested  $S$ -parameters of their performance.

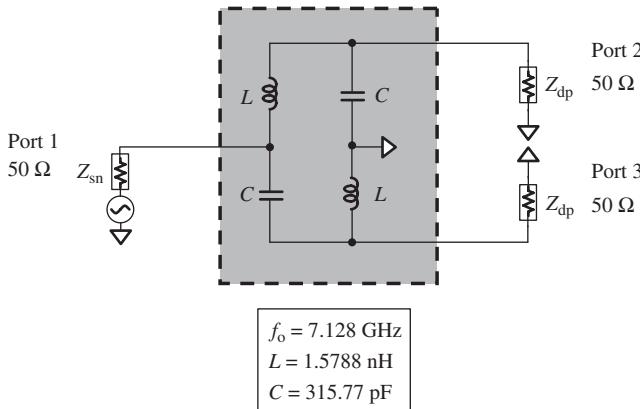


Figure 15.34. Setup of simulation for a group 3 UWB system.

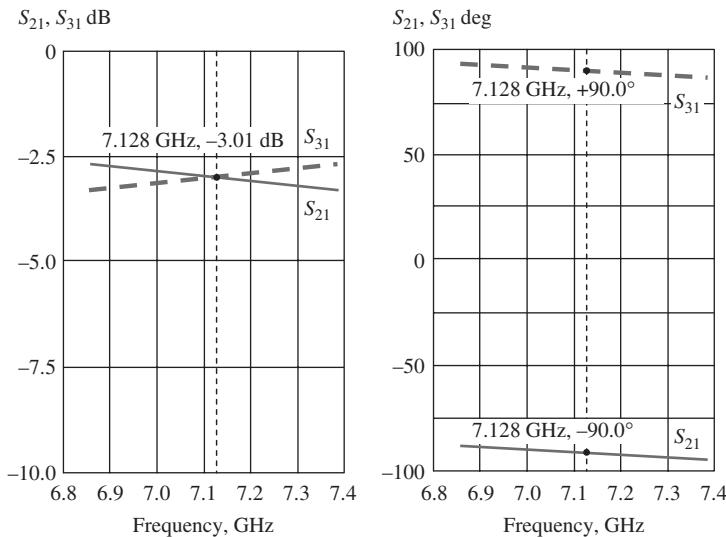
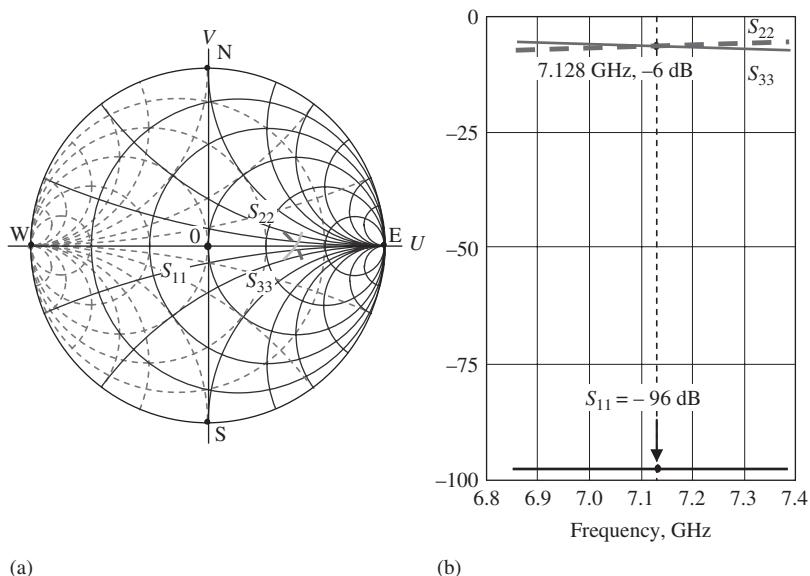


Figure 15.35. Insertion loss and phase shift of LC balun for a group 3 UWB system ( $S_{21}, S_{31}$ ).

At the operating frequency  $f_0 = 7128 \text{ MHz}$ ,  $S_{21} = S_{31} = -3.01 \text{ dB}$ , which means that, at each port of the differential pair, the impedance is matched with the LC balun well. In addition, the phase at one port of the differential pair is  $90^\circ$  and the phase at the other is  $-90^\circ$ , so that the phase difference between the two ports is  $180^\circ$ . The magnitude of the return loss at port 1,  $S_{11}$ , is dropped down to  $-96 \text{ dB}$  so that it is not shown on the plot. At ports 2 and 3,  $S_{22} = S_{33} = -6 \text{ dB}$ . These values are not lowered as is  $S_{11}$ , because of the existence of three ports rather than two ports. However, this imperfection does not impact the simulation at all.

At other frequencies within the frequency bandwidth  $f = 6864\text{--}7392 \text{ MHz}$ , the frequency responses of  $S_{21}$  and  $S_{31}$  are not flat, but tilted, since the capacitor  $C$  and the inductor  $L$  are dependent on the frequency. The tilted magnitude is less than  $0.5 \text{ dB}$  and the phase shift is less than  $5^\circ$ . Interestingly, the tilt directions of the phases of  $S_{21}$  and  $S_{31}$  are the same, so that the phase difference of the two differential ports is kept around  $180^\circ$ . The magnitude of the return loss at port 1,  $S_{11}$ , is dropped down to about  $-96 \text{ dB}$  so that it is not shown on the plot. At ports 2 and 3,  $S_{22}$  and  $S_{33}$  are approximately  $-6 \text{ dB}$ , with a variation of less than  $0.75 \text{ dB}$ . These values are not lowered as is  $S_{11}$ ,



**Figure 15.36.** Return loss of LC balun for a group 3 UWB system ( $S_{11}, S_{22}, S_{33}$ ). (a) On Smith chart. (b) By rectangular coordination [dB].

**TABLE 15.4.** Values of Interested S-Parameters of the Ideal Transformer Balun Operated in the Frequency Range of a Group 3 UWB system:  $f = 6864\text{--}7128\text{--}7392$  MHz;  $Z_S = Z_L = 50 \Omega$

Frequencies	$f$	6864–7128–7392		MHz
Bandwidth	$\Delta f$		528	MHz
Insertion loss	$S_{21}$	Magnitude	$-3.01 \pm 0.5$ (maximum)	dB
		Phase	$-90^\circ \pm 5^\circ$ (maximum)	—
	$S_{31}$	Magnitude	$-3.01 \pm 0.5$ (maximum)	dB
		Phase	$+90^\circ \pm 5^\circ$ (maximum)	—
Return loss	$S_{11}$	Magnitude	$-96 \pm \dots$	dB
	$S_{22}$	Magnitude	$-6.00 \pm 0.75$ (maximum)	dB
	$S_{33}$	Magnitude	$-6.00 \pm 0.75$ (maximum)	dB

because of the existence of three ports rather than two ports. However, this imperfection, just like that at the operating frequency  $f_o = 7128$  MHz, does not impact the simulation at all.

### 15.3.3 A Practical LC Balun

In Section 15.3.1, the simplicity of an LC balun was shown, and in Section 15.3.2 its performance was demonstrated.

One might feel that the design of an LC balun is an easy task. It seems that all we need to do is to calculate the values of  $L$  and  $C$  from equations (15.8) and (15.9) according to the impedance of the single-ended and differential pair ports,  $Z_{sn}$  and  $Z_{dp}$ . Unfortunately, it is generally not so simple.

First, it should be noted that, in the demonstration of LC balun in Section 15.3.2, the impedances of all three ports were  $50 \Omega$ . This means that the LC balun applied in the demonstration was designed in a special case in which  $Z_{sn}$  and  $Z_{dp}$  in the design formulas (15.8) and (15.9) were both  $50 \Omega$ . In most cases,  $Z_{sn}$  and  $Z_{dp}$  are complex values.

It follows that there is a problem with formulas (15.8) and (15.9). In general cases,  $Z_{sn}$  and  $Z_{dp}$  are complex values, and the product of  $Z_{sn}^*$  and  $Z_{dp}$  under the sign of square root is a complex number, so it is impossible to get real values for the inductor  $L$  and the capacitor  $C$ . But, in reality, only inductors and capacitors with real values, not complex values, are available.

As a matter of fact, this implies that there is a restriction in the design of the simple balun. The product  $Z_{sn}^*$  and  $Z_{dp}$  can be expressed as

$$Z_{dp}Z_{sn}^* = (R_{dp} + jX_{dp})(R_{sn} - jX_{sn}) = (R_{dp}R_{sn} + X_{dp}X_{sn}) + j(X_{dp}R_{sn} - R_{dp}X_{sn}). \quad (15.12)$$

In order to ensure that the product of  $Z_{sn}^*$  and  $Z_{dp}$  under the square root symbol is a complex number, the imaginary portion of the product  $Z_{dp}Z_{sn}^*$  must be a real number. This means that

$$X_{dp}R_{sn} - R_{dp}X_{sn} = 0, \quad (15.13)$$

or

$$\frac{R_{sn}}{R_{dp}} = \frac{X_{sn}}{X_{dp}}. \quad (15.14)$$

The single-ended impedance  $Z_{sn}$  must be proportional to  $Z_{dp}$  so that the ratio of their real parts is equal to the ratio of their imaginary parts, as shown in equation (15.14).

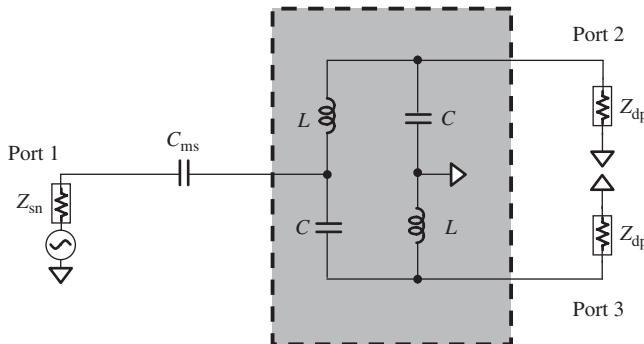
Now, before the calculation for the values of  $L$  and  $C$  from equations (15.8) and (15.9), we must first check whether condition (15.14) is satisfied or not. In special cases, if both the given  $Z_{sn}$  and the given  $Z_{dp}$  are pure resistances or reactances, this limitation will automatically disappear. However, for general cases, both the given  $Z_{sn}$  and  $Z_{dp}$  are neither pure resistance nor pure reactance, and the problem remains.

The unique solution to the problem is to modify either  $Z_{sn}$  or  $Z_{dp}$  by the addition of an extra part, either an inductor or capacitor, so that the modified  $Z'_{sn}$  or  $Z'_{dp}$  satisfies condition (15.14). There are three considerations in the process of modification:

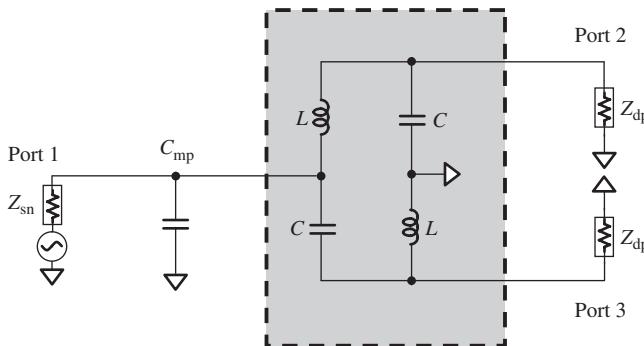
1. In order to avoid the increase of unnecessary insertion loss, only the imaginary part but not the real part of  $Z_{sn}$  or  $Z_{dp}$  is modified.
2. It is more economical if the modification is done at the single-ended side instead of the differential pair side.
3. It is more economical if a capacitor instead of an inductor is used for the modification.

In general cases, the original LC balun as shown in Figure 15.30 must be modified to that as shown in Figure 15.37 or 15.38. Again, we are going to focus on the balun as a splitter but not as a combiner.

In Figure 15.37, the capacitor  $C_{ms}$  is employed to modify the simple balun. It is inserted between port 1 and the input of the balun in series at the single-ended side. In Figure 15.39, the capacitor  $C_{mp}$  is employed to modify the simple balun. It is connected between port 1 and the ground in parallel at the single-ended side.



**Figure 15.37.** Simple LC balun modified by inserting a capacitor in series  $C_{ms}$  at the single-ended side.



**Figure 15.38.** Simple LC balun modified by inserting a capacitor in parallel  $C_{mp}$  at the single-ended side.

Let us discuss the case as shown in Figure 15.37, where the capacitor  $C_{ms}$  is added in series to modify the simple LC balun. The other case, as shown in Figure 15.38, where the capacitor  $C_{mp}$  is added *in parallel* to modify the simple LC balun will be left to the reader as an exercise.

Assume that the original  $Z_{sn}$  and  $Z_{dp}$  are

$$Z_{sn} = r_{sn} + jx_{sn}, \quad (15.15)$$

$$Z_{dp} = r_{dp} - jx_{dp}, \quad (15.16)$$

where  $r_{sn}, x_{sn}, r_{dp}$ , and  $x_{dp}$  are assumed to be positive values.

At the single-ended port, the impedance will be changed from the original value of  $Z_{sn}$  as shown in (15.15) to the modified value  $Z'_{sn}$  by the insertion of the capacitor  $C_{ms}$  in series, such that

$$Z'_{sn} = r_{sn} - jx'_{sn}, \quad (15.17)$$

where  $x'_{sn}$  is assumed to be a positive value as well, so that the values of  $Z'_{sn}$  and  $Z_{dp}$  satisfy limitation (15.14). This implies that

$$x'_{sn} = \frac{x_{dp}}{r_{dp}} r_{sn}. \quad (15.18)$$

The value of  $C_{ms}$  can be calculated from the difference  $\Delta Z_{sn}$  between  $Z_{sn}$  and  $Z'_{sn}$ : that is

$$\Delta Z_{sn} = Z'_{sn} - Z_{sn} = (-jx'_{sn}) - jx_{sn} = -j(x'_{sn} + x_{sn}) = -j \left( \frac{x_{dp}}{r_{dp}} r_{sn} + x_{sn} \right), \quad (15.19)$$

$$C_{ms} = \frac{1}{\left(\frac{x_{dp}}{r_{dp}}r_{sn} + x_{sn}\right)\omega}, \quad (15.20)$$

where  $\omega$  = the operating angular frequency.

Let us have an example with detailed values. Assume that

$$Z_{sn} = r_{sn} + jx_{sn} = 80 \Omega + j35 \Omega,$$

$$Z_{dp} = r_{dp} - jx_{dp} = 100 \Omega - j65 \Omega,$$

$$f = 500 \text{ MHz}.$$

Then

$$x'_{sn} = \frac{x_{dp}}{r_{dp}}r_{sn} = 52 \Omega,$$

$$C_{ms} = \frac{1}{\left(\frac{x_{dp}}{r_{dp}}r_{sn} + x_{sn}\right)\omega} = 6.12 \text{ pF},$$

$$Z'_{sn} = r_{sn} - jx'_{sn} = 80 \Omega - j52 \Omega,$$

$$Z_{dp}Z'^*_{sn} = r_{dp}r_{sn} + x_{dp}x'_{sn} = 11,380 \Omega^2,$$

$$L = \frac{\sqrt{2Z_{dp}Z'^*_{sn}}}{\omega} = \frac{\sqrt{2(r_{dp}r_{sn} + x_{dp}x'_{sn})}}{\omega}$$

$$= \frac{\sqrt{22,760}}{3141.6 \times 10^6} = \frac{150.864}{3141.6} \times 10^{-6} \text{ H} = 48.02 \text{ nH},$$

and

$$C = \frac{1}{\sqrt{2Z_{dp}Z'^*_{sn}}\omega} = \frac{1}{\sqrt{2(r_{dp}r_{sn} + x_{dp}x'_{sn})\omega}} = \frac{10^{-6}}{150.864 \times 3141.6} = 2.11 \text{ pF}.$$

It should be noted that the frequency response of the modified LC balun shown in Figure 15.38 may be somewhat degraded from that of the simple balun shown in Figure 15.30(a). However, this modification is necessary because the product of  $Z_{sn}^*$  and  $Z_{dp}$  is not a real value in most cases.

Compared with the transformer balun, the drawback of the LC balun is its narrow-band frequency response. This is an inherent property of the LC balun because its basic parts, two inductors and two capacitors, are dependent on frequency. Nevertheless, the LC balun is widely employed in RF circuit design, especially in the testing of differential circuitry today. Beside its many merits mentioned above, the most attractive point of the LC balun is that it can be built by hand in minutes as long as the calculations and preparations for the two inductors and three capacitors as shown in Figure 15.38 or 15.39 have been done well. If the frequency bandwidth of the circuit block is wide, more than one LC balun can be built so as to cover the entire bandwidth in the testing.

## 15.4 MICROSTRIP LINE BALUN

In the low RF range, the application of the microstrip line in the circuit design is restricted by its size. In high RF range, however, the application of microstrip line is receiving more and more attention every year. A microstrip line can replace individual parts, such as inductors and capacitors, and even a simple network. For example, in many RF circuit blocks such as the LNA or PA, the recommendation is to implement all the input and output impedance matching networks by microstrip lines because

- in general, the simulation model of a microstrip line is more accurate than that of discrete parts;
- the tolerance of a microstrip line is more easily controlled than that of other parts;
- Most importantly, the application of the microstrip line is the most cost effective, especially in RF module design or in RF block design by discrete parts.

In the last decades, many baluns of the microstrip line type were developed. It would be impossible to describe all of them in this section. Instead, we will introduce a few types that have been quite welcomed by RF designers.

### 15.4.1 Ring Balun

The ring microstrip line balun is a very simple but sophisticated balun applied in the frequency range from UHF to tens of gigahertz. It is an important part in RF circuit design or testing of RF blocks.

As shown in Figure 15.39, the main body of the ring microstrip line balun is a metallic ring on the PCB in an RF module design or on the IC substrate in an IC chip. The ring's circumference is equal to 1.5 times the wavelength of the electric length, and its width must be chosen so that its characterized impedance is equal to  $2^{1/2}Z_0$ . The characterized impedance of the four tapes must be equal to the source or load impedance  $Z_0 = 50 \Omega$ . The arc of the ring is one-quarter wavelength from port 1 to port 3, port 3 to port 2, and port 2 to port 4, but is three-quarters wavelengths from port 4 to port 1. If port 1 is chosen as the single-ended input port, then port 2 is a virtual grounding because

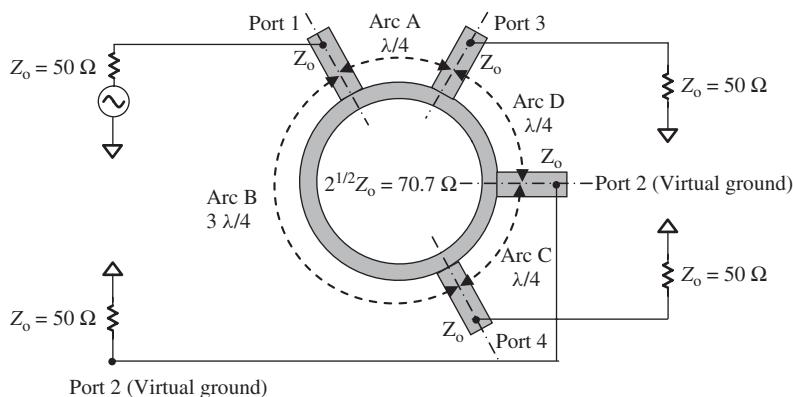


Figure 15.39. Layout of a ring microstrip line balun.

it is at a distance of one-half wavelength from port 1. Ports 3 and 4 are assigned as the two differential ports.

Figure 15.40 shows its equivalent circuit.

- Port 2 is a virtual ground.
- The impedance looking from port 3 toward port 2 is

$$Z_3 = Z_o = 50 \Omega, \quad (15.21)$$

where  $Z_3$  = the impedance looking from port 3 toward port 2, because the arc d is one-quarter of the wavelength from the virtual grounded port 2, so that its impedance is infinite.

- The impedance looking from port 4 toward port 2 is

$$Z_4 = Z_o = 50 \Omega, \quad (15.22)$$

where  $Z_4$  = the impedance looking from port 4 toward port 2, because the arc c is one-quarter of the wavelength from the virtual grounded port 2, so that its impedance is infinite.

- The impedance looking from port 1 toward arc a or arc b is

$$Z_a = Z_b = 2Z_o = \frac{(2^{1/2}Z_o)^2}{Z_o} = 100 \Omega, \quad (15.23)$$

because arc a and arc b are odd multiples of quarter wavelength, where

$Z_a$  = the impedance looking from port 1 toward arc a,

$Z_b$  = the impedance looking from port 1 toward arc b, and

$2^{1/2}Z_o$  = the characteristic impedance of the microstrip line.

- The impedance looking outward from port 1 is

$$Z_1 = Z_a // Z_b. \quad (15.24)$$

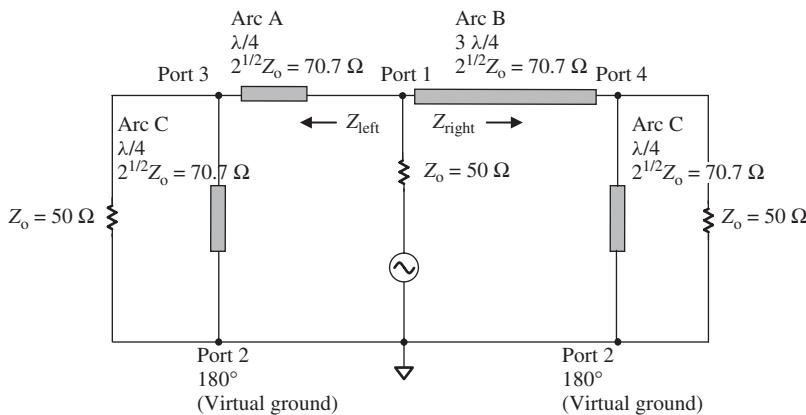


Figure 15.40. Equivalent circuit of ring microstrip balun.

From expression (15.23), we have

$$Z_1 = Z_0 = 50 \Omega. \quad (15.25)$$

It is therefore concluded that the impedance looking at port 1 into the ring is matched with the source impedance  $Z_0$ .

Relative to the virtual grounded port 2, the phase shift at port 1 is  $180^\circ$ . The phase shift at port 3 from the virtual grounded port 2 must be  $90^\circ$  because the length of arc a is one-quarter wavelength from port 1, while the phase shift at port 4 from the virtual grounded port 2 is  $-90^\circ$  or  $270^\circ$  because the length of arc b is one-quarter wavelength from port 2 but three-quarters wavelengths from port 1.

The ring microstrip line balun is also called the *rat-race* balun. It is a narrowband balun, with a relative bandwidth usually less than 15%. It has been employed extensively in the development of cellular phones and other communication systems today.

### 15.4.2 Split Ring Balun

For convenience in layout, the circular ring in a ring microstrip line balun can be cut into smaller pieces. Figure 15.41 is one design example, in which the ring is cut into 10 pieces. It is implemented on an alumina ceramic substrate. Usually, the alumina ceramic substrate has a higher permittivity, so that the size of ring balun can be reduced. In this example, the permittivity  $\epsilon_r$  of the alumina ceramic substrate is 10.5, while the permittivity of a plastic substrate is about 4.5. This means that the size of a module built by the ceramic substrate will be significantly reduced from that of a module built by a plastic substrate. The thickness of the alumina ceramic substrate is 25 mils.

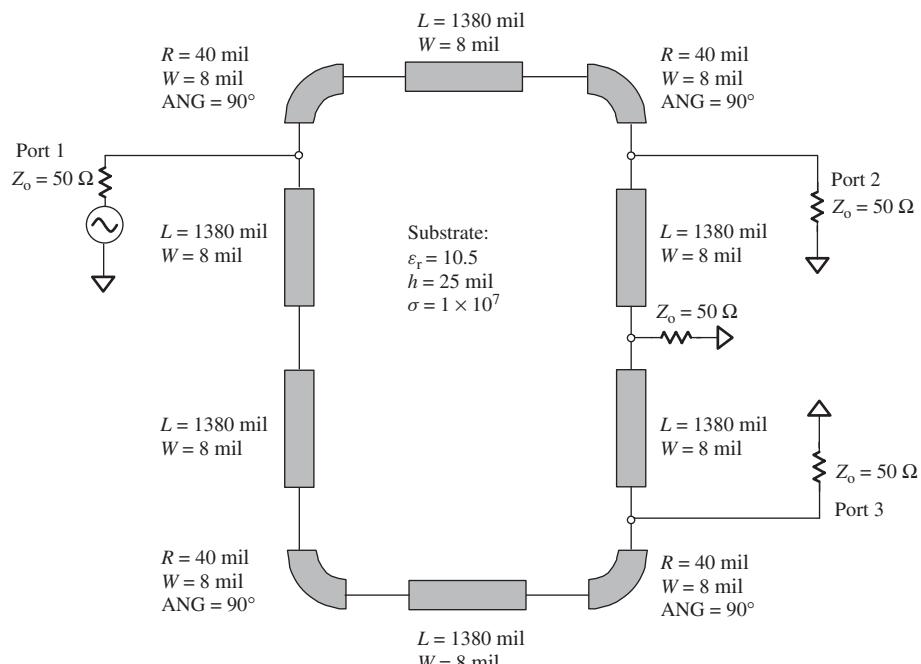


Figure 15.41. Split ring microstrip line balun;  $f = 800\text{--}900 \text{ MHz}$ .

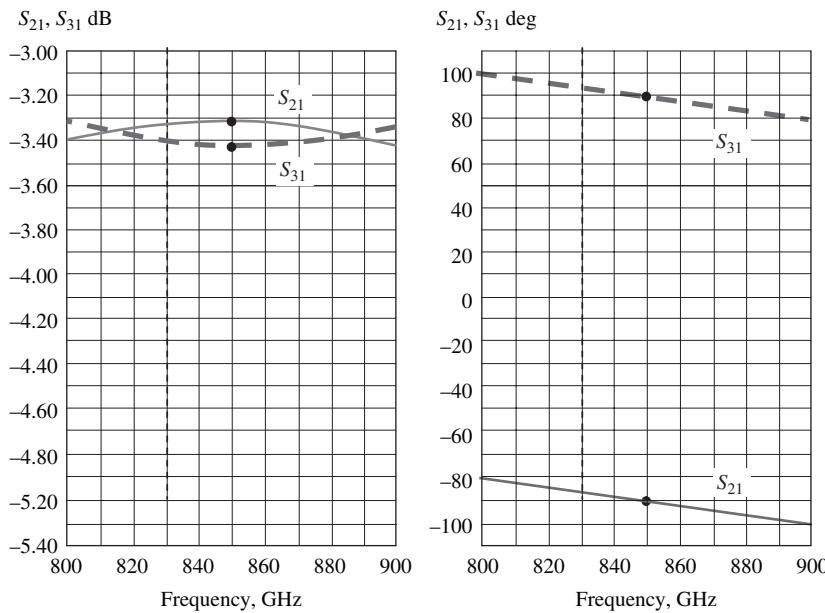


Figure 15.42. Insertion loss and phase shift of split ring balun;  $f = 800\text{--}900$  MHz.

The performance of this balun is shown in Figure 15.42. The magnitudes of  $S_{21}$  and  $S_{31}$  are between  $-3.58$  and  $-3.32$  dB in the frequency range from  $800$  to  $900$  MHz. Thus, the additional attenuation is only  $0.3\text{--}0.6$  dB throughout the entire frequency bandwidth. At the central frequency, namely,  $f = 850$  MHz, the phases of  $S_{21}$  and  $S_{31}$  are exactly  $90^\circ$  and  $-90^\circ$ , respectively, while at other frequencies the frequency response of the phase is tilted over the bandwidth. Interestingly, the tilted directions of the phases of  $S_{21}$  and  $S_{31}$  are the same, so that the phase difference is kept at around  $180^\circ$ .

## 15.5 MIXING TYPE OF BALUN

In order to reduce size, microstrip lines can be partially replaced by capacitors. In the first example, the microstrip line will be replaced by a chip capacitor. Furthermore, if all the microstrip lines are replaced by inductors and capacitors, the size will be significantly reduced, which will be demonstrated in the second example.

### 15.5.1 Balun Built by Microstrip Line and Chip Capacitor

In the balun shown in Figure 15.41, the total length of the microstrip line is 8440 mils or 8.44 in. This is too long for some smaller products. It is therefore required to reduce this length so that the size of product can be shrunk.

The balun shown in Figure 15.43 is modified from the balun shown in Figure 15.41, in which partial microstrip lines are replaced by six chip capacitors. The total length of the microstrip line now is 5280 mils or 5.28 in., so that 37.44% of the total length has been saved. Its performance is shown in Figure 15.44. Comparing Figures 15.44 and 15.42, it

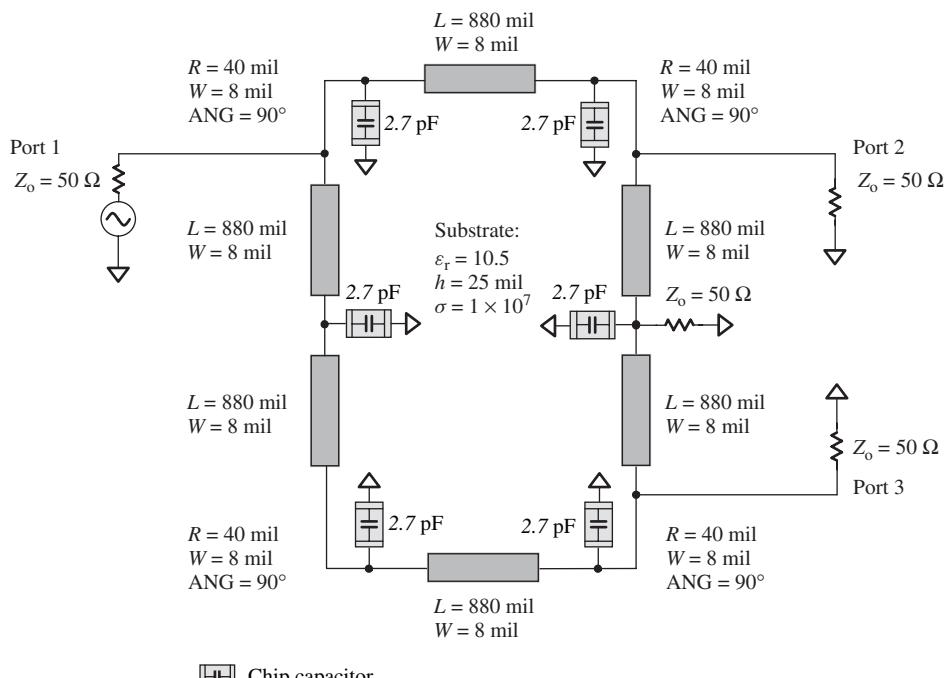


Figure 15.43. Modified split ring microstrip line balun;  $f = 800\text{--}900\text{ MHz}$ .

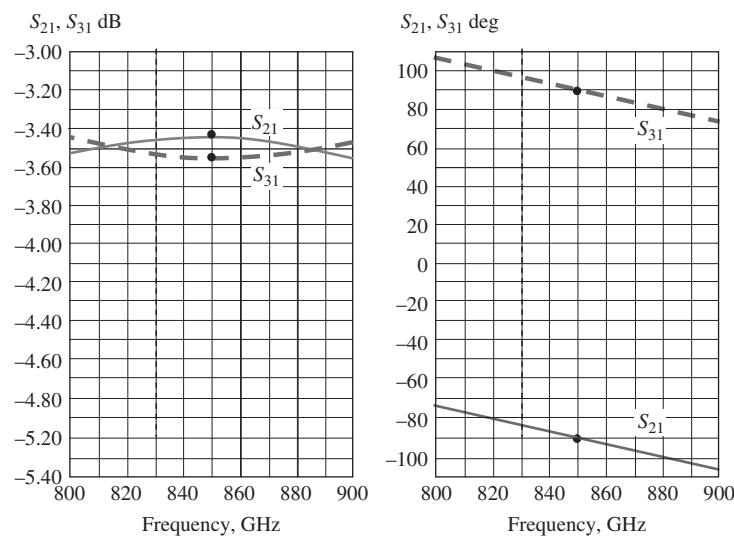


Figure 15.44. Insertion loss and phase shift of the modified split ring balun;  $f = 800\text{--}900\text{ MHz}$ .

is found they are similar. In the modified ring balun, the additional attenuation is about 0.1–0.2 dB and the phase shift is tilted 10°–20° more, but the difference of the phase in the differential pair is still kept at around 180°.

It should be noted that the actual model of chip capacitors is adapted in the simulation of the circuitry so that the actual performance is close to the simulation result, as shown in Figure 15.42.

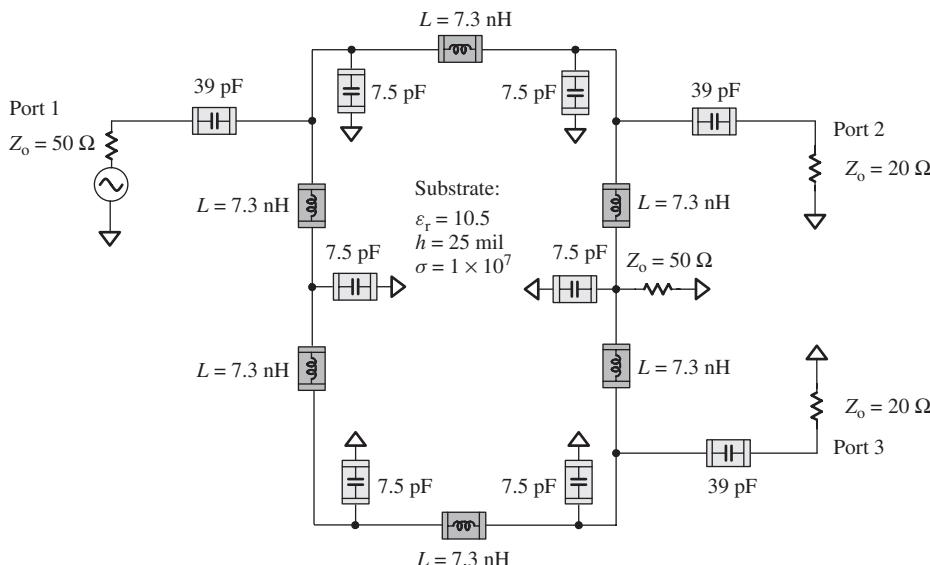
### 15.5.2 Balun Built by Chip Inductors and Chip Capacitors

The successful partial replacement of microstrip lines by chip capacitors shown in Figures 15.43 and 15.44 is encouraging. Is it possible for all the microstrip lines to be replaced by inductors and capacitors?

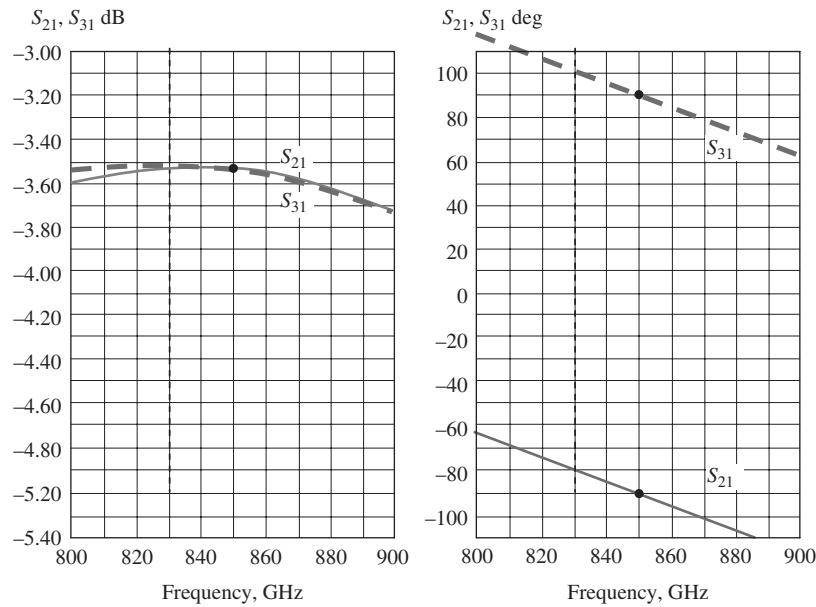
The answer is positive. Figure 15.45 shows a balun built by only chip inductors and capacitors. There are six chip inductors with 7.3 nH and six capacitors with 7.5 pF, which form a “ring.” The three capacitors with 39 pF are “zero” capacitors within the frequency bandwidth. In the simulation, actual models of chip inductor and chip capacitors are adapted so that the simulated performance and actual test results are closer together.

In Figure 15.45, the impedance at the single-ended port is 50 Ω, but the impedance at the differential pair ports is 20 Ω. This is an actual design for a differential PA and is implemented on the alumina ceramic substrate. Its performance is shown in Figure 15.46. The additional attenuation is about 0.6–0.8 dB and the phase shift is tilted more than that of the balun shown in Figure 15.43; however, the phase difference between differential pair ports is still kept around 180°.

This kind of balun should be quite useful in the products with RF module design.



**Figure 15.45.** Ring balun built by chip inductors and chip capacitors;  $f = 800\text{--}900 \text{ MHz}$ .



**Figure 15.46.** Insertion loss and phase shift of a balun built by chip inductors and chip capacitors;  $f = 800\text{--}900 \text{ MHz}$ .

## APPENDICES

### 15.A.1 Transformer Balun Built by Two Stacked Transformers

Figure 15.A.1 shows four ways to build a transformer balun by two stacked transformers with their turns ratio  $1 : n$ . The  $1\text{-G}\Omega$  resistor is a dummy part to satisfy the computer requirement for a DC circuit loop in either transformer 1 or transformer 2.

The relationship between  $R_L$ ,  $R_S$ , and the turns ratio  $n$  can be derived and expressed as follows:

1. **SS-DS Connection Mode** “Single-ended, primary windings stacked in Series, goes to Differential, secondary windings stacked in Series.”

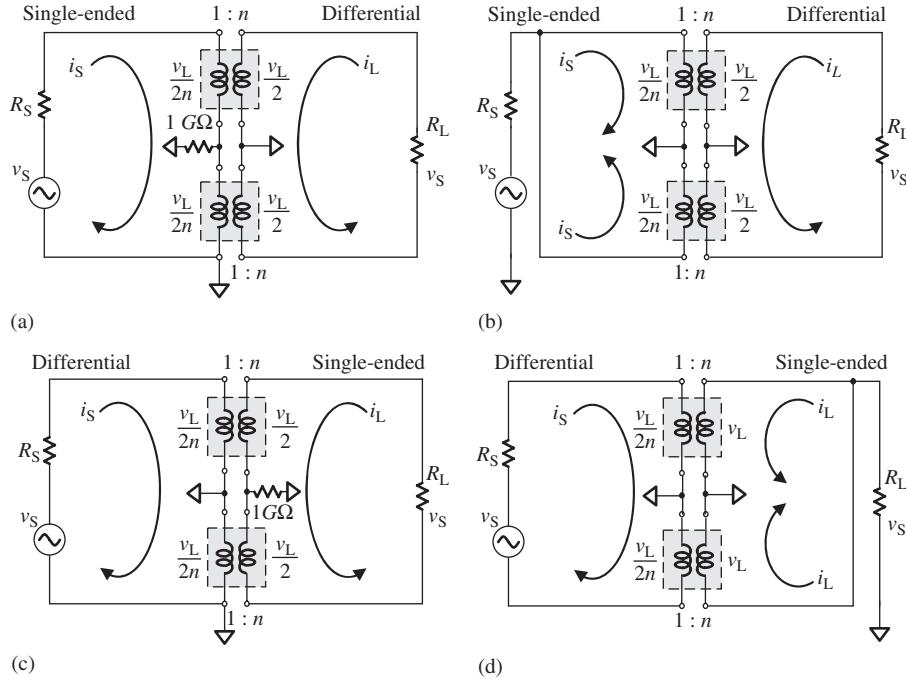
From Figure 15.A.1(a), we have

$$v_S = \frac{2v_L}{2n} = \frac{v_L}{n} = \frac{i_L R_L}{n} = \frac{i_S R_L}{n} = \frac{i_S R_L}{n^2}, \quad (15.A.1)$$

and then

$$R_S = \frac{v_S}{i_S} = \frac{R_L}{n^2}. \quad (15.A.2)$$

2. **SP-DS Connection Mode** “Single-ended, primary windings stacked in Parallel, goes to Differential, secondary windings stacked in Series.” From



**Figure 15.A.1.** Four connection modes to build a transformer balun by two stacked transformers.  
 (a) SS-DS connection mode. (b) SP-DS connection mode. (c) DS-SS connection mode. (d) DS-SP connection mode.

Figure 15.A.1(b), we have

$$v_S = \frac{v_L}{2n} = \frac{i_L R_L}{2n} = \frac{i_S R_L}{2n^2}, \quad (15.A.3)$$

and then,

$$R_S = \frac{v_S}{2i_S} = \frac{R_L}{4n^2}. \quad (15.A.4)$$

3. **DS-DS Connection Mode** “Differential, primary windings stacked in Series, goes to Single-ended, secondary windings stacked in Series.” From Figure 15.A.1(c), we have

$$v_S = \frac{2v_L}{2n} = \frac{v_L}{n} = \frac{i_L R_L}{n} = \frac{i_S R_L}{n n} = \frac{i_S R_L}{n^2}, \quad (15.A.5)$$

and then,

$$R_S = \frac{v_S}{i_S} = \frac{R_L}{n^2}. \quad (15.A.6)$$

4. **DS-SP Connection Mode** “Differential, primary windings stacked in Series, goes to Single-ended, secondary windings stacked in Parallel.” From Figure 15.A.1(d), we have

$$v_S = \frac{2v_L}{n} = 2\frac{2i_L R_L}{n} = 2\frac{2\frac{i_S}{n} R_L}{n} = 4\frac{i_S R_L}{n^2}, \quad (15.A.7)$$

and then,

$$R_S = \frac{v_S}{i_S} = 4\frac{R_L}{n^2}. \quad (15.A.8)$$

### 15.A.2 Analysis of a Simple LC Balun

The following analysis will be aimed at the splitter as shown in Figure 15.A.2, because the combiner is simply a reversed splitter. (This analysis was developed by the author in the 1990s and is abstracted from Li, 2005, pp. 220–225.)

There is a single port with its impedance  $Z_{sn}$  and two balanced ports with their impedance  $Z_{dp}$ . They are

$$Z_{sn} = R_{sn} + jX_{sn}, \quad (15.A.9)$$

$$Z_{dp} = R_{dp} + jX_{dp}, \quad (15.A.10)$$

where

$Z_{sn}$  = the impedance of a single-ended port,

$R_{sn}$  = the real part of a single-ended port impedance,

$X_{sn}$  = the imaginary part of a single-ended port impedance,

$Z_{dp}$  = the impedance of one differential pair port,

$R_{dp}$  = the real part of impedance at one differential pair port, and

$X_{dp}$  = the imaginary part of impedance at one differential pair port.

For the analysis, let us redraw Figure 15.A.2 as Figure 15.A.3.

The impedances looking into the right-hand and left-hand sides from the single-ended port are

$$Z_a = Z_c + Z_L // Z_{dp}, \quad (15.A.11)$$

$$Z_b = Z_L + Z_c // Z_{dp}, \quad (15.A.12)$$

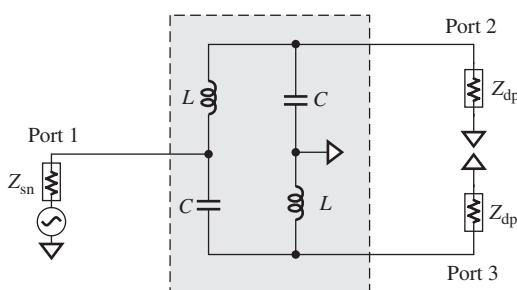
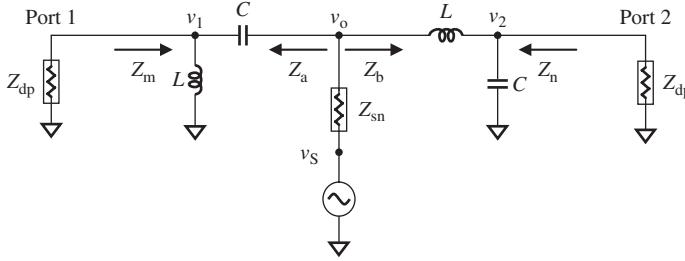


Figure 15.A.2. LC balun functioning as a splitter.



**Figure 15.A.3.** The splitter of Figure 15.A.2 redrawn.

where

$$Z_c = \frac{1}{j\omega C}, \quad (15.A.13)$$

$$Z_L = j\omega L, \quad (15.A.14)$$

and  $\omega$  = the operating angular frequency.

Then,

$$Z_a = \frac{(1 - \omega^2 LC) Z_{dp} + j\omega L}{-\omega^2 LC + j\omega CZ_{dp}}, \quad (15.A.15)$$

$$Z_b = \frac{(1 - \omega^2 LC) Z_{dp} + j\omega L}{1 + j\omega CZ_{dp}}. \quad (15.A.16)$$

These two impedances are connected together in parallel from the single-ended port to the ground. The resulting impedance must be matched and is equal to the conjugated value of  $Z_{sn}$ , that is,

$$Z_{sn}^* = R_{sn} - jX_{sn} = \frac{Z_a Z_b}{Z_a + Z_b}. \quad (15.A.17)$$

By substituting (15.A.15) and (15.A.16) into (15.A.17), we have

$$Z_{sn}^* = \frac{(1 - \omega^2 LC) Z_{dp} + j\omega L}{(1 - \omega^2 LC) + j2\omega CZ_{dp}}, \quad (15.A.18)$$

$$Z_{sn}^* = \frac{(1 - \omega^2 LC) R_{dp} + j[(1 - \omega^2 LC) X_{dp} + \omega L]}{(1 - \omega^2 LC - 2\omega CX_{dp}) + j2\omega CZ_{dp}}. \quad (15.A.19)$$

From (15.A.17) and (15.A.19),

$$R_{sn} = \frac{(1 - \omega^2 LC) R_{dp} (1 - \omega^2 LC - 2\omega CX_{dp}) + 2\omega CR_{dp} [(1 - \omega^2 LC) X_{dp} + \omega L]}{(1 - \omega^2 LC - 2\omega CX_{dp})^2 + (2\omega CR_{dp})^2}, \quad (15.A.20)$$

$$X_{sn} = \frac{2\omega CR_{dp} (1 - \omega^2 LC) R_{dp} - (1 - \omega^2 LC - 2\omega CX_{dp}) [(1 - \omega^2 LC) X_{dp} + \omega L]}{(1 - \omega^2 LC - 2\omega CX_{dp})^2 + (2\omega CR_{dp})^2}. \quad (15.A.21)$$

The expressions (15.A.19)–(15.A.21) can be simplified if

$$\omega^2 LC = 1. \quad (15.A.22)$$

This condition is expected. The power can be well transmitted from the single port to the differential ports or vice versa only when the balun core is a resonant-type core. Under condition (15.A.22), expressions (15.A.15), (15.A.16), and (15.A.19), (15.A.20), and (15.A.21) can be simplified as

$$Z_a = \frac{j\omega L}{-1 + j\omega CZ_{dp}}, \quad (15.A.23)$$

$$Z_b = \frac{j\omega L}{1 + j\omega CZ_{dp}}, \quad (15.A.24)$$

$$Z_{sn}^* = \frac{L}{2CZ_{dp}}, \quad (15.A.25)$$

$$R_{sn} = \frac{L}{2C} \frac{R_{dp}}{(R_{dp}^2 + X_{dp}^2)}, \quad (15.A.26)$$

$$X_{sn} = \frac{L}{2C} \frac{X_{dp}}{(R_{dp}^2 + X_{dp}^2)}. \quad (15.A.27)$$

From expressions (15.A.22) and (15.A.25), two conclusive expressions listed below are obtained:

$$\omega L = \sqrt{2Z_{dp}Z_{sn}^*}, \quad (15.A.28)$$

$$\omega C = \frac{1}{\sqrt{2Z_{dp}Z_{sn}^*}}. \quad (15.A.29)$$

These two equations are very useful to the design because the values of  $L$  and  $C$  can be easily calculated from the given impedances  $Z_{dp}$  and  $Z_{sn}$ .

Now, let us examine the impedance looking from the balanced ports,  $Z_m$  and  $Z_n$ .

$$Z_m = (Z_b // Z_{sn} + Z_c) // Z_L, \quad (15.A.30)$$

$$Z_n = (Z_a // Z_{sn} + Z_L) // Z_c. \quad (15.A.31)$$

From expressions (15.A.11), (15.A.12) and (15.A.28), (15.A.29),

$$Z_m = \left(1 + \frac{2Z_{sn}^*}{Z_{sn}}\right) Z_{dp}, \quad (15.A.32)$$

$$Z_n = \left(1 + \frac{2Z_{sn}^*}{Z_{sn}}\right) Z_{dp}. \quad (15.A.33)$$

Therefore,

$$Z_m = Z_n, \quad (15.A.34)$$

which proves that the two ports are balanced.

It should be noted that, from equation (15.A.32) or (15.A.33),

$$Z_m = Z_n = 3Z_{dp}, \quad (15.A.35)$$

and if  $Z_{sn}$  is a pure resistor,

$$Z_{sn} = R_{sn}. \quad (15.A.36)$$

Expression (15.A.35) shows that  $Z_m$  and  $Z_n$  are not conjugate-matched to the balanced impedance of the two balanced ports, respectively. This is not surprising because the LC balun is a network with three ports.

Furthermore, we will examine the phase relation between the single-ended port and the differential pair ports. Referring to Figure 15.A.3,

$$v_1 = \frac{Z_{dp}/Z_L}{Z_{dp}/Z_L + Z_c v_o}, \quad (15.A.37)$$

$$v_2 = \frac{Z_{dp}/Z_c}{Z_{dp}/Z_c + Z_L} v_o. \quad (15.A.38)$$

By means of equations (15.A.11)–(15.A.14), (15.A.28), and (15.A.29), and noting that

$$v_o = \frac{Z_{sn}^*}{Z_{sn} + Z_{sn}^*} v_s, \quad (15.A.39)$$

we have

$$v_1 = j \frac{\sqrt{2(R_{dp}R_{sn} + X_{dp}X_{sn})}}{4R_{dp}} v_s, \quad (15.A.40)$$

$$v_2 = -j \frac{\sqrt{2(R_{dp}R_{sn} + X_{dp}X_{sn})}}{4R_{dp}} v_s. \quad (15.A.41)$$

It means that the voltage at differential pair port 1 is  $90^\circ$  ahead of that at the single-ended port and the voltage at the differential pair port 2 is  $90^\circ$  behind that at the single-ended port. It is therefore concluded that the two pair ports are differential as long as condition (15.A.22) is satisfied.

### 15.A.3 Example of Calculating of $L$ and $C$ Values for a Simple LC Balun

TABLE 15.A.1 Calculated  $L$  and  $C$  values of a simple  $LC$  balun for Groups 1 and 3 in the UWB system

$Z_{sn}$	$Z_{dp}$	$(2 \times Z_{sn}Z_{dp})^{0.5}$	$f$ , Hz	Angular $f$ , rad/s	$L$ , nH	$C$ , pF
<b>Group 1</b>						
50	50	70.71067812	3,168,000,000	19,905,177,600	3.55	0.71
50	50	70.71067812	3,432,000,000	21,563,942,400	3.27	0.65
50	50	70.71067812	3,696,000,000	23,222,707,200	3.04	0.60
50	50	70.71067812	3,696,000,000	23,222,707,200	3.04	0.60
50	50	70.71067812	3,960,000,000	24,881,472,000	2.84	0.56
50	50	70.71067812	4,224,000,000	26,540,236,800	2.66	0.53
50	50	70.71067812	4,224,000,000	26,540,236,800	2.66	0.53
50	50	70.71067812	4,488,000,000	28,199,001,600	2.50	0.50
50	50	70.71067812	4,752,000,000	29,857,766,400	2.36	0.47
<b>Group 3</b>						
50	50	70.71067812	6,336,000,000	39,810,355,200	1.77	0.35
50	50	70.71067812	6,600,000,000	41,469,120,000	1.70	0.34
50	50	70.71067812	6,864,000,000	43,127,884,800	1.63	0.32
50	50	70.71067812	6,864,000,000	43,127,884,800	1.63	0.32
50	50	70.71067812	7,128,000,000	44,786,649,600	1.57	0.31
50	50	70.71067812	7,392,000,000	46,445,414,400	1.52	0.30
50	50	70.71067812	7,392,000,000	46,445,414,400	1.52	0.30
50	50	70.71067812	7,656,000,000	48,104,179,200	1.46	0.29
50	50	70.71067812	7,920,000,000	49,762,944,000	1.42	0.28

### 15.A.4 Equivalence of Parts between Single-Ended and Differential Pair with Respect to a Simple LC Balun

Just like the transformer balun, the equivalence of parts between the single-ended and differential pair also exists in a simple  $LC$  balun, so that these parts can be interpreted from each other in terms of the following rules:

- An inductor  $L/2$  in parallel at the single side is interpreted as an inductor  $L$  in series at the differential branch, or vice versa.
- A capacitor  $2C$  in parallel at the single side is interpreted as a capacitor  $C$  in series at the differential branch, or vice versa.
- An inductor  $L/2$  in series at the single side is interpreted as an inductor  $L$  in parallel at the differential branch, or vice versa.
- A capacitor  $2C$  in series at the single side is interpreted as a capacitor  $C$  in parallel at the differential branch, or vice versa.

- The value of the inductor or capacitor at the differential branch is determined by the impedances of the single-ended source and the differential pair load, such that,

$$L\omega_o = \sqrt{2Z_{dp}Z_{sn}^*}, \quad (15.A.28)$$

and

$$C\omega_o = \frac{1}{\sqrt{2Z_{dp}Z_{sn}^*}}, \quad (15.A.29)$$

where

- $L$  = the inductance of the inductor in the simple LC balun,
- $C$  = the capacitance of the capacitor in the simple LC balun,
- $Z_{sn}$  = the source impedance at the single-ended port,
- $Z_{dp}$  = the load impedance at each port of the differential pair, and
- $\omega_o$  = the operating angular frequency.

It should be noted that unlike the interpretation in transformer balun, if a capacitor is added to the single-ended side, then the corresponding part to be added to the differential pair side is also a capacitor; likewise, if a inductor is added to the single-ended side, then the corresponding part to be added to the differential pair side is also an inductor.

Figure 15.A.4 shows five cases in which a part is added to the single-ended port and the interpreted part is added to each port of the differential pair, respectively. Formulas (15.8) and (15.9) define the relationship between the added parts, by which the added parts can be interpreted from each other. Simulations prove that, before and after the parts are added to the transformer balun, the *S*-parameters or impedances at all three ports are unchanged at the operating frequency  $\omega_o$  or  $f_o$  and are mostly unchanged or only changed slightly within a certain bandwidth around the operating frequency.

In order to verify the interpretation regulation and to check the equivalence of the impedances looking into and looking outward from the transformer balun, let us conduct another five experimental simulations for the verification of the interpretation regulation. The number of parts added will be increased from 0 to 4 in steps so that they are not the same as shown in Figure 15.A.4, where only one part is added to either the single-ended port or each port of the differential pair.

Simulation A is conducted for the original LC balun with no added parts.

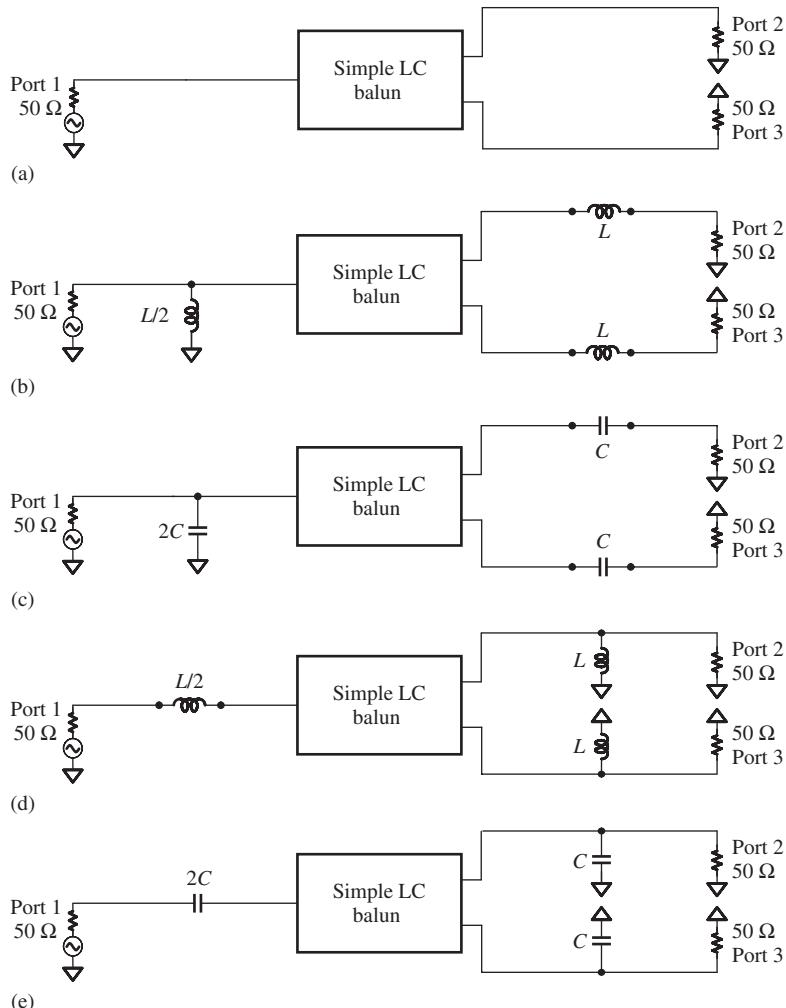
Simulation B is conducted for the original LC balun with one added part.

Simulation C is conducted for the original LC balun with two added parts.

Simulation D is conducted for the original LC balun with three added parts.

Simulation E is conducted for the original LC balun with four added parts.

It should be noted that



**Figure 15.A.4.** Five ways to add parts to a simple LC balun, in which the  $S$ -parameters or impedances at all three ports as shown in this figure are kept unchanged before and after these parts are added (note that  $LC = 1/\omega_0^2$ ). (a) Nothing is added to the single-ended side. (b) An inductor is added to the single-ended side in parallel. (c) An capacitor is added to the single-ended side in parallel. (d) An inductor is added to the single-ended side in series. (e) A capacitor is added to the single-ended side in series.

- in the case of a simple LC balun, either  $L$  or  $C$  must be selected using formulas (15.8) and (15.9), which are dependent on the impedances of the single-ended and differential pair ports;
- on the contrary, in the case of the transformer balun, either  $L$  or  $C$  can be selected with any value as long as the following relation is satisfied:

$$L\omega_0 = \frac{1}{C\omega_0}, \quad (15.A.42)$$

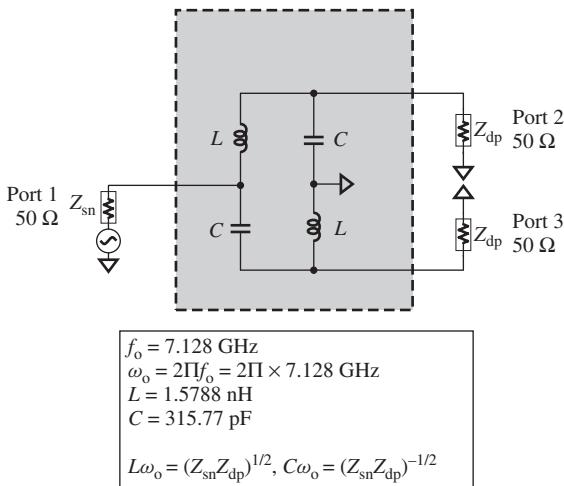


Figure 15.A.5. Simulation A: LC balun with no added parts.

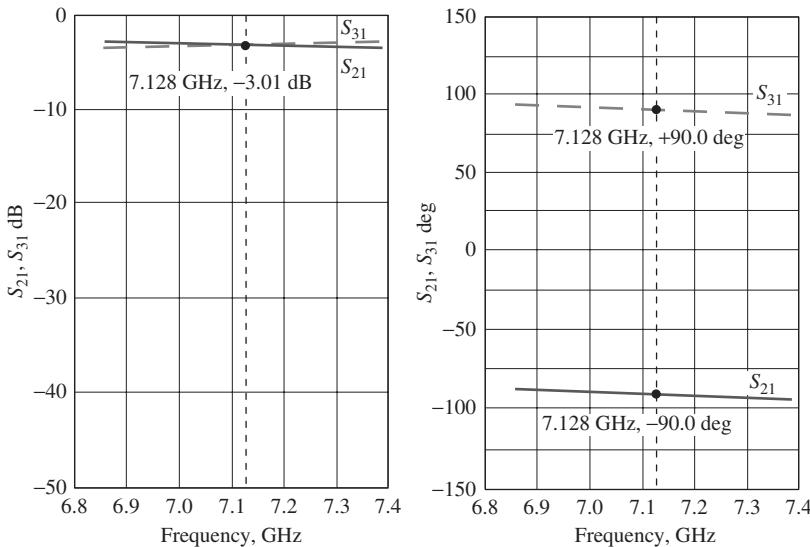


Figure 15.A.6. Insertion loss and phase shift of LC balun for a group 3 UWB system ( $S_{21}, S_{31}$ ) in simulation A: LC balun with no added parts.

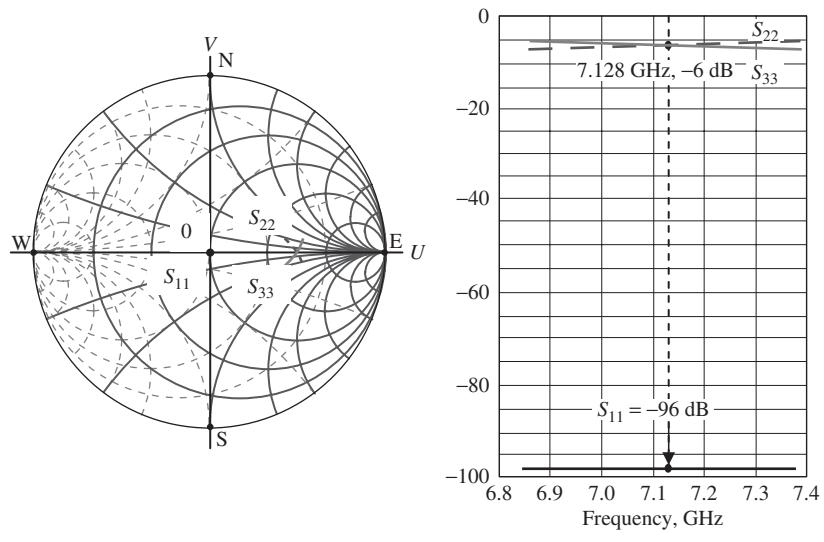
or

$$LC = \frac{1}{\omega_0^2}, \quad (15.A.43)$$

where  $\omega_0$  = the operating angular frequency.

This leads the essential difference about interpretation in the transformer balun and in the simple LC balun;

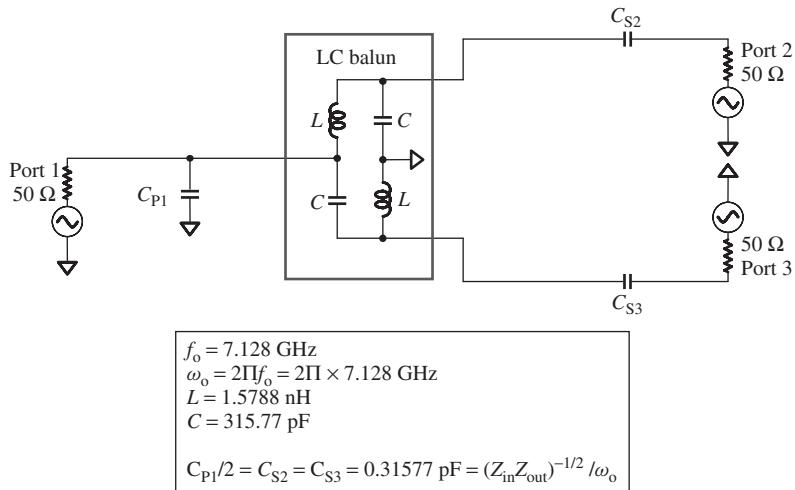
- in the case of transformer balun, the interpretation can be made for any value of  $L$  or  $C$  as long as both of them obey the relation (15.5) or (15.6);



(a)

(b)

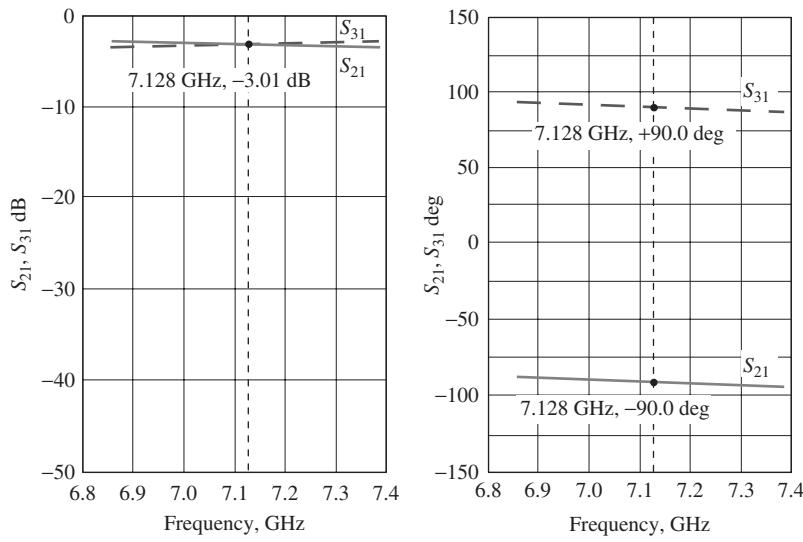
**Figure 15.A.7.** Return loss of LC balun for a group 3 UWB system ( $S_{11}$ ,  $S_{22}$ ,  $S_{33}$ ) in simulation A: LC balun with no added parts. (a) On Smith Chart. (b) By rectangular coordination [dB].



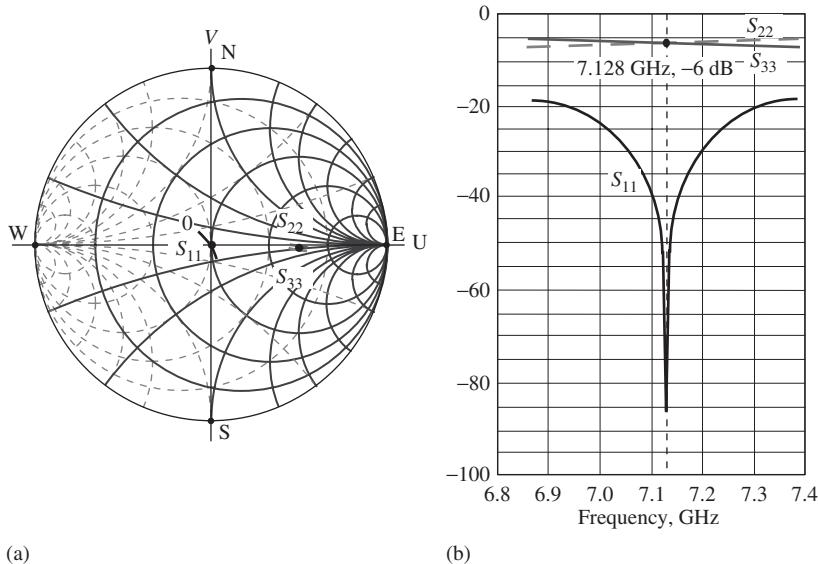
**Figure 15.A.8.** Simulation B: LC balun with one added part.

- In the case of a simple LC balun, the interpretation can be conducted only for values of  $L$  and  $C$  calculated from formulas (15.8) and (15.9), respectively.

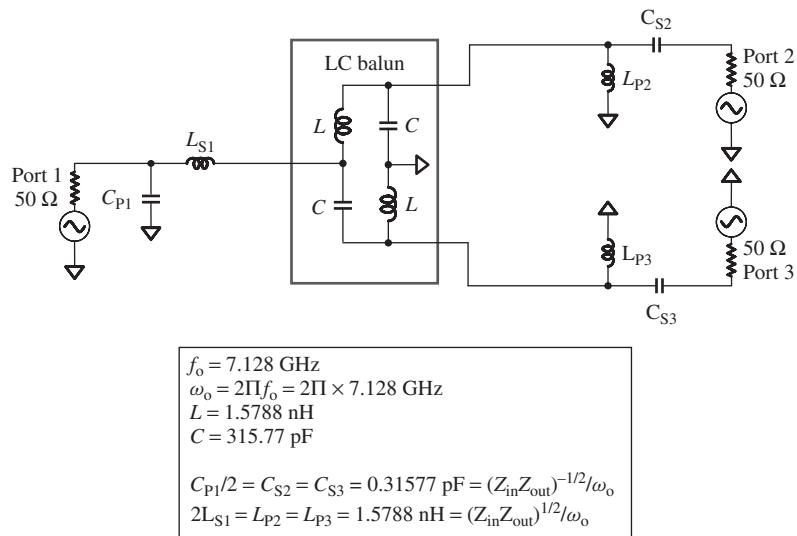
Nevertheless, the interpretation feature of a LC balun does exist and might be applied for some other purpose. It is therefore kept as one of the appendices to this chapter.



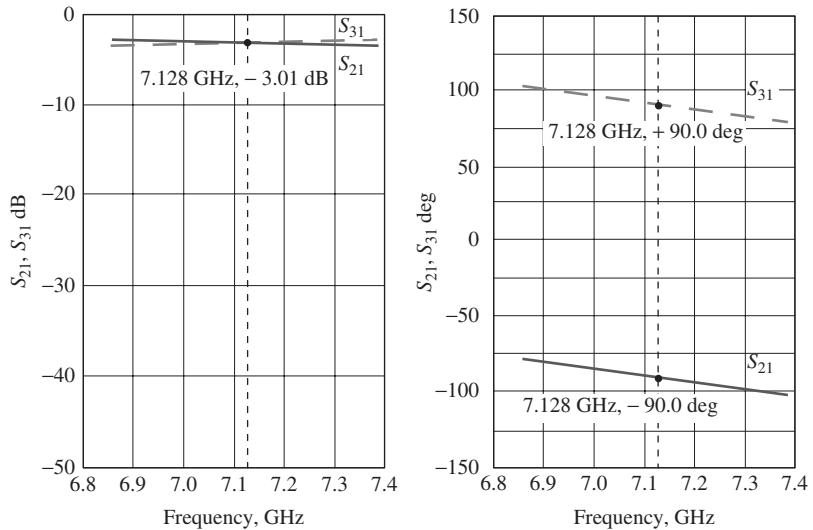
**Figure 15.A.9.** Insertion loss and phase shift of LC balun for a group 3 UWB system ( $S_{21}, S_{31}$ ) in simulation B: LC balun with one added part.



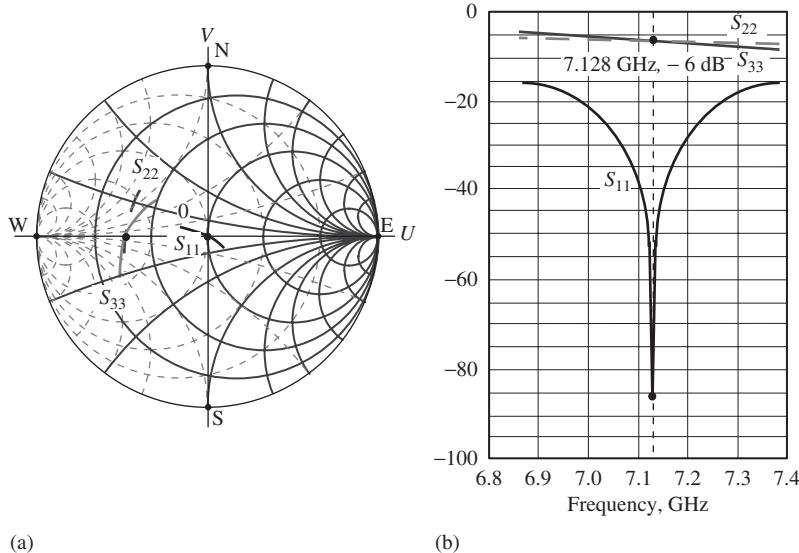
**Figure 15.A.10.** Return loss of LC balun for a group 3 UWB system ( $S_{11}, S_{22}, S_{33}$ ) in simulation B: LC balun with one added part. (a) On Smith Chart. (b) By rectangular coordination [dB].



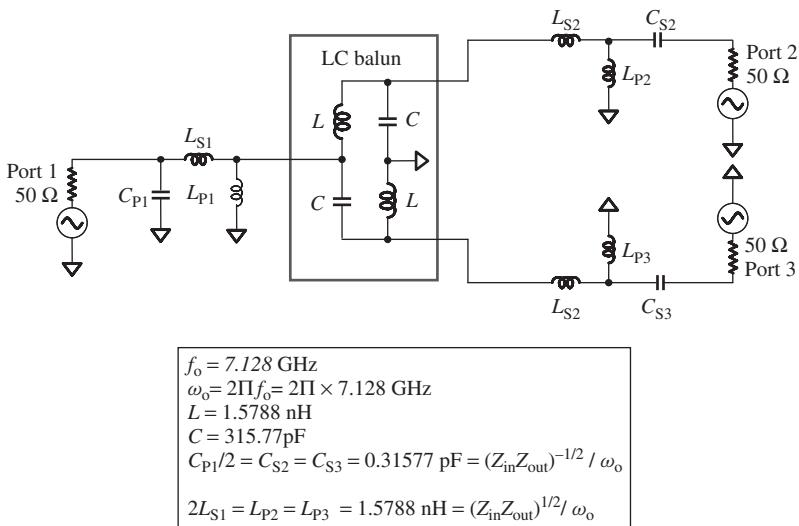
**Figure 15.A.11.** Simulation C: LC balun with two added parts  $C_{P1}/2 = C_{S2} = C_{S3}$  and  $2L_{S1} = L_{P2} = L_{P3}$ .



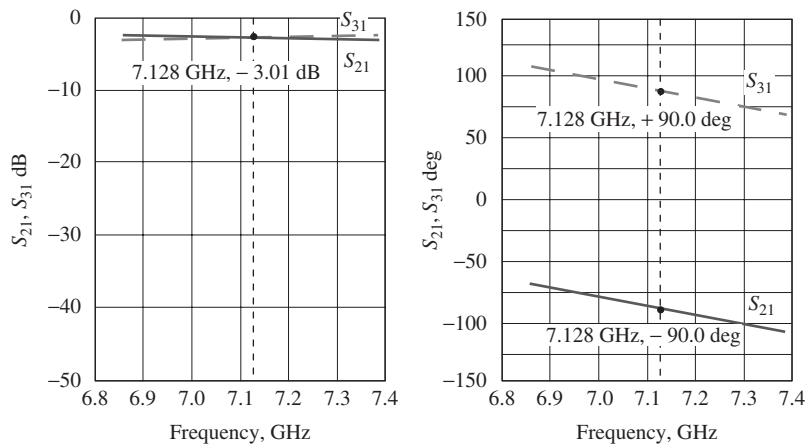
**Figure 15.A.12.** Insertion loss and phase shift of LC balun for a group 3 UWB system ( $S_{21}, S_{31}$ ) in simulation C: LC balun with two added parts  $C_{P1}/2 = C_{S2} = C_{S3}$  and  $2L_{S1} = L_{P2} = L_{P3}$ .



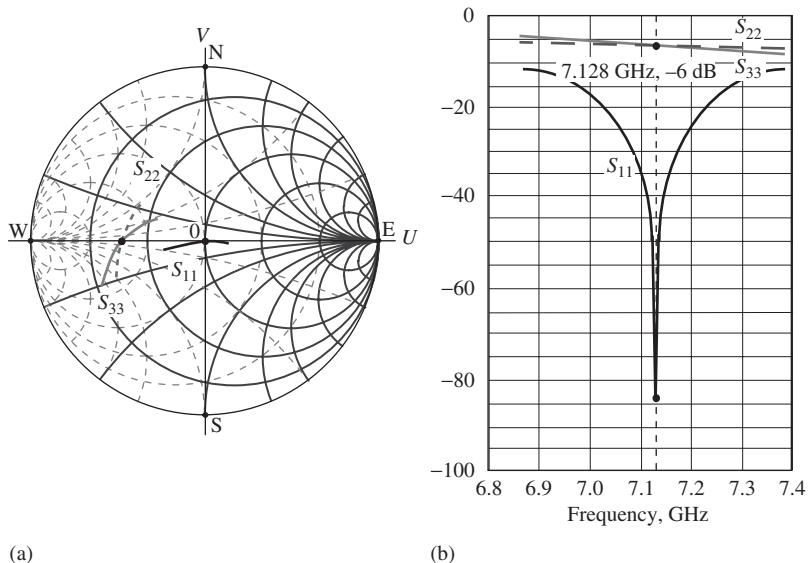
**Figure 15.A.13.** Return loss of LC balun for a group 3 UWB system ( $S_{11}, S_{22}, S_{33}$ ) in simulation C: LC balun with two added parts  $C_{P1}/2 = C_{S2} = C_{S3}$  and  $2L_{S1} = L_{P2} = L_{P3}$ . (a) On Smith Chart. (b) By rectangular coordination [dB].



**Figure 15.A.14.** Simulation D: LC balun with three added parts  $C_{P1}/2 = C_{S2} = C_{S3}$ ,  $2L_{S1} = L_{P2} = L_{P3}$ , and  $2L_{P1} = L_{S2} = L_{S3}$ .



**Figure 15.A.15.** Insertion loss and phase shift of LC balun for a group 3 UWB system ( $S_{21}, S_{31}$ ) in simulation D: LC balun with three added parts  $C_{P1}/2 = C_{S2} = C_{S3}$ ,  $2L_{S1} = L_{P2} = L_{P3}$ , and  $2L_{P1} = L_{S2} = L_{S3}$ .



**Figure 15.A.16.** Return loss of LC balun for a group 3 UWB system ( $S_{11}, S_{22}, S_{33}$ ) in simulation D: LC balun with three added parts  $C_{P1}/2 = C_{S2} = C_{S3}$ ,  $2L_{S1} = L_{P2} = L_{P3}$ , and  $2L_{P1} = L_{S2} = L_{S3}$ . (a) On Smith Chart. (b) By rectangular coordination [dB].

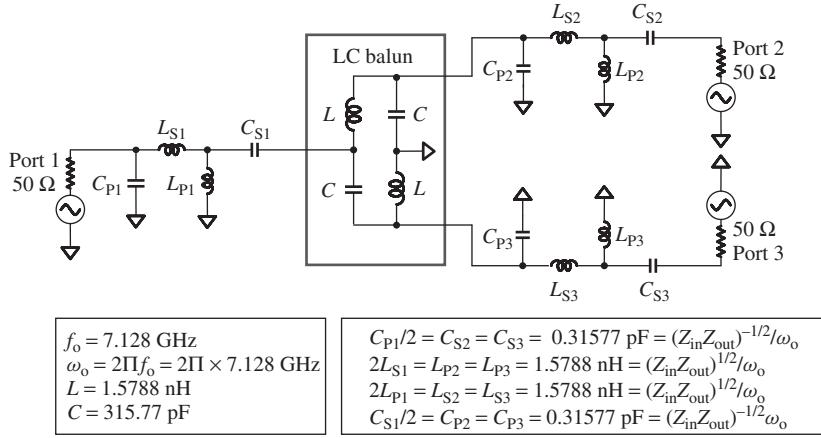


Figure 15.A.17. Simulation E: LC balun with four added parts  $C_{P1}/2 = C_{S2} = C_{S3}$ ,  $2L_{S1} = L_{P2} = L_{P3}$ ,  $2L_{P1} = L_{S2} = L_{S3}$ , and  $C_{S1}/2 = C_{P2} = C_{P3}$ .

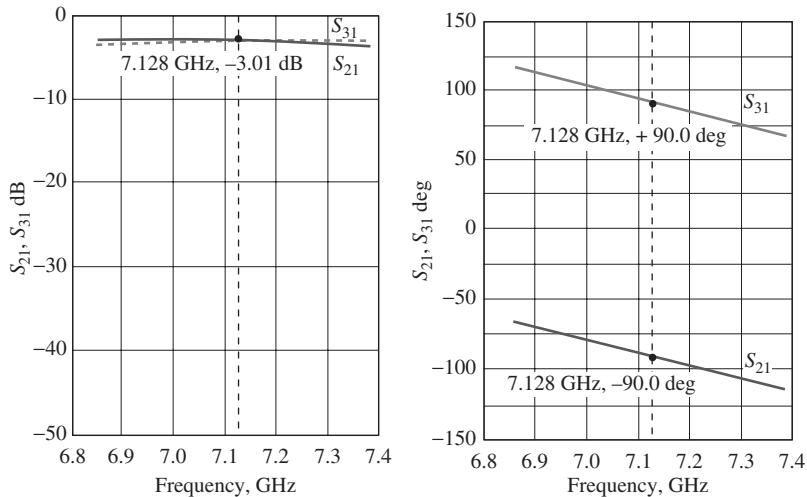
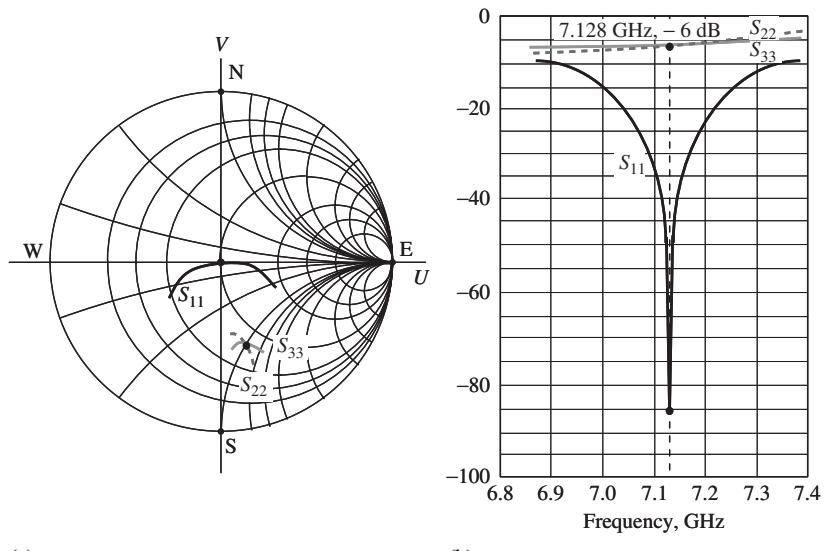


Figure 15.A.18. Insertion loss and phase shift of LC balun for a group 3 UWB system ( $S_{21}, S_{31}$ ) in simulation E: LC balun with four added parts  $C_{P1}/2 = C_{S2} = C_{S3}$ ,  $2L_{S1} = L_{P2} = L_{P3}$ ,  $2L_{P1} = L_{S2} = L_{S3}$ , and  $C_{S1}/2 = C_{P2} = C_{P3}$ .



**Figure 15.A.19.** Return loss of LC balun for a group 3 UWB system ( $S_{11}, S_{22}, S_{33}$ ) in simulation E: LC balun with four added parts  $C_{P1}/2 = C_{S2} = C_{S3}$ ,  $2L_{S1} = L_{P2} = L_{P3}$ ,  $2L_{P1} = L_{S2} = L_{S3}$ ,  $C_{S1}/2 = C_{P2} = C_{P3}$ . (a) On Smith Chart. (b) By rectangular coordination [dB].

## 15.A.5 Some Useful Couplers

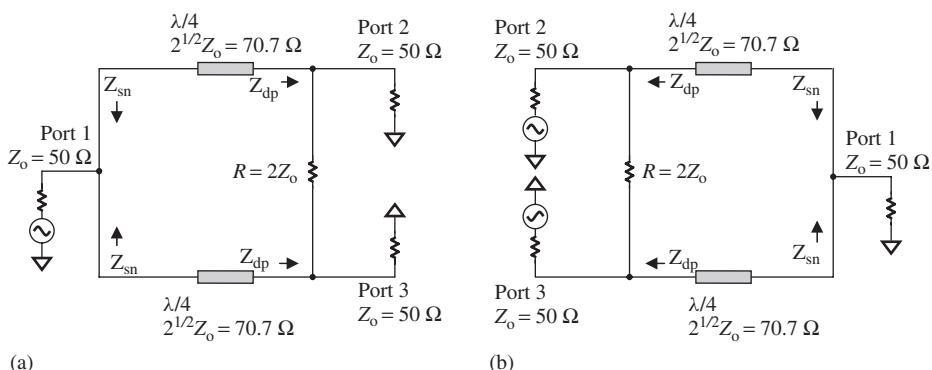
**15.A.5.1 Wilkinson Coupler.** The Wilkinson coupler is a balun with a very simple configuration. As shown in Figure 15.A.20, it consists of only two microstrip lines of a quarter-wavelength length and a  $100\Omega$  resistor. Figure 15.A.20(a) and (b) shows a Wilkinson splitter and combiner, respectively.

The impedance  $Z_{dp}$  looking from one of the microstrip lines toward the adjacent differential port is

$$Z_{dp} = Z_0 // (2Z_0 + Z_0 // Z_{dp}). \quad (15.A.44)$$

The impedance  $Z_{sn}$  looking from the microstrip line toward single-ended port is

$$Z_{sn}Z_{dp} = (\sqrt{2}Z_0)^2. \quad (15.A.45)$$



**Figure 15.A.20.** Wilkinson coupler. (a) Wilkinson splitter. (b) Wilkinson combiner.

From (15.A.44), we have

$$Z_{dp} = \frac{Z_0}{\sqrt{2}}. \quad (15.A.46)$$

From (15.A.45) and (15.A.46), we have

$$Z_{sn} = 2\sqrt{2}Z_0. \quad (15.A.47)$$

**15.A.5.2 Microstrip Lange Coupler.** The microstrip Lange coupler is a quadrature hybrid balun. Figure 15.A.21 and (b) shows its four-strip and six configurations, respectively. The coupling between strips is very tight due to the interdigitated structure. Consequently, it achieves a very wide bandwidth performance. The relative bandwidth reached can be greater than 100%. For instance, if the operating frequency of a Lange coupler is 5–20 GHz, then the relative bandwidth is  $\Delta f/f_o = 15/12.5 = 120\%$ .

Over a wide frequency range, the phase performance is very close to  $90^\circ$  and the variation of magnitude is usually less than 2 dB.

### 15.A.6 Cable Balun

The cable balun was developed a couple of decades ago. At present, it is almost never employed in actual circuitry because of its large size.

However, it is still sometimes useful in the test laboratory where the size of a part is not a problem. Figure 15.A.22 shows a simplest cable balun. The length of the cable should be a quarter wavelength of the operating frequency. For  $f = 850$  MHz, a regular RF cable of approximately 2.3 in. is required. This can be verified and tested by the setup as shown in Figure 15.A.23.

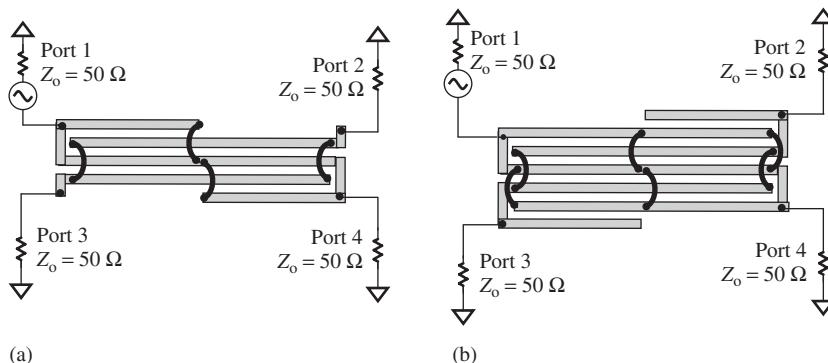


Figure 15.A.21. Microstrip Lange coupler. (a) Four-strip Lange coupler. (b) six-strip Lange coupler.

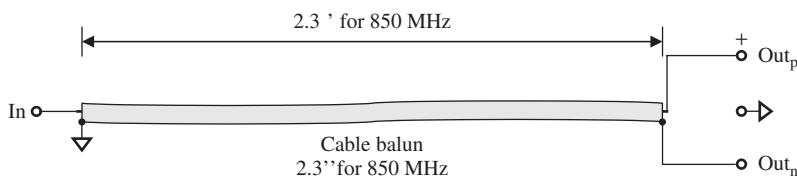


Figure 15.A.22. A simplest cable balun.

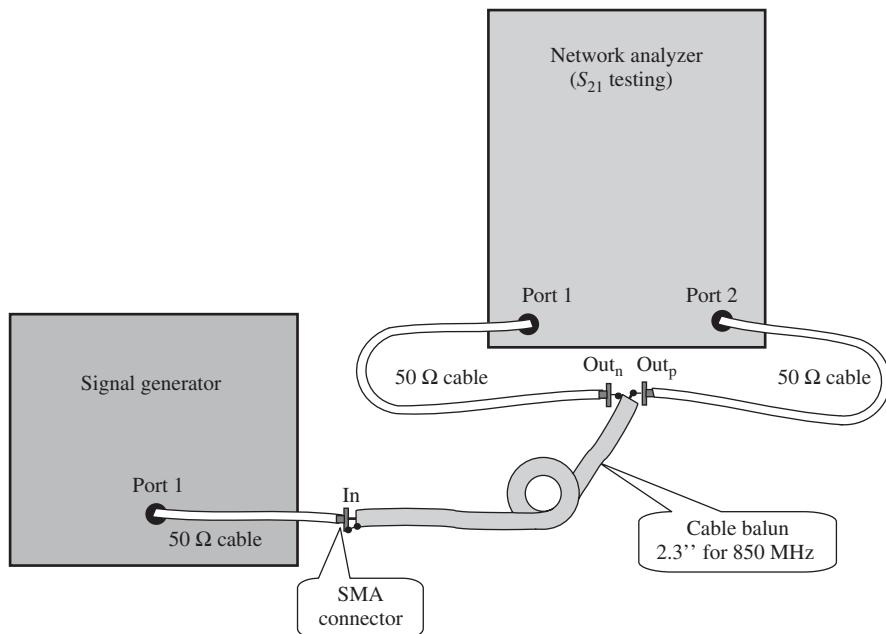


Figure 15.A.23. Verification of a simplest cable balun.

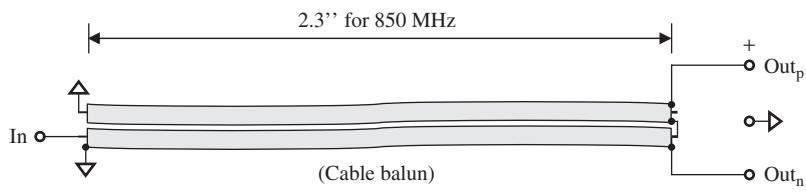


Figure 15.A.24. Cable balun with additional cable for compensation.

The simplest cable balun is not very well balanced and can be improved as shown as Figure 15.A.24.

## REFERENCE

Li R. Key issues in RF/RFIC circuit design. Beijing: Higher Education press; 2005.

## FURTHER READING

- Antoniades MA, Eleftheriades GV. A broadband wilkinson balun using microstrip metamaterial lines. *IEEE Antennas Wirel Propag Lett* 2005;4:209–212.
- Bakalski W, Simbirgerl W, Knapp H, et al. Lumped and Distributed Lattice-type LC-Baluns. *IEEE MTT-S Digest*, Seattle, WA; 2002. pp. 209–212.
- Costantini A, Lawrence B, Mahon S, et al. Broadband active and passive balun circuits: functional blocks for modern millimeter-wave radio architectures. *Proceedings of the 1st European Microwave Integrated Circuits Conference*, Manchester; 2006 Sep; pp. 421–424.

- Huang KY, Hsu CJ, Leu LY. Modeling methodology of integrated five-port balun using two-port RF measurement. 2005 IEEE Radio Frequency Integrated Circuits Symposium; 2005 pp. 295–298.
- Kawashima M, Nakagawa T, Araki K. A novel broadband active balun. 33rd European Microwave Conference; Munich: 2003. pp. 495–498.
- Lee JW, Webb KJ. Analysis and design of low-loss planar microwave baluns having three symmetric coupled lines. Volume 1, IEEE MTT-S International Microwave Symposium Digest, Seattle, WA; 2002 June 2–7. pp 117–120.
- Li RC. Double line RF balun transformer US patent 5,477,204. Motorola Inc; 1993.
- Rogers J, Bhatia R. A 6 to 20 GHz planar balun using a wilkinson divider and lange couplers. Volume 2, Microwave Symposium Digest, IEEE MTT-S International, Seattle, WA; 1991 June 10–14. pp 865–868.
- Rotella F, Tkachenko G, Cheng Y. Characterization, design, modeling, and model validation of silicon on-wafer M:N Balun components under match and un-matched conditions. 2005 IEEE Radio Frequency Integrated Circuits Symposium; 2005. pp. 291–294.
- Rutkowski T, Zieniutycz W, Joachimowski K. Wideband coaxial balun for antenna application. Volume 2, 12th International Conference on Microwaves and Radar, MIKON '98, Krakow; 1998 May 20–22; 1998. pp. 389–392.
- Sheen JW, Tang CW. LTCC-MLC Balun for WLAN/Bluetooth. IEEE MIT-S Digest, Pheonix, AZ; 2001. pp. 315–318.
- Sun JS, Lee TL. Design of a planar balun. Volume 2, Microwave Conference, APMC 2001, Asia-Pacific; 2001 Dec 3–6. pp. 535–538.
- Tao L, Gillis J, Wang D, et al. Design and modeling of compact on-chip transformer/balun using multi-level metal windings for RF integrated circuits. Radio Frequency Integrated Circuits (RFIC) Symposium, IEEE, Pheonix, AZ; 2001 May 20–22.
- Tiiharju E, Halonen KAI. An active differential broad-band phase splitter for quadrature-modulator applications. IEEE Trans Microwave Theory Tech 2005;53(2):679–686.
- Tutt MN, Tserng HQ, Ketterson A. A low-loss, 5.5 GHz–20 GHz monolithic balun. Volume 2, IEEE MTT-S International Microwave Symposium Digest, Denver, CO; 1997 June 8–13; pp. 933–936.
- Yang HYD, Zhang L, Castaneda JA. Design and analysis of a multi-layer transformer balun for silicon RF integrated circuits. Radio Frequency Integrated Circuits (RFIC) Symposium, IEEE, Seattle, WA; 2002 June 2–4.

## EXERCISES

1. What is a balun and what is its main function?
2. What is the main problem of the transformer balun implemented in an RFIC die or by discrete parts?
3. An ideal transformer balun is good for simulation. Why would its turn ratio between single-ended and differential pair be set to  $1 : 2^{1/2}$ ?
4. What is the equivalence of parts between a single-ended and a differential pair in an ideal transformer balun? In other words, what are the rules to convert the value of parts from a single-ended to a differential pair, or vice versa?
5. Impedance matching for differential pair can be replaced by impedance matching for single-ended in terms of a transformer balun. Describe the procedures.
6. A simplest LC balun consists of two identical inductors and two identical capacitors. Write down two equations to calculate the inductor and capacitor values.

7. What is the limitation in the design of the simplest LC balun?
8. Design a LC balun in the case of
  - (a) central operating frequency  $f = 800$  MHz;
  - (b) terminal impedance at single-ended;
  - (c) terminal impedance at each branch of differential pair.
9. Why is the simplest LC balun good for application in testing?
10. What is the equivalence of parts between single-ended and differential pair in the simplest LC balun? In other words, what are the rules to convert the value of parts from a single-ended port to a differential pair, or vice versa?
11. What is the specialty of the Wilkinson coupler?

## ANSWERS

1. The balun converts a single-ended block to a differential pair, or vice versa. For an ideal balun,
  - (a) at two differential terminals their impedance is equal but their phase difference is  $180^\circ$  for all the frequencies;
  - (b) the insertion loss from the single-ended to each differential terminal is  $-3$  dB for all the frequencies.

For an actual balun, all the behaviors mentioned above are approximately true within a certain frequency bandwidth.
2. The main problem of the transformer balun implemented in an RFIC die is that the  $Q$  value of coils is very low, usually not more than 10, which brings about significant attenuation. On the contrary, the main problem of transformer balun implemented by discrete parts is the upper limit of the operating frequency, which can be up to 1 GHz today.
3. An ideal transformer balun is good for simulation. Its turns ratio between a single-ended and a differential pair is set to  $1 : 2^{1/2}$  because it results in the same impedance in all three terminals, one is in single-ended and other two are in a differential pair.
4. (a) A capacitor in parallel at the single-ended port is interpreted as an inductor in parallel at each port of differential pair, or vice versa.  
 (b) A capacitor in series at the single-ended port is interpreted as an inductor in series at each port of differential pair, or vice versa.  
 (c) An inductor in parallel at the single-ended port is interpreted as a capacitor in parallel at each port of differential pair, or vice versa.  
 (d) An inductor in series at the single-ended port is interpreted as a capacitor at in series each port of differential pair, or vice versa.  
 (e) The reactance of the inductor must be equal to the negative reactance of capacitor, that is,

$$L\omega_0 = \frac{1}{C\omega_0}.$$

5. In the simulation, impedance matching for differential pair can be replaced by impedance matching for single-ended port in terms of a transformer balun with following procedures:

- (a) Connect two differential ends of an ideal transformer balun with the turns ratio  $1 : 2^{1/2}$  to two terminals of differential pair.
- (b) Build an impedance matching network at the single end of the balun.
- (c) Based on the rules mentioned above, interpret the impedance matching network from the single end to the two differential ends.
- (d) Remove the transformation balun and the impedance matching network at the single end away.
6. A simple LC balun consists of two identical inductors and two identical capacitors. The two equations for calculating inductor and capacitor values are as follows:

$$L\omega_o = \frac{1}{\sqrt{2Z_{dp}Z_{sn}^*}}$$

$$C\omega_o = \frac{1}{\sqrt{2Z_{dp}Z_{sn}^*}},$$

where

$$Z_{sn} = R_{sn} + jX_{sn},$$

$$Z_{dp} = R_{dp} + jX_{dp}.$$

7. In the design of a simplest LC balun, there is a limitation or a condition. The design becomes possible only if this limitation or condition is satisfied. It is

$$\frac{R_{sn}}{R_{dp}} = \frac{X_{sn}}{X_{dp}}.$$

8.  $Z_{sn} = r_{sn} + jx_{sn} = 75 \Omega + j25 \Omega,$   
 $Z_{dp} = r_{dp} - jx_{dp} = 120 \Omega - j48 \Omega,$   
 $Z'_{sn} = r_{sn} - jx'_{sn} = 75 \Omega - j30 \Omega,$   
 $f = 800 \text{ MHz};$

$$x'_{\text{sn}} = \frac{x_{\text{dp}}}{r_{\text{dp}}} r_{\text{sn}} = \frac{48}{120} \times 75 = 30.$$

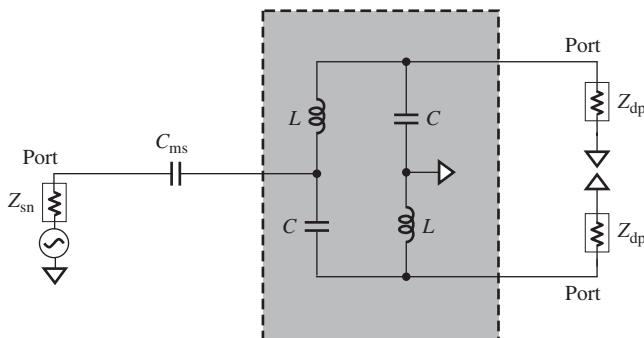
$$C_{\text{ms}} = \frac{1}{\left(\frac{x_{\text{dp}}}{r_{\text{dp}}} r_{\text{sn}} + x_{\text{sn}}\right) \omega} = \frac{10^{12}}{55 \times 6.2832 \times 800 \times 10^6} = 3.62 \text{ pF}$$

$$Z_{\text{dp}} Z'_{\text{sn}}^* = r_{\text{dp}} r_{\text{sn}} + x_{\text{dp}} x'_{\text{sn}} = 120 \times 75 + 48 \times 30 = 10,440 \Omega^2$$

$$L = \frac{\sqrt{2} Z_{\text{dp}} Z'_{\text{sn}}^*}{\omega} = \frac{\sqrt{2(r_{\text{dp}} r_{\text{sn}} + x_{\text{dp}} x'_{\text{sn}})}}{\omega} = \frac{\sqrt{20,880} \times 10^9}{6.2832 \times 800 \times 10^6} = 26.75 \text{ nH}$$

$$C = \frac{1}{\sqrt{2 Z_{\text{dp}} Z'_{\text{sn}}^* \omega}} = \frac{1}{\sqrt{2(r_{\text{dp}} r_{\text{sn}} + x_{\text{dp}} x'_{\text{sn}}) \omega}}$$

$$= \frac{10^{12}}{\sqrt{20880} \times 6.2832 \times 800 \times 10^6} = 1.37 \text{ pF}$$



**Figure 15.P.1.** Simple LC balun modified by inserting of a capacitor in series  $C_{\text{ms}}$  at the single-ended side.

9. The simplest LC balun is good to be applied in the testing, because
  - (a) it is easily implemented,
  - (b) its bandwidth is reasonable, and
  - (c) its insertion loss is low.
10. The rules to convert the value of parts from single-ended to differential pair or vice versa are as follows:
  - (a) An inductor  $L/2$  in parallel at the single-ended port is interpreted as an inductor  $L$  in series at each port of differential pair, or vice versa.
  - (b) A capacitor  $2C$  in parallel at the single-ended port is interpreted as a capacitor  $C$  in series at each port of differential pair, or vice versa.
  - (c) An inductor  $L/2$  in series at the single-ended port is interpreted as an inductor  $L$  in parallel at each port of differential pair, or vice versa.
  - (d) A capacitor  $2C$  in series at the single-ended port is interpreted as a capacitor  $C$  in parallel at each port of differential pair, or vice versa.

- (e) The reactance of the inductor must be equal to the negative reactance of capacitor, that is,

$$L\omega_o = \sqrt{2Z_{dp}Z_{sn}^*},$$

$$C\omega_o = \frac{1}{\sqrt{2Z_{dp}Z_{sn}^*}},$$

$$Z_{sn}Z_{dp} = (\sqrt{2}Z_o)^2,$$

$$Z_{sn} = 2\sqrt{2}Z_o.$$

11. The specialty of the Wilkinson coupler is

$$Z_{dp} = \frac{Z_o}{\sqrt{2}}.$$



# SOC (SYSTEM-ON-A-CHIP) AND NEXT

## 16.1 SOC

### 16.1.1 Basic Concept

The development of the RFIC was a milestone in the history of the electronic enterprise. The next milestone will be the SOC (system-on-a-chip). On an SOC, not only are all RF circuit blocks integrated on the same chip substrate, but also are all the digital and analog circuit blocks. In short, an SOC chip contains all

1. RFICs,
2. Digital ICs, and
3. Analog ICs.

As an alternative to the SOC solution, the SIP (system-in-a-package) solution has been suggested and fabricated for some electronic products or communication systems. Instead of one IC chip substrate containing all ICs, there are three individual IC chip substrates containing RFICs, digital ICs, and analog ICs, respectively. They are packaged by means of stacking these three IC chips together. Obviously, SIP is different from SOC.

SOC is the highest goal in circuit design. We face quite a lot of challenges today in order to reach the SOC design goal, though many achievements have been made.

### 16.1.2 Remove Bottlenecks in Approach to RFIC

In order to reach the SOC goal, the first task is to remove the main bottlenecks in RFIC design, which are as follows:

- Enhancing the low  $Q$  value of the spiral inductor,
- Developing a “zero” capacitor directly inside the RFIC chip,
- Researching the conductive via, and
- Modeling the bonding wire with higher accuracy.

### 16.1.3 Study Isolation between RFIC, Digital IC, and Analog IC

- *Studying the Isolation between RF Blocks.* By means of double guard rings, P+ and N wells, the isolation between RF blocks can probably reach 70–90 dB, depending on the width of guard rings and the spacing between the rings. Isolation between the RF blocks in the 70–90 dB range is basically satisfactory in the fabrication of RFIC for general purposes, such as for cellular phones, portable radios, and so on. However, this is not satisfactory for the fabrication of RFICs needed for the most advanced electronic products, such as 128QAM or military communication systems.
- *Studying Isolation between Digital Blocks.* In contrast to an RF signal, a digital signal contains many significant harmonics. The study of isolation between the digital blocks with high data rate is important because the frequency of their harmonics can be much higher than the RF frequency range.
- *Studying Isolation between RF and Digital Blocks.* On average, the power of the RF block is much higher than that of the digital block since the current in the RF block is on the order of milliamperes while that in the digital block is microamperes. From the power viewpoint, a digital block is more easily disturbed than an RF block. From the frequency viewpoint, an RF block is more easily disturbed than a digital block because a digital signal is basically a pulse that contains many harmonics. A narrow pulse is a wideband signal. It contains high frequencies, including RF components, even if its pulse repetition rate is lower than the RF frequency. Therefore, isolation between RF and digital blocks is somewhat more complicated than other isolation cases.

In a wireless communication system, the ideal isolation level should be around 130 dB, which is 10 dB higher than the total gain of the useful signal from the antenna to the data output terminal. Assuming that an interference with  $0 \text{ dB}_m$  power appears somewhere in the system, the maximum power of interference at the antenna is less than  $-130 \text{ dB}_m$  because the isolation in the system is 130 dB. The interference with  $-130 \text{ dB}_m$  power at the antenna is lower than the sensitivity, say,  $-120 \text{ dB}_m$ , of a communication system. Consequently, the communication system never perceives the existence of the interference. Of course, if the power of the interference is much greater than  $10 \text{ dB}_m$ , then 130 dB of isolation is not enough for the system; however, cases like this seldom occur.

## 16.2 WHAT IS NEXT

Digital and analog ICs have been developing since the 1960s and the RFIC has been developing worldwide since 1995. For the reduction of cost and size and the enhancement

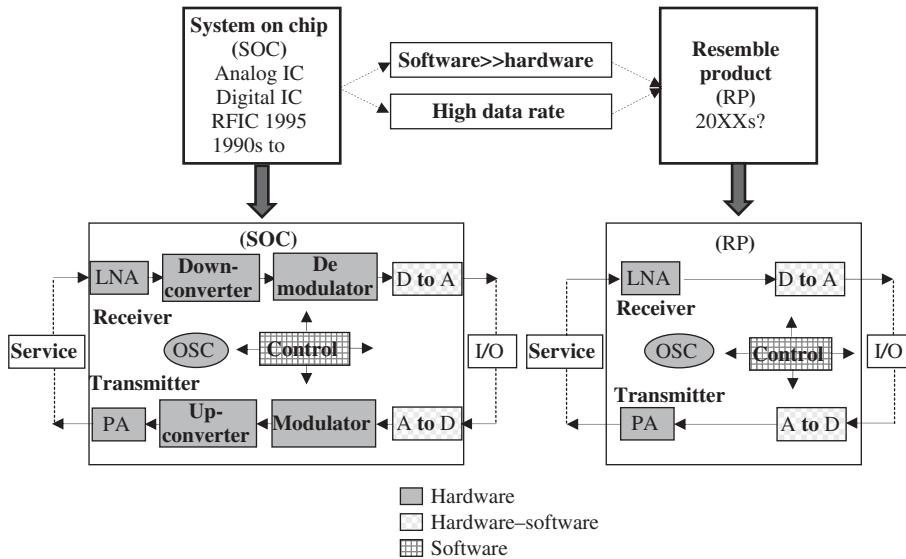


Figure 16.1. Prospect of a communication system.

of reliability of the product, our goal is the development of the SOC IC, which gained common recognition or common agreement over the world in the 1990s. An SOC IC chip is supposed to contain three kinds of integrated circuits together: the digital IC, the analog IC, and the RFIC.

Today, we are just on the way to approaching the goal of the SOC, which is expected to be reached in the near future if all the tasks listed in Section 16.1 are fulfilled.

As a matter of fact, the concept of the SOC concerns only hardware but not software. What will be the relationship between hardware and software in the future? This is a very important and interesting question in the circle of electronic engineers. Let us take a communication system as an example and try to figure it out in Figure 16.1.

As shown in Figure 16.1, an SOC communication system consists of 10 big blocks, they are as follows:

<b>Receiver</b>	
• LNA (low-noise amplifier)	(Hardware)
• Down-converter	(Hardware)
• Demodulator	(Hardware)
• D to A converter	(Software–hardware)
<b>Synchronization and Control</b>	
• Oscillator	(Hardware)
• $\mu$ P (Microprocessing) control	(Software)
<b>Transmitter</b>	
• A to D converter	(Hardware–software)
• Modulator	(Hardware)
• Up-converter	(Hardware)
• PA (power amplifier)	(Hardware)

In recent years, replacement of hardware by software is a general and irresistible tendency. The reasons are very simple. Software is much superior to hardware in many aspects, such as listed below:

- *Lower Cost.* The physical cost of software is merely that of memory chips, while the cost of hardware is that of a bunch of parts, devices, PCB, and so on. The large difference in cost between software and hardware is obvious.
- *Smaller Physical Size.* The size of software is only the size of the memory chip it resides in, while the size of hardware is that of a bunch of parts, devices, PCB, and so on. The large difference in size between software and hardware is also obvious.
- *Higher Reliability.* The software statement is executed by two symbols, “1” and “0,” which can be distinguished with quite a high reliability, while hardware operates with voltage and current, which are very dependent on the impedance, type and size of device, and many environmental factors, so that its reliability is much lower than that of software.
- *Power Consumption.* Power consumption for software is almost negligible, while power consumption for hardware constitutes the main power consumption of the system.

In terms of DSP (digital signaling processing) software, almost all the hardware for gate circuits such as the AND gate, OR gate, NAND gate, and NOR gate are replaced by software. In terms of DSP software, most receiver back ends and transmitter front ends have been replaced by software. Traditionally, the QPSK modulator and demodulator were built by hardware. Nowadays, software QPSK modulators and demodulators are available for application.

The transition of circuit blocks from hardware to software is restricted by the data rate or clock frequency. If the data rate or clock frequency is high enough, so that within one cycle the microprocessor can sample an RF signal many times and thus can manipulate the RF signal effectively, more RF circuit blocks built by hardware can be replaced by software. Going along with this general tendency of the circuit block built by hardware being replaced by software, not only modulators and demodulators but also filters, down-converters, up-converters, and other circuit blocks may be “swallowed up” by software sooner or later.

The “software radio” has sounded loudly since 1990s. Yes, it represents the general tendency that circuit blocks built by hardware will be “swallowed” by software. However, is the software radio 100% implemented by software without any hardware? The answer is no!

No matter how sophisticated the software is, it cannot intensify a very weak signal sensed by the antenna of the receiver. The LNA cannot be replaced by software, but must be fabricated by hardware. Similarly, the PA cannot be replaced by software but must be fabricated by hardware. Furthermore, let us examine the total 10 large blocks of the SOC communication system as shown in Figure 16.1. Two converters, D to A and A to D, are transformers between external analog parameters and internal digital sequences. It is pointless to talk about replacement of hardware by software here because the blocks themselves function as the transition between software to hardware. The unique block, the oscillator or VCO (voltage-controlled oscillator), must be fabricated by hardware and can by no means be built by software. Finally, the entire system must be controlled a microprocessor, which is indeed a software block.

Consequently, the number of circuit blocks in a SOC communication system is revised from the existing 10 large blocks to the 6 big blocks, which are as follows:

---

<b>Receiver</b>	
• LNA	(Hardware)
• D to A converter	(Software–hardware)
<b>Synchronization and Control</b>	
• Oscillator	(Hardware)
• $\mu$ P (Microprocessing) control	(Software)
<b>Transmitter</b>	
• A to D converter	(Hardware–software)
• PA	(Hardware)

---

In Figure 16.1, the SOC system becomes the RP (resemble product). The RP integrated circuit consists of six big blocks as shown above. It can be produced as different products. The essential difference between its products is the software block and the  $\mu$ P (Microprocessing) control, while all the other hardware blocks basically “resemble” each other. This is what the terminology “resemble” stands for.

## APPENDICES

### 16.A.1 Packaging

Compared to IC fabrication technology or RFIC design, insufficient attention is given to packaging technology. As a matter of fact, the performance of an RFIC can be dramatically affected by the IC package. Study and research of IC packaging becomes an important task nowadays.

The 1960s were the milestone years for the electronic industry. Apart from traditional technology, in which the circuitry is fabricated with the discrete components, the digital circuitry was initially fabricated with new IC technology.

Packaging is one of the IC technologies to be developed. The in-line package was the first type applied in IC products. Figure 16.A.1 shows its DIP (dual in-line package) configuration, in which the input/outputs of the IC die are arranged and assigned to the pins which are dual-in-lined along two lateral sides of the IC packaged block.

Meanwhile, plastic packaging has been widely applied in the IC for many years. This is due primarily to the low cost and relatively acceptable performance. Ceramic

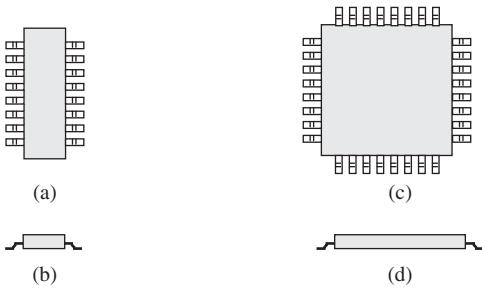


(a)



(b)

Figure 16.A.1. DIP (dual in-line) type of package. (a) Top view of in-line package. (b) Profile of in-line package.



**Figure 16.A.2.** Small outline (SOP) type of package. (a) Top view of SSOP package. (b) Profile of SSOP package. (c) Top view of QSOP package. (d) Profile of QSOP package.

packages have been developed for higher performance applications, such as for DBS (direct broadcast satellite) system. However, ceramic packages are more expensive than plastic ones.

The maximum operating frequency for the dual in-line type of package is about 2 GHz. The great inconvenience is that it requires through-hole mounting. In a massive production line, this is not an effective packaging method because it takes a long time slot to place and solder the IC package on the PCB; however, the dual in-line package is still being applied in the IC packaging for the circuit blocks operating in the low-frequency range. Therefore, the dual in-line package was gradually replaced by the SOP (small outline package) in the past decades. Figure 16.A.2 shows two configurations of package: SSOP (shrink small outline package) and QSOP (quad small outline package).

The packages with SOP type are smaller than the dual in-line types and are surface mountable. The maximum operating frequency is enhanced from 2 GHz in a dual in-line package to about 5 GHz in an SSOP. This is due to the shrinking of the length and width of the leads from dual in-line to SSOP.

The evidence is provided as shown in Figure 16.A.3, in which the measured and simulated performance of an SSOP 8 package are plotted when the input is shorted to the ground through a bonding wire interconnect. Owing to the additional inductive inductances,  $Z_{11}$  rises dramatically and exhibits a very strong resonance between 6.5 and 8 GHz. This means that this package is impossible to be applied to any circuitry operating in this frequency range. In addition, the performance is degraded from SSOP to QSOP. The maximum operating frequency for QSOP is dropped down to 3 GHz from about 5 GHz for SSOP as a result of the increase in the pin number.

Many studies and much research on SOP have been conducted in the past decades. One crucial aspect in the application of these packages is obtaining an accurate equivalent circuit model.

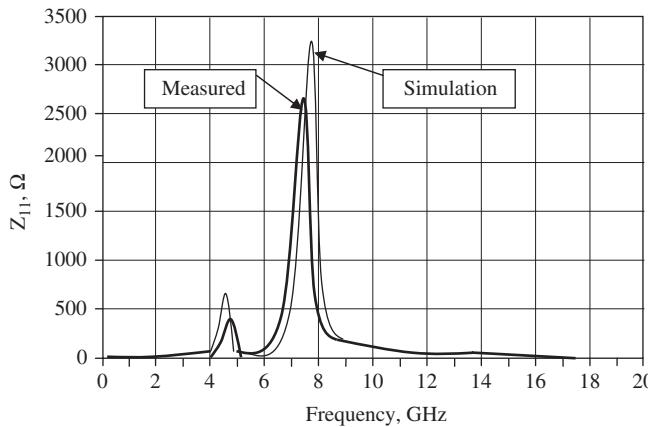
Figure 16.A.4(a) shows the cross-section of a SOP-type package, while Figure 16.A.4(b) shows the lead frame equivalent circuit, which is modeled as a lumped circuit array where every lead and bond wire is coupled to every other lead and bond wire.

Lead frame inductance directly depends on the length and the cross-section of the conductor. A short, thick conductor has less inductance than a long, thin conductor. According to Greenhouse (1974), at low frequencies

$$L_{Lf}(\mu\text{H}) = 0.002l \left( \ln \frac{2l}{w+t} + 0.50049 + \frac{w+t}{3l} \right), \quad (16.A.1)$$

where

$L_{Lf}$  = the lead frame inductance, in  $\mu\text{H}$  (typically,  $L_{Lf} \sim 0.9 \text{ nH/mm}$ ),  
 $l$  = the length of the conductor in cm,



**Figure 16.A.3.** Input impedance when the input terminal is short-circuited by a bonding wire in an SSOP 8 package. (Source: Jessie D, Larson, LE, Improved techniques for the measurement and modelling of Plastic surface mount packages to 20 GHz, IEEE Radio and Wireless Conference (RAWCON). 2000. p. 243–246.)

$w$  = the width of the conductor in cm, and  
 $t$  = the thickness of the conductor in cm.

It can be seen that a shorter lead will reduce lead-frame inductance directly. This is the motivation to develop SSOP based on SOP-type package.

Wire-bond inductance  $L_{wb}$  at low frequencies is

$$L_{wb}(\mu\text{H}) = 0.002l \left( \ln \frac{2l}{\rho} - 0.75 \right), \quad (16.A.2)$$

where

$L_{wb}$  = the wire-bond inductance in  $\mu\text{H}$  (typically,  $L_{wb} \sim 1 \text{ nH/mm}$ ),  
 $l$  = length of the wire bond in cm,  
 $\rho$  = radius of the wire bond in cm,

And at high frequencies, it is

$$L_{wb}(\mu\text{H}) = L_{wb} - L_2 f^2. \quad (16.A.3)$$

Typically,  $L_2 \sim 0.1 \text{ pH}/(\text{GHz}^2 \text{ mm})$  for single bond wire, and  $L_2 \sim 0.06 \text{ pH}/(\text{GHz}^2 \text{ mm})$  for double bond wires.

The mutual inductance between adjacent pins and bond wires increases the total inductance of any pin. In addition, it leads to serious coupling between adjacent pins because of the induced currents in the coupled inductors through the well-known relation

$$V_L = sLI_1 + sMI_2, \quad (16.A.4)$$

where

$V_L$  = the induced voltage in the inductor L,  
 $I_1$  = the current flowing through the conductor 1, and  
 $I_2$  = the current flowing through the conductor 2.

$$k_{12} = \frac{M_{12}}{\sqrt{L_1 L_2}}, \quad (16.A.5)$$

where

- $k_{12}$  = the mutual coupling coefficient,
- $M_{12}$  = the mutual inductance [ $\mu\text{H}$ ],
- $L_1$  = the self-inductance of the conductor 1 [ $\mu\text{H}$ ], and
- $L_2$  = self-inductance of the conductor 2 [ $\mu\text{H}$ ].

$$M_{12}(\mu\text{H}) = 0.002l \left( \ln \left[ \frac{l}{d} + \sqrt{1 + \frac{l^2}{d^2}} \right] - \sqrt{1 + \frac{l^2}{d^2} + \frac{d}{l}} \right), \quad (16.\text{A}.6)$$

where

- $M_{12}$  = the mutual inductance [ $\mu\text{H}$ ],
- $l$  = the length of the conductor [cm], and
- $d$  = distance apart of the conductor [cm].

For I/O leads greater than 20, the quad surface mount type of packages (QSOP) is more appropriate.

In order to enhance the symmetry of differential pair and the isolation between pins or RF blocks, it is suggested to assign pins as shown in Figure 16.A.5:

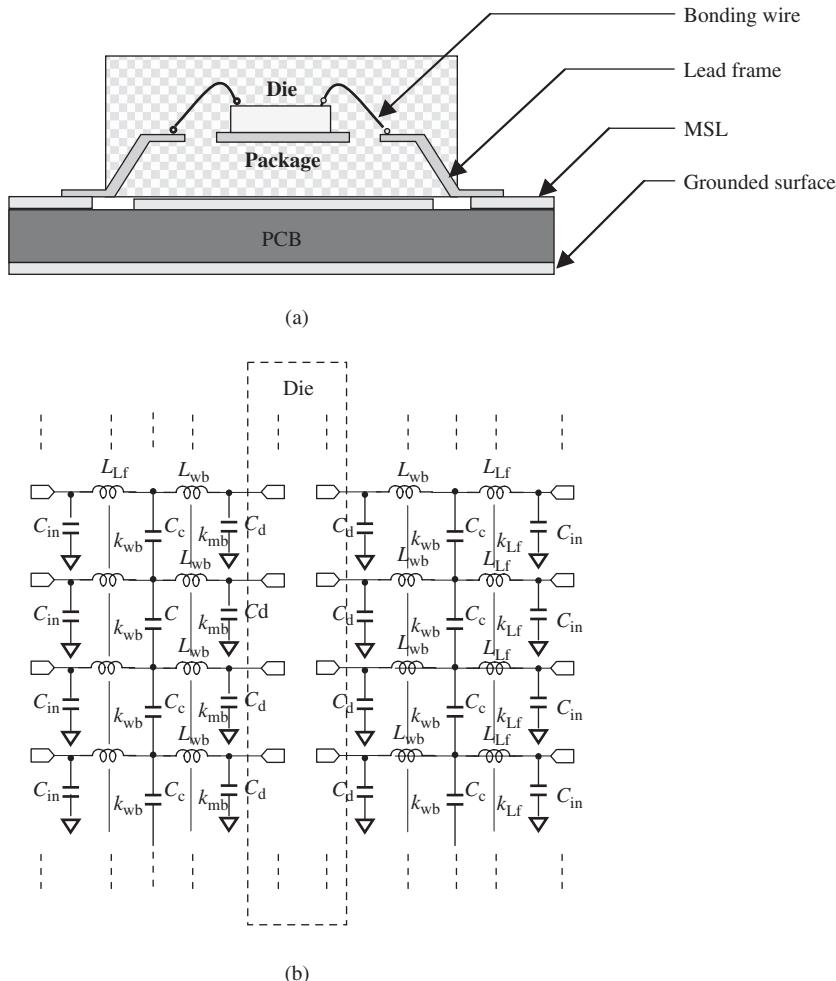
- insert one “GND” pad in the middle of differential pair pads,
- insert two GND pads between two groups of differential pair pads,
- insert three GND pads between two different RF blocks.

SOPs have been widely applied in IC packaging in the last decades, especially in the early stages of RFIC implementation in the 1980s. The lead frame inductance  $L_{\text{Lf}}$  can be significantly reduced if its configuration is modified to a transmission-line-like structure.

The most recent developed BCC (bump chip carrier) and BGA (ball grid array) packages are applied in RFIC or digital IC products. These exhibit improved performance, with an RL (return loss) greater than 20 dB and IL (insertion loss) less than 1 dB. And, its maximum of operating frequency approaches 7 GHz or higher. Figures 16.A.6 and 16.A.7 show the configuration of the QFN (quad flat no-lead) package and the BCC packages, respectively, in which the leads are replaced by resin bumps. As a matter of fact, the resin bump is a resin protrusion with metal. They are also well suited to BGA packages, which exhibit very low inductance. The resin bump inductance is reduced to approximately 50 pH, which is significantly less than  $\sim 1 \text{ nH/mm}$  of a wire bond. However, its cost is increased two to three times of that by wire-bond technology. In addition, its equivalent flip-chip pitch is substantially wider than that by wire-bond technology.

Figure 16.A.8 is a conceptual drawing of BGA technology. The gold bond wire in a BCC package is much longer than in a BGA package if most of the gold bond wire is replaced by a combination of runners and balls as shown in Figure 16.A.8. Consequently, the additional inductance could be significantly reduced from a BCC to a BGA package.

Finally, it should be noted that an alternative package technology, SIP, has been developing since the SOC goal was defined and pursued. In the SIP, many kinds of ICs, including RFIC, digital ICs, and analog ICs, are stacked together. It is not a real SOC package, although it looks like one.



$C_{in}$  = Capacitance between the lead frame solder pad and the backplane ground. It is independent on the dielectric constant, thickness of PCB, and solder pad area.

$L_{Lf}$  = The lumped inductance of the lead frame conductor and some of parasitic inductance between PCB and the lead conductor in  $\mu\text{H}$ ,

$L_{wb}$  = The lumped inductance of the bond wire in  $\mu\text{H}$ ,

$C_d$  = The bond pad capacitance where the bond wire attached to die,

$k_{Lf}$  = The magnetic coupling coefficient between the lead frame pins,

$k_{wb}$  = The magnetic coupling coefficient between the bondwires,

$C_c$  = The coupling capacitance between the pins,

**Figure 16.A.4.** Traditional SOP-type leaded package. (a) Cross-section of SO-type plastic-lead package. (b) The package parasitics.

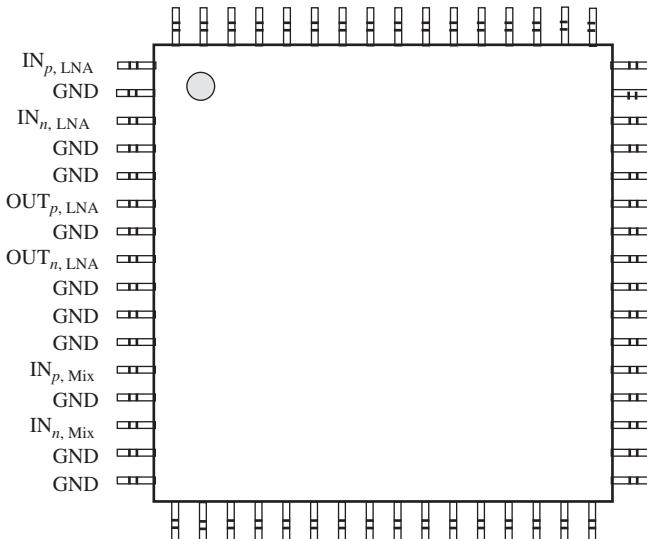


Figure 16.A.5. Example of pin distribution on an IC chip.

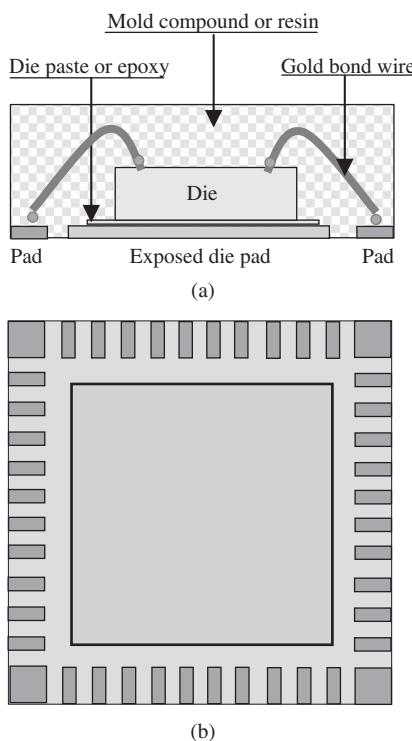
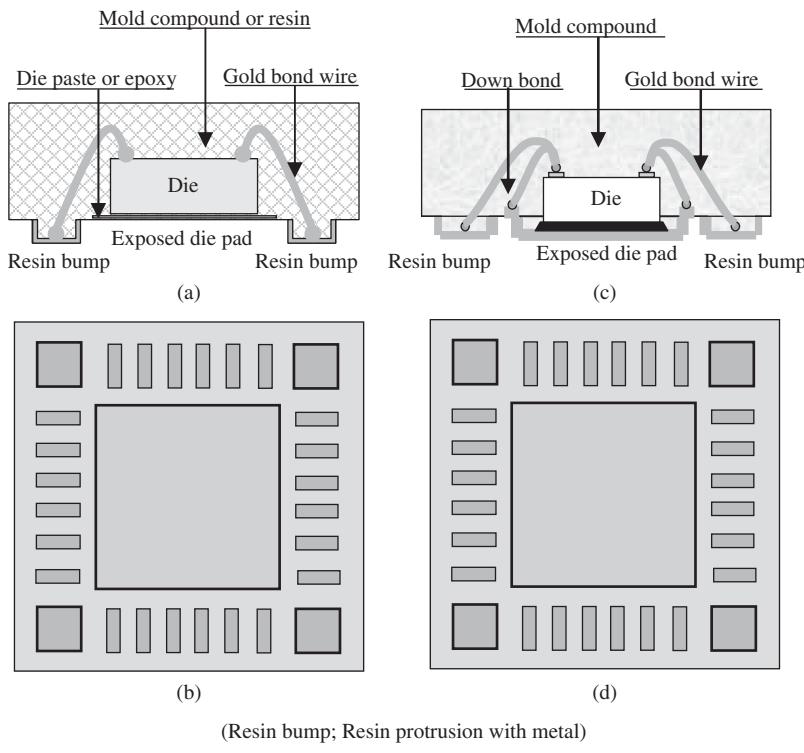
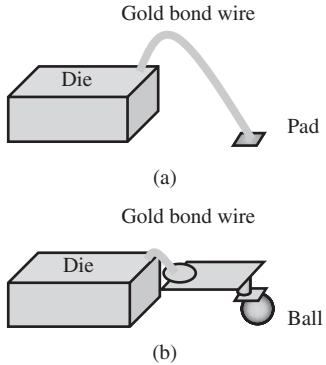


Figure 16.A.6. QFN (quad flat no-lead) package.  
(a) Cross-section of the package. (b) Bottom view of the package.



**Figure 16.A.7.** BCC (bump chip carrier) package (resin bump: resin protrusion with metal. (a) Top view of the BCC package. (b) Cross-section of the BCC++ package. (c) Bottom view of the BCC package. (d) Bottom view of the BCC++ package.



**Figure 16.A.8.** Concept of BGA (ball grid array) technology.  
 (a) Connection between die and pad on PCB by bond wire in BCC package.  
 (b) Connection between die and PCB by bond wire and ball in BGA package.

## REFERENCES

- Greenhouse HM. Design of planar rectangular microelectronic inductors. *IEEE Trans Parts Hybrids Packag* 1974;PHP-10(2):101–109.
- Jessie D, Larson LE. Improved techniques for the measurement and modelling of plastic surface mount packages to 20 GHz. *IEEE Radio and Wireless Conference (RAWCON)*. 2000. pp. 243–246.

## FURTHER READING

- EIA/JESD51-1 to 11, EIA/JEDEC Standard, 1995–2001. Integrated Circuits measurement method-Electrical test method, regulated by EIA/JEDEC since December 1995.
- Alimenti F, Mezzanotte P, Roselli L, et al. Modelling and characterization of the bond-wire interconnection. *IEEE Trans Microw Theory Tech* 2001;49(1):142–150.
- Consolazio S, Nguyen K, Biscan D, et al. Low-temperature co-fired ceramic(LTCC) for wireless applications. Proceeding 2001 IEEE Topical Workshop on Technologies for Wireless Applications, Vancouver, BC. 2001. pp. 201–205.
- Grover FW. Inductance calculations: working formulas and tables. New York: Dover; 1973.
- Horng TS, Wu SM, Chiu CT, et al. Electrical performance improvements on RFICs using bump chip carrier packages as compared to standard thin shrink small outline packages. *IEEE Trans Adv Packag* 2001;24(4):548–554. 1521-3323/01\$10.00©2001 IEEE.
- Jessie D, Larson L. An X-Band small outline leaded plastic package for MMIC applications. *IEEE Trans Adv Packag* 2002;25(3):439–447.
- Khan A. Recent developments in high-performance system-on-chip IC design. International Conference on IC Design and Technology; 2004 May 17–19; Austin (TX). Austin, Texas: Cadence Design Systems, Inc; 2004. pp. 151–157. 0-7803-8528-4/04/\$20.00 2004 IEEE.
- Larson L, Jessie D. Advances in RF packaging technologies for next-generation wireless communications applications. (Invited paper), IEEE 2003 Custom Integrated Circuits Conference. 2003. pp. 323–330. 0-7803-7842-3/03/\$17.00© 2003 IEEE.
- Larson L, Jessie D, University of California, San Diego, Qualcomm Inc. Advances in RF packaging technologies for next-generation wireless communication applications. (Invited papers), IEEE 2003 Custom Integrated Circuits Conference. 2003. pp. 323–329. 0-7803-7842-3/03/\$17.00 2003 IEEE.
- Low Y, Degani Y, Guinn K, et al. RF flip-module BGA package. *IEEE Trans Adv Packag* 1999;22(2):111–115.
- Ofner E, Nutmi J, Madsen J, et al. SoC-Mobinet, R&D and education in system-on-chip design. International Conference on IC Design and Technology; 2004 May 17–19; Austin (TX). 2004. pp. 77–79. 0-7803-8558-6/04/\$20.00 02004 IEEE.
- Scranton C, Lawson J. LTCC materials: where we are and where we are going? Proceedings of 1999 Radio and wireless Conference (RAWCON). 1999. pp. 194–200.
- Sullivan P, Xavier B, Ku W. An integrated CMOS distributed amplifier utilizing packaging inductance. *IEEE Trans Microwave Theory Tech* 1997;45(10Pt.2):1969–1976.
- Svelto F, Deantonio S, Castello R. A 1mA, −120.5 dBc/Hz at 600 kHz from 1.9 GHz fully tunable LC CMOS VCO. IEEE Radio and wireless Conference (RAWCON) 2000, Proceeding Custom Integrated Circuits Conference. 2000. pp. 577–580.

## EXERCISES

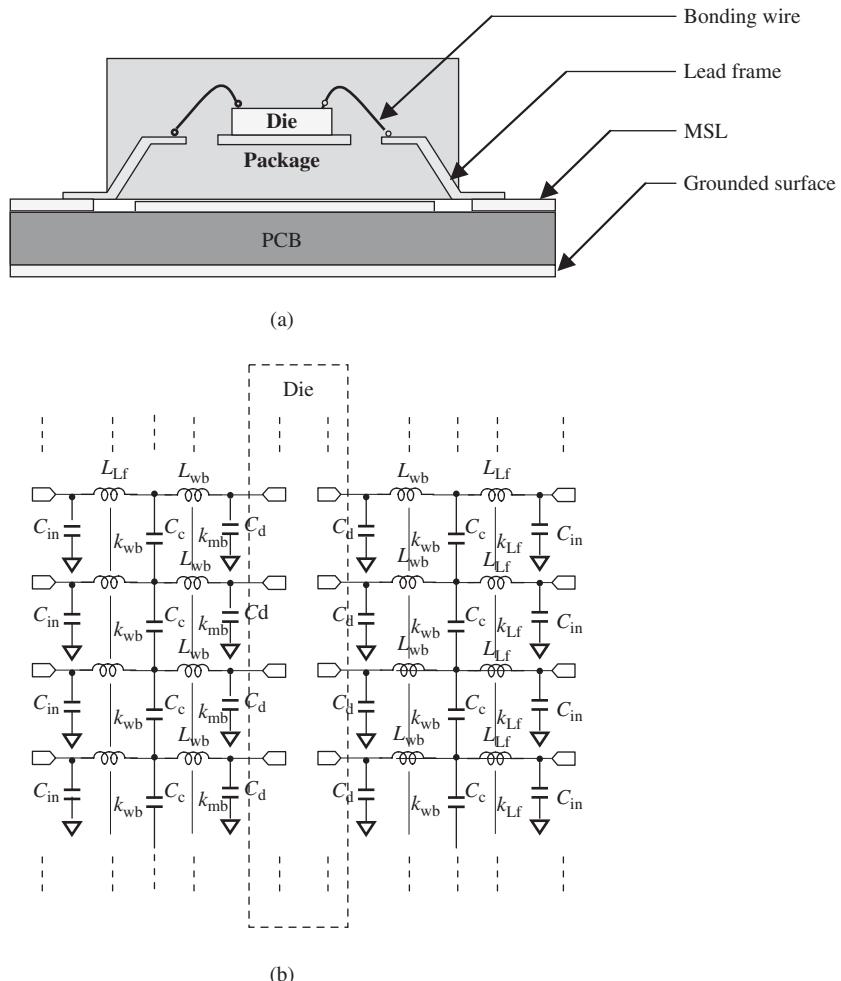
1. What is the SOC goal?
2. How many bottlenecks block the way to approaching SOC?
3. Comment on the insight of the RP (resemble product) era.
4. How many types of RF packaging are there? What is their upper frequency limit? Which is the best package for RF circuitry currently?
5. What kinds of degradation happen because of imperfect packaging?

6. How difficult is it to establish the model for a bond wire?
7. Briefly draw the package parasitics in a leaded package.

## ANSWERS

1. The SOC goal is a single IC chip with only one substrate containing all ICs, including
  - (a) RFICs,
  - (b) digital ICs, and
  - (c) analog ICs.
2. The main bottlenecks in RFIC design are
  - (a) enhancing the low  $Q$  value of the spiral inductor,
  - (b) developing a zero capacitor directly on the RFIC chip,
  - (c) modeling the bonding wire with higher accuracy, and
  - (d) studying and researching of via.The main bottlenecks in isolation research are
  - (a) isolation between RF blocks,
  - (b) isolation between digital blocks, and
  - (c) isolation between RF and digital blocks.
3. The insight about the RP (resemble product) era is reasonable and would become reality after several decades.
4. There are four types of RF packaging:
  - (a) Dual in-line package (DIP),
  - (b) Small outline package (SOP),
  - (c) Quad flat no-lead mount (QFN), and
  - (d) Bump chip carrier (BCC) and Ball grid array (BGA).Their upper frequency limit is 2, 5, 3, and 7 GHz, respectively.  
The grid array type is the best package for RF circuitry currently.
5. Imperfect packaging would cause degradation of input signal in terms of
  - (a) excessive IL,
  - (b) insufficient RL,
  - (c) pin-to-pin isolation problem, and
  - (d) self-resonance.
6. It is difficult to establish the model for a bond wire. A more accurate model for bond-wire inductance requires simulation with full three-dimensional analysis, and must account for
  - (a) the curvature of the bond wire,
  - (b) the tilted angle of the bond wire, as well as
  - (c) the height above the ground plane.

7. Figure 16.P.1 shows the package parasitics in a leaded package.



**Figure 16.P.1.** Traditional SOP-type leaded package. (a) Cross-section of the package. (b) The package parasitics.

---

# PART 3

---

---

## INDIVIDUAL RF BLOCKS

---



# LNA (LOW-NOISE AMPLIFIER)

## 17.1 INTRODUCTION

In a wireless communication system, the LNA (low-noise amplifier) is the first circuit block in the receiver. It is one of the most important blocks in RF circuit design, because of the following:

- The sensitivity of the receiver is mainly determined by the LNA noise figure and power gain. The noise figure of the LNA significantly impacts the overall noise performance of the receiver. On the other hand, the power gain of the LNA significantly suppresses noise contributions from subsequent stages, so that it also impacts the overall noise performance of the receiver.
- The LNA plays an important role in the linearity of the entire system. Its nonlinearity must be reduced as much as possible.
- In a CDMA (code division multiplex access) wireless communication system, the LNA takes care of AGC (automatic gain control) in the entire system as well.

This chapter covers the following:

- *Typical Design Procedures.* These include selection of device size, raw device testing, input and output impedance matching, stability checking, and linearity examination.

- *Cascode LNA*. As the wireless bandwidth is raised up to gigahertz (GHz) or tens of GHz, the performance of the LNA is restricted by the input Miller capacitance. Increasing isolation between the input and output in an LNA would be helpful to an advanced communication system. The cascode LNA would improve the isolation performance from single-ended LNA.
- *AGC (Automatic Gain Control)*. Without AGC capability, it is impossible for the wireless CDMA communication system to operate well.

In recent years, the differential LNA is specially required for the direct conversion or “zero IF” wireless communication system. This has been discussed in Chapter 13, where the differential pair discussed applies not only to the differential LNA but also to other RF circuit blocks.

The LNA has been developed over several decades. However, as the progress of electronic products moved forward, LNA design was asked to reach higher and higher goals. For example, the voltage of DC power supply became lower and lower, from 3 to 1 V in a cellular phone design. The current drain had to be reduced as much as possible so that the standby current of the overall receiver could be just a few milliamperes to conserve battery consumption. It must be small, the cost must be low, and the performance must be maintained at a high level. LNA design becomes more complicated if trade-offs must be taken between size, cost, and performance.

It is well known that the LNA must magnify the weak signal from the antenna and intensify the signal up to an appropriate power level required by subsequent stages. This implies that an LNA must be

1. a low noise block so that the weak signal will not be “submerged” by noise;
2. a high power gain block so that its output can drive the following stage well.

An LNA with maximum gain may not be in the state of minimum noise, or vice versa. A trade-off is usually taken between maximizing gain and minimizing the noise figure. In past decades, much effort has been put into designing an LNA to reach both maximum gain and minimum noise figure simultaneously. This is a great challenge in LNA circuit design. This jeopardy was solved a couple of years ago in my designs but has not been reported yet. Now I am going to share it with our readers.

## 17.2 SINGLE-ENDED SINGLE DEVICE LNA

In this section, the design procedures and schemes are illustrated through a design example, in which a MOSFET (metal–oxide–semiconductor field-effect transistor) is selected as the single-ended device (it can of course be replaced by other types of devices).

A single-end LNA with a single device is the simplest LNA. Nevertheless, it is the essence or core in all other types of LNA designs, including cascode and differential designs. The design procedures and schemes described in this section are suitable to all types of LNA design.

The main goals for the design example are

- $V_{cc} = 3.0 \text{ V}$ ,
- $I_{cc} < 3.0 \text{ mA}$ ,

- frequency range = 850–940 MHz,
- NF < 2.5 dB,
- gain > 10 dB,
- IP<sub>3</sub> > 0 dB<sub>m</sub>, and
- IP<sub>2</sub> > 40 dB<sub>m</sub>.

### 17.2.1 Size of Device

The first step in LNA circuit design is to decide the size of the device. Many trade-offs must be taken into account between size, cost, performance, and so on. In this subsection, only performance is counted in the selection of device size.

In digital IC circuit design, the MOSFET has become dominant in recent years because the size of the device can be shrunk and the current drain can be reduced more than with other devices. Among MOSFETs, device length therefore becomes the key parameter in the selection of IC foundry and processing because it strongly impacts the total area of the IC die and therefore the cost, speed of performance, maximum data rate, current drain, and so on. The reason is simple: in digital IC circuit design, hundreds and even thousands of transistors are needed. The total area of the IC die, and therefore the cost, is significantly reduced as the device length decreases. IC scientists and engineers have worked very hard to shrink the size of transistors, which now approach unbelievably tiny sizes. In the 1990s, the length of a MOSFET device was in the order of micrometers. From 2000 to 2005, the IC world entered the so-called nanometers era. Many foundries today have the capability of manufacturing MOSFET ICs with lengths of 0.5, 0.35, 0.25, 0.18, and 0.11 μm. In 2006, the length of a MOSFET device was furthermore shrunk to 90, 45, 22.5 nm. The progress of IC processing is moving forward very fast, and consequently, IC circuit design work becomes more and more challenging.

In the RF circuit design, bipolar transistors were applied to RFIC development in the 1990s. Meanwhile, the MOSFET device has been applied to the RFIC as well. The smaller size of MOSFET devices bring about the same advantages to RF circuit design as to digital circuit design, such as the reduction of cost and the increase in operating frequencies. It must be pointed out, however, that smaller size of the device is not the main objective pursued in RF circuit design because the total number of devices applied in RF circuits is much lower than the number of devices applied in digital circuits. Instead of pursuing smaller size, RF engineers prefer to select device lengths for which the technology of IC processing in the foundry is more mature and the device model for simulation is more accurate. In addition, there are two important factors to be considered in the selection of the MOSFET device's size: one, the restriction of device size because of  $V_{gs}$  limitation, and another because of expectations of NF<sub>min</sub>.

**17.2.1.1 Restriction of W/L Due to Limitation of  $V_{gs}$ .** In LNA design, the MOSFET is usually operated in its active region. The characteristics of a MOSFET can be expressed as

$$I_d = \frac{\mu_n C_{ox}}{2} \frac{W}{L} (V_{gs} - V_{th})^2, \quad (17.1)$$

$$g_m = \frac{\partial I_d}{\partial V_{gs}} = \mu_n C_{ox} \frac{W}{L} (V_{gs} - V_{th}), \quad (17.2)$$

where

- $I_d$  = drain current,
- $g_m$  = transconductance of MOSFET,
- $W$  = width of MOSFET,
- $L$  = length of MOSFET,
- $V_{gs}$  = gate-source voltage for  $n$ -channel MOSFET,
- $V_{th}$  = threshold voltage for  $n$ -channel MOSFET, the minimum gate-to-channel voltage needed to produce an inversion layer beneath the gate,
- $V_{ds}$  = drain-source voltage for  $n$ -channel MOSFET,
- $\mu_n$  = channel mobility, typically  $700 \text{ cm}^2/\text{V s}$ , and
- $C_{ox}$  = capacitance per unit area of the gate oxide.

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}, \quad (17.3)$$

where  $t_{ox}$  is the thickness of the gate oxide.

From (17.1) and (17.2) we have

$$g_m = \sqrt{2\mu_n C_{ox} \frac{W}{L} I_d}, \quad (17.4)$$

$$V_{gs} = 2 \frac{I_d}{g_m} + V_{th}. \quad (17.5)$$

Equation (17.4) shows that  $g_m$  is proportional to the square root of the ratio  $W/L$ . The increase in the ratio  $W/L$  is equal to the increase in  $g_m$ . On the other hand, from equation (17.5), it can be seen that there are two ways to make  $I_d$  reach to a certain amount, either by increasing  $g_m$  through the increase in the ratio  $W/L$  or by increasing  $V_{gs}$  through the factor of  $(V_{gs} - V_{th})$ . Should the selected value of the ratio  $W/L$  be too small,  $V_{gs}$  must be increased to an unacceptable value.

In order to illustrate the relationships between  $g_m$ ,  $I_d$ , and  $W/L$ , and the corresponding values of  $V_{gs}$ , Table 17.1 lists the calculated  $V_{gs}$  values when  $I_d$  and the ratio  $W/L$  are selected in different levels or amounts and when the basic parameters applied in the calculations are assumed as follows:

$$\epsilon_{ox} = 3.45 \times 10^{-13} \text{ F/cm}, \quad (17.6)$$

$$t_{ox} = 23.3 \text{ \AA} = 23.3 \times 10^{-8} \text{ cm}, \quad (17.7)$$

$$\mu_n = 170 \text{ cm}^2/\text{V s}, \quad (17.8)$$

$$C_{ox} = 14.81 \text{ fF}/\mu\text{A}^2, \quad (17.9)$$

$$V_{tn} = 0.49 \text{ V}, \quad (17.10)$$

then

$$\mu_n C_{ox} = 251.72 \text{ } \mu\text{A/V}^2. \quad (17.11)$$

In Table 17.1, the calculations are conducted for the cases of  $I_d = 1, 2, 5, 10, 20$ , and  $50 \text{ mA}$  with the different levels of  $W/L = 10, 100, 1000, 2000, 5000, 10,000$ , and  $20,000$ .

TABLE 17.1.  $V_{gs}$  Limitation in the Selection of Device Size

$I_d$ , mA	$W$ , $\mu\text{m}$	$L$ , $\mu\text{m}$	$W/L$	$g_m$ , mA/V	$V_{gs}$ , V
1.00	0.9	0.09	10.00	2.24	<u>1.38</u>
1.00	9	0.09	100.00	7.10	<u>0.77</u>
1.00	90	0.09	1000.00	22.44	0.58
1.00	180	0.09	2000.00	31.73	0.55
1.00	450	0.09	5000.00	50.17	0.53
1.00	900	0.09	10,000.00	70.95	0.52
1.00	1800	0.09	20,000.00	100.34	0.51
2.00	0.9	0.09	10.00	3.17	1.75
2.00	9	0.09	100.00	10.03	0.89
2.00	90	0.09	1000.00	31.73	0.62
2.00	180	0.09	2000.00	47.79	0.58
2.00	450	0.09	5000.00	70.95	0.55
2.00	900	0.09	10,000.00	100.34	0.53
2.00	1800	0.09	20,000.00	141.91	0.52
5.00	0.9	0.09	10.00	5.02	<u>2.48</u>
5.00	9	0.09	100.00	15.87	<u>1.12</u>
5.00	90	0.09	1000.00	50.17	0.69
5.00	180	0.09	2000.00	75.56	0.63
5.00	450	0.09	5000.00	112.19	0.58
5.00	900	0.09	10,000.00	158.66	0.55
5.00	1800	0.09	20,000.00	224.37	0.53
10.00	0.9	0.09	10.00	7.10	<u>3.31</u>
10.00	9	0.09	100.00	22.44	<u>1.38</u>
10.00	90	0.09	1000.00	70.95	<u>0.77</u>
10.00	180	0.09	2000.00	106.86	0.69
10.00	450	0.09	5000.00	158.66	0.62
10.00	900	0.09	10,000.00	224.37	0.58
10.00	1800	0.09	20,000.00	317.31	0.55
20.00	0.9	0.09	10.00	10.03	4.48
20.00	9	0.09	100.00	31.73	1.75
20.00	90	0.09	1000.00	100.34	0.89
20.00	204	0.09	2000.00	151.13	0.77
20.00	450	0.09	5000.00	224.37	0.67
20.00	900	0.09	10,000.00	317.31	0.62
20.00	1800	0.09	20,000.00	448.75	0.58
50.00	0.9	0.09	10.00	15.87	<u>6.79</u>
50.00	9	0.09	100.00	50.17	<u>2.48</u>
50.00	90	0.09	1000.00	158.66	<u>1.12</u>
50.00	204	0.09	2000.00	238.95	<u>0.94</u>
50.00	450	0.09	5000.00	354.77	<u>0.77</u>
50.00	900	0.09	10,000.00	501.71	0.69
50.00	1800	0.09	20,000.00	709.53	0.63

The underlined values of  $V_{gs}$  in the rightmost column in Table 17.1 are unacceptable because they are higher than 0.7 V, which is considered the highest acceptable value of  $V_{gs}$  when the DC power supply is low, say, 1.0–1.8 V. Therefore, the rows containing underlined values of  $V_{gs}$  in Table 17.1 must be abandoned in the selection of the ratio  $W/L$ . Hence, the values of the ratio  $W/L$  are restricted for the given values of  $I_d$  and

$g_m$  because of the constraint on  $V_{gs}$ . All other rows and their candidates in Table 17.1 are acceptable. They will be further narrowed down in the consideration of the so-called power-constrained noise optimization.

It should be noted that Table 17.1 is an example only. The selection of the ratio  $W/L$  for the device must be conducted by designers based on the basic parameters,  $\epsilon_{ox}$ ,  $t_{ox}$ ,  $\mu_n$ ,  $C_{ox}$ , and  $V_{tn}$ , which actually apply to the device.

**17.2.1.2 Optimum Width  $W_{opt}$  of Device.** On the basis of theoretical derivation (Lee, 1998), the size selection of the device in LNA design is more reasonably considered from the expectation of a minimum of noise. By explicitly taking power consumption into account, the optimal width of a device  $W_{opt}$  for the minimum noise figure  $NF_{min}$  can be expressed as

$$W_{opt} = \frac{1}{3\omega LC_{ox}R_S}, \quad (17.12)$$

where

- $W_{opt}$  = optimal width of device (MOSFET),
- $\omega$  = operation angular frequency,
- $L$  = length of device (MOSFET),
- $C_{ox}$  = capacitance per unit area of the gate oxide, and
- $R_S$  = source resistance.

This results from the power-constrained noise optimization.

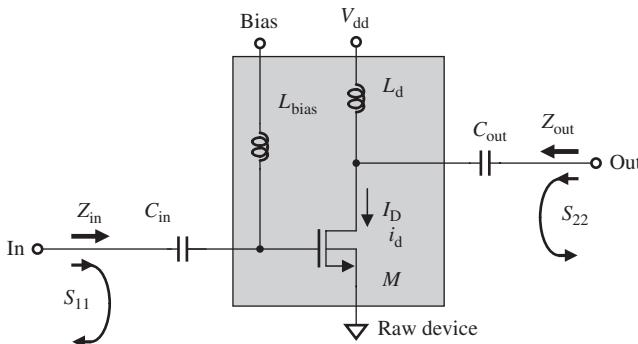
The value of the optimized width of the device is inversely proportional to the operating frequency,  $\omega$ , the source resistance,  $R_S$ , the capacitance of the gate oxide area,  $C_{ox}$ , and the length of the device,  $L$ . The designer knows the first two parameters,  $\omega$  and  $R_S$ . The other two parameters,  $C_{ox}$  and  $L$ , are provided by the IC foundry, which may have a couple of choices. For instance, device lengths of 0.25, 0.18, 0.13, 0.11, and 0.09  $\mu m$  are available in most MOS IC foundries at present. On the basis of data that the IC foundry provides, the corresponding values of  $W_{opt}$  can be calculated from expression (17.12). Then, these  $W_{opt}$  and  $L$  values can be examined for a reasonable value of  $V_{gs}$  as in Table 17.1 and the best set of  $W_{opt}$  and  $L$  can be determined. Finally, the decision of IC processing can be made.

## 17.2.2 Raw Device Setup and Testing

Raw device testing is the second step in the block circuit design. It should be noted that it is a key step in a good LNA circuit design.

In the circuit design, a “device” is a general name for a transistor. The transistor can be bipolar, or a MOSFET, or GaAs, or some other type. The purpose of raw device testing is to determine the operating characteristics only of the device, and nothing else. However, an operating transistor must be provided with the DC power supply and bias, and therefore, some additional parts such as the RF choke and DC blocking or AC bypass capacitors must be connected. The impedance of the additional parts must therefore approach either zero when they are connected in series or infinity when they are connected in parallel. If so, the tested characteristics of the transistor are not disturbed by the addition of those parts.

Figure 17.1 shows the setup for raw device testing. The capacitors  $C_{in}$  and  $C_{out}$  are “zero” capacitors, whereas the inductors  $L_{bias}$  and  $L_d$  are “infinite” inductors. They are



$C_{in}, C_{out}$ : “zero” capacitor.  
 $L_{bias}, L_d$ : “Infinite” Inductor.

Figure 17.1. Setup for raw device testing,  
 $f = 850\text{--}940 \text{ MHz}$ ,  $I_D = 2.6 \text{ mA}$ .

discussed in Chapter 7 where the “zero” capacitor and “infinite” inductor are selected from discrete chip parts. In the actual simulation for IC circuitry,  $C_{in}$  and  $C_{out}$  can be large capacitors with a high value of capacitance so that their impedance approaches zero at operating frequencies, whereas  $L_{bias}$  and  $L_d$  can be large inductors with a high value of inductance so that their impedance approaches infinity at operating frequencies. The desired current drain of the transistor,  $I_D + i_d$ , can be adjusted by the bias voltage, where  $I_D$  is the DC drain portion and  $i_d$  is AC drain portion of the MOSFET.

In Figure 17.1, the device is a MOSFET with CMOS (complementary metal-oxide semiconductor) IC processing. Its size has been selected based on the considerations of  $V_{gs}$  and  $NF_{min}$  as discussed in Section 17.2.1. As mentioned above, the DC power supply is 3 V, and the drain current, adjusted by the bias, is 2.6 mA.

The operating frequency range is from 850 to 940 MHz. Its relative bandwidth is

$$\frac{\Delta f}{f_0} = \frac{940 - 850}{(940 + 850)/2} = 10.05\%. \quad (17.13)$$

This is a narrow band block. Usually, a block or a system with a relative bandwidth greater than 15% is considered a wide band block or system. On the contrary, a block or a system with a relative bandwidth less than 15% is considered a narrow band block or system.

The purpose of raw device testing is twofold:

1. To create a starting point for impedance matching in order to continue the next design step.
2. To see if the raw device can approach a good LNA design. A good LNA design suggests that a minimum of noise figure and a maximum of gain can be obtained simultaneously.

It is easy to understand that the first purpose of raw device testing is to create a starting point for matching input and output impedance. The input and output impedances,  $Z_{in}$  and  $Z_{out}$ , are approximately related to the  $S$  parameters by the following expressions:

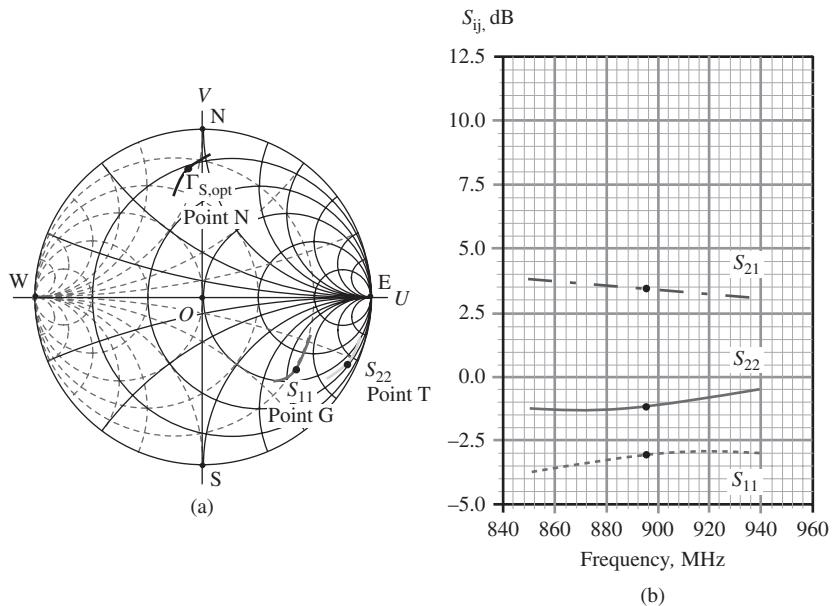
$$z_{in} = \frac{1 + S_{11}}{1 - S_{11}}, \quad (17.14)$$

$$z_{\text{out}} = \frac{1 + S_{22}}{1 - S_{22}}. \quad (17.15)$$

This approximation is usually correct if the transistor's isolation between input and output is good and the testing calibration is well done. Through the testing of  $S_{11}$  and  $S_{22}$ , the input and output impedances,  $Z_{\text{in}}$  and  $Z_{\text{out}}$ , can be read directly from the Smith Chart at the same locations of  $S_{11}$  and  $S_{22}$ , respectively.

Figure 17.2(a) shows the test results of  $S_{11}$  and  $S_{22}$ , and hence  $Z_{\text{in}}$  and  $Z_{\text{out}}$  on the Smith Chart. The input and output impedances of a MOSFET are usually capacitive and are located in the bottom half of Smith Chart, while the input and output impedances of a bipolar transistor can be either capacitive or inductive, depending on the device size, current drain, and operating frequency. Another difference between the MOSFET and bipolar transistor is that the input impedances of a MOSFET are usually located in the relatively higher impedance area, while the input and output impedances of a bipolar transistor are usually located in the relatively lower impedance area. This difference implies that impedance matching is more difficult for the MOSFET than for the bipolar transistor and that the isolation between input and output in the MOSFET is better than in the bipolar transistor.

Figure 17.2(a) also shows that the location of  $S_{22}$  is much farther from  $50 \Omega$  and is in the very high impedance area, while  $S_{11}$  is located somewhat closer to  $50 \Omega$  than  $S_{22}$ . Correspondingly, Figure 17.2(b) shows that the magnitude of  $S_{22}$  is almost close to  $-1 \text{ dB}$ , of  $S_{11}$  around  $-3 \text{ dB}$ , and of  $S_{21}$  around  $3.5 \text{ dB}$ , which is much lower than expected. We do not mind  $S_{21}$  too much because it is tested under an unmatched case. The magnitude of  $S_{12}$  is usually lower than  $-20 \text{ dB}$  and therefore disappears from the



**Figure 17.2.**  $S$  parameters from raw device testing,  $f = 850\text{--}940 \text{ MHz}$ ,  $I_D = 2.6 \text{ mA}$  (the central frequency 895 MHz is marked with dot on each trace) (a)  $S_{11}$ ,  $S_{22}$ , and  $\Gamma_{S,\text{opt}}$  on Smith Chart; (b) magnitude of  $S_{ij}$ , dB.

plot. Likewise, we do not mind  $S_{12}$  too much because the isolation in today's devices is usually sufficient unless a feedback circuit is added. A remarkable feature shown in Figure 17.2 is that the frequency response for all the  $S$  parameters is flattened, so that one does not need to worry about bandwidth at this point.

The second purpose of raw device testing is to examine the performance of the noise figure.

On the basis of Haus' theory (1960), the noise figure of a noisy block can be expressed by

$$NF = NF_{\min} + \frac{R_n}{G_s} [(G_s - G_{s,\text{opt}})^2 + (B_s - B_{s,\text{opt}})^2]. \quad (17.16)$$

where

- $NF$  = noise figure of noisy block,
- $NF_{\min}$  = minimum of noise figure of noisy block,
- $R_n$  = equivalent noise resistance,
- $Y_s$  = admittance of input source,
- $G_s$  = conductance of input source,
- $B_s$  = susceptance of input source,
- $Y_{s,\text{opt}}$  = optimal admittance of input source,
- $G_{s,\text{opt}}$  = optimal conductance of input source, and
- $B_{s,\text{opt}}$  = optimal susceptance of input source.

The noisy two-port block can reach a minimum of noise figure.

$$NF = NF_{\min}, \quad (17.17)$$

when

$$G_s = G_{s,\text{opt}}, \quad (17.18)$$

$$B_s = B_{s,\text{opt}}. \quad (17.19)$$

Expressions (17.18) and (17.19) can be written together, that is,

$$Y_s = Y_{s,\text{opt}}, \quad (17.20)$$

where

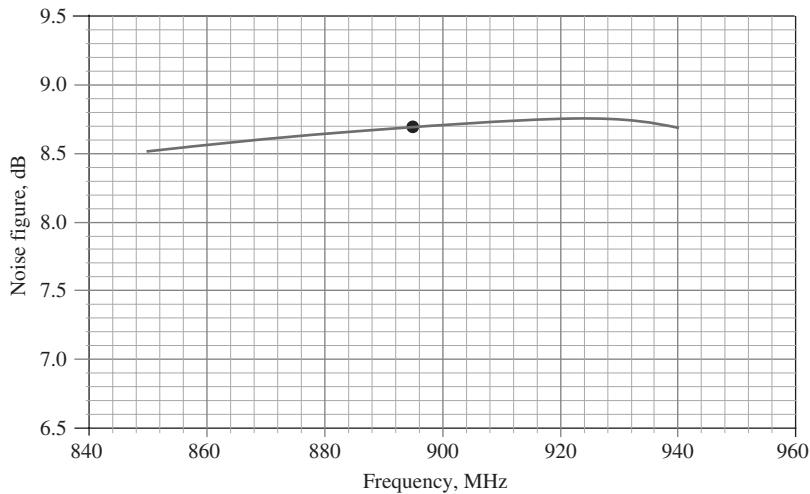
$$Y_s = G_s + jB_s, \quad (17.21)$$

$$Y_{s,\text{opt}} = G_{s,\text{opt}} + jB_{s,\text{opt}}. \quad (17.22)$$

On the Smith Chart, the optimal condition (17.20) is usually labeled by the corresponding reflection coefficient,  $\Gamma_{s,\text{opt}}$ , corresponding to  $Y_{s,\text{opt}} = G_{s,\text{opt}} + jB_{s,\text{opt}}$ .

$$\Gamma_s = \Gamma_{s,\text{opt}}. \quad (17.23)$$

Of course,  $\Gamma_{s,\text{opt}}$  is a complicated function mainly determined by the type, size, and transconductance of the raw device. It has been formularized in some technical books.



**Figure 17.3.** Noise figure from 850 to 940 MHz,  $I_D = 2.6$  mA and NF = 8.7 dB, when  $f = 895$  MHz.

Usually, an optimal source reflection coefficient,  $\Gamma_{S,\text{opt}}$ , is computed by the computer simulation program and can be displayed on the Smith Chart as shown in Figure 17.2(a). On the Smith Chart, its corresponding parameters,  $G_{\text{opt}}$ ,  $B_{\text{opt}}$ ,  $R_{\text{opt}}$ , and  $X_{\text{opt}}$ , can be read from the same point.

Noise figure would be expected to be at a minimum if the input impedance were adjusted to point N where its input impedance corresponds to  $\Gamma_{S,\text{opt}}$ . However, noise figure would be much higher than the expected minimum if the input impedance were adjusted to point G where its input impedance corresponds to  $S_{11}$ . Figure 17.3 shows the tested noise figure. In the entire frequency range, it is around 8.7 dB.

At this point, a question might be raised: Through impedance matching, the trace  $S_{11}$  can be pulled to  $50 \Omega$ , the center of Smith Chart. What would happen to the trace of  $\Gamma_{S,\text{opt}}$  then? We expect that the trace of  $\Gamma_{S,\text{opt}}$  would also be pulled to  $50 \Omega$ , the center of Smith Chart. Can we control the change of the trace of  $\Gamma_{S,\text{opt}}$  when the impedance matching network is implemented?

Let us take a look at the performance of the noise figure in the entire frequency range from the raw device testing. Figure 12.3 shows that the noise figure in the operating frequency range is

$$\text{NF} = 8.52 - 8.77 \text{ dB}, \quad \text{when } 850 \text{ MHz} < f < 940 \text{ MHz}, \quad (17.24)$$

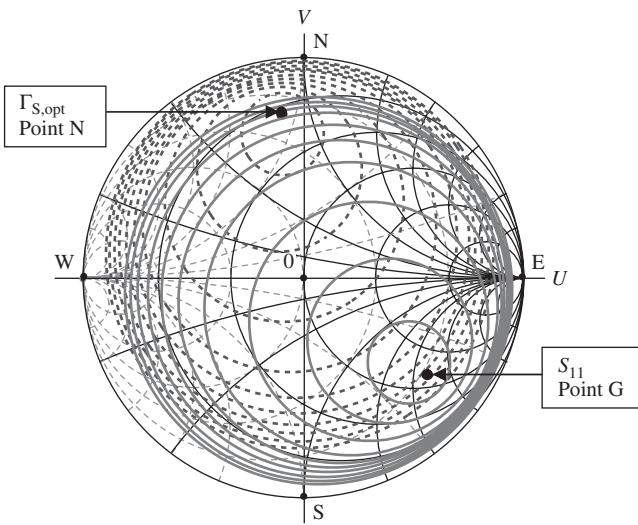
$$\text{NF} = 8.7 \text{ dB}, \quad \text{when } f = 895 \text{ MHz}. \quad (17.25)$$

The goal is

$$\text{NF} < 2.5 \text{ dB}, \quad \text{when } 850 \text{ MHz} < f < 940 \text{ MHz}. \quad (17.26)$$

It can be seen that in the entire operating frequency range, the noise figure is unacceptable.

Furthermore, let us examine the gain circles and noise figure circles at one frequency, say,  $f = 895$  MHz, on the input reflection coefficient plane. Figure 17.4 plots both the gain circles and noise figure circles together.



- Gain circles:  $G_{\max} = 3.0 \text{ dB}$  at point G, step =  $-1.0 \text{ dB}$ ,
- Noise figure circles:  $NF_{\min} = 5 \text{ dB}$  at point N, step =  $0.5 \text{ dB}$ .

**Figure 17.4.** Constant gain circles and constant noise figure circles when  $f = 895 \text{ MHz}$ .

The maximum of gain,  $G = G_{\max}$ , is located at point G, which is 3.0 dB, as shown in Figure 17.2(b). However, its noise figure is higher than the minimum  $NF_{\min} = 5 \text{ dB}$  as shown at point N, that is, at point G,

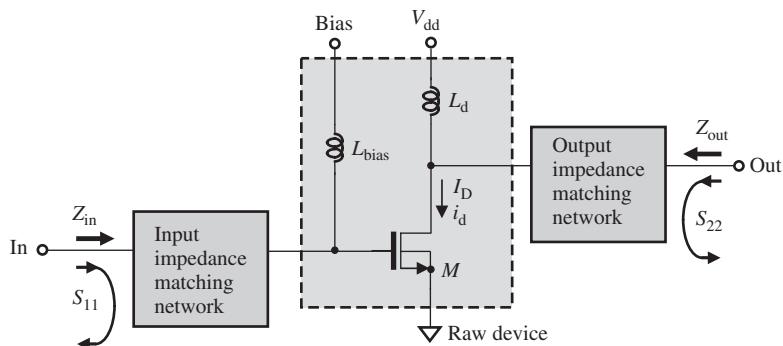
$$G = G_{\max} = 3.0 \text{ dB}, \quad \text{and} \quad NF = 8.7 \text{ dB}. \quad (17.27)$$

The minimum of the noise figure,  $NF = NF_{\min}$ , is located at point N, which is quite far from location of point G. Its gain is, of course, much lower than the maximum gain of 3 dB at point G, that is, at point N,

$$G = -4.8 \text{ dB}, \quad \text{and} \quad NF = NF_{\min} = 5 \text{ dB}. \quad (17.28)$$

Instead of point G or N, the raw device could be operating at the point O, the center of the Smith Chart, if the internal impedance of the signal source is  $50 \Omega$ . Its gain would be higher than  $-4.8 \text{ dB}$  but lower than  $3.0 \text{ dB}$ , whereas its noise figure would be lower than  $8.7 \text{ dB}$  but higher than  $5 \text{ dB}$ . As a matter of fact, the raw device can be operated with any source impedance, and therefore its impedance can be correspondingly adjusted to any point on the Smith Chart. The actual values of gain and noise figure can be read from Figure 17.4. These gains will not be higher than  $3.0 \text{ dB}$ , and the noise figures will not be lower than  $5 \text{ dB}$ . The ideal case is to find a raw device in which the maximum of gain,  $G_{\max}$ , and the minimum of noise figure,  $NF_{\min}$ , come together at one point on the Smith Chart. This seems almost impossible without a special scheme being involved. However, we should ask the following questions:

Might this be a temporary outcome due to the design work being in the preliminary stage? The next step is to build the input and output impedance matching networks based on the raw device testing. Is it possible to pull points G and N together after the input impedance matching network is built?



**Figure 17.5.** Input and output impedance matched networks are added to the raw device.  
 $f = 850\text{--}940 \text{ MHz}$ ,  $I_D = 2.6 \text{ mA}$ .

We will temporarily submit the noise figure to the will of Heaven, take care of the gain only, and move on to the task of input and output impedance matching as shown in Figure 17.5.

Impedance matching is a special scheme and key technology in RF circuit design. It is discussed somewhat in detail in this book. Because there are many ways to do impedance matching, many different results could be found. Figure 17.6 shows one such result. Point G of maximum gain, along with its gain circles, is moved from its original high-impedance location as shown in Figure 17.4 to a location near the center of the Smith Chart,  $50 \Omega$ . The maximum of gain is increased from 3.0 dB of Figure 17.4 to 13 dB of Figure 17.6 as a result of the impedance matching. However, its noise figure has not reached its minimum,  $\text{NF}_{\min} = 1.8 \text{ dB}$ , as shown at point N in Figure 17.6.

At point G,

$$G = G_{\max} = 13 \text{ dB}, \quad \text{and} \quad \text{NF} = 3.5 \text{ dB}. \quad (17.29)$$

In addition, owing to impedance matching, the point N of minimum noise figure, along with its noise figure circles, is moved from point N in the upper left area of Smith Chart in Figure 17.4 to its new location in Figure 17.6. The minimum of noise figure is dropped from the 5 dB of Figure 17.4 to 1.8 dB of Figure 17.6, as a result of the impedance matching. That is, at point N,

$$G = 8.2 \text{ dB}, \quad \text{and} \quad \text{NF} = \text{NF}_{\min} = 1.8 \text{ dB}. \quad (17.30)$$

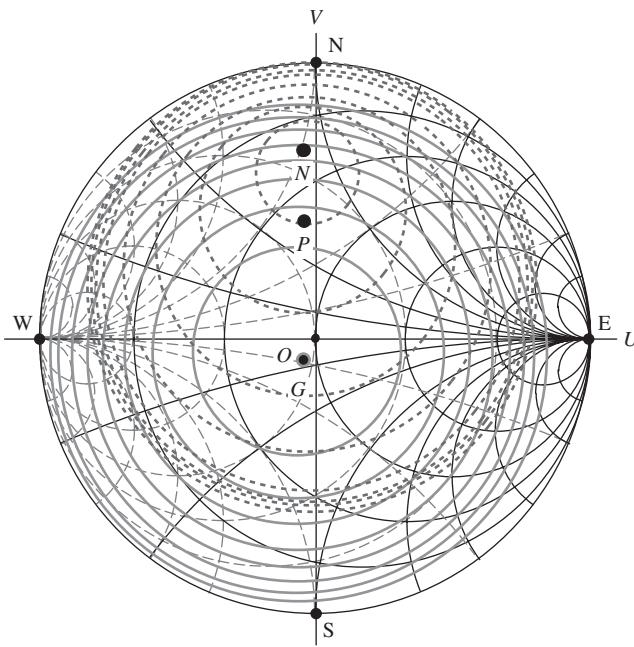
The goals as mentioned previously were

$$G > 10 \text{ dB}, \quad \text{when } 850 \text{ MHz} < f < 940 \text{ MHz}; \quad (17.31)$$

$$\text{NF} < 2.5 \text{ dB}, \quad \text{when } 850 \text{ MHz} < f < 940 \text{ MHz}. \quad (17.32)$$

It can be seen that the goals at either point G or point N are still insufficient. In order to satisfy these goals, a trade-off is usually taken between the maximum gain,  $G_{\max}$ , and the minimum of noise figure,  $\text{NF}_{\min}$ . For instance, if we force the input impedance matching to point P, where the gain is lower than  $G_{\max}$  and the noise figure is higher than  $\text{NF}_{\min}$ , we satisfy the goals because at point P,

$$G = 11.3 \text{ dB}, \quad \text{and} \quad \text{NF} = 2.3 \text{ dB}. \quad (17.33)$$

Input reflection coefficient  $\Gamma_S$  plane

- Gain circles:  $G_{\max} = 13$  dB at point G, step = 1.0 dB,
- Noise figure circles:  $NF_{\min} = 1.8$  dB at point N, step = 0.5 dB,

**Figure 17.6.** Constant gain circles and constant noise figure circles when  $f = 895$  MHz.

It can be seen that the trade-off sacrifices gain to obtain low noise. This is, of course, adapted reluctantly. The probability of success without a trade-off just relies on luck, although with hard work, the probability is higher than winning a lottery. There is very little chance for a designer to obtain a maximum of gain and a minimum of noise figure at the same time.

However, this is what designers have been attempting in the past decades.

The ideal case to be pursued is that point N and G, and hence the gain circles and noise figure circles exactly overlap together at the center of the Smith Chart when the input and output impedances are matched to  $50 \Omega$ .

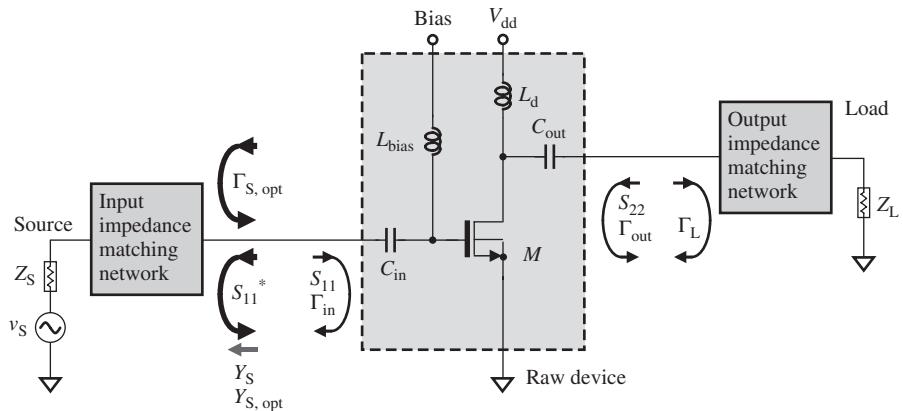
This is a very challenging but exciting project.

### 17.2.3 Challenge for a Good LNA Design

In order to solve the dilemma of overlapping the maximum gain with the minimum of noise figure at  $50 \Omega$  in the LNA circuit design, I suggest a repeated study of Haus' articles to thoroughly understand the meaning of the "optimum of source reflection coefficient,  $\Gamma_{S,\text{opt}}$ ." In the history of the sciences of developing technology, every forward step requires a correct concept and diligent practice.

**17.2.3.1 Optimum of Source Voltage Reflection Coefficient,  $\Gamma_{\text{opt}}$ .** Let us return to raw device testing.

To clarify and to distinguish the various reflection coefficients, Figure 17.7 emphasizes the directions of  $\Gamma_{S,\text{opt}}$ ,  $\Gamma_S$ ,  $S_{11}$ ,  $\Gamma_{\text{in}}$ , and  $S_{11}^*$  immediately at the input of the raw device ( $\Gamma_{S,\text{opt}}$  should not be misunderstood as the input reflection coefficient of the raw



**Figure 17.7.** Directions of  $\Gamma_{S,\text{opt}}$ ,  $\Gamma_{\text{in}}$  and  $\Gamma_{\text{out}}$ , and  $S_{11}$  and  $S_{11}^*$  in the schematic for the raw device testing.

device,  $\Gamma_{\text{in}}$ , which is almost equal to  $S_{11}$  if  $S_{12}$  is negligible). The following concepts are essential to solving the problem:

- In order to enable the raw device to approach the minimum of noise figure,  $\Gamma_{S,\text{opt}}$  is a required optimum of source voltage reflection coefficient,  $\Gamma_S$ , looking from the raw device toward the source.
- The actual value of source voltage reflection coefficient,  $\Gamma_S$ , looking from the raw device toward the source is  $S_{11}^*$  because the actual voltage reflection coefficient looked from the source toward the raw device is  $S_{11}$ .

Then, based on Haus' theory, the minimum noise figure of the raw device could be obtained if

$$\Gamma_{S,\text{opt}} = S_{11}^*. \quad (17.34)$$

Conversely, it is impossible to approach the minimum noise figure of the raw device with the maximum gain if

$$\Gamma_{S,\text{opt}} \neq S_{11}^*. \quad (17.35)$$

In the LNA circuit design the 1<sup>st</sup> objective is to get minimum noise figure. Once the condition (17.34) is satisfied, by which the minimum noise figure could be obtained. We then pursue the 2<sup>nd</sup> objective, the maximum gain by means of implementing input and output impedance matching network. Consequently, both of  $G_{\text{max}}$  and  $\text{NF}_{\text{min}}$  are simultaneously achieved.

I developed the condition (17.34) in the 1990s and have been applying it in many successful design projects, although it has not been published. We will now pursue expression (17.34). This expression represents the condition by which a perfect LNA with a minimum noise figure and maximum gain can be simultaneously approached after the input and output impedance matching networks are implemented.

Let us return to Figure 17.2 or 17.4 for raw device testing. It can be found that  $\Gamma_{S,\text{opt}}$  and  $S_{11}^*$  do not satisfy condition (17.34), but are in the condition of (17.35). This is the reason of the problem in Figure 17.6, where the points G of maximum gain and N of

minimum noise figure were very far apart and not in conjugate locations to each other; therefore, the gain circles and noise figure circles did not overlap together after the implementation of the input and output impedance matching networks. Consequently, it was impossible to simultaneously achieve a maximum of gain and a minimum of noise figure.

It should be reiterated that the implementation of the input impedance matching network cannot change the deviation status between  $\Gamma_{S,\text{opt}}$  and  $S_{11}^*$ , since  $\Gamma_{S,\text{opt}}$  and  $S_{11}^*$  are determined by the raw device only and are basically independent of the impedance matching network.

**17.2.3.2 Simultaneous Approach to Both  $NF_{\min}$  and  $G_{\max}$ .** On the surface, the purpose of raw device testing is the starting point for impedance matching. Actually, the more important purpose of this testing is to judge whether a raw device could build a good LNA or not, that is, to check whether condition (17.34) is satisfied or not. Should condition (17.34) not be satisfied, impedance matching design should be halted and possible means to satisfy or approach condition (17.34) should be sought.

Three major schemes are used to satisfy condition (17.34) in the step of raw device testing:

1. *Increasing or Decreasing the Current Drain,  $I_D$ .* The  $S$  parameters as well as the values of  $\Gamma_{S,\text{opt}}$  change as the current drain is varied. The condition  $\Gamma_{S,\text{opt}} = S_{11}^*$  could be reached with an appropriate amount of current drain.
2. *Changing the Device Size.* The  $S$  parameters as well as the values of  $\Gamma_{S,\text{opt}}$  change as the device size is varied. The condition  $\Gamma_{S,\text{opt}} = S_{11}^*$  could be reached with an appropriate device size. Of course, this scheme is available only to the IC designer and not to the designer implementing a circuit by discrete parts.
3. *Addition of Degeneration Part.* According to empirical design experience in practical design, this is an easy way to achieve success, and so on.

Owing to space limitations, we are going to apply only the third scheme to the design example. Readers are encouraged to apply the first and second one in their designs since both of them are very effective as well. Usually, it is easier to approach the condition  $\Gamma_{S,\text{opt}} = S_{11}^*$  by the use of multiple combined schemes rather than only one.

Figure 17.8 shows a degeneration inductor,  $L_{\text{degen}}$ , applied to the source of the MOSFET.

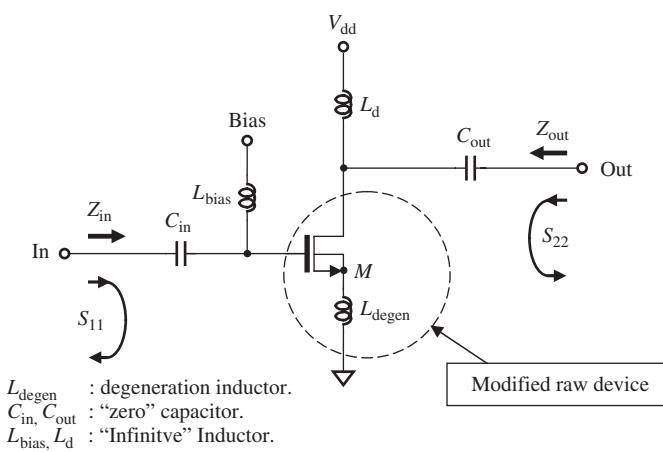


Figure 17.8. Setup for modified raw device testing by adding of degeneration inductor.

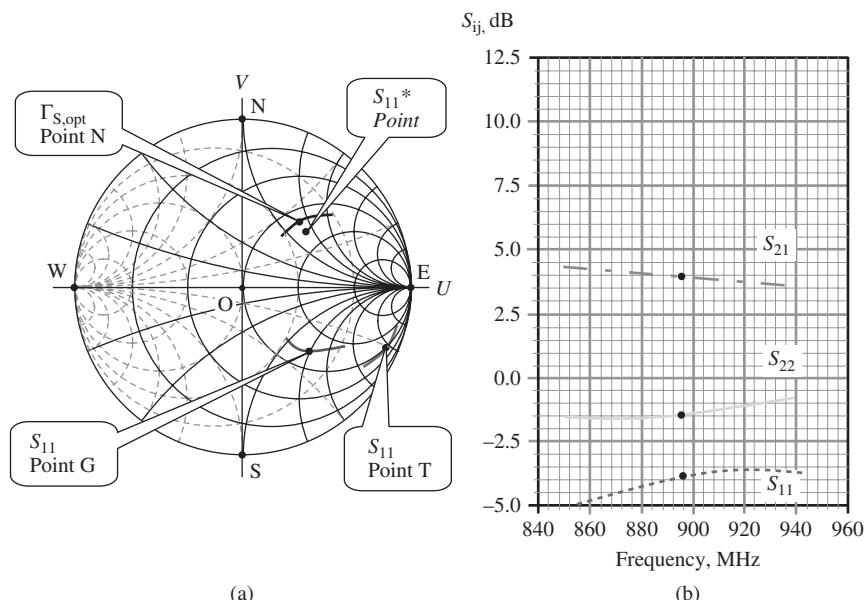
The degeneration inductor,  $L_{\text{degen}}$ , is a simple and tiny part. For example, in the RFIC circuit design for gigahertz of operating frequency range, it is about 0.5–5 turns. The actual value of the degeneration inductor applied to this design sample is 10 nH.

In the simulation, it is connected to the source of the device; the device combined with the degeneration inductor is a modified raw device. The input impedance, of course, is different from the original impedance without the degeneration inductor.

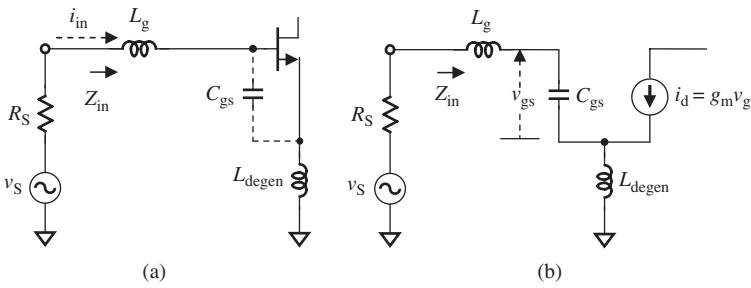
By adjusting the value of the degeneration inductor to change its input impedance, condition (17.34) can be satisfied. As the degeneration inductor is added, the trace of  $S_{11}$  is moved from its original location as shown in Figure 17.2(a) to a new location as shown in Figure 17.9(a). Consequently, the traces of  $S_{11}^*$  and  $\Gamma_{S,\text{opt}}$  on the input reflection coefficient plane are close to each other. The corresponding variation of impedance can be determined by the equation derived in the following subsection, in which both resistance and reactance are increased.

Figure 17.9(a) shows that the location of  $S_{11}$  is somewhat closer to  $50 \Omega$  than in Figure 17.2(a). Correspondingly, Figure 17.9(b) shows that the magnitude of  $S_{11}$  is now around -3.75 dB, which is 0.7 dB improved from that of Figure 17.2(b). The magnitude of  $S_{21}$  is about 4.0 dB, which is 0.5 dB higher than the original shown in Figure 17.2(b). Again, Figure 17.9 shows that the frequency response for all the  $S$  parameters is flat; we need not worry about the bandwidth for now. As a matter of fact, it does not matter too much in the variation of magnitude of  $S$  parameters because the input and output impedances are not matched yet.

From Figure 17.9(a) it is easy to imagine that  $S_{11}$  and  $\Gamma_{S,\text{opt}}$  could be pulled close together near the center of Smith Chart after impedance matching is done although  $S_{11}^*$  and  $\Gamma_{S,\text{opt}}$  are still not at the same point.



**Figure 17.9.**  $S$  parameters from modified raw device testing,  $f = 850$ –940 MHz,  $I_D = 2.6$  mA. (The intermediate frequency 895 MHz is marked with dot on each trace.) (a)  $S_{11}$ ,  $S_{22}$ , and  $\Gamma_{S,\text{opt}}$  on Smith Chart; (b) magnitude of  $S_{ij}$ , dB.



**Figure 17.10.** The input stage and its equivalent of a modified raw device.

Before the discussion of input and output impedance matching, we are going to examine the variation of the input impedance due to the addition of the degeneration inductor.

### 17.2.3.3 Variation of Input Impedance due to Degeneration Inductor.

Figure 17.10 plots the input stage of a modified raw device and its equivalent.

From Figure 17.10(b), we have

$$\begin{aligned} v_{in} &= i_{in} \left( jL_g \omega + \frac{1}{jC_{gs}\omega} \right) + (i_{in} + i_d) jL_{degen} \\ \omega &= i_{in} j \left( L_g \omega - \frac{1}{C_{gs}\omega} \right) + (i_{in} + g_m V_{gs}) jL_{degen} \omega, \end{aligned} \quad (17.36)$$

$$v_{in} = i_{in} j \left( L_g \omega - \frac{1}{C_{gs}\omega} \right) + \left( i_{in} + g_m i_{in} \frac{1}{jC_{gs}\omega} \right) jL_{degen} \omega, \quad (17.37)$$

$$Z_{in} = \frac{v_{in}}{i_{in}} = R_{in} + jX_{in}, \quad (17.38)$$

$$\begin{aligned} Z_{in} &= j(L_g + L_{degen})\omega + \frac{1}{jC_{gs}\omega} + \frac{g_m}{C_{gs}} L_{degen} \\ &= j \left[ (L_g + L_{degen})\omega - \frac{1}{C_{gs}\omega} \right] + L_{degen} \omega_T, \end{aligned} \quad (17.39)$$

where

$$\omega_T = \frac{g_m}{C_{gs}}, \quad (17.40)$$

which is called the *cut-off frequency*.

$$R_{in} = L_{degen} \omega_T, \quad (17.41)$$

$$X_{in} = (L_g + L_{degen})\omega - \frac{1}{C_{gs}\omega}. \quad (17.42)$$

On the other hand,

$$G_{m,\text{eff}} = g_m Q_{\text{in}} = \frac{g_m}{\omega C_{\text{gs}}(R_S + L_{\text{degen}}\omega_T)} = \frac{\omega_T}{\omega R_S \left(1 + \frac{L_{\text{degen}}\omega_T}{R_S}\right)} = \frac{\omega_T}{2\omega R_S}, \quad (17.43)$$

when the input impedance is matched, that is,

$$R_s = L_{\text{degen}}\omega_T, \quad (17.44)$$

where  $Q_{\text{in}}$  is the effective input  $Q$  of the circuit. Note that  $R_g$  has been neglected relative to  $R_s$ . This expression is valid at resonance where the signal voltage across  $C_{\text{gs}}$  is equal to  $Q_{\text{in}}$  times the input signal. It should be noted that the overall transconductance is apparently independent of  $g_m$ , the intrinsic device transconductance.

From expressions (17.41) and (17.42) it can be seen that the input impedance is increased

$$\Delta Z_{\text{in}} = L_{\text{degen}}\omega_T + jL_{\text{degen}}\omega, \quad (17.45)$$

because of the presence of  $L_{\text{degen}}$ .

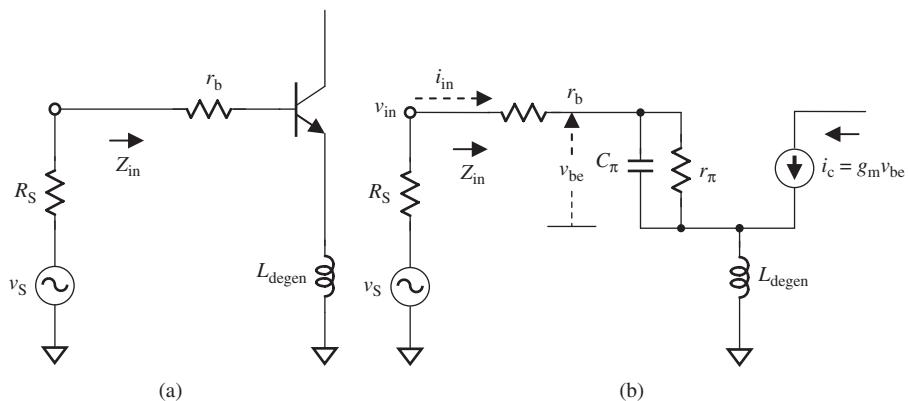
If the device is a bipolar transistor, the input portion can be drawn as shown in Figure 17.11.

Figure 17.11 shows the degeneration inductor connected from emitter to ground. The input impedance can be calculated as follows:

$$r_\pi//C_\pi = \frac{r_\pi}{1 + jC_\pi\omega r_\pi} = \frac{r_\pi}{1 + (C_\pi\omega r_\pi)^2} - j\frac{C_\pi\omega r_\pi^2}{1 + (C_\pi\omega r_\pi)^2}, \quad (17.46)$$

$$v_{\text{in}} = i_{\text{in}}(r_b + r_\pi//C_\pi) + (i_{\text{in}} + i_c)L_{\text{degen}}\omega, \quad (17.47)$$

$$i_c = g_m v_{\text{be}} = g_m i_{\text{in}}(r_\pi//C_\pi), \quad (17.48)$$



**Figure 17.11.** The input stage of a bipolar transistor with a degeneration inductor. (a) Schematic of input stage; (b) equivalent of input stage.

$$Z_{in} = \frac{v_{in}}{i_{in}} = r_b + r_\pi // C_\pi + j[1 + (r_\pi // C_\pi)g_m]L_{degen}\omega, \quad (17.49)$$

$$Z_{in} = r_b + r_\pi + \frac{L_{degen}C_\pi\omega^2g_m r_\pi^2}{1 + (C_\pi\omega r_\pi)^2} + j \left[ 1 + \frac{\left(g_m - \frac{C_\pi}{L_{degen}}r_\pi\right)r_\pi}{1 + (C_\pi\omega r_\pi)^2} \right] L_{degen}\omega. \quad (17.50)$$

$$R_{in} = r_b + r_\pi + \frac{L_{degen}C_\pi\omega^2g_m r_\pi^2}{1 + (C_\pi\omega r_\pi)^2}, \quad (17.51)$$

$$X_{in} = \left[ 1 + \frac{\left(g_m - \frac{C_\pi}{L_{degen}}r_\pi\right)r_\pi}{1 + (C_\pi\omega r_\pi)^2} \right] L_{degen}\omega. \quad (17.52)$$

It should be noted that when

$$g_m = \frac{C_\pi}{L_{degen}}r_\pi, \quad (17.53)$$

then

$$X_{in} = L_{degen}\omega. \quad (17.54)$$

In cases with low frequency, the capacitor  $C_\pi$  can be neglected. Then (17.50), (17.51), and (17.52) become

$$Z_{in} = r_b + r_\pi + j(1 + g_m r_\pi)L_{degen}\omega, \quad (17.55)$$

$$R_{in} = r_b + r_\pi, \quad (17.56)$$

$$X_{in} = (1 + r_\pi g_m)L_{degen}\omega. \quad (17.57)$$

In cases with high frequency,  $R_{in}$  is increased, while  $X_{in}$  is decreased from those in cases with low frequency.

From expressions (17.51) and (17.52) it can be seen that the input impedance is increased

$$\Delta Z_{in} = \frac{L_{degen}C_\pi\omega^2g_m r_\pi^2}{1 + (C_\pi\omega r_\pi)^2} + j \left[ 1 + \frac{g_m r_\pi}{1 + (C_\pi\omega r_\pi)^2} \right] L_{degen}\omega, \quad (17.58)$$

because of the presence of  $L_{degen}$ .

By comparing (17.45) with (17.58), the increased resistance is almost the same, while the increased reactance of the bipolar transistor has a higher slope than that of the MOSFET if  $C_\pi$  in the bipolar transistor is equivalent to  $C_{gs}$  in the MOSFET and  $C_\pi\omega r_\pi \gg 1$ .

On the basis of these two expressions, the designer is able to estimate how far the trace of  $S_{11}$  will be moved on the Smith Chart, so as to satisfy condition (17.34), by which the maximum gain and minimum noise figure circles can be almost entirely overlapped near the center of the Smith Chart.

### 17.2.4 Input and Output Impedance Matching

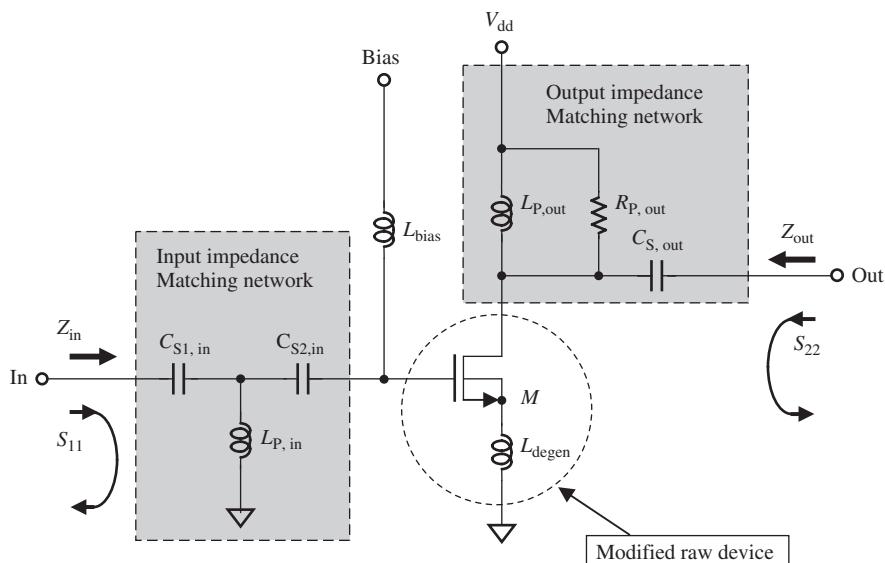
Impedance matching is a core technology in RF circuit design and is discussed somewhat in detailed in this book. The process of impedance matching for the design sample shown above will be neglected. Instead, the final input and output impedance matching networks and their performance are shown in Figures 17.12 and 17.13, respectively.

Figure 17.12 plots the impedance matching networks, which consist of the following parts:

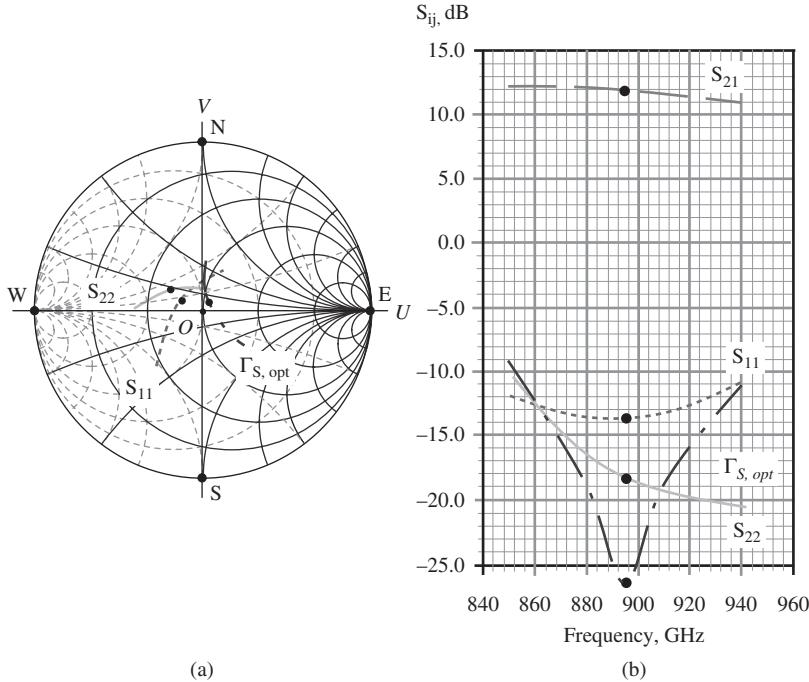
- In the input impedance network:  $C_{S1,\text{in}} = 1 \text{ pF}$ ,  $L_{P,\text{in}} = 20 \text{ nH}$ ,  $C_{S2,\text{in}} = 39 \text{ pF}$ ;
- In the output impedance network:  $L_{P,\text{out}} = 15 \text{ nH}$  and  $R_{P,\text{out}} = 1500 \Omega$ ,  $C_{S,\text{out}} = 1.6 \text{ pF}$ .

The input impedance matching network is built with a T type circuitry of  $C-L-C$ , by which the trace of  $S_{11}$  is pulled to a location near the center of the Smith Chart, which corresponds to the reference impedance point of  $50 \Omega$ . The capacitor  $C_{S2,\text{in}} = 39 \text{ pF}$  is actually a “zero” capacitor and functions as the DC blocking part. By means of  $L_{P,\text{in}}$  in parallel, the trace of  $S_{11}$  is pulled counterclockwise to the resistance circle of  $50 \Omega$  along the constant conductance circle, and then, by means of  $C_{S1,\text{in}}$  in series, the trace of  $S_{11}$  is pulled counterclockwise to the area near the reference impedance point  $50 \Omega$  along the constant resistance circle.

The output impedance matching network consists of an inductor, a resistor, and a capacitor. The inductor,  $L_{P,\text{out}} = 15 \text{ nH}$ , is the main part in the output impedance matching network, by which the trace of  $S_{22}$  is pulled counterclockwise to a location near the  $50\text{-}\Omega$  resistance circle along the constant conductance circle. The capacitor  $C_{S,\text{out}} = 1.6 \text{ pF}$  is the next part, by which the  $S_{22}$  is drawn counterclockwise to a location near the center of Smith Chart corresponding to the reference impedance point  $50 \Omega$  along the constant



**Figure 17.12.** Impedance matching of modified raw device by parts. At input:  $C_{S1,\text{in}} = 1 \text{ pF}$ ,  $L_{P,\text{in}} = 20 \text{ nH}$ , and  $C_{S2,\text{in}} = 39 \text{ pF}$ ; at output:  $L_{P,\text{out}} = 15 \text{ nH}$ ,  $R_{P,\text{out}} = 1500 \Omega$ , and  $C_{S,\text{out}} = 1.6 \text{ pF}$ .



**Figure 17.13.**  $S$  parameters of the design example,  $f = 850\text{--}940\text{ MHz}$ ,  $I_D = 2.6\text{ mA}$ . (The intermediate frequency 895 MHz is marked with a dot on each trace.) The input and output impedances are matched by the following parts. At input:  $C_{S1,\text{in}} = 1\text{ pF}$ ,  $L_{P,\text{in}} = 20\text{ nH}$ , and  $C_{S2,\text{in}} = 39\text{ pF}$ ; at output:  $L_{P,\text{out}} = 15\text{ nH}$ ,  $R_{P,\text{out}} = 1500\text{ }\Omega$ , and  $C_{S,\text{out}} = 1.6\text{ pF}$ . (a)  $S_{11}$ ,  $S_{22}$ , and  $\Gamma_{S,\text{opt}}$  on Smith Chart; (b) magnitude of  $S_{ij}$ , dB.

resistance circle. The resistor  $R_{P,\text{out}}$  is a de- $Q$  part, by which the bandwidth is widened and the LNA is changed from an unstable state to a stable state.

From Figure 17.13(a) it can be seen that  $S_{11}$  and  $\Gamma_{S,\text{opt}}$  are not completely superimposed on each other. The deviation arises from the fact that  $S_{11}^*$  and  $\Gamma_{S,\text{opt}}$  of the modified raw device shown in Figure 17.8 are not exactly overlapped together. However, this is normal phenomenon in the actual design because all parts have tolerance, and in addition, absolutely ideal cases never happen in actual engineering design, and the absolute accuracy of simulation processing is never realistic. As long as the two points  $S_{11}^*$  and  $\Gamma_{S,\text{opt}}$  are pulled close enough to each other, it will be sufficient to satisfy the goals of the design.

Figure 17.13(b) shows that

- *Input Return Loss,  $S_{11}$*

$$-13.8\text{ dB} < S_{11} < -11.0\text{ dB}, \quad \text{when } 850\text{ MHz} < f < 940\text{ MHz}, \quad (17.59)$$

$$S_{11} = -13.7\text{ dB}, \quad \text{when } f = 895\text{ MHz}. \quad (17.60)$$

- *Output Return Loss,  $S_{22}$*

$$-20.5\text{ dB} < S_{22} < -10.5\text{ dB}, \quad \text{when } 850\text{ MHz} < f < 940\text{ MHz}, \quad (17.61)$$

$$S_{22} = -18.5\text{ dB}, \quad \text{when } f = 895\text{ MHz}. \quad (17.62)$$

- Gain,  $S_{21}$

$$11.0 \text{ dB} < S_{21} < -12.2 \text{ dB}, \quad \text{when } 850 \text{ MHz} < f < 940 \text{ MHz}, \quad (17.63)$$

$$S_{21} = -12.0 \text{ dB}, \quad \text{when } f = 895 \text{ MHz}. \quad (17.64)$$

- Optimum of Source Reflection Coefficient,  $\Gamma_{S,\text{opt}}$

$$-26.5 \text{ dB} < \Gamma_{S,\text{out}} < -9.0 \text{ dB}, \quad \text{when } 850 \text{ MHz} < f < 940 \text{ MHz}, \quad (17.65)$$

$$\Gamma_{S,\text{out}} = -26.5 \text{ dB}, \quad \text{when } f = 895 \text{ MHz}. \quad (17.66)$$

The isolation  $S_{12}$  is not shown in Figure 17.13(b) because it is lower than  $-25 \text{ dB}$  in the entire frequency range.

It can be seen that the gain is reasonable and the bandwidth is wide enough. In addition, the values of the parts are appropriate. Now let us check its noise performance.

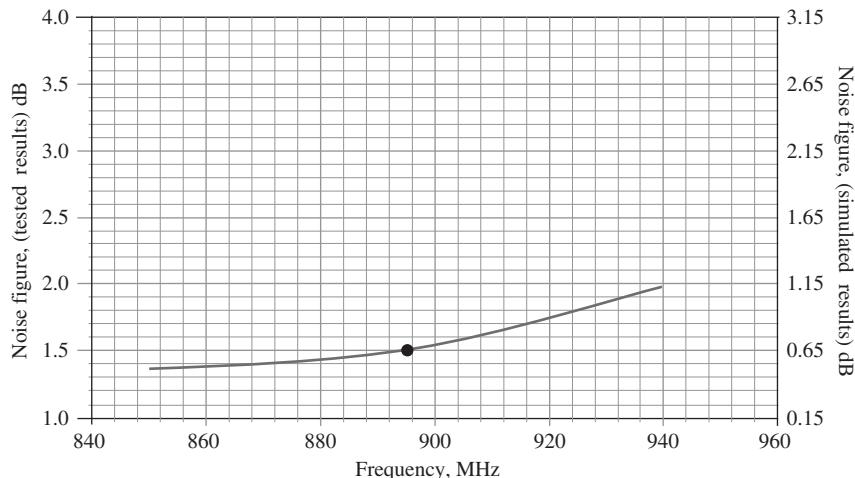
### 17.2.5 Gain Circles and Noise Figure Circles

The noise performance of the design sample is good after impedance matching is done. This is expected because the modified raw device is designed for simultaneously approaching maximum gain and minimum noise figure. Figure 17.14 shows the performance of the noise figure in the entire frequency range, that is,

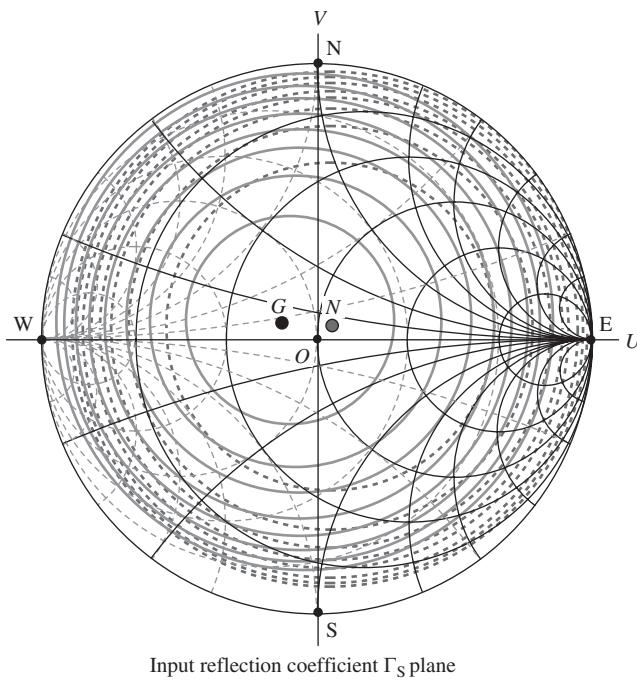
$$1.37 \text{ dB} < \text{NF} < 1.97 \text{ dB}, \quad \text{when } 850 \text{ MHz} < f < 940 \text{ MHz}, \quad (17.67)$$

$$\text{NF} = 1.5 \text{ dB}, \quad \text{when } f = 895 \text{ MHz}. \quad (17.68)$$

The values of the noise figure shown in Figure 17.14 are denoted with two coordinates in the vertical direction, one for simulated values and another for actual tested



**Figure 17.14.** Noise figure of the design example,  $f = 850\text{--}940 \text{ MHz}$ ,  $I_D = 2.6 \text{ mA}$ . (The central frequency 895 MHz is marked with a dot on the trace.)  $\text{NF} = 1.5 \text{ dB}$  when  $f = 895 \text{ GHz}$ .

Input reflection coefficient  $\Gamma_S$  plane

- Gain circles:  $G_{\max} = 12$  dB at point G, step = 1.0 dB,
- Noise figure circles:  $NF_{\min} = 1.5$  dB at point N, step = 0.5 dB.

**Figure 17.15.** Constant gain circles and constant noise figure circles when  $f = 895$  MHz.

results. It can be seen that the actual tested values of the noise figure are 0.85 dB higher than the simulated values. The good news is that the noise figure in the entire frequency range is better than the previously stated goals.

Exciting results are found from the plot of the gain circles and noise figure circles as shown in Figure 17.15. This is a plot for the operating frequency of  $f = 895$  MHz.

After the input and output impedances are matched, the center of the gain circles is point G, which is quite close to the center of the Smith Chart. The maximum of gain at this operating frequency is 12 dB, which is consistent with the results shown in Figure 17.13(b). On the other hand, the center of the noise figure circles is point N, which is also very close to the center of the Smith Chart. The minimum of noise figure at this operating frequency is 1.5 dB, which is consistent with the results shown in Figure 17.14. The wonderful feature of Figure 17.15 is that the gain circles and noise circles almost overlap each other perfectly. The two centers are very close to each other and near the center of the Smith Chart.

As mentioned above, the deviation between point N and point O is due to the incomplete overlapping in the modified raw device so that condition (17.34),  $\Gamma_{S,\text{opt}} = S_{11}^*$ , is not perfectly satisfied. In reality, it is impossible to reach an ideal goal predicted by theory. Deviation between the practical design and the circuit theory always existed.

### 17.2.6 Stability

One of the important specifications for an LNA design is stability. It is obvious that an LNA may become an oscillator if it is unstable in the circuit performance. Therefore, the

designer must examine its stability after considering the topology of the circuit and the value of parts, such as the design sample shown in Figure 17.12 are confirmed.

It is well known that the Smith Chart is a reflection coefficient plane, called a  $\Gamma$  plane, in which the impedance,  $z$ , is plotted by means of the following relationship:

$$z = \frac{Z}{Z_0} = \frac{1 + \Gamma}{1 - \Gamma}, \quad (17.69)$$

where  $z$  is a normalized impedance by the reference impedance,  $Z_0$ , which is usually  $50 \Omega$ .

Expression (17.69) indicates an important fact, that is,

$$z > 0, \quad \text{if } |\Gamma| < 0. \quad (17.70)$$

Otherwise,

$$z < 0, \quad \text{if } |\Gamma| > 0. \quad (17.71)$$

This is a universal relationship for any kind of impedance and its corresponding reflection coefficient. If an impedance looking into a two-port block or a terminal is positive, then the block or terminal is stable; if an impedance looking into a two-port block or a terminal is negative, then the block or terminal is unstable. As shown in Figure 17.16, there are four reflection coefficients,  $\Gamma_S$ ,  $\Gamma_L$ ,  $\Gamma_{in}$ , and  $\Gamma_{out}$ , and four  $S$  parameters,  $S_{11}$ ,  $S_{22}$ ,  $S_{21}$ , and  $S_{12}$ .

The conditions for unconditional stability for all of the reflection coefficients shown in Figure 17.16 are

$$|\Gamma_S| < 1, \quad (17.72)$$

$$|\Gamma_L| < 1, \quad (17.73)$$

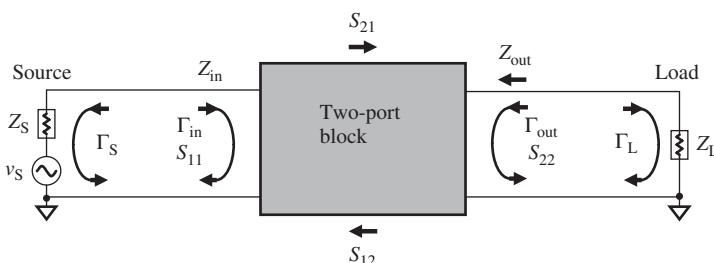
$$|\Gamma_{in}| < 1, \quad (17.74)$$

$$|\Gamma_{out}| < 1. \quad (17.75)$$

The two-port block is unstable if the magnitude of any reflection coefficient above is greater than 1. A critical case is that in which the input and output reflection coefficients are equal to 1, that is,

$$|\Gamma_{in}| = 1, \quad (17.76)$$

$$|\Gamma_{out}| = 1. \quad (17.77)$$



**Figure 17.16.** Different reflection coefficients and  $S$  parameters in a two-port block.

It is well known that

$$\Gamma_{\text{in}} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L}, \quad (17.78)$$

$$\Gamma_{\text{out}} = S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1 - S_{11}\Gamma_S}. \quad (17.79)$$

The conditions in the critical case therefore are

$$\left| S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \right| = 1, \quad (17.80)$$

$$\left| S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1 - S_{11}\Gamma_S} \right| = 1. \quad (17.81)$$

The critical values of  $\Gamma_S$  and  $\Gamma_L$  can be solved from the equations (17.80) and (17.81); they are

$$\left| \Gamma_L - \frac{(S_{22} - \Delta S_{11}^*)^*}{|S_{22}|^2 - |\Delta|^2} \right| = \left| \frac{S_{12}S_{21}}{|S_{22}|^2 - |\Delta|^2} \right|, \quad (17.82)$$

$$\left| \Gamma_S - \frac{(S_{11} - \Delta S_{22}^*)^*}{|S_{11}|^2 - |\Delta|^2} \right| = \left| \frac{S_{12}S_{21}}{|S_{11}|^2 - |\Delta|^2} \right|. \quad (17.83)$$

where

$$|\Delta| = |S_{11}S_{22} - S_{12}S_{21}|. \quad (17.84)$$

On the Smith Chart, they appear as two circles and are called the output and input stability circles, respectively. These two stability circles are very useful in the analysis of potentially unstable blocks. However, we are not going to repeat this analysis since it has already been discussed in many textbooks in great detail. What we are interested in is how to quickly judge the stability of a designed block.

Historically, a  $K$  factor, which is a function of  $S$  parameters, has been defined as

$$K = 1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2. \quad (17.85)$$

In terms of mathematical processing, the necessary and sufficient conditions for a two-port block to be unconditionally stable are

$$K > 1 \quad (17.86)$$

and

$$|\Delta| < 1. \quad (17.87)$$

For many years, the stability of a two-port block has been judged by the  $K$  factor and the intermediate parameter  $\Delta$  in terms of expressions (17.86) and (17.87).

In recent years, however, they have been replaced by the  $\mu$  factor, which is defined as

$$\mu = \frac{1 - [\text{mag}(S_{11})]^2}{\text{mag}[S_{22} - \Delta \text{ conj}(S_{11})] + \text{mag}(S_{21}S_{12})}. \quad (17.88)$$

In expression (17.88), the symbols “ $\text{mag}(\dots)$ ” and “ $\text{mag}[\dots]$ ” denote the magnitude of the parameters in the parenthesis or brace, and the symbol “ $\text{conj}(\dots)$ ” denotes the conjugated value of the parameters in the parenthesis.

The block is unconditionally stable if

$$\mu > 1. \quad (17.89)$$

Otherwise, it is potentially unstable.

Rather than the two conditions of the  $K$  and  $\Delta$  factors, condition (17.89) using the factor of  $\mu$  is simplest and most convenient criterion. From (17.88) it can be seen that the value of  $\mu$  is dependent on the tested  $S$  parameters only. In today's simulation by ADS and Cadence simulation tools, the values of  $\mu$  for the entire frequency range can be promptly and easily displayed on screen.

The stability of the raw device shown in Figure 17.8 is unstable because the simulation shows its  $\mu$  value as

$$\mu = 0.95 < 1. \quad (17.90)$$

After the steps of raw device testing, one should perform analysis on the sources of instability.

In the schematic shown in Figure 17.8, the instability may come from the drain inductor,  $L_d$ , and the degenerator inductor,  $L_{\text{degen}}$ .

In order to retain as much voltage drop across the device as possible, so as to reduce intermodulation, it is preferred to use a drain inductor,  $L_d$ , rather than a resistor, although an inductor is much more expensive than a resistor. In order to satisfy the overlapping condition (17.34) for  $\text{NF}_{\min}$  without a significant increase in the noise figure, it is preferred to use a degeneration inductor,  $L_{\text{degen}}$ , rather than a degeneration resistor,  $R_{\text{degen}}$ .

The drain inductor  $L_d$  in Figure 17.8 is adjusted and renamed as  $L_{P,\text{out}}$  as shown in Figure 17.12. The resistor,  $R_{P,\text{out}}$ , is used to de- $Q$  the load inductor  $L_{P,\text{out}}$ , in order to prevent oscillation. On the other hand, the value of  $L_{\text{degen}}$  should not be too high; otherwise, the gain and stability both suffer.

In our design sample,  $L_{\text{degen}} = 10 \text{ nH}$  is chosen. The resistor  $R_{P,\text{out}}$  is the key part to keep the LNA block stable, while the lower value of the degeneration inductor  $L_{\text{degen}}$  provides good assistance.

The values of all the parts shown in the schematic as shown in Figure 17.12 are simultaneously optimized to obtain stability with a reasonable gain and noise figure level, that is,

$$\mu = 1.1 > 1, \quad (17.91)$$

when

$$G = 12.0 \text{ dB}, \quad (17.92)$$

and

$$\text{NF} = 1.5 \text{ dB}, \quad (17.93)$$

at  $f = 895 \text{ MHz}$ .

The  $\mu$  value shown in (17.91) indicates that the designed LNA is in an unconditionally stable state.

### 17.2.7 Nonlinearity

**17.2.7.1 Spectrum at LNA Output.** An overview of the spectrum at the LNA output can provide an intuitive feeling about the linearity of the design. Figure 17.17 shows that

$$P_{\text{out}} = -38.0 \text{ dB}_m, \quad \text{at } f = 895 \text{ MHz}, \quad (17.94)$$

$$P_{\text{out}} = -91.7 \text{ dB}_m, \quad \text{at } f = 1790 \text{ MHz}, \quad (17.95)$$

$$P_{\text{out}} = -112.6 \text{ dB}_m, \quad \text{at } f = 2685 \text{ MHz}, \quad (17.96)$$

when

$$P_{\text{in}} = -50 \text{ dB}_m. \quad (17.97)$$

Then the gain of LNA is

$$P_{\text{out}} - P_{\text{in}} = -38 \text{ dB}_m - (-50 \text{ dB}_m) = 12 \text{ dB}, \quad (17.98)$$

which is consistent with the gain testing shown in Figure 17.13(b).

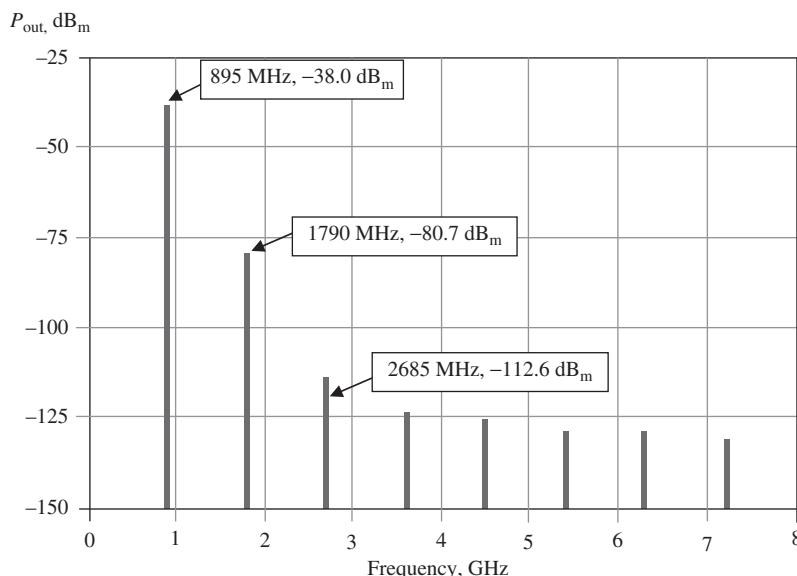


Figure 17.17. Spectrum at LNA output,  $f_o = 895 \text{ MHz}$ ,  $P_{\text{in}} = -50 \text{ dB}_m$ .

The second harmonic is lower than the output power at the operating frequency by

$$\Delta P_{\text{out},2} = P_{\text{out}}|_{1790 \text{ MHz}} - P_{\text{out}}|_{895 \text{ MHz}} = -80.7 \text{ dB}_m - (-38.0 \text{ dB}_m) = -42.7 \text{ dB}, \quad (17.99)$$

which implies that the spurious products at the half IF frequency, or the Able-Baker spurious products, would be low, and that the second-order intercept point would be high. According to design experience, if the LNA block is applied to a popular cellular phone communication system,  $\Delta P_{\text{out},2}$  should be suppressed from output power more than 30 dB at least at the operating frequency so that the distortion of the useful signal will not have conceivable distortion due to the second-order spurious products. The second-order nonlinearity is of priority in a direct conversion or zero IF communication system because it produces DC offset to disturb the desired signal.

The third harmonic is lower than the output power at the operating frequency by

$$\Delta P_{\text{out},3} = P_{\text{out}}|_{1790 \text{ MHz}} - P_{\text{out}}|_{895 \text{ MHz}} = -116.0 \text{ dB}_m - (-38.0 \text{ dB}_m) = -78.0 \text{ dB}, \quad (17.100)$$

which implies that the third-order nonlinearity is very small, so that the probability of interference from adjacent channels will be very low or negligible.

Other harmonics higher than the third-order harmonic are at least 70 dB lower than the output power at the operating frequency. There is nothing to worry about here.

**17.2.7.2 1-dB Compression Point.** The 1-dB compression point is the second easy method to overview the nonlinearity of a circuit block. Figure 17.18 shows the 1-dB compression point of our design sample, that is,

$$P_{1 \text{ dB}} = -1.25 \text{ dB}_m, \quad \text{when } f = 895 \text{ MHz}, \quad (17.101)$$

which satisfies the desired goal.

An intuitive feeling is that the nonlinearity would have occurred if the LNA is operated with an input power higher than  $-5 \text{ dB}_m$ . The 1-dB compressed point looks like the demarcation point of the input power. The LNA performs as a linear unit when

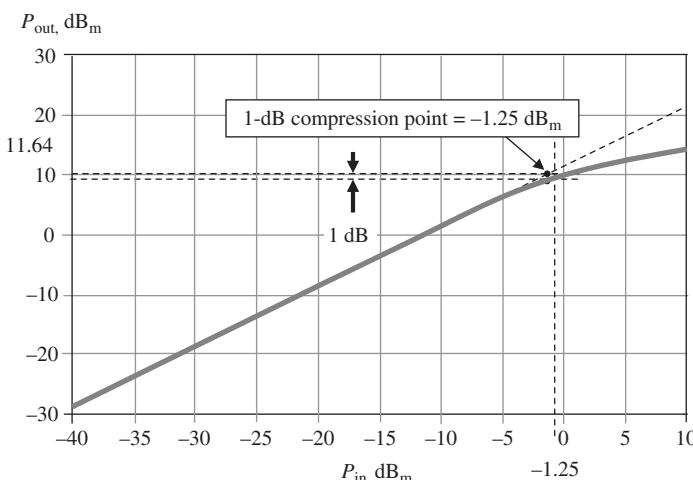


Figure 17.18. A 1-dB compression point when  $f_o = 895 \text{ MHz}$ .

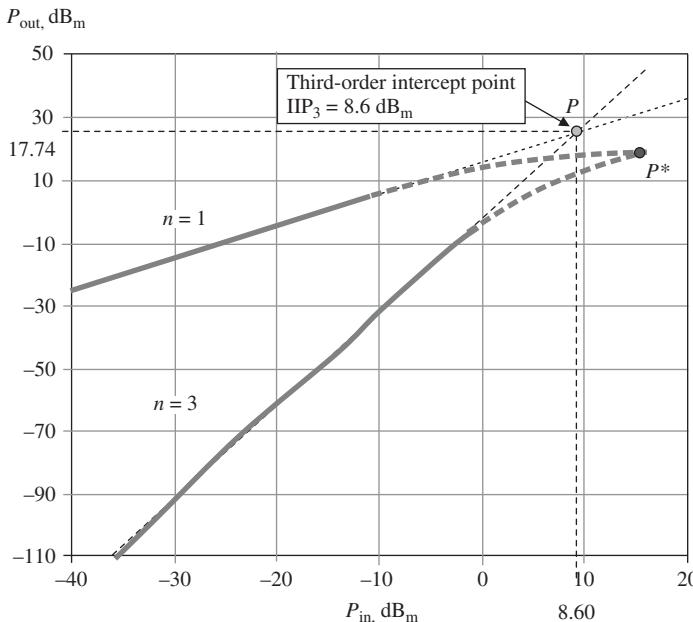


Figure 17.19. Third-order input intercept point when \$f\_o = 895\$ MHz.

its input power is below the 1-dB compression point. On the contrary, the LNA performs as a nonlinear unit when its input power is beyond the 1-dB compression point.

The third-order intercept point can be estimated to be 5–10 dB higher than the 1-dB compression point, which might be a value between 5 and 10 dB<sub>m</sub>.

**17.2.7.3 IP<sub>3</sub> and IP<sub>2</sub>.** Figure 17.19 shows the third-order intercept point, IP<sub>3</sub>, that is,

$$\text{IIP}_3 = 8.6 \text{ dB}_m, \quad \text{and} \quad \text{OIP}_3 = \text{IIP}_3 + G = 20.6 \text{ dB}_m. \quad (17.102)$$

In Figure 17.19, the tested values of input and output powers are plotted with bold lines. The asymptote with the expected slope are plotted with dash lines. Point P\* is the actual intercept point. However, point P is the expected intercept point, which is the intercept point of the lines with slopes \$n = 1\$ and \$n = 3\$. The bent portion of the dash lines indicates nonlinearity of the designed block.

Figure 17.20 shows the second-order intercept point, IP<sub>2</sub>, that is,

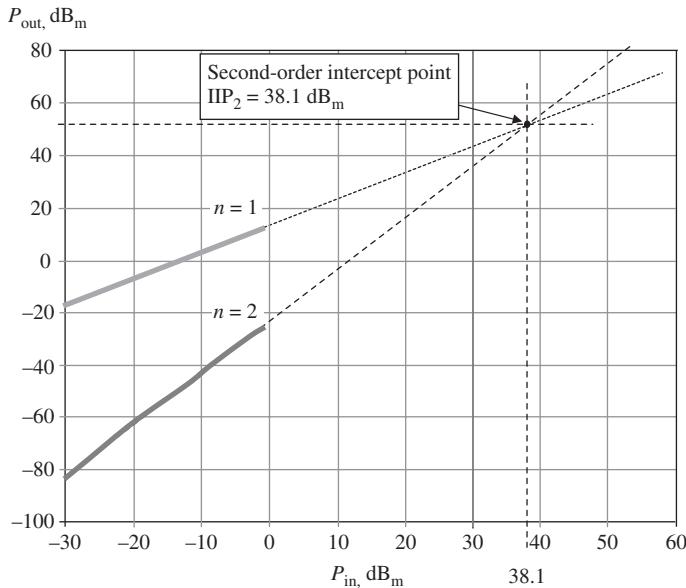
$$\text{IIP}_2 = 47.47 \text{ dB}_m, \quad \text{and} \quad \text{OIP}_2 = \text{IIP}_2 + G = 61.47 \text{ dB}_m. \quad (17.103)$$

Testing is conducted in the input power range below 0 dB<sub>m</sub>, and the intercept point is obtained by the extension of the two lines with slopes \$n = 1\$ and \$n = 2\$.

These results satisfy the aforementioned goals. Should the tested results be unsatisfactory, extra linearization work for the LNA must be conducted and the simulation might have to start from the beginning.

## 17.2.8 Design Procedures

On the basis of design experience, a flow chart of LNA design procedures is drawn in Figure 17.21. It should be noted that



**Figure 17.20.** Second-order input intercept point when  $f_o = 895$  MHz.

- Raw device testing is the key step on which success or failure largely depends.
- The first iteration loop related to the raw device is to judge the condition of  $S_{11}^* = \Gamma_{S,\text{opt}}$ . There are three options to choose from to satisfy this condition. If the design work in this step is executed well, a good LNA design should be able to obtain maximum gain and minimum noise figure simultaneously, and consequently the gain and noise figure circles should almost overlap entirely near the center of the Smith Chart.
- The second iteration loop related to the raw device is to examine the bandwidth in the process of input and output impedance matching.
- The third iteration loop related to the raw device is to judge the stability by the condition  $\mu > 1$ .
- The fourth iteration loop related to the raw device is to reach or exceed all the performance parameters.

### 17.2.9 Other Examples

The following three LNAs are actually implemented by discrete chip parts for different frequency bands. They are good examples to verify and illustrate the technology of the simultaneous approach to both  $NF_{\min}$  and  $G_{\max}$  as discussed in Section 17.2.3, even though they were designed by the author many years ago.

**17.2.9.1 LNA Design for VHF Band.** The LNA designed for VHF (very high frequency) band is outlined in the following three portions:

1. Electrical features (Table 17.2)
2. The schematic (Figure 17.22)
3. The gain and noise figure circles at the  $\Gamma_{\text{in}}$  (input voltage reflection coefficient) plan (Figure 17.23).

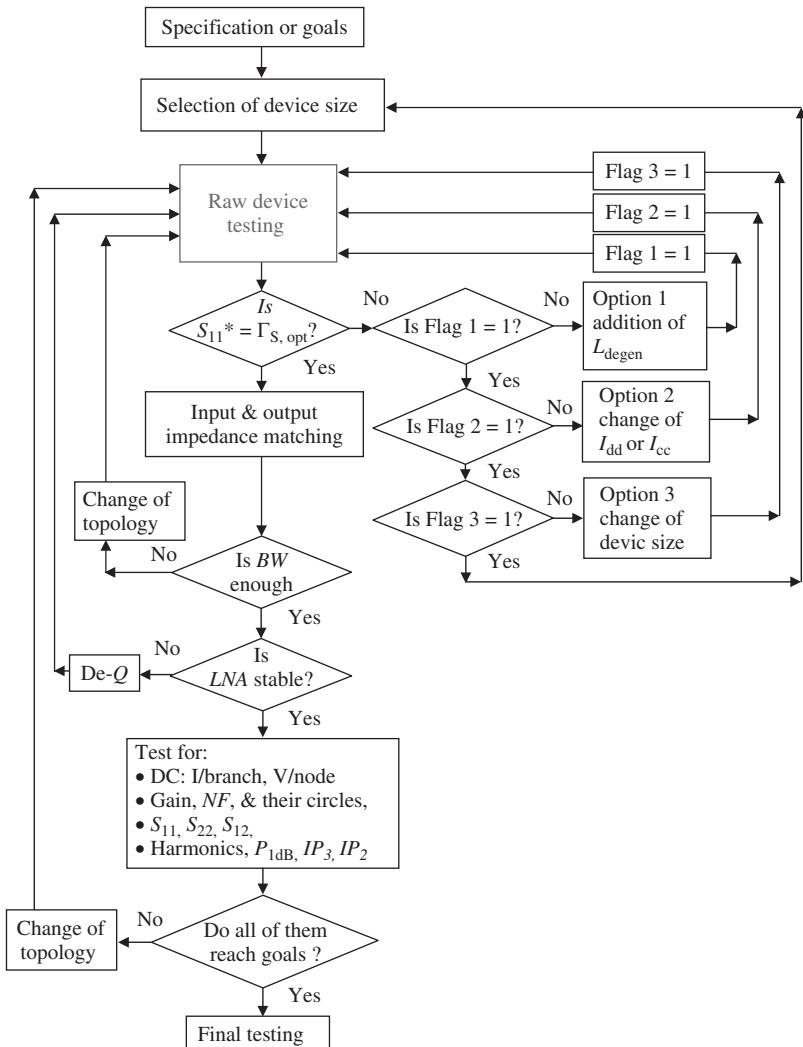


Figure 17.21. Flowchart for LNA design procedures.

## ELECTRICAL FEATURES

TABLE 17.2. Electrical Features of LNA for VHF Radio

	Specification	Final Tested Result
Frequency range	130–180 MHz	—
Device	BFQ67 (Manufacturer: Siemens, Munich, Germany)	—
DC power supply	3 V	—
Current drain	<4 mA	3.47 mA
Gain	>10 dB	15.0 dB
Noise figure	<2 dB	1.75 dB
IIP <sub>3</sub>	>0 dB <sub>m</sub>	2.5 dB <sub>m</sub>
Input return	<-10 dB	-40.0 dB
Output return	<-10 dB	-12.2 dB

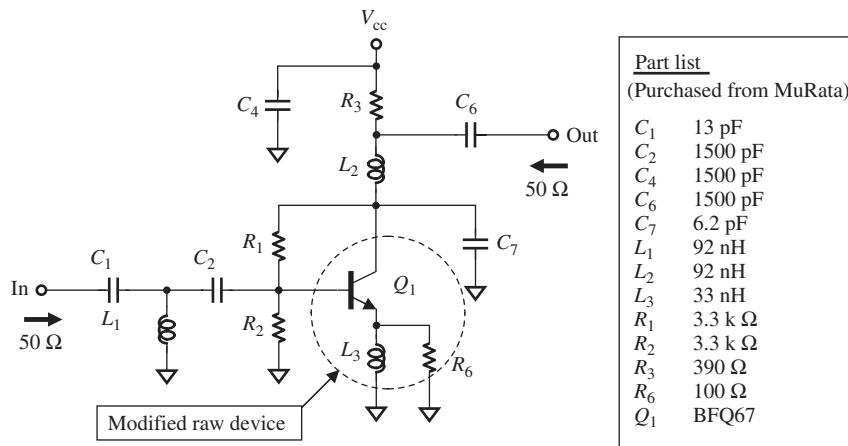


Figure 17.22. LNA designed for VHF radio.

SCHEMATIC. The following should be noted:

- The inductor  $L_3 = 33$  nH is the degeneration inductor mentioned above. The resistor  $R_6 = 100 \Omega$  is the “de- $Q$ ” resistor. The modified raw device consists of  $Q_1, L_3$ , and  $R_6$ .
- Capacitors,  $C_4, C_2$ , and  $C_6$  are “zero” capacitors. Their specified values should be 1300 pF, and actual values are 1500 pF.
- The resistors  $R_3$  and  $R_6$  are 390  $\Omega$  and 100  $\Omega$ , respectively; they are both “de- $Q$ ” resistors.
- The resistors  $R_1$  and  $R_2$  are the combination of a voltage divider. They take the DC bias for the transistor from  $V_{cc}$ ; they function as a feedback branch as well.
- The output impedance network consisting of  $R_3, L_2$ , and  $C_6$  confers two advantages, lower part count and wider bandwidth.
- All parts in Figure 17.22 are chip parts and are purchased from MuRata.

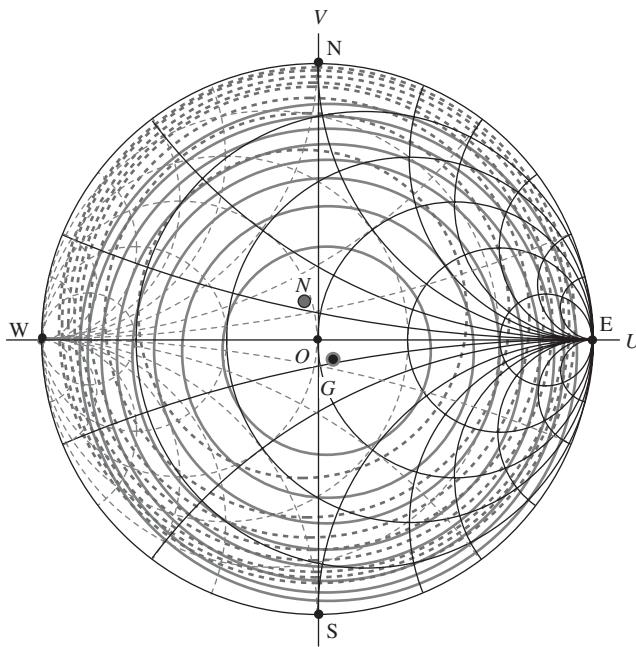
GAIN CIRCLES AND NOISE CIRCLES. Figure 17.23 shows that for this circuit the noise figure circles overlap with the gain circles well enough, although not perfectly. This verifies the assertion of simultaneously approaching both  $NF_{min}$  and  $G_{max}$ , because Figure 17.23 is obtained because of the satisfaction of condition (17.34).

**17.2.9.2 LNA Design for UHF Band.** The LNA designed for UHF (ultrahigh frequency) band can be outlined in the following three items:

1. Electrical features (Table 17.3)
2. The schematic (Figure 17.24)
3. The gain and noise figure circles at the  $\Gamma_{in}$  (input voltage reflection coefficient) plan (Figure 17.25).

ELECTRICAL FEATURES. The following should be noted:

- The inductor  $L_3 = 5.6$  nH is the degeneration inductor mentioned above. The resistor  $R_6 = 390 \Omega$  is the “de- $Q$ ” resistor. The modified raw device consists of  $Q_1, L_3$ , and  $R_6$ .



- Gain circles:  $G_{\max} = 15$  dB at point G, step = 1.0 dB,
- Noise figure circles:  $NF_{\min} = 1.75$  dB at point N, step = 0.25 dB,

Figure 17.23. Constant gain circles and constant noise figure circles when  $f = 150$  MHz.

TABLE 17.3. Electrical Features of LNA for UHF Radio

	Specification	Final Tested Result
Frequency range	400 to 470 MHz	—
Device	BFQ67 (Manufacturer: Siemens)	—
DC power supply	3 V	—
Current drain	<4 mA	3.43 mA
Gain	>10 dB	12.0 dB
Noise figure	<2 dB	1.5 dB
IIP <sub>3</sub>	>0 dB <sub>m</sub>	5.0 dB <sub>m</sub>
Input return	<-10 dB	-17.8 dB
Output return	<-10 dB	-16.0 dB

- The capacitors  $C_4$ ,  $C_2$ , and  $C_6$  are “zero” capacitors. Their value is 150 pF.
- The resistors  $R_3 = R_6 = 390 \Omega$  are “de-Q” resistors.
- The resistors,  $R_1$  and  $R_2$ , are the combination of a voltage divider. They take the DC bias for the transistor from  $V_{cc}$ ; they function as a feedback branch as well.
- The output impedance network consisting of  $R_3$ ,  $L_2$ ,  $C_6$ , and  $L_4$  confers two advantages, lower part count and wider bandwidth.
- All parts in Figure 17.24 are chip parts and are purchased from MuRata.

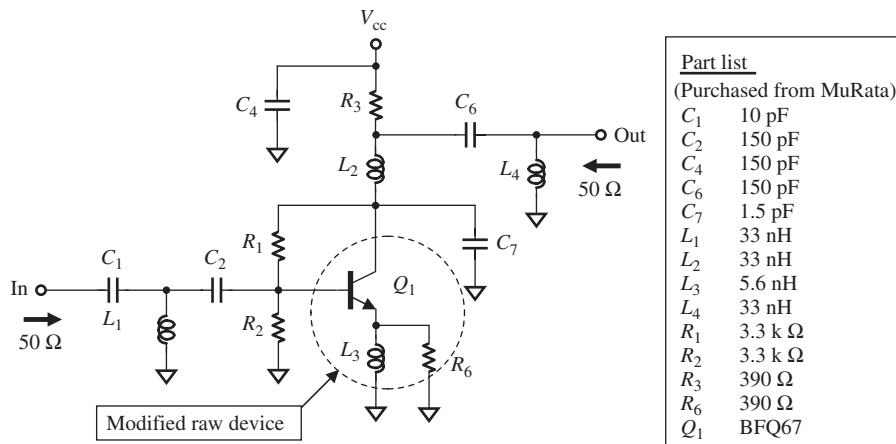
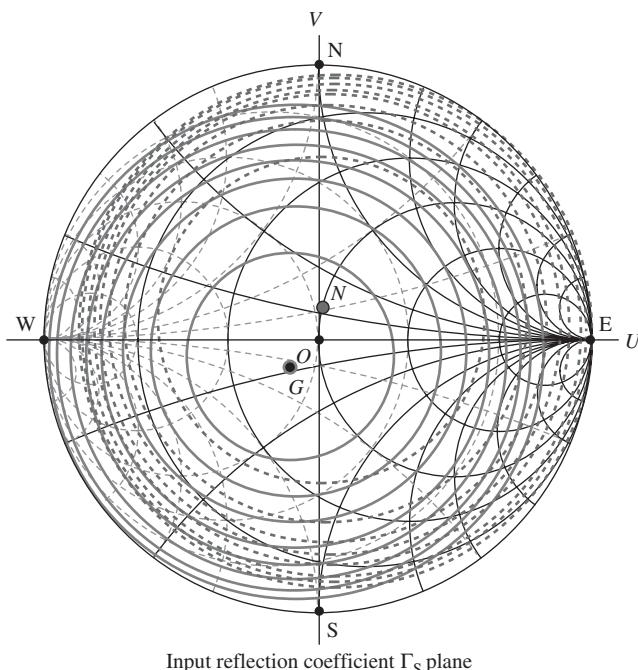


Figure 17.24. LNA designed for UHF radio.

Figure 17.25. Constant gain circles and constant noise figure circles when  $f = 450$  MHz.

**GAIN CIRCLES AND NOISE CIRCLES.** Figure 17.25 shows that for this circuit the noise figure circles overlap with the gain circles well enough, although not perfectly. This verifies the assertion of simultaneously approaching both  $NF_{\min}$  and  $G_{\max}$ , because Figure 17.25 is obtained because of the satisfaction of condition (17.34).

**17.2.9.3 LNA Design for 800/900 MHz Radio.** The LNA designed for 800/900 MHz radio can be outlined in the following three portions:

1. Electrical features (Table 17.4)
2. The schematic (Figure 17.26)
3. The gain and noise figure circles at the  $\Gamma_{in}$  (input voltage reflection coefficient) plan (Figure 17.27).

#### ELECTRICAL FEATURES

TABLE 17.4. Electrical Features of LNA for 800/900 MHz Radio

	Specification	Final Tested Result
Frequency range	850–940 MHz	—
Device	BFQ67 (Manufactured by Siemens)	—
DC power supply	3 V	—
Current drain	<4 mA	3.48 mA
Gain	>10 dB	11.0 dB
Noise figure	<2 dB	1.8 dB
IIP <sub>3</sub>	>0 dB <sub>m</sub>	8.6 dB <sub>m</sub>
Input return	<−10 dB	−11.7 dB
Output return	<−10 dB	−17.0 dB

**SCHEMATIC.** The following should be noted:

- The inductor  $L_3 = 4.7 \text{ nH}$  is the degeneration inductor mentioned above. The resistor  $R_6 = 390 \Omega$  is the “de- $Q$ ” resistor. The modified raw device consists of  $Q_1, L_3$ , and  $R_6$ .
- The capacitors,  $C_4, C_2$ , and  $C_6$  are “zero” capacitors. Their value is 39 pF.
- The resistors  $R_3 = R_6 = 390 \Omega$  are “de- $Q$ ” resistors.
- The resistors,  $R_1$  and  $R_2$ , are the combination of a voltage divider. They take the DC bias for the transistor from  $V_{cc}$ ; they function as a feedback branch as well.

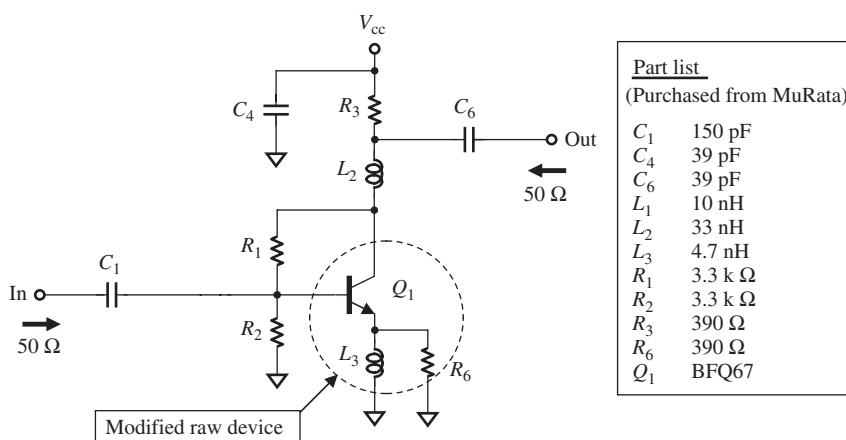
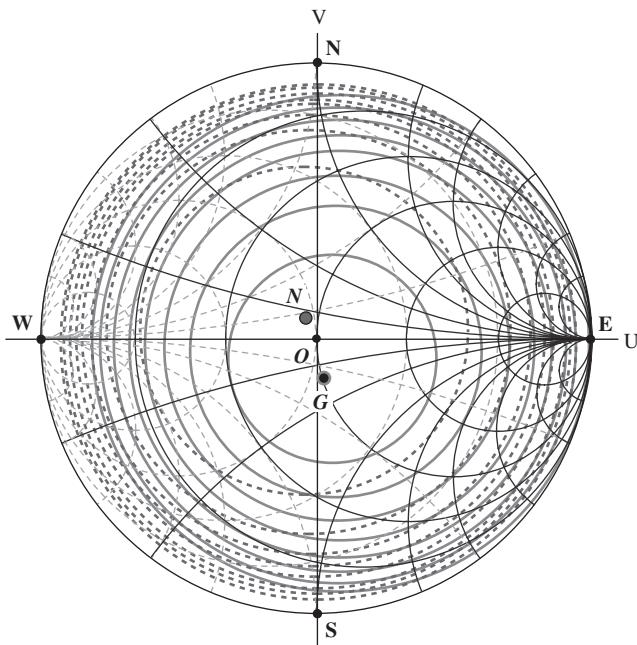


Figure 17.26. LNA designed for 800/900 MHz radio.

Input reflection coefficient  $\Gamma_S$  plane

Gain circles:  $G_{\max} = 11$  dB at point G, step = 1.0 dB,

Noise figure circles:  $NF_{\min} = 1.8$  dB at point N, step = 0.25 dB.

**Figure 17.27.** Constant gain circles and constant noise figure circles when  $f = 850$  MHz.

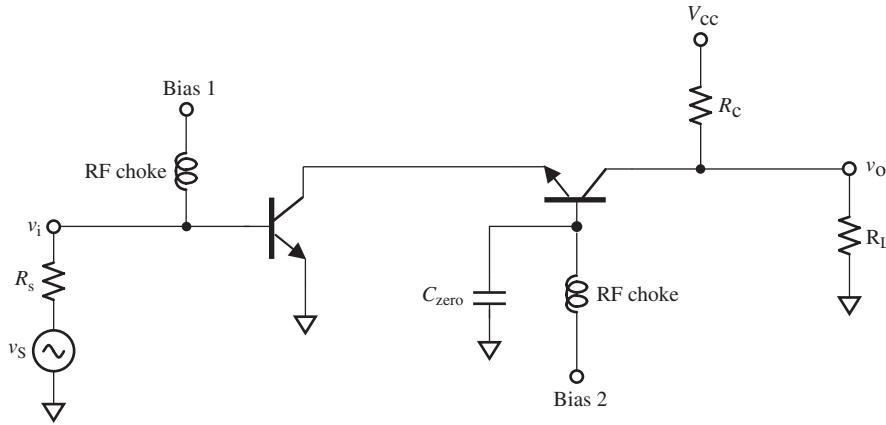
- The output impedance network consisting of  $R_3$ ,  $L_2$ , and  $L_6$  confers two advantages, lower part count and wider bandwidth.
- All parts in Figure 17.26 are chip parts and are purchased from MuRata.

**GAIN CIRCLES AND NOISE CIRCLES.** Figure 17.27 shows that for this circuit the noise figure circles are good enough overlapped with the gain circles, although not perfect. It verifies that the correctness of the assertion of simultaneous approach to both  $NF_{\min}$  and  $G_{\max}$ , because Figure 17.27 is obtained from the satisfaction of the condition (17.34).

## 17.3 SINGLE-ENDED CASCODE LNA

### 17.3.1 Bipolar CE–CB Cascode Voltage Amplifier

Figure 17.28 shows a bipolar CE–CB cascode amplifier. The first stage is a voltage amplifier with a CE (common emitter) configuration because its emitter is the common AC and DC grounded terminal of the input and output. The second CB stage is a voltage amplifier with a CB (common base) configuration because its base is the common AC grounded terminal of the input and output. The collector of the first CE stage is connected to the emitter of the second CB stage. The input terminal is the base of the first CE stage, and the output terminal is the collector of the second CB stage. RF chokes and a “zero” capacitor,  $C_{\text{zero}}$ , are connected to the devices for DC bias. In the operating frequency range, the impedance of the RF choke is assumed to approach infinity and the impedance of the capacitor  $C_{\text{zero}}$  is assumed to approach zero.



**Figure 17.28.** A bipolar cascode amplifier.

The equivalent of the bipolar cascode amplifier is shown in Figure 17.29(a). In order to simplify the analysis, let us consider the cases of low frequencies only so as to neglect all the capacitors in the transistors. Also, the resistors  $r_b$ ,  $r_c$ , and  $r_\mu$  are neglected. In this section, the subscript “1” denotes the parameter in the first stage, the subscript “2” denotes the parameter in the second stage; the subscript “i” denotes the input parameter, the subscript “o” denotes the output parameter.

The input resistance of the cascode amplifier is the input resistance of the first CE stage. Obviously, from Figure 17.29(b), it can be seen that

$$R_i = R_{i1} = r_{\pi 1}. \quad (17.104)$$

The output resistance of the first CE stage is contributed by the first CE portion only when  $v_S = v_1 = 0$  so that the generator  $g_{m1}v_1$  is inactive. Then,

$$R_{o1} = r_{o1}. \quad (17.105)$$

The input resistance of the second CB amplifier is

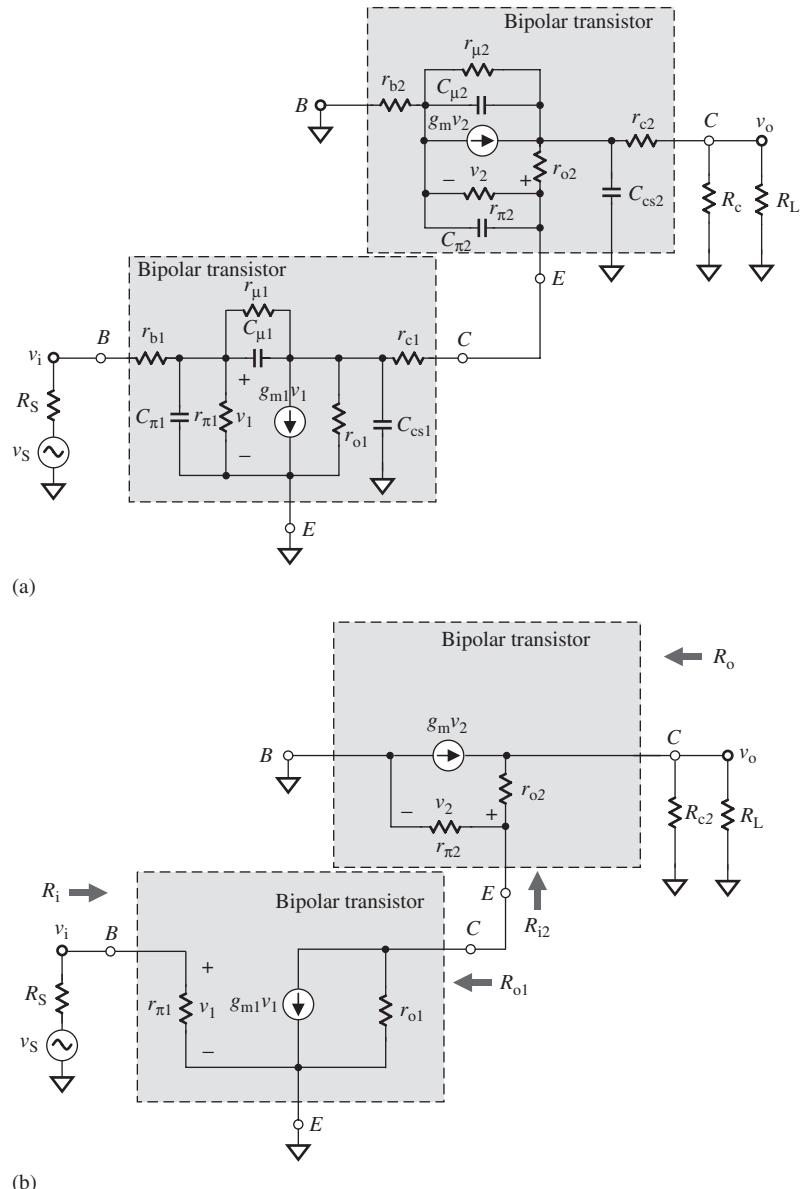
$$R_{i2} = r_{e2} = \frac{1}{g_{m2} + \frac{1}{r_{\pi 2}}} = \frac{1}{1 + \beta_2} r_{\pi 2} \approx \frac{r_{\pi 2}}{\beta_2}, \quad (17.106)$$

which has been discussed in Section 5.5. It should be noted that the input resistance of the second CB stage is dropped by about  $\beta$  times from  $r_\pi$  in CE stage alone.

Figure 17.29(b) shows the equivalent circuit of the bipolar cascode amplifier with a CE–CB configuration at low frequencies.

The output resistance is contributed by the upper CB portion only when  $v_S = v_1 = 0$ , so that the generator  $g_{m1}v_1$  is inactive. The bottom portion looks just like a single resistor,  $r_{o1}$ . The entire CE–CB amplifier is equivalent to a CB stage with a degeneration resistor  $r_{o1}$  and an AC grounded input terminal. The output resistance has been formulized as

$$R_o = r_{o2} \left( 1 + \frac{g_{m2}r_{o1}}{1 + \frac{g_{m2}r_{o1}}{\beta_2}} \right). \quad (17.107)$$



**Figure 17.29.** Equivalent circuit of bipolar cascode amplifier with CE-CB configuration. (The resistors,  $r_b$ ,  $r_\mu$ , and  $r_c$ , are neglected.) (a) In general case; (b) in low-frequency case.

(Refer to Gray et al., 2004.)

If

$$g_{m2}r_{o1} \gg \beta_2 \gg 1, \quad (17.108)$$

then

$$R_o \approx \beta_2 r_{o2}. \quad (17.109)$$

It should be noted that in the derivation of (17.107), the resistor  $R_c$  in the second CB stage and the load  $R_L$  are ignored. They, of course, can never be neglected in the RF circuit design.

From expression (17.109) it can be found that the CE–CB cascade amplifier displays an output resistance about  $\beta$  times higher than that in the CE stage alone.

Now let us consider the voltage and current gain.

The voltage gain of the first CE stage is not that as expected from a normal CE stage,  $A_{v1} = -g_{m1}r_{o1}$ . The output resistor  $r_{o1}$  of the first CE stage is connected with the input resistor  $R_{i2}$  of the second CB stage in parallel. As shown in expression (17.106), the input resistance of the second CB stage is dropped by about  $\beta$  times from the  $r_\pi$  in a normal CE stage; consequently,

$$A_{v1} = -g_{m1}r_{o1}/R_{i2} \approx -g_{m1}\frac{r_{\pi2}}{\beta_2}, \quad (17.110)$$

if the input voltage  $v_i$  but not  $v_S$  is considered as the reference for the voltage gain.

Should the main parameters of the CE and CB transistor be equal, that is,

$$g_{m1} = g_{m2}, \quad (17.111)$$

$$r_{\pi1} = r_{\pi2}, \quad (17.112)$$

and

$$\beta_1 = \beta_2, \quad (17.113)$$

then expression (17.110) becomes

$$A_{v1} = -1. \quad (17.114)$$

The current gain of the first CE stage is

$$A_{i1} = \beta. \quad (17.115)$$

The current gain of the second stage is

$$A_{i2} \approx 1, \quad (17.116)$$

because the input current is almost equal to its output current in the second CB stage

$$I_{i2} \approx I_{o2} \quad \text{and} \quad i_{i2} \approx i_{o2}, \quad (17.117)$$

where  $I$  denotes DC and  $i$  denotes AC.

From (17.115) and (17.116), the total current gain is

$$A_i \approx A_{i1}A_{i2} \approx \beta. \quad (17.118)$$

The transconductance from the input to output is

$$G_m = g_{m1}, \quad (17.119)$$

since  $A_{i2} \approx 1$  as shown in (17.116).

The total voltage gain is

$$A_v = \frac{v_o}{v_i} \approx -G_m R_o \approx -g_{m1} r_{o2} \beta_2. \quad (17.120)$$

### 17.3.2 MOSFET CS-CG Cascode Voltage Amplifier

Figure 17.30 shows a MOSFET CS-CG cascode amplifier. The first stage is a voltage amplifier with a CS (common source) configuration because its source is the common AC and DC grounded terminal of the input and output. The second stage is a voltage amplifier with a CG (common gate) configuration because its gate is the common AC grounded terminal of the input and output. The drain of the first CS stage is connected with the source of the second CG stage. The input terminal is the gate of the first CS stage, and the output terminal is the drain of the second CG stage. RF chokes and a “zero” capacitor,  $C_{\text{zero}}$ , are connected to the devices for DC bias. In the operating frequency range, the impedance of the RF choke is assumed to approach infinity and the impedance of the capacitor  $C_{\text{zero}}$  is assumed to approach zero.

The equivalent of the MOSFET cascode amplifier is shown in Figure 17.31. In order to simplify the analysis, let us consider the cases of low frequencies only so as to neglect all the capacitors in the transistors. Also, the resistors  $r_b$ ,  $r_c$ , and  $r_\mu$  are neglected. Figure 17.32 shows the equivalent circuit of the MOSFET cascode amplifier with a CS-CG configuration at low frequencies.

The input resistance of the CS-CG cascode amplifier is the input resistance of the first CS stage. Obviously, from Figure 17.32, it can be seen that

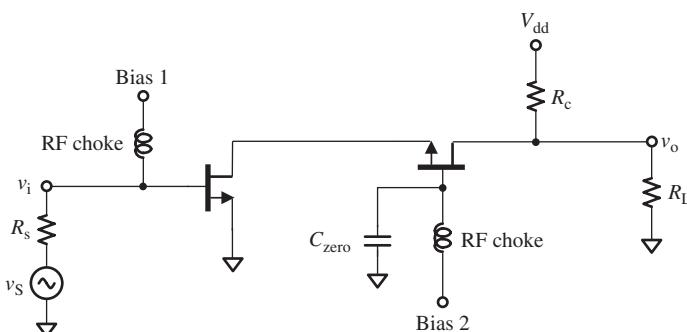
$$R_i \rightarrow \infty. \quad (17.121)$$

The output resistance of the first CS stage is contributed by the first CS portion only when  $v_S = v_1 = 0$  so that the generator  $g_{m1}v_1$  is inactive; then,

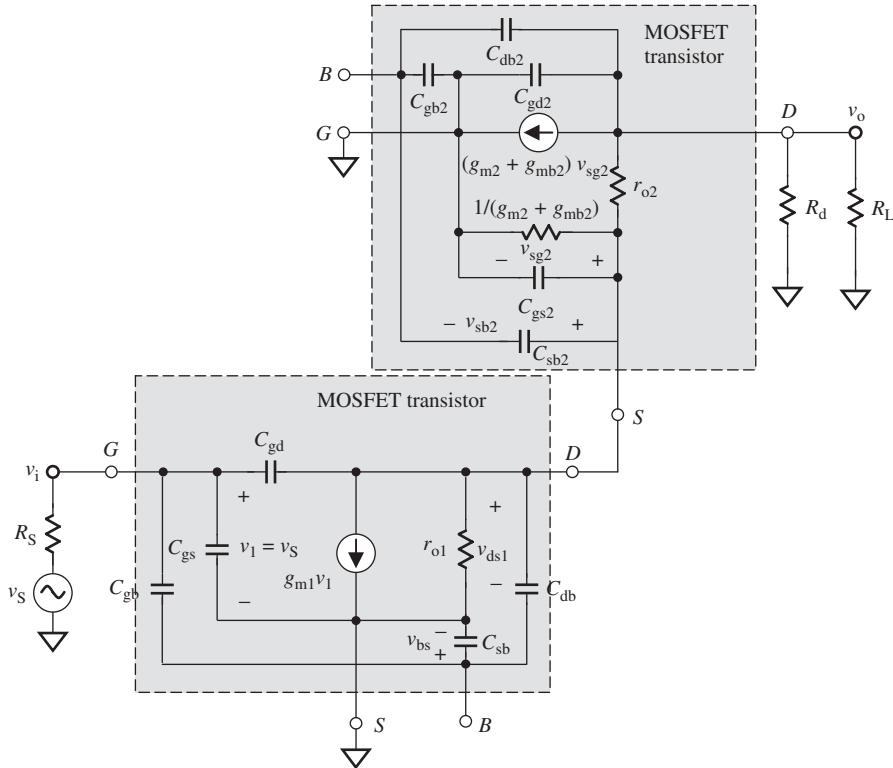
$$R_{o1} = r_{o1}. \quad (17.122)$$

In terms of KCL (Kirchoff's current law) at the output and at the source of the second CG stage, the input resistance of the second CG amplifier is

$$R_{i2} = \frac{1}{(g_{m2} + g_{mb2})} + \frac{R'_L}{(g_{m2} + g_{mb2})r_{o2}}, \quad (17.123)$$



**Figure 17.30.** A MOSFET cascode amplifier.



**Figure 17.31.** Equivalent circuit of MOSFET cascode amplifier with CS–CG configuration.

where

$$R'_L = R_d // R_L = \frac{R_d}{R_d + R_L} R_L, \quad (17.124)$$

and the transconductance is derived when  $R'_L = 0$  so that  $v_o = 0$ , that is,

$$G_m \approx g_{m1}. \quad (17.125)$$

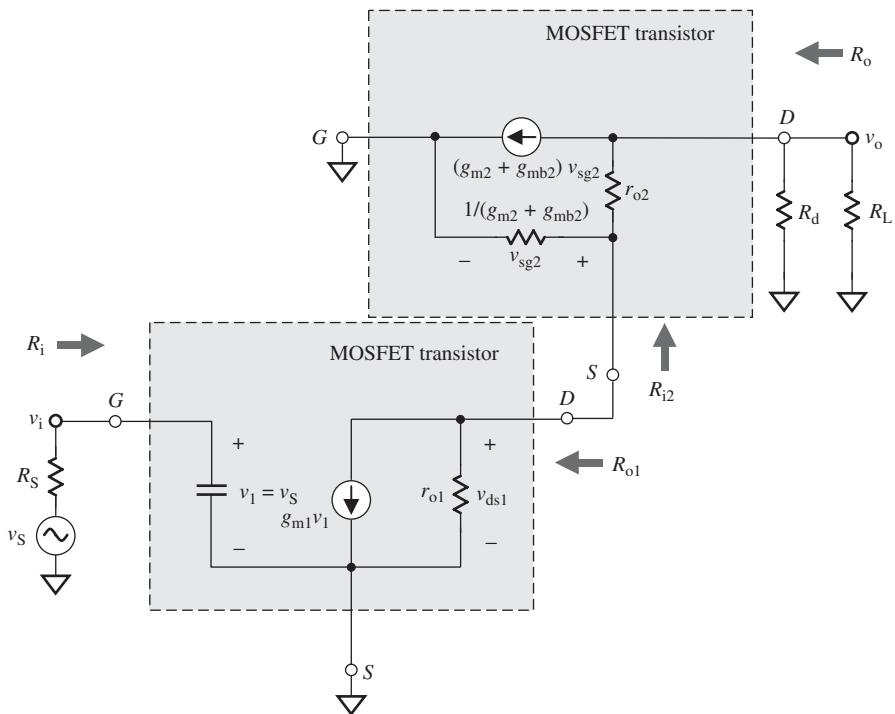
The output resistance can be found when the  $g_{m1}$  generator is inactive, that is  $v_i = 0$ . Consequently, the CS portion becomes a simple resistor  $r_{o1}$ , and the entire cascode looks like a CG device with a source degeneration resistor. Then,

$$R_o = r_{o2} + r_{o1}[1 + (g_{m2} + g_{mb2})]r_{o2} \approx (g_{m2} + g_{mb2})r_{o1}r_{o2}, \quad (17.126)$$

if

$$R'_L \rightarrow \infty. \quad (17.127)$$

(Refer to Gray et al., 2004b.)



**Figure 17.32.** Equivalent circuit of MOSFET cascode amplifier with CS–CG configuration at low frequencies.

Similar to the CE–CB bipolar cascode amplifier, the input resistance of the second CG stage is quite low because the second term in expression (17.123) disappears when

$$R'_L = 0. \quad (17.128)$$

It becomes

$$R_{i2} \approx \frac{1}{(g_{m2} + g_{mb2})}, \quad (17.129)$$

which usually happens in cases where

$$R_{i2} \ll r_{o1}. \quad (17.130)$$

Now let us consider the voltage and current gain.

The voltage gain of the first CS stage is not that as expected from a normal CS stage,  $A_{v1} = -g_{m1}r_{o1}/R_d$ . The output resistor  $r_{o1}$  of the first CS stage is connected with the input resistor  $R_{i2}$  of the second CG stage in parallel. As shown in expression (12.129), the input resistance of the second CG stage is dropped down a lot and is much lower than that of  $r_{o1}$  in a normal CS stage; consequently,

$$A_{v1} = -g_{m1}r_{o1}/R_{i2} \approx -g_{m1}R_{i2} \approx -\frac{g_{m1}}{g_{m2} + g_{mb2}}, \quad (17.131)$$

if the input voltage  $v_i$  but not  $v_S$ , is considered as the reference for the voltage gain.

If the transconductances of the CS and CG transistors are the same, that is,

$$g_{m1} = g_{m2}, \quad (17.132)$$

then expression (17.131) becomes

$$A_{v1} \approx -1. \quad (17.133)$$

The current gain of the first CS stage is

$$A_{i1} \rightarrow \infty. \quad (17.134)$$

The current gain of the second CG stage is

$$A_{i2} \approx 1, \quad (17.135)$$

because the input current is almost equal to its output current in the second CG stage, that is,

$$I_{i2} \approx I_{o2} \quad \text{and} \quad i_{i2} \approx i_{o2}, \quad (17.136)$$

where  $I$  denotes DC and  $i$  denotes AC.

Finally, the total voltage gain can be evaluated from the expressions (17.125) and (17.126)

$$A_v = \frac{v_o}{v_i} \approx -G_m R_o \approx -g_{m1}(g_{m2} + g_{mb2})r_{o1}r_{o2}. \quad (17.137)$$

### 17.3.3 Why Cascode?

Let us summarize the comments why the cascode amplifier is superior to other amplifiers.

It increases the output impedance, which is particularly useful in desensitizing bias references from variations in power supply voltage and in achieving large amounts of voltage gain.

This is very beneficial to the digital circuit design. Digital circuits always operate under high impedances; the high output impedance of a cascode amplifier provides a powerful way to reach the goal of high voltage gain.

However, this remarkable advantage is not exciting to the RF circuit designers because RF circuits always operate under low impedances. RF designers pursue power gain but not voltage gain; here, high voltage gain does not make too much sense if its current gain is low.

However, both RF circuit designers and digital circuit designers are excited about these three advantages of the cascode amplifier.

- 1. It Alleviates the Miller Effect in a Voltage Amplifier.* Unwanted capacitive feedback always exists in a voltage amplifier with a CE or CS configuration. Alleviating the Miller effect therefore allows the amplifier operation at higher frequencies than would otherwise be possible.

The input impedance of the second CG stage is

$$R_{i2} = \frac{r_{\pi2}}{\beta_2}, \quad (17.138)$$

which is  $\beta_2$  times lower than the input impedance of a CE stage if the same device is applied to both stages. The voltage gain of the first stage is low because of its output impedance being pulled down by the low input impedance of the second stage, that is,

$$A_{v1} = -g_{m1}R'_{L1}/R_{i2} = -g_{m1}R'_{L1}/\frac{r_{\pi2}}{\beta_2} \approx -g_{m1}\frac{r_{\pi2}}{\beta_2} \approx 1, \quad (17.139)$$

if

$$g_{m1} = g_{m2}. \quad (17.140)$$

Consequently, the input Miller capacitance in the first stage is kept the same as  $C_\mu$  due to the low voltage gain.

$$C_{i,miller} = C_\mu(1 + g_m R_L') \approx C_\mu. \quad (17.141)$$

Owing to the low Miller capacitance, the bandwidth in the first stage will be increased from

$$\omega_{T1} = \frac{g_{m1}}{C_{\pi1} + C_{\mu1}(1 + g_{m1}R_{L1}')} \quad (17.142)$$

to

$$\omega_{T1} = \frac{g_{m1}}{C_{\pi1} + C_{\mu1}}. \quad (17.143)$$

The bandwidth of the entire CS–CB cascode amplifier is then mainly determined by the second stage, that is,

$$\omega_{T2} = \frac{g_{m2}}{C_{\mu2}}. \quad (17.144)$$

It is therefore concluded that the bandwidth of the CS–CB cascode amplifier is approximately equivalent to the bandwidth of the single CB amplifier, which is much higher than that of a single CE amplifier.

This wide bandwidth is an outstanding advantage, which is why the cascode amplifier is widely applied in the amplifier design.

2. *It Provides Better Isolation.* The CB or CG configuration of the second stage guarantees isolation between output and input. The base of the bipolar transistor or the gate of the MOSFET is grounded, and the capacitive feedback from the output to input is reduced to an insignificant level. In addition, the low input impedance of the second stage is also beneficial to the isolation between output and input.

3. *It can Magnify not only Voltage but also the Power of Signall.* In a cascade amplifier, the current is magnified in the first CE stage and the voltage is magnified in the second CB stage. Consequently, the voltage and power of the signal can be magnified simultaneously, therefore satisfying both digital and RF circuit designers.

### 17.3.4 Example

This is an LNA RFIC design example, which is designed for group #1, band #2 of UWB (ultrawide band) system. The main electrical features are the following:

- DC power supply:  $V_{dd} = 1.2$  V;
- Current drain:  $I_{total} < 5$  mA;
- Operating frequency range:  $f = 3.696\text{--}4.4224$  GHz;
- Gain:  $G > 12$  dB;
- Input return loss:  $S_{11} < -10$  dB;
- Output return loss:  $S_{22} < -10$  dB;
- Noise figure:  $NF < 2.5$  dB;
- Third-order input intercept point:  $IIP_3 > 5$  dB<sub>m</sub>;
- Second-order input intercept point:  $IIP_2 > 35$  dB<sub>m</sub>.

In consideration of the bandwidth, it is decided to use the cascode configuration, because its relative bandwidth is

$$BW = \frac{\Delta f}{f} = \frac{4224 - 3696}{(4224 + 3696)/2} = \frac{528}{3960} = 13.3\%. \quad (17.145)$$

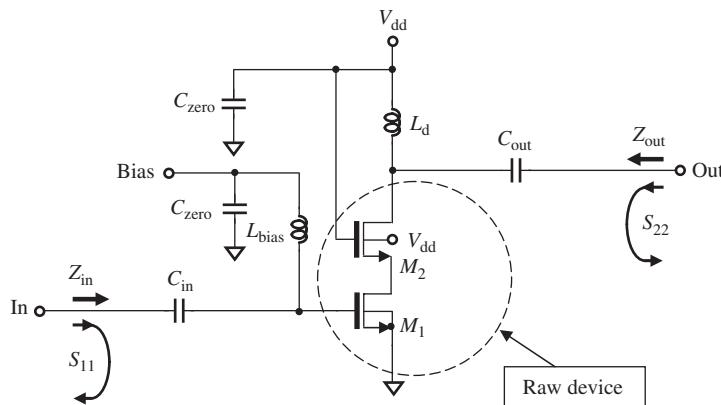
This can be categorized as neither a narrow band block nor a wide band block. Usually, a block or a system with a relative bandwidth greater than 15% is considered a wide band block or system; conversely, a block or a system with a relative bandwidth of less than 15% is considered a narrow band block or system. We are hence quite hesitant to consider this design sample as a narrow band block because 13.3% is close to 15%.

TSMC's 90-nm CMOS is selected to be the IC processing device, and *n*-channel MOSFET are chosen to be the CS–CG device. The size of the device is calculated based on the considerations described in Section 17.2.1.

**17.3.4.1 Raw Device Testing.** Figure 17.33 shows the setup for raw device testing. The raw device is a combination of two transistors,  $M_1$  and  $M_2$ . All the capacitors are “zero” capacitors, and all the inductors are “infinite” inductors. At the operating frequencies, the impedance of a “zero” capacitor approaches zero and the impedance of an “infinite” inductor approaches infinity.

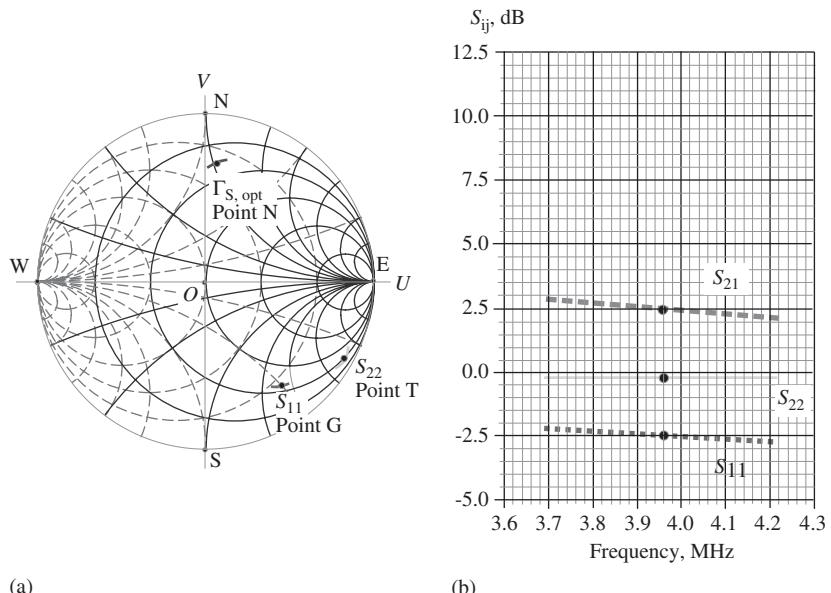
Figure 17.34(a) shows the tested *S* parameters and the optimal input reflection coefficient  $\Gamma_{S,\text{opt}}$  on the Smith Chart; Figure 17.34(b) shows the frequency response of *S* parameters. The intermediate frequency,  $f = 3.96$  GHz, is marked with a dot on each trace.

The first impression is that the CS–CG configuration of the raw devices exhibits wideband behavior as expected. In Figure 17.34(a), all of the traces corresponding to the entire frequency band, 3.696–4.224 GHz, are squeezed to a small trace segment on the Smith Chart, and in Figure 17.34(b) the trace is flat over the entire frequency band.



- $C_{in}, C_{out}, C_{zero}$  : “zero” capacitor.
- $L_{bias}, L_d$  : “Infinitive” inductor.

**Figure 17.33.** Schematic of cascode LNA for raw device testing.



**Figure 17.34.**  $S$  parameters from raw device testing. (The intermediate frequency  $f = 3.96$  GHz is marked by a dot on each trace.) (a)  $S_{11}$ ,  $S_{22}$ , and  $\Gamma_{S,\text{opt}}$  on Smith Chart; (b) magnitude of  $S_{ij}$ , dB.

The second impression is that, unfortunately, the optimal input voltage reflection coefficient  $\Gamma_{S,\text{opt}}$  is not at the expected location, which is supposed to be conjugate to  $S_{11}$ . The raw devices must be modified so that the overlapping condition of maximum gain and minimum noise figure, (17.34), can be satisfied.

The magnitudes of the  $S$  parameters shown in Figure 17.34(b) can be outlined as follows:

$$2.45 \text{ dB} < S_{21} < 2.58 \text{ dB}, \quad \text{when } 3.696 \text{ GHz} < f < 4.224 \text{ GHz}, \quad (17.146)$$

$$S_{21} = 2.50 \text{ dB}, \quad \text{when } f = 3.960 \text{ GHz}, \quad (17.147)$$

$$-2.75 \text{ dB} < S_{11} < -2.20 \text{ dB}, \quad \text{when } 3.696 \text{ GHz} < f < 4.224 \text{ GHz}, \quad (17.148)$$

$$S_{11} = -2.5 \text{ dB}, \quad \text{when } f = 3.960 \text{ GHz}, \quad (17.149)$$

$$-0.19 \text{ dB} < S_{22} < -0.17 \text{ dB}, \quad \text{when } 3.696 \text{ GHz} < f < 4.224 \text{ GHz}, \quad (17.150)$$

$$S_{22} = -0.18 \text{ dB}, \quad \text{when } f = 3.960 \text{ GHz}, \quad (17.151)$$

For now, we will not worry about the low gain, poor return losses, stability, or other issues, because the input and output impedances are not matched yet. Let us go ahead to modify the raw devices. This modification is usually conducted many times, depending on how much experience the designer has. Figure 17.35 shows the final attempt, in which a degeneration inductor is connected between the source of the CS device and ground:

$$L_{\text{degen}} = 0.45 \text{ nH}. \quad (17.152)$$

The degeneration inductor,  $L_{\text{degen}}$ , is a simple and tiny part. In the RFIC circuit design for the operating frequency range of gigahertz, it is about 0.5–3 turns, with approximately a tenth to a couple nH of inductance.

Figure 17.36(a) shows the tested  $S$  parameters and the optimal input reflection coefficient  $\Gamma_{S,\text{opt}}$  on the Smith Chart from the modified raw device testing. The optimal input voltage reflection coefficient  $\Gamma_{S,\text{opt}}$  is at the expected location, which is very close to  $S_{11}^*$ . This means that the raw devices are quite close to fitting the overlapping condition of maximum gain and minimum of noise figure, (17.34), although  $\Gamma_{S,\text{opt}}$  does not exactly overlap  $S_{11}^*$ . A little deviation is inevitable, mirroring a philosophy of life: “Nobody is perfect and nothing is perfect!”

Again, at this point, we are not worrying too much about the magnitude of the  $S$  parameters shown in Figure 17.36(b), which are far from the stated goals, because the input and output impedances are still not matched.

**17.3.4.2 Gain and Bandwidth.** Impedance matching is one of the key issues in RF circuit design. A new designer without RF circuit design experience might need to spend a lot of time on it. Let us spend some time to demonstrate the process of impedance matching in this example.

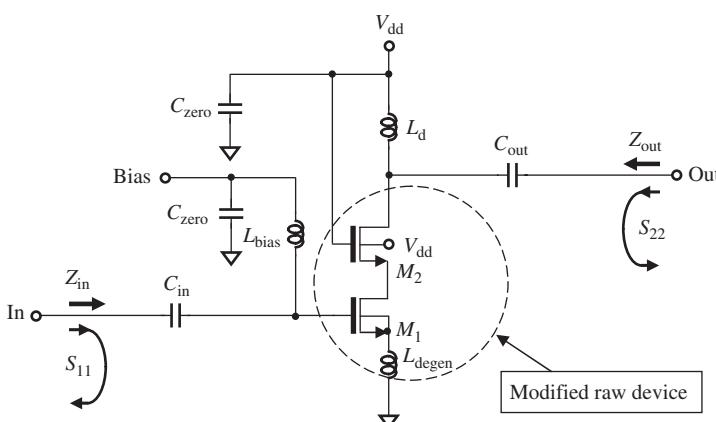
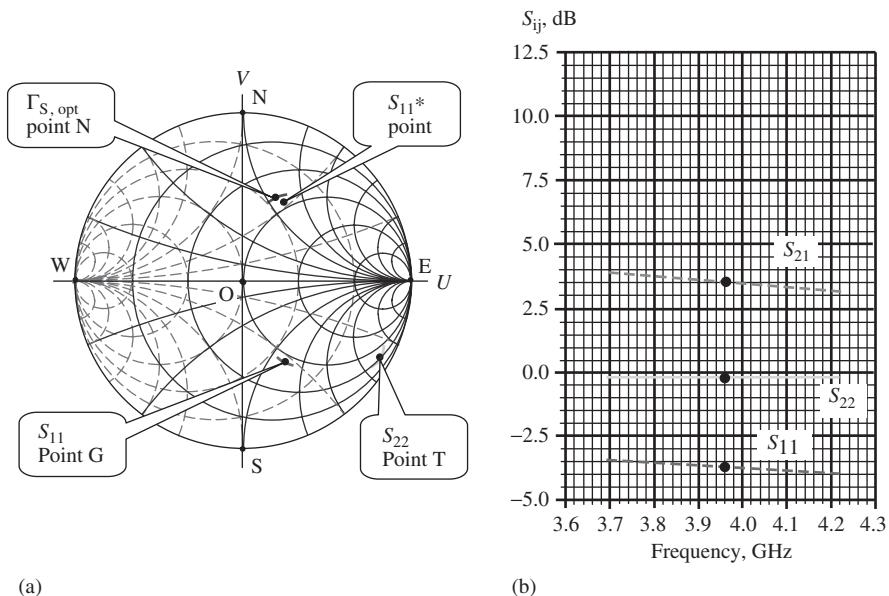
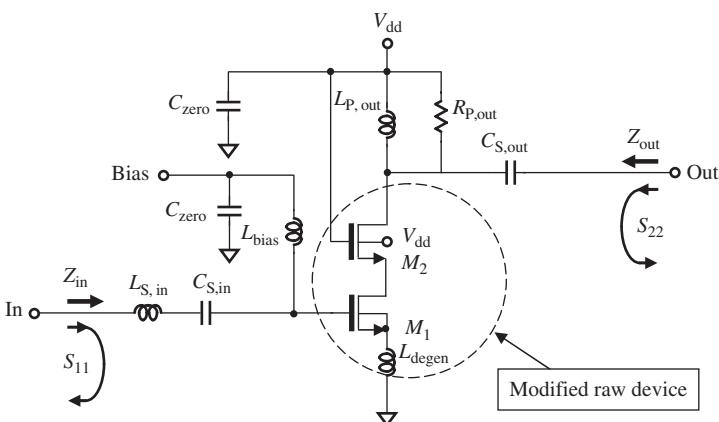


Figure 17.35. Schematic of setup for modified raw device testing.



**Figure 17.36.**  $S$  parameters from modified raw device testing. (The intermediate frequency  $f = 3.96$  GHz is marked by a dot on each trace.) (a)  $S_{11}$ ,  $S_{22}$ , and  $\Gamma_{S,\text{opt}}$  on the Smith Chart; (b) magnitude of  $S_{ij}$ , dB.



**Figure 17.37.** Impedance matching by parts. At input:  $L_{S,in} = 4$  nH and  $C_{S,in} = 10$  pF; at output:  $L_{P,out} = 10$  nH,  $R_{P,out} = 50 \Omega$ , and  $C_{S,out} = 1$  pF.

The first try to implement the impedance matching network is shown in Figure 17.37. It consists of the following parts:

- In the input impedance network:  $L_{S,in} = 4$  nH and  $C_{S,in} = 10$  pF.
- In the output impedance network:  $L_{P,out} = 10$  nH,  $R_{P,out} = 50 \Omega$ , and  $C_{S,out} = 1$  pF.

The inductor  $L_{S,in} = 4$  nH is the main part in the input impedance matching network, by which the trace of  $S_{11}$  is supposed to be pulled to a location near the center of the Smith Chart, which corresponds to the reference impedance point  $50 \Omega$ . The capacitor

$C_{S,in} = 10 \text{ pF}$  functions as the DC blocking part because at the operating frequency range its impedance is very low, such that

$$Z_{10 \text{ pF}} = \frac{1}{j\omega C} = -j \frac{1}{2\pi f C} = 4.02 \Omega, \text{ when } f = 3960 \text{ MHz.} \quad (17.153)$$

The inductor  $L_{P,out} = 10 \text{ nH}$  is the main part in the output impedance matching network, by which the trace of  $S_{22}$  is supposed to be pulled counterclockwise up to a location near the  $50\text{-}\Omega$  resistance circle along the constant conductance circle; then, the capacitor  $C_{S,out} = 1 \text{ pF}$  is the next part, by which the  $S_{22}$  is drawn counterclockwise to a location near the center of the Smith Chart, which corresponds to the reference impedance point  $50 \Omega$  along the constant resistance circle. The difference between these two inductors,  $L_{S,in}$  and  $L_{P,out}$ , should be noted. One is in series and the other in parallel. By means of  $L_{S,in}$  in series, the trace of  $S_{11}$  is pulled clockwise to the area near the reference impedance point  $50 \Omega$  along the constant resistance circle, while by means of  $L_{P,out}$  in parallel, the trace of  $S_{22}$  is pulled counterclockwise to the area near the reference impedance point  $50 \Omega$  along the constant conductance circle.

The raw device shown in Figure 17.35 is unstable because the simulation shows its  $\mu$  value as

$$\mu = 0.95 < 1. \quad (17.154)$$

Looking at the schematic shown in Figure 17.35, the instability may come from the load inductor,  $L_{P,out}$ , and the degenerator inductor,  $L_{degen}$ . However, in order to retain as much voltage drop across the device as possible, so as to reduce the intermodulation, we prefer to use a load inductor,  $L_{P,out}$ , rather than a resistor, even though an inductor is much more expensive than a resistor. On the other hand, in order to satisfy the overlapping condition (17.34) for  $NF_{min}$  without significant increase of the noise figure, we prefer to use a degeneration inductor,  $L_{degen}$ , rather than a degeneration resistor,  $R_{degen}$ . The value of  $L_{degen}$ , however, should not be too high; otherwise, the gain and stability would both suffer. In our design sample,  $L_{degen} = 0.45 \text{ nH}$ . The resistor,  $R_{P,out}$ , is used to de-Q the load inductor  $L_{P,out}$  so as to prevent oscillation; hence, it is the key part in keeping the LNA block stable, while the low value of the degeneration inductor  $L_{degen}$  provides good assistance.

The values of all the parts shown in the schematic in Figure 17.37 are simultaneously optimized to obtain stability with a reasonable gain and noise figure level, that is,

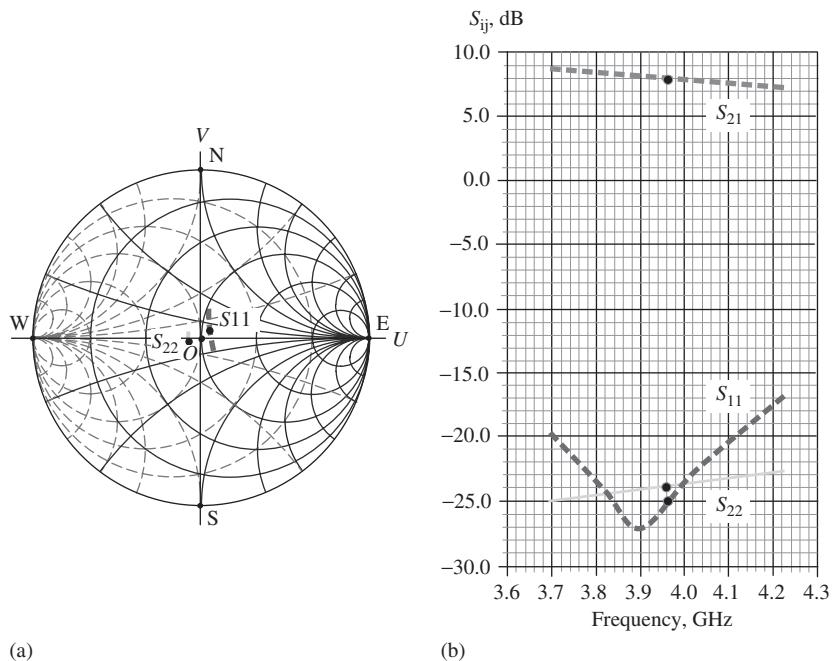
$$\mu = 1.23 > 1, \quad (17.155)$$

when

$$G = 8.0 \text{ dB, and } f = 3.960 \text{ GHz.} \quad (17.156)$$

The  $\mu$  value shown in (17.155) indicates that the unconditionally stable condition is satisfied.

In summary, the input impedance matching network in this example basically consists of only one part. It is a simple impedance matching network indeed. The output impedance matching network consists of three parts, which is a typical part count in an LNA design. Speaking in general, impedance matching is a difficult task and it is one of the main tasks of an RF circuit designer. In the chapters about impedance matching, some



**Figure 17.38.**  $S$  parameters after input and output impedances are matched by parts. (The intermediate frequency  $f = 3.96$  GHz is marked by a dot on each trace.) At input:  $L_{S,\text{in}} = 4$  nH and  $C_{S,\text{in}} = 10$  pF; at output:  $L_{P,\text{out}} = 10$  nH,  $R_{P,\text{out}} = 500 \Omega$ , and  $C_{S,\text{out}} = 1$  pF. (a)  $S_{11}$  and  $S_{22}$  on the Smith Chart; (b) magnitude of  $S_{ij}$ , dB.

useful schemes or technologies are presented for more complicated impedance matching networks.

Figure 17.38 displays the tested  $S$  parameters after the input and output impedance matching networks are implemented.

In the entire operating frequency range, the results are

$$7.2 \text{ dB} < S_{21} < 8.8 \text{ dB}, \quad \text{when } 3.696 \text{ GHz} < f < 4.224 \text{ GHz}, \quad (17.157)$$

$$S_{21} = 8.0 \text{ dB}, \quad \text{when } f = 3.960 \text{ GHz}, \quad (17.158)$$

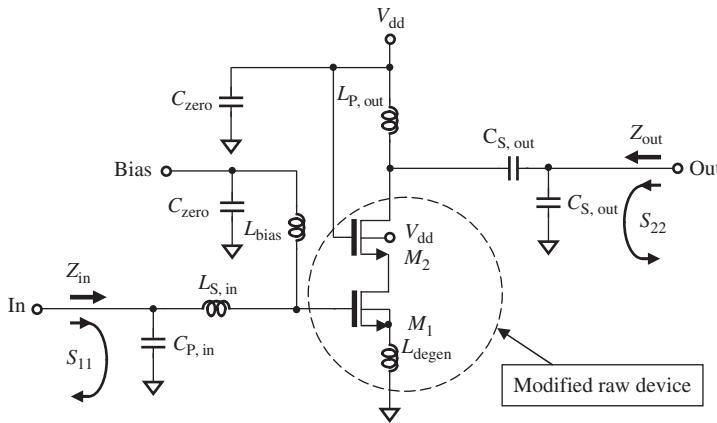
$$-27.1 \text{ dB} < S_{11} < 16.9 \text{ dB}, \quad \text{when } 3.696 \text{ GHz} < f < 4.224 \text{ GHz}, \quad (17.159)$$

$$S_{11} = -25.0 \text{ dB}, \quad \text{when } f = 3.960 \text{ GHz}, \quad (17.160)$$

$$-25.0 \text{ dB} < S_{22} < -23.7 \text{ dB}, \quad \text{when } 3.696 \text{ GHz} < f < 4.224 \text{ GHz}, \quad (17.161)$$

$$S_{22} = -23.9 \text{ dB}, \quad \text{when } f = 3.960 \text{ GHz}. \quad (17.162)$$

By comparing the locations of  $S_{11}$  and  $S_{22}$  shown in Figures 17.36 and 17.38, it can be found that either the  $S_{11}$  or  $S_{22}$  traces are not moved exactly along the resistance or conductance circles. Both traces are moved to the location with higher resistance. In other words, the corresponding resistances of  $S_{11}$  and  $S_{22}$  as shown in Figure 17.38 are increased from those in Figure 17.36. This is due to the inductors  $L_{S,\text{in}}$  and  $L_{P,\text{out}}$ , which are low  $Q$  parts and bring about additional resistance. The  $Q$  value of inductors applied in this section is assumed to be 10, which is a reasonable value in today's RFIC design.



**Figure 17.39.** Modified impedance matching by parts. At input:  $C_{P,\text{in}} = 0.3$  pF and  $L_{S,\text{in}} = 3.3$  nH; at output:  $L_{P,\text{out}} = 5$  nH,  $C_{S,\text{out}} = 0.13$  pF, and  $C_{P,\text{out}} = 0.3$  pF.

The input and output impedances are matched well, and the LNA is in a stable state. However, the gain,  $S_{21} = 8.0\text{--}9.0$  dB, is too low.

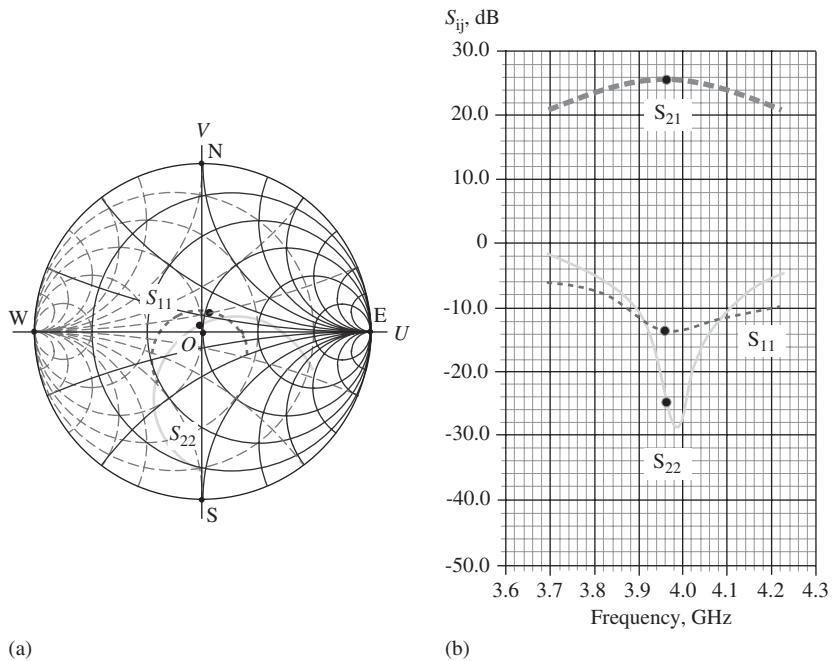
One solution might be to change the topologies of the input and output impedance networks, even though they are quite simple as shown in Figure 17.37 and have the feature of wideband performance, as shown in Figure 17.38.

Figure 17.39 plots the modified impedance matching networks, which consists of the following parts:

- In the input impedance network:  $C_{P,\text{in}} = 0.3$  pF,  $L_{S,\text{in}} = 3.3$  nH, and  $C_{S,\text{in}} = 5$  pF.
- In the output impedance network:  $L_{P,\text{out}} = 5$  nH,  $C_{S,\text{out}} = 0.13$  pF, and  $C_{P,\text{out}} = 0.3$  pF.

The topologies of the input and output impedance matching networks shown in Figure 17.39 are different from those shown in Figure 17.37. In the input impedance network of Figure 17.38, the inductor  $L_{S,\text{in}} = 3.3$  nH plays the same role as the inductor  $L_{S,\text{in}} = 4$  nH in the input impedance network of Figure 17.37. However, after the inductor  $L_{S,\text{in}} = 3.3$  nH is attached to the gate of the device, the location of  $S_{11}$  still does not exactly overlap with the center of the Smith Chart. This implies that further improvements could be made if the trace  $S_{11}$  is made to exactly overlap the center of the Smith Chart. This is the clue or motivation to apply a capacitor,  $C_{P,\text{in}} = 0.3$  pF, to the input impedance network, by which the trace  $S_{11}$  can be pulled clockwise toward the center of the Smith Chart along the constant conductance circle. This will definitely help enhance the power gain of LNA.

In the output impedance network shown in Figure 17.37, the de- $Q$  resistor,  $R_{P,\text{out}} = 500 \Omega$ , is removed. The DC blocking capacitor in Figure 17.37,  $C_{S,\text{out}} = 1$  pF, is replaced by a smaller capacitor,  $C_{S,\text{out}} = 0.13$  pF, and an additional capacitor,  $C_{P,\text{out}} = 0.3$  pF, is added at the output terminal as shown in Figure 17.39. By the addition of  $L_{P,\text{out}} = 5$  nH, the trace  $S_{22}$  is pulled counterclockwise to somewhere in the upper portion of the Smith Chart along the constant conductance circle. Then, by the addition of  $C_{S,\text{out}} = 0.13$  pF, the trace  $S_{22}$  is drawn counterclockwise to somewhere in the upper portion of the Smith Chart along the constant resistance circle. Finally, by the addition of  $C_{P,\text{out}} = 0.3$  pF, the trace  $S_{22}$  is pulled clockwise toward the center of the Smith Chart along the constant conductance circle.



**Figure 17.40.**  $S$  parameters after input and output impedances are matched by parts. (The intermediate frequency  $f = 3.96$  GHz is marked by a dot on each trace.) At input:  $C_{P,in} = 0.3$  pF and  $L_{S,in} = 3.3$  nH; at output:  $L_{P,out} = 5$  nH,  $C_{S,out} = 0.13$  pF, and  $C_{P,out} = 0.3$  pF.

The resulting  $S$  parameters are plotted in Figure 17.40:

$$20.8 \text{ dB} < S_{21} < 25.5 \text{ dB}, \quad \text{when } 3.696 \text{ GHz} < f < 4.224 \text{ GHz}, \quad (17.163)$$

$$S_{21} = 25.5 \text{ dB}, \quad \text{when } f = 3.960 \text{ GHz}, \quad (17.164)$$

$$-13.8 \text{ dB} < S_{11} < -6.2 \text{ dB}, \quad \text{when } 3.696 \text{ GHz} < f < 4.224 \text{ GHz}, \quad (17.165)$$

$$S_{11} = -13.6 \text{ dB}, \quad \text{when } f = 3.960 \text{ GHz}, \quad (17.166)$$

$$-28.8 \text{ dB} < S_{22} < -1.8 \text{ dB}, \quad \text{when } 3.696 \text{ GHz} < f < 4.224 \text{ GHz}, \quad (17.167)$$

$$S_{22} = -25.0 \text{ dB}, \quad \text{when } f = 3.960 \text{ GHz}. \quad (17.168)$$

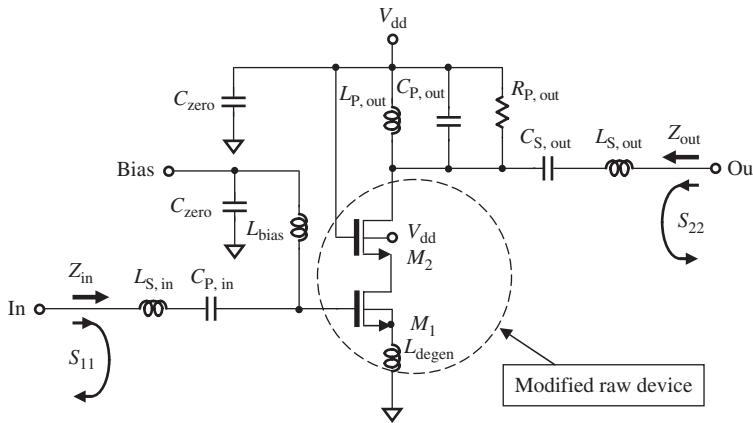
In the entire frequency range, the gain,  $S_{21} > 20.8$  dB, is satisfactory.

Unfortunately, from the frequency response of  $S_{11}$  and  $S_{22}$ , the bandwidth becomes narrow and does not cover the desired frequency range:  $3.696 \text{ GHz} < f < 4.224 \text{ GHz}$ .

Another issue concerns the capacitors  $C_{S,out} = 0.13$  pF and  $C_{P,out} = 0.3$  pF. Their capacitances are too small and therefore unrealistic. One cannot find any discrete capacitor with 0.13 pF capacitance in the market at all. A 0.3-pF capacitor is in the same order as the parasitic or spray capacitance that exists in the circuit layout. These capacitors are therefore unacceptable parts in this circuit implementation.

So, more effort must be put into the impedance matching work. Let us try again!

In general, it is expected that input and output impedance networks could be constructed to satisfy both the goals of power gain and the bandwidth of LNA. Figure 17.41 shows a successful attempt, which consists of the following parts:



**Figure 17.41.** Remodified impedance matching by parts. At input:  $L_{S,\text{in}} = 3.8 \text{ nH}$  and  $C_{S,\text{in}} = 5 \text{ pF}$ ; at output:  $R_{P,\text{out}} = 200 \Omega$ ,  $L_{P,\text{out}} = 2 \text{ nH}$ ,  $C_{P,\text{out}} = 0.97 \text{ pF}$ ,  $C_{S,\text{out}} = 10 \text{ pF}$ , and  $L_{S,\text{out}} = 3.3 \text{ nH}$ .

- At input impedance matching network:  $L_{S,\text{in}} = 3.8 \text{ nH}$  and  $C_{S,\text{in}} = 5 \text{ pF}$ ,
- At output impedance matching network:  $R_{P,\text{out}} = 200 \Omega$ ,  $L_{P,\text{out}} = 2 \text{ nH}$ ,  $C_{P,\text{out}} = 0.97 \text{ pF}$ ,  $C_{S,\text{out}} = 10 \text{ pF}$ , and  $L_{S,\text{out}} = 3.3 \text{ nH}$ .

Special topologies are applied to both input and output impedance matching networks.

The input impedance network is an LC “arm,” in which two parts,  $L_{S,\text{in}}$  and  $C_{S,\text{in}}$ , are connected in series. The output impedance matching network consists of one “branch” and one “arm.” The “branch” is built by three parts,  $L_{P,\text{out}}$ ,  $C_{P,\text{out}}$ , and  $R_{P,\text{out}}$ , which are connected in parallel, while the “arm” is built by two parts,  $C_{S,\text{out}}$  and  $L_{S,\text{out}}$ , which are connected in series. As discussed in Chapter 4, either the “arm” or the “branch” is a special “part,” which is effectively applied in the wideband impedance matching network. Readers are encouraged to read the related chapter about wideband impedance matching in this book so as to understand more schemes on wideband impedance matching. Figure 17.42 shows its performance:

$$13.0 \text{ dB} < S_{21} < 14.3 \text{ dB}, \quad \text{when } 3.696 \text{ GHz} < f < 4.224 \text{ GHz}, \quad (17.169)$$

$$S_{21} = 14.0 \text{ dB}, \quad \text{when } f = 3.960 \text{ GHz}, \quad (17.170)$$

$$-26.2 \text{ dB} < S_{11} < -15.9 \text{ dB}, \quad \text{when } 3.696 \text{ GHz} < f < 4.224 \text{ GHz}, \quad (17.171)$$

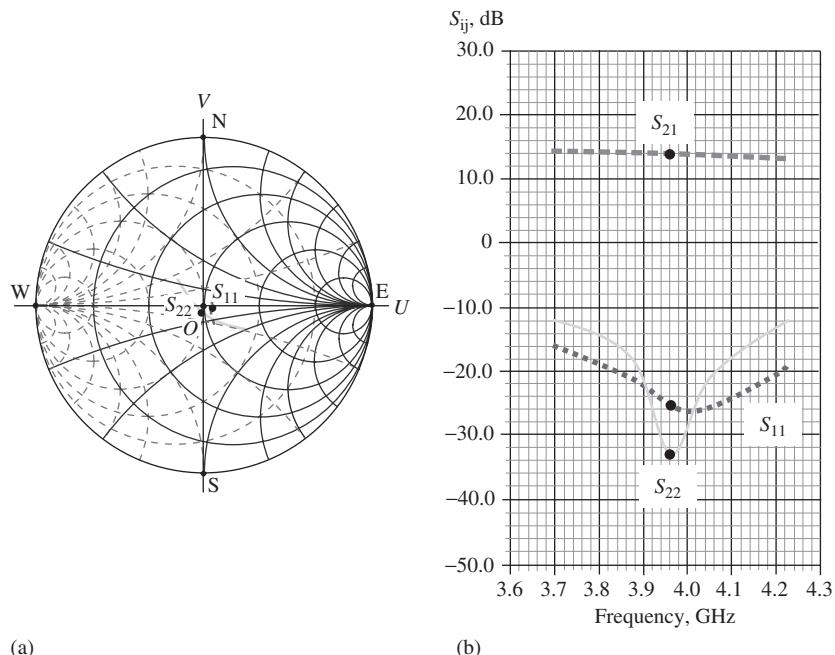
$$S_{11} = -25.5 \text{ dB}, \quad \text{when } 3.830 < f < 4.200 \text{ GHz}, \quad (17.172)$$

$$-33.2 \text{ dB} < S_{22} < -11.9 \text{ dB}, \quad \text{when } 3.696 \text{ GHz} < f < 4.224 \text{ GHz}, \quad (17.173)$$

$$S_{22} = -33.1 \text{ dB}, \quad \text{when } 3.895 < f <= 4.035 \text{ GHz}. \quad (17.174)$$

Now the gain is reasonable and the bandwidth is wide enough. In addition, the values of the parts are appropriate.

From the demonstrations above, we recognized that impedance matching is not an easy task. Not only the gain but also the bandwidth needs to be taken care of. In addition, there are many solutions to the same problem. The designer must select one of the possible solutions with the least part count, smallest size, lowest cost and current drain, and best performance.



**Figure 17.42.**  $S$  parameters after input and output impedances are matched by parts. (The intermediate frequency  $f = 3.96$  GHz is marked by a dot on each trace.) At input:  $L_{S,\text{in}} = 3.8$  nH and  $C_{S,\text{in}} = 5$  pF; at output:  $R_{P,\text{out}} = 200 \Omega$ ,  $L_{P,\text{out}} = 2$  nH,  $C_{P,\text{out}} = 0.97$  pF,  $C_{S,\text{out}} = 10$  pF, and  $L_{S,\text{out}} = 3.3$  nH. (a)  $S_{11}$  and  $S_{22}$  on the Smith Chart; (b) magnitude of  $S_{ij}$ , dB.

**17.3.4.3 Noise Figure.** The noise performance of the design sample is good after impedance matching is done. This is expected because the modified raw device is designed to satisfy the condition (17.34) so as to simultaneously approach maximum gain and minimum noise figure. Figure 17.43 shows the performance of the noise figure over the entire frequency range:

$$\text{NF} = 1.37 \text{ dB}, \quad \text{when } f = 3.696 \text{ GHz}, \quad (17.175)$$

$$\text{NF} = 1.80 \text{ dB}, \quad \text{when } f = 3.960 \text{ GHz}, \quad (17.176)$$

and

$$\text{NF} = 2.40 \text{ dB}, \quad \text{when } f = 4.224 \text{ GHz}. \quad (17.177)$$

Exciting results are found from the plot of the gain circles and noise figure circles as shown in Figure 17.44. This is a plot for the operating frequency  $f = 3.960$  GHz. After the input and output impedances are matched, the center of the gain circles is point G, which is very close to the center of the Smith Chart. The maximum gain at this operating frequency is 14 dB, which is consistent with the results shown in Figure 17.42(b). On the other hand, the center of the noise figure circles is point N, which is also very close to the center of the Smith Chart. The minimum noise figure at this operating frequency is 1.8 dB, which again is consistent with the results shown in Figure 17.43.

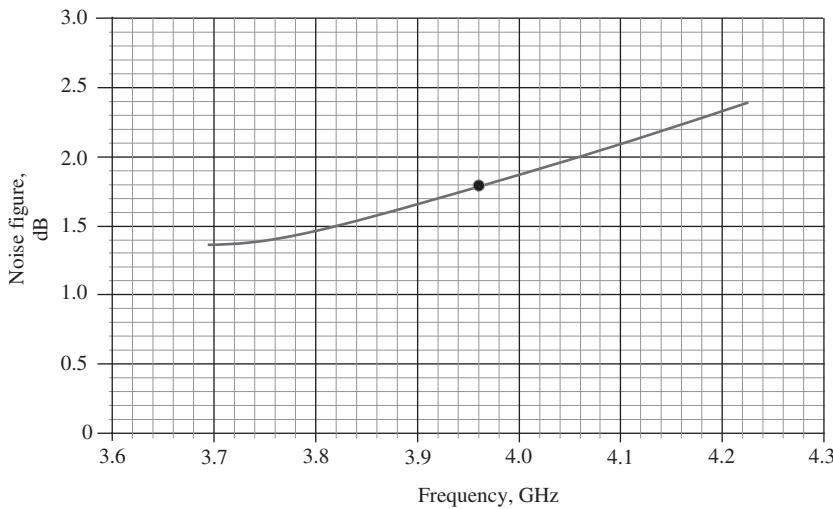
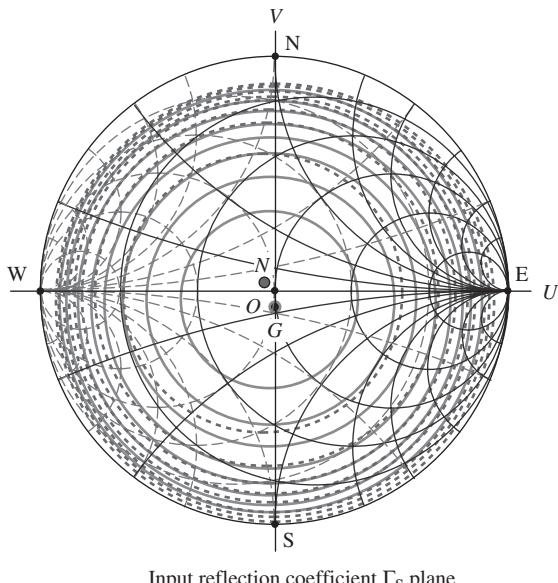


Figure 17.43. Noise figure from 3.696 to 4.224 GHz.  $NF = 1.8$  dB when  $f = 3.960$  GHz.



Input reflection coefficient  $\Gamma_S$  plane

- Gain circles:  $G_{\max} = 14$  dB at point G, step = 1.0 dB,
- Noise figure circles:  $NF_{\min} = 1.8$  dB at point N, step = 0.5 dB,

Figure 17.44. Constant gain circles and constant noise figure circles when  $f = 3.96$  GHz.

The wonderful feature in Figure 17.44 is that the circles and noise circles overlap each other almost completely. The two centers are very close to each other near the center of the Smith Chart. The deviation between point N and point O is due to the tolerance existing in the condition (17.34),  $\Gamma_{S,\text{opt}} = S_{11}^*$ , while the deviation between point G and point O is due to the tolerance of parts in the impedance matching networks. In reality, it is impossible to reach a perfect goal predicted by theory. Deviation between the practical design and the circuit theory always exists.

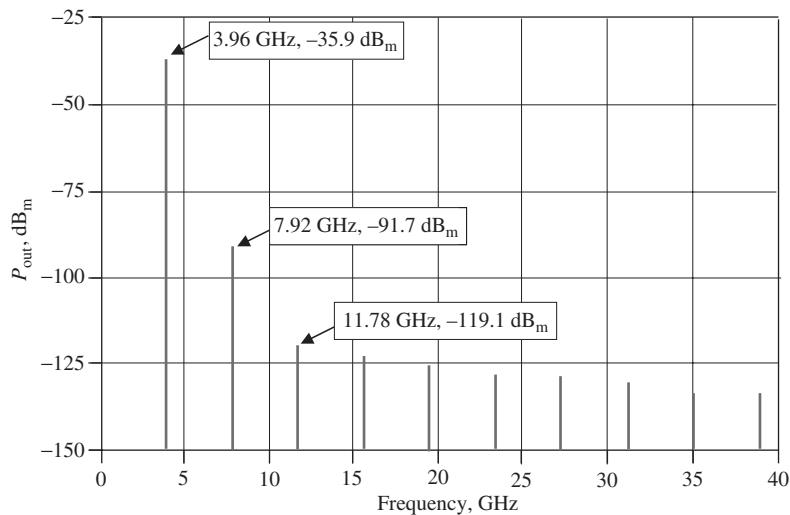


Figure 17.45. Spectrum at LNA output,  $f_o = 3.96$  GHz,  $P_{in} = -50$  dB<sub>m</sub>.

**17.3.4.4 Nonlinearity.** At the output of the LNA, the frequency spectrum is shown in Figure 17.45. The first (the desired signal), second, and third harmonics are:

$$P_{out} = -35.9 \text{ dB}_m, \quad \text{at } f = 3.96 \text{ GHz}, \quad (17.178)$$

$$P_{out} = -91.7 \text{ dB}_m, \quad \text{at } f = 7.92 \text{ GHz}, \quad (17.179)$$

$$P_{out} = -119.1 \text{ dB}_m, \quad \text{at } f = 11.78 \text{ GHz}, \quad (17.180)$$

when  $P_{in} = -50$  dB<sub>m</sub>.

The gain of the LNA is

$$P_{out} - P_{in} = -35.9 \text{ dB}_m - (-50 \text{ dB}_m) = 14.1 \text{ dB}, \quad (17.181)$$

which is approximately consistent with the gain testing of  $S_{21} = 14.0$  dB as shown in Figure 17.42(b).

The second and third harmonics are lower than the output power at the operating frequency by

$$\Delta P_{out,2} = P_{out}|_{7.92 \text{ GHz}} - P_{out}|_{3.96 \text{ GHz}} = -91.7 \text{ dB}_m - (-35.9 \text{ dB}_m) = -55.8 \text{ dB}, \quad (17.182)$$

and

$$\Delta P_{out,3} = P_{out}|_{11.78 \text{ GHz}} - P_{out}|_{3.96 \text{ GHz}} = -119.1 \text{ dB}_m - (-35.9 \text{ dB}_m) = -83.2 \text{ dB}, \quad (17.183)$$

respectively, which implies that low-value spurious products existed in this LNA block. It is quite good news that we need not worry about DC offset and the interference of adjacent channels anymore.

Figures 17.46–17.48 show the 1-dB compression point, third-order intercept point, and second-order intercept point, respectively.

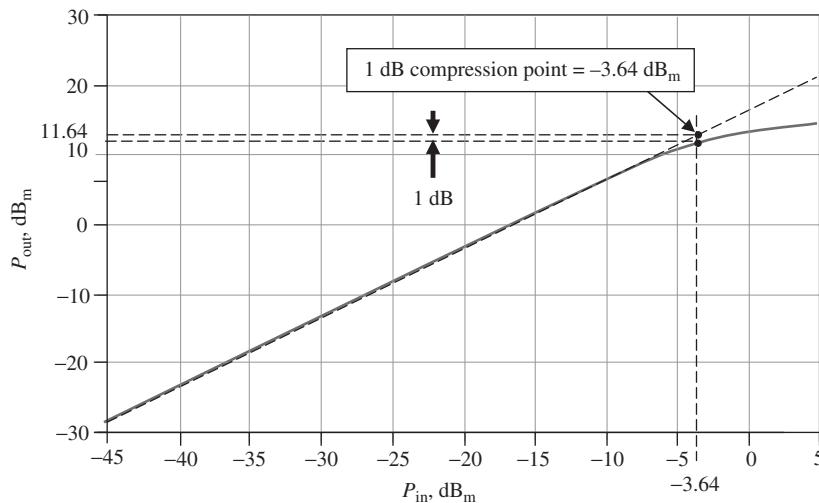


Figure 17.46. A 1-dB compression point when \$f\_o = 3.96 \text{ GHz}\$.

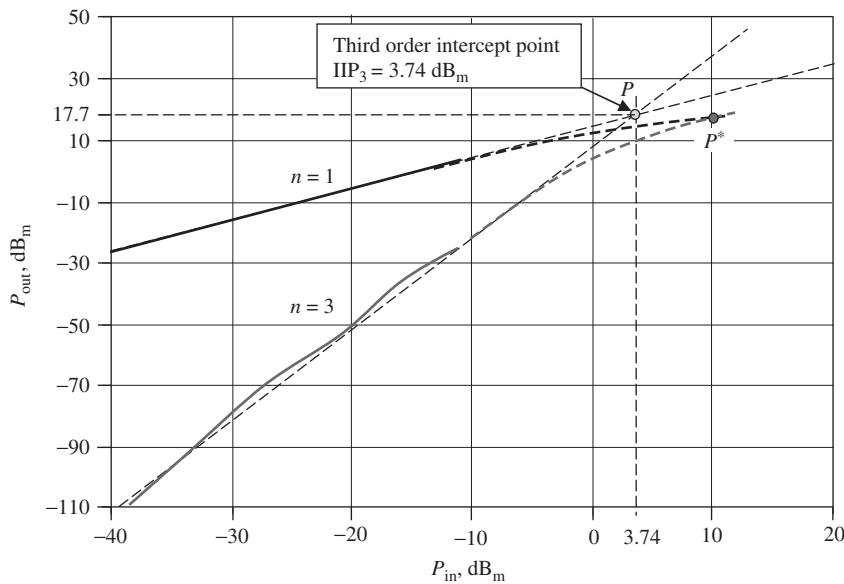


Figure 17.47. Third-order input intercept point when \$f\_o = 3.96 \text{ GHz}\$.

The 1-dB compression point is

$$P_{1 \text{ dB}} = -3.64 \text{ dB}_m, \quad \text{when } f = 3.96 \text{ GHz}. \quad (17.184)$$

The third-order and second-order intercept points are

$$\text{IIP}_3 = 3.74 \text{ dB}_m, \quad \text{when } f = 3.96 \text{ GHz}, \quad (17.185)$$

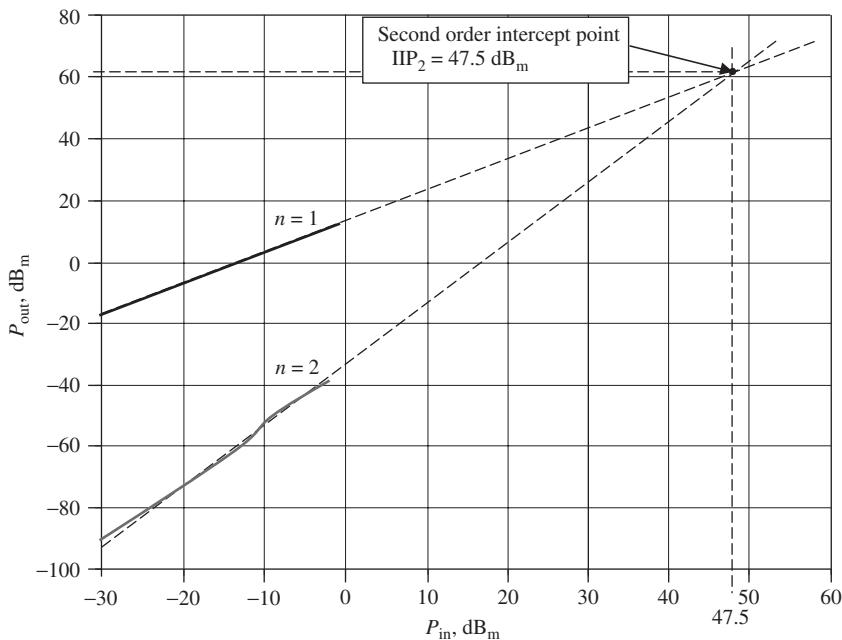


Figure 17.48. Second-order input intercept point when  $f_o = 3.96$  GHz.

and

$$\text{IIP}_2 = -47.5 \text{ dB}_m, \quad \text{when } f = 3.96 \text{ GHz}, \quad (17.186)$$

respectively.

## 17.4 LNA WITH AGC (AUTOMATIC GAIN CONTROL)

### 17.4.1 AGC Operation

Received signals very often suffer from fading, so that the fluctuation of the field strength of radio signals at the receiver antenna always exists. In order to maintain a relatively constant output to the sensor, such as the speaker, it is necessary to detect the strength of the received signal and then adjust the gain of the receiver automatically. This is the subject of AGC operation.

In the early developing stages of portable radios or cellular phones, modulation technology was relatively simple and therefore it was not necessary to apply AGC technology in developed products. As advanced modulation technology arose, AGC technology became required and sometimes played a decisive role in a receiver.

In the early development stages of the cellular phone, a 20-dB dynamic range of AGC in LNA design was occasionally required. However, in today's CDMA or WCDMA receivers, an AGC range of up to 90 dB is required. This is really an astronomic number! It is well known that every 3-dB variation corresponds to a 50% or double power variation. A 90-dB variation implies that the received power or field strength around the

antenna could vary by as many times as

$$2^{\frac{90}{3}} = 1,073,741,824$$

The wide dynamic range of AGC is indeed a challenge to circuit designers.

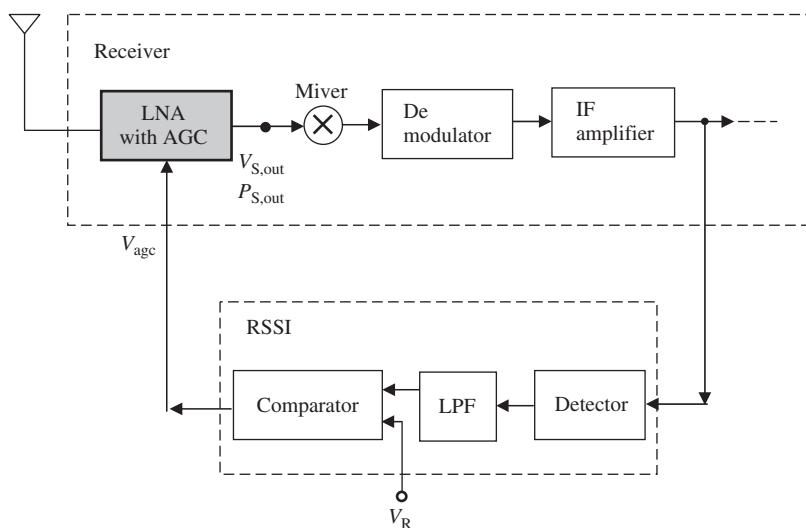
It is well known that the amount of power gain in a receiver is mostly contributed by the back end after the demodulator. In order to be able to set up 90 dB of AGC in the back end, the total power gain in the back end must roughly be higher than 90 dB. This will easily and inevitably bring about oscillation in those amplification blocks because of the high gain required. Eventually, the 90 dB of power variation cannot rely on the AGC in the back end only. The front end or RF designers must share the task of AGC for a 90-dB power variation. At present, the task of AGC in a radio or a cellular phone is distributed to both the front end and the back end in an approximately 50:50 ratio. The LNA at the front end must then have an AGC of 40–50 dB, as other blocks in the front end are difficult to work with for AGC.

An LNA with AGC is sometimes called a VGA (variable gain amplifier).

An important concern is that the LNA with AGC should not produce additional distortion on the modulated signal. At present, AGC requirements for AM receivers are usually more stringent than those for FM or PM receivers.

The AGC loop is usually implemented in the receiver of a wireless communication system, which can be illustrated in Figure 17.49. There is a special block called the RSSI (received signal strength indicator) for AGC in the receiver. It detects the received signal from the IF amplifier output and compares it with a reference signal,  $V_R$ , in a comparator. The difference between these two signals is the output of the RSSI,  $V_{AGC}$ , which is used to control the gain of the LNA.

The analysis and design for the AGC loop must be conducted at the system level. Here we focus on the LNA with AGC block only. The characteristics of the LNA with an AGC block are shown in Figure 17.50. For low input received signals the AGC is



**Figure 17.49.** An AGC loop in the receiver of a communication system.

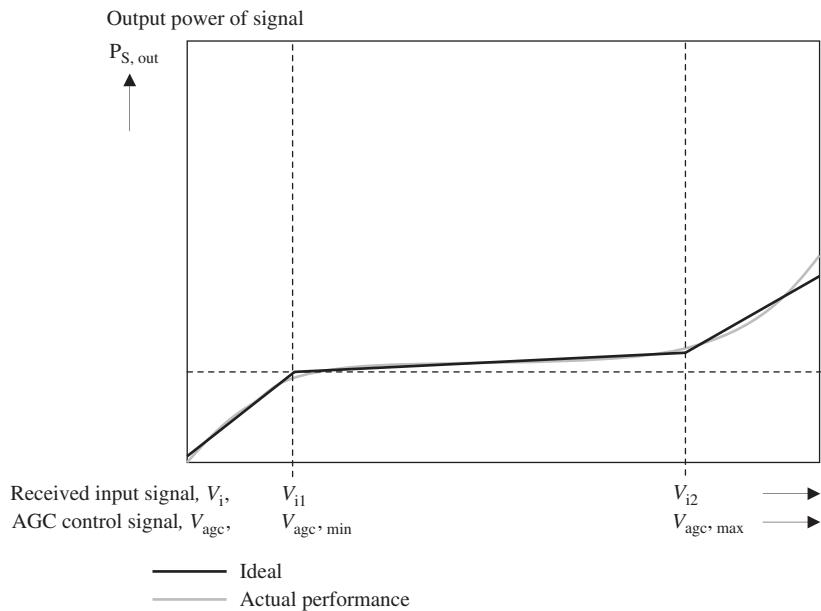


Figure 17.50. Expected characteristics of AGC.

inactivated, so that the output of LNA,  $V_{S,\text{out}}$  or  $P_{S,\text{out}}$  is linearly related to the input received signal  $V_i$ . When the input received signal is increased above  $V_{i1}$ , the AGC loop is activated and will maintain the output level at a constant value or with a tiny increase until the input received signal increases above  $V_{i2}$ . The minimum and maximum output voltages from the RSSI block,  $V_{\text{AGC},\text{min}}$  and  $V_{\text{AGC},\text{max}}$ , correspond to  $V_{i1}$  and  $V_{i2}$ , respectively. When the input received signal is stronger than  $V_{i2}$ , the AGC loses its control capability and the LNA output,  $V_{S,\text{out}}$  or  $P_{S,\text{out}}$ , returns to the linear relationship with the input received signal  $V_i$ . Instead of suddenly activating or inactivating the AGC, the actual performance of the AGC produces a somewhat smooth curve as shown in Figure 17.50.

The dynamic range of the AGC voltage is

$$\text{DR}_{\text{AGC}} = V_{i,2} - V_{i,1} = V_{\text{AGC},\text{max}} - V_{\text{AGC},\text{min}}, \quad (17.187)$$

in which the power variation of received signal at antenna, say, 40–50 dB, can be “swallowed.”

#### 17.4.2 Traditional LNA with AGC

Figure 17.51 shows a traditional LNA with AGC. Its configuration is a differential type with a cascode LNA branch, which consists of  $M_1$  and  $M_2$ , and an AGC branch, which consists of  $M_1$  and  $M_3$ . The function of each part is as follows:

- $M_1, M_2$ —cascoded devices;
- $C_1, C_2$ —AC short-circuited or “zero” capacitor;
- $L_1$ —load inductor;

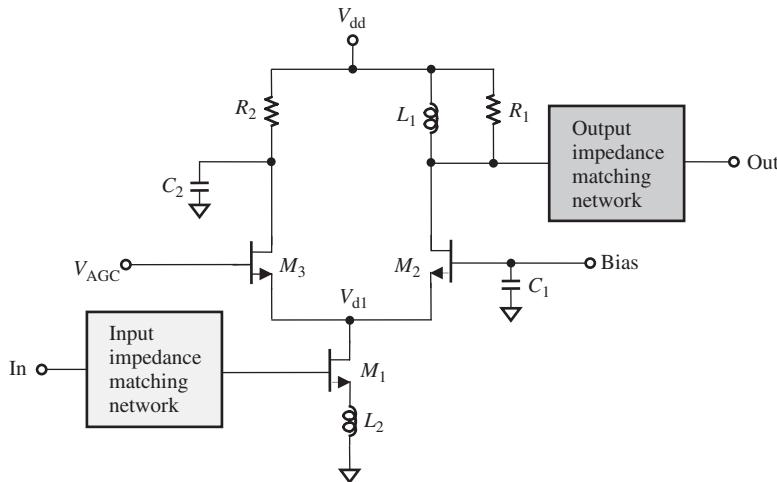


Figure 17.51. Traditional LNA with AGC.

- $R_1$ —stability resistor;
- $L_2$ —degeneration inductor;
- $M_3$ —AGC effective device;
- $R_2$ —current control resistor.

The input to the gate of the transistor  $M_3$ ,  $V_{AGC}$ , is a DC voltage from the RSSI output in the radio. When the RSSI voltage is lower than  $V_{AGC,\min}$ ,  $M_3$  is turned off and AGC function is not activated. However, if the RSSI output voltage,  $V_{AGC}$ , is increased above  $V_{AGC,\min}$ ,  $M_3$  is turned on and shares the AC from the  $M_1$ – $M_2$  cascode LNA branch. As the RSSI voltage continuously rises,  $M_3$  will share more AC. The AC shared by  $M_3$  is shorted to the ground by the capacitor  $C_2$ . Consequently, the gain of the cascode LNA branch  $M_1$ – $M_2$  will be reduced because of the AC being partially shared and shorted to the ground by the AGC branch  $M_1$ – $M_3$ . Intuitively, the higher the RSSI control voltage  $V_{AGC}$ , the more the reduction of LNA gain.

The normal gain of the LNA is defined as the gain when the AGC branch is not activated, that is, the RSSI voltage  $V_{AGC}$  is lower than  $V_{AGC,\min}$ . The LNA gain can be reduced when the AGC branch becomes effective.

As mentioned above, the AGC dynamic range is required to be at least 40 dB. However, with the traditional LNA with AGC design shown in Figure 17.51, the AGC dynamic range only approaches 20 dB. This limitation is due to the restrictions of the topology. Let us take a look at Figure 17.51. Should the drain voltage of  $M_1$ ,  $V_{d1}$ , always trace the variation of RSSI output voltage  $V_{AGC}$  at the gate of  $M_3$ , the LNA gain might be continuously reduced as the RSSI output voltage  $V_{AGC}$  at the gate of  $M_3$  is increased, and then the AGC dynamic range could be much higher than 20 dB.

Unfortunately, owing to the existence of the  $M_1$ – $M_2$  branch, the variation of voltage  $V_{d1}$  is not as expected. At the beginning,  $V_{d1}$  increases as the RSSI voltage increases, so that the AC is shared by the AGC branch and the LNA gain is lowered. However, as the RSSI output voltage  $V_{AGC}$  increased past a certain value,  $V_{d1}$  stops increasing, or even inversely decreases. This leads to the increase in LNA gain back in the direction toward its normal gain without AGC. It is therefore that the AGC dynamic range is limited.

### 17.4.3 Increase in AGC Dynamic Range

In order to increase the AGC dynamic range, there are three improvements to be adapted:

1. *Diode Clamping.* In order to force the voltage  $V_{d1}$  to trace the variation of the RSSI voltage, a device  $M_4$  with a diode connection is added to the gate and source of  $M_3$  as shown in Figure 17.52.
  2. *Avoiding RSSI Operation as a Current Sink or Source.* Without the additional resistor  $R_3$  and the additional device  $M_4$  as shown in Figure 17.52, the gate of  $M_3$  is the load of the RSSI output, which is extremely light since the input impedance of  $M_3$  is quite high. Now, the addition of the device  $M_4$  brings about a problem to the RSSI output. The device  $M_4$  is DC connected from RSSI output to the source of  $M_3$  and the rest of the circuit. This makes the RSSI output port a DC source or sink. The current from or into the RSSI output could be in the order of milliamperes or more, which the RSSI output terminal cannot usually afford.

The solution to this problem is to add a resistor  $R_3$  between the  $V_{dd}$  and the gate of  $M_3$  as shown in Figure 17.53. An appropriate adjustment of the value of  $R_3$  can make the current either flowing from or into the RSSI output less than 0.5 mA.

3. *Selection of Big Device  $M_3$ .* In the design of LNA with AGC, it is found that the selection of the device  $M_3$  is very important. From a system design viewpoint, it is desirable to reduce the current drain as much as possible. Therefore, the  $I_{ds}$  current of  $M_3$  should not be too high when it is turned on. On other hand, its  $I_{ds}$  current should be very low when it is turned off. In order to satisfy the above two requirements simultaneously, one cannot but select a big device  $M_3$  with large number of “fingers.”

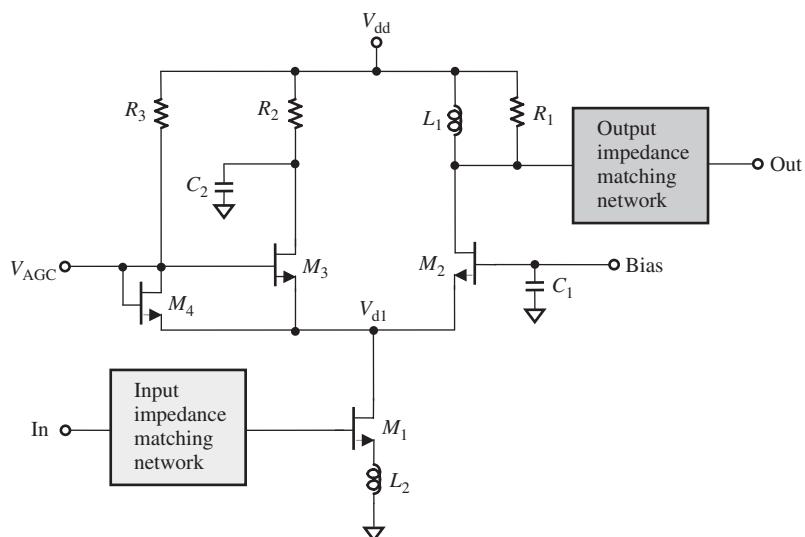


Figure 17.52. LNA with large dynamic range AGC.

### 17.4.4 Example

The schematic as shown in Figure 17.52 has been simulated with the CMOS process, and its corresponding work on the bench has been done.

In this design, the number of fingers is 8 and 16 for  $M_1$  and  $M_2$ , respectively, whereas the number of fingers for  $M_3$  is selected as 100. (In CMOS processing, the width of one finger is 16  $\mu\text{m}$ .) The DC characteristics of  $M_3$  are shown in Figure 17.53 and can be

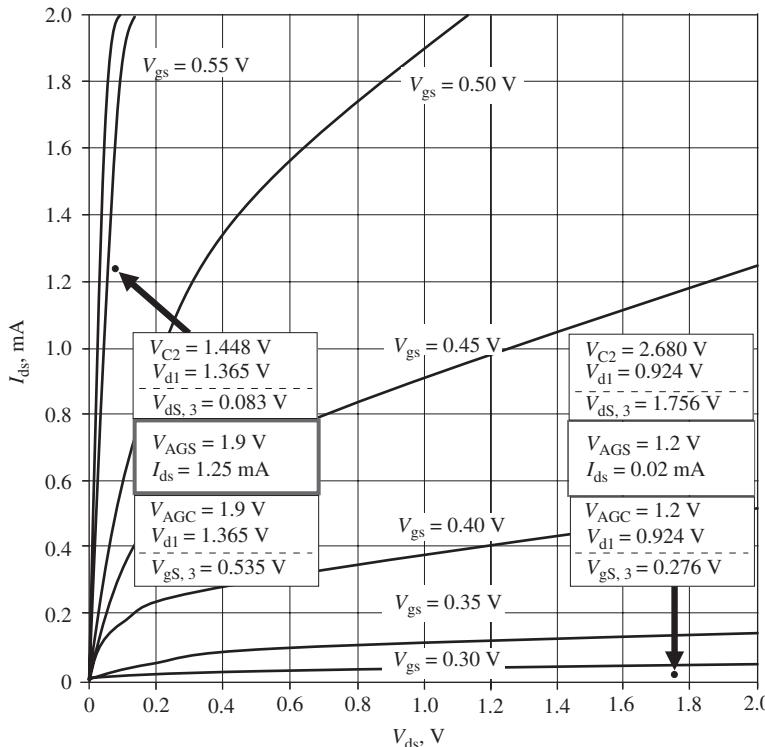


Figure 17.53. DC characteristics of transistor  $M_3$ .

TABLE 17.5. Part List of LNA with AGC

Part	Value	
$M_1$	Fingers = 8	$W_{\text{total}} = 128 \mu\text{m}$
$M_2$	Fingers = 16	$W_{\text{total}} = 256 \mu\text{m}$
$M_3$	Fingers = 100	$W_{\text{total}} = 1600 \mu\text{m}$
$M_4$	Fingers = 16	$W_{\text{total}} = 1256 \mu\text{m}$
$R_1$	700 $\Omega$	
$R_2$	1 k $\Omega$	
$R_3$	10 k $\Omega$	
$C_1$	20 pF	
$C_2$	40 pF	
$L_1$	50 nH	
$L_2$	30 nH	

TABLE 17.6. Goals, Simulation, and Performance of an LNA with AGC

	Goal	Simulation			Performance		
DC power supply, $V$	3.0		3.0		3.0		V
Frequency, $f$	460–470		460–470		460–470		MHz
RSSI control voltage, $V_{AGC}$	1.2	1.9	1.2	1.9	1.2	1.9	V
AGC dynamic range, $DR_{AGC}$	0.00	−40	0.00	−43.7	0.0	−42	dB
Current drain, $I_{ds}$	2.0	*	1.87	1.48	2.0	1.4	mA
Gain, $G$	12.0	−28	12.6	−31.1	15.0	−27.0	dB
Noise figure, NF	2.0	*	1.26	29.4	2.7	*	dB
Third-order intercept point, $IIP_3$	−15.0	*	−10.2	−7.5	−10.5	−8.0	$dB_m$
Input return loss, $S_{11}$	*	*	−15.3	−10.4	*	*	dB
Output return loss, $S_{22}$	*	*	−23.2	−11.7	*	*	dB

\*indicates that the value is not available due to reading loss or reading absence.

outlined as follows:

$$I_{ds} \approx 0.02 \text{ mA or } 20 \mu\text{A}, \quad \text{when } V_{AGC} = 1.2 \text{ V}, \quad (17.188)$$

$$I_{ds} \approx 1.25 \text{ mA}, \quad \text{when } V_{AGC} = 1.9 \text{ V}. \quad (17.189)$$

Table 17.5 lists the parts, and Table 17.6 lists their performance.

This LNA with a large dynamic range of AGC design has been integrated with a 1-W PA in an IC chip, which has been successfully applied to a radio production line.

## REFERENCES

- Gray PR, Hurst PJ, Lewis SH, Meyer RG. *Analysis and Design of Analog Integrated Circuits*. 4th ed. John Wiley & Sons, Inc.; 2004, p. 197–207.
- Gray PR, Hurst PJ, Lewis SH, Meyer RG. *Analysis and Design of Analog Integrated Circuits*. 4th ed. John Wiley & Sons, Inc.; 2004b, p. 208–209.
- Lee TH. *The Design of CMOS Radio-frequency Integrated Circuits*. Cambridge University Press; 1998.

## FURTHER READING

- Cassan DJ, Long JR. A 1V 0.9 dB NF Low noise Amplifier for 5–6 GHz WLAN in 0.18  $\mu\text{m}$  CMOS. IEEE 2002 Custom Integrated Circuits Conference, Orlando, FL, 2000. pp. 419–422.
- Cassan DJ, Long JR. A1-V transformer feedback low noise amplifier for 5 GHz wireless LAN in 0.18  $\mu\text{m}$  CMOS, IEEE J Solid-State Circuits 2003; 38(3): 427–435.
- Chen J, Shi B. Impact of intrinsic channel resistance on noise performance of CMPS LNA. IEEE Electron Device Lett 2002; 23(1): 34–36.

- Gatta F, Sacchi E, Svelto F, et al. A 2dB Noise Figure 900 MHz Differential CMOS LNA. *IEEE J Solid-State Circuits* 2001; 36(10): 1444–1452.
- Girlando G, Palmisano G Noise figure and impedance matching in RF cascode amplifiers. *IEEE Trans Circuits Syst Analog Digital Signal Process (IEEE Trans Circuits Syst Express Briefs)* 1999; 46(11): 1388–1396.
- Gonzalez G. *Microwave transistor amplifiers*. Englewood Cliffs, NJ: Prentice-Hall, Inc; 1984.
- Goo JS, Ahn HT, Donald J, et al. A noise optimization technique for integrated low noise amplifier. *IEEE Microw Guid Wave Lett* 2002;37(8): 994–1002.
- Gray PR, Hurst PJ, Lewis SH, et al. *Analysis and design of analog integrated circuits*. 4th ed. New York, NY: John Wiley & Sons, Inc; 2001. pp. 206–212, 522–532.
- Gramagna G, Magazza A, Sclafani C, et al. Ultra-wide dynamic range 1.75 dB noise figure, 900 MHz CMOS LNA. *ISSCC 2000, IEEE International Solid-State Circuit Conference*, San Francisco, CA, 2000. pp. 380–381.
- Haus HA, et al. Representation of noise in linear two ports. *Proc IRE* 1960; 48: 69–74.
- Huang Qiuting, Orsatti Paolo, Piazza Francesco. GSM transceiver front-end circuits in 0.25  $\mu\text{m}$  CMOS. *IEEE J Solid-State Circuits* 1999; 34(3): 292–302.
- Karanicolas AN. A 2.7V 900 MHz CMOS LNA and Mixer. *1996 IEEE International Solid State Circuits Conference*, San Francisco, CA. 1996. pp. 50–53.
- Karimi-Sanjaani A, Sjoland H, Asad, AA. A 2 GHz merged CMOS LNA and mixer for WCDMA. *2001 Symposium on VLSI Circuit Digest of Technical Papers*, Kyoto, Japan, 2001. pp. 19–22.
- Kim CS, Park M, Kim CH, et al. Design guide of coupling between inductors and its effect on reverse isolation of a CMOS LNA. *2000 IEEE MTT\_S Digest*, Boston, MA, 2000. pp. 225–228.
- Lin Y, Obrecht M, Manku T. RF noise characterization of MOS devices for LNA design using a physical-based quasi-3-D approach. *IEEE Trans Circ Syst Analog Digital Signal Process* 2001; 48(10): 972–984.
- Svelto F, Montagna G, Deantoni S, et al. Solution for image rejection CMOS LNA. *ISCAS IEEE International Symposium On Circuits And Systems*; 2000 May 28–31; Geneva, Switzerland. 2000. pp. III49–III52.
- Svelto F, Deantoni S, Montagna G, et al. Implementation of a CMOS LNA plus mixer for GPS applications with no external components. *IEEE Trans Very Large Scale Integr VLSI Syst* 2001; 9(1): 100–104.
- Tsui H-Y, Lau J, SPICE simulation and trade-offs of CMOS LNA performance with source degeneration inductor. *IEEE Trans Circuits Systems-II: Analog and Digital Signal Processing* 2000; 47(1): 62–65.

## EXERCISES

1. What is the purpose of raw device testing?
2. How is raw device testing setup?
3. What is the variation of input impedance before and after a degeneration inductor is added to the device if the device is MOSFET and bipolar respectively.
4. What is the key criterion to get  $G_{\max}$  and  $NF_{\min}$  simultaneously in the LNA design?
5. List the possible ways to force the raw device into satisfying the condition  $S_{11}^* = \Gamma_{S,\text{opt}}$ .
6. Describe the gain circle and noise circles in the Smith Chart.
7. What does the optimized  $\Gamma_{S,\text{opt}}$  for a raw device mean?
8. What is the key step in the new design scheme for LNA suggested by Richard Li?

9. Summarize the main features of Haus' theory?
10. The noise figure of an LNA depends mainly on its source but not on its load, why?
11. If the condition  $S_{11}^* = \Gamma_{S,\text{opt}}$  is not satisfied from raw device testing, could this condition be approached through the input impedance matching work?
12. What is the purpose of a cascode LNA?
13. Describe the  $K$  factor and  $\mu$  factor for the stability of LNA.
14. What is the meaning of AGC? Why does a receiver need LNA with AGC function?

## ANSWERS

1. The purpose of raw device testing is twofold:
  - (a) To see if this raw device can approach a good LNA design or not. A good LNA design is that the minimum of noise figure and the maximum of gain can be obtained simultaneously.
  - (b) To create a starting point for impedance matching so as to continue the next design step.
2. The setup for raw device testing is as follows (Fig. 17.P.1):
  - (a) On the basis of the design specification, choose the expected DC voltage and bias of power supply or battery.
  - (b) Connect DC voltage and bias to the device and setup the DC to the expected value by adjustment of the bias.
  - (c) In order to feed DC and bias, two “zero” capacitors and two “infinite” inductors must be connected to the device.
  - (d) Test for  $S$  parameters.

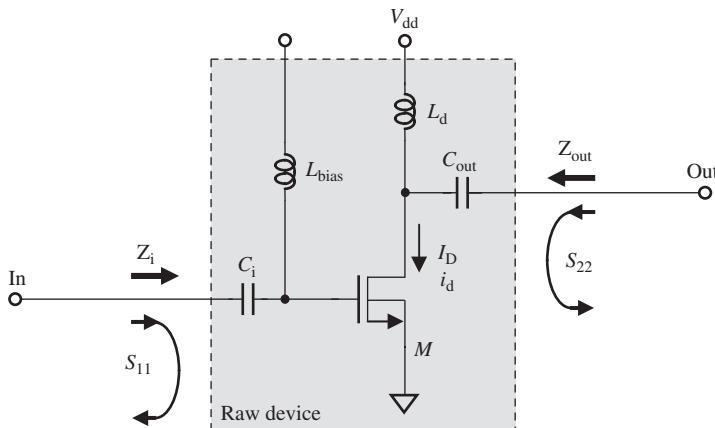


Figure 17.P.1. Setup for raw device testing.  $C_{\text{in}}, C_{\text{out}}$ : “zero” capacitor;  $L_{\text{bias}}, L_{\text{d}}$ : “infinite” inductor.

3. The variation of input impedance before and after a degeneration inductor  $L_{\text{degen}}$  is added to the device is

(a) *MOSFET*

$$\begin{aligned} Z_{\text{in}} &= j(L_g + L_{\text{degen}})\omega + \frac{1}{jC_{\text{gs}}\omega} + \frac{g_m}{C_{\text{gs}}} L_{\text{degen}} \\ &= j \left[ (L_g + L_{\text{degen}})\omega - \frac{1}{C_{\text{gs}}\omega} \right] + L_{\text{degen}}\omega_T \end{aligned}$$

$$\begin{aligned} Z_{\text{in}} &= j(L_g + L_{\text{degen}})\omega + \frac{1}{jC_{\text{gs}}\omega} + \frac{g_m}{C_{\text{gs}}}L_{\text{degen}} \\ &= j \left[ (L_g + L_{\text{degen}}\omega) - \frac{1}{C_{\text{gs}}\omega} \right] + L_{\text{degen}}\omega_T \end{aligned}$$

(b) *Bipolar*

$$\begin{aligned} Z_{\text{in}} &= r_b + r_\pi + \frac{L_{\text{degen}}C_\pi\omega^2g_m r_\pi^2}{1 + (C_\pi\omega r_\pi)^2} + j \left[ 1 + \frac{\left(g_m - \frac{C_\pi}{L_{\text{degen}}}r_\pi\right)r_\pi}{1 + (C_\pi\omega r_\pi)^2} \right] L_{\text{degen}}\omega \\ Z_{\text{in}} &= r_b + r_\pi + \frac{L_{\text{degen}}C_\pi\omega^2g_m r_\pi^2}{1 + (C_\pi\omega r_\pi)^2} + j \left[ 1 + \frac{\left(g_m - \frac{C_\pi}{L_{\text{degen}}}r_\pi\right)r_\pi}{1 + (C_\pi\omega r_\pi)^2} \right] L_{\text{degen}}\omega_T \end{aligned}$$

4. The key criterion to get  $G_{\max}$  and  $\text{NF}_{\min}$  simultaneously in the LNA design is

$$S_{11}^* = \Gamma_{S,\text{opt}}.$$

5. List the possible ways to force the raw device into satisfying the condition  $S_{11}^* = \Gamma_{S,\text{opt}}$ .

- (a) *Increase or Decrease of Current Drain,  $I_D$* . The  $S$  parameters as well as the values of  $\Gamma_{S,\text{opt}}$  will be changed as the current drain is varied. The condition  $\Gamma_{S,\text{opt}} = S_{11}^*$  could be reached at an appropriate amount of current drain.
- (b) *Change of Device Size*. The  $S$  parameters as well as the values of  $\Gamma_{S,\text{opt}}$  will be changed as the device size is varied. The condition  $\Gamma_{S,\text{opt}} = S_{11}^*$  could be reached at an appropriate device size. Of course, this scheme is only possible to the IC designer but not to the designer who implements the circuits by discrete parts.
- (c) *Addition of Degeneration Part*. Up to the design experience in the practical design, this is an easy way to get success, and so on.

6. In the Smith Chart, gain circles are a family of circles with a constant  $G$  (gain) value. The difference of  $G$  values between two adjacent circles is the step of  $G$  value's variation. The "focused center" of the gain circles is maximum of  $G$ . In the Smith Chart, noise figure circles are a family of circles with a constant NF (noise figure) value. The difference of NF values between two adjacent circles is the step of NF value's variation. The "focused center" of the noise figure circles is minimum of NF.

7. The optimized  $\Gamma_{S,\text{opt}}$  for a raw device means that the minimum of noise figure can be realized if

$$\Gamma_S = \Gamma_{S,\text{opt}}$$

$$\Gamma_S = \Gamma_{S,\text{opt}}.$$

8. The key step in the new design scheme for LNA suggested by Richard Li is the raw device testing.

9. The main features of Haus' theory are as follows:
- A minimum of noise figure,  $NF_{min}$ , exists in a noisy block with two ports, when an optimized source admittance,  $Y_{S,opt}$ , or an optimized source conductance and susceptance,  $G_{S,opt}$  and  $B_{S,opt}$ , or an optimized source voltage reflection coefficient,  $\Gamma_{S,opt}$ , is reached.
  - The minimum of noise figure,  $NF_{min}$ , the equivalent noise resistor,  $R_N$ , and the optimized source voltage reflection coefficient,  $\Gamma_{S,opt}$ , depend on the type of device, the size of the device, other electromagnetic parameters of the device, and the current drain flowing through the device.
  - The noise figure seems to be not related to the output of the noisy block.
  - The source impedance can be any value and is not restricted at  $50 \Omega$ .
10. The noise figure of an LNA depends mainly on its source but not on its load, because the noise figure is to characterize the additional noise produced by the noisy block LNA itself. This is why the noise figure is basically independent of the out put side or load side.
11. If the condition  $S_{11}^* = \Gamma_{S,opt}$  is not satisfied from raw device testing, one should work out for its satisfaction by some ways. It must be pointed out that this condition cannot be approached through the input impedance matching work because both  $S_{11}$  and  $S_{11}^*$  would be “pulled” to  $50 \Omega$  simultaneously, whereas  $\Gamma_{S,opt}$  would be moved to somewhere after the input impedance matching network is attached to the raw device if  $S_{11}^* \neq \Gamma_{S,opt}$ .
12. The main purpose of a cascode LNA is to reduce the input Miller capacitance of the raw device and to enhance the isolation between input and output of LNA.
13. Describe the  $K$  factor and  $\mu$  factor for the stability of LNA.

$K$ factor	$\mu$ factor
$K = 1 -  S_{11} ^2 -  S_{22} ^2 +  \Delta ^2$	
$ \Delta  =  S_{11}S_{22} - S_{12}S_{21} $	$\mu = \frac{1 - [\text{mag}(S_{11})]^2}{\text{mag}[S_{22} - \Delta \text{conj}(S_{11})] + \text{mag}(S_{21}S_{12})}$
Criteria of stability	Criterion of stability
$K > 1$	$\mu > 1$
$ \Delta  < 1$	

14. AGC means automatic gain control. A receiver needs LNA with AGC function because in a communication system with CDMA technology, the field strength around the antenna could be varied up to 90 dB. Without AGC function available in a receiver, the communication system with CDMA technology would be out of work.

## 18.1 INTRODUCTION

There are two kinds of mixers used to convert a signal from an RF to an IF in a communication system, that is, the active mixer and the passive mixer. Both have been developed and applied for several decades, and either can be used in a communication system in the frequency range of VHF, UHF, and GHz. However, in the microwave frequency range, it is more realistic to apply the passive mixer in a communication system.

Figures 18.1 and 18.2 show two typical active and passive mixers, respectively. The main device applied to an active mixer is the MOSFET (metal–oxide–semiconductor field-effect transistor) or bipolar transistor, whereas the main device applied to a passive mixer is a matched quad-diode.

The active mixer shown in Figure 18.1 is quite simple. It contains only one MOSFET. The RF signal is fed to a self-coupling transformer and then goes to the gate of the MOSFET. The self-coupling transformer or, say, an inductor with central tape, functions as an impedance matching part because the input impedance at the gate of the MOSFET is usually quite high. The LO injection is fed to the source of the MOSFET. At the drain of the MOSFET, an inductor and three capacitors form an output impedance matching network.

The core of the passive mixer shown in Figure 18.2 is the quad-diode. The four diodes are usually packaged together. They are, of course, expected to be as identical as possible, because the performance of the mixer largely depends on the uniformity of

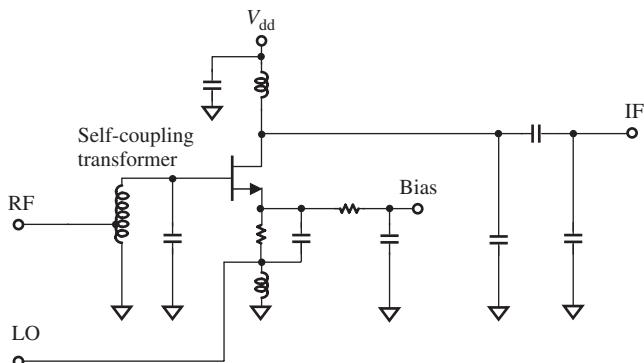


Figure 18.1. A typical MOSFET active mixer.

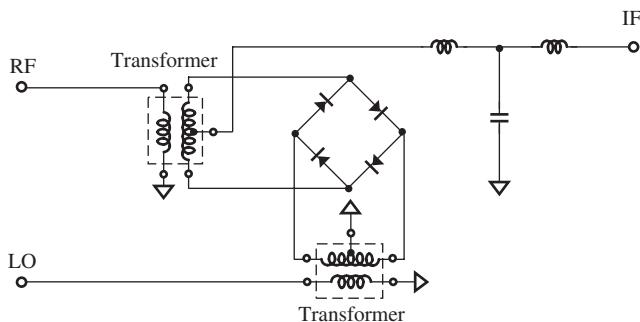


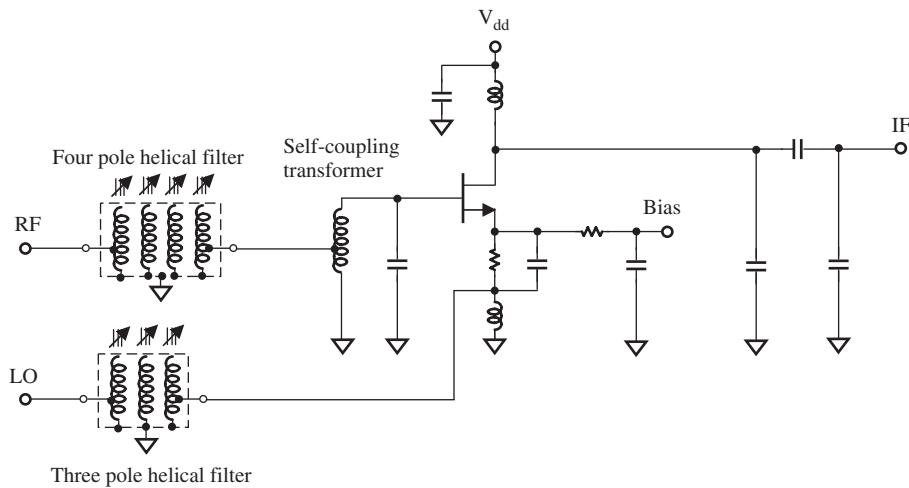
Figure 18.2. A typical quad-diodes passive mixer.

these four diodes. Both the RF signal and LO injection are fed to a transformer first and then go to the quad-diode. This is due to the differential requirement from the quad-diode: the input RF signal and LO injection into the quad-diode must be as perfectly differential as possible. Two transformers function as a balun to transfer the single-ended input into a differential pair output. In addition, they function as an input impedance matching network. The central tap of the RF transformer is the IF output. The *T*-type output impedance matching network consists of two inductors and one capacitor.

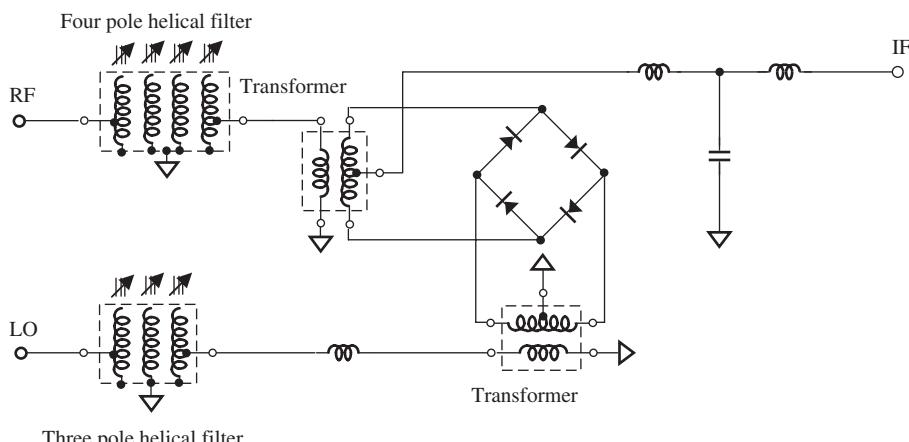
In a dual-conversion portable radio or communication system, the BPF (band pass filter) is always set in the path of the input RF signal and in the path of the LO injection before they get into the mixer, so that the spurious products can be considerably reduced. The helical filter has been applied for such a purpose. Figures 18.3 and 18.4 show the active and passive mixer with BPF helical filters, respectively.

The helical filters shown in Figures 18.3 and 18.4 are the adjustable BPFs. They are filters with a cavity-type configuration, constructed as a casting aluminum box. There are a couple of small closets in the box. The helical coil is circled around a cylindrical plastic frame and then put into each small closet in the box. A ferrite core is inserted into each cylindrical frame and can be adjusted up and down along the cylindrical axis. A four-pole helical filter indicates that there are four small closets with four helical coils contained in the metallic box; a three-pole helical filter indicates three small closets with three helical coils contained in a metallic box, and so on. In the UHF range, the size of a helical filter is about  $L \times W \times H = 2.0 \times 0.5 \times 1.5$  cm for a four-pole helical filter and about  $L \times W \times H = 1.5 \times 0.5 \times 1.5$  cm for a three-pole helical filter.

At present, for the considerations of cost and size, most helical filters have been replaced by other filters, such as the SAW (surface acoustic wave) filter. However, the



**Figure 18.3.** A typical MOSFET active mixer with BPF helical filter.



**Figure 18.4.** A typical quad-diode passive mixer with BPF helical filter.

performance of the helical filter is superior to other filters in terms of a high  $Q$  value, low-insertion loss, and wide bandwidth.

It is interesting, of course, to compare respective advantages and disadvantages of mixers in order to decide which type of mixer to use. Table 18.1 lists the main items for comparison on the basis of Figures 18.1 and 18.2.

With respect to the active mixer, the advantages of a passive mixer are no current drain, low-noise figure, wide bandwidth, and lower part count (and hence high reliability); the disadvantages are a higher required LO injection, negative conversion gain, and high cost. It should be noted that in a passive mixer, two transformers with a central tap must be provided. Together with a matched quad-diode device, symmetry becomes a key factor in their performance.

At present, by means of RFIC technology, the resistive mixer is being developed as a new configuration of a passive mixer, whereas the Gilbert cell is developed as the core of the active mixer. These are discussed in the following sections.

TABLE 18.1. Comparison between Active and Passive Mixer

Item	Active Mixer	Passive Mixer	Unit
Current drain	~2–5	~0	mA
LO injection	~-5–0	~5–10	dB <sub>m</sub>
Conversion gain	~5–10	~-5 to -3	dB
Noise figure	~10–15	~3–5	dB
Bandwidth	Narrower	Wider	—
Part count	~13	~6	—
Reliability	Lower	Higher	—
Cost	Lower	Higher	—

Note: The values listed in Table 18.1 are approximate.

## 18.2 PASSIVE MIXER

A passive mixer implies that the current drain in the mixer circuitry is zero.

### 18.2.1 Simplest Passive Mixer

The simplest passive mixer can be implemented by a diode and a compound transformer as shown in Figure 18.5.

The operating principle of this mixer is based on the even-order nonlinearity of the diode, that is,

$$i = a_0 + a_1(k_r v_{RF} + k_l v_{LO}) + a_2(k_r v_{RF} + k_l v_{LO})^2 + a_3(k_r v_{RF} + k_l v_{LO})^3 \\ + a_4(k_r v_{RF} + k_l v_{LO})^4 + \dots, \quad (18.1)$$

where

$i$  = RF current flowing through the diode,

$a_i$  =  $i$ th-order nonlinearity coefficient,

$k_r$  = coupling coefficient between RF and IF port of transformer,

$k_l$  = coupling coefficient between LO and IF port of transformer,

$v_{RF}$  = input voltage of RF signal, and

$v_{LO}$  = voltage of LO injection.

If

$$v_{RF} = v_{RF0} \cos(\omega_{RF}t + \varphi), \quad (18.2)$$

$$v_{LO} = v_{LO0} \cos \omega_{LO} t, \quad (18.3)$$

where

$\omega_{RF}$  = angular frequency of RF signal,

$\omega_{LO}$  = angular frequency of LO injection,

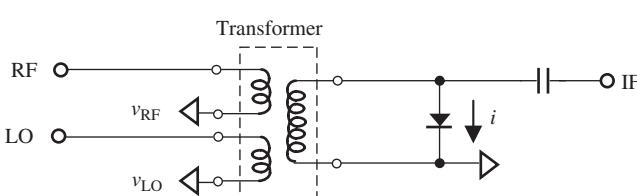


Figure 18.5. A simplest passive mixer.

$v_{RFo}$  = amplitude of RF signal, and  
 $v_{LOo}$  = amplitude of LO injection.

then the expression (18.1) becomes

$$\begin{aligned}
 i &= a_0 + a_1[k_r v_{RFo} \cos(\omega_{RF}t + \varphi) + k_l v_{LOo} \cos \omega_{LO}t] \\
 &+ a_2 \left[ (k_r v_{RFo})^2 \frac{1 + \cos 2(\omega_{RF}t + \varphi)}{2} + (k_l v_{LOo})^2 \frac{1 + \cos 2\omega_{LO}t}{2} \right] \\
 &+ a_2 [k_r k_l v_{RFo} v_{LOo} (\cos\{(\omega_{RF} + \omega_{LO})t + \varphi\} + \cos\{(\omega_{RF} - \omega_{LO})t + \varphi\})] \\
 &+ a_3 [k_r v_{RFo} \cos(\omega_{RF}t + \varphi) + k_l v_{LOo} \cos \omega_{LO}t]^3 \\
 &+ a_4 [k_r v_{RFo} \cos(\omega_{RF}t + \varphi) + k_l v_{LOo} \cos \omega_{LO}t]^4 + \dots
 \end{aligned} \tag{18.4}$$

It should be noted that in expression (18.4), one of the terms with second-order nonlinearity in the low-frequency range is

$$k_r k_l v_{RF} v_{LO} \cos\{(\omega_{RF} - \omega_{LO})t + \varphi\} = k_r k_l v_{RF} v_{LO} \cos\{\omega_{IF}t + \varphi\}, \tag{18.5}$$

where  $\omega_{IF}$  is the angular frequency of IF signal, and

$$\omega_{RF} - \omega_{LO} = \omega_{IF}. \tag{18.6}$$

Expressions (18.4) to (18.6) indicate that the terms with second-order nonlinearity produce the IF components. This is the operational principle of a mixer.

### 18.2.2 Double-Balanced Quad-Diode Mixer

Figure 18.6 shows a typical double-balanced quad-diode mixer, in which the main part is a quad-diode or ring diode. The transformer at the RF port couples the RF signal to the ring diodes, whereas the transformer at the LO port couples the LO injection to the ring diodes.

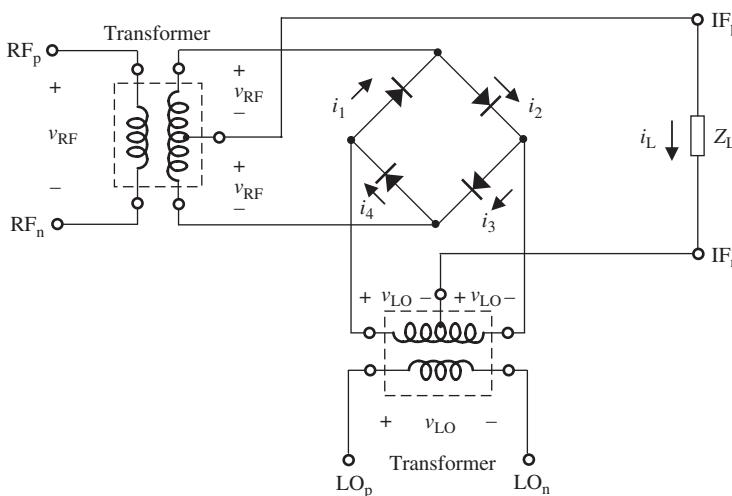


Figure 18.6. A typical double-balanced quad-diode mixer with ring configuration.

When the RF signal voltage and LO injection are coupled to the quad-diodes, the corresponding currents go through all the PN junctions with different paths in the ring configuration. Owing to the nonlinearity of the diodes, these currents interacted with each other.

There are four subcurrents,  $i_1, i_2, i_3$ , and  $i_4$ , flowing through the diodes and the load,  $Z_L$ , with different respective paths. They are determined by the voltages from the RF signal and LO injection,  $v_{RF}$  and  $v_{LO}$ , on the individual loop containing the load  $Z_L$ . They are

$$\begin{aligned} i_1 = & a_0 + a_1(-v_{RF} + v_{LO}) + a_2(-v_{RF} + v_{LO})^2 \\ & + a_3(-v_{RF} + v_{LO})^3 + a_4(-v_{RF} + v_{LO})^4 + \dots, \end{aligned} \quad (18.7)$$

$$\begin{aligned} i_2 = & a_0 + a_1(v_{RF} + v_{LO}) + a_2(v_{RF} + v_{LO})^2 \\ & + a_3(v_{RF} + v_{LO})^3 + a_4(v_{RF} + v_{LO})^4 + \dots, \end{aligned} \quad (18.8)$$

$$\begin{aligned} i_3 = & a_0 + a_1(v_{RF} - v_{LO}) + a_2(v_{RF} - v_{LO})^2 \\ & + a_3(v_{RF} - v_{LO})^3 + a_4(v_{RF} - v_{LO})^4 + \dots, \end{aligned} \quad (18.9)$$

$$\begin{aligned} i_4 = & a_0 + a_1(-v_{RF} - v_{LO}) + a_2(-v_{RF} - v_{LO})^2 \\ & + a_3(-v_{RF} - v_{LO})^3 + a_4(-v_{RF} - v_{LO})^4 + \dots, \end{aligned} \quad (18.10)$$

where  $a_k$  is the  $k$ th-order nonlinearity coefficient of diode.

Finally, the current flow through the load,  $i_L$ , consists of four subcurrents,  $i_1, i_2, i_3$ , and  $i_4$ , that is,

$$i_L = (i_1 - i_2) - (i_3 - i_4) = -a_2 8v_{RF}v_{LO} - a_4 16(v_{RF}^3v_{LO} + v_{RF}v_{LO}^3) + \dots \quad (18.11)$$

Expression (18.11) implies that the odd-order nonlinearity terms are cancelled by each other. The spurious products can be produced only by the even-order nonlinearity terms.

Assuming that

$$v_{RF} = v_{RF0} \cos(\omega_{RF}t + \varphi), \quad (18.12)$$

$$v_{LO} = v_{LO0} \cos \omega_{LO}t, \quad (18.13)$$

then,

$$\begin{aligned} i_L = & -a_2 4v_{RF0}v_{LO0}[\cos\{(\omega_{RF} + \omega_{LO})t + \varphi\} + \cos\{(\omega_{RF} - \omega_{LO})t + \varphi\}] \\ & - a_4 8v_{RF0}v_{LO0}[\cos\{(\omega_{RF} + \omega_{LO})t + \varphi\} + \cos\{(\omega_{RF} - \omega_{LO})t + \varphi\}], \\ & \times [v_{RF0}^2\{1 + \cos 2(\omega_{RF}t + \varphi)\} + v_{LO0}^2\{1 + \cos 2\omega_{LO}t\}] - \dots \end{aligned} \quad (18.14)$$

or

$$\begin{aligned} i_L = & -[a_2 4v_{RF0}v_{LO0} + a_4 8v_{RF0}v_{LO0}(v_{RF0}^2 + v_{LO0}^2) + \dots] \cos\{\omega_{IF}t + \varphi\} \\ & - [a_2 4v_{RF0}v_{LO0} + a_4 8v_{RF0}v_{LO0}(v_{RF0}^2 + v_{LO0}^2) + \dots] \cos\{(\omega_{RF} + \omega_{LO})t + \varphi\} - \dots \end{aligned} \quad (18.15)$$

In the right side of expression (18.15), the first term is the desired IF signal and will be preserved, while the other terms on the right side are high-frequency components and

will be filtered out. Consequently, from expression (18.15), the IF current,  $i_{IF}$ , on the load,  $Z_L$ , is

$$i_{IF} = -[a_2 4v_{RFo} v_{LOo} + a_4 8v_{RFo} v_{LOo}(v_{RFo}^2 + v_{LOo}^2) + \dots] \cos\{\omega_{IF}t + \varphi\}. \quad (18.16)$$

It can be seen that the second-order nonlinearity of the diode is the main part or key contribution in producing the IF signal. Other even-order nonlinearities contribute just a few percentage to the IF signal but produce a lot of spurious products.

Also, from expression (18.16) it can be seen that the performance of a double-balanced mixer with ring quad-diodes depends on the following:

- The nonlinearity of diodes, such as  $a_2, a_4, a_6, \dots$
- LO injection,  $v_{LO}$ ,
- RF signal itself,  $v_{RFo}$ , and
- Frequency,  $\omega$ .

Its typical performance, for example, is

- LO injection: 5 dB<sub>m</sub>
- Conversion gain: -4.0–4.5 dB,
- Noise figure: 4.0–4.5 dB,
- IIP<sub>3</sub>: -5 to -2 dB<sub>m</sub>, and
- Relative bandwidth: 10%.

In the quad-diode mixer, instead of conversion gain, there is always conversion loss. The conversion loss depends on many factors, such as the LO injection, the nonlinearity of device, the input and output impedance matching status, the  $Q$  values of the parts in the mixer, and so on. It should be especially noted that the conversion loss is dependent on the LO injection level. Figure 18.7 shows an example of this relationship.

The conversion loss is high when the LO injection is low. It is decreased when the LO injection is increased. Usually, a passive mixer begins to reach its normal operation state only after the LO injection is increased to over 3 dB<sub>m</sub>.

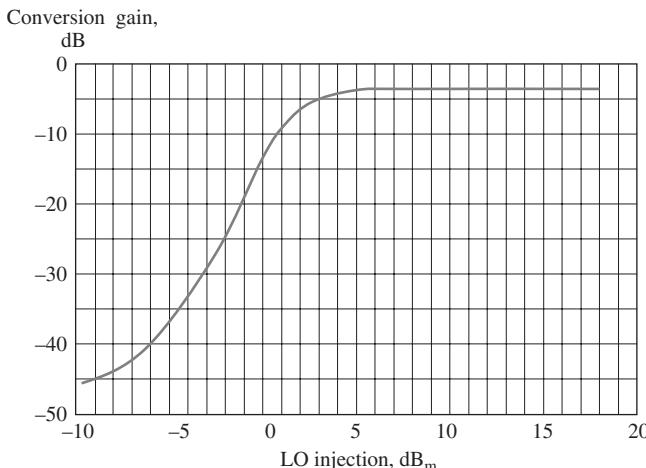
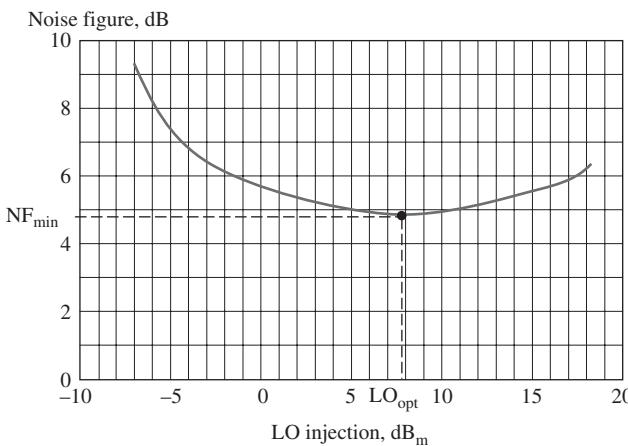


Figure 18.7. An example of the relationship between conversion gain and LO injection in a passive mixer.



**Figure 18.8.** An example of the relationship between conversion gain and noise figure in a passive mixer.

For a passive RF block, the noise figure is reasonably equivalent to the conversion loss. However, in the quad-ring mixer, a remarkable feature of a passive mixer is that its noise figure depends on the LO injection. Figure 18.8 shows that there is a minimum of the noise figure as the LO injection is varied. As shown in Figure 18.8, the minimum noise figure is 4.8 dB when the LO injection is 7.8 dB<sub>m</sub>.

The nonlinearity is mainly determined by the expressions of even orders of nonlinearity of quad-diodes except the second-order of nonlinearity.

It should be noted that in the expressions from (18.7) to (18.10), the nonlinearity characteristics of each diode in the quad-diode are assumed to be perfectly identical, so that their nonlinearity coefficients are the same. Obviously, imperfect uniformity among the quad-diodes brings about more spurious problems, in which the odd-order nonlinearity of the diodes does not cancel with each other. Then, the linearity of the mixer would be degraded.

In addition, it is also assumed that the central tap of transformers at either the RF or LO port is at a perfect symmetrical position so that the voltage at the secondary winding of transformer is exactly divided by two identical voltages,  $v_{RF} + v_{LO}$  or  $v_{LO} + v_{RF}$ . Should the perfectly symmetrical condition of the transformers be unsatisfied, the final expressions of the current flowing on the load or IF current, (18.15) and (18.16), would be much complicated, depending on how much asymmetry is present.

### 18.2.3 Double-Balanced Resistive Mixer

Figure 18.9 shows the core of a double-balanced resistive mixer. Figure 18.10 is exactly the same as Figure 18.9, but is drawn in a different style.

The capacitors  $C_1 - C_6$  block all DC so that they are “zero” capacitors in the operating frequency range. The special feature of a resistive mixer is that there is no DC flow through it.

At first glance, the resistive mixer looks like a ring diode mixer. However, this is not true because it is not a real ring but a circuit with a double-balanced configuration. Essentially, it is a Gilbert cell without DC power supply and bias.

This mixer’s basic operating principle is that the RF signal current flows through the resistive channels under the gate of the double-balanced transistors and are “modulated” by, or interact with, the LO injection voltage at the gates of the double-balanced transistors.

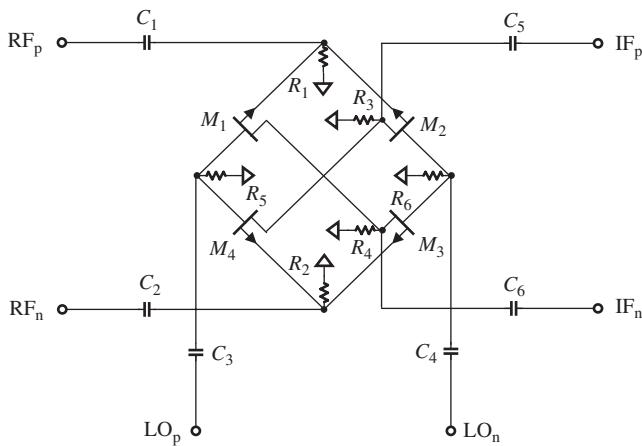


Figure 18.9. Core of a passive resistive mixer.

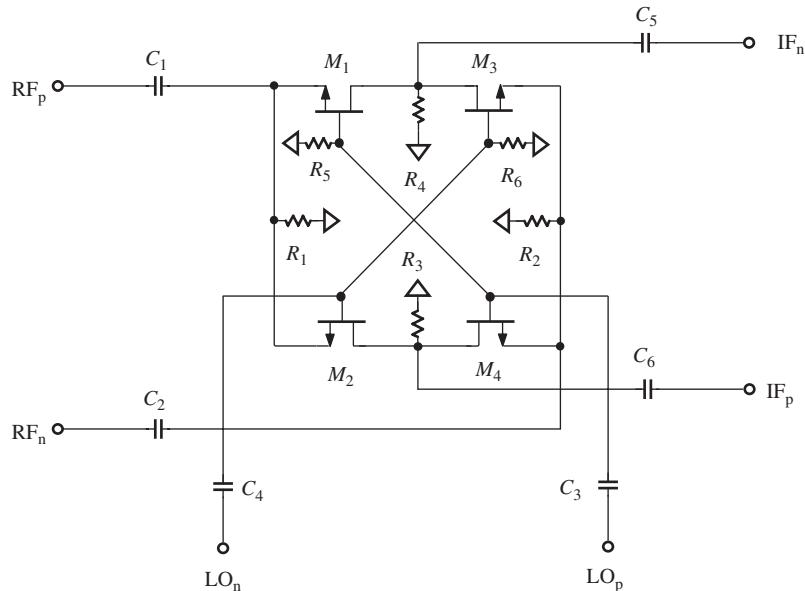


Figure 18.10. Alternative drawing of a passive resistive mixer core.

The differences between the quad-diode and resistive mixer can be found from their operating principles, that is,

- in the quad-diode mixer, the current of RF signal flows through the PN junction of the diode, whereas in the resistive mixer, the current of RF signal flows through the resistive channel of the MOSFET under the gate;
  - in the quad-diode mixer, the LO injection current flows through the PN junction of the diode, just like the RF signal current, whereas in the resistive mixer, the LO injection appears as a control voltage to control the resistive channel of the MOSFET;

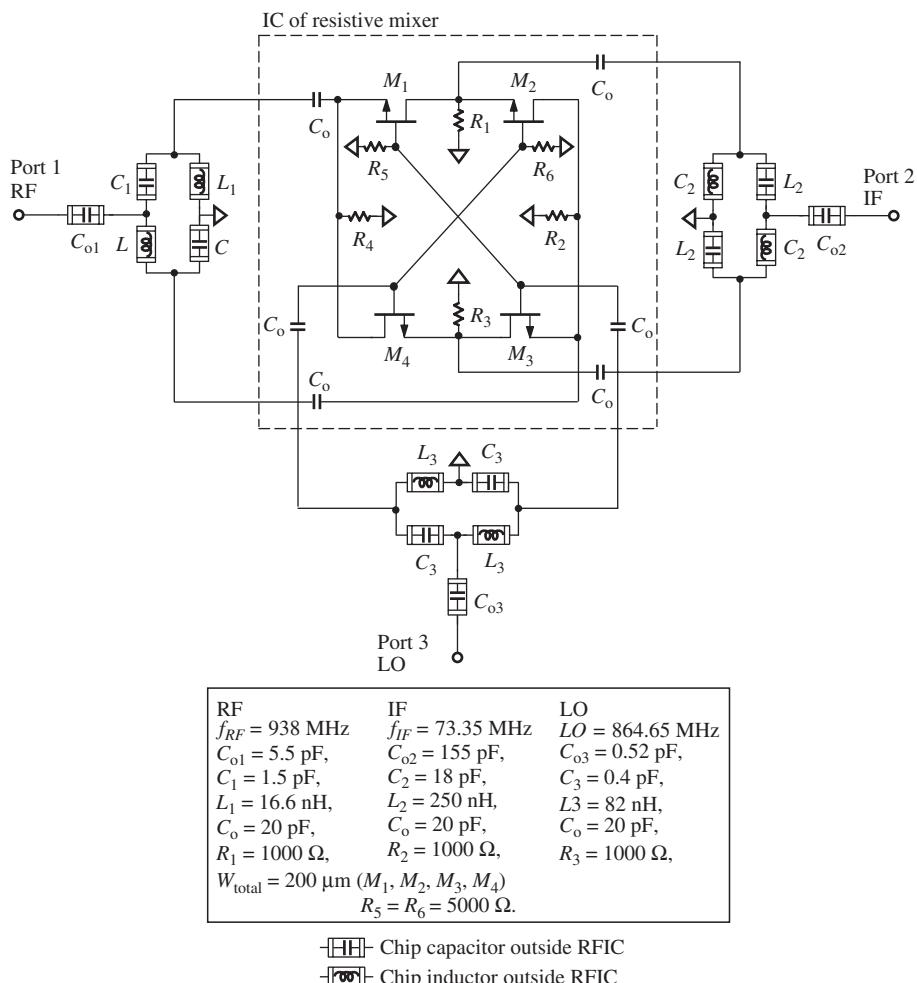


Figure 18.11. An 800–900-MHz MOSFET resistive mixer.

- in the quad-diode mixer, the IF signal is a product resulting from the interaction of the RF signal current and LO injection current flowing through the PN junction together, whereas in the resistive mixer, the IF signal is a product resulting from the RF signal current flowing through the resistive channel and modulated by the LO injection control voltage at the gate of the MOSFET.

In summary, the same IF signal is produced by both, but the mechanism of production is different.

The following is a design example for 800–900-MHz radios. Figure 18.11 is the schematic of the resistive mixer with three simple LC baluns for the RF input signal, LO injection, and IF output signal.

The core of the resistive mixer is designed for the RFIC chip. The three baluns for the RF, LO, and IF ports are designed with discrete parts, chip capacitors, and chip inductors. The design of the balun is described in chapter 15 in this book, and the values of their parts are listed in Figure 18.11.

The main features of this resistive mixer are

- $P_{LO} = -5 \text{ dB}_m$ ,
- $CG = -7.7 \text{ dB}$ ,
- $IIP_3 = 2.4 \text{ dB}_m$ , and
- $NF = 4.1 \text{ dB}$ .

By comparing its performance with that of the quad-diode mixer introduced above, the apparent differences are the following:

- The required LO injection in resistive mixer is approximately  $-5 \text{ dB}_m$ , while it is  $+5 \text{ dB}_m$  in the quad-diode mixer. This is a big difference. The quad-diode mixer needs a more powerful LO injection because it directly promotes the LO injection current flowing through the diodes. The resistive mixer needs a low LO injection because the LO injection is only applied for modulating the RF signal. In the resistive mixer, the LO injection is a control voltage appearing on the gate of the MOSFET. So far, the resistive mixer is much superior to the quad-diode mixer.
- The resistive mixer has better linearity than the quad-diode mixer: the  $IIP_3$  in the resistive mixer is  $2.4 \text{ dB}_m$ , while it is  $-5$  to  $-2 \text{ dB}_m$  in the quad-diode mixer. The reason for this is that in the resistive mixer, the resistive channel under the gate of transistor is much more linear than a diode's character. This is another advantage that the resistive mixer holds.
- The main drawback of the resistive mixer is its conversion gain. Its conversion gain in the design example for 800–900-MHz radios is  $-7.7 \text{ dB}$ , while the quad-diode's conversion gain is  $-4.0$  to  $-4.5 \text{ dB}$ .

Is it necessary to take a trade-off between the quad-diode mixer and resistive mixer because of the advantages and disadvantages of the above comparison?

The best solution, of course, is to preserve all the advantages and remove all the disadvantages. Figure 18.12 shows an alternative way to overcome the drawback of the resistive mixer. The new mixer consists of two blocks: the first block is a resistive mixer and the second block is an IF amplifier. Let us call it a compound resistive mixer.

Figure 18.13 is the desired IF amplifier. It consists of a MOSFET, chip capacitors, and chip inductors. It has a  $T$ -type input impedance matching network built using two capacitors,  $C_1$  and  $C_2$ , and one inductor,  $L_1$ . The output impedance matching network consists of only the inductor  $L_2$  and a capacitor  $C_3$ . Their values are listed in Figure 18.13.

The performance of this IF amplifier is

- $G = 16.6 \text{ dB}$ ,
- $RL_{in} = -27 \text{ dB}$ ,
- $RL_{out} = -28 \text{ dB}$ ,

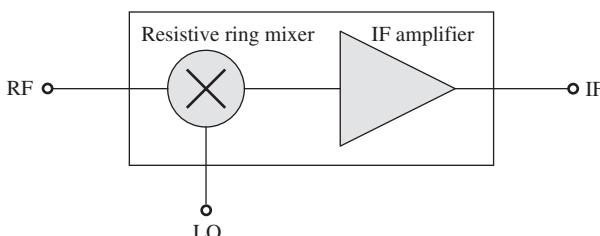


Figure 18.12. A compound resistive mixer built using a resistive mixer and an IF amplifier connected together in series.

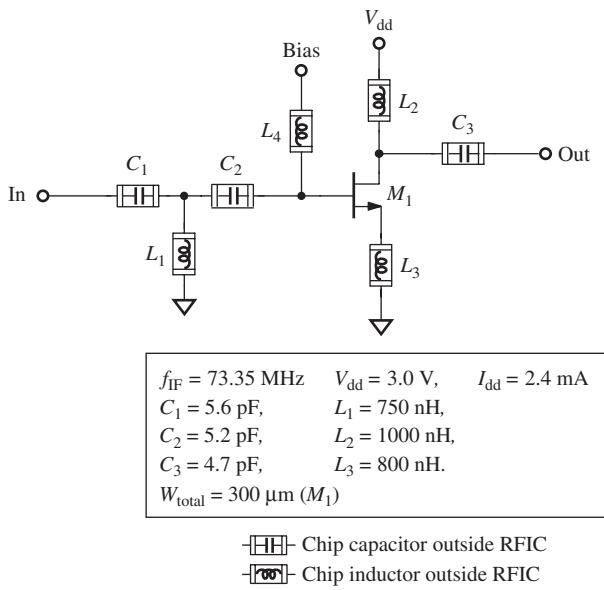


Figure 18.13. IF amplifier after 800–900-MHz MOSFET resistive mixer.

- $\mu = 1.16$ ,
- $\text{IIP}_3 = 4.4 \text{ dB}_m$ , and
- $\text{NF} = 0.76 \text{ dB}$ .

The performance of the full compound mixer is

- $V_{\text{dd}} = 3.0 \text{ V}$ ,  $I_{\text{dd}} = 2.4 \text{ mA}$ ,
- $f_{\text{RF}} = 938 \text{ MHz}$ ,  $P_{\text{RF}} = -40 \text{ dB}_m$ ,
- $f_{\text{LO}} = 864.65 \text{ MHz}$ ,  $P_{\text{LO}} = -5 \text{ dB}_m$ ,
- $f_{\text{IF}} = 73.35 \text{ MHz}$ ,
- $\text{CG} = 8.9 \text{ dB}$ ,
- $\text{IIP}_3 = 7.1 \text{ dB}_m$ , and
- $\text{NF} = 5.8 \text{ dB}$ .

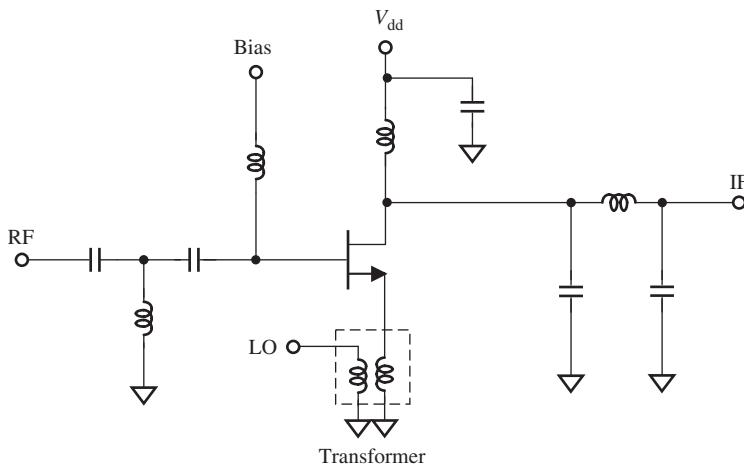
## 18.3 ACTIVE MIXER

### 18.3.1 Single-End Single Device Active Mixer

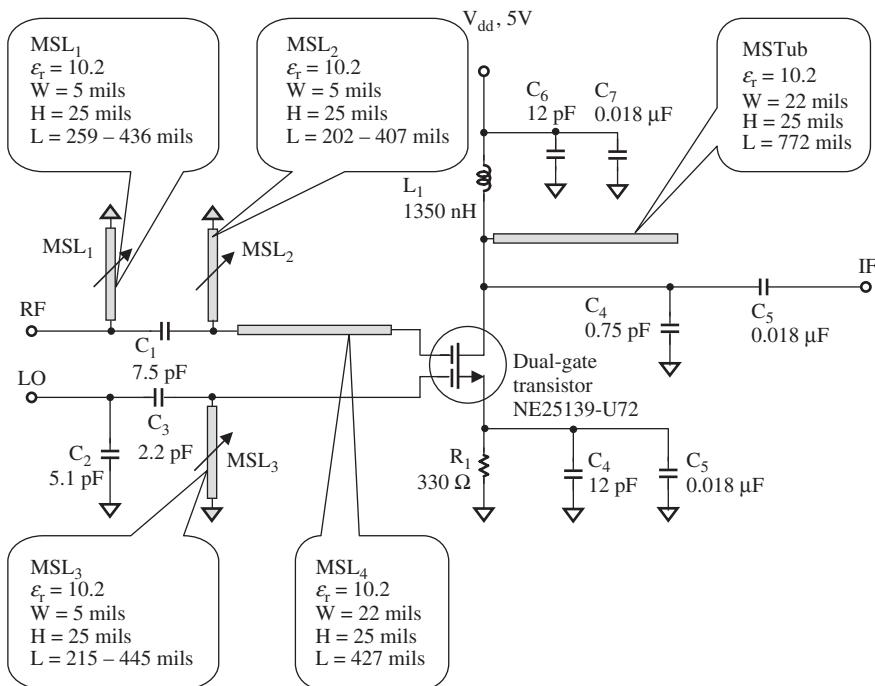
Figure 18.14 shows a simple active mixer built using a single MOSFET device.

The RF signal is matched to the gate of the MOSFET, and the LO injection goes to the source of the MOSFET through a transformer. The output IF signal is coupled from the drain of MOSFET through a  $\Pi$ -type impedance matching network. The expensive part of this mixer, of course, is the transformer in either the RFIC or module configuration with discrete parts.

Figure 18.15 shows a simple active mixer built using a dual-gate MOSFET device. It is still a mixer built using a single device, but it is a special device. The RF signal is matched to the first gate, and the LO injection goes to the second gate of the MOSFET. The output IF signal is coupled from the drain of the MOSFET through a  $\Pi$ -type impedance matching network. Its operating frequency range is around 1.5 GHz.



**Figure 18.14.** A simple active mixer built using a single device.



**Figure 18.15.** A simple active mixer built using a dual-gate GaAs device (except the dual-gate device and microstrip lines, all discrete parts in the schematic are chip parts).

- $V_{dd} = 5.0 \text{ V}$ ,  $I_{dd} = 2.9 \text{ mA}$ ,
- $f_{RF} = 1513\text{--}1525 \text{ MHz}$ ,  $P_{RF} = -40 \text{ dB}_m$ ,
- $f_{LO} = 1459.45\text{--}1471.45 \text{ MHz}$ ,  $P_{LO} = -4.5 \text{ dB}_m$ , and
- $f_{IF} = 53.55 \text{ MHz}$ .

In order to lower the cost, it is intentional to apply microstrip lines as much as possible in this design. Furthermore, in order to shrink the size of the module, the substrate is not a plastic PCB (printed circuit board), but an aluminum ceramic PCB

with a high permittivity,  $\epsilon_r = 10.7$ . On the other hand, all the microstrip lines in the layout have a “worm” shape. The adjustable microstrip lines, MSL<sub>1</sub>, MSL<sub>2</sub>, and MSL<sub>3</sub>, are trimmed by a laser trimmer so that the input impedances at the RF and LO port can be well matched from  $50 \Omega$  to the dual-gate device.

There is some trouble in the isolation between the input LO and input RF signal and between the input LO and output IF signal. The first isolation problem between the input LO and RF signal is imaginable since the two gates are crowded in a common device. In the layout, the two input impedance matching networks built using microstrip lines are kept separated as far as possible. This does not cause any big trouble. However, the second isolation problem between the input LO and output IF signal stubbornly persists. It was found that the LO signal “penetrates” the device and appears at the drain of the device without being attenuated enough from the original LO injection level. It was therefore decided to apply a microstrip line stub at the drain of the device. The electrical length of the microstrip stub is a quarter wavelength of the LO injection frequency. Its one end is connected to the drain of the device, and its other end is open circuited. Consequently, this quarter wavelength open-circuited microstrip line stub forces the drain of device to be grounded at the LO injection frequency so that it is impossible for the LO injection to establish its voltage there. The isolation problem is thus solved in a very satisfying way. The LO injection signal at the drain of the device is lowered by 30 dB before and after the open-circuited microstrip line stub with a quarter wavelength is applied.

The final performance from the actual testing is

- CG = 8.0 dB,
- RL<sub>RF</sub> = -11 dB,
- RL<sub>LO</sub> = -8 dB,
- IIP<sub>2</sub> = 3.0 dB<sub>m</sub>,
- IIP<sub>3</sub> = -8.0 dB<sub>m</sub>, and
- NF = 8.0 dB.

### 18.3.2 Gilbert Cell

Figure 18.16 shows a typical MOSFET Gilbert cell.

As shown in Figure 18.16, the basic element of a Gilbert cell is two differential pairs on the top and a differential pair in the bottom. The two differential pairs on the top form a double-balanced configuration. The outputs are a cross combination of the drains of two differential pairs on the top. The current drain is controlled by the tail current  $I_{EE}$ .

Figure 18.17 is exactly the same as Figure 18.16, but it is drawn like a passive mixer built using a ring or quad-diode as shown in Figure 18.6. However, the Gilbert cell is not a cell with a “ring” configuration, but is a cell with a dual-balanced configuration. As a matter of fact, the Gilbert cell is a cell that is not only “double balanced” but also “double differential.”

Theoretically, the apparent advantages of an ideal differential pair are

- The odd orders of nonlinearity of the devices produce IF product and a number of spurious products while the even orders of nonlinearity of the devices are cancelled with each other;
- an ideal differential pair has the capability to reject the common mode components of the incoming signal.

The Gilbert cell has all these features since its configuration is double differential.

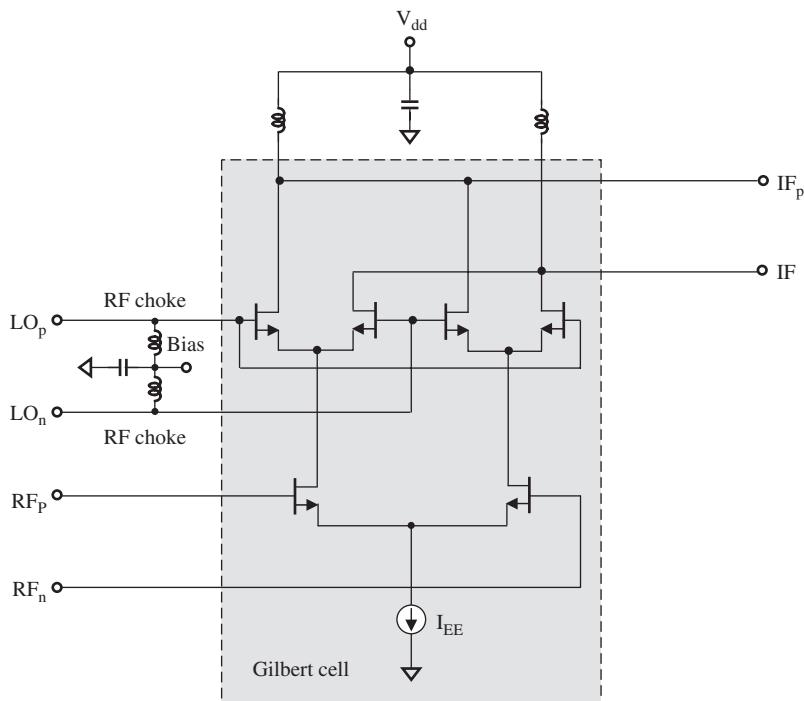


Figure 18.16. An active mixer with MOSFET Gilbert cell.

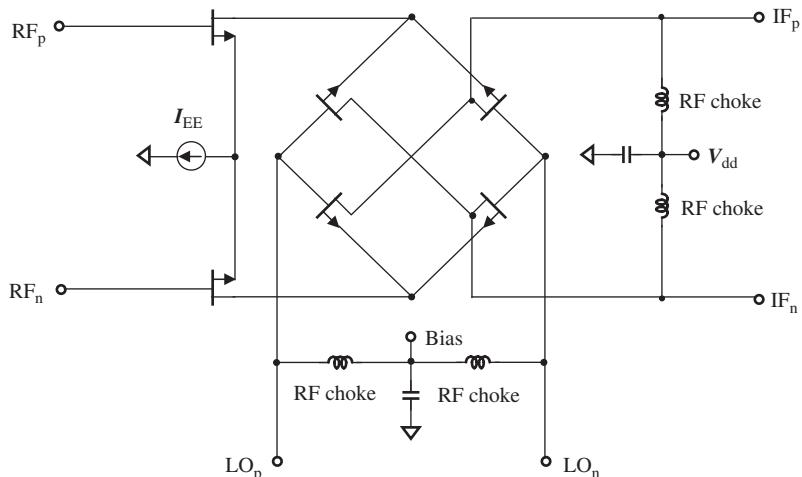
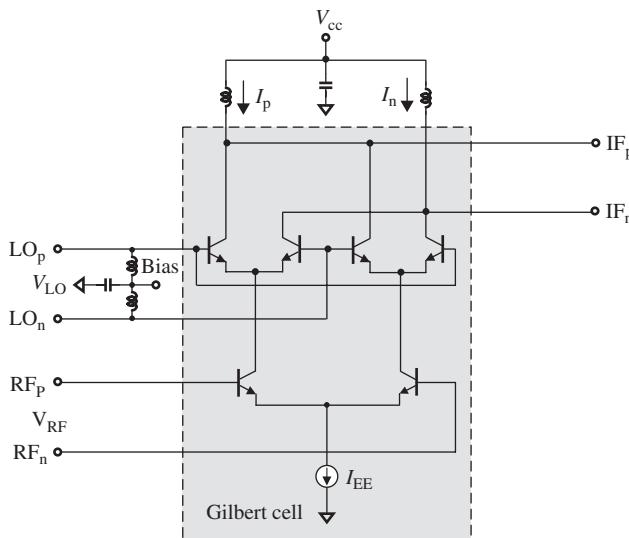


Figure 18.17. Alternative drawing of an active mixer with a MOSFET Gilbert cell.

From another viewpoint, that is, in comparison to a passive mixer,

- one of the outstanding features in an active mixer built using a Gilbert cell is the positive conversion gain; this could alleviate the problem of the gain required in the LNA (low-noise amplifier) block of a receiver;
- the LO injection level in an active mixer required is much lower than that of a passive mixer.



**Figure 18.18.** An active mixer with a bipolar Gilbert cell.

The disadvantages of an active mixer, as mentioned earlier, are

- its noise figure is higher than that of a passive mixer;
- an active mixer needs a DC power supply.

If the MOSFETs are simply replaced by bipolar transistors, a MOSFET Gilbert cell becomes a bipolar Gilbert cell, and the active mixer with a MOSFET Gilbert cell as shown in Figure 18.16 becomes an active mixer with a bipolar Gilbert cell as shown in Figure 18.18.

An active mixer with bipolar Gilbert cell has been well analyzed (Gray et al., 2001). In a bipolar Gilbert cell, the differential output currents can be expressed as

$$\Delta I = I_P - I_n = I_{EE} \tanh\left(\frac{v_{RF}}{2V_T}\right) \tanh\left(\frac{v_{LO}}{2V_T}\right), \quad (18.17)$$

where  $v_{LO}$  and  $v_{RF}$  are the differential voltages at the RF and LO differential port, respectively.

Note that

$$\tanh x = x - \frac{1}{3}x^3 + \frac{2}{15}x^5 - \frac{17}{315}x^7 + \frac{62}{2835}x^9 - \dots, \quad (18.18)$$

if

$$x < 1. \quad (18.19)$$

Usually  $v_{RF}$  is a small signal, that is,

$$v_{RF} \ll V_T, \quad (18.20)$$

so that expression (18.17) can be approximated as

$$\Delta I = I_{EE} \frac{v_{RF}}{2V_T} \tanh\left(\frac{v_{LO}}{2V_T}\right). \quad (18.21)$$

However,  $v_{LO}$  is usually not a small signal, but comparable with or larger than  $V_T$ , that is,

$$v_{LO} \approx \text{or} > V_T. \quad (18.22)$$

In order to linearize the output differential current  $\Delta I$ , the input differential voltage at the LO port must be predistorted so as to compensate for the hyperbolic tangent transfer characteristic. In other words, a “ $\tanh^{-1}$ ” block must be added to the LO port before the LO injection  $v_{LO}$  enters the Gilbert cell. This  $\tanh^{-1}$  block can be implemented by two diode-connected transistors based on the relationship between  $\tanh^{-1}$  and the natural logarithm.

$$\tanh^{-1} x = \frac{1}{2} \ln \left( \frac{1+x}{1-x} \right). \quad (18.23)$$

The addition of the  $\tanh^{-1}$  block could be called the linearization processing of the Gilbert cell. Implementation of the  $\tanh^{-1}$  block is introduced in Appendix 18.A.2.

Consequently, expression (18.21) becomes

$$\Delta I = I_{EE} \frac{v_{RF}}{2V_T} \frac{v_{LO}}{2V_T}. \quad (18.24)$$

If the RF signal and LO injection are sinusoidal as expressed in equations (18.2) and (18.3), that is,

$$v_{RF} = v_{RFo} \cos(\omega_{RF} t + \varphi), \quad (18.2)$$

$$v_{LO} = v_{LOo} \cos \omega_{LO} t. \quad (18.3)$$

And note that if the LO injection is a low-side LO injection,

$$\omega_{RF} - \omega_{LO} = \omega_{IF}, \quad (18.6)$$

then (18.24) becomes

$$\begin{aligned} \Delta I &= I_{EE} \frac{v_{RFo} v_{LOo}}{2V_T^2} \{ \cos[(\omega_{RF} - \omega_{LO})t + \varphi] + \cos[(\omega_{RF} + \omega_{LO})t + \varphi] \}, \\ \Delta I &= I_{EE} \frac{v_{RFo} v_{LOo}}{2V_T^2} \{ \cos(\omega_{IF} t + \varphi) + \cos[(\omega_{RF} + \omega_{LO})t + \varphi] \}. \end{aligned} \quad (18.25)$$

If the high RF component, which is the second term in expression (18.25), is filtered out, then the remaining component is the desired IF signal, that is,

$$\Delta I = I_{EE} \frac{v_{RFo} v_{LOo}}{2V_T^2} \cos(\omega_{IF} t + \varphi). \quad (18.26)$$

This is the basic principle by which the Gilbert cell operates as a mixer.

### 18.3.3 Active Mixer with Bipolar Gilbert Cell

Let us present an example of an active mixer with a bipolar Gilbert cell. Figure 18.19 is a plan to design an active mixer for an 800-MHz radio.

The main body is the active mixer with a bipolar Gilbert cell. The RF input and LO injection are single ended, and the IF output is asked to be single ended. Therefore, three baluns for the RF, LO, and IF ports must be included. Ideally, everything might be contained in an RFIC chip. However, the IF balun is hard to design as an on-chip block, because the size of the inductors and capacitors required for the low IF are unacceptable on an IC chip. It therefore is designed as an off-chip block. In addition, for considerations of cost saving, the LC-type balun is selected to implement the IF balun.

The LO and RF baluns can be put on an RFIC chip. An active balun has been developed for such a purpose. Figure 18.20 shows this schematic of an active balun.

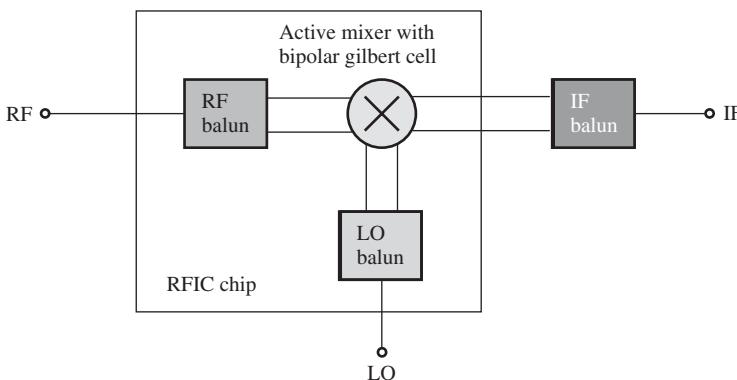


Figure 18.19. A plan to design an active mixer with a bipolar Gilbert cell.

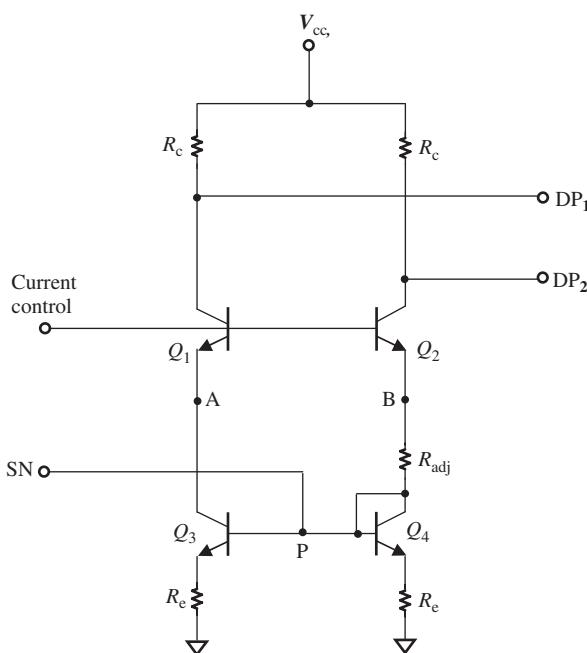


Figure 18.20. Schematic of an active balun.

In Figure 18.20, the collector and the base of  $Q_4$  are connected together. They function as a current mirror between the left and right branches of the circuitry. More importantly, they function to reverse the phase for the single-ended input signal, SN, between the left and right branches: at point B, the phase is basically the same as that in the input point P, whereas at point A, the phase is reversed  $180^\circ$  from the phase at point P. Consequently, the phase between points A and B is  $180^\circ$ . It ensures the output is differential.

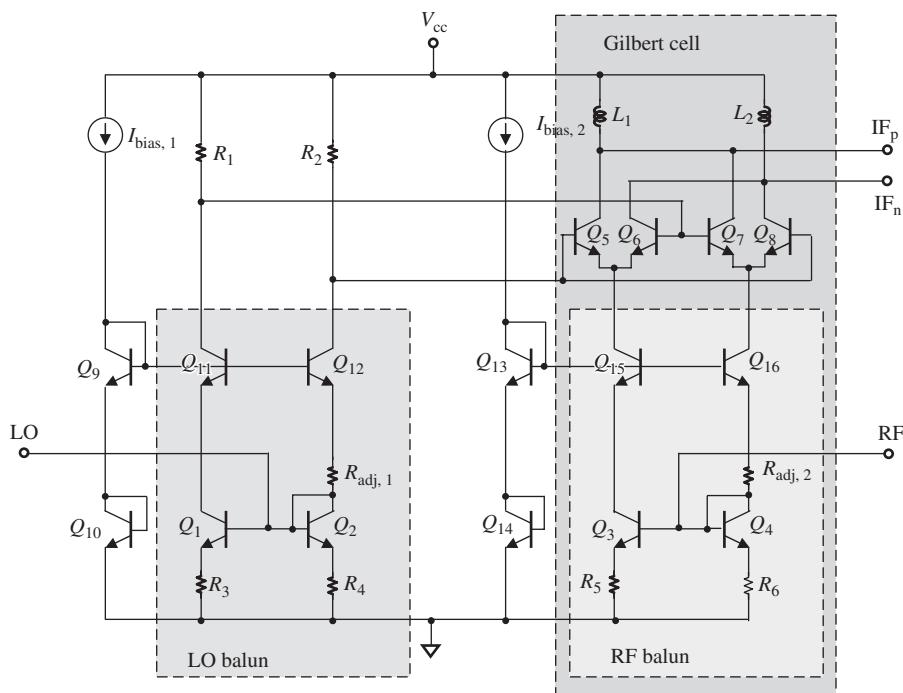
As the single-ended input point of a balun, the impedance at point P must satisfy two conditions:

- At point P, the impedance must be  $50 \Omega$ .
- From point P, the impedance looking toward the left-hand side must be equal to that looking toward the right-hand side.

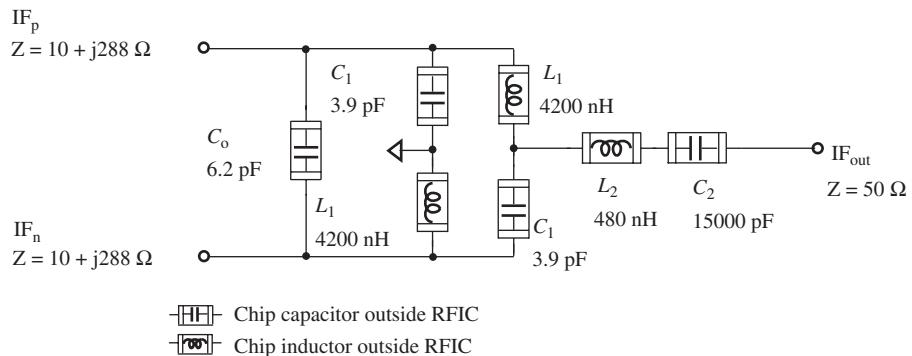
Without the adjusting resistor,  $R_{\text{adj}}$ , the second condition is impossible to satisfy. On the contrary, by adjusting the resistors  $R_{\text{adj}}$  and  $R_e$  the two conditions above can be reached.

The transistors  $Q_1$  and  $Q_2$  are identical. They are applied for current control; they also play the role of isolation between the differential output and the single-ended input. Usually,  $Q_3$  and  $Q_4$  are also identical. However, they could be different as long as the above two conditions are satisfied.

Figure 18.21 shows the active mixer with a bipolar Gilbert cell and active LO and RF baluns.



**Figure 18.21.** Active mixer with a bipolar Gilbert cell and active LO and RF baluns.



**Figure 18.22.** LC balun for IF differential terminals in the active mixer with a bipolar Gilbert cell.

The left-hand side of Figure 18.21 is the LO balun, and the right-hand side of Figure 18.21 is the Gilbert cell and RF balun. The RF balun is based on the same operating principle as the LO balun. It can be seen that the RF balun is also part of the Gilbert cell. The current flowing through the LO balun is controlled by the current source  $I_{\text{bias},1}$ , and the current flowing through the Gilbert cell and RF balun is controlled by the current source  $I_{\text{bias},2}$ . The transistors  $Q_9-Q_{16}$  are the parts of the current mirror and solely serve for current control.

The outstanding achievement in this design is that no capacitors are needed and the current reuse is applied to the Gilbert cell and RF balun. Since there are no capacitors, the RFIC chip's size is reduced.

Let us examine the off-chip IF balun. Figure 18.22 shows its schematic: it is constructed by chip inductors and chip capacitors. The LC balun design is based on the input and output impedances and the operating frequencies, which is introduced and discussed in Section 15.3. The operating frequency in this design is 45 MHz, and the input and output impedances are shown in Figure 18.22.

The active mixer with a bipolar Gilbert cell and active baluns introduced in this section have been successfully packaged into actual radios.

The operating frequencies and DC power supply are

- $f_{\text{RF}} = 860 \text{ MHz}$ ,
- $f_{\text{LO}} = 905 \text{ MHz}$ ,
- $f_{\text{IF}} = 45 \text{ MHz}$ ,
- $V_{\text{cc}} = 3.0 \text{ V}$ , and
- $I_{\text{cc}} = 3.14 \text{ mA}$ .

Its performance is

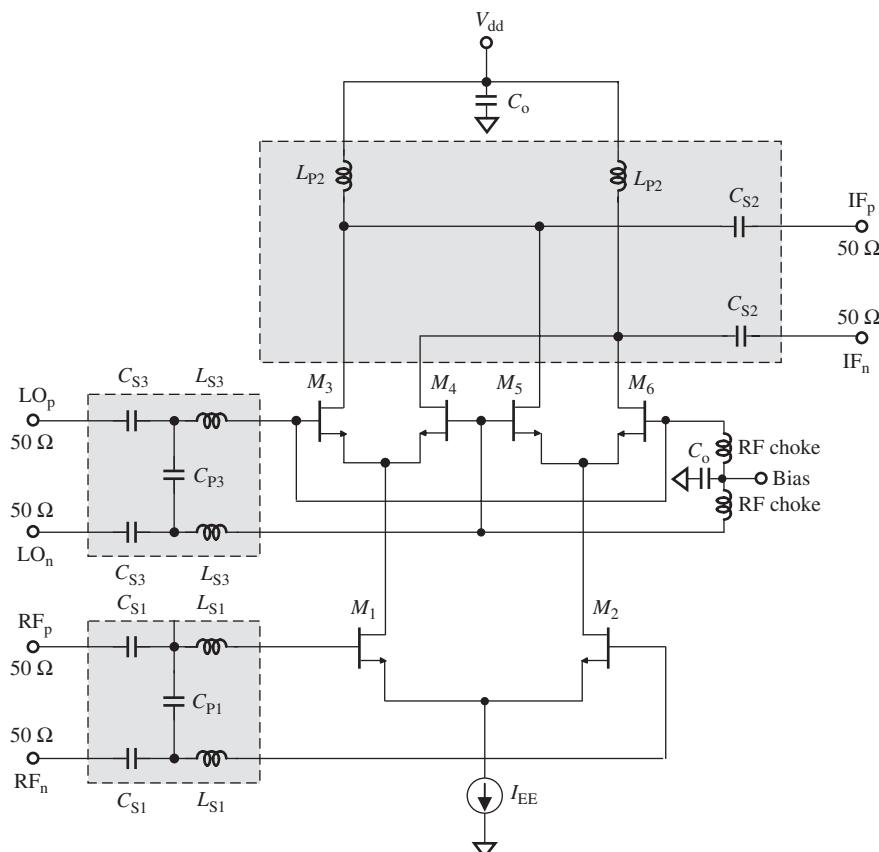
- $P_{\text{LO}} = -10 \text{ dB}_m$ ,
- $\text{CG} = 3.0 \text{ dB}$ ,
- $\text{RL}_{\text{RF}} = -12.8 \text{ dB}$ ,
- $\text{RL}_{\text{LO}} = -13.0 \text{ dB}$ ,
- $\text{NF} = 11.18 \text{ dB}$ , and
- $\text{IIP}_3 = 3.98 \text{ dB}_m$ .

### 18.3.4 Active Mixer with MOSFET Gilbert Cell

The MOSFET and bipolar transistor mainly differ in input impedance. The impedance at the gate of the MOSFET is much higher than the impedance at the base of the bipolar transistor.

Owing to the high impedance at the gate of a MOSFET, building an active balun by the MOSFET device seems to be difficult compared to building one by bipolar transistors. Passive baluns might be good candidates for the active mixer with a MOSFET Gilbert cell. The difficult part of this design task is the implementation of the input impedance matching network. In narrow band cases, a reasonable topology as shown in Figure 18.23 is recommended.

Both the LO and RF impedance matching networks are connected from  $50 \Omega$  to the gates of MOSFETs. As mentioned above, the impedance at the gate of the MOSFET is quite high and usually is located in region 4 on the Smith Chart. It is reasonable to insert an inductor,  $L_{S3}$  or  $L_{S1}$ , so as to move the original impedance at the gates of the transistors to region 1 on the Smith Chart. Then, using the two capacitors,  $C_{P3}$  and  $C_{S3}$ , or  $C_{P1}$  and  $C_{S1}$ , the impedance is pulled to approach  $50 \Omega$ . Two capacitors in both the LO and RF impedance matching networks play the role of DC blocking as well since DC bias must be provided to the gate of the transistors. It is therefore concluded that the

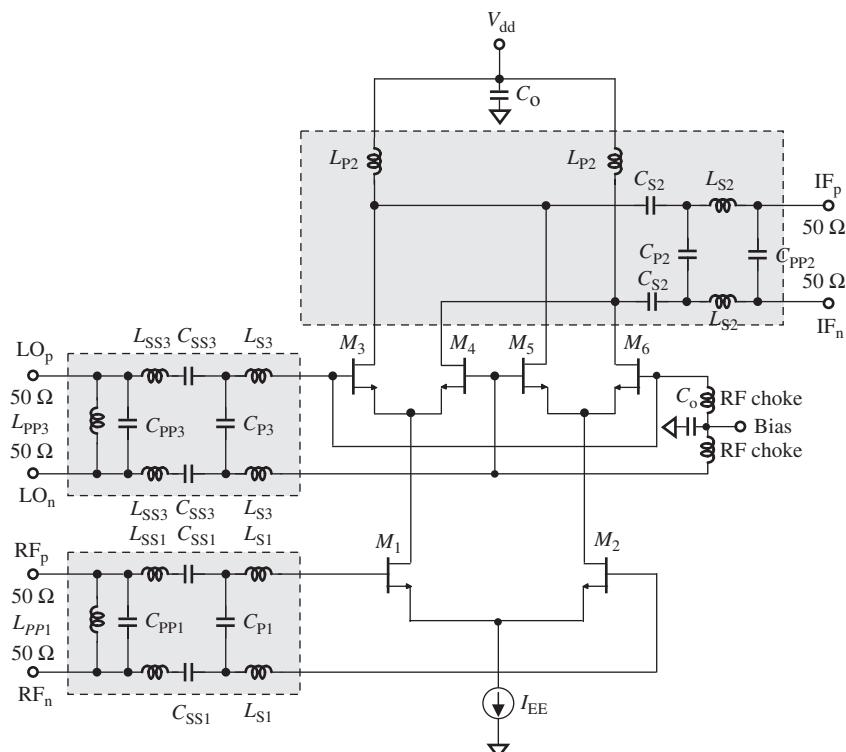


**Figure 18.23.** Recommended impedance matching networks for an active mixer with MOSFET Gilbert cell in narrow band cases.

topology of the impedance matching network for LO or RF ports is the most economic, with a minimum part count.

At the IF output, the IF impedance matching network is inserted from the drain of the transistor to  $50 \Omega$ . The first part is an inductor  $L_{P2}$  in parallel. This can definitely not be replaced by a capacitor because this part is expected to feed the DC from the DC power supply as well. However, it may be replaced by a resistor. Indeed, using a resistor to provide DC to the drain of the transistor should not be a problem. However, this resistor would occupy quite a high percentage of the DC voltage drop and thus reduce the DC voltage drop from the drain of the transistor to the ground, eventually degrading the linearity of the transistor. Therefore, the first part in the IF impedance matching network must be an inductor. There seems to be no other choice. After this part is inserted between the drain and  $V_{dd}$ , the original impedance is expected to move to the vicinity of the circle  $r = 1$ . Then, in terms of the capacitor  $C_{S2}$  in series, the impedance is pulled down to around  $50 \Omega$ . It is concluded that such an IF impedance matching network is the most reasonable and economic.

In wide band or UWB cases a reasonable topology is recommended, as shown in Figure 18.24, which is described in chapter 4 discussing impedance matching in wide band cases. The unsatisfied point here is that there are five inductors appearing in the impedance matching network. This is not good for either cost saving or noise performance. Our excuse is that it is a design for a wide band circuit, and therefore is allowed a greater cost and higher part count. However, the potential for improvement exists, and further efforts should be applied.



**Figure 18.24.** Recommended impedance matching networks for an active mixer with a MOSFET Gilbert cell in wide band cases

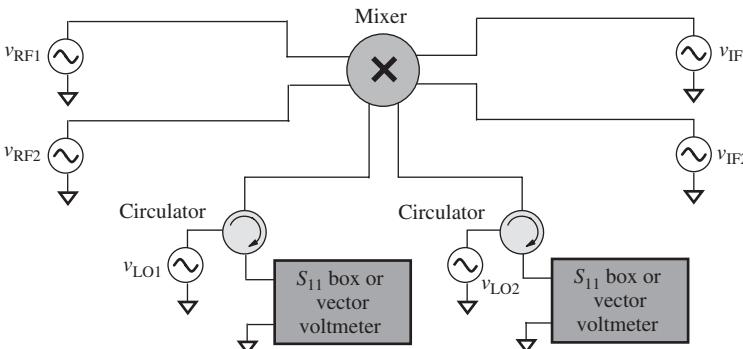
## 18.4 DESIGN SCHEMES

### 18.4.1 Impedance Measuring and Matching

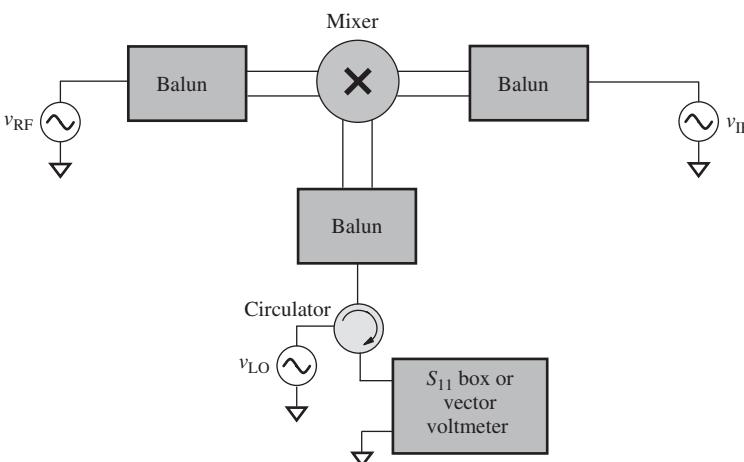
Among RF blocks, one of the more complicated blocks is the mixer, because it has three ports with three different frequencies, the RF, LO, and IF ports. This complexity is enhanced if all three ports are differential, since the number of ports increase from three to six, as shown in Figure 18.25. The power sources  $v_{RF1}$  and  $v_{RF2}$ ,  $v_{LO1}$  and  $v_{LO2}$ , and  $v_{IF1}$  and  $v_{IF2}$  all are differential sources and must be set up with phase differences of  $180^\circ$  between the source pairs.

Among these six ports, the RF and IF ports have small signals so that their impedances can be measured directly by the network analyzer. The LO injection is a large signal, so it is necessary to apply the circulator for an accurate impedance measurement. In the simulation stage, an  $S_{11}$  box can be used to measure the voltage returned from the port in the mixer. In actual testing, a vector voltmeter would be a good assistant to measure the voltage returned from the port in the mixer.

As a matter of fact, as shown in Figure 18.26, the differential configuration of a mixer can be simplified from six differential ports to three single ports by means of the balun either in the simulation phase or in the testing stage. This greatly reduces the complexity of measurement.



**Figure 18.25.** Impedance measurement for a differential mixer by means of a circulator.



**Figure 18.26.** Change of impedance measurement for a differential mixer by means of a balun.

Of course, the impedances of the baluns from the single-ended to the differential must be well known before they are applied to the test setup.

In the impedance testing or measuring, special attention must be paid to the IF portions. The incident power from the power sources,  $v_{IF1}$  and  $v_{IF2}$ , or  $v_{IF}$ , would be reflected back to the power sources vector voltmeter. One can calculate the impedance from the ratio of the reflected and incident power, or the reflection coefficient. However, there is another unexpected IF power, the product of the RF input and LO injection, which also flows into the power source  $v_{IF}$  and consequently disturbs the reflected power. In order to separate the unexpected IF power from the reflected power flowing into the power source  $v_{IF}$ , there are two ways to measure the IF impedance measurement:

- The IF at the IF power sources can be shifted a little bit so as to avoid the unexpected IF power overlapping with the reflected power flowing into the power source  $v_{IF}$ . The measured impedance at the IF ports, of course, corresponds to the shifted frequency but not to the expected IF. We can approximately treat the measured impedance at the shifted frequency as the IF impedance because the frequency shift is very small.
- Instead of shifting the IF, the unexpected IF product due to the RF input and LO injection can be prohibited if the RF input is turned off.

Another possible way to prohibit the unexpected IF product due to the RF input and LO injection is to turn-off the LO injection. However, this is not allowed do so because the real operating status of a mixer is determined and maintained by the LO injection.

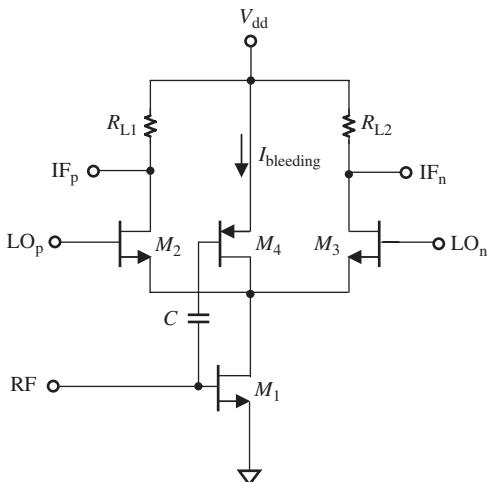
Impedance matching is one of the key subjects in RF circuit design. It is not an easy task sometimes. One of the examples is to impedance-match a differential mixer. Owing to the imperfect isolation between the RF, LO, and IF ports, impedance matching must be conducted with many iterations back and forward between three ports, especially between the LO and the RF ports, because the isolation between them is poor. The matching procedures should be as follows:

1. Matching of the LO port first: the LO injection is the main source controlling the operating status of a mixer. The operating status of a mixer never approaches a normal state until the LO port is well matched.
2. Matching of the RF port second.
3. Repeating the two steps above until the variation between steps is negligible: owing to the imperfect isolation between the LO and RF port, the variation resulting from the matching at the RF port would change the matching status at the LO port, and vice versa.
4. Finally, matching the IF port: usually, the isolation between the LO and IF ports or between the RF and IF ports is good enough so that the variation resulting from the matching at the IF port will not impact the matching status in either the LO or RF ports.

### 18.4.2 Current Bleeding

Lee and Choi (2000) proposed a new mixer topology with current bleeding as shown in Figure 18.27.

It is a single-balanced mixer, in which a p-channel transistor  $M_4$  functions as a bleeding current source as well as a part of the driver amplifier. This technology leads to better performance of conversion gain, linearity, noise figure, and LO isolation.



**Figure 18.27.** Schematic diagram of single-balanced mixer with current bleeding.

Conversion gain and  $IP_3$  are proportional to the square root of the current flowing into  $M_1$ . Without the bleeding current flowing through  $M_4$ , the increase in the current flowing into  $M_1$  must be traded off with the reduction of the load resistors  $R_{L1}$  and  $R_{L2}$ . This results in the reduction in conversion gain and the value of  $IP_3$ . With the bleeding current flowing through  $M_4$ , the increase in current flowing into  $M_1$  does not have to be traded off with the reduction of the load resistors  $R_{L1}$  and  $R_{L2}$  since the increased current is flowing through  $M_4$  but not  $M_2$  and  $M_3$ . This results in an increase in conversion gain and the value of  $IP_3$ .

Secondly,  $M_4$  is a part of the driver amplifier together with  $M_1$ ; the higher overall transconductance reduces the noise figure.

Finally, the switching transistors  $M_2$  and  $M_3$  can be operated at a lower gate-source voltage or smaller size due to the bleeding current introduced, so that the lower LO injection is good enough to have the same performance as that in the case without bleeding current introduced.

The experiment indicated that the performance of the mixer topology with bleeding current demonstrates an approximately 4 dB higher conversion gain, 0.9 dB lower noise figure, 2.4 dB higher  $IP_3$  at the output, and 3.3 dB lower LO injection than that of the conventional mixer.

#### 18.4.3 Multi-tanh Technique

In order to enhance the linearity of the mixer with a Gilbert cell, the multi-tanh technique was developed by Gilbert (1998).

The key idea of the multi-tanh technique is to connect multiple differential pairs together in parallel, in which the transconductance function of each pair is arranged to cover a different range of input voltage. A much more linear overall function is approached as all transconductance functions from all the individual pairs are summed together. In other words, the individually nonlinear transconductance functions are separated along the input voltage axis so that the transconductance can be kept constant over a wide range of the input voltage.

For an individual differential pair, let us recall expression (18.21)

$$\Delta I = I_{EE} \frac{v_{RF}}{2V_T} \tanh\left(\frac{v_{LO}}{2V_T}\right), \quad (18.21)$$

where  $v_{RF}$  is assumed to be a small signal and condition (18.20) is satisfied.

Let us focus on the linear relationship between  $\Delta I$  and  $v_{LO}$ .

For simplicity, let us redenote

$$I_{out} = \Delta I, \quad (18.27)$$

$$I_E = I_{EE} \frac{v_{RF}}{2V_T}, \quad (18.28)$$

$$v_{in} = v_{LO}. \quad (18.29)$$

Then, expression (18.21) becomes

$$I_{out} = I_E \tanh \frac{v_{in}}{2V_T}. \quad (18.30)$$

Its transconductance is

$$g_m = \frac{\partial I_{out}}{\partial V_{in}} = \frac{I_E}{2V_T} \sec h^2 \frac{v_{in}}{2V_T}. \quad (18.31)$$

The hyperbolic tangent relationship between  $g_m$  and  $v_{in}$  is plotted in Figure 18.28. The value of  $g_m$  changes considerably as  $v_{in}$  is varied. This indicates that it is difficult to keep the transconductance constant over a wide range of input voltage, resulting in the nonlinearity between the input and output signal.

Using the multi-tanh technique, in which the linearity between the input and output can be improved significantly, the Gilbert cell is no longer a simple differential pair but is a combination of  $n$  differential pairs so that the resultant transconductance is

$$g_m = \sum_{j=1}^n \frac{I_j}{2V_T} \sec h^2 \left( \frac{v_{in} + v_j}{2V_T} \right). \quad (18.32)$$

Figure 18.29 shows the schematic of  $n$  multi-tanh differential pairs. The outputs of all the differential pairs are connected together as the terminal  $I_{out}$  in parallel. The resultant transconductance,  $g_m$ , is a simple sum of all the individual transconductances,  $g_{mj}$ . The

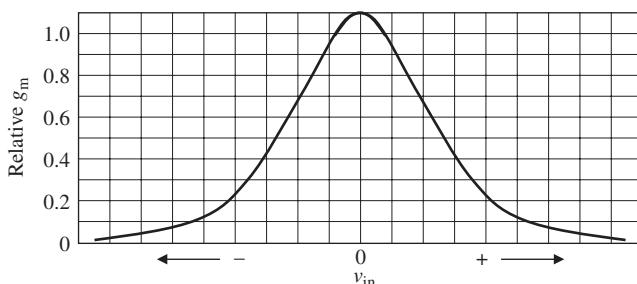
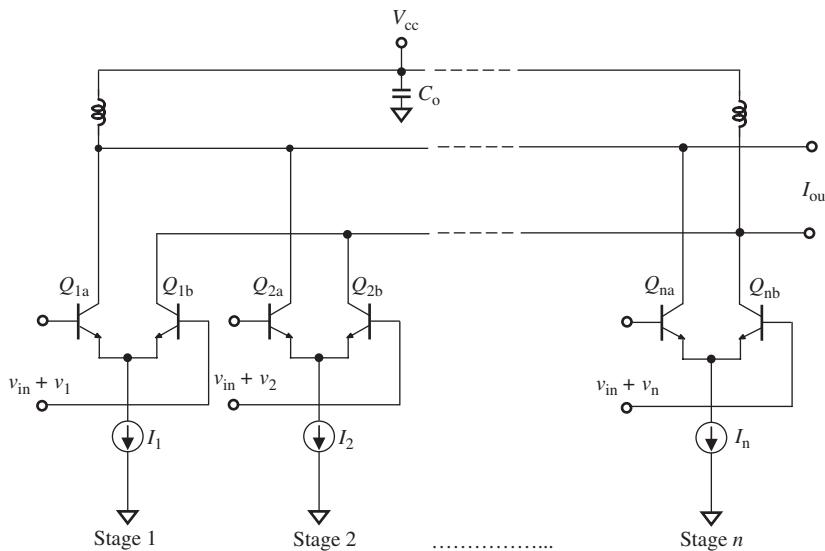
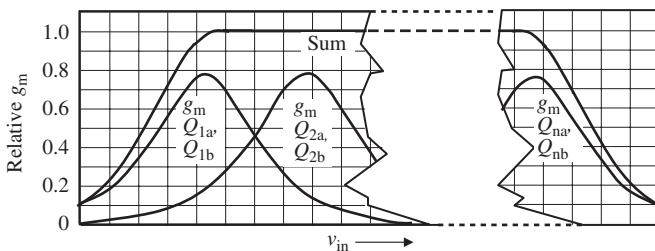


Figure 18.28. Typical transconductance  $g_m$  of a Gilbert cell.



**Figure 18.29.**  $n$  Multi-tanh differential pairs. (Source: Gilbert B. The multi-tanh principle: a tutorial overview. IEEE J of Solid-State Circuits January 1998;33(1):4; Figure 1).

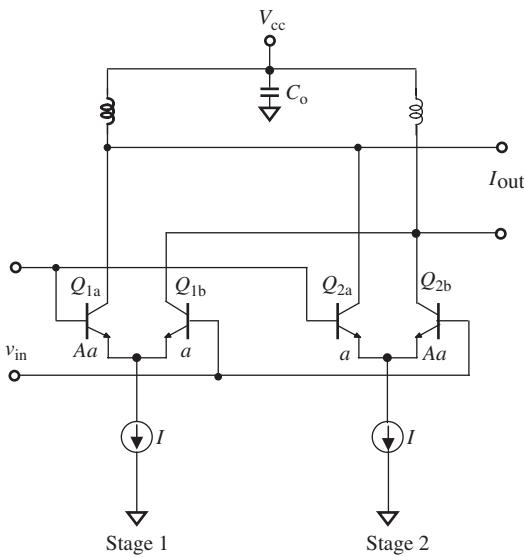


**Figure 18.30.** Multi- $g_m$  components of  $n$  differential pairs.

individual input voltage  $v_{\text{in}} + v_j$  must and can be produced by  $v_{\text{in}}$  through a voltage amplifier/follower or a transformer. By adjusting all the individual input voltages  $v_{\text{in}} + v_j$ , the individual transconductance  $g_{mj}$  can be reasonably arranged over a different range of input voltages so that the sum of the transconductances  $g_m$  are kept constant over a wide range of input voltages as shown in Figure 18.30.

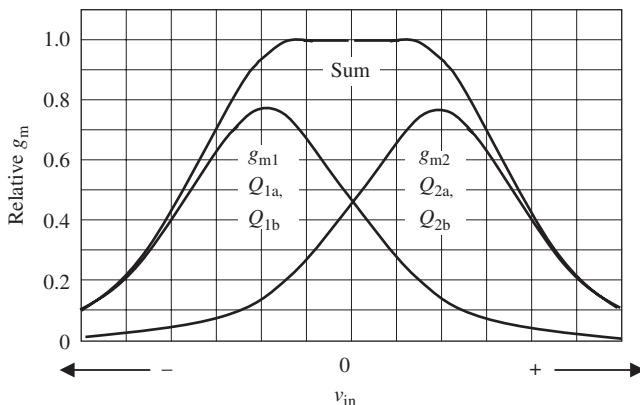
Double-tanh differential pairs are the simplest multi-tanh differential pairs; one is shown in Figure 18.31. Instead of different individual input voltages  $v_{\text{in}} + v_j$ , the input voltage for both differential pairs is kept the same, but the emitter area of the transistors is different in each differential pair. As shown in Figure 18.31, the emitter area of  $Q_{1b}$  and  $Q_{2a}$  is kept the same and is equal to  $a$ , while the emitter area of  $Q_{1a}$  and  $Q_{2b}$  is kept the same and is equal to  $Aa$ , which is  $A$  times larger than the emitter area of  $Q_{1b}$  and  $Q_{2a}$ .

As a matter of fact, the difference of emitter area is equivalent to the difference of the different individual input voltages  $v_{\text{in}} + v_j$ . The first differential pair, marked "Stage 1" in Figure 18.31, results in the  $g_{m1}$  curve in Figure 18.32, which is shifted to the left-hand or negative direction of the input voltage  $v_{\text{in}}$  from the case when  $v_j = 0$ . The second differential pair, marked "Stage 2" in Figure 18.31, results in the  $g_{m2}$  curve in Figure 18.32, which is shifted to the right-hand or positive direction of the input voltage  $v_{\text{in}}$  from the case when  $v_j = 0$ . The sum of  $g_m$  is therefore kept constant over a wider range of input voltages.



$Aa$  = Emitter area of  $Q_{1a}$  and  $Q_{1b}$   
 $a$  = Emitter area of  $Q_{1b}$  and  $Q_{2a}$

**Figure 18.31.** Doublet-tanh differential pairs. (Source: Gilbert B. The multi-tanh principle: a tutorial overview. IEEE J of Solid-State Circuits January 1998;33(1):4; Figure 5).



**Figure 18.32.** Dual  $g_m$  components of the doublet-tanh differential pairs. (Source: Gilbert B. The multi-tanh principle: a tutorial overview. IEEE J of Solid-State Circuits January 1998;33(1):7; Figure 6).

#### 18.4.4 Input Types

There are two RF input types of mixers built using Gilbert cells: the CS (common source) and CG (common gate) types of RF input. Figure 18.33 shows a CS-type RF input, and Figure 18.34 shows a CG-type RF input. The impedance of the mixer at the RF ports with a CG-type RF input is much lower than that of the mixer with a CS-type RF input since the input impedance for a MOSFET with a CG configuration is much lower than the input impedance of that with a CS configuration. Low impedance is easier for RF circuit design than high impedance. As shown in Figure 18.34, there is a transformer applied in the RF ports in the mixer with a CG-type RF input. In addition, a mixer with a CG-type RF input provides a wider bandwidth of frequency response than a CS-type RF input. Consequently, in most practical mixer designs, the CG-type RF input is adapted.

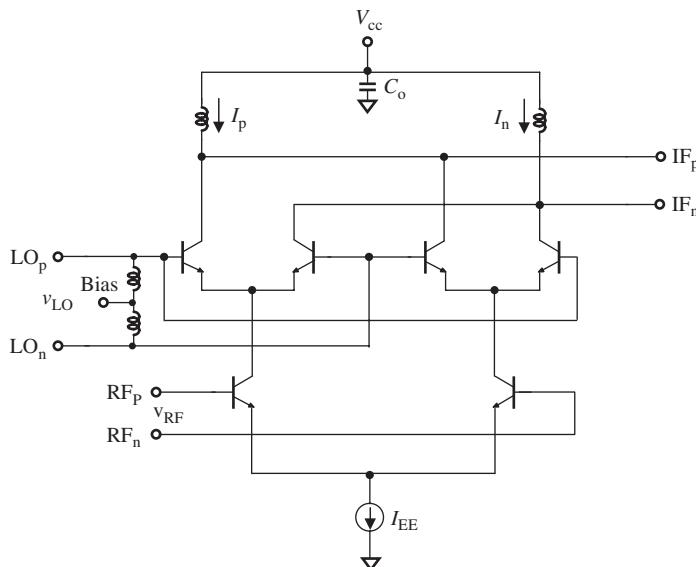
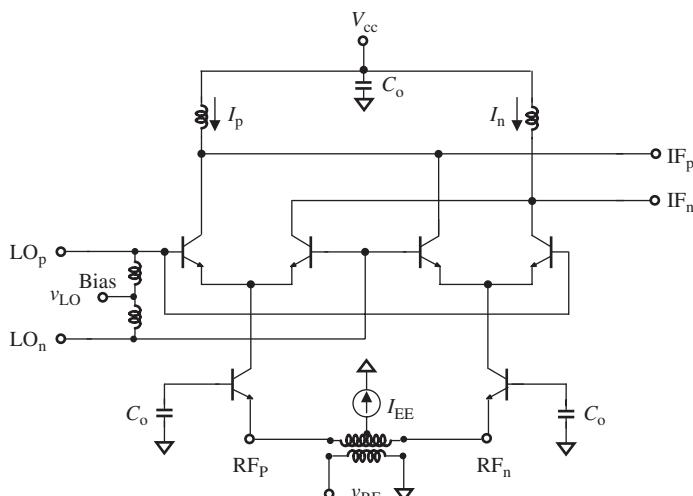


Figure 18.33. An active mixer with a bipolar Gilbert cell and with CS input for RF signal.



**Figure 18.34.** An active mixer with bipolar Gilbert cell and with CG input for RF signal.

Similarly, the CB (common base) type of RF input in a mixer built using bipolar transistors with a Gilbert cell has the same advantages of the CE (common emitter) type RF input.

## APPENDICES

## 18.A.1 Trigonometric and Hyperbolic Functions

$$e^{j\theta} = \cos \theta + j \sin \theta, \quad (18.A.1)$$

$$e^{-j\theta} = \cos \theta - j \sin \theta, \quad (18.A.2)$$

$$\cos \theta = \frac{e^{j\theta} + e^{-j\theta}}{2}, \quad (18.A.3)$$

$$\sin \theta = \frac{e^{j\theta} - e^{-j\theta}}{2j}, \quad (18.A.4)$$

$$\cosh \theta = \frac{e^\theta + e^{-\theta}}{2}, \quad (18.A.5)$$

$$\tanh \theta = \frac{\sinh \theta}{\cosh \theta} = \frac{e^\theta - e^{-\theta}}{e^\theta + e^{-\theta}} = \frac{e^{2\theta} - 1}{e^{2\theta} + 1}, \quad (18.A.6)$$

$$\sinh \theta = \frac{e^\theta - e^{-\theta}}{2}, \quad (18.A.7)$$

$$\sin(\alpha \pm \beta) = \sin \alpha \cos \beta \pm \cos \alpha \sin \beta, \quad (18.A.8)$$

$$\cos(\alpha \pm \beta) = \cos \alpha \cos \beta \mp \sin \alpha \sin \beta, \quad (18.A.9)$$

$$\tan(\alpha \pm \beta) = \frac{\tan \alpha \pm \tan \beta}{1 \mp \tan \alpha \tan \beta}, \quad (18.A.10)$$

$$\sin \alpha \cos \beta = \frac{1}{2} \sin(\alpha - \beta) + \frac{1}{2} \sin(\alpha + \beta), \quad (18.A.11)$$

$$\cos \alpha \sin \beta = -\frac{1}{2} \sin(\alpha - \beta) + \frac{1}{2} \sin(\alpha + \beta), \quad (18.A.12)$$

$$\sin \alpha \sin \beta = \frac{1}{2} \cos(\alpha - \beta) - \frac{1}{2} \cos(\alpha + \beta), \quad (18.A.13)$$

$$\cos \alpha \cos \beta = \frac{1}{2} \cos(\alpha - \beta) + \frac{1}{2} \cos(\alpha + \beta), \quad (18.A.14)$$

$$\sin \alpha + \sin \beta = 2 \sin \left( \frac{\alpha + \beta}{2} \right) \cos \left( \frac{\alpha - \beta}{2} \right), \quad (18.A.15)$$

$$\sin \alpha - \sin \beta = 2 \cos \left( \frac{\alpha + \beta}{2} \right) \sin \left( \frac{\alpha - \beta}{2} \right), \quad (18.A.16)$$

$$\cos \alpha + \cos \beta = 2 \cos \left( \frac{\alpha + \beta}{2} \right) \cos \left( \frac{\alpha - \beta}{2} \right), \quad (18.A.17)$$

$$\cos \alpha - \cos \beta = -2 \sin \left( \frac{\alpha + \beta}{2} \right) \sin \left( \frac{\alpha - \beta}{2} \right), \quad (18.A.18)$$

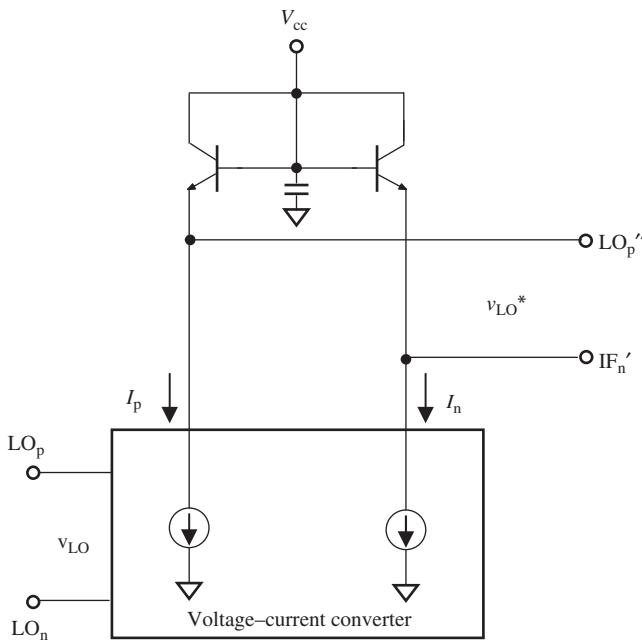
$$\tanh x = x - \frac{1}{3}x^3 + \frac{2}{15}x^5 - \frac{17}{315}x^7 + \frac{62}{2835}x^9 - \dots, \quad (18.A.19)$$

$$\tanh^{-1} x = x + \frac{1}{3}x^3 + \frac{1}{5}x^5 + \frac{1}{7}x^7 + \frac{1}{9}x^9 - \dots, \quad (18.A.20)$$

$$\tanh^{-1} x = \frac{1}{2} \ln \left( \frac{1+x}{1-x} \right). \quad (18.A.21)$$

## 18.A.2 Implementation of $\tanh^{-1}$ Block

The  $\tanh^{-1}$  block can be implemented by two diode-connected transistors, which are shown in Figure 18.A.1 (Gray et al., 2001).



**Figure 18.A.1.** An inverse hyperbolic tangent circuit is implemented by two diode-connected transistors. (Source: Gray PR and Meyer RG. *Analysis and design of analog integrated circuits*. 3rd ed. New York: John Wiley & Sons, Inc; 1963, p.674, Figure 10.12).

The function of the voltage–current converter is

$$I_p = I_o + K v_{LO}, \quad (18.A.22)$$

$$I_n = I_o - K v_{LO}, \quad (18.A.23)$$

where

$I_o$  = DC component of  $I_p$  or  $I_n$ , and

$K$  = transconductance of the voltage–current converter.

The differential voltage developed across the two diode-connected transistor is

$$v_{LO*} = V_T \ln \frac{I_o + K v_{LO}}{I_S} - V_T \ln \frac{I_o - K v_{LO}}{I_S} = V_T \ln \frac{I_o + K v_{LO}}{I_o - K v_{LO}}, \quad (18.A.24)$$

$$v_{LO*} = 2V_T \tanh^{-1} \left( \frac{K v_{LO}}{I_o} \right). \quad (18.A.25)$$

Replacing  $v_{LO}$  in expression (18.21) by  $v_{LO*}$  as shown in the expression (18.A.25), we have

$$\Delta I = I_{EE} \frac{v_{RF}}{2V_T} \frac{K v_{LO}}{I_o}, \quad (18.A.26)$$

which is the desired differential current.

## REFERENCES

- Gray PR, Hurst PJ, Lewis SH, Meyer RG. *Analysis and Design of Analog Integrated Circuits*. 4th ed. New York: John Wiley & Sons, Inc., 2001.
- Gilbert B. The multi-tanh principle: a tutorial overview. *IEEE J Solid-State Circ* 1998;33(1): 2–17.
- Lee S-G, Choi J-K. Current-reuse bleeding mixer. *Electron Lett* 2000;36(8):696–697.

## FURTHER READING

- Bautista EE, Bastini B, Heck J. A high IIP2 down-conversion mixer using dynamic matching. *IEEE J Solid-State Circ* 2000;35(12):1934–1941.
- Chen TH, Chang KW, Bui SBT, Liu LCT, Pak GS. Broadband single- and double-balanced resistive HEMT monolithic mixers. *IEEE Trans Microw Theory Tech* 1995;43(3):477–484.
- Darabi H, Chiu J. A noise cancellation technique in active RF-CMOS mixers. *IEEE J Solid-State Cir* 2005;40(12):2628–2632.
- Hamed KW, Freundorfer AP, Antar YMM. A monolithic double-balanced direct conversion mixer with an integrated wideband passive balun. *IEEE J Solid-State Circ* 2005;40(3):622–629.
- Kim W, Keum D, Kang S, Park B-H. A mixer with third-order nonlinearity cancellation technique for CDMA applications. *IEEE Microw Wireless Compon Lett* 2007;17(1):76–78.
- Kim Y, Kim Y, Lee S. Linearized mixer using predistortion technique. *IEEE Microw Wireless Compon Lett* 2002;12(6):204–205.
- Kivekäs K, Pärssinen A, Halonen KAI. Characterization of IIP2 and DC-offsets in transconductance mixers. *IEEE Trans Circ Syst-II: Analog Digit Signal Process* 2001;48(11):1028–1038.
- Margraf M, Boeck G. Analysis and modeling of low-frequency noise in resistive FET mixers. *IEEE Trans Microw Theory Tech* 2004;52(7):1709–1718.
- Otaka S, Ashida M, Ishii M, Itakura T. A +10-dBm IIP3 SiGe mixer with IM3 cancellation technique. *IEEE J Solid-State Circ* 2004;39(12):2333–2339.
- Safarian AQ, Yazdi A, Heydari P. Design and analysis of an ultrawide-band distributed CMOS mixer. *IEEE Trans Very Large Scale Integr (VLSI) Syst* 2005;13(5):618–629.
- Sheng L, Larson LE. An Si-SiGe BiCMOS direct-conversion mixer with second-order and third-order nonlinearity cancellation for WCDMA applications. *IEEE Trans Microw Theory Tech* 2003;51(11):2211–2220.
- Siddiqi AA, Kwasniewski T. *2.4GHz RF Down-Conversion Mixers in Standard CMOS Technology*. Ottawa, ON, Canada: Department of Electronics, Carleton University; 2004. 0-7803-8251-X/04/\$17.00 © 2004 IEEE, IV 321–324, ISCAS 2004.
- Svitek R, Raman S. 5–6GHz SiGe active I/Q subharmonic mixers with power supply noise effect characterization. *IEEE Microw Wireless Compon Lett* 2004;14(7):319–321.
- Valla M, Montagna G, Castello R, Tonietto R, Bietti I. A 72-mW CMOS 802.11a direct conversion front-end With 3.5-dB NF and 200-kHz 1/f noise corner. *IEEE J Solid-State Circ* 2005;40(4):970–976.
- Wentzloff DD, Chandrakasan AP. A 3.1–10.6GHz Ultra-Wideband Pulse-Shaping Mixer. In: *IEEE Radio Frequency Integrated Circuits Symposium*; 2005. pp. 83–86.
- Zhou S, Chang M-CF. A CMOS passive mixer with low flicker noise for low-power direct-conversion receiver. *IEEE J Solid-State Circ* 2005;40(5):1084–1092.

## EXERCISES

1. Compare the advantages and disadvantages between the active and passive mixer.
2. What is the main improvement from the basic passive mixer to the ring diode mixer?

3. How does the LO injection level impact on the performance of a passive mixer?
4. Both the passive resistive mixer with Gilbert cell and the quad-diode ring mixer have the same double-balanced configuration. What is the difference in their operating principle?
5. What are the advantages and disadvantages of a double-balanced passive resistive mixer?
6. The operating principle of a mixer is based on the second-order nonlinearity of the device. What does its linearity mean?
7. How would an inverse hyperbolic tangent circuit be implemented so as to linearize the hyperbolic function existed in Gilbert cell?
8. How to implement an active balun for a mixer with Gilbert cell in the RFIC design?
9. What is current breeding technology applied in an active mixer design?
10. Describe the principle of the multi-tanh technology.

## ANSWERS

1. The advantages and disadvantages between active and passive mixer can be listed as follows:

Item	Active Mixer	Passive Mixer	Unit
Current drain	~2–5	~0	mA
LO Injection	~-10~-0	~5–10	dBm
Conversion gain	~5–10	~-5 to -3	dB
Noise figure	~10–15	~3–5	dB
Bandwidth	Narrower	Wider	—
Part count	~13	~6	—
Reliability	Lower	Higher	—
Cost	Lower	Higher	—

2. The main improvement from the basic passive mixer to the ring diode mixer are
  - (a) DC disappears;
  - (b) all odd-order nonlinearity disappear.
3. The impact of LO injection level on the performance of a passive mixer is twofold:
  - (a) In order to get a normal conversion gain, LO injection must exceed over a threshold.
  - (b) For a passive mixer, there is an optimum of LO injection, at which the noise figure approaches to a minimum.
4. Both the passive resistive mixer with a Gilbert cell and the passive quad-diode ring mixer have the same double-balanced configuration. However, their operating principle is essentially different. In the passive resistive mixer with a Gilbert cell, the RF current that flows through the activated channel of the device is modulated by the LO injection, which is applied to the gate of the device. The activated channel under the gate of device behaves like a resistor. In the passive quad-diode ring mixer, the RF current and LO injection current flow directly through the diode.

5. Comparing with the passive quad-diode ring mixer, the advantages of a double-balanced passive resistive mixer are (i) no DC, (ii) low LO injection, (iii) low noise, and (iv) better linearity; the disadvantage is low conversion gain.
6. The operating principle of a mixer is based on the second-order nonlinearity of the device. Its linearity is determined by the higher orders' nonlinearity of the device, such as third-order and above for unbalanced mixer, fourth-order and above for balanced mixer. Therefore, the terms of "linearity" and "nonlinearity" are relative.
7. An inverse hyperbolic tangent circuit is shown as follows (Fig. 18.P.1):

$$I_p = I_o + K v_{LO}$$

$$I_n = I_o - K v_{LO}$$

$$v_{LO^*} = V_T \ln \frac{I_o + K v_{LO}}{I_S} - V_T \ln \frac{I_o - K v_{LO}}{I_S} = V_T \ln \frac{I_o + K v_{LO}}{I_o - K v_{LO}}$$

$$\tanh^{-1} x = \frac{1}{2} \ln \left( \frac{1+x}{1-x} \right)$$

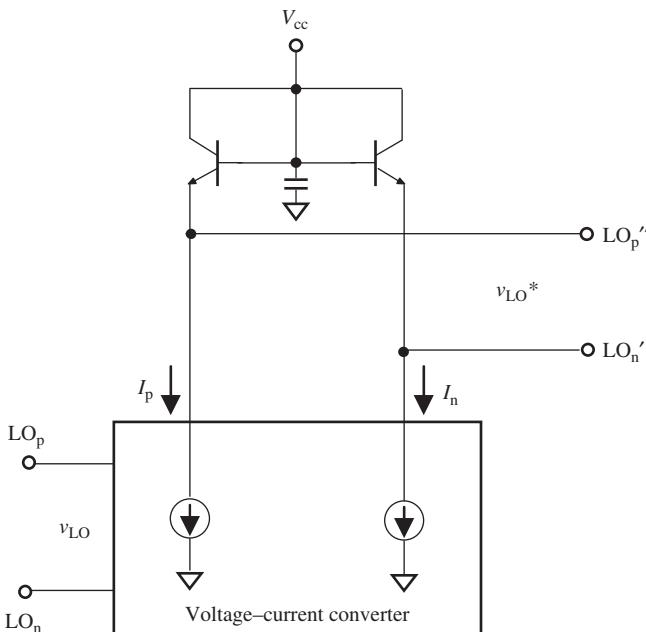


Figure 18.P.1. An inverse hyperbolic tangent circuit implemented by two diode-connected transistors.

8. An active balun can be sketched as Figure 18.P.2. In terms of a current mirror, which is basically formed by Q3 and Q4, the single-ended signal, SN, is converted into a differential pair of signal at point A and B. Q1 and Q2 then magnify and deliver this differential pair of signal to the output terminals, DP1 and DP2. The function of  $R_{adj}$  is to adjust the differential pair of signal so that an ideal differential state can be approached as possible.

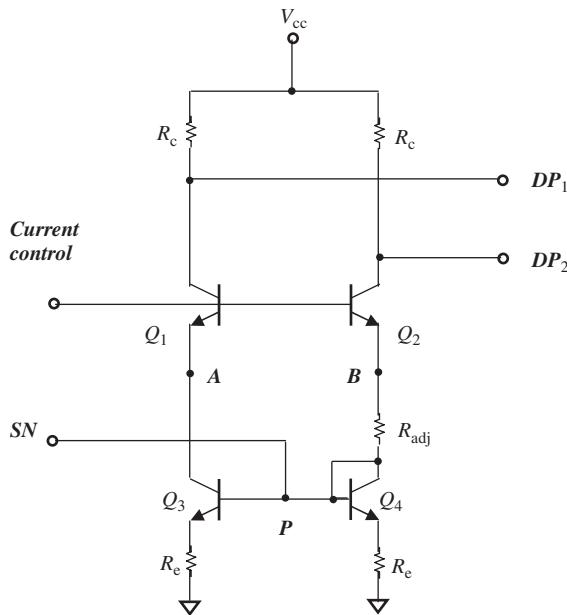


Figure 18.P.2. Sketching of an active balun.

9. The advantages of current breeding technology applied in an active mixer design are as follows (Fig. 18.P.3):

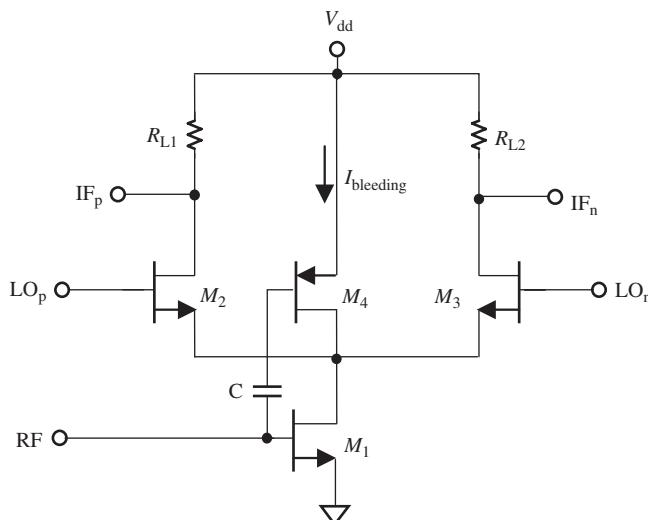


Figure 18.P.3. Schematic diagram of single-balanced mixer with current bleeding.

- (a) The increase of current flowing into  $M_1$  will enhance conversion gain and  $IP_3$ . With bleeding current this increase does not need to take a trade-off with reduction of  $R_{L1}$  and  $R_{L2}$ ;
  - (b) The addition of  $M_4$  enhances the  $g_m$  value of the mixer.
10. The principle of the multi-tanh technology comes from the understanding about the source of nonlinearity of a mixer. The nonlinearity of a mixer is due to the variation of  $g_m$  within the dynamic range of  $v_{in}$ . The variation of  $g_m$  is a tanh function of

$v_{in}$  as shown in Figure 18.P.4. If the device is replaced by multiple devices and the resultant  $g_m$  is kept constant over the dynamic range of  $v_{in}$  as shown in Figure 18.P.5, the linearity of the mixer would be significantly improved.

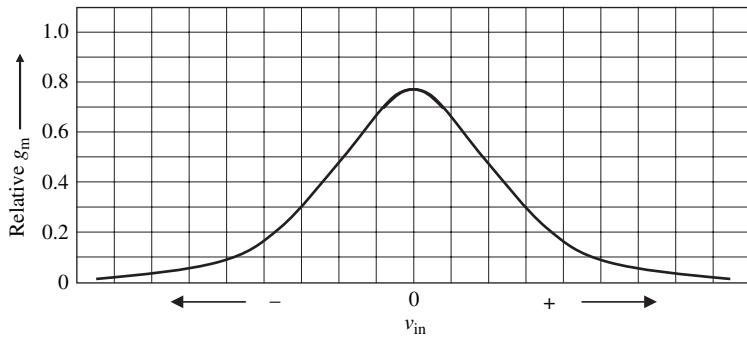


Figure 18.P.4. Typical transconductance  $g_m$  of a Gilbert cell.

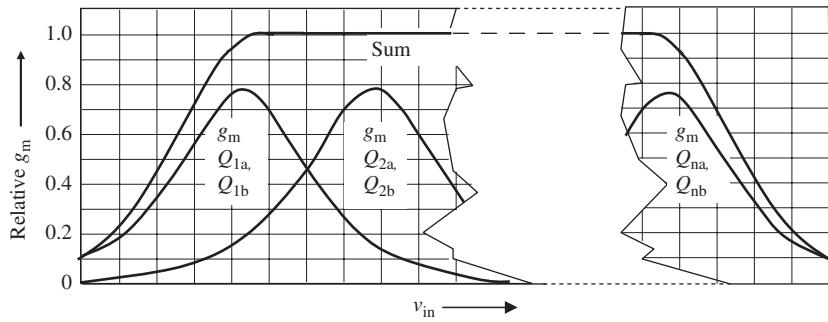


Figure 18.P.5. Multi- $g_m$  components of  $n$  tanh differential pairs.

# TUNABLE FILTER

## 19.1 TUNABLE FILTER IN A COMMUNICATION SYSTEM

There are two kinds of filters in electrical circuit design: the fixed filter and the tunable filter. Intuitively, a fixed filter's passband is fixed and well defined, while a tunable filter's passband can be tuned over a certain frequency range.

Theory for a fixed filter design has been comprehensively developed many decades ago. The engineering developments of passive filters, including the LPF (low pass filter), HPF (high pass filter), BPF (band pass filter), and BRF (band reject filter), have been formulized and tabulated for the design with discrete parts. This simplifies the design procedures: the only thing remaining is how to convert the values of parts into reasonable ones through the so-called  $\Delta$ -\* or  $\pi$ - $T$  transformation.

In past years, the frequency bands assigned for a communication system rose higher and higher. It became more and more difficult to implement filters by discrete parts because the required values of either capacitors or inductors were too small, which are unacceptable for the production line. Many types of filters with distributed parameters were developed, such as the crystal filter, ceramic block filter, microstrip line filter, SAW (surface acoustic wave) filter, and so on.

The theory for a tunable filter design was developed many decades ago. Its implementation, however, faces some barriers. This is one of the reasons why this, and not another type of filter, is selected to be the topic in this book. Today, the SAW filter is very often applied to a communication system, whether it is a dual conversion or a direct conversion system. As a matter of fact, a lot of effort is being put into the development

of a SAW tunable filter at present. In this chapter, we are going to show how to remove the main obstacles in the tunable filter design. It might be helpful to provide some useful clues so as to assist the design of a SAW tunable filter or other type of tunable filter, although in our example the tunable filter is implemented by discrete parts. This is the second reason to include this chapter (The content of this chapter is abstracted from a US patent that was obtained by the author in 1992.).

### 19.1.1 Expected Constant Bandwidth of a Tunable Filter

In a communication system, a customer may occupy only one channel, although the number of available channels in a communication system is usually huge. In other words, the bandwidth that a customer needs is much less than the entire bandwidth of the system. For example, in a UHF portable radio, the entire bandwidth is assigned to be from 403 to 520 MHz, while a customer occupying bandwidth for voice conversation needs no more than 25 kHz. To approach good performance, the ideal condition is to assign 25 kHz only to each customer by a tunable filter, while the rest of the bandwidth is rejected. All of the hardware is the same for each customer; the difference lies only in the control voltage being provided to the tunable filter for each customer. Consequently, the goals of good performance and mass production can be realized simultaneously.

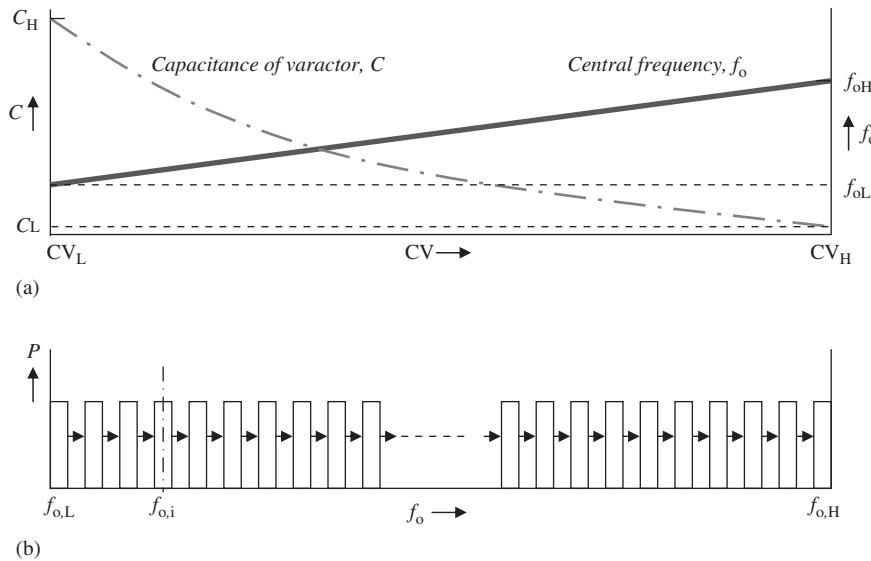
In practical RF tunable filter design, it is possible to request that the bandwidth be narrowed down to 10 MHz or so, but it is almost impossible to narrow the bandwidth down to the order of around 25 kHz. Nevertheless, narrowing the bandwidth is still very helpful in improving the selectivity of the receiver, reducing noise, and preventing a variety of spurious interference. As long as the bandwidth can be narrowed down somewhat from the entire bandwidth of the system, the performance of the portable radio could be appreciably improved. Without a tunable filter functioning in the front end of the receiver, spurious interference and noise might be significant.

Figure 19.1 shows how the tunable filter operates in a portable radio. When the control voltage,  $CV$ , is moved from its minimum,  $CV_L$ , to its maximum,  $CV_H$ , the capacitance of the varactor,  $C$ , is changed from its maximum,  $C_H$ , to its minimum,  $C_L$ , and the central frequency of the tunable filter,  $f_o$ , moves from its minimum,  $f_{oL}$ , to its maximum,  $f_{oH}$ , correspondingly.

To enable a tunable filter tuned over a wide frequency range to be realistic, the bandwidth should be kept as constant as possible, as shown in Figure 19.1, when the central frequency,  $f_o$ , is tuned.

### 19.1.2 Variation of Bandwidth

Unfortunately, many designs of the tunable filter are far from expected, in which the bandwidth is varied significantly as shown in Figure 19.2 when the central frequency is tuned. In Figure 19.2(b), at the low end of the frequency range, the bandwidth of the tunable filter is narrow as expected. As the control voltage is increased, the capacitance of the varactor is decreased and hence the central frequency of the tunable filter increases, but, on the other hand, its bandwidth also increases. Up to the high end of the frequency range, the bandwidth is widened to an unacceptable amount, and the “tunable” in “tunable filter” becomes meaningless. In Figure 19.2(c), the variation of bandwidth is just opposite from that in Figure 19.2(b). At the high end of the frequency range, the bandwidth of the tunable filter is narrow as expected. As the control voltage is decreased, the capacitance of the varactor is increased, and hence the central frequency of the tunable filter is lowered,



**Figure 19.1.** Expected constant bandwidth as the control voltage is tuned from low to high voltage. (a) Variation of the varactor capacitance and the central frequency as the control voltage is varied. (b) Central frequency is moved to higher frequency with constant bandwidth as the control voltage is increased.

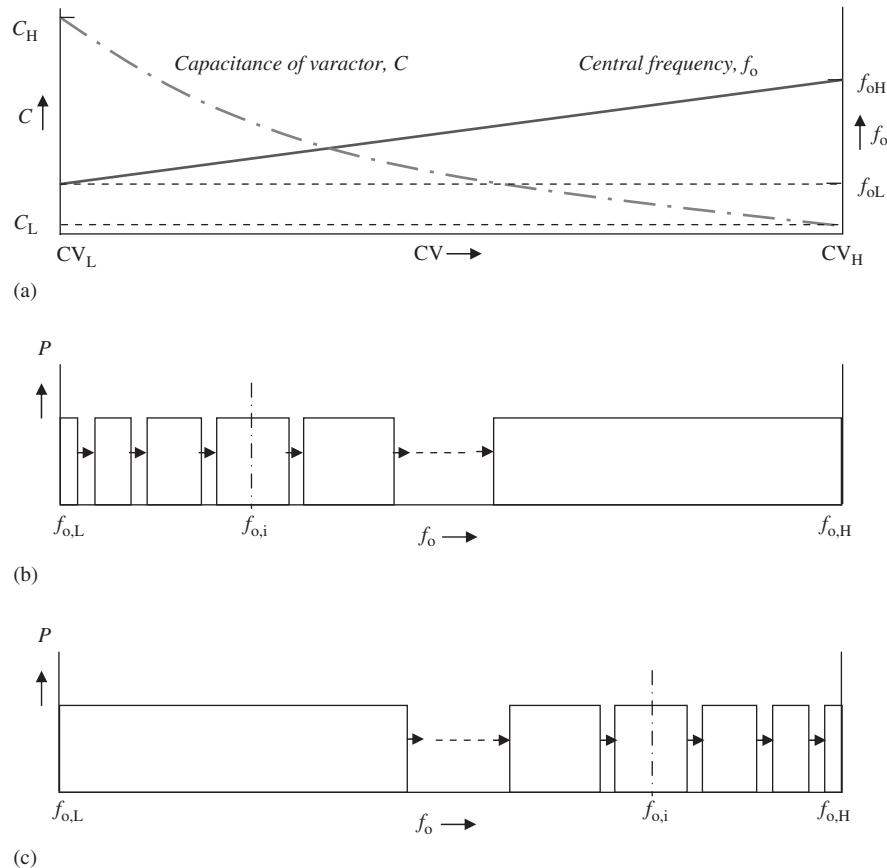
but, on the other hand, its bandwidth is increased. At the low end of the frequency range, the bandwidth is widened to an unacceptable amount, by which the tunable in tunable filter likewise becomes meaningless.

## 19.2 COUPLING BETWEEN TWO TANK CIRCUITS

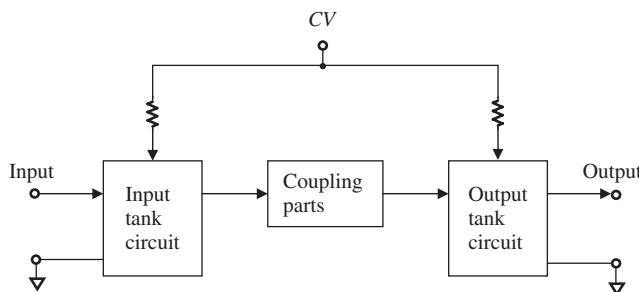
A tunable filter is a BPF with a variable central frequency. Typically, an RF tunable filter in a communication system consists of two tank circuits coupled by one or more coupling parts. Figure 19.3 shows its blocks.

The input and output tank circuits are implemented by an inductor and a capacitor in parallel. In order to tune the central frequency, the capacitor is a special capacitor called a *varactor*, in which its capacitance is varied and tuned by a common control voltage, CV. Both the input and output tank circuits must be resonant at the same frequency and hence must be identical. Should the tunable filter be built by more than one tank circuit, its frequency response would be rolled up and down more sharply, and within the bandwidth of the passband its frequency response would be more flattened. However, the uniform requirement of the tank circuits restricts their number. The more the number of tank circuits, the more difficult it is to maintain their uniformity. This is why most RF tunable filters are implemented by only two tank circuits.

Figure 19.4 shows the schematics of an RF tunable filter with three different coupling styles. Instead of one inductor and one capacitor applied in one tank circuit, the tank circuit consists of two inductors and two varactors. This makes it convenient to apply the control voltage to the varactors and has the additional advantage of impedance matching to the input and output terminals. Figure 19.4(a) shows a tunable filter implemented by two tank circuits and coupled only by a capacitor, and Figure 19.4(b) and (c) shows



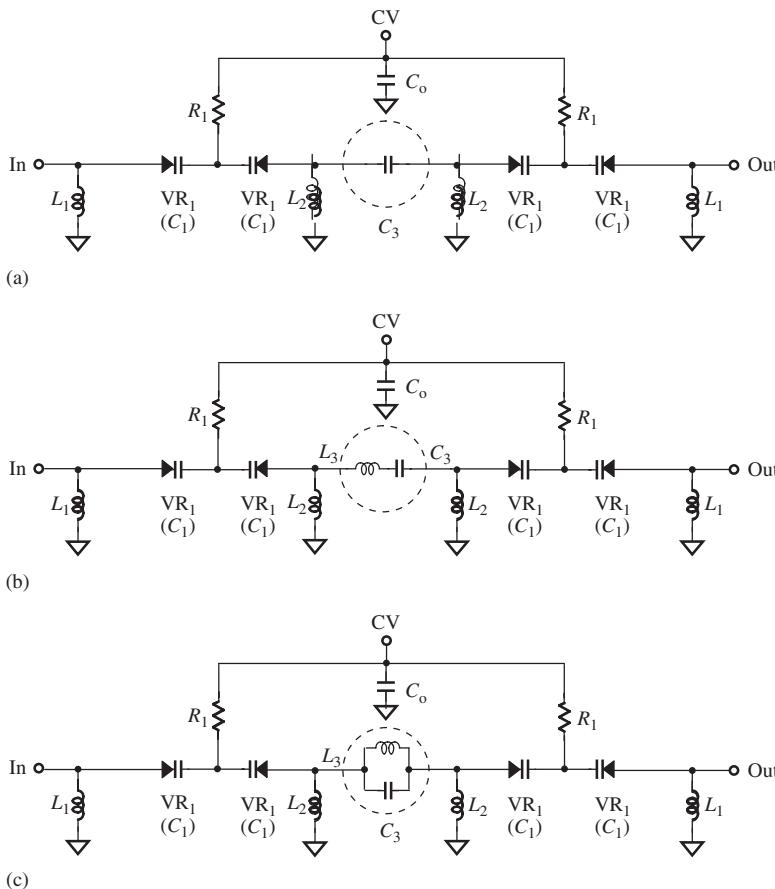
**Figure 19.2.** Constant and varied bandwidth as the control voltage is increased. (a) Variation of the varactor capacitance and the central frequency as the control voltage is varied. (b) Central frequency is moved to higher frequency with wider bandwidth as the control voltage is increased. (c) Central frequency is moved to higher frequency with narrower bandwidth as the control voltage is increased.



**Figure 19.3.** Blocks of a tunable filter.

tunable filters implemented by two tank circuits and coupled by an LC in series and in parallel.

It should be noted that the tank circuits shown in Figure 19.4 have the same topology. The capacitors or varactors of the tank circuit are arranged as an arm in series with input or output terminals while most of the inductors are connected as branches in parallel with



**Figure 19.4.** Three coupling types of a tunable filter. (a) Tunable filter implemented by two tank circuits and coupled by capacitor only. (b) Tunable filter implemented by two tank circuits and coupled by LC in series. (c) Tunable filter implemented by two tank circuits and coupled by LC in parallel.

the input or output terminals. Such a topology leads a frequency response with a higher slope to the lower side of the passband and a response with a lower slope to the higher side of passband. Such a tunable filter is therefore more appropriate to be cooperated with a low-side injection mixer. On the contrary, if the capacitors or varactors of the tank circuit are arranged as a branch in parallel with the input or output terminals while most of inductors are connected as arm in series with the input or output terminals, it would result in a frequency response with a higher slope in the higher side of the passband and a lower slope in the lower side of the passband. Such a tunable filter is more appropriate to be cooperated with a high-side injection mixer. In the following discussion, we focus on the topology appropriate to the low-side injection mixer only.

### 19.2.1 Inappropriate Coupling

The tunable filter with three coupling types as shown in Figure 19.4 was designed in years past. One of their common problems is, as mentioned above, that the bandwidth changes significantly from the low end to high end of the frequency range.

Through a simple circuit analysis one can understand why this is so. As a matter of fact, the tank circuits are always tuned up to resonance at the central frequency. It means that the inductor and the capacitor in the tank circuits are in resonant status so that their reactances are “neutralized” with each other. Therefore, the two tank circuits look like two resistors,  $r_o$ , connected with the main coupling parts in series. The tunable filters with three kinds of main coupling types as shown in Figure 19.4 can be replaced by their equivalents as shown in Figure 19.5 when the tank circuits are resonant at the central frequency.

The bandwidth of the tunable filter is mainly determined by the  $Q$  value of the tank circuit. In the cases when the main coupling part is the capacitor  $C$  only, as in Figure 19.4(a) or 19.5(a),

$$Q = \frac{f_o}{\text{BW}} = \frac{1}{2r_o C \omega_o} = \frac{1}{4\pi r_o C f_o}, \quad (19.1)$$

where

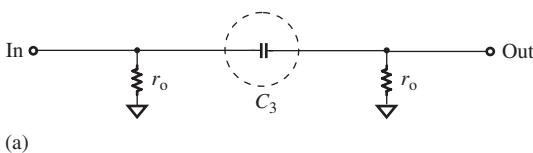
- $Q$  = unloaded  $Q$  of the tunable filter,
- $f_o$  = central frequency of the tunable filter,
- $\omega_o$  = central angular frequency of the tunable filter,
- BW = frequency bandwidth of the tunable filter, and
- $r_o$  = equivalent resistor when tank circuit is resonant at the central frequency.

From (19.1),

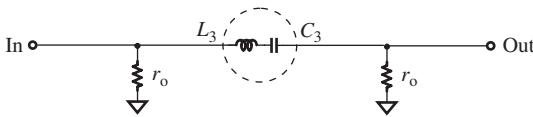
$$\text{BW} = 4\pi r_o C f_o^2, \quad (19.2)$$

Then,

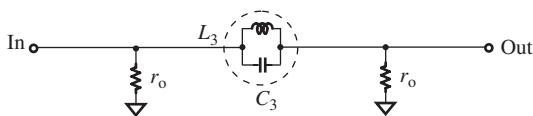
$$\frac{\partial(\text{BW})}{\partial f_o} = 8\pi r_o C f_o. \quad (19.3)$$



(a)



(b)



(c)

**Figure 19.5.** Equivalent circuit of a tunable filter with three coupling types in resonant state when  $f = f_o$ . (a) At  $f_o$ , two tank circuits are in resonant state and coupled by  $C$  only. (b) At  $f_o$ , two tank circuits are in resonant state and coupled by LC in series. (c) At  $f_o$ , two tank circuits are in resonant state and coupled by LC in parallel.

In the cases when the main coupling part is an LC in series as in Figure 19.4(b) or 19.5(b),

$$Q = \frac{f_o}{\text{BW}} = \frac{\frac{L\omega_o}{C} - \frac{1}{C\omega_o}}{2r_o} = \frac{4LC\pi^2 f_o^2 - 1}{4\pi r_o Cf_o}, \quad (19.4)$$

From (19.4),

$$\text{BW} = \frac{4\pi r_o Cf_o^2}{4\pi^2 LC f_o^2 - 1}, \quad (19.5)$$

Then,

$$\frac{\partial(\text{BW})}{\partial f_o} = -\frac{8\pi r_o Cf_o}{(4\pi^2 LC f_o^2 - 1)^2}. \quad (19.6)$$

In the cases when the main coupling part is an LC in parallel, as in Figure 19.4(c) or 19.5(c),

$$Q = \frac{f_o}{\text{BW}} = \frac{\frac{L\omega_o}{1 - LC\omega_o^2}}{2r_o} = \frac{\pi L f_o}{r_o (1 - 4\pi^2 LC f_o^2)}. \quad (19.7)$$

From (19.7),

$$\text{BW} = \frac{(1 - 4\pi^2 LC f_o^2) r_o}{\pi L}. \quad (19.8)$$

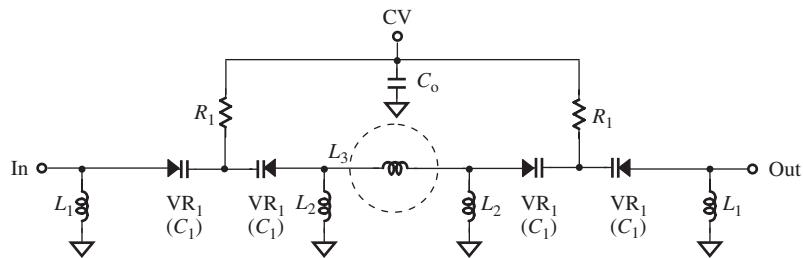
Then,

$$\frac{\partial(\text{BW})}{\partial f_o} = -8\pi r_o Cf_o. \quad (19.9)$$

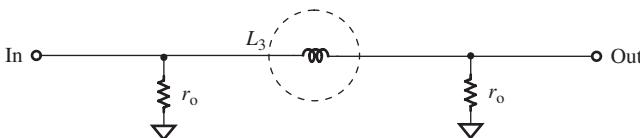
As shown in expressions (19.3), (19.6), and (19.9), in all three coupling types of tunable filters shown in Figure 19.4, the variation of the bandwidth is dependent on the central frequency,  $f_o$ .

In the case when the main coupling part is the capacitor  $C$  only, as shown in Figure 19.4(a) or 19.5(a), the variation of bandwidth is proportional to the central frequency,  $f_o$ . The bandwidth becomes wider and wider as the central frequency increases. It might become unacceptably wide at the high-frequency end as shown in Figure 19.2(b). In the case when the main coupling part is an LC in series as shown in Figure 19.4(b) or 19.5(b), the variation of bandwidth is dependent on the central frequency by a complicated function. In the case when the main coupling part is an LC in parallel as shown in Figure 19.5(c), the variation of bandwidth is negatively proportional to the central frequency. In both the cases shown in Figures 19.4(b) or 19.5(b) and 19.4(c) or 19.5(c), the bandwidth becomes wider and wider as the central frequency is decreased. It might become unacceptably wide at the low-frequency end as shown in Figure 19.2(c).

It is therefore concluded that, generally speaking, all three coupling styles shown in Figure 19.4 are inappropriate in the tunable circuit design.



**Figure 19.6.** Tunable filter implemented by two tank circuits and coupled by  $L$  only.



**Figure 19.7.** At  $f_o$ , two tank circuits are in resonant state and coupled by  $L$  only.

### 19.2.2 Reasonable Coupling

A reasonable coupling approach is shown in Figure 19.6, in which the coupling part between the two tank circuits consists of only an inductor  $L$ . When this filter is operating at its central frequency, its equivalent circuit operates as shown in Figure 19.7.

Similar to the derivation in Section 19.2.1, we have

$$Q = \frac{f_o}{\text{BW}} = \frac{L\omega_o}{2r_o} = \frac{L\pi f_o}{r_o}. \quad (19.10)$$

From (19.10),

$$\text{BW} = \frac{r_o}{L\pi}. \quad (19.11)$$

Then,

$$\frac{\partial(\text{BW})}{\partial f_o} = 0. \quad (19.12)$$

The expression (19.12) leads to an important conclusion: the bandwidth of a tunable filter can be kept unchanged over the entire frequency range if the main coupling parts between the two tank circuits consist of only a pure inductor  $L$ .

## 19.3 CIRCUIT DESCRIPTION

Figure 19.8 shows the schematic of a tunable filter, which will be used with a low-side injection mixer in a UHF portable radio.

The main feature of the circuit is a pair of identical tank circuits coupled by an inductor,  $L_3$ . The simplest tank circuit consists of one inductor and one capacitor in parallel. As shown in Figure 19.8, the tank circuit consists of three inductors, two  $L_1$  inductors and one  $L_2$  inductor, as well as two  $VR_1$  varactors. The purpose of applying three inductors, instead of only one, is twofold: (i) to match the impedance of the input

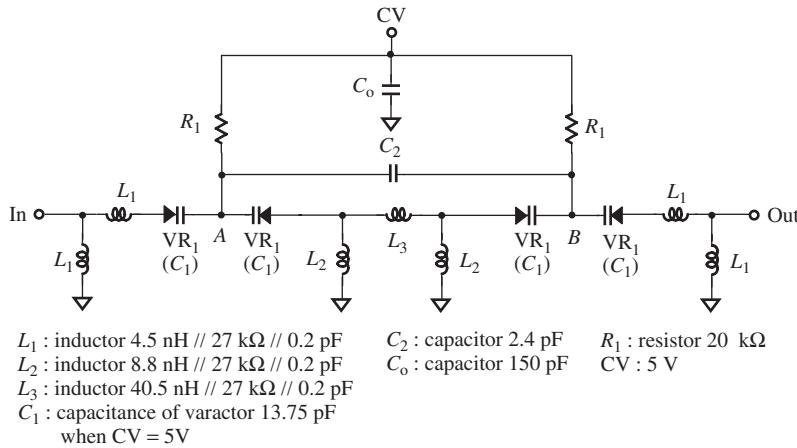


Figure 19.8. Schematic of a UHF tunable filter.

or output, 50  $\Omega$ , without additional impedance matching parts; (ii) to avoid too tight of a coupling between the two tank circuits. In each tank circuit, two identical varactors are piggy-backed together. The tuning function is performed by these varactors. The capacitances of these varactors are controlled by the control voltage, CV, via the resistor  $R_1 = 20$  k $\Omega$ , through which there is no current flowing.

The capacitor  $C_o$  is a “zero” capacitor in the UHF frequency range.

Another remarkable achievement in this tunable circuit design is that there is a second coupling capacitor,  $C_2$ , which creates two “zeros” in the skirt portion of the frequency response plot. One is located below the passband and the other above the passband. These two “zeros” can be adjusted by  $C_2$ ,  $L_2$ ,  $L_3$ , and  $VR_1$ , and can be applied to trace the imaginary spurious products in either the high-side or the low-side injection if the receiver is not operating in the direct conversion mode. In the passband, the effect of the second coupling is negligible. Now, let us analyze how the second coupling works. (Note: A “zero” means that there is a deep notch point in the frequency response curve, which is  $S_{21}$  vs.  $f$  curve usually.)

## 19.4 EFFECT OF SECOND COUPLING

Let us introduce the “ $\pi$ -T” or “ $\Delta$ - $*$ ” transformation of the impedance.

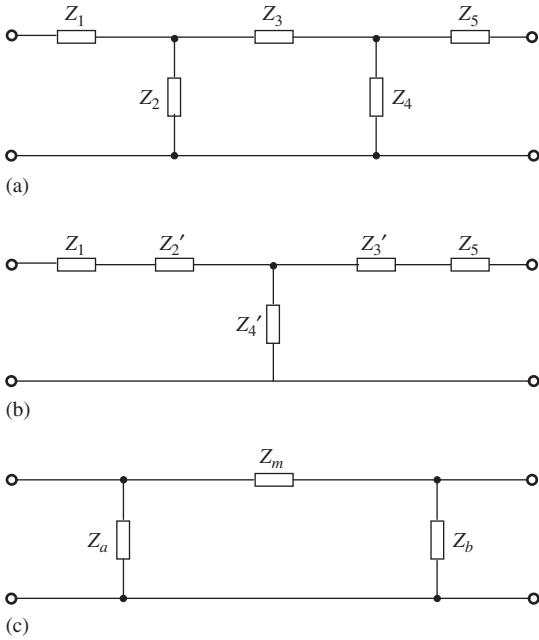
Figure 19.9 shows the process of impedance transformation from the original network (a) to the final network (c).

The corresponding expressions between the impedances shown in Figure 19.9 are

$$Z'_2 = \frac{Z_2 Z_3}{Z_2 + Z_3 + Z_4}, \quad (19.13)$$

$$Z'_3 = \frac{Z_3 Z_4}{Z_2 + Z_3 + Z_4}, \quad (19.14)$$

$$Z'_4 = \frac{Z_4 Z_2}{Z_2 + Z_3 + Z_4}, \quad (19.15)$$



**Figure 19.9.** “ $\pi$ -T” or “ $\Delta^*$ ” transformation of impedance. (a) Original network. (b) Intermediate network. (c) Final network.

and

$$Z_a = Z_1 + Z_2' + Z_4' + Z_4 \frac{Z_1 + Z_2'}{Z_3' + Z_5}, \quad (19.16)$$

$$Z_b = Z_5 + Z_3' + Z_4' + Z_4 \frac{Z_3' + Z_5}{Z_1 + Z_2'}, \quad (19.17)$$

$$Z_m = Z_1 + Z_5 + Z_2' + Z_3' + \frac{(Z_1 + Z_2')(Z_3' + Z_5)}{Z_4'}. \quad (19.18)$$

If

$$Z_1 = Z_5, \quad (19.19)$$

$$Z_2 = Z_4, \quad (19.20)$$

then

$$Z_2' = Z_3' = \frac{Z_2 + Z_3}{2Z_2 + Z_3}, \quad (19.21)$$

$$Z_4' = \frac{Z_2^2}{2Z_2 + Z_3}, \quad (19.22)$$

$$Z_a = Z_b = Z_1 + Z_2, \quad (19.23)$$

$$Z_m = 2Z_1 \left( 1 + \frac{Z_3}{Z_2} \right) + Z_3 \left[ 1 + \frac{Z_1^2}{Z_2^2} \left( 1 + 2 \frac{Z_2}{Z_3} \right) \right]. \quad (19.24)$$

Now let us apply the “ $\pi$ -T” or “ $\Delta^*$ ” transformation of impedance to the network between the nodes of A and B in Figure 19.8. The corresponding impedances are

$$Z_1 = Z_5 \Rightarrow \frac{1}{j\omega C_1}, \quad (19.25)$$

$$Z_2 = Z_4 \Rightarrow j\omega L_2, \quad (19.26)$$

$$Z_3 \Rightarrow j\omega L_3. \quad (19.27)$$

Substituting the relationships (19.25), (19.26), and (19.27) into (19.23) and (19.24), we have

$$Z_a = Z_b = \frac{1}{j\omega C_1} + j\omega L_3, \quad (19.28)$$

$$Z_m = \frac{1}{j\omega C_m} + j\omega L_m, \quad (19.29)$$

where

$$C_m = \frac{C_1}{2\left(1 + \frac{L_3}{L_2}\right)}, \quad (19.30)$$

$$L_m = \left(1 + \frac{1 + 2\frac{L_2}{L_3}}{L_2^2 C_1^2 \omega^4}\right) L_3. \quad (19.31)$$

The equivalent schematic of the tunable filter for UHF portable radio as shown in Figure 19.8 can be depicted as Figure 19.10, which implies two possible zeros in the tunable filter. One is contributed by the series network,  $L_m$ ,  $C_m$ , and  $C_2$ , and the other exists in the parallel branch,  $VR_1(C_1)$  and  $L_2$ . As a matter of fact, the original idea to have the second coupling is due to the recognition of these two zeros.

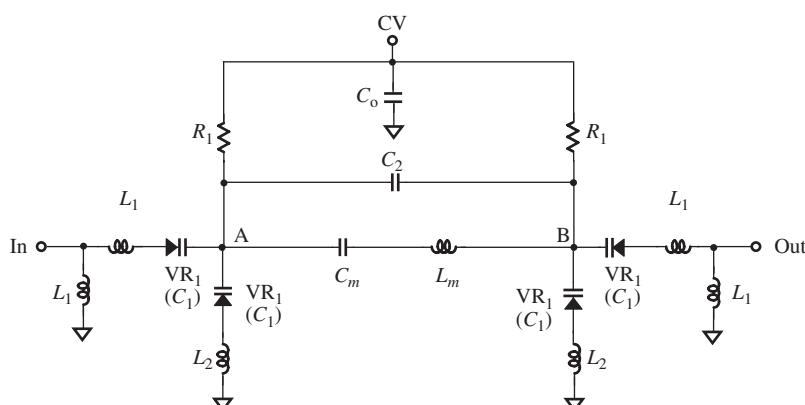


Figure 19.10. Equivalent schematic of a UHF tunable filter.

The equivalent coupling network consists of  $L_m$ ,  $C_m$ , and  $C_2$ . Its maximum impedance corresponds to a zero of the tunable filter at the frequency

$$\omega_{Lo}^2 = \frac{C_2 + C_m}{L_m C_2 C_m}. \quad (19.32)$$

And, it is always lower than the central frequency of the tunable filter,  $\omega_o$ , that is,

$$\omega_{Lo} < \omega_o, \quad (19.33)$$

The notation  $\omega_{Lo}$  is therefore called the *low-side zero frequency* where a zero appears.

In a practical design, the difference between  $\omega_o$  and  $\omega_{Lo}$  is adjusted to about the value of  $2\omega_{IF}$ , so that  $\omega_{Lo}$  could be treated as the imaginary frequency of the RF input,  $\omega_{Lir}$ , when a mixer is operated in a low-side injection mode. If the values of the parts in the tunable filter are selected and adjusted carefully, the difference between  $\omega_o$  and  $\omega_{Lo}$  is kept almost unchanged when the central frequency is tuned from the low-frequency end to the high-frequency end. This is a very useful behavior of the imaginary rejection in cooperation with a mixer design.

Another zero of the tunable filter is created by the equivalent branch in parallel at either node A or node B as shown in Figure 19.10, which is formed by the parts  $C_1$  and  $L_2$ , that is,

$$\omega_{Ho}^2 = \frac{1}{L_2 C_1}. \quad (19.34)$$

It should noted that the frequency  $\omega_{Ho}$  is always higher than  $\omega_o$ , that is,

$$\omega_{Ho} > \omega_o, \quad (19.35)$$

The notation  $\omega_{Ho}$  is therefore called the *high-side zero frequency* where a zero appears.

When the tunable filter is cooperated with a mixer operating in the low-side injection mode, the zero at the frequency  $\omega_{Ho}$  is not helpful to the imaginary rejection. However, it does help narrow the skirt of the entire frequency response curve.

Figure 19.11 compares the frequency responses between two cases with and without second coupling, or in other words, with or without the existence of the capacitor  $C_2$ . Both frequency response curves are tested when  $f_o = 435.43$  MHz.

Without second coupling, the skirt of the frequency response curve is wide open and the corresponding low-side imaginary rejection is poor, that is,

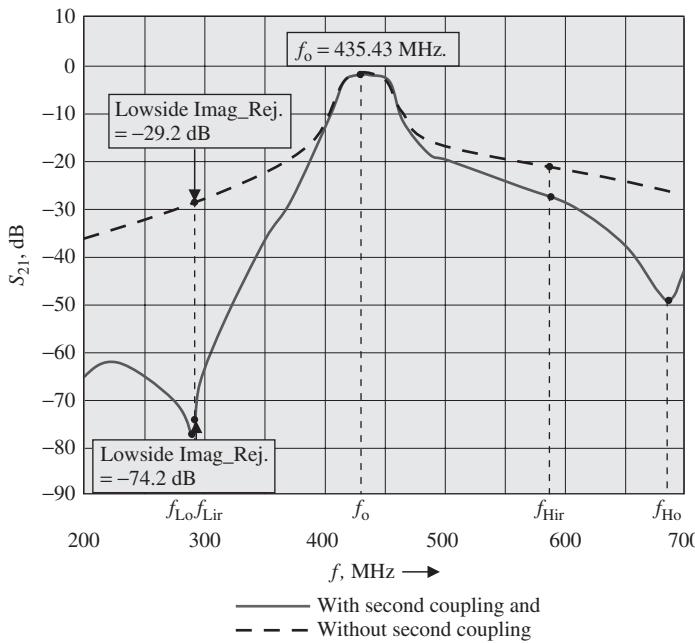
$$\text{Imag\_Rej}_{\text{at low side}} = -29.2 \text{ dB}, \quad (19.36)$$

down from the input signal level.

With second coupling, the skirt of frequency response curve is much narrowed since two zeros are created. The corresponding low-side imaginary rejection is much better, that is,

$$\text{Imag\_Rej}_{\text{at low side}} = -74.2 \text{ dB}, \quad (19.37)$$

down from the input signal level. This is a 45-dB improvement. It is quite an encouraging work since a small capacitor  $C_2$  brings about such a huge advantage!



**Figure 19.11.** Comparison of frequency response of tunable filter.

## 19.5 PERFORMANCE

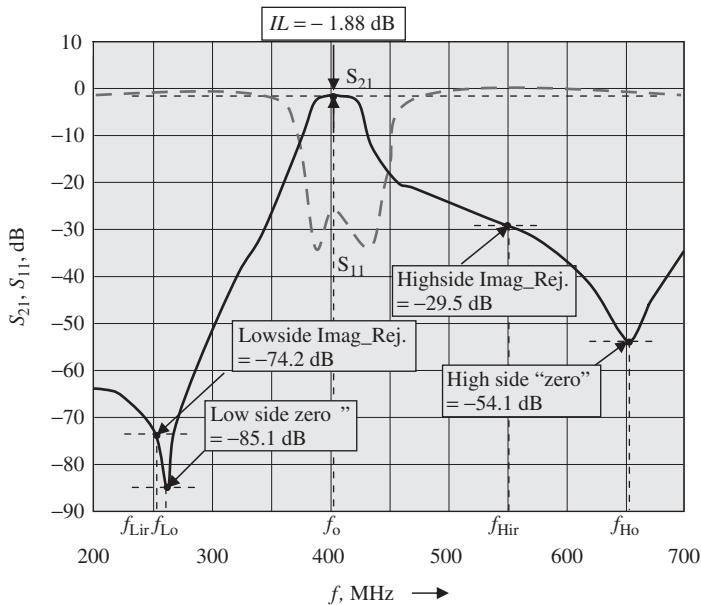
The tested results of the tunable filter for UHF portable radio are presented with four cases when the central frequency is tuned at

- Low end of frequency,  $f_{o1} = 403$  MHz,
- Intermediate frequency 1,  $f_{o2} = \sqrt{f_{o1}f_{o3}} = 435.43$  MHz,
- Intermediate frequency 2,  $f_{o2} = 470$  MHz, and
- High end of frequency,  $f_{o3} = 512$  MHz.

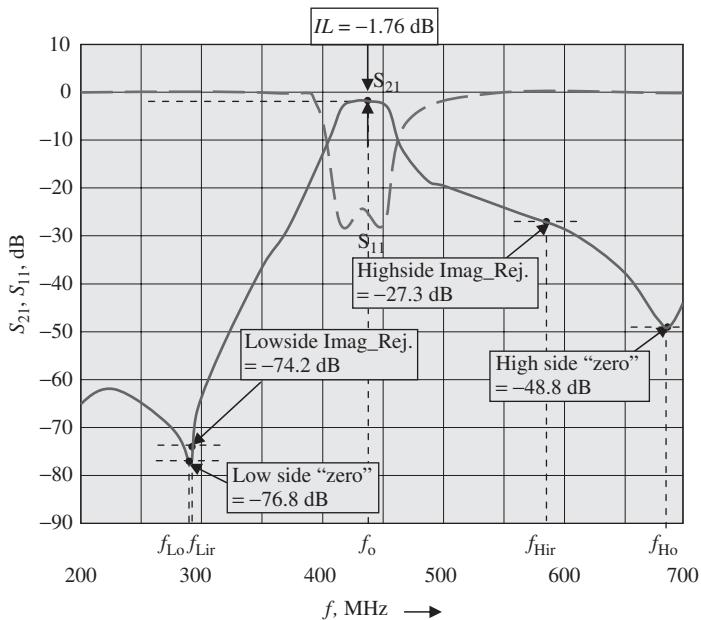
The frequency response of this tunable filter when  $f_0 = 403.00$  MHz is depicted in Figure 19.12. It shows that

- Central frequency,  $f_0 = 403.00$  MHz,
- Bandwidth, BW = 33.5 MHz,
- Insertion loss, IL = -1.88 dB,
- Image rejection applied for low-side injection, when  $f_{IF} = 73.35$  MHz,

Low side:	$\text{Imag\_Rej} = -74.2$ dB,	at $f_{Lir} = 256.3$ MHz,
	“zero” = -85.1 dB,	at $f_{Lo} = 267.5$ MHz.
High side:	$\text{Imag\_Rej} = -29.5$ dB,	at $f_{Hir} = 549.7$ MHz,
	“zero” = -54.1 dB,	at $f_{Ho} = 653.1$ MHz.



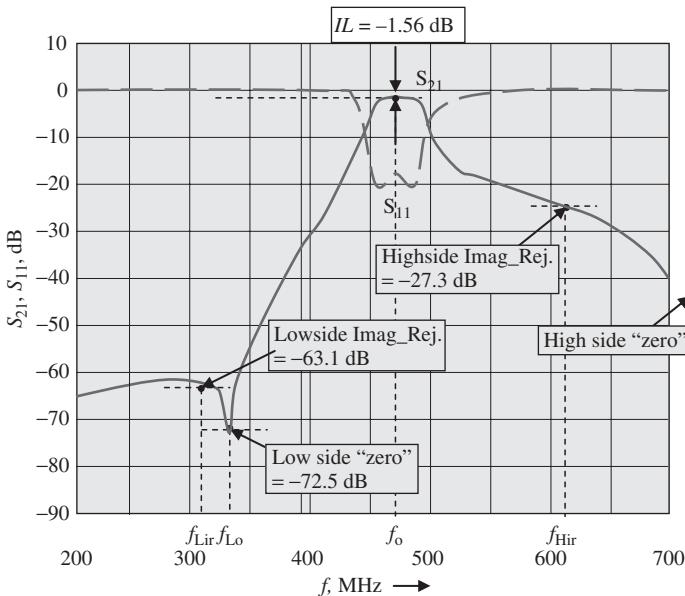
**Figure 19.12.** Frequency response of a tunable filter when  $f_o = 403.00$  MHz.



**Figure 19.13.** Frequency response of a tunable filter when  $f_o = 435.43$  MHz.

The frequency response of this tunable filter when  $f_o = 435.43$  MHz is depicted in Figure 19.13. It shows that

- Central frequency,  $f_o = 435.43$  MHz,
- Bandwidth, BW = 34.3 MHz,
- Insertion loss, IL = -1.76 dB,



**Figure 19.14.** Frequency response of a tunable filter when  $f_o = 470.00$  MHz.

- Image rejection applied for low-side injection, when  $f_{IF} = 73.35$  MHz,

$$\text{Low side: } \text{Imag_Rej} = -74.2 \text{ dB, at } f_{Lir} = 288.73 \text{ MHz,}$$

$$\text{"zero"} = -76.8 \text{ dB, at } f_{Lo} = 267.5 \text{ MHz.}$$

$$\text{High side: } \text{Imag_Rej} = -27.3 \text{ dB, at } f_{Hir} = 582.13 \text{ MHz,}$$

$$\text{"zero"} = -48.8 \text{ dB, at } f_{Ho} = 653.1 \text{ MHz.}$$

The frequency response of this tunable filter when  $f_o = 470.00$  MHz is depicted in Figure 19.14. It shows that

- Central frequency,  $f_o = 470.00$  MHz,
- Bandwidth,  $BW = 35.1$  MHz,
- Insertion loss,  $IL = -1.56$  dB,
- Image rejection applied for low-side injection, when  $f_{IF} = 73.35$  MHz,

$$\text{Low side: } \text{Imag_Rej} = -63.1 \text{ dB, at } f_{Lir} = 323.3 \text{ MHz,}$$

$$\text{"zero"} = -72.5 \text{ dB, at } f_{Lo} = 267.5 \text{ MHz,}$$

$$\text{High side: } \text{Imag_Rej} = -24.8 \text{ dB, at } f_{Hir} = 616.7 \text{ MHz.}$$

"zero" cannot be read because, it is out of the Figure 19.14.

The frequency response of this tunable filter when  $f_o = 512.00$  MHz is depicted in Figure 19.15. It shows that

- Central frequency,  $f_o = 512.00$  MHz,
- Bandwidth,  $BW = 35.8$  MHz,

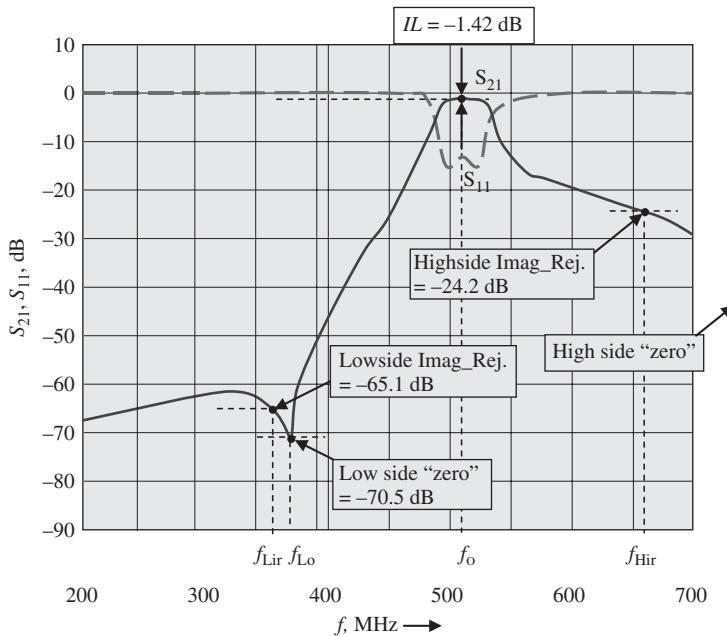


Figure 19.15. Frequency response of a tunable filter when  $f_0 = 470.00$  MHz.

- Insertion loss,  $IL = -1.42 \text{ dB}$ ,
- Image rejection applied for low-side injection, when  $f_{IF} = 73.35 \text{ MHz}$ ,

Low side:       $\text{Imag_Rej} = -65.1 \text{ dB}$ ,      at  $f_{Lir} = 365.3 \text{ MHz}$ ,  
                   "zero" =  $-70.5 \text{ dB}$ ,      at  $f_{Lo} = 377.5 \text{ MHz}$ ,

High side:       $\text{Imag_Rej} = -24.2 \text{ dB}$ ,      at  $f_{Hir} = 658.7 \text{ MHz}$ .  
                   "zero" cannot be read because it is out of the Figure 19.15.

It can be concluded from Figures 19.12–19.15 that when  $403 < f_0 < 512 \text{ MHz}$ ,

1. the insertion loss is low,

$$IL < 1.88 \text{ dB},$$

2. low-side imaginary rejection is excellent if  $f_{IF} = 73.35 \text{ MHz}$ ,

$$\text{Imag_Rej}_{\text{lowside}} < -60 \text{ dB},$$

3. the insertion loss and the imaginary rejection are tunable or traceable.

## FURTHER READING

Beyer WH. *Standard Mathematical Tables*. 24th ed. Cleveland, OH: CRC Press; 1976.

Li RC. Tunable filter having capacitive coupled tuning elements. US patent 5,392,011. Motorola Inc; 1992.

White DRJ. *Electrical Filters*. Gainesville, VA: Don White Consultants, Inc; 1980.

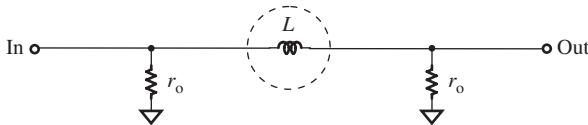
Zverev AI. *Handbook of Filter Synthesis*. New York, NY: John Wiley & Sons Inc; 1967.

## EXERCISES

1. What is the main function of an RF tunable filter in a communication system?
2. How does an RF tunable filter impact a communication system?
3. How do we ensure that the operating frequency in a tunable filter can be tuned linearly as the control voltage is varied?
4. Why is the tunable filter usually constructed by only two tank circuits?
5. Why is the inductor in each tank circuit of the tunable filter cut into two or three pieces?
6. Prove that in order to keep the bandwidth of a tunable filter unchanged when the operating frequency is tuned, the satisfied and necessary condition is that the main coupling part must be an inductor.
7. There is a second coupling part  $C_2$  in the designed tunable filter. Explain how it can create two zeros for the purpose of image rejection.

## ANSWERS

1. The main function of an RF tunable filter in a communication system is to remove many spurious products because of too wide a bandwidth for an individual user. It results in better linearity in a receiver.
2. In a communication system the application of a tunable filter does improve the linearity so that distortion is reduced. However, on the other hand, it increases the noise so that it reduces the sensitivity of receiver. A trade-off between distortion and sensitivity must be taken.
3. To ensure that the operating frequency in a tunable filter can be tuned linearly as the control voltage is varied, the selection of varactors applied in the tunable filter plays an important role. The variation of the varactors' capacitance should be reciprocally proportional to the square of the control voltage.
4. The tunable filter is usually constructed by only two tank circuits. This is due to the difficulty of adjusting more than two tank circuits to the same frequency simultaneously.
5. The inductor in each tank circuit of the tunable filter is cut into two or three pieces. There are two reasons:
  - (a) To fix input and output impedance to  $50 \Omega$ , so that it is not necessary to build input and output impedance matching network;
  - (b) To make the parts' value as similar as possible.
6. In order to keep the bandwidth of a tunable filter unchanged when the operating frequency is tuned, the satisfied and necessary condition is that the main coupling part must be an inductor, because when two tank circuits are adjusted to the operating frequency  $f_0$ , its equivalent circuit is shown in Figure 19.P.1.



$$Q = \frac{f_0}{\text{BW}} = \frac{L\omega_0}{2r_o} = \frac{L\pi f_0}{r_o}$$

$$\text{BW} = \frac{r_o}{L\pi}$$

$$\frac{\partial(\text{BW})}{\partial f_0} = 0$$

Figure 19.P.1. Equivalent circuit of two tank circuits when they are in resonant state.

7. There is a second coupling part  $C_2$  in the designed tunable filter. In order to understand its function, the schematic can be converted to the equivalent as shown in Figure 19.P.2 in terms of “ $T - \pi$ ” transformation. It shows two zeros in the Figure 19.P.2; they are

$$\omega_{Lo}^2 = \frac{C_2 + C_m}{L_m C_2 C_m}; \quad \omega_{Lo} < \omega_0,$$

$$\omega_{Ho}^2 = \frac{1}{L_2 C_1}; \quad \omega_{Ho} > \omega_0,$$

$$Z_m = \frac{1}{j\omega C_m} + j\omega L_m; \quad C_m = \frac{C_1}{2\left(1 + \frac{L_3}{L_2}\right)}; \quad L_m = \left(1 + \frac{1 + 2\frac{L_2}{L_3}}{L_2^2 C_1^2 \omega^4}\right) L_3.$$

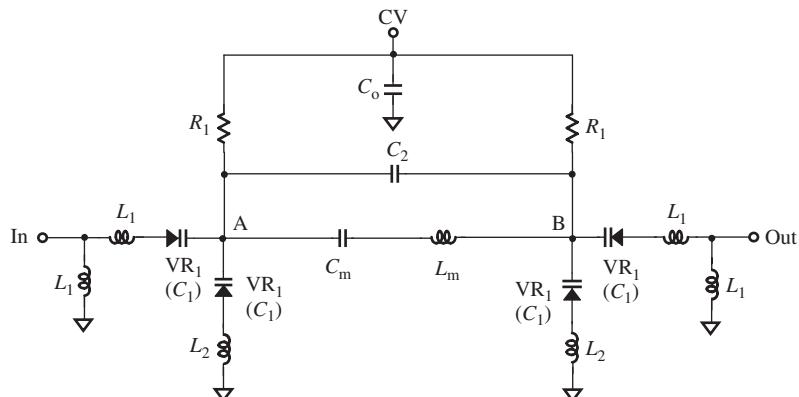


Figure 19.P.2. Equivalent schematic of a UHF tunable filter.

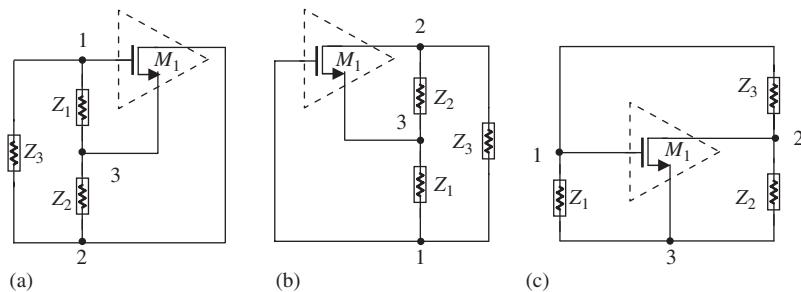
# VCO (VOLTAGE-CONTROLLED OSCILLATOR)

## 20.1 “THREE-POINT” TYPES OF OSCILLATOR

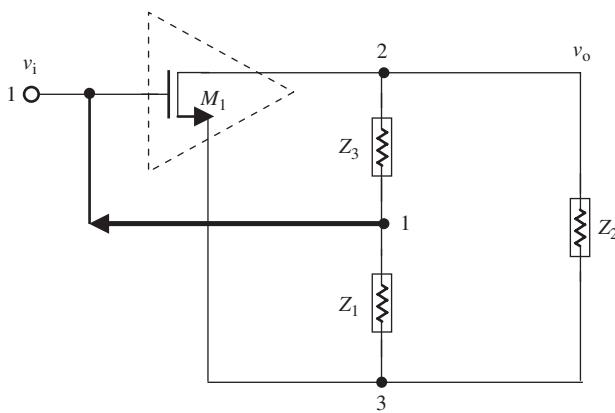
There are many single-ended oscillators in RF circuit design. In this subsection, only the “three-point” type oscillators are discussed. Among the three-point type oscillators, the most popular one applied in communication systems is the Clapp oscillator, which will therefore be discussed in Section 20.1.3 in somewhat detail.

A three-point type oscillator contains only one device, either a bipolar transistor or a MOSFET (metal–oxide–semiconductor field-effect transistor). The AC equivalents of these three points are the base, collector, and emitter for a bipolar transistor, or the gate, drain, and source for a MOSFET. Figure 20.1 shows the AC equivalent of a three-point oscillator built by a MOSFET. The three nodes marked with 1, 2, and 3, are connected to the gate, drain, and source of the MOSFET, respectively. On the other hand, the three parts with impedances  $Z_1$ ,  $Z_2$ , and  $Z_3$  are connected between nodes 1 and 3, 3 and 2, and 2 and 1, respectively.

These three parts form a tank circuit loop. The AC equivalents plotted in Figure 20.1(a)–(c) are exactly the same circuit. The only difference between Figure 20.1(a), (b), and (c) is the location where the tank circuit is drawn; it is at the input side in Figure 20.1(a), at the output side in Figure 20.1(b), and at both the input and output sides in Figure 20.1(c).



**Figure 20.1.** AC equivalent of a three-point type of oscillator. (a) Tank circuit in input side. (b) Tank circuit in output side. (c) Tank circuit in both of input and output side.



**Figure 20.2.** A three-point type oscillator that looks like a CS (common source) amplifier with a feedback from output to input.

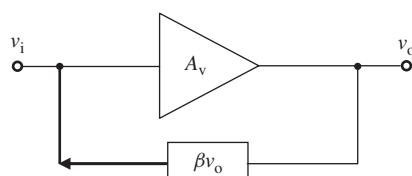
Figure 20.2 shows that this oscillator looks like a CS (common source) amplifier with a feedback factor  $\beta$ , which is

$$\beta = \frac{Z_1}{Z_1 + Z_3}. \quad (20.1)$$

Figure 20.3 represents a voltage amplifier with feedback. The voltage gain  $A_v$  without feedback becomes the voltage gain  $A_f$  with feedback, that is,

$$A_v = \frac{v_o}{v_i}, \quad (20.2)$$

$$A_f = \frac{v_o}{v_i + \beta v_o} = \frac{A_v}{1 + \beta A_v}, \quad (20.3)$$



**Figure 20.3.** Voltage amplifier with feedback.

where

- $v_o$  = the output voltage,
- $v_i$  = the input voltage,
- $A_v$  = the voltage gain without feedback,
- $A_f$  = the voltage gain with feedback, and
- $\beta$  = ratio of feedback voltage to output voltage.

In the tank circuit, the impedance of a loop must be real or its phase shift must be zero, that is,

$$Z_1 + Z_2 + Z_3 = 0. \quad (20.4)$$

Then, from expressions (20.1) and (20.4), we have

$$\beta A_v = A_v \frac{Z_1}{Z_1 + Z_3} = -A_v \frac{Z_1}{Z_2}. \quad (20.5)$$

From equations (20.3) and (20.5), it can be seen that in order to ensure that the amplifier with feedback is an oscillator, the factor  $\beta A_v$  must be negative so that it is an amplifier with positive feedback. Consequently, from equation (20.5) it can be concluded that  $Z_1$  and  $Z_2$  must have the same sign since  $A_v$  is positive. In other words, they must be the same kind of reactance, either both inductive or both capacitive. On the other hand,  $Z_3$  must have a reactance with the opposite sign of the reactance of  $Z_1$  and  $Z_2$ . This was the basic clue in the development of the Hartley, Colpitts, and Pierce oscillators.

### 20.1.1 Hartley Oscillator

If  $Z_1$  and  $Z_2$  are inductors and  $Z_3$  is a capacitor, the circuit is called a *Hartley oscillator*. Figure 20.4 shows the schematic and its AC equivalent circuit.

The coupling between  $L_1$  and  $L_2$  can be described by their coupling coefficient, that is,

$$k_c = \frac{M}{\sqrt{L_1 L_2}}, \quad (20.6)$$

where

- $k_c$  = the mutual coupling coefficient and
- $M$  = the mutual inductance between  $L_1$  and  $L_2$ .

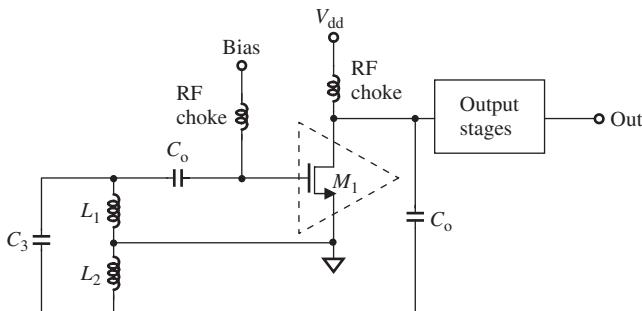
In the case of strong coupling, that is,

$$k_c \approx 1, \quad (20.7)$$

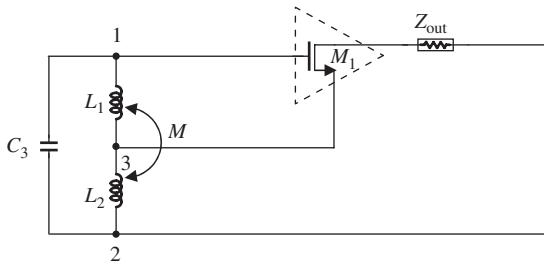
the oscillation frequency can be approximated as

$$f = \frac{1}{2\pi\sqrt{C(L_1 + L_2 + 2M)}}. \quad (20.8)$$

The condition of strong coupling (20.7) is fully satisfied if the two inductors in Figure 20.4 can be replaced by one inductor with an intermediate tap as shown in Figure 20.5. This replacement is beneficial to both cost and part count.

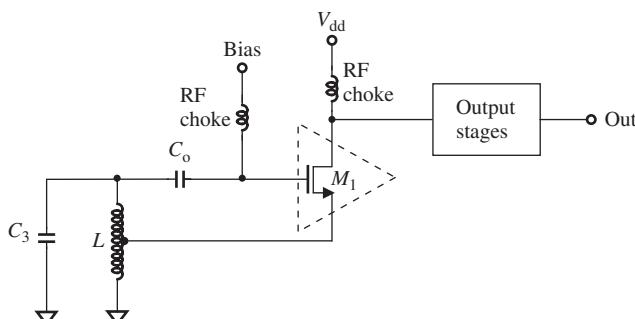


(a)

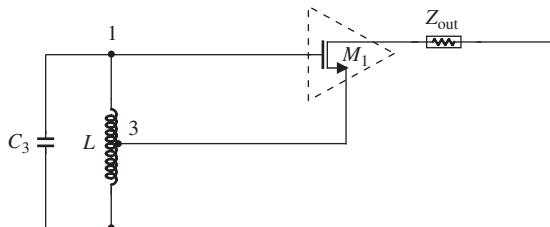


(b)

**Figure 20.4.** Hartley oscillator ( $C_o$ : “zero” capacitor). (a) Schematic. (b) AC equivalent circuit.



(a)



(b)

**Figure 20.5.** Hartley oscillator with an intermediate tapped inductor. (a) Schematic. (b) AC equivalent circuit.

The Hartley oscillator has been applied to RF circuitry from the very early days of radio. One of its advantages is that the oscillation is easily excited. In addition, the two inductors  $L_1$  and  $L_2$  can be combined as one tapped inductor as shown in Figure 20.5. However, in this case the mutual inductance becomes an important factor to the oscillation, so the tapped point of the inductor must be selected carefully.

Today, the Hartley oscillator is much less common. One of the reasons for this is that two inductors or one tapped inductor must be used. Another reason is that its phase noise is higher than that of other types of oscillators.

### 20.1.2 Colpitts Oscillator

If  $Z_1$  and  $Z_2$  are capacitors and  $Z_3$  is an inductor, the circuit is called a *Colpitts oscillator*, which is shown in Figure 20.6.

The oscillation frequency can be approximated as

$$f = \frac{1}{2\pi \sqrt{L \frac{C_1 C_2}{C_1 + C_2}}}. \quad (20.9)$$

Compared to the Hartley oscillator, one of the Colpitts oscillator's advantages is that only one inductor is needed in the tank circuit, instead of two inductors as in the Hartley oscillator. In addition, the phase noise in a Colpitts oscillator is much lower than that in a Hartley oscillator or other oscillators. This is why the Colpitts oscillator has become a popular type of VCO (voltage-controlled oscillator) core today.

### 20.1.3 Clapp Oscillator

If  $Z_1$  and  $Z_2$  are capacitors and  $Z_3$  is an inductor connected with two capacitors in series, the circuit is called a *Clapp oscillator*, which is shown in Figure 20.7.

As a matter of fact, the Clapp oscillator is a modified Colpitts oscillator. The improvement from the Colpitts to the Clapp oscillator is that the frequency bandwidth is more easily covered by adjusting one of the two capacitors in series with the inductor. In

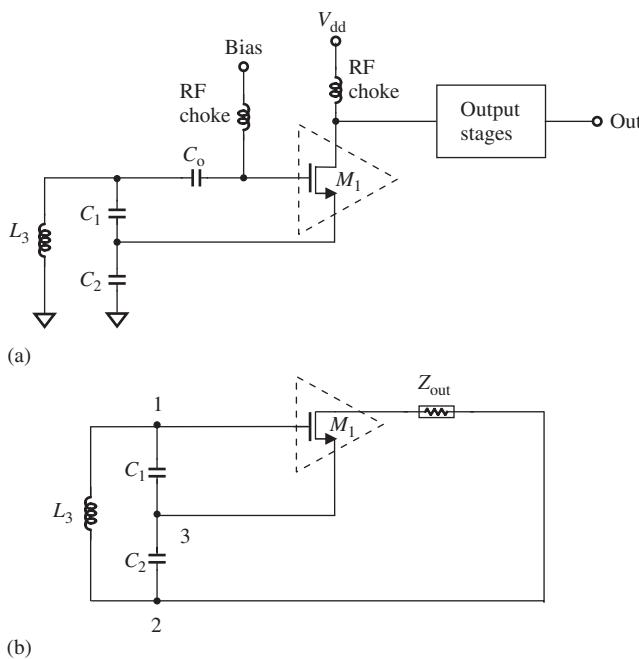
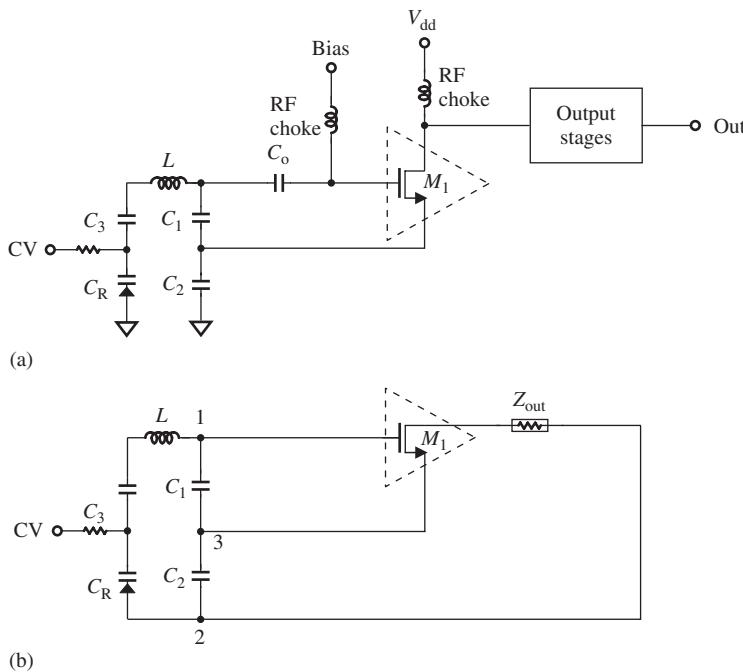


Figure 20.6. Colpitts oscillator. (a) Schematic.  
(b) AC equivalent circuit.



**Figure 20.7.** Clapp oscillator.  
(a) Schematic. (b) AC equivalent circuit.

Figure 20.7,  $C_R$  is an adjustable capacitor called a *varactor*, a capacitor in which the capacitance is adjusted or controlled by a control voltage (CV), so that the Clapp oscillator shown in Figure 20.7 is a VCO.

From Figure 20.7, we have

$$Z_3 = j \left( L_3 \omega - \frac{C_3 + C_R}{C_3 C_R \omega} \right). \quad (20.10)$$

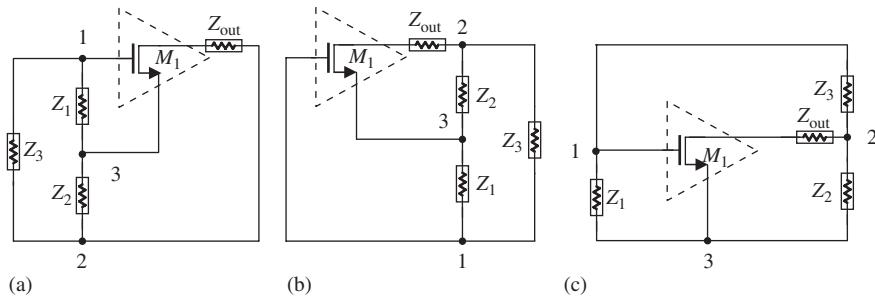
The oscillation frequency can be approximated as

$$f = \frac{1}{2\pi \sqrt{L \left( \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \frac{1}{C_R} \right)}}, \quad (20.11)$$

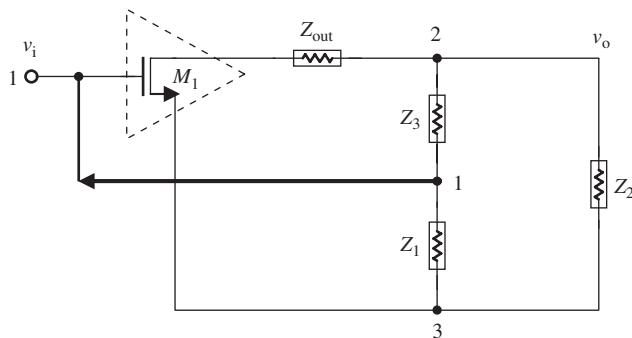
where  $C_R$  is the capacitance of varactor for a given value of the CV.

As a matter of fact, either expression (20.10) or (20.11) is only a rough approximation because the actual inductor  $L$  is a complicated part with stray capacitance.

It should be noted that, in the practical design of the Hartley, Colpitts, and Clapp oscillators, as shown in Figures 20.4–20.7, the drain of the MOSFET is not AC-grounded, as in Figures 20.1 and 20.2. Instead, the output impedance  $Z_{out}$  at the drain of the MOSFET is not zero. Consequently, the AC equivalents of the three-point type oscillators shown in Figures 20.1 and 20.2 must be modified by adding the output impedance at the drain of MOSFET; they then become Figures 20.8 and 20.9, respectively. Figures 20.8 and 20.9 still behave as oscillators although their performances are somewhat different from those shown in Figures 20.1 and 20.2. Additional attenuation or frequency shift may occur in the practical designs shown in Figures 20.8 and 20.9.



**Figure 20.8.** Modified AC equivalent of three-point type of oscillator by adding the output impedance  $Z_{out}$ . (a) Tank circuit in input side. (b) Tank circuit in output side. (c) Tank circuit in both of input and output sides.



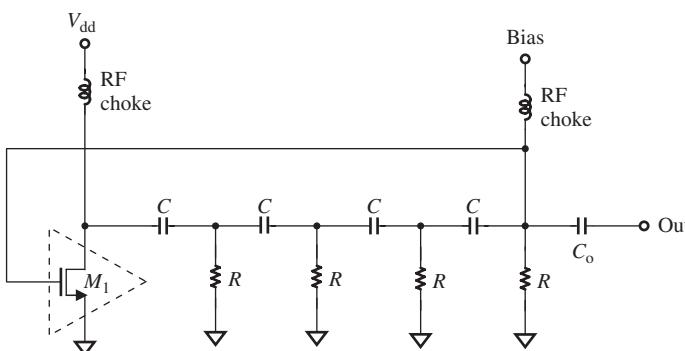
**Figure 20.9.** Modified Figure 20.2 by adding the output impedance  $Z_{out}$  between the gate of the MOSFET and node 2.

## 20.2 OTHER SINGLE-ENDED OSCILLATORS

### 20.2.1 Phase-Shift Oscillator

The simplest RC phase-shift oscillator is shown in Figure 20.10. Its oscillation frequency can be approximated as

$$f = \frac{4}{RC}. \quad (20.12)$$



**Figure 20.10.** A simple RC phase-shift oscillator ( $C_o$ : "zero" capacitor).

In Figure 20.10, only one output is shown. Theoretically, there could be a quad-phase output from each node of the RC branches. However, because of the existence of  $M_1$ , RF chokes, and other additional parts which are not shown in Figure 20.10, the expected quad-phases are not kept at a  $90^\circ$  difference between the RF branches over the frequency bandwidth.

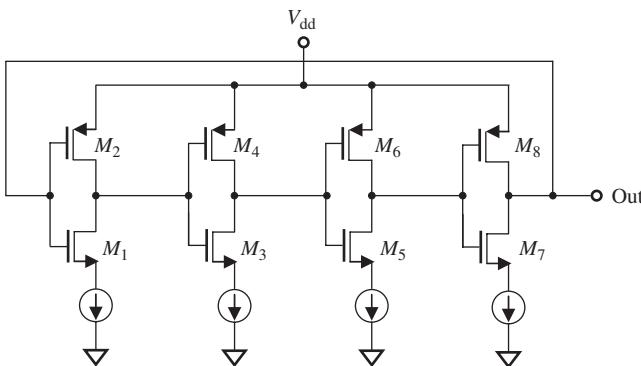
Instead of the RC shift, a simple phase-shift oscillator can be built by transistors only, which is shown in Figure 20.11. There are four stages in series circularly. The phase is shifted  $90^\circ$  in each stage. This oscillator can have quad-phases outputs from each connecting node between stages.

Figure 20.12 shows a differential type of phase-shift oscillator.

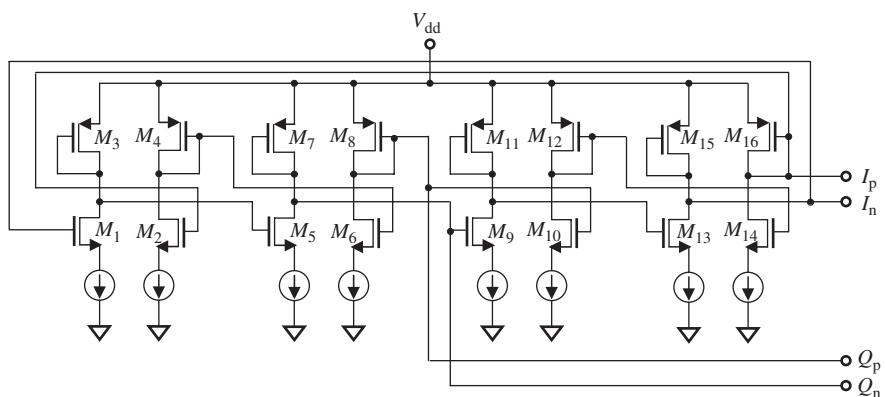
The advantages of the phase-shift oscillator are the following:

1. It has lower current drain;
2. It is considerably smaller in size than an LC VCO;
3. Its quadrature signals are inherently produced from a four-stage ring VCO. Neither a PPF (polyphase filter) nor a frequency divider to produce the quadrature signals is required.

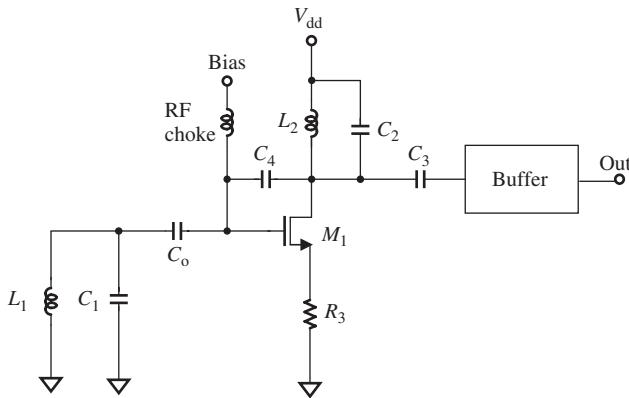
Its disadvantage is that its phase noise is higher than that of the LC VCO.



**Figure 20.11.** A simple phase-shift oscillator built by transistors only (current source is two digits controlled by voltage  $V_c$  from the phase lock loop (PLL)).



**Figure 20.12.** A simple differential phase-shift oscillator (current source is two digits controlled by voltage  $V_c$  from the PLL).



**Figure 20.13.** TITO oscillator ( $C_o$ : "zero" capacitor).

### 20.2.2 TITO (Tuned Input and Tuned Output) Oscillator

Figure 20.13 shows a TITO (tuned input and tuned output) oscillator, in which  $C_4$  is a coupling capacitor. The oscillation frequency can be approximated as the resonant frequency of the input tank or the output tank circuit, that is,

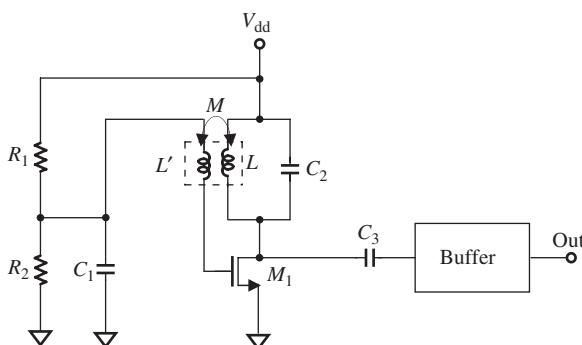
$$f \approx \frac{1}{2\pi\sqrt{L_1 C_1}} \approx \frac{1}{2\pi\sqrt{L_2 C_2}}. \quad (20.13)$$

### 20.2.3 Resonant Oscillator

Figure 20.14 shows a resonant oscillator. As a matter of fact, it is a positive-feedback amplifier. There is a transformer connected between the drain and the gate. Some output power is coupled from the inductance  $L$  of the secondary winding of the transformer by the inductance  $L'$  of the primary winding of transformer through their mutual inductance  $M$ . The feedback power maintains the oscillation of the oscillator.

The oscillation frequency can be approximated as

$$f = \frac{1}{2\pi\sqrt{LC}}. \quad (20.14)$$



**Figure 20.14.** Resonant oscillator.

### 20.2.4 Crystal Oscillator

The key element in a crystal oscillator is a piezoelectric crystal. Figure 20.15 shows its characteristics. The  $Q$  value for a piezoelectric crystal is very high, usually in the range of 5,000–10,000. For a 45-MHz crystal, the value of  $L$  is in the order of 100 mH,  $C$  is in the order of 0.0001 pF, and  $C'$  is in the order of 0.1 pF (which is much higher than the value of  $C$ ). In Figure 20.15(c),

$$jX = -j \frac{1}{\omega C'} \frac{\omega^2 - \omega_s^2}{\omega^2 - \omega_p^2}, \quad (20.15)$$

where

$\omega_s$  = the series resonant angular frequency,

$\omega_p$  = the parallel resonant angular frequency,

and

$$\omega_s^2 = \frac{1}{LC}, \quad (20.16)$$

$$\omega_p^2 = \frac{1}{L} \left( \frac{1}{C} + \frac{1}{C'} \right). \quad (20.17)$$

Usually

$$\omega_s^2 \approx \omega_p^2, \quad (20.18)$$

because

$$C' \gg C. \quad (20.19)$$

Figure 20.16 shows a crystal oscillator in which the crystal is used for  $Z_1$  as shown in Figure 20.2 or 20.9, while the tuned LC combination  $L_2//C_2$  is used for  $Z_2$ , and  $C_4$  is used for  $Z_3$ , which is

$$C_4 = C_{dg} + C_{stray}, \quad (20.20)$$

where

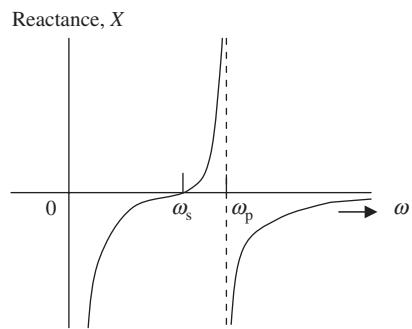
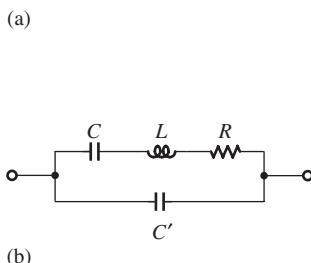


Figure 20.15. Characteristics of a piezoelectric crystal. (a) Symbol. (b) Electrical model. (c) Reactance versus frequency when  $R = 0$ .

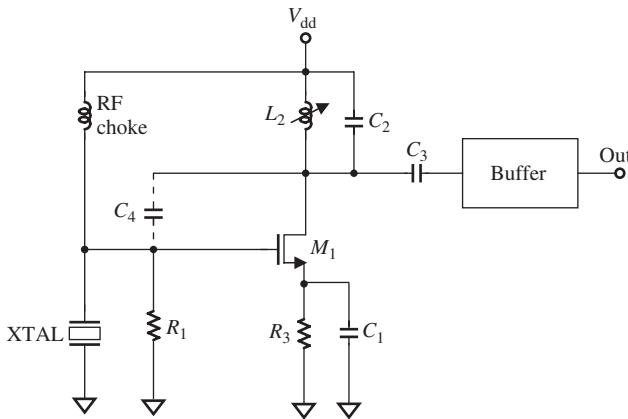


Figure 20.16. Crystal oscillator.

$C_{dg}$  = the capacitance between drain and gate of the MOSFET and

$C_{stray}$  = the stray capacitance existing in the device and layout.

The oscillator frequency is essentially determined by the crystal, and not by the rest of the circuit.

### 20.3 VCO AND PLL (PHASE LOCK LOOP)

The main function of the VCO is to control the frequency of an oscillator by a control voltage, which is applied to a PLL (phase lock loop) in most cases for frequency synthesizer, modulator, demodulator, and other applications. In order to understand the VCO better, it is necessary to introduce the fundamentals of the PLL and the relationship between the VCO and PLL.

#### 20.3.1 Implication of VCO

A VCO is a frequency modulator (see Figure 20.17). The output voltage  $v_{VCO}$  of a VCO can be expressed by

$$v_{VCO}(t) = V_o \sin[\omega_c t + \theta(t)]. \quad (20.21)$$

The frequency deviation of the output,  $d\theta/dt$ , is proportional to the control voltage  $v_c$ , that is,

$$\frac{d\theta(t)}{dt} = K_v v_c(t). \quad (20.22)$$

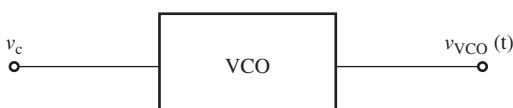


Figure 20.17. Function of a VCO.

Its integral is the phase of VCO:

$$\theta(t) = K_v \int_0^t v_c(x) dx, \quad (20.23)$$

where

$v_{VCO}$  = the output voltage of the VCO,

$V_o$  = the voltage amplitude of the VCO output,

$\omega_c$  = angular frequency of the VCO,

$\theta$  = the phase of the VCO signal,

$K_v$  = the VCO constant, measured in radians per second per unit of input, and

$v_c$  = the control voltage.

### 20.3.2 Transfer Function of PLL

Figure 20.18 shows a basic block diagram of a PLL, which includes a PD (phase detector), a loop filter, a VCO, and a divider.

The PD is a multiplier when the two inputs are small signals. If these two inputs are assumed as

$$v_i(t) = V_s \sin(\omega_i t + \theta_i), \quad (20.24)$$

$$v_o(t) = V_o \cos(\omega_i t + \theta_o), \quad (20.25)$$

where

$v_i$  = the input signal of the PD,

$v_o$  = another input signal of the PD from divider,

$\omega_i$  = the frequency of input signal,

$\theta_i$  = the phase of input signal, and

$\theta_o$  = phase of output signal from divider.

The basic loop equation is

$$v_d(t) = K_m v_i(t) \cdot v_o(t) = \frac{1}{2} K_m V_s V_o \sin(2\omega_i t + \theta_i + \theta_o) + \frac{1}{2} K_m V_s V_o \sin(\theta_i - \theta_o), \quad (20.26)$$

where  $K_m$  is the multiple factor of multiplier with dimensions of  $[V^{-1}]$ .

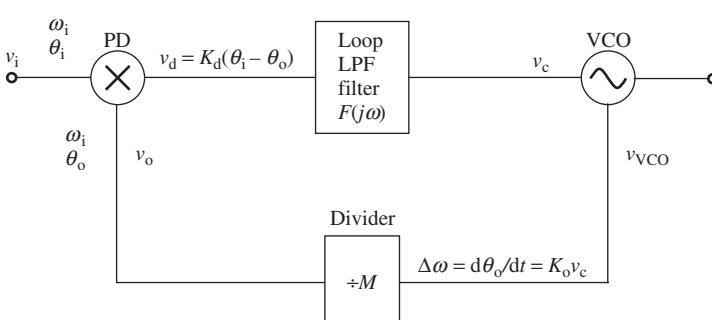


Figure 20.18. Basic block diagram of a PLL (phase lock loop).

Owing to the existence of the loop low-pass filter, the high-frequency term in (20.25) can be neglected, so

$$v_d(t) \approx K_m v_i(t) \cdot v_o(t) = \frac{1}{2} K_m V_s V_o \sin(\theta_i - \theta_o) = K_d \sin(\theta_i - \theta_o), \quad (20.27)$$

where  $K_d = \frac{1}{2} K_m V_s V_o$  is the PD gain factor of the multiplier and is measured in unit of volts per radian V/rad.

The output of the loop BPF (band-pass filter) is

$$v_c(t) = v_d(t)F(j\omega). \quad (20.28)$$

The variation of the phase at the VCO output is linearly related to the control voltage at the VCO input, that is,

$$M \frac{d\theta_o}{dt} = K_o v_c, \quad (20.29)$$

where

$K_o$  = the VCO gain factor and is measured unit of rad  $(s \cdot V)^{-1}$  and

$M$  = the divisor of the divider.

By using Laplace notation, the basic loop equations (20.27), (20.28), and (20.29) become

$$v_d(s) = K_d[\theta_i(s) - \theta_o(s)], \quad (20.30)$$

$$v_c(s) = v_d(s)F(s), \quad (20.31)$$

$$\theta_o = \frac{K_o v_c(s)}{sM}. \quad (20.32)$$

The open-loop transfer function or the open-loop gain is

$$G(s) = \frac{\theta_o(s)}{\theta_e(s)} = \frac{\theta_o(s)}{\theta_i(s) - \theta_o(s)} = \frac{K_o K_d F(s)}{sM}, \quad (20.33)$$

where the phase difference between the two input signals into the PD is denoted by  $\theta_e$ , that is,

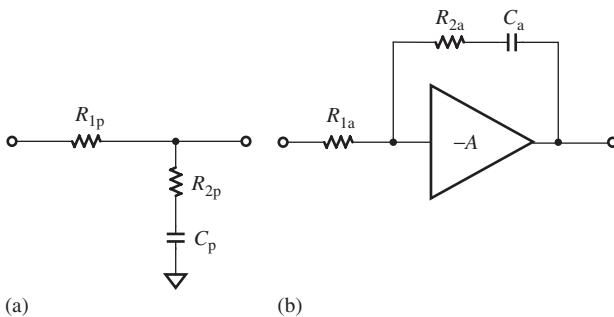
$$\theta_e(s) = \theta_i(s) - \theta_o(s). \quad (20.34)$$

The closed-loop transfer function or closed-loop gain is

$$H(s) = \frac{\theta_o(s)}{\theta_i(s)} = \frac{K_o K_d F(s)}{sM + K_o K_d F(s)} = \frac{G(s)}{1 + G(s)}, \quad (20.35)$$

$$1 - H(s) = \frac{\theta_e(s)}{\theta_i(s)} = \frac{\theta_i(s) - \theta_o(s)}{\theta_i(s)} = \frac{sM}{sM + K_o K_d F(s)}, \quad (20.36)$$

$$v_c(s) = v_d(s)F(s) = K_d \theta_e(s)F(s) = \frac{sMK_d F(s)\theta_i(s)}{sM + K_o K_d F(s)} = \frac{sM\theta_i(s)}{K_o} H(s). \quad (20.37)$$



**Figure 20.19.** Filters applied in a second-order PLL loop. (a) Passive lag-lead filter. (b) active filter.

The low-pass loop filter is very often a passive or an active second-order loop filter. Figures 20.19(a) and (b) show their respective schematics.

The transfer function of the passive filter as shown in Figure 20.19(a) is

$$F_{\text{psv}}(s) = \frac{sC_p R_{2p} + 1}{sC_p(R_{1p} + R_{2p}) + 1} = \frac{s\tau_{2p} + 1}{s\tau_{1p} + 1}, \quad (20.38)$$

where

$$\tau_{1p} = C_p(R_{1p} + R_{2p}), \quad (20.39)$$

$$\tau_{2p} = C_p R_{2p}. \quad (20.40)$$

The transfer function of the active filter as shown in Figure 20.19(b) is

$$F_{\text{atv}}(s) = \frac{-A(sC_a R_{2a} + 1)}{sC_a R_{2a} + 1 + (1 + A)sC_a R_{1a}}. \quad (20.41)$$

If

$$A \gg 1, \quad (20.42)$$

then

$$F_{\text{atv}}(s) = \frac{sC_a R_{2a} + 1}{sC_a R_{1a}} = \frac{s\tau_{2a} + 1}{s\tau_{1a}}, \quad (20.43)$$

where

$$\tau_{1a} = C_a R_{1a}, \quad (20.44)$$

$$\tau_{2a} = C_a R_{2a}. \quad (20.45)$$

The closed-loop transfer functions with passive and active second-loop filters are as follows:

For the passive filter,

$$H_{\text{psv}}(s) = \frac{K_o K_d (s \tau_{2p} + 1) / \tau_{1p}}{s^2 + s(1 + K_o K_d \tau_{2p}) / \tau_{1p} + K_o K_d / \tau_{1p}}; \quad (20.46)$$

for the active filter when condition (20.42) is satisfied,

$$H_{\text{atv}}(s) = \frac{K_o K_d (s \tau_{2a} + 1) / \tau_{1a}}{s^2 + s(K_o K_d \tau_{2a} / \tau_{1a}) + K_o K_d / \tau_{1a}}. \quad (20.47)$$

These two closed-loop transfer functions can be rewritten as

$$H_{\text{psv}}(s) = \frac{s(2\zeta_p \omega_{np} - \omega_{np}^2 / K_o K_d) + \omega_{np}^2}{s^2 + 2\zeta_p \omega_{np} s + \omega_{np}^2}, \quad (20.48)$$

where

$\omega_{np}$  = the natural frequency of the loop, and

$\zeta_p$  = the damping factor.

$$\omega_{np} = \sqrt{\frac{K_o K_d}{\tau_{1p}}}, \quad (20.49)$$

$$\zeta_p = \frac{1}{2} \sqrt{\frac{K_o K_d}{\tau_{1p}}} \left( \tau_{2p} + \frac{1}{K_o K_d} \right), \quad (20.50)$$

$$H_{\text{atv}}(s) = \frac{2\zeta_a \omega_{na} s + \omega_{na}^2}{s^2 + 2\zeta_a \omega_{na} s + \omega_{na}^2}, \quad (20.51)$$

$$\omega_{na} = \sqrt{\frac{K_o K_d}{\tau_{1a}}}, \quad (20.52)$$

and

$$\zeta_a = \frac{\tau_{2a}}{2} \sqrt{\frac{K_o K_d}{\tau_{1a}}} = \frac{\tau_{2a}}{2} \omega_{na}. \quad (20.53)$$

### 20.3.3 White Noise from the Input of the PLL

Assuming that the input  $v_i$  of the PLL, as shown in Figure 20.18, is a sinusoidal signal plus a stationary Gaussian bandpass noise, expression (20.24) is modified as

$$v_i(t) = V_s \sin(\omega_i t + \theta_i) + n(t), \quad (20.54)$$

while expression (20.25) is kept unchanged, that is,

$$v_o(t) = V_o \cos(\omega_i t + \theta_o). \quad (20.55)$$

Expression (20.27) is modified to

$$v_d(t) = K_d[\sin(\theta_i - \theta_o) + n'(t)]. \quad (20.56)$$

If the noise power spectral density of  $n(t)$  is

$$W_n(f) = N_o(V^2/\text{Hz}), \quad (20.57)$$

it can be proved that the normalized spectrum density of  $n'(t)$  is

$$\Phi_{n'}(f) = \frac{2N_o}{V_s^2}(\text{Hz}^{-1}). \quad (20.58)$$

The variance of the output phase is

$$\overline{\theta_{no}^2} = \int_0^\infty \Phi_{n'}(f) |H(j\omega)|^2 df, \quad (20.59)$$

$$\overline{\theta_{no}^2} = \frac{2N_o}{V_s^2} \int_0^\infty |H(j2\pi f)|^2 df = \frac{2N_o}{V_s^2} B_L, \quad (20.60)$$

where the noise bandwidth is defined as

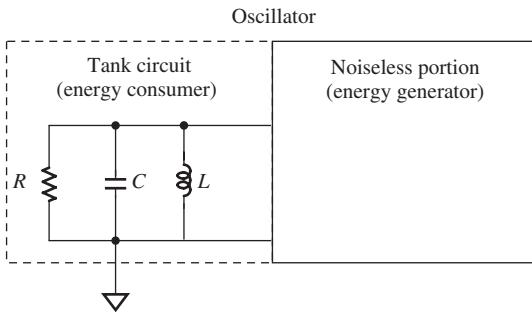
$$B_L = \int_0^\infty |H(j\omega)|^2 df. \quad (20.61)$$

### 20.3.4 Phase Noise from a VCO

The white noise from the input of the PLL or at the terminal  $v_i$  has just been considered, and expression (20.24) has been modified to expression (20.54). The impact on the PLL loop is that there is a phase variance  $\overline{\theta_{no}^2}$ , as shown in expression (20.59), appearing at the output of the divider or at the other input of the PD.

In the derivation of expression (20.60), it was assumed that there is no noise source inside the PLL loop, so that expression (20.25) is unchanged. This is, of course, not true in reality. As a matter of fact, another noise source exists in the PLL loop. In other words, all the blocks shown in Figure 20.18 can produce noise and bring about other phase variances. Among these noise sources, the phase noise from the VCO dominates over the others. Studies on the phase noise produced by the VCO have been conducted for many years.

The oscillation operation can be considered as a process of charging and discharging its tank circuit. The oscillator can be thought of as equivalent to two parts: One portion is the tank circuit, which consists of a resistor  $R$ , a capacitor  $C$ , and an inductor  $L$ . The resistor  $R$  is an equivalent part that causes all the energy loss in the oscillation, so that the tank circuit is an energy consumer portion. In order to maintain oscillation, the other portion of the oscillator must provide the energy to the tank circuit, so that it is called the *energy generator portion*. For simplicity, this portion is assumed to be noiseless because



**Figure 20.20.** Two equivalent portions of the oscillator: energy generator and energy consumer.

all possible resistance in this portion can be converted into an equivalent resistance and added to the value of the resistor  $R$  in the tank circuit. This is a simple model of an oscillator.

The noise produced by  $R$  in the tank circuit is a thermal noise with a mean square spectral density of

$$\frac{\overline{i_n^2}}{\Delta f} = \frac{4kT}{R}, \quad (20.62)$$

where

$\overline{i_n^2}$  = the mean square of thermal noise current of a resistor,

$k$  = the Boltzmann constant,

$T$  = the temperature,

$R$  = the resistance of  $R$ , and

$\Delta f$  = the operating frequency bandwidth.

The impedance of a tank circuit is

$$Z(\omega_0 + \Delta\omega) \approx j \frac{\omega_0 L}{2\Delta\omega/\omega_0}, \quad (20.63)$$

$$\omega_0 = \frac{1}{\sqrt{LC}}, \quad (20.64)$$

if  $\Delta\omega$  is much smaller than  $\omega_0$ , and

where

$Z(\omega_0 + \Delta\omega)$  = the impedance of tank circuit at  $\omega = \omega_0 + \Delta\omega$ ,

$\omega_0$  = the operating angular frequency,

$L$  = the inductance in tank circuit, and

$C$  = the capacitance in tank circuit.

Noting that the unloaded  $Q$  is

$$Q = \frac{R}{L\omega_0}, \quad (20.65)$$

the impedance of the tank circuit can be expressed in terms of  $Q$ , that is, from (20.63) and (20.65) we have

$$Z(\omega_0 + \Delta\omega) = j \frac{\omega_0 R}{2Q\Delta\omega}. \quad (20.66)$$

The spectral density of the mean square noise voltage can be obtained by multiplying the spectral density of the mean square noise current with the squared magnitude of the tank impedance. In terms of (20.62) and (20.66), we have

$$\frac{\overline{v_n^2}}{\Delta f} = \frac{\overline{i_n^2}}{\Delta f} Z^2 = 4kTR \left( \frac{\omega_0}{2Q\Delta\omega} \right)^2. \quad (20.67)$$

The normalized single-sideband noise spectral density is the ratio of the noise power density divided by the power of signal, that is

$$L(\Delta\omega) = \frac{\frac{1}{2R} \frac{\overline{v_n^2}}{\Delta f}}{P_{\text{sig}}} = \frac{2kT}{P_{\text{sig}}} \left( \frac{\omega_0}{2Q\Delta\omega} \right)^2, \quad (20.68)$$

where

$L(\Delta\omega)$  = the normalized single-sideband noise spectral density and  
 $P_{\text{sig}}$  = the power of signal.

Or, in the logarithmic scale with the unit of dB,

$$L(\Delta\omega) = 10 \cdot \log \left[ \frac{2kT}{P_{\text{sig}}} \left( \frac{\omega_0}{2Q\Delta\omega} \right)^2 \right]. \quad (20.69)$$

Figure 20.21 plots  $L(\Delta\omega)$  versus  $\Delta\omega$  in the logarithmic scale based on equation (20.69), which is a straight line. When

$$\Delta\omega = \frac{\omega_0}{2Q}, \quad (20.70)$$

the normalized single-sideband noise spectral density  $L(\Delta\omega)$  is

$$L(\Delta\omega)|_{\Delta\omega=\omega_0/2Q} = 10 \cdot \log \left( \frac{2kTR}{P_{\text{sig}}} \right). \quad (20.71)$$

The plot as shown in Figure 20.21 is usually called the *phase noise plot* because the ordinate represents the normalized single-sideband noise spectral density  $L(\Delta\omega)$  and, instead of the oscillation frequency  $\omega_0$ , the abscissa scaled by  $\Delta\omega$  represents the variation of the oscillation which is controlled by the phase  $\theta_e = \theta_i - \theta_o$  in the PLL.

As shown in Figure 20.21, the theoretical phase noise predicted by equation (20.69) is roughly close to the actual tested phase noise in the region when  $\Delta\omega$  is neither too low nor too high. This is a region where  $L(\Delta\omega)$  is approximately proportional to  $1/(\Delta\omega)$  or  $1/f^2$ . (Note: traditionally, the symbol “ $f$ ” is used in the phase noise plot. As a matter of fact, it is the offset frequency  $\Delta\omega$ .) In the two extreme cases, that is, when  $\Delta\omega$  is quite low or high, the theoretical phase noise predicted by equation (20.69) is far from the values of the actual tested phase noise.

Leeson (1966) modified the phase noise from the simple model expressed by the equation (20.69). According to Leeson’s model, the phase noise can be expressed by

$$L(\Delta\omega) = 10 \cdot \log \left[ \frac{2FkT}{P_{\text{sig}}} \left( 1 + \frac{\omega_0}{2Q\Delta\omega} \right)^2 \left( 1 + \frac{\Delta\omega_{1/f^3}}{|\Delta\omega|} \right) \right], \quad (20.72)$$

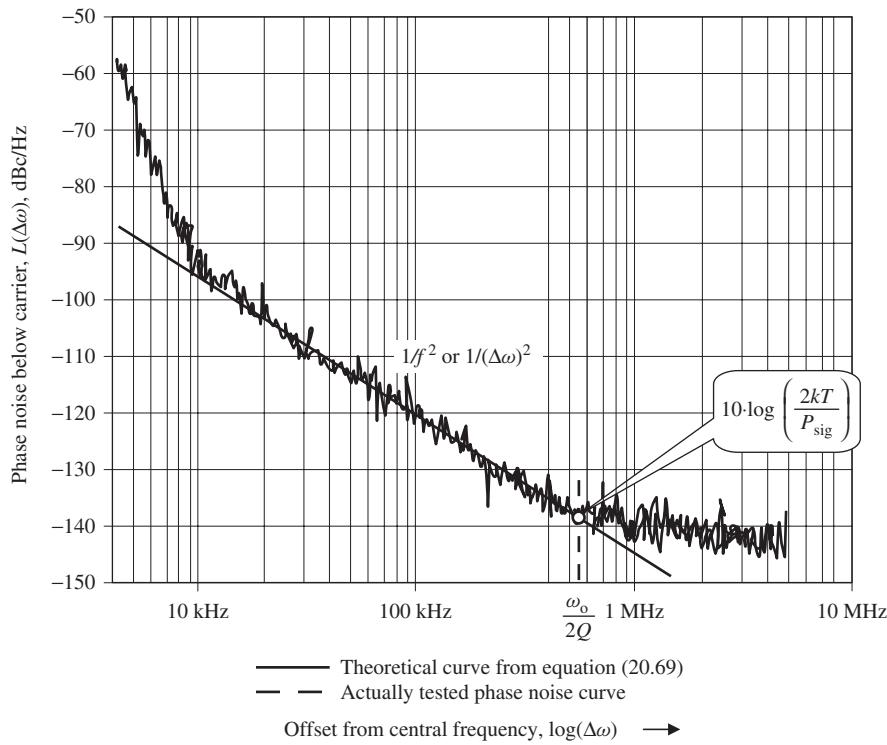


Figure 20.21. Spectrum of the VCO phase noise.

where

$F$  = a factor to account for the increased noise in the  $1/(\Delta\omega)^2$  or  $1/f^2$  region and

$\Delta\omega_{1/f^3}$  = the offset frequency turning point from  $1/(\Delta\omega)^3$  or  $1/f^3$  region to  $1/(\Delta\omega)^2$  or  $1/f^2$  region.

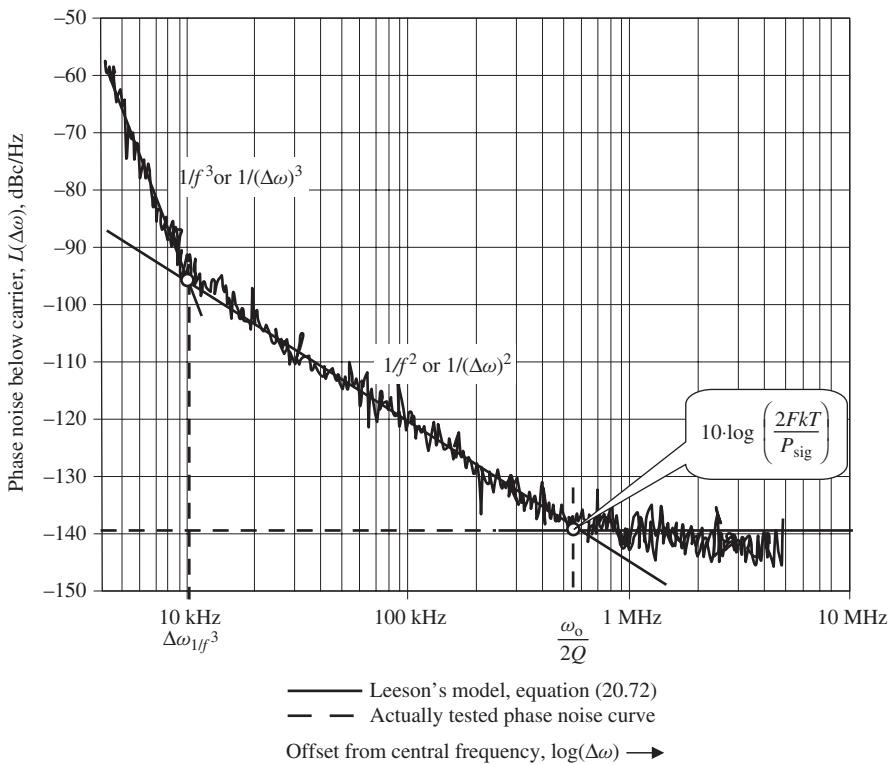
Comparing equation (20.72) with (20.69), there are three differences in equation (20.72):

1. The factor  $F$  is to account for the increased noise in the  $1/(\Delta\omega)^2$  or  $1/f^2$  region.
2. The additive factor of unity in the first parenthesis is to account for the noise floor.
3. The multiplicative factor  $(1 + \Delta\omega_{1/f^3}/|\Delta\omega|)$  is to provide the  $1/(\Delta\omega)^3$  or  $1/f^3$  behavior in the  $1/(\Delta\omega)^3$  or  $1/f^3$  region.

Figure 20.20 plots  $L(\Delta\omega)$  versus  $\Delta\omega$  in logarithmic scale. The solid line segments are asymptotes of equation (20.72).

The first line segment in Figure 20.22 is located in the  $1/(\Delta\omega)^3$  or  $1/f^3$  region. In this region, the flicker noise of the device dominates over other noises. It is well known that the flicker noise is proportional to the reciprocal of the frequency, that is,

$$\overline{i_{n,\text{flick}}^2} \propto \frac{I^a}{f}, \quad (20.73)$$



**Figure 20.22.** Spectrum of the VCO phase noise. (Source: Leeson DB. A simple model of feedback oscillator noise spectrum. Proc. IEEE, February 1966; 54:329–330).

where

- $i_{n,\text{flick}}$  = the random flicker noise current,
- $f$  = the frequency,
- $I$  = the DC flowing through the device, and
- $a$  = a constant in the range 0.5–2.

The noise with  $1/(\Delta\omega)^2$  or  $1/f^2$  behavior shown in equation (20.69) is converted to the noise with the  $1/(\Delta\omega)^3$  or  $1/f^3$  behavior. This is why the multiplicative factor  $(1 + \Delta\omega_{1/f^3}/|\Delta\omega|)$  is added in Leeson's equation (20.72).

The second line segment in Figure 20.22 is located in the  $1/(\Delta\omega)^2$  or  $1/f^2$  region. It is the same as that in the simple model described by equation (20.69).

The third line segment in Figure 20.22 is located in a flat region. It represents the noise floor of the oscillator.

As shown in Figure 20.22, the actual tested phase noise curve can be approximated by Leeson's model. In recent years, efforts have been made to theoretically study the factor  $F$  and the offset frequency turning point  $\Delta\omega_{1/f^3}$  in equation (20.72).

Much effort has been placed on the reduction of the phase noise of the oscillation because the phase noise directly impacts the performance of the PD and thus the entire PLL loop. Phase noise can cause the PLL to be locked for an unacceptably long time,

or the PLL being locked at unexpected frequencies, or, in extreme cases, put the PLL out of functioning altogether. Some special schemes have been reported to reduce phase noise, such as by means of an on-chip filter or off-chip low-frequency inductor, modify the oscillator topology, and so on.

## 20.4 DESIGN EXAMPLE OF A SINGLE-ENDED VCO

### 20.4.1 Single-Ended VCO with Clapp Configuration

This is a practical VCO design example with discrete parts that has been applied to a communication system in the UHF (ultrahigh frequency) range, from 440 to 470 MHz. Most of the design considerations and schemes are universally applicable to today's RFIC design, even though the VCO in this example is implemented by discrete parts.

Figure 20.23 shows the schematic of the VCO with Clapp configuration. The tank circuit consists of  $C_{R1}$ ,  $C_1$ ,  $L_2$ ,  $C_2$ , and  $C_3$ , and the varactor  $C_{R1}$  is controlled by the control voltage. At the output portion, the  $\Pi$  type of Chebyshev filter consists of  $C_4$ ,  $L_5$ , and  $C_5$  and is inserted between the collector of the bipolar transistor and the buffer. The inductors  $L_3$  and  $L_4$  adjust the input and output impedance, while the resistor  $R_4$  partially plays the self-bias role. The primary goals are listed in Table 20.1.

In order to obtain a better coverage of the frequency bandwidth, one must be careful to measure, test the characteristics of the parts in the tank circuit, and then evaluate the frequency bandwidth through the adjustment of the parts' value.

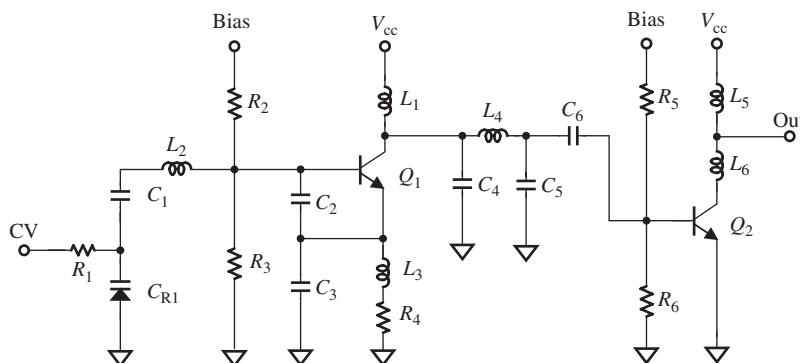
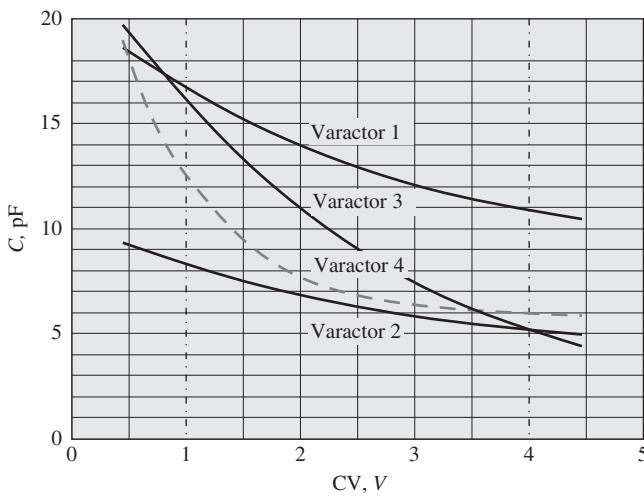


Figure 20.23. Single-ended VCO with Clapp configuration.

TABLE 20.1. Primary Goals of Single-Ended VCO with Clapp Configuration

Frequency range	440–470 MHz
DC power supply	5 V
Current drain	<15 mA
Output power	>12 dB <sub>m</sub> (16 mW)



**Figure 20.24.** Capacitance versus control voltage of a varactor.

### 20.4.2 Varactor

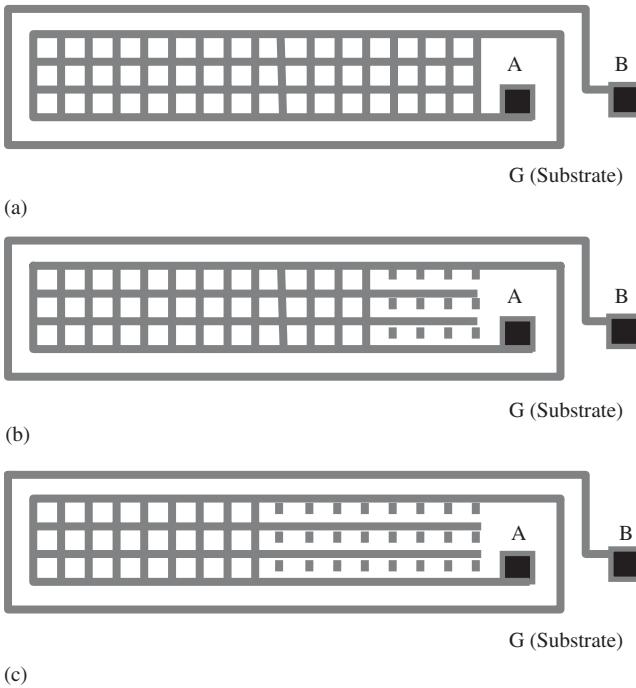
The capacitance of a varactor changes as its control voltage is varied. As shown in Figure 20.24, the capacitance of a varactor is high when the control voltage is low and vice versa. The operating frequency of the VCO is therefore controlled by the control voltage through the variation of the varactor's capacitance.

Generally, the characteristics of a varactor are functions of the frequency. Therefore, the variation of capacitance versus the control voltage must be carefully measured and evaluated as the frequency is varied. Fortunately, it has been found that such a frequency dependence for actual varactors is very small within the range of experimental error. The values shown in Figure 20.24 are applicable for the frequency range to be covered. In other words, within experimental error, they are independent of the frequency.

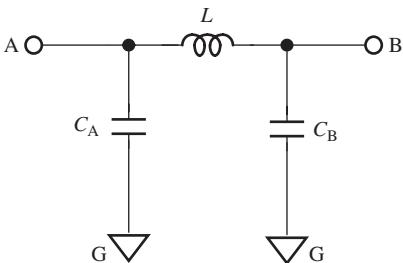
The test results for four varactors are shown in Figure 20.24. Varactor 1 might be applied for cases when the VCO is operated in the low-frequency range, since its capacitances in the range of control voltages are higher, while varactor 2 might be applied for cases where the VCO is operated in the high-frequency range since its capacitances in the range of control voltages are relatively low. Varactor 3 seems to be overall the best candidate because its variation of capacitance versus of control voltage,  $\Delta C / \Delta(CV)$ , is higher than that of either varactor 1 or 2. On the contrary, varactor 4 is a bad candidate because its capacitance versus of control voltage,  $\Delta C / \Delta(CV)$ , changes too much over the range of the control voltage. The slope  $\Delta C / \Delta(CV)$  is too steep when the control voltage is low, whereas it is too flat when the control voltage is high.

### 20.4.3 Printed Inductor

For the sake of cost saving, printed inductors are adapted for  $L_2$ ,  $L_4$  to  $L_6$ . Their configurations are shown in Figure 20.25. Their inductances can be adjusted by trimming the grill portion of the printed inductor. Figure 20.25(a) shows the primary printed inductor before trimming. Figures 20.25(b) and 20.25(c) show the printed inductor after it is



**Figure 20.25.** Printed inductors before and after trimming. (a) Printed inductor before trimming. (b) Trimming for low-frequency range. (c) Trimming for high-frequency range.



**Figure 20.26.** Equivalent circuit of a printed inductor.

trimmed for cases when the VCO operates in low- and high-frequency ranges, respectively. Its  $Q$  value is approximately 60–90, which is lower than that of an air coil but much higher than that of a spiral inductor in an RFIC. Figure 20.25 shows the printed inductors before and after the trimming is done.

Measuring the inductance of a printed inductor in the laboratory is a problem not only in theory but also in practice. At present, some calculation methods are available, but only for printed inductors of symmetrical or uniform geometry and not for irregular or unsymmetrical printed inductors. It has been shown that any printed inductor can be modeled by the equivalent circuit as shown in Figure 20.26. Instead of distributed parameters, an equivalent circuit with lumped parameters is preferred.

The values of the parts in the equivalent circuit in Figure 20.26 can be obtained by measuring the impedance three times.

The measured impedance  $C'_A$  between the nodes A and G is contributed by two branches in parallel. One branch is the capacitor  $C_A$ . Another branch is  $C_{L+C_B}$ , the

inductor  $L$  connected with the capacitor  $C_B$  in series:

$$C_{L+C_B} = \frac{C_B}{1 - \omega^2 L C_B}. \quad (20.74)$$

Then, the measured  $C'_A$  should be

$$C'_A = C_A // C_{L+C_B} = C_A + \frac{C_B}{1 - \omega^2 L C_B}. \quad (20.75)$$

Similarly, the measured capacitance  $C'_B$  between the nodes B and G is contributed by two branches in parallel. One branch is the capacitor  $C_B$ . Another branch is  $C_{L+C_A}$ , the inductor  $L$  connected with the capacitor  $C_A$  in series:

$$C_{L+C_A} = \frac{C_A}{1 - \omega^2 L C_A}. \quad (20.76)$$

Then, the measured  $C'_B$  should be

$$C'_B = C_B // C_{L+C_A} = C_B + \frac{C_A}{1 - \omega^2 L C_A}. \quad (20.77)$$

The measured inductance  $L'$  between nodes A and B is contributed by two branches in parallel. One branch is the inductor  $L$ . The other branch is  $C_A + C_B$ , the capacitor  $C_A$  connected with the capacitor  $C_B$  in series:

$$C_{C_A+C_B} = \frac{C_A C_B}{C_A + C_B}. \quad (20.78)$$

Then, the measured  $L'$  should be

$$L' = L // (C_A + C_B) = L \left( 1 + \omega^2 L \frac{C_A C_B}{C_A + C_B} \right). \quad (20.79)$$

By means of some simple but tedious algebraic manipulations, from expressions (20.74) to (20.79), we have

$$C_A = \frac{2C'_A C'_B (\omega^2 L' C'_A C'_B + C'_B - C'_A)}{(\omega^2 L' C'_A C'_B + C'_B - C'_A)^2 + 4\omega^2 L' C'_A C'_B}, \quad (20.80)$$

$$C_B = \frac{2C'_A C'_B (\omega^2 L' C'_A C'_B - C'_B + C'_A)}{(\omega^2 L' C'_A C'_B + C'_B - C'_A)^2 + 4\omega^2 L' C'_A C'_B}, \quad (20.81)$$

$$L = \frac{(\omega^2 L' C'_A C'_B + C'_B - C'_A)^2 + 4\omega^2 L' C'_A C'_B}{2\omega^2 C'_A C'_B (\omega^2 L' C'_A C'_B + C'_B + C'_A)^2}. \quad (20.82)$$

The value of the parts in the equivalent circuit,  $C_A$ ,  $L$ , and  $C_B$ , can be calculated from the measured values  $C'_A$ ,  $L'$ , and  $C'_B$  by means of expressions (20.80)–(20.82).

Expressions (20.80) to (20.82) are applicable to the spiral inductor implemented in the RFIC as long as the spiral inductor can be measured or trimmed just like the printed inductor shown in Figure 20.26.

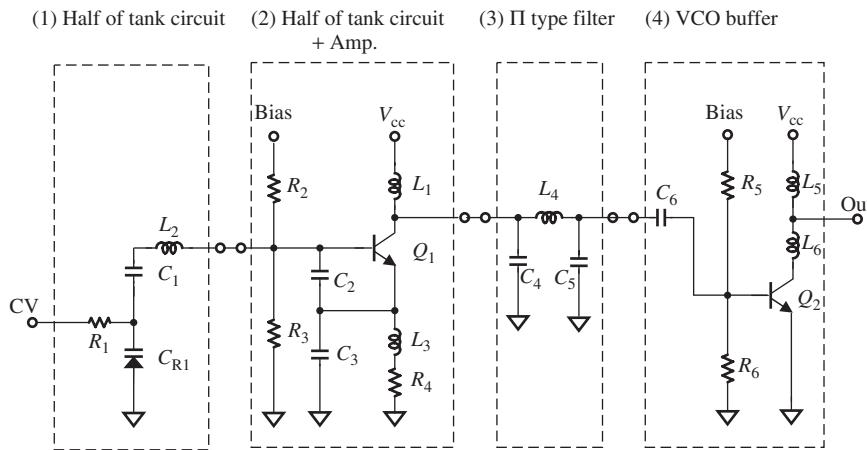


Figure 20.27. Subblocks of a single-ended single device VCO with Colpitts configuration.

#### 20.4.4 Simulation

Simulation for the VCO is somewhat different from the simulation for other individual blocks. Instead a simulation for the entire block, the VCO block is divided into four portions as shown in Figure 20.27:

1. Half of tank circuit
2. Another half of tank circuit plus amplifier
3.  $\Pi$ -type Chebyshev filter
4. VCO buffer.

The main purpose of dividing the entire VCO block into subblocks is to examine the optimized oscillation condition between the tank circuit and the rest of the circuitry. The additional purpose is to optimize the performance of each subblock more carefully and in detail.

The tank circuit consists mainly of  $C_{R1}$ ,  $C_1$ ,  $C_2$ ,  $C_3$ , and  $L_2$ . However, in order to maintain the subblocks with only one input port and one output port so as to apply the  $S$  parameter testing for a two-port block, the tank circuit is divided into two halves. The first half consists of  $C_{R1}$ ,  $C_1$ , and  $L_2$ . The second half consists of  $C_2$  and  $C_3$  and is combined with the bipolar transistor and other parts, which function as an amplifier element and form the second subblock.

The first simulation run is to optimize the values of the parts in subblock 2 as shown in Figure 20.28. The goals are

1. to make  $S_{11} = S_{11,\max}$  and
2. to get (phase of  $S_{11}$ ) when  $S_{11} = S_{11,\max}$ .

The second simulation run is to optimize the values of the parts in subblock 1 as shown in Figure 20.29. The goals are

1. to make (phase of  $\Gamma$ ) = -(phase of  $S_{11,\max}$ ) and
2. to have  $|\Gamma| > |1/S_{11,\max}|$ ,

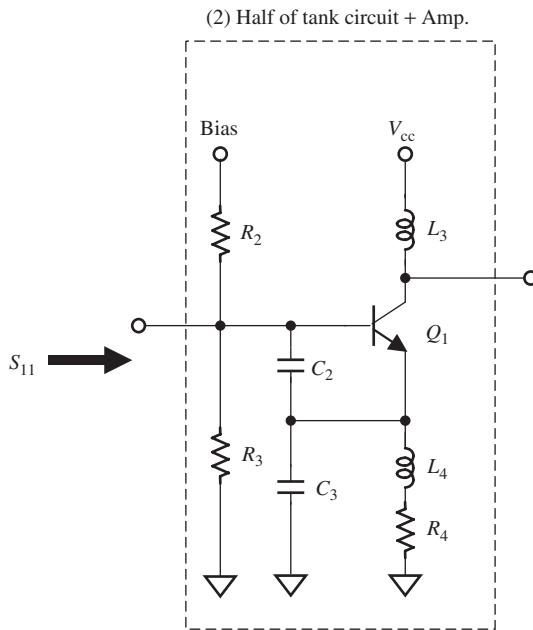


Figure 20.28. Optimization of the subblock 2 for  
 $S_{11} = S_{11,\max}$ .

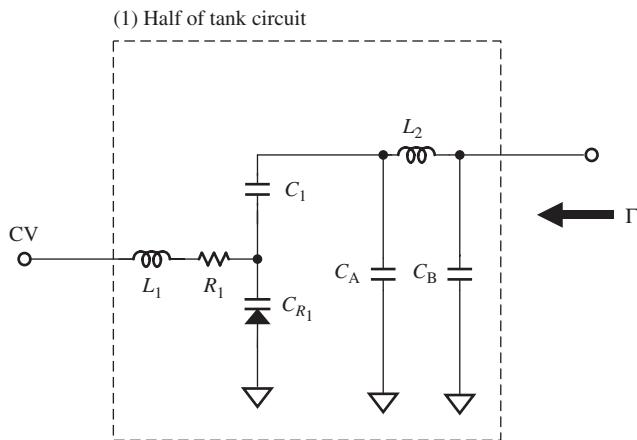


Figure 20.29. Optimization of the subblock 1  
for  $S_{11} = S_{11,\max}$  (where  $C_A$  and  $C_B$  are additional  
capacitors of the printed inductor  $L_2$  as shown in  
Figure 20.26).

where

$S_{11,\max}$  = obtained in the first simulation as shown in Figure 20.28 and  
 $\Gamma$  = the voltage reflection coefficient as shown in Figure 20.29.

It should be noted that in Figure 20.29 the printed inductor  $L_2$  is replaced by its equivalent circuit as shown in Figure 20.26, where the additional capacitors  $C_A$  and  $C_B$  are added.

Also, it should be noted that the range of the control voltage must be set so that either the tank circuit or the amplifier stage can be operated over the desired frequency bandwidth. In other words, in the first and second simulation steps for subblocks 1 and 2, simulations must be conducted for three control voltages:  $CV_{\max}$ ,  $CV_o$ , and  $CV_{\min}$ .

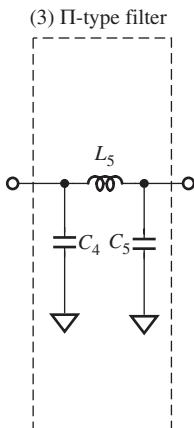


Figure 20.30. Optimization of subblock 3 for impedance matching.

These three control voltages correspond to three frequencies  $f_{\min}$ ,  $f_0$ , and  $f_{\max}$ , which cover the desired frequency bandwidth.

The third simulation run is to optimize the values of the parts in subblock 3 as shown in Figure 20.30. The goals are

1. to obtain high input impedance, so that it is a light load of the subblock 2;
2. to match its output to the input of subblock 4;
3. to filter the unexpected harmonics as much as possible.

The fourth simulation run is to optimize the values of the parts in subblock 4 as shown in Figure 20.31. The goals are

1. to match the output of subblock 4 to  $50 \Omega$  so that its power output reaches the maximum; and
2. to see if its frequency response is flattened over the desired bandwidth.

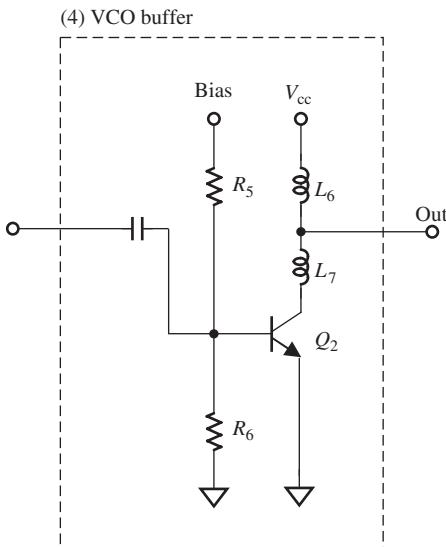


Figure 20.31. Optimization of subblock 4 for maximum of output.

The final simulation run is to check if all the blocks of VCO are in their normal operational states so that all the desired goals are reached or more than reached, including the following:

1. Its power output reaches the maximum and its frequency response is flattened over the desired bandwidth;
2. The operating frequencies cover the entire desired frequency bandwidth;
3. Its phase noise is low enough so that the specification is satisfied over the desired frequency bandwidth;
4. The load-pulling test is especially important to the stability of an oscillator.

The third and fourth items listed above will be introduced and illustrated in the following subsections.

After repeating the above optimization simulations, the optimized part list is obtained, as shown in Table 20.2.

#### 20.4.5 Load-Pulling Test and VCO Buffer

Another factor causing the instability of the oscillation frequency is the variation of the load. The load-pulling test examines the variation of the oscillation frequency as the load impedance is changed.

From Table 20.3 it can be seen that when

$$R_L = R_o = 50 \Omega, \quad (20.83)$$

TABLE 20.2. Part List of the Single-Ended VCO with Clapp Configuration shown in Figure 20.23

Item	Value/type	Item	Value/type	Item	Value/type
$R_1$	1 kΩ	$C_1$	62 pF	$L_1$	300 nH
$R_2$	1.5 kΩ	$C_2$	12 pF	$L_2$	(Printed and trimmed)
$R_3$	820 Ω	$C_3$	4.3 pF	$L_3$	130 nH
$R_4$	51 Ω	$C_4$	22 pF	$L_4$	(Printed and trimmed)
$R_5$	2.7 kΩ	$C_5$	30 pF	$L_5$	(Printed and trimmed)
$R_6$	1.5 kΩ	$C_6$	11800 PF	$L_6$	(Printed and trimmed)
$R_7$	2N3948	$C_7$	MV2103		

TABLE 20.3. An Example of VCO Load-Pulling Test When  $P_{R_L} = 1 \text{ W}$  or  $30 \text{ dB}_m$  ( $R_o = \text{characteristic Impedance} = 50 \Omega$ ) (VSWR, voltage standing wave ratio)

$R_o$	$R_L$	$\Gamma_L$	VSWR	$\text{VSWR}_{\text{dB}}$	$P_{R_L}, \text{dB}_m$	$f, \text{MHz}$
50.00	2500.00	0.96	50.00	33.98	0.3	463.846
50.00	1000.00	0.90	20.00	26.02	0.8	464.265
50.00	500.00	0.82	10.00	20.00	1.7	464.673
50.00	100.00	0.33	2.00	6.02	9.5	464.875
50.00	50.00	0.00	1.00	0.00	(16 mW) 12.0	465.000
50.00	25.00	-0.33	2.00	6.02	9.7	465.282
50.00	5.00	-0.82	10.00	20.00	1.8	465.453

where  $R_o$  is the characteristic impedance, the output frequency,

$$f = 465 \text{ MHz}; \quad (20.84)$$

and when

$$25 \Omega \leq R_L \leq 100 \Omega, \quad (20.85)$$

or

$$\text{VSWR} = 2, \quad (20.86)$$

or

$$\text{VSWR}_{\text{dB}} = 6.02 \text{ dB}, \quad (20.87)$$

the maximum of frequency variation,

$$|\Delta f| \leq 0.282 \text{ MHz}, \quad (20.88)$$

or

$$\frac{|\Delta f|}{f} \leq \frac{0.282}{465} \approx 0.061\%. \quad (20.89)$$

This is a good result in the load-pulling test.

It must be noted that impedance matching in most RF blocks is important and must be emphasized in RF circuit design. However, an exception exists in the VCO buffer design.

In order to isolate the VCO, it is necessary to have a VCO buffer inserted between the VCO output and the load  $R_L$  as shown in Figure 20.33. However, the VCO output need not be impedance-matched with its load  $R_L$ . The reasons are as follows:

- The output impedance of the VCO is negative because, in an oscillator, the impedance at any node is negative except at the nodes for the DC power supply or grounded points. The oscillation is maintained by the negative resistance. Impedance matching to a node or port with negative impedance is meaningless.
- Rather than talking about impedance matching, let us say that the input of the VCO buffer is trying to couple some of the oscillating power from the VCO. There is expected to be as little coupling as possible because a strong coupling brings about disturbance to the VCO due to the variation of the load impedance and, consequently, brings about the degradation of the stability of the oscillation frequency. It is well known that the stability of the oscillation frequency is the most important goal in a VCO or any oscillator design.

Consequently, the input impedance of the VCO buffer will be kept high so that the VCO output will not be disturbed considerably. The RF power coupled from the VCO is kept at a minimum or at “conceivable” level. On the other hand, the sensed RF power must be intensified in the VCO buffer appropriately so as to provide enough RF power of the oscillating signal to other blocks. Figure 20.33 illustrates these basic ideas.

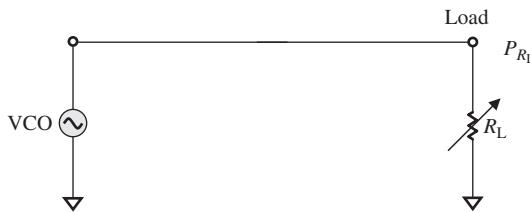


Figure 20.32. VCO load-pulling test.

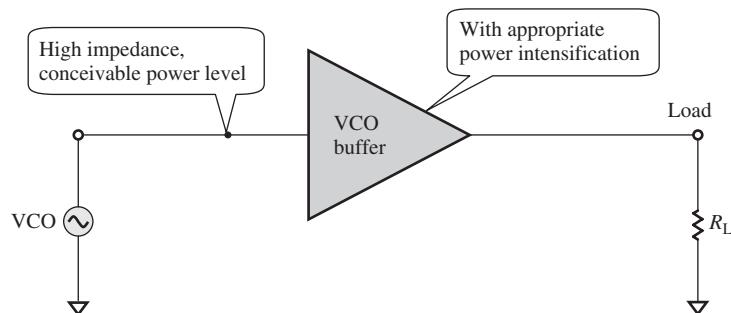


Figure 20.33. Unmatched RF block VCO coupled with a buffer.

TABLE 20.4. Primary Goals of Single-Ended VCO with the Clapp Configuration

Frequency range	440–470 MHz
DC power supply	5 V
Current drain	13 mA
Output power	12 dB <sub>m</sub> (16 mW)
Frequency controlled range	440–470 MHz
Corresponding control voltage	1.0–3.0 V
Control voltage sensitivity	13–15–17 MHz/V
Output power	12 dB <sub>m</sub> (16 mW)
Output impedance	50 Ω
Pulling figure when ratio of standard wave VSWR = 2.0	0.282 MHz
Phase noise when offset frequency $f_{\text{offset}} = 10 \text{ kHz}$	-95 dB <sub>c</sub> /Hz

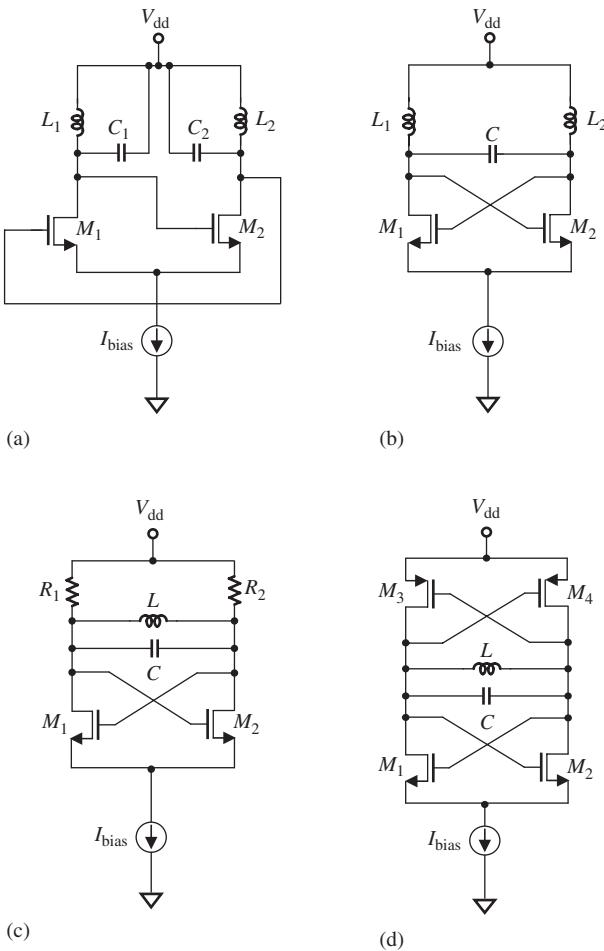
Now we can also explain why the simulation in this design example is unconventional. Instead of conducting a simulation of the entire block, the VCO block is divided into four portions as shown in Figure 20.27. By doing so, the function of each portion can be optimized with respect to its special target. On the contrary, if simulation is conducted for an entire block, the special targets, including the special load-pulling concern for the VCO buffer, would be more or less neglected or ignored.

In addition to the load-pulling test, other tests have been conducted for this VCO design example. The performance can be summarized as shown in Table 20.4.

## 20.5 DIFFERENTIAL VCO AND QUAD-PHASES VCO

Figure 20.34 shows the evolution of the differential VCO.

Figure 20.34(a) shows two stages of the amplifier with positive feedback.



**Figure 20.34.** Evolution of Differential VCO.  
 (a) Two stages of amplifier with positive feedback.  
 (b)  $C_1$  and  $C_2$  are combined as one part  $C$ . (c)  $L_1$  and  $L_2$  are combined as one part  $L$ .  $R_1$  and  $R_2$  represent loss of parts. (d) Piggy-back pairs with current reused.

Figure 20.34(b) looks like two stages of the amplifier with positive feedback modified from Figure 20.34(a) by combining the two capacitors  $C_1$  and  $C_2$  in Figure 20.34(a) into one capacitor  $C$  in series.

Figure 20.34(c) shows two stages of amplifier with positive feedback modified from Figure 20.34(b) by combining the two inductors  $L_1$  and  $L_2$  in Figure 20.34(b) into one inductor  $L$  in series in Figure 20.34(c), while the two resistors represent the loss of parts.

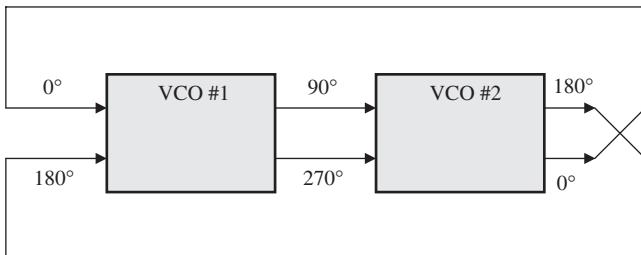
Figure 20.34(d) is a VCO with a pair of two stages of amplifiers stacked together. Instead of only one pair of n-channel MOSFETs as in Figure 20.34(a)–(c), there is also a pair of p-channel MOSFETs in the VCO block. Comparing the VCOs shown in Figure 20.34(c) and 20.34(d), it can be seen that the pair of p-channel MOSFETs look like the two resistors shown in Figure 20.34 (c). On the other hand, the pair of n-channel MOSFETs looks like two resistors if the VCO is mainly built by the upper pair of p-channel MOSFET. In other words, the VCO shown in Figure 20.34(d) consists of two sub-VCOs. One is mainly built by the bottom pair of n-channel MOSFETs, while the upper pair of p-channel MOSFETs functions as its load. Another is mainly built by the upper pair of p-channel MOSFETs, while the bottom pair of n-channel MOSFETs functions as its load. The remarkable advantage in such a stacked configuration is that the current drain can

be reused, so that the output voltage swing or output power is almost doubled from that of the VCOs without stacked configurations shown in Figure 20.34(a)–(c).

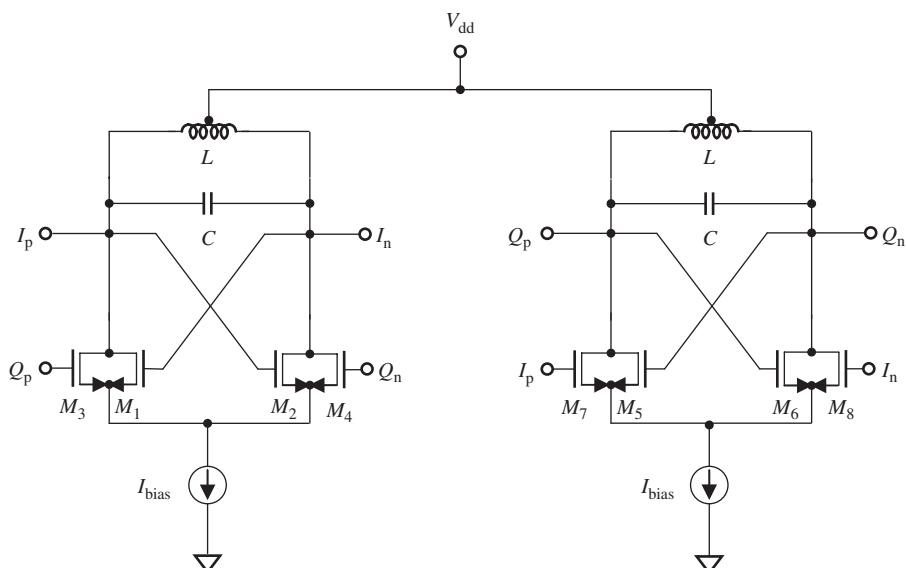
A quad-phase VCO can be built by two differential VCOs. Figure 20.35 shows its block diagram. If the phases of input in the first VCO are  $0^\circ$  and  $180^\circ$ , its output phases are shifted  $90^\circ$  and become  $90^\circ$  and  $270^\circ$  respectively. Then, if the outputs of the first VCO are connected to the corresponding inputs of second VCO inputs, its output phases are  $180^\circ$  and  $0^\circ$ , which are exactly the opposite of the inputs of the first VCO. To positively feedback to the first VCO, the outputs of the second VCO are exchanged with each other and then fed to the inputs of the first VCO.

Figure 20.36 shows the schematic of the quadrature-phase VCO built by two differential VCOs. The individual VCOs in Figure 20.36 are basically the same as that shown in Figure 20.34(b) except for two modifications:

1. The two inductors  $L_1$  and  $L_2$  in Figure 20.34(b) are replaced by one inductor  $L$  with a central tap in Figure 20.36.
2. There are four coupling MOSFETs added to the two VCOs for interconnection between  $I_p$ ,  $I_n$ ,  $Q_p$ , and  $Q_n$ .



**Figure 20.35.** Block diagram of a quadrature-phase VCO built by two differential VCOs.

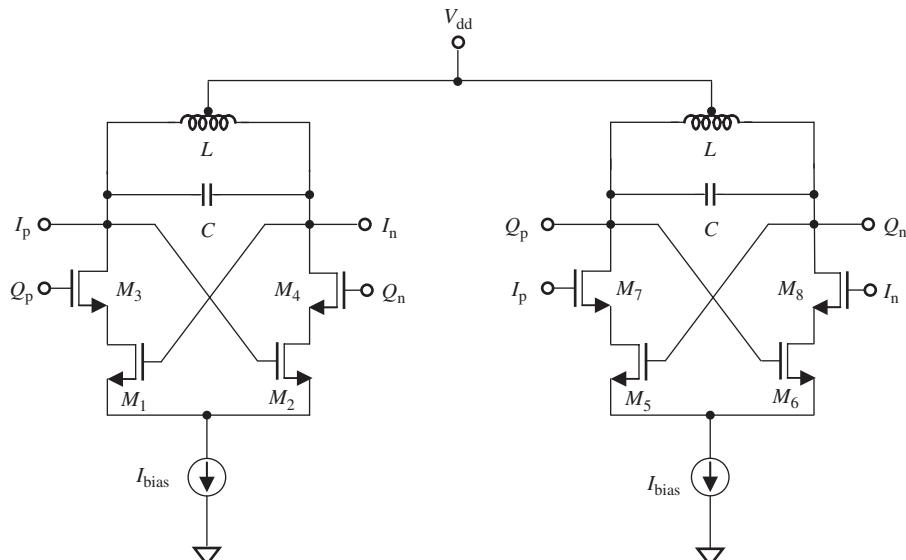


**Figure 20.36.** Schematic of quadrature-phase VCO built by two differential VCOs.  $M_1$  and  $M_3$ ,  $M_2$  and  $M_4$ ,  $M_5$  and  $M_6$ , and  $M_7$  and  $M_8$  are connected in parallel.

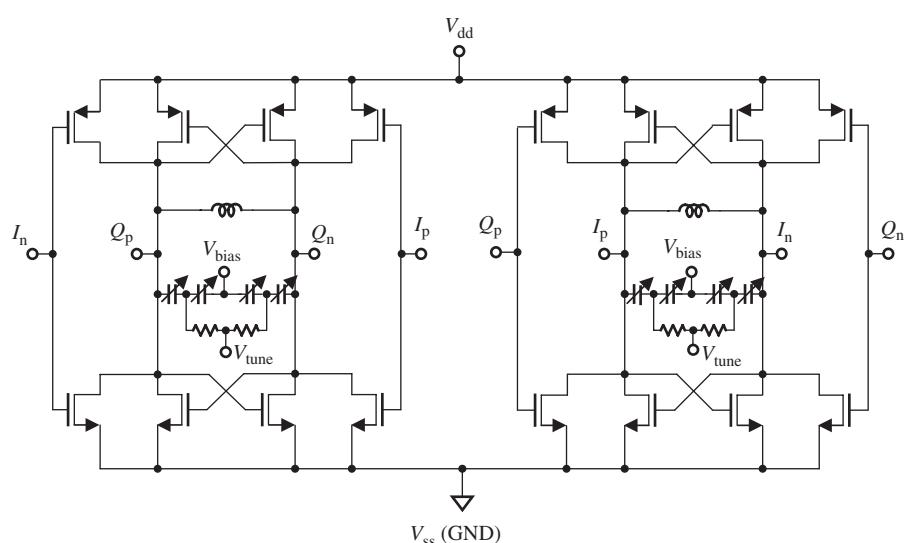
The quad-phase VCO shown in Figure 20.37 is basically the same as that one shown in Figure 20.36. The difference is that the connection mode of the coupling transistors  $M_3$ ,  $M_4$ ,  $M_7$ , and  $M_8$ , with  $M_1$ ,  $M_2$ ,  $M_5$ , and  $M_6$  in Figure 20.37, respectively, in Figure 20.36 is in parallel, while their connection mode in Figure 20.37 is in series. The phase noise in the quad-phase VCO shown in Figure 20.37 is much lower than that in Figure 20.36.

Figures 20.38 and 20.39 show the quad-phases VCO with current reuse.

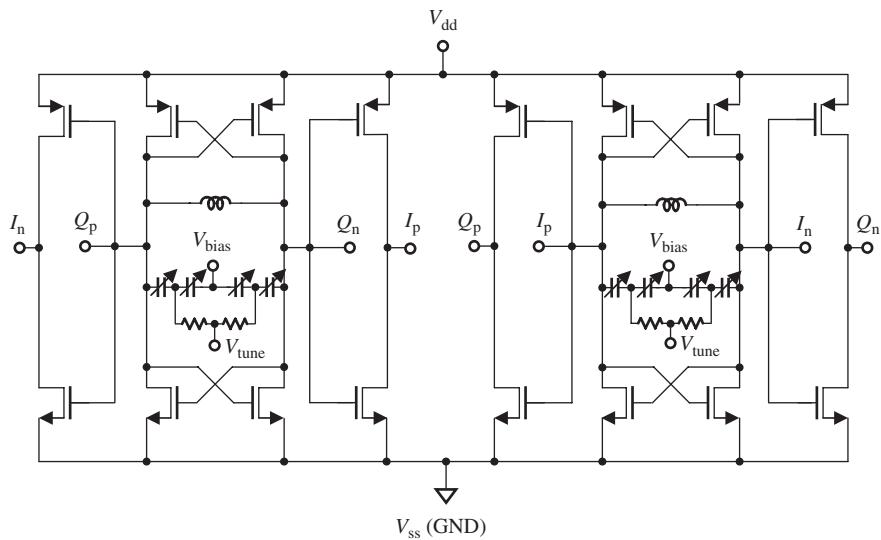
As mentioned in Section 20.4.5, a VCO buffer must be designed for the four terminals,  $I_p$ ,  $I_n$ ,  $Q_p$ , and  $Q_n$ , of the quad-phase VCO as shown in Figure 20.38. The coupling



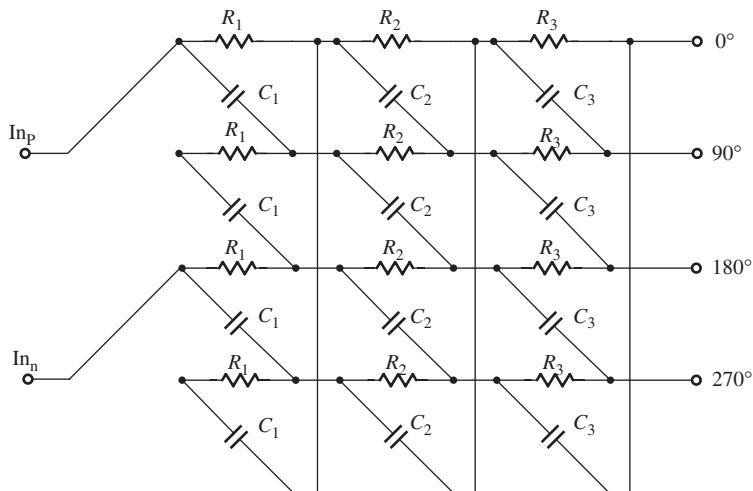
**Figure 20.37.** Schematic of quadrature-phase VCO built by two differential VCOs.  $M_1$  and  $M_3$ ,  $M_2$  and  $M_4$ ,  $M_5$  and  $M_7$ , and  $M_6$  and  $M_8$  are connected in series.



**Figure 20.38.** A quadrature-phase VCO with coupling transistors.



**Figure 20.39.** A quadrature-phase VCO with coupling buffer transistors.



**Figure 20.40.** A quadrature RC phase split network.

transistors in Figure 20.38 function as coupling parts only, and nothing else. On the contrary, the coupling transistors in Figure 20.39 function not only as coupling parts but also as buffer stages. Consequently, it may not be necessary to design a buffer additionally for the quad-phase VCO. Therefore, the quad-phase VCO shown in Figure 20.39 is more economical than that shown in Figure 20.38.

It should be noted that in either Figure 20.38 or 20.39, the capacitors are varactors with control voltages  $V_{bias}$  and  $V_{tune}$ . The VCO operating frequency can be adjusted by the control voltage by varying the varactors' capacitance. The varactors may be divided into two branches in parallel, one for coarse adjustment and another for fine adjustment. They must be carefully selected so that the variation of the VCO frequency can cover all the expected range.

From Figures 20.38 and 20.39, it can be seen that many transistors must be employed. If the performance of a quad-phase VCO is not expected to be paramount, it can be built by any differential VCO as shown in Figure 20.34 and then cooperated with a quadrature RC phase-shift network such as the one shown in Figure 20.40.

The quadrature RC phase-shift network shown in Figure 20.40 is one with two differential inputs and four phase outputs. In each row, there are three RC circuit sectors connected together in series. Any number of sectors can be chosen by the designer depending on the bandwidth required. The more the sectors, the wider the bandwidth.

In general, the phase noise of the VCO with a quadrature RC phase-shift network is much higher than the quad-phase VCO with LC tanks shown in either Figure 20.38 or 20.39.

## REFERENCE

Leeson DB. A simple model of feedback oscillator noise spectrum. Proc IEEE 1966; 54: 329–330.

## FURTHER READING

- Andreani P, Bonfanti A, Romanò L, Samori C. Analysis and design of a 1.8-GHz CMOS LC quadrature VCO. IEEE J Solid-State Circ 2002;37(12):1737–1747.
- Andreani P, Wang X. On the phase-noise and phase-error performances of multiphase LC CMOS VCOs. IEEE J Solid-State Circuits 2004;39(11):1883–1893.
- Berny AD, Niknejad AM, Meyer RG. A wideband low-phase-noise CMOS VCO. IEEE 2003 Custom Integrated Circuits Conference, San Jose, CA; 2003. pp. 555–558.
- Gierkink SLJ, Levantino S, Frye RC, Samori C, Bocuzzi V. A low-phase-noise 5-GHz CMOS quadrature using superharmonic coupling. IEEE J Solid-State Circ 2003;38(7):1148–1154.
- Hajimiri A. *Noise in Phase-Locked Loops* [Invited], 2001 Southwest Symposium on Mixed-Signal Design, SSMSD, Austin, Texas, 2001.
- Hajimiri A, Lee TH. A general theory of phase noise in electrical oscillators. IEEE J Solid-State Circ 1998;33(2):179–194.
- Hajimiri A, Lee TH. Design issues in CMOS differential LC oscillators. IEEE J Solid-State Circ 1999;34(5):717–724.
- Ham D, Hajimiri A. Concepts and methods in optimization of integrated LC VCOs. IEEE J Solid-State Circ 2001;36(6):896–909.
- Jeon S, Jung S, Lee D, Lee H. A fully integrated CMOS LC VCO and frequency divider for UHF RFID reader. pp. 117–120.
- Jia L, Ma JG. A novel methodology for the design of LC tank VCO with low phase noise. pp. 376–379.
- Ko S, Hong S. Noise property of a quadrature balanced VCO. IEEE Microw Wireless Compon Lett 2005;15(10):673–675.
- Koukab A, Lei Y, Declercq M. Design and optimization of a linear wide-band VCO for multimode applications. 2005 IEEE Radio Frequency integrated Circuits Symposium, Long Beach, CA; 2005. pp. 527–530.
- Kwok KC, Luong HC. A 0.3-V 1.46-mW low phase noise oscillator with transformer feedback in standard 0.18- $\mu$ m CMOS process. IEEE 2003 Custom Integrated Circuits Conference, San Jose, CA; 2003. pp. 551–554.
- Lee TH, Hajimiri A. Oscillator phase noise: a tutorial. IEEE J Solid-State Circ 2000;35(3): 326–335.
- Li S, Kipnis I, Ismail M. A 10-GHz CMOS Quadrature LC-VCO for multirate optical applications. IEEE J Solid-State Circ 2003;38(10):1626–1634.
- Moon H, Kang S, Kim YT, et al. A fully differential LC-VCO using a new varactor control structure. IEEE Microw Wireless Compon Lett 2004;14(9):410–412.

- Muer BD, Borremans M, Steyaert M, et al. A 2-GHz low-phase-noise integrated LC-VCO set with flicker-noise up-conversion minimization. *IEEE J Solid-State Circ* 2000;35(7):1034–1038.
- Navid R, Lee TH, Dutton RW. Lumped inductorless oscillators: how far can they go? *IEEE 2003 Custom Integrated Circuits Conference*, San Jose, CA; 2003 pp. 543–546.
- Razavi B. A study of phase noise in CMOS oscillators. *IEEE J Solid-State Circ* 1996;31(3):331–343.
- Ryu S, Chung Y, Kim H, et al. Phase noise optimization of CMOS VCO through harmonic tuning. *2005 IEEE Radio Frequency Integrated Circuits Symposium Long Beach, CA*; 2005. pp. 403–406.
- Soltanian B, Kinget P. A tail current-shaping technique to reduce phase noise in LC VCOs. *IEEE 2005 Custom Integrated Circuits Conference*, Lille, France; 2005. pp. 579–582.
- Yoon SW, Pinel S, Laskar J. High-performance 2-GHz CMOS LC VCO with high-q embedded inductors using wiring metal layer in a package. *IEEE Trans Adv Packag* 2006;29(3):639–646.

## EXERCISES

1. Explain the difference between a Hartley oscillator, a Colpitts oscillator, and a Clapp oscillator.
2. What is the operating frequency of a phase-shift oscillator?
3. What is the operating frequency of a crystal oscillator?
4. Why the impedance matching in the tank circuit or in the subcircuitry containing the VCO device has not been conducted?
5. Explain the open-loop transfer function or open-loop gain in the PLL loop.
6. Explain closed-loop transfer function or closed-loop gain in the PLL loop.
7. What does “phase noise” stand for?
8. What are the special features of Leeson’s model?
9. What is the equivalent circuit of a printed inductor on a PCB (printed circuit board)? How to get the values of its parts from testing?
10. In the design of VCO, the first two design procedures are (i) the simulation for the second half of tank circuit, in which the maximum of  $S_{11}$  is pursued and its phase is read when  $S_{11} = S_{11,\max}$ ; and (b) the simulation for the first half of tank circuit, in which the phase  $\Gamma$  is forced to be equal to the negative phase value of  $S_{11,\max}$  and the absolute value of  $\Gamma$  is tried to be as higher than the absolute value of  $1/S_{11,\max}$  as possible. Why?
11. Compare the advantages and disadvantages of the quadrature-phase LC VCO and the quadrature-phase split RC VCO.
12. What is the improvement in the VCO design from a quadrature-phase VCO built by two differential VCOs with only n-channel coupling transistors to a quadrature phase VCO built by two differential VCOs with both of p- and n-channel coupling transistors?

## ANSWERS

1. Hartley oscillator, Colpitts oscillator, and Clapp oscillator are “Three-point” type of oscillators. The advantage of the Hartley oscillator is that the oscillator is easily excited. Its disadvantage is that two inductors or one tapped inductor must be

employed and that its phase noise is higher than that of other types of oscillators. Compared to the Hartley oscillator, the advantage of the Colpitts oscillator is that only one inductor is needed and that its phase noise is much lower than that of the Hartley and other oscillators. Essentially, the Clapp oscillator is a modified Colpitts oscillator. The improvement is that the frequency bandwidth is more easily covered by adjusting one of two capacitors.

2. The operating frequency of a phase-shift oscillator is

$$f = \frac{4}{RC}.$$

3. The operating frequency of a crystal oscillator (Fig. 20.P.1) is

$$\omega_p^2 = \frac{1}{L} \left( \frac{1}{C} + \frac{1}{C'} \right).$$

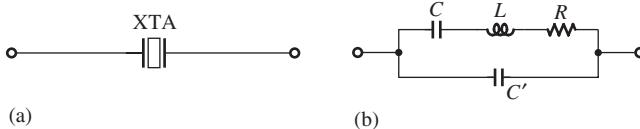


Figure 20.P.1. Characteristics of a piezoelectric crystal. (a) Symbol. (b) Electrical model.

4. VCO is a special RF block. Impedance matching in the tank circuit, in the subcircuitry containing the VCO device, or in the VCO output is not necessary, because  
 (a) VCO is an independent and indivisible power transportation unit;  
 (b) except at the node  $V_{dd}$  or  $V_{cc}$ , bias, and grounded points, the impedance is negative in other nodes.
5. Open-loop transfer function or open-loop gain (Fig. 20.P.2) is

$$G(s) = \frac{\theta_o(s)}{\theta_e(s)} = \frac{\theta_o(s)}{\theta_i(s) - \theta_o(s)} = \frac{K_o K_d F(s)}{sM}$$

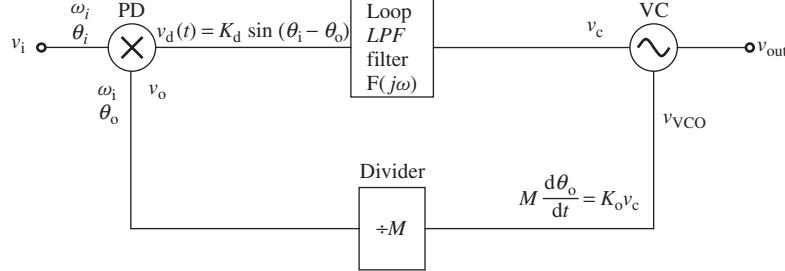


Figure 20.P.2. Basic block diagram of PLL (phase lock loop).

6. Closed-loop transfer function or closed-loop gain is

$$H(s) = \frac{\theta_o(s)}{\theta_i(s)} = \frac{K_o K_d F(s)}{sM + K_o K_d F(s)} = \frac{G(s)}{1 + G(s)}$$

7. “Phase noise” is the noise in the PLL. In the input of PLL, the PD detects the phase difference between the input signal and the output signal of divider, and the noise is therefore “attached” to its output voltage which is proportional to the phase difference. It is therefore called *phase noise*. At the output of VCO, phase noise is converted as the variation of the spectral line around the VCO operating frequency.
8. By Leeson’s model, the phase noise can be expressed as

$$L(\Delta\omega) = 10 \cdot \log \left[ \frac{2FkT}{P_{\text{sig}}} \left( 1 + \frac{\omega_0}{2Q\Delta\omega} \right)^2 \left( 1 + \frac{\Delta\omega_{1/f^3}}{|\Delta\omega|} \right) \right].$$

The special features of Leeson’s model are as follows:

- (a) The factor  $F$  is to account for the increased noise in the  $1/(\Delta\omega)^2$  or  $1/f^2$  region;
  - (b) The additive factor of unity in the first parenthesis is to account for the noise floor;
  - (c) The multiplicative factor  $(1 + \Delta\omega_{1/f^3}/|\Delta\omega|)$  is to provide  $1/(\Delta\omega)^3$  or  $1/f^3$  behavior in the  $1/(\Delta\omega)^3$  or  $1/f^3$  region.
9. The equivalent circuit of a printed inductor on a PCB is shown in Figure 20.P.3. In terms of the following equations, the values of its parts  $C_A$ ,  $C_B$ , and  $L$  can be calculated from the testing values  $C'_A$ ,  $C'_B$ , and  $L'$  at the tested terminals A–G, B–G, and A–B, respectively.

$$C_{L+C_B} = \frac{C_B}{1 - \omega^2 L C_B},$$

$$C_{L+C_A} = \frac{C_A}{1 - \omega^2 L C_A},$$

$$C_{C_A+C_B} = \frac{C_A C_B}{C_A + C_B},$$

$$C'_A = C_A // C_{L+C_B} = C_A + \frac{C_B}{1 - \omega^2 L C_B},$$

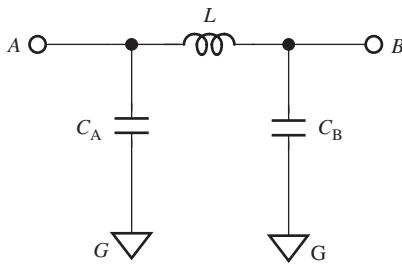
$$C'_B = C_B // C_{L+C_A} = C_B + \frac{C_A}{1 - \omega^2 L C_A},$$

$$L' = L // (C_A + C_B) = L \left( 1 + \omega^2 L \frac{C_A C_B}{C_A + C_B} \right),$$

$$C_A = \frac{2C'_A C'_B (\omega^2 L' C'_A C'_B + C'_B - C'_A)}{(\omega^2 L' C'_A C'_B + C'_B - C'_A)^2 + 4\omega^2 L' C'_A C'_B},$$

$$C_B = \frac{2C'_A C'_B (\omega^2 L' C'_A C'_B - C'_B + C'_A)}{(\omega^2 L' C'_A C'_B + C'_B - C'_A)^2 + 4\omega^2 L' C'_A C'_B},$$

$$L = \frac{(\omega^2 L' C'_A C'_B + C'_B - C'_A)^2 + 4\omega^2 L' C'_A C'_B}{2\omega^2 C'_A C'_B (\omega^2 L' C'_A C'_B + C'_B + C'_A)^2}.$$



**Figure 20.P.3.** Equivalent circuit of a printed inductor.

10. The reason is that it is expected that the oscillation can be maintained forever, so that within the tank circuit, the following goals are set:
- For the second half of tank circuit including the device, the maximum of  $S_{11}$  is pursued. When  $S_{11} = S_{11,\max}$ , the magnitude and phase are read for the next simulation step;
  - For the first half of tank circuit, the phase  $\Gamma$  is forced to be equal to the negative phase value of  $S_{11,\max}$  and the absolute value of  $\Gamma$  is tried to be as much higher than the absolute value of  $1/S_{11,\max}$  as possible.
11. The comparison of the advantages and the disadvantages between the quadrature-phase LC VCO and the quadrature phase split RC VCO is shown in the following table:

Item	Quadrature-Phase LC VCO	Quadrature-Phase split RC VCO
Upper frequency limit	Higher	Lower
Attenuation	Lower	Higher
Stability	Higher	Lower
Cost	Higher	Lower

12. The improvements are
- current reuse
  - increase of  $g_m$ , resulting in high-output power
  - improvement of carrier's purity (lower harmonics' power).



# PA (POWER AMPLIFIER)

## 21.1 CLASSIFICATION OF PA

A PA (power amplifier) is an amplifier that intensifies the power of an input signal (Fig. 21.1).

Its output power is  $G$  times its input power. The added portion of the output power is AC power converted from the DC power supply. The ratio of the added portion of the power to the input power is called its PAE (*power added efficiency*), that is,

$$G_{\text{watt}} = \frac{P_{\text{out,watt}}}{P_{\text{in,watt}}}, \quad (21.1)$$

$$G_{\text{dB}} = P_{\text{out,dB}} - P_{\text{in,dB}}, \quad (21.2)$$

$$\Delta P = P_{\text{out,watt}} - P_{\text{in,watt}}, \quad (21.3)$$

$$\text{PAE} = \frac{\Delta P}{P_{\text{DC}}} = \frac{P_{\text{out,watt}} - P_{\text{in,watt}}}{P_{\text{DC}}}, \quad (21.4)$$

where

$P_{\text{out,watt}}$  = the output power in watts,

$P_{\text{in,watt}}$  = the input power in watts,

$P_{\text{out,dB}}$  = the output power in decibels,

$P_{\text{in,dB}}$  = the input power in decibels,

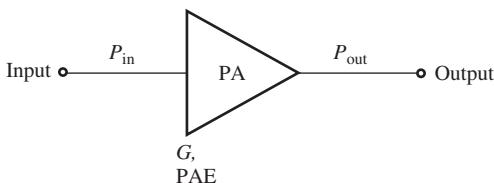


Figure 21.1. Representation of a power amplifier.

$G_{\text{watt}}$  = the power gain in watts,

$G_{\text{dB}}$  = the power gain in decibels,

$\Delta P$  = the difference between output and input power in watts,

$P_{\text{DC}}$  = the DC power from the DC power supply, and

PAE = the power added efficiency in percent.

In a PA design, the most important goal is the power conversion efficiency from DC power to AC power, which combined with the input power forms the output power.

The classification of power amplifier is based on the PAE.

### 21.1.1 Class A Power Amplifier

Figure 21.2 shows the operating principle of a class A power amplifier. The inductor  $L_o$  and capacitor  $C_o$  are resonant at the operating frequency.

When the device is working as a class A power amplifier, the device is conducted in all the cycle.

The waveform of the current drain  $i_D$  is a normal sinusoidal wave and the waveform of the output voltage  $v_o$  is a normal sinusoidal wave as well.

The PAE for a class A power amplifier is about 30%, although theoretically it can approach a higher value.

### 21.1.2 Class B Power Amplifier

Figure 21.3 shows the operating principle of a class B power amplifier. The inductor  $L_o$  and capacitor  $C_o$  are resonant at the operating frequency.

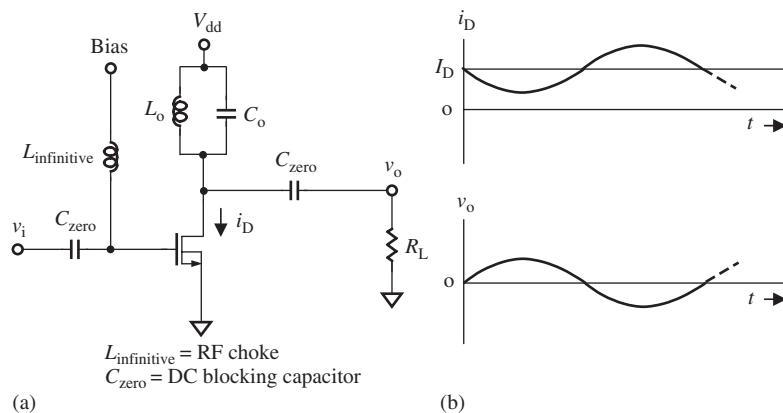
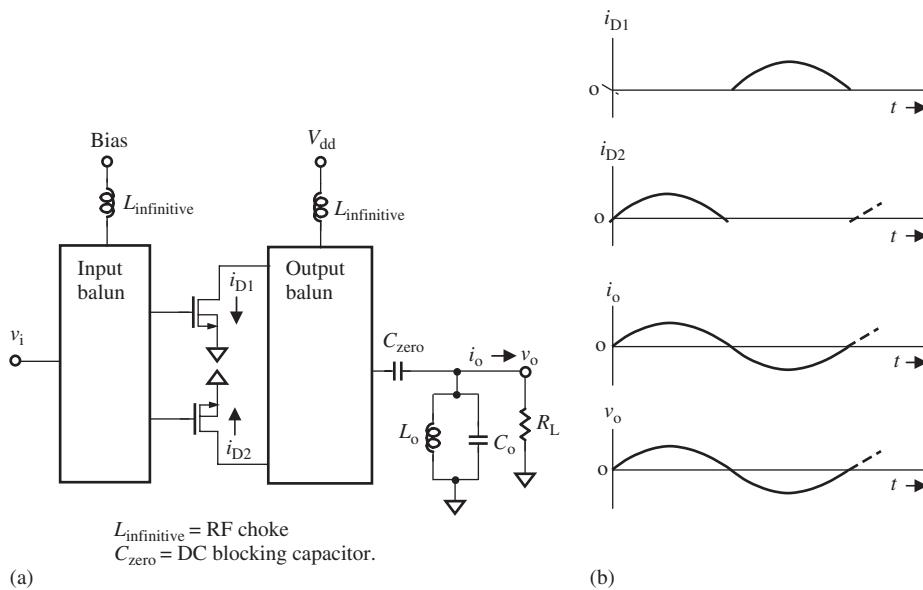


Figure 21.2. Class A power amplifier and its waveforms. (a) Simple circuit of a class A power amplifier. (b) Waveform of  $i_D$  and  $v_o$ .



**Figure 21.3.** Class B power amplifier and its waveforms. (a) Simple circuit of a class B power amplifier. (b) Waveform of  $i_{D1}$ ,  $i_{D2}$ ,  $i_o$ , and  $v_o$ .

When the device is working as a class B power amplifier, the device is conducted for more than one-half cycle but less than a whole cycle.

The waveforms of the current drains  $i_{D1}$  and  $i_{D2}$  have a normal but “broken” sinusoidal wave lasting a little bit more than one-half cycle. However, the waveform of the output current  $i_o$  and voltage  $v_o$  is a normal sinusoidal wave.

The PAE for a class B power amplifier is about 60%, although theoretically it can approach a higher value.

### 21.1.3 Class C Power Amplifier

Figure 21.4 shows the operating principle of a class C power amplifier. The inductor  $L_o$  and capacitor  $C_o$  are resonant at the operating frequency.

When the device is working as a class C power amplifier, the device is conducted for less than half a cycle.

The waveform of the current drain  $i_D$  is a portion of a sinusoid, but the waveform of output voltage  $v_o$  is a normal sinusoid.

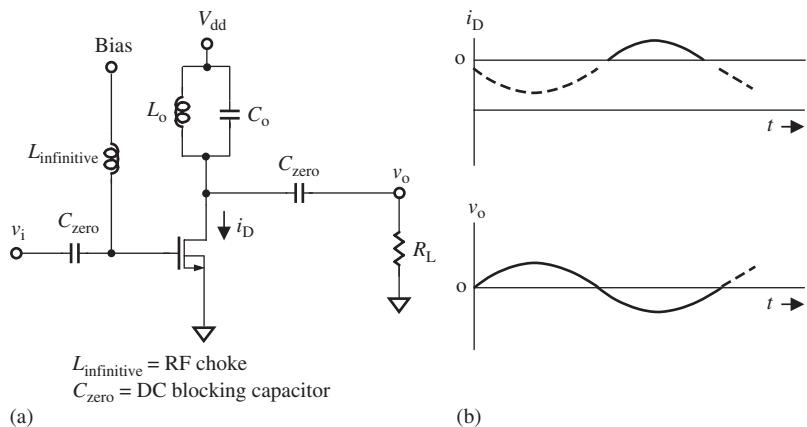
The PAE for a class C power amplifier is about 80%, although theoretically it can approach a higher value.

### 21.1.4 Class D Power Amplifier

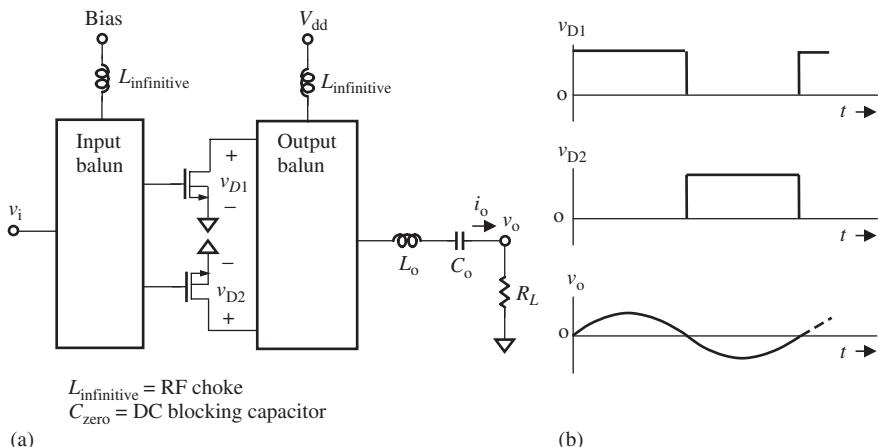
Figure 21.5 shows the operating principle of a class D power amplifier. The inductor  $L_o$  and capacitor  $C_o$  are resonant at the operating frequency.

A class D power amplifier is a pair of devices acting as a two-pole switch, which defines a rectangular voltage or current.

The drain voltages,  $v_{D1}$  and  $v_{D2}$ , are rectangular waveforms. However, the waveform of the output voltage  $v_o$  is a normal sinusoid.



**Figure 21.4.** Class C power amplifier and its waveforms. (a) Simple circuits of a class C power amplifier. (b) Waveform of  $i_D$  and  $v_o$ .



**Figure 21.5.** Class D power amplifier and its waveforms. (a) Simple circuit of a class D power amplifier. (b) Waveform of  $v_{D1}$ ,  $v_{D2}$ , and  $v_o$ .

The PAE for a class D power amplifier is more than 80%, although theoretically it can approach a higher value.

### 21.1.5 Class E Power Amplifier

Figure 21.6 shows the operating principle of a class E power amplifier. The inductor  $L_o$  and capacitor  $C_o$  are resonant at the operating frequency.

When the device is working as a class E power amplifier, it is driven to act as a switch.

The waveforms of the current  $i_D$  and voltage  $v_D$  are conjugate to each other but the waveform of output voltage  $v_o$  is a normal sinusoid with a phase delay  $\varphi$ .

The PAE for a class E power amplifier is about 90%, although theoretically it can approach a higher value.

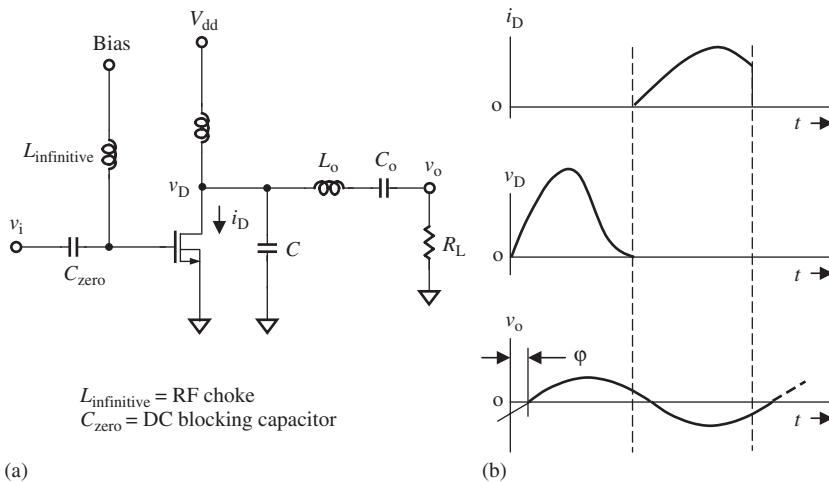


Figure 21.6. Class E power amplifier and its waveforms. (a) Simple circuit of a class E power amplifier. (b) Waveform of  $i_D$ ,  $v_D$ , and  $v_o$ .

### 21.1.6 Third-Harmonic-Peaking Class F Power Amplifier

Figure 21.7 shows the operating principle of a class F power amplifier. The inductor  $L_o$  and capacitor  $C_o$  are resonant at the operating frequency and the inductor  $L_3$  and capacitor  $C_3$  are resonant at the third harmonic frequency.

When the device is working as a class F power amplifier, the device is driven to approximately act as a switch. The waveform of the current  $i_D$  is about half that of the sinusoidal wave but the waveform of the output voltage  $v_o$  is a normal sinusoid.

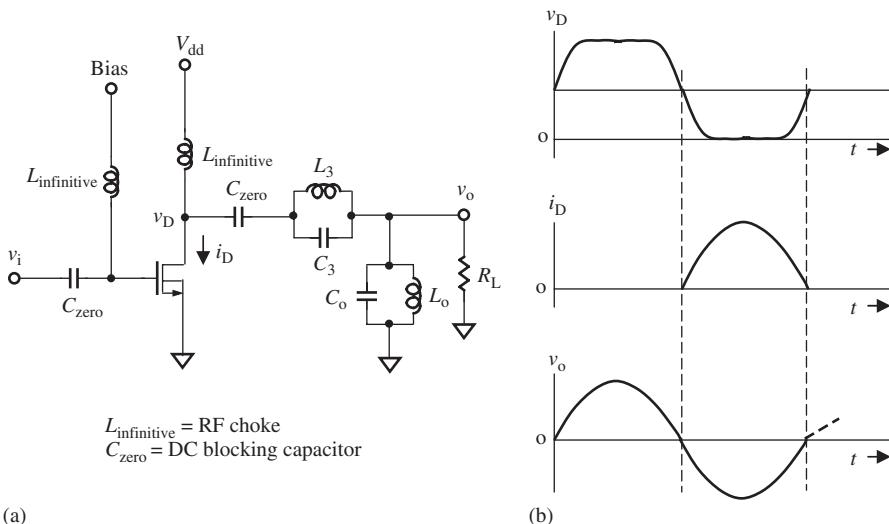


Figure 21.7. Third-harmonic-peaking class F power amplifier and its waveforms. (a) Simple circuit of a class F power amplifier. (b) Waveform of  $i_D$ ,  $v_D$ , and  $v_o$ .

The PAE for a class F power amplifier is about 80%, although theoretically it can approach a higher value.

### 21.1.7 Class S Power Amplifier

Figure 21.8 shows the operating principle of a class S power amplifier. An LPF (low-pass filter) is built by the inductor  $L_o$  and the capacitor  $C_o$ .

A class S power amplifier is a two-position switch with a rectangular waveform applied to an LPF, which allows the slowly varying DC or average component to appear on the load.

The drain voltage  $v_{D2}$  has a rectangular waveform. However, the waveform of the output voltage  $v_o$  is a normal sinusoid.

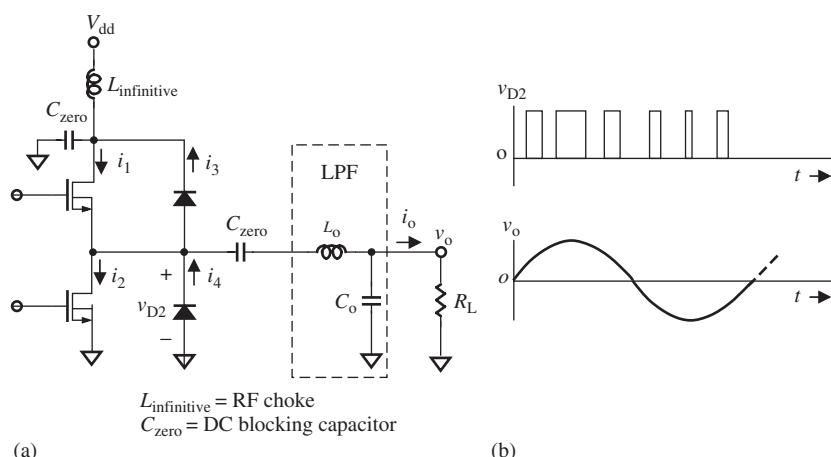
The PAE for a class S power amplifier is more than 90%, although theoretically it can approach a higher value.

## 21.2 SINGLE-ENDED PA

Usually, the main design procedures for an RF block are

- selection of device and obtaining the model of device from vendor or foundry,
- simulation,
- layout, and
- testing.

Very often, the problem exists in the second step. The success or failure of simulation relies on the accuracy of the device model. There is little chance that the model obtained from the vendor or foundry is accurate enough to get a reasonable simulation result. The main task of simulation is to implement the input and output impedance matching networks. Instead of simulation, an alternative way is to adapt so-called “tuning-on-the-bench” technique.



**Figure 21.8.** Class S power amplifier and its waveforms. (a) Simple circuit of a class S power amplifier. (b) Waveform of  $v_{D1}$ ,  $v_{D2}$ , and  $v_o$ .

### 21.2.1 Tuning on the Bench

The input and output impedance matching networks for the device can be implemented by means of the tuning-on-the-bench technique. Figure 21.9 is a tuning setup for the device.

The test fixture contains the tested device M, two “zero” capacitors, and two “infinite” inductors. The device shown in Figure 21.9 is assumed to be a MOSFET (metal–oxide–semiconductor field-effect transistor). All the parts are soldered on a small PCB. Two “zero” capacitors and two “infinite” inductors serve to supply DC power to the device. Owing to the fact that their impedances are, respectively, 0 and  $\infty$ , their existence does not disturb the AC or RF characteristics of the tested device. There are two RF tuners. The input tuner is inserted between the signal generator and the input of the test fixture, and the output tuner is inserted between the output of the test fixture and the power meter.

For a specified DC power supply voltage, the DC current flowing through the device must be adjusted to the specified value by the DC bias. Then, the work of tuning on the bench can be started.

The input and output tuners can be slid up and down to adjust its impedance. Therefore, the input and output impedances of the device can be regulated by adjusting the input and output tuners so that the output power displayed in the power meter approaches a maximum, which implies that either the input impedance or the output impedance of the device is matched well. The maximum must be verified very carefully and patiently by several repeated adjustments of the tuners’ positions.

When the power reading in the power meter reaches a maximum, the sliding positions of input and output tuners must be fixed and then detached from the test setup. After the detachment is done, the two impedances  $Z_{in}^*$  and  $Z_{out}^*$  can be tested and obtained by means of the new test setup as shown in Figure 21.10. The network analyzer can read the values of impedances  $Z_{in}^*$  and  $Z_{out}^*$ .

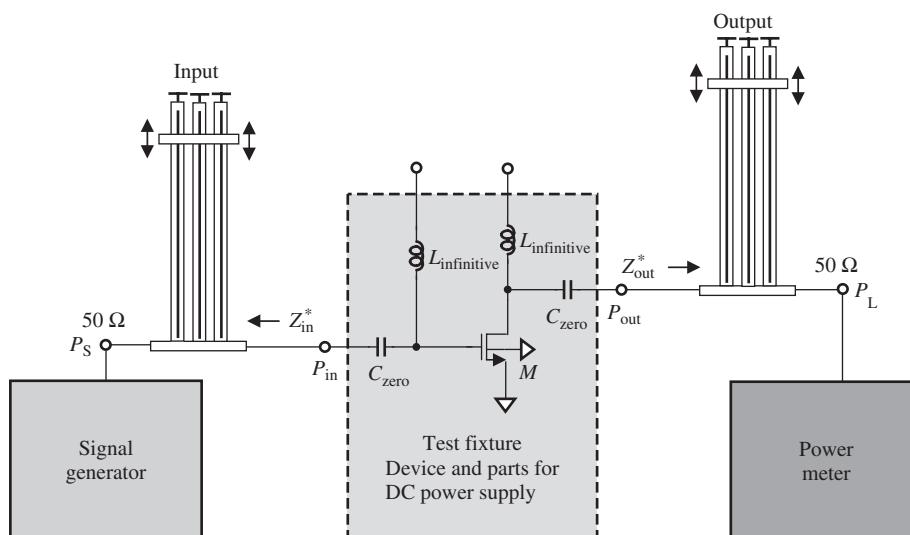


Figure 21.9. Device tuned by two tuners.

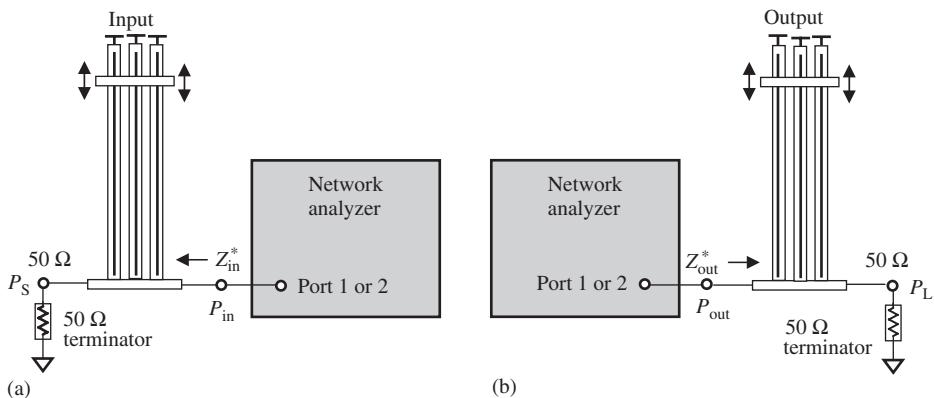


Figure 21.10. Test setup for  $Z_{in}^*$  and  $Z_{out}^*$ . (a) Test for  $Z_{in}^*$ . (b) Test for  $Z_{out}^*$ .

### 21.2.2 Simulation

Having obtained the values of  $Z_{in}^*$  and  $Z_{out}^*$ , we can now design and implement the input and output matching networks. In the ADS (advanced design system) simulation system, there is a “1 Port Opt” box by which the corresponding impedance matching network can be optimized if the original impedance to be matched is known. For instance, if

$$z_{in}^* = r_{in}^* + jx_{in}^*, \quad (21.5)$$

$$z_{out}^* = r_{out}^* + jx_{out}^*, \quad (21.6)$$

and if  $x_{in}^*$  is capacitive and  $x_{out}^*$  is inductive, that is

$$x_{in}^* < 0, \quad (21.7)$$

$$x_{out}^* > 0, \quad (21.8)$$

then,  $x_{in}$  is inductive and  $x_{out}$  is capacitive.

Figures 21.11 and 21.12 show, respectively, the simulation setup for the optimization of the input and output impedance matching networks by means of the “1 Port Opt” box. As an example, the operating frequency is assumed to be

$$f = 800 - 900 \text{ MHz}, \quad (21.9)$$

$$f_c = 850 \text{ MHz}. \quad (21.10)$$

All the parts of the input impedance matching network are assumed to be soldered on a plastic PCB. The main parameters of the plastic PCB are listed in both Figures 21.11 and 21.12. The actual models for the chip inductor and chip capacitor are employed in the simulation.

In order to reduce part count and cost, it is desirable that the “zero” capacitor at the input of the device be a part of the input impedance matching network. The capacitor  $C_{S12}$  in Figure 21.11 is arranged such that it can function as an impedance matching

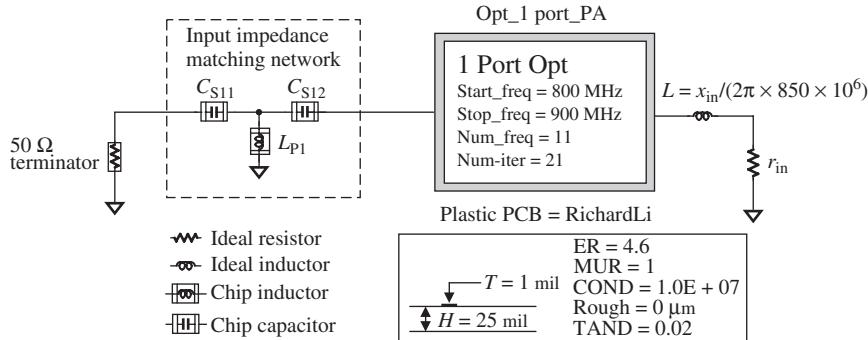


Figure 21.11. Simulation setup for the optimization of input impedance matching network.

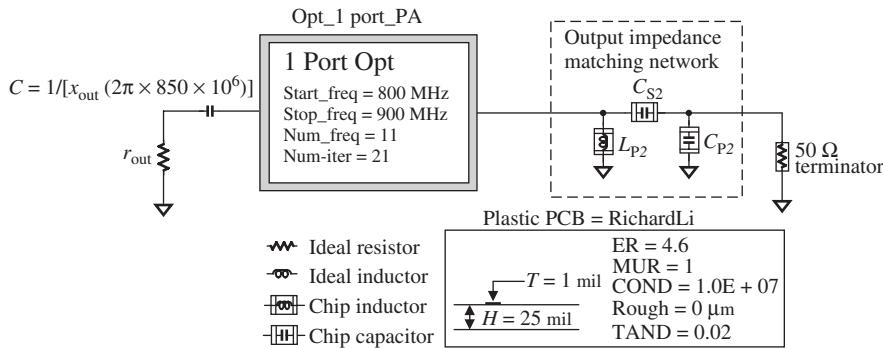


Figure 21.12. Simulation setup for the optimization of output impedance matching network.

part and as a DC blocking capacitor as well. On the other hand, it is desirable that the “infinite” inductor and “zero” capacitor at the output of the device become parts of the output impedance matching network. The inductor  $L_{P2}$  and capacitor  $C_{S2}$  in Figure 21.12 are arranged such that the inductor  $L_{P2}$  can function as an impedance matching part and as a DC voltage provider, and the capacitor  $C_{S2}$  can function as an impedance matching part and as a DC blocking capacitor as well.

Figure 21.13 shows the schematic after the impedance matching is done. The remaining design procedures are layout and testing, which are basically the same as in other RF circuit block designs.

Finally, there is an important remark to be made about “tuning on the bench,” that is, the special care that must be paid to the “security.” The test setup shown in Figure 21.9 is dangerous to the device because it is in a situation where the input and output impedances of the device are unmatched. The returned power from the load, which is the power meter in Figure 21.9, can burn and destroy the device. The solution to this problem is to insert a circulator between the output tuner and the power meter so as to avoid the return power from the power meter entering the device. Another alternative solution is to turn on the signal from the signal generator carefully. It is better to turn it on starting from zero and then gradually increase the signal step by step to the desired level. The increase of power in each step is kept to a minimum.

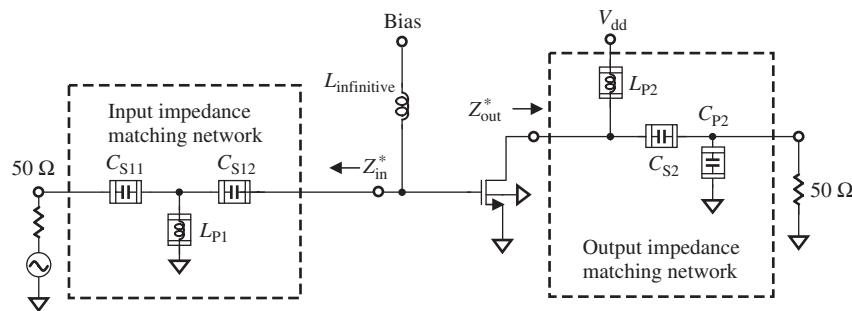


Figure 21.13. Two tuners replaced by an input and output impedance matching network.

### 21.3 SINGLE-ENDED PA IC DESIGN

The methodology of “tuning on the bench” applied to PA design seems all right for a PA module or a PA built by discrete parts. However, it is almost impossible to apply it in a PA IC design. The bonding pads, bonding wires, and test fixture may bring about many unidentifiable and unforeseen problems.

How about forgetting the circuit simulation and directly jumping to the layout procedure? This is a simple but “unorthodox” method. Surprisingly, it works!

Figure 21.14 shows schematic of a 1–2-W PA built on a CMOS IC chip.  $M_1$  and  $M_2$  constitute an electronic switch controlled by the control voltage  $V_{\text{CTL}}$ . When  $V_{\text{CTL}} = 0$ , the PA is “ON”, and when  $V_{\text{CTL}} = 1$ , the PA is “OFF.”  $M_3$  is in the preamplifier stage and  $M_4$  is in the final stage.  $M_4$  is a huge transistor with 256 fingers, which is four times larger than  $M_3$ .

The main specifications and performance are as follows:

- DC power supply = 3 V

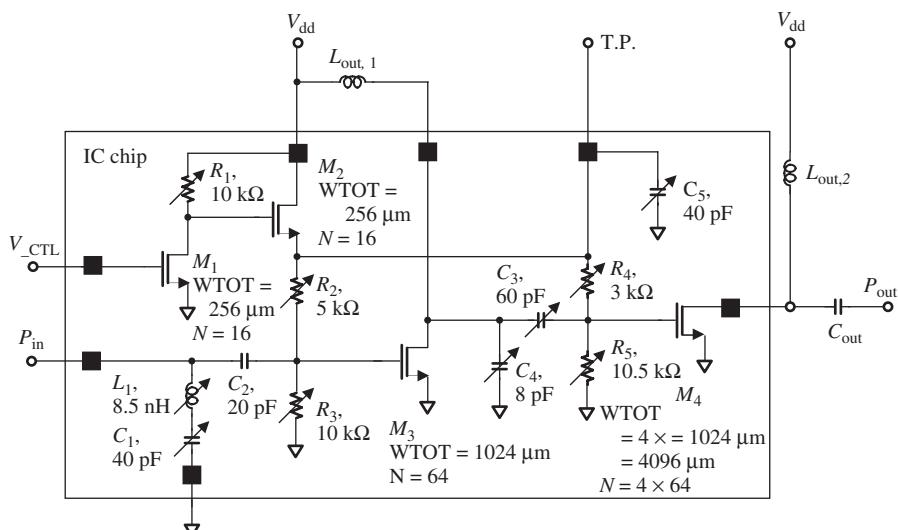


Figure 21.14. Schematic of a 3-V, 1-W PA IC (integrated circuit) with CMOS processing.

- Frequency range = UHF with central frequency  $f_c = 450$  MHz
- Input power = 150 mW
- Output power = 1–2 W
- Current drain = 690 mA when output power = 1 W
- PAE >50%.

The key issues for success are

- Selecting the expected topology,
- Hand-calculating the variable range for every part's value, including device size and the value of capacitors, inductors, and resistors,
- Designing all the parts in the circuit as variable parts. This requires a special layout scheme,
- Trimming every part to an expected value by a laser trimmer (also called a laser cutting system) on the IC die or on the wafer.

The outstanding achievement is that this IC chip is made with only one tape-out. It saves a lot of design time, at least a couple of months, and greatly reduces the design cost.

The detailed layout scheme is described in chapter 9 in this book.

## 21.4 PUSH-PULL PA DESIGN

We now describe a sample push-pull PA design. (The content of this section is abstracted from the author's 1992 design work.)

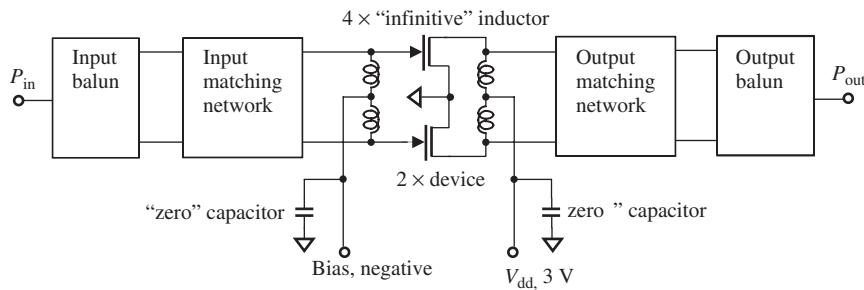
### 21.4.1 Main Specification

- Frequency range: 800–900 MHz ( $f_c = 850$  MHz)
- Output power: 2 and 5 W
- PAE: >50%
- DC power supply: 3 V
- Selected device: GaAs FET, CLY10 (by Siemens).

### 21.4.2 Block Diagram

It is well known that a class A power amplifier has the best linearity but the lowest efficiency. Considering a trade-off between efficiency and linearity, it is decided to design this PA with a push-pull style and with class B operation. Input and output baluns are needed for the push-pull configuration. Figure 21.15 shows the block diagram, which consists of five portions, including the input and output impedance matching networks.

This is a push-pull but not a differential PA. The difference between the push-pull and the differential PA is the CMRR capability. A PA with push-pull configuration does not have this capability, while a PA with differential configuration has the capability to reject the common mode input signal. The difference of topology is that a PA with a



**Figure 21.15.** Block diagram of 3-V push–pull PA. Note: “zero” capacitor = the bypass capacitor. “Infinite” inductor = the RF choke.

differential configuration has a common tail resistor between the two devices while a PA with push–pull configuration does not.

A question might be raised: why is a push–pull PA preferred to a differential PA?

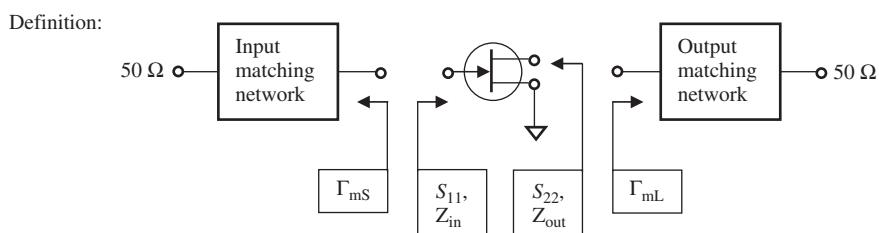
Usually, the power efficiency of a PA with a push–pull configuration is higher than a PA with a differential configuration. In a differential PA, the tail resistor is a negative feedback part. It lowers the output power for better performance in nonlinearity and DC offset. In a push–pull PA, the tail resistor does not exist so that its output power is higher than the differential one. Unless a design for a linear PA, efficiency is the most important factor for a PA design.

Of course, in the design of a linear PA, the priorities are reversed. The PA with a differential but not a push–pull configuration is preferred because the first priority is now to have good performance with respect to nonlinearity and DC offset.

### 21.4.3 Impedance Matching

Siemens provides the primary data for a single device CLY10 as shown in Figure 21.16.

The first row of measured data in Figure 21.16 is close to the expected frequency range, 800–900 MHz. From the values of  $\Gamma_{mS,\text{mag}}$ ,  $\Gamma_{mS,\text{ang}}$ ,  $\Gamma_{mL,\text{mag}}$ , and  $\Gamma_{mL,\text{ang}}$  in this row, the input and output impedances of the device CLY10 at the frequency  $f = 850$  MHz



Measured data:

$f$ , GHz	$V_{dd}$ , V	$I_d$ , mA	$P_{1dB}$ , dBm	Gain, dB	$\Gamma_{mS,\text{mag}}$	$\Gamma_{mS,\text{ang}}$	$\Gamma_{mL,\text{mag}}$	$\Gamma_{mL,\text{ang}}$
0.9	3	700	26.7	15.3	0.58	169	0.68	-156
1.8	3	700	28.5	9.0	0.79	-110	0.75	-87
2.4	3	700	27.9	7.2	0.77	-86	0.73	-107

**Figure 21.16.** Siemens CLY10 power GaAs FET matching conditions.

can be approximately evaluated from the Smith Chart:

$$Z_{in} = 13.5 + j4.5 \Omega, \quad (21.11)$$

$$Z_{out} = 10.0 - j10.0 \Omega. \quad (21.12)$$

In order to reduce size and cost, the input and output impedance matching networks will be implemented by only capacitors and microstrip lines. An inductor is not welcome because the size of an inductor is much larger than that of a capacitor, and the cost of a chip inductor is about 10–20 times higher than that of a chip capacitor.

The input and output impedance networks can be optimized in terms of the “1 Port Opt” simulation. In this simulation, we assume that all the parts, including the capacitor and microstrip lines, are placed on a ceramic alumina substrate. As shown in Figures 21.16 and 21.17, the relative permeability of the substrate is  $\epsilon_r = 10.5$  and its thickness is 25 mil.

Figure 21.17 shows the setup for optimizing the input impedance matching network. At the right-hand side of the “1 Port Opt” is  $Z_{in} = 13.5 + j4.5 \Omega$  as shown in expression (21.11). At the left-hand side of the “1 Port Opt” is the input impedance matching network.

Figure 21.18 shows the setup for optimizing the output impedance matching network. At the left-hand side of the “1 Port Opt” is  $Z_{out} = 10 - j10.0 \Omega$  as shown in expression (21.12). At the right-hand side of the “1 Port Opt” is the output impedance matching network.

The primary consideration in the topology of the input and output impedance matching networks is that they are simple T-type matching networks. That is, one microstrip line segment is in series, one capacitor is in parallel, and one microstrip line segment is in series again. In the process of optimization, it is found that a microstrip open-stub can improve the impedance matching and widen the bandwidth. In addition, a bias and a  $V_{dd}$  providing point must be setup in the input and output impedance matching networks, respectively. This is why there are five pieces of microstrip line segments in total in either the input or output impedance matching network.

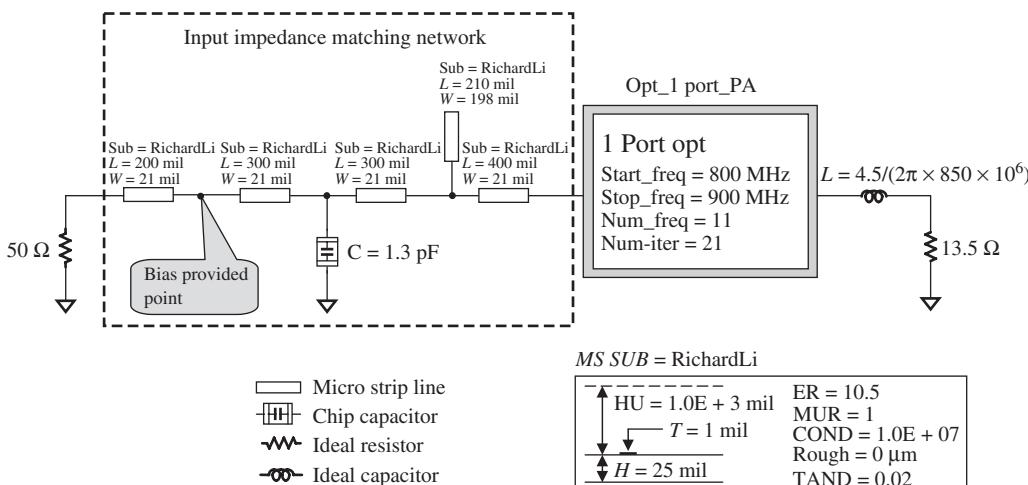


Figure 21.17. Optimization of CLY10 input impedance matching.

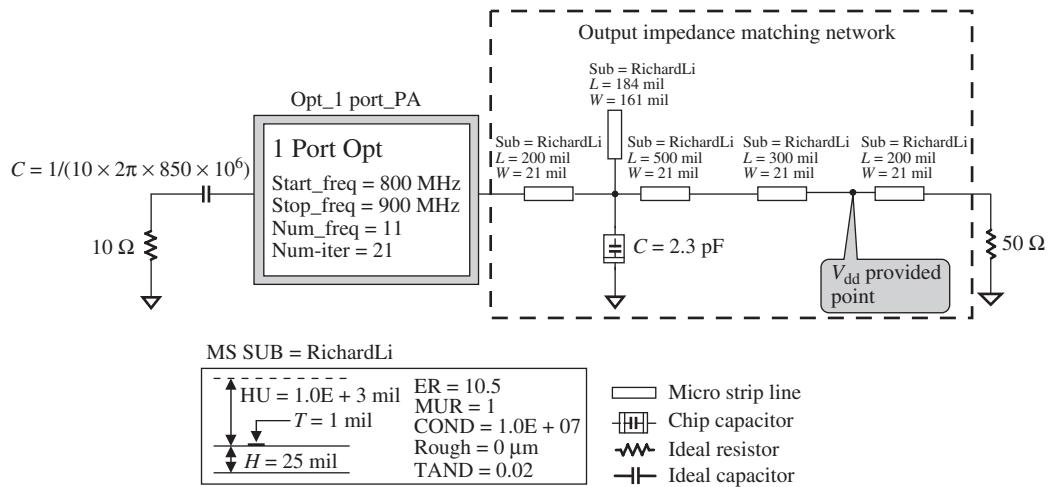


Figure 21.18. Optimization of CLY10 output impedance matching.

The final optimized values of the capacitor and microstrip lines are shown in Figures 21.17 and 21.18.

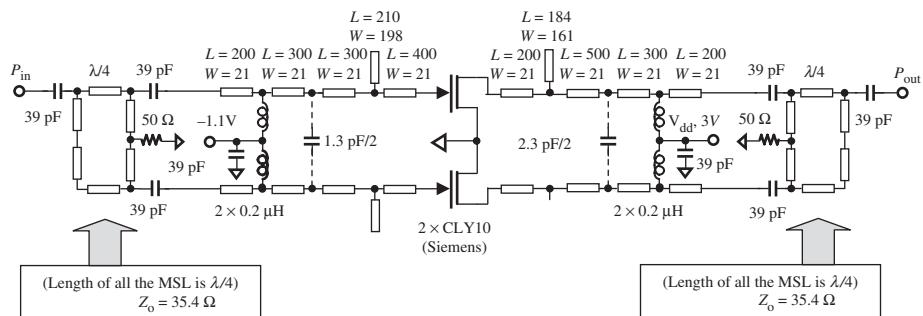
The impedance matching networks shown in Figures 21.17 and 21.18 are only half of the impedance matching network of the entire push–pull PA. Figure 21.19 shows the schematic of the entire push–pull 2-W PA.

The four inductors with values of  $L = 0.2 \mu\text{H}$  function as RF chokes. The input and output matching networks are connected to the input and output baluns, respectively.

The impedances of all the terminals in both the input and output baluns are  $50\ \Omega$ . Both baluns look like branch-type baluns built by microstrip lines. The length of each microstrip line segment is a quarter wavelength corresponding to the central frequency,  $f = 850\ \text{MHz}$ . The characteristic of the microstrip line segment is  $35.4\ \Omega$ . Figure 21.20 depicts the layout of the input or output balun on the ceramic alumina board.

Figure 21.21 depicts the entire layout of the PA on the ceramic alumina board.

The performance of this PA is outlined in Table 21.1. In addition, Figure 21.22 shows the curves of the output power and PAE versus input power. It can be seen that, when the input power less than 20 dB<sub>m</sub>, the output power and PAE are approximately



**Figure 21.19.** Schematic of 3-V, 2-W push-pull PA (unit of length: mil).  $f_o = 850$  MHz, substrate:  $H = 25$  mil,  $T = 1$  mil,  $\epsilon_r = 10.5$ .

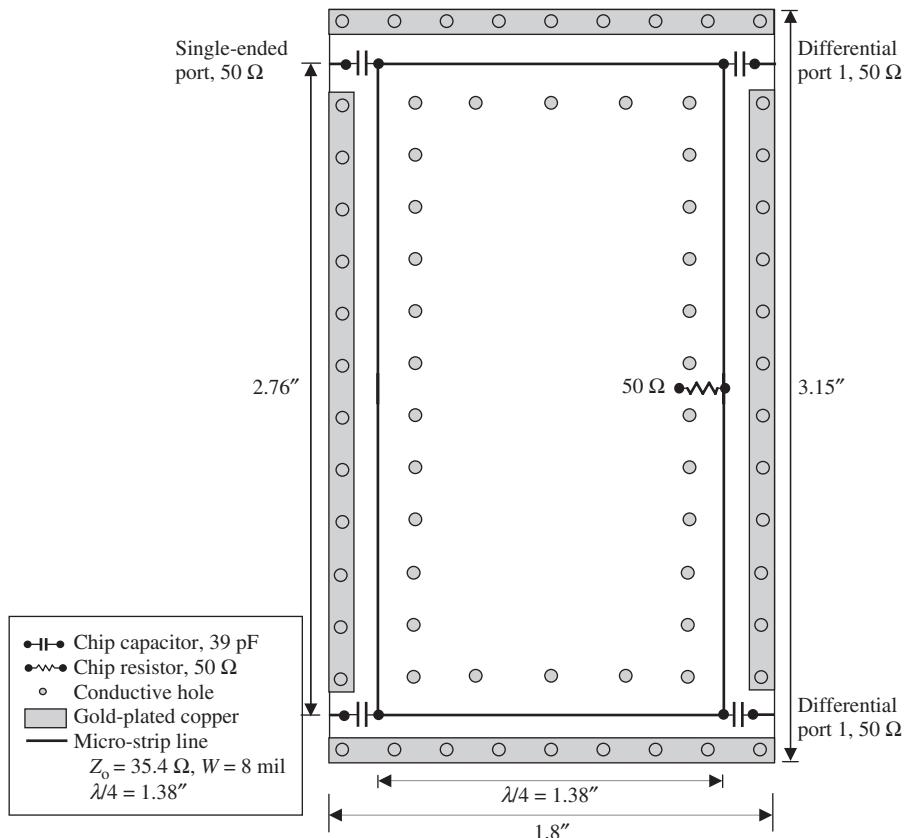


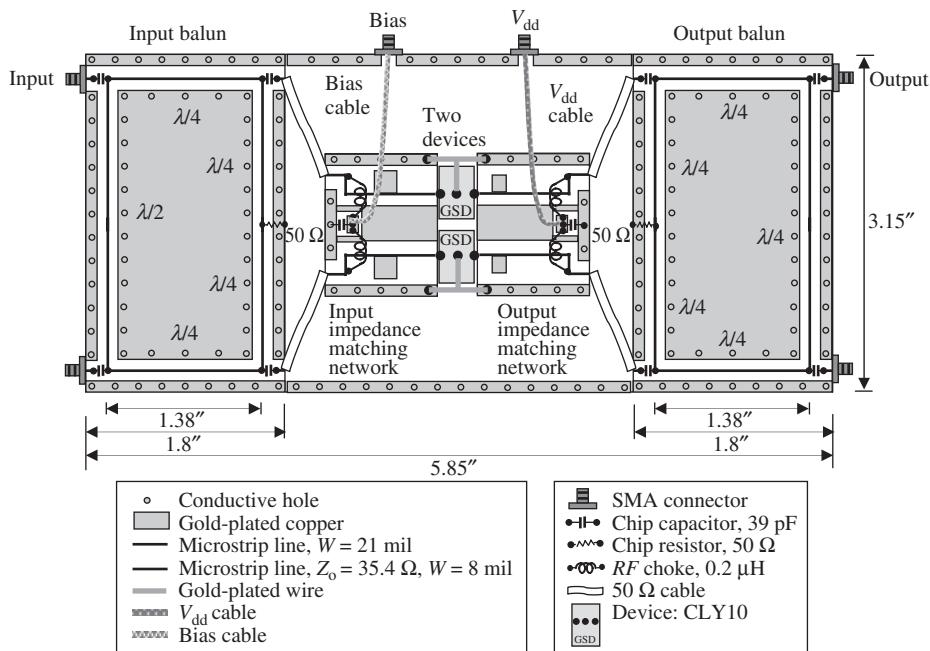
Figure 21.20. Layout of balun on a ceramic alumina board. Thickness = 25 mil,  $\epsilon_r = 10.5$ ,  $\mu_r = 1$ ,  $f = 800\text{--}900$  MHz.

proportional to the input power. The output power is saturated when the input power is nearly equal to or greater than  $20 \text{ dB}_m$ .

Figure 21.23 plots the curve of output power versus the DC power supply voltage  $V_{dd}$ . When  $V_{dd}$  is lower than 3 V, the output power is poor but is proportional to the value of  $V_{dd}$ . When  $V_{dd}$  is higher than 3 V, the output power does not increase much as  $V_{dd}$  increases. In parallel with the design for the 2-W PA, the design for a 5-W PA is also conducted in a similar way. The devices CS-1 are selected, which are manufactured by Motorola. In order to get more output power, the bias is changed from  $-1.1$  to  $-1.58$  V. Figure 21.24 shows the schematic of the entire push-pull 5-W PA after the input and output impedances of the devices are matched.

The performance of this PA is outlined in Table 21.2. In addition, Figure 21.25 shows the curves of the output power and PAE versus the input power. It can be seen that, when the input power is less than  $25 \text{ dB}_m$ , the output power and PAE are approximately proportional to the input power. The output power approaches saturation when the input power is nearly equal to or greater than  $25 \text{ dB}_m$ .

Figure 21.26 plots the curve of the output power versus the DC power supply voltage  $V_{dd}$ . When  $V_{dd}$  is lower than 3 V, the output power is poor, but is proportional to the value of  $V_{dd}$ . When  $V_{dd}$  is higher than 3 V, the output power does not increase much as  $V_{dd}$  increases.



**Figure 21.21.** Layout of push–pull PA module on a ceramic alumina board. Thickness = 25 mil,  $\epsilon_r = 10.5$ ,  $\mu_r = 1$ ,  $f = 800\text{--}900$  MHz.

**TABLE 21.1.** Main Performance of a 3-V, 2-W Push–Pull PA

$f_o = 850$ MHz
$P_{in} = 150$ mW
$P_{out} = 2.2$ W
$G = 12$ dB
PAE = 56.8%
$I = 1682$ mA

#### 21.4.4 Reducing the Block Size

Looking at Figure 21.20, one can find that the size of the entire PA block is unacceptably large. The largest parts in Figure 21.20 are the input and output baluns. We will therefore focus on shrinking the balun size. In the following discussion, only a 2-W PA is considered since a 5-W PA could be designed in the same way.

Figure 21.27 shows a modified balun built by a combination of microstrip lines and capacitors; its performance is shown in Figure 21.28. The magnitudes of  $S_{21}$  and  $S_{31}$  are kept within  $-3.5 \pm 0.2$  dB, and their phase difference is kept within  $180^\circ$  over the entire frequency bandwidth. The length of microstrip lines has been partially replaced by capacitors so that the total length of the microstrip line is reduced from 5520 to 3520 mil. Then, the layout of the push–pull PA is changed from Figures 21.21 to 21.29. The size of the entire the PA module is reduced by about 40%.

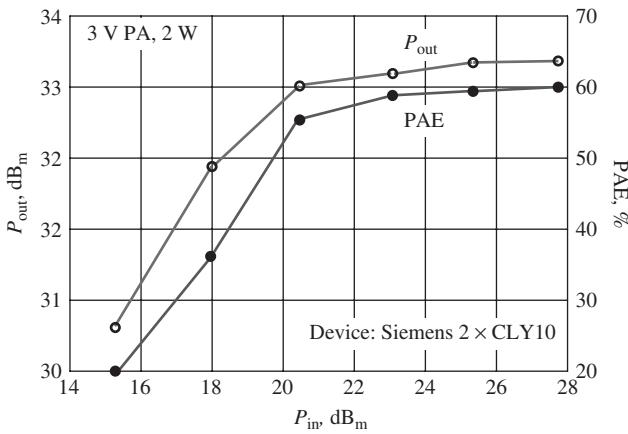


Figure 21.22. Plot of  $P_{\text{out}}$ , PAE versus  $P_{\text{in}}$ ,  $f_o = 850 \text{ MHz}$ , 2-W push-pull PA.

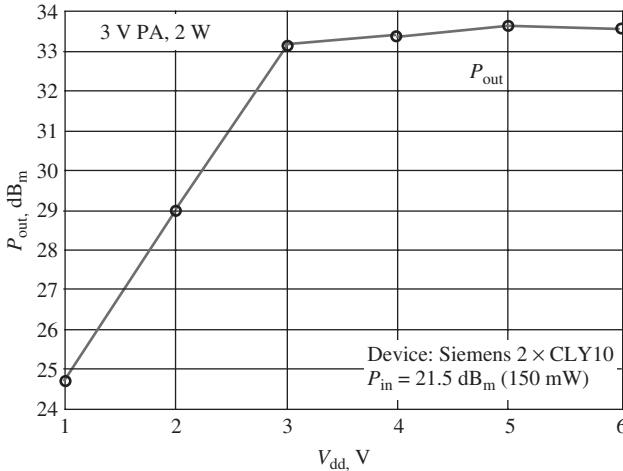


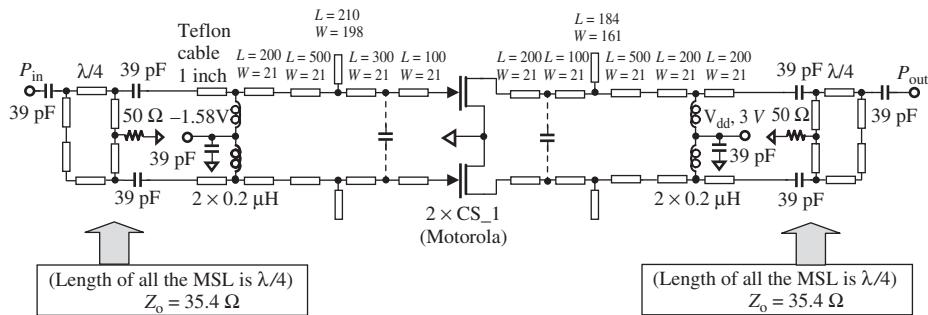
Figure 21.23. Plot of  $P_{\text{out}}$  versus  $V_{\text{dd}}$ ,  $f_o = 850 \text{ MHz}$ , 2-W push-pull PA.

TABLE 21.2. Main Performance of 3-V, 5-W Push-Pull PA

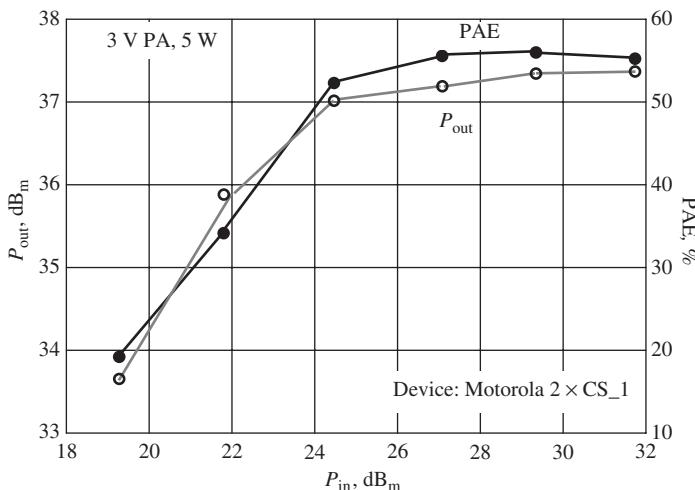
$f_o = 850 \text{ MHz}$
$P_{\text{in}} = 640 \text{ mW}$
$P_{\text{out}} = 5.5 \text{ W}$
$G = 9.5 \text{ dB}$
PAE = 58%
$I = 2162 \text{ mA}$

Despite these changes, the size of the push-pull PA shown in Figure 21.29 is still not reduced down to what is expected. Further effort is concentrated on three aspects:

1. Changing the input and output impedances of the device,  $Z_{\text{in}}$  and  $Z_{\text{out}}$ , to pure resistances  $R_{\text{in}}$  and  $R_{\text{out}}$  by connecting a capacitor or an inductor in parallel at the input or output, respectively. Recall the expression of  $Z_{\text{in}}$  (21.11) at the gate



**Figure 21.24.** Schematic of a 3-V, 5-W push-pull PA (unit of length: mil).  $f_o = 850$  MHz, substrate: ceramic alumina,  $H = 25$  mil,  $T = 1$  mil,  $\epsilon_r = 10.5$ .



**Figure 21.25.** Plot of  $P_{\text{out}}$ , PAE versus  $P_{\text{in}}$ ,  $f_o = 850$  MHz, 5-W differential PA.

$G$  of the device, and the expression of  $Z_{\text{out}}$  (21.12) at the drain D of device, that is,

$$Z_{\text{in}} = 13.5 + j4.5 \Omega, \quad (21.11)$$

$$Z_{\text{out}} = 10.0 - j10.0 \Omega. \quad (21.12)$$

From the Smith Chart, as shown in Figure 21.30, if  $Z_{\text{in}}$  at the gate G of the device is connected to a capacitor  $C = 4.12$  pF in parallel,  $Z_{\text{in}}$  will change from  $(13.5 + j4.5)\Omega$  to

$$Z_{\text{in}}' = R_{\text{in}}' = 15 \Omega. \quad (21.13)$$

And, if  $Z_{\text{out}}$  at the drain D of the device is connected to an inductor  $L = 3.75$  nH in parallel,  $Z_{\text{out}}$  will change from  $(10 - j10)\Omega$  to

$$Z_{\text{out}}' = R_{\text{out}}' = 20 \Omega. \quad (21.14)$$

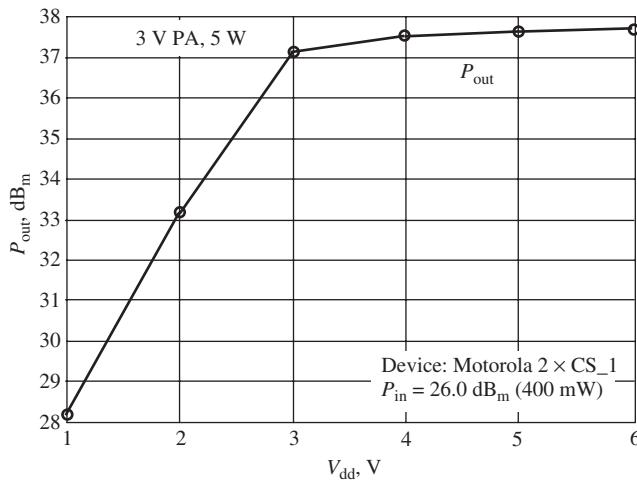


Figure 21.26. Plot of  $P_{\text{out}}$  versus  $V_{\text{dd}}$ ,  $f_o = 850$  MHz, 5-W differential PA.

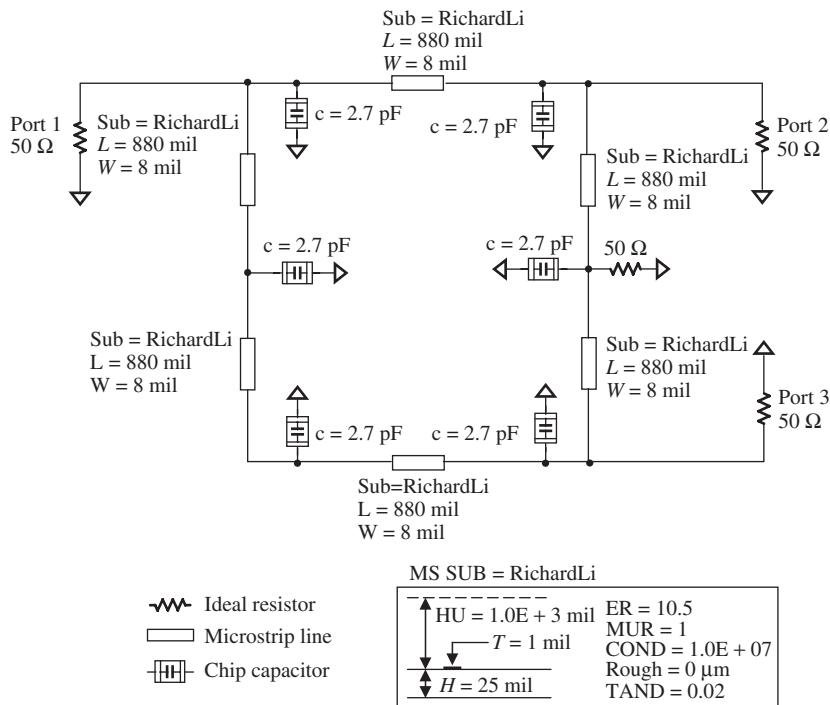
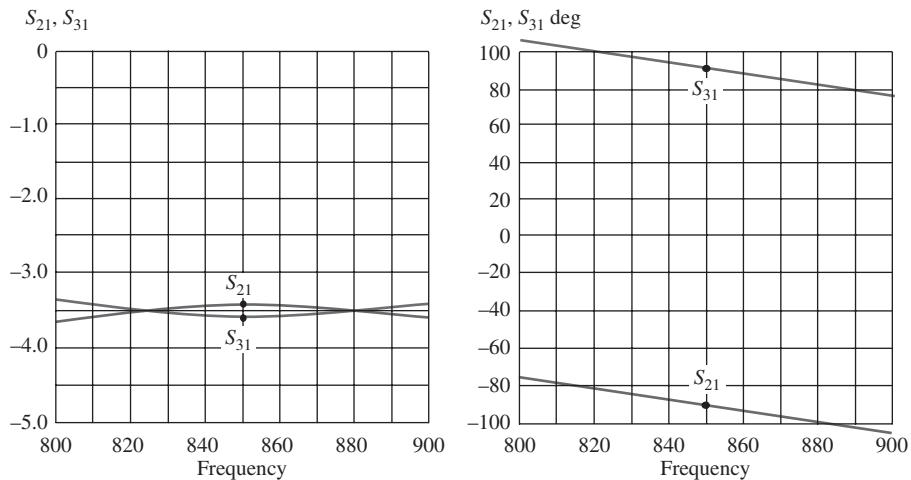


Figure 21.27. Modified balun by a combination of microstrip lines and capacitors.  $f = 800$ – $900$  MHz, balun:  $50 \Omega$  to  $2 \times 50 \Omega$ .

2. Removing the special impedance matching network as shown in Figure 21.21 or 21.29, and matching the new input and output impedances  $R_{\text{in}}$  and  $R_{\text{out}}$  directly by the input and output baluns, respectively.
3. Modifying the input and output baluns as shown in Figure 21.19 again to further decrease the size. All the microstrip line segments will be replaced by chip capacitors and inductors.



**Figure 21.28.**  $S_{21}$  and  $S_{31}$  of the balun built by microstrip lines and capacitors  $f = 800\text{--}900\text{ MHz}$ , balun:  $50\ \Omega$  to  $2 \times 50\ \Omega$ .

Figures 21.31 and 21.32 show the input balun and its performance, respectively. Figures 21.33 and 21.34 show the output balun and its performance, respectively.

The magnitudes of  $S_{21}$  and  $S_{31}$  in Figure 21.32 are kept within  $-3.75 \pm 0.5\text{ dB}$ , and their phase difference is kept within  $180^\circ$  over the entire frequency bandwidth. The magnitude of  $S_{21}$  and  $S_{31}$  in Figure 21.34 is kept within  $-3.75 \pm 0.5\text{ dB}$ , and their phase difference is also kept within  $180^\circ$  over the entire frequency bandwidth. Within the frequency bandwidth, their performances are acceptable.

Figure 21.35 shows that the size of the balun is reduced down from  $3.15 \times 1.8\text{ in.}^2 = 5.67\text{ in.}^2$  to  $0.9 \times 0.92\text{ in.}^2 = 0.83\text{ in.}^2$ . Figure 21.36 shows the size of the entire PA with the minimized balun. The capacitor  $C = 4.12\text{ pF}$  changes the input impedance of the device  $Z_{in}$  to a pure resistance  $R_{in}$ ; the inductor  $L = 3.75\text{ nH}$  changes the output impedance of the device  $Z_{out}$  to a pure resistance  $R_{out}$ . In addition, in order to avoid a crowded layout drawing, the parts in the input and output baluns are replaced by the parts' symbols.

As shown in Figure 21.21, the original size of the entire PA was  $5.85 \times 3.15\text{ in.}^2 = 18.43\text{ in.}^2$ . Now, from Figure 21.36, it can be seen that the size of entire PA is  $2.88 \times 1.10\text{ in.}^2 = 3.17\text{ in.}^2$ . The size of entire PA has been reduced by approximately 5.8 times its original size.

### 21.4.5 Double Microstrip Line Balun

As shown in Figure 21.36, the size of a push-pull PA can be shrunk to a reasonably small size if the baluns are completely built by chip capacitors and chip inductors.

However, this is not completely satisfactory because there are some drawbacks in the balun implementation. For instance, as shown in Figure 21.36, an input or an output balun will need at least six chip inductors and six capacitors, which would lead to

- a high cost for a total of 12 chip inductors;
- additional attenuation due to the 12 chip inductors;

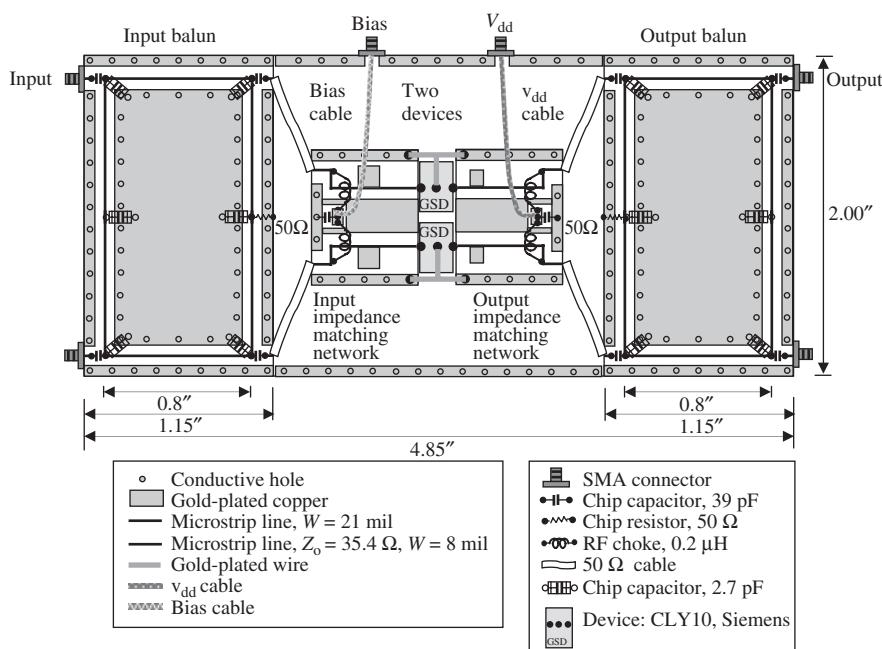
- the degradation of the performance of the balun, including inaccuracy or fluctuation of magnitude and phase, due to the parts' nonuniformity;
- the chip inductor being unable to bear a high PA current on it.

Therefore, it is very much desirable to develop a good balun for a push-pull PA. A good balun employed in a push-pull PA must have the following features:

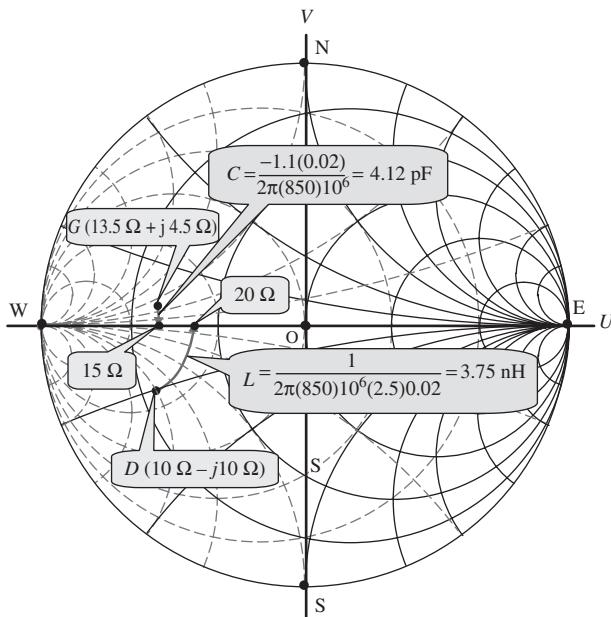
- low cost,
- very low insertion loss,
- ability to operate in high current or high power status,
- small size,
- simple configuration.

The primary idea to develop a double microstrip line balun is illustrated in Figure 21.37. (The content of this section is abstracted from a U.S. patent awarded to the author in 1993.)

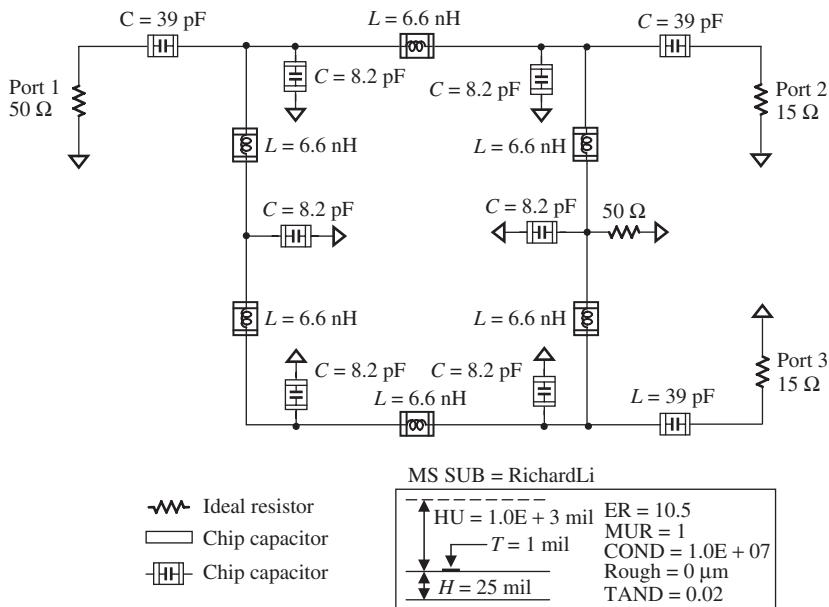
The model of the microstrip line can be represented by many “distributed” inductances in series as well as many “distributed” capacitances in parallel. The simple model of a double microstrip line in parallel can be represented by two individual microstrip line models connected together with many “distributed” coupling capacitors as shown in Figure 21.37. It is assumed that, if one of double microstrip lines is the primary path for the RF signal, its power can be coupled to the second microstrip line by those “distributed” coupling capacitors. These double lines could function like a transformer. If one end of the first microstrip line is grounded and the center of the second microstrip



**Figure 21.29.** Modified layout of push-pull PA on a ceramic alumina board. Thickness = 25 mil,  $\epsilon_r = 10.5$ ,  $\mu_r = 1$ ,  $f = 800\text{--}900$  MHz.



**Figure 21.30.** The drain D is impedance-matched by an inductor  $L = 3.75 \text{ nH}$  and the gate G is impedance-matched by a capacitor  $C = 4.12 \text{ pF}$ .



**Figure 21.31.** Input balun built by inductors and capacitors.  $f = 800\text{--}900 \text{ MHz}$ , balun:  $50 \Omega$  to  $2 \times 15 \Omega$ ; "zero" capacitor:  $C = 39 \text{ pF}$ .

line is also taped to the ground, then these double lines will look like single-ended to the differential pair balun.

If such a concept can be realized, its cost should become low. If the double microstrip lines are copper wires on a PCB or gold wires on an IC substrate, the insertion loss could be low because the resistance on the double lines is negligible. In addition, it can

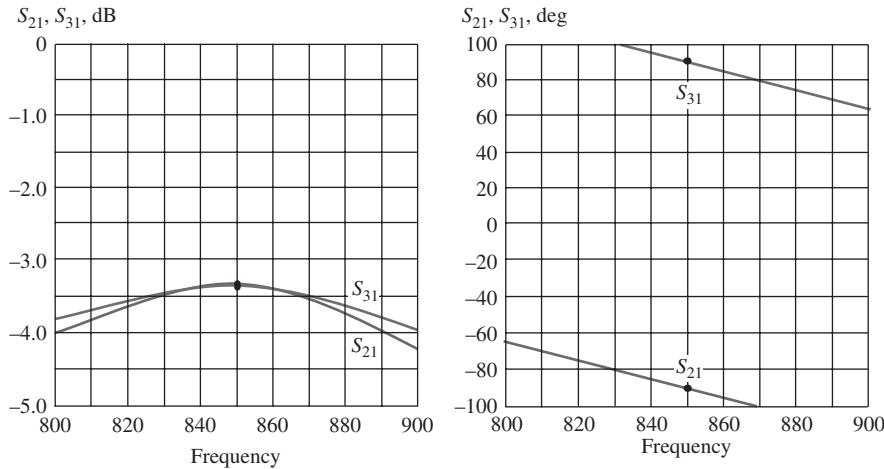


Figure 21.32.  $S_{21}$  and  $S_{31}$  of the input balun built by inductors and capacitors.  $f = 800\text{--}900 \text{ MHz}$ , balun:  $50 \Omega$  to  $2 \times 50 \Omega$ ; “zero” capacitor:  $C = 39 \text{ pF}$ .

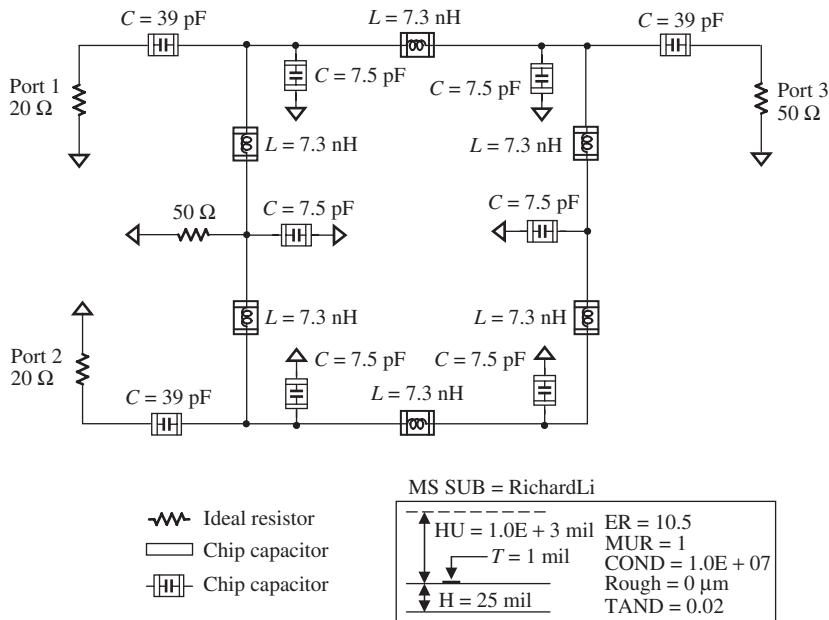
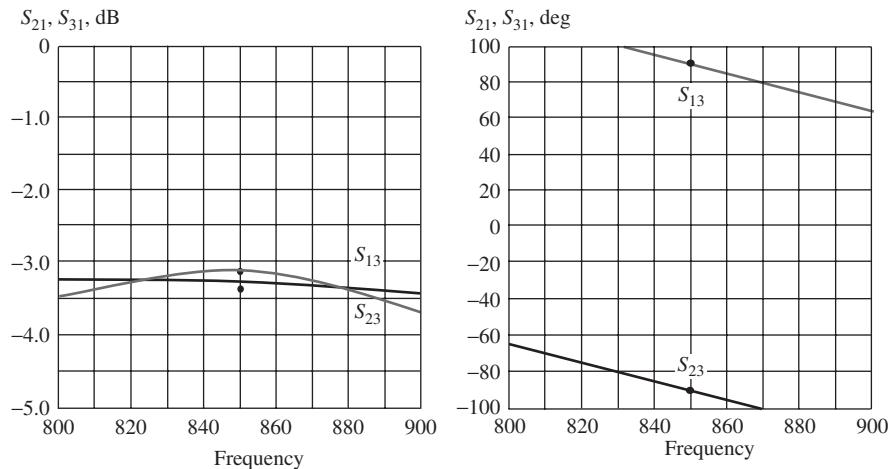


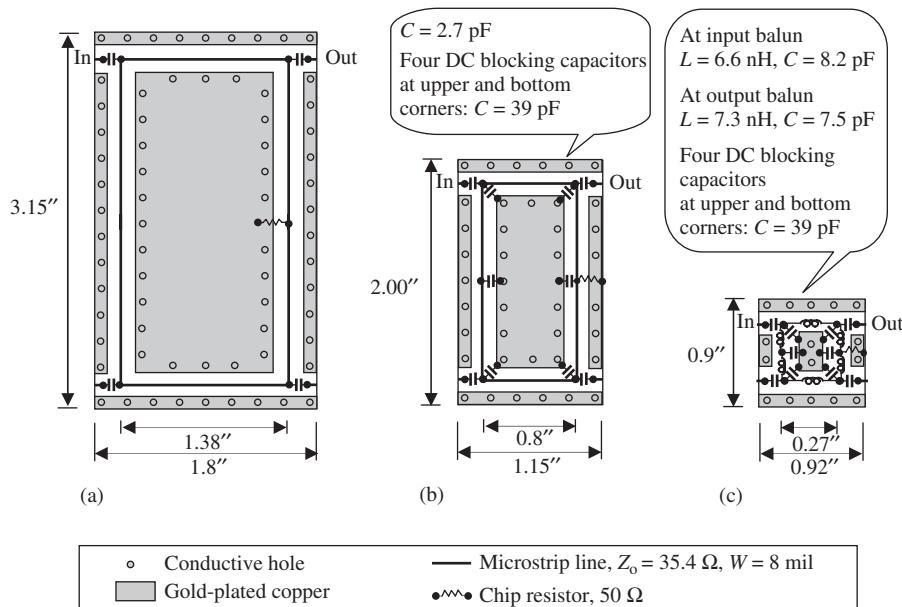
Figure 21.33. Output balun built by inductors and capacitors.  $f = 800\text{--}900 \text{ MHz}$ , balun:  $2 \times 20 \Omega$  to  $50 \Omega$ ; “zero” capacitor:  $C = 39 \text{ pF}$ .

afford high current or high power. The most outstanding advantage would be its simple configuration. Should small size be an issue, this setup can be built on a PCB or IC substrate with higher permeability.

The next thing to worry about is radiation, and hence interference with other parts in the circuitry. It would be helpful if these microstrip lines are enclosed by a grounded frame in order to prevent the magnetic flux from flying out to the sky. Figure 21.38 shows



**Figure 21.34.** S<sub>13</sub> and S<sub>23</sub> of output balun built by inductors and capacitors f = 800–900 MHz, balun: 2 × 20 Ω to 50 Ω; “zero” capacitor: C = 39 pF.



**Figure 21.35.** Evolution of balun applied to PA on a ceramic alumina board. Thickness = 25 mil;  $\epsilon_r = 10.5$ ;  $\mu_r = 1$ ; f = 800–900 MHz. (a) Balun built by MSL only. (b) Balun built by MSL. (c) Balun built by inductors and capacitors.

such a configuration. The double microstrip lines in parallel are printed and gold-plated on a small piece of alumina ceramic. They have the same width and length and are bent into a rectangular “S” shape. One of them functions as a single-ended or unbalanced port, while the other with a central grounded tap forms two balanced ports. Obviously, the operating principle of this balun is based on the strong edge coupling between the double lines.

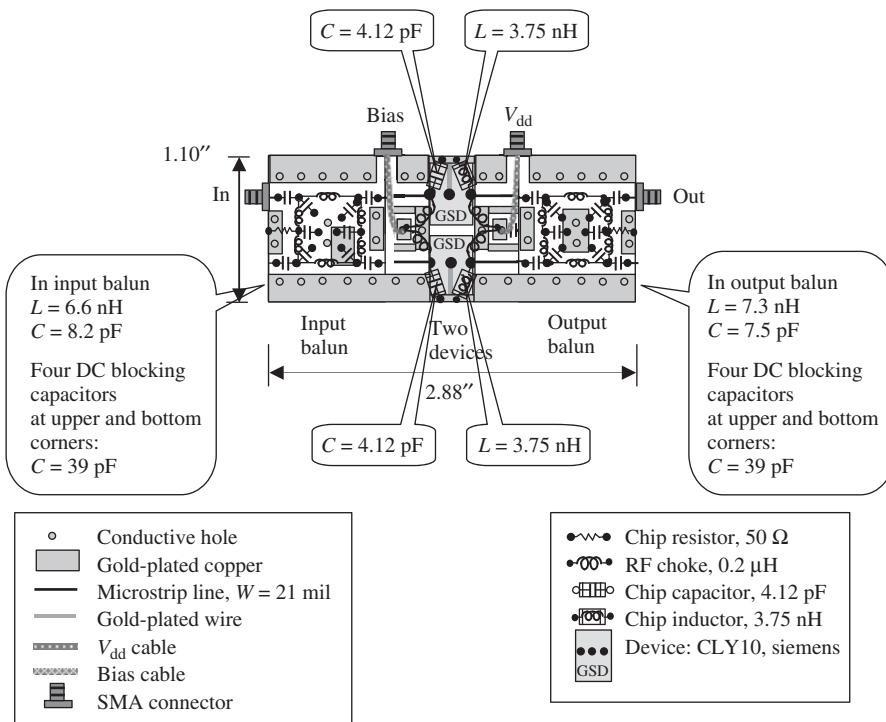


Figure 21.36. Minimized layout of push-pull PA on a ceramic alumina board. Thickness = 25 mil;  $\epsilon_r = 10.5$ ;  $\mu_r = 1$ ;  $f = 800\text{--}900 \text{ MHz}$ .

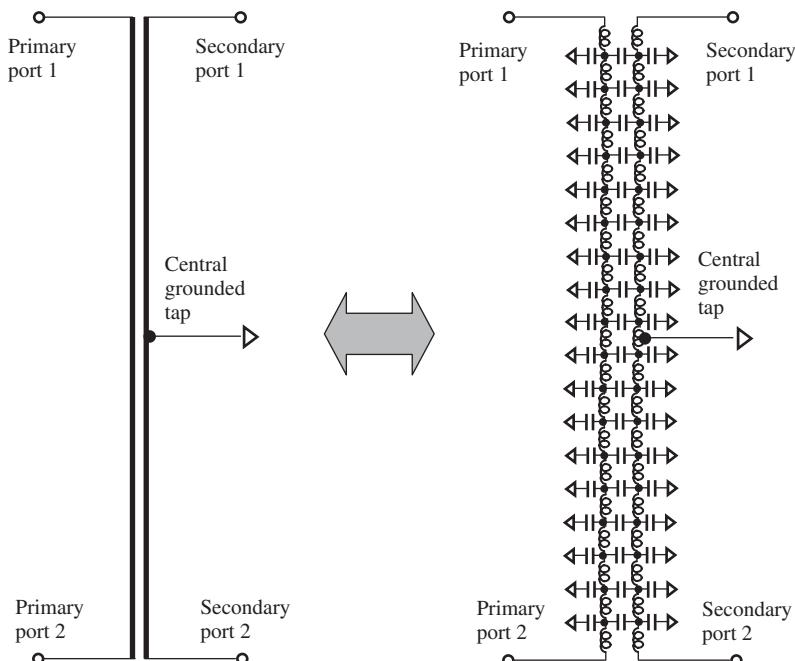
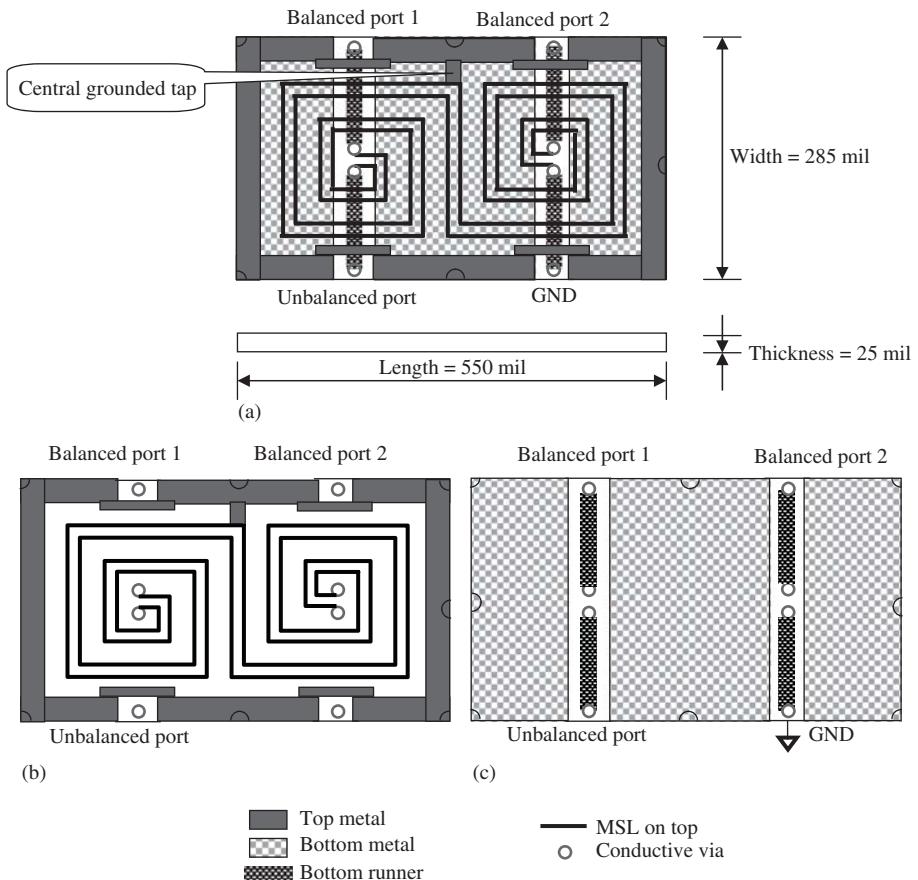


Figure 21.37. Conceptual drawing of double-line balun.



**Figure 21.38.** Double-line balun built on an alumina ceramic substrate. Thickness of substrate: 25 mil,  $\epsilon_r = 10.5$ , width of wire = 5 mil, spacing between double wire = 10 mil. (a) Double microstrip line balun. (b) Top side. (c) Bottom side.

In order to minimize insertion loss, two issues are important:

1. One of them is its “S-type” configuration. Because of the “S” shape, the windings are clockwise on the left-hand side and counterclockwise on the right. Consequently, the magnetic flux generated by the loop on the left-hand side will automatically bend down into the loop on the right-hand side. The energy will not be scattered out of the balun block, thereby lowering the insertion loss and preventing interference with other parts. As a matter of fact, in the first design attempt, the loops on both left- and right-hand sides had the same clockwise winding and resulted in a higher insertion loss.
2. On the other hand, the total electrical length of the double line is about half the wavelength of the central frequency. The advantage of using the half wavelength is that the energy of the standing wave is confined to the central area of the substrate. In other words, the leakage of energy, and hence the insertion loss, can be reduced to a lower level.

In order to minimize the size, four aspects of the design have been considered:

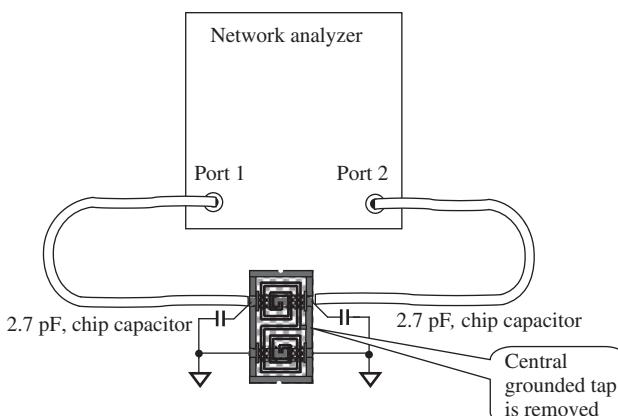
1. The material of the substrate is chosen as ceramic alumina because its dielectric constant is high, which in our case is  $\epsilon_r = 10.5$ . Theoretically, the higher the dielectric constant, the smaller the size of the balun. However, a substrate with a higher dielectric constant brings about higher insertion loss and higher tolerance. As a matter of fact, there is a trade-off between the size and the insertion loss or tolerance.
2. Instead of being circular, the shape of the loop is rectangular.
3. The width of the lines and the spacing between the lines are 5 mil each, which is the limit of the size in a thin-film lab.
4. The use of double lines alone is not enough to achieve low insertion loss and to minimize the size at the same time. A resonant condition could be created by adding capacitors to assist the two windings. The designed double-line balun is supposed to be applied in practical implementations of circuitry together with capacitors.

Many attempts have been made -to optimize its size, including the length of the double lines, the width of the microstrip line, and the spacing between double lines. In order to simplify the testing, one of the balanced ports is grounded while its central grounded tap is removed. The testing is then changed from testing for three ports to testing for two ports without loss of generality. In actual application, the test result corresponds to the condition where the balanced ports are treated as a terminal with  $25 \Omega$  of impedance, instead of  $50 \Omega$ . The test setup is shown in Figure 21.39, where the capacitors (2.7 pF) are the additional parts as mentioned previously and are specially chosen for these specific dimensions of dual line balun:  $L \times W \times T = 550 \times 285 \times 25$  mil.

Figure 21.40 shows their typical frequency response and return loss. The successful modules are listed in Table 21.3. The outstanding performance is the insertion loss,

$$\text{IL} < 0.6 \text{ dB, when } 485 \text{ MHz} < f < 940 \text{ MHz.} \quad (21.15)$$

As is well known, in the high RF frequency range the insertion loss of a ferrite transformer balun is about 2–3 dB. Therefore, it is impossible to employ it in a PA



**Figure 21.39.** Setup for testing of double-line balun.

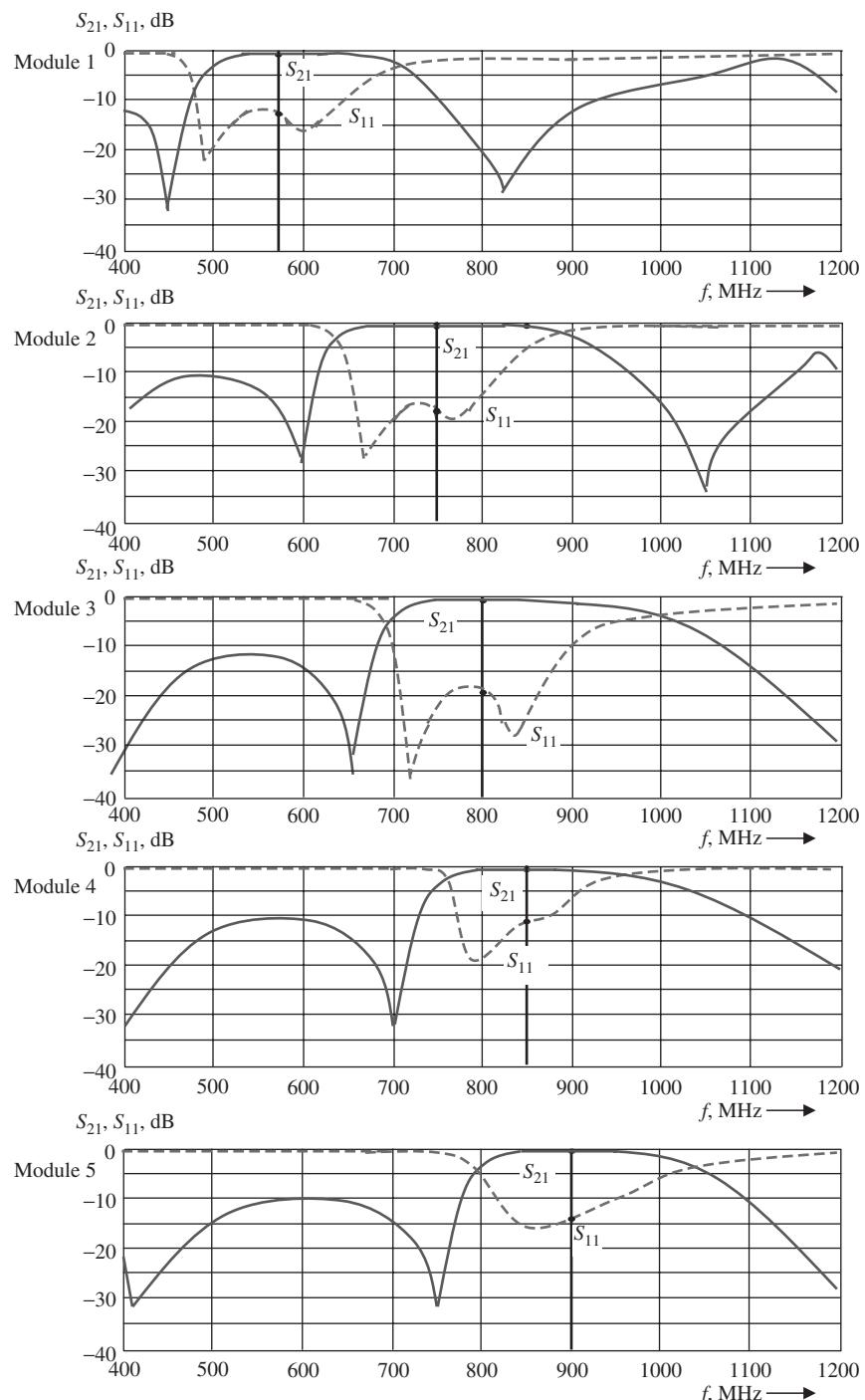


Figure 21.40.  $S_{21}$  and  $S_{11}$  of double-line balun.

TABLE 21.3. Configuration and Performance of Double Microstrip Line Balun

	Module 1	Module 2	Module 3	Module 4	Module 5
<i>Substrate</i>					
Material	Alumina ceramic	-----	-----	-----	→
Size, mil	$L \times W \times T = 550 \times 285 \times 25$	-----	-----	-----	→
<i>Microstrip Line</i>					
Length, mil	5456	3463	3174	3034	2813
Width, mil	5	10	10	10	10
Spacing	5	5	5	5	5
<i>Performance</i>					
Frequency, MHz	485–615	685–815	785–885	800–900	850–940
BW, MHz	130	130	100	100	90
Insertion loss, dB	0.40	0.53	0.46	0.53	0.59
Return loss, dB	< -11	-15	< -18	< -12	< -11

circuit block. Some other baluns, such as the ring or “Ratrace” baluns, might have the same advantage of low insertion loss. However, they are usually much larger than the double-line balun. With the exception of these disadvantages in a double-line balun, its simple configuration and low cost are superior to other types of baluns in the application in a PA.

As a matter of fact, the size of the double microstrip line balun can be further reduced if the dielectric constant of the alumina ceramic can be further increased. A size reduction of more than 50% is possible.

#### 21.4.6 Toroidal RF Transformer Balun

The idea of the toroidal RF transformer balun comes directly from that of the ferrite transformer. Figure 21.41 is a conceptual drawing in which the primary clue to implementing a transformer on a piece of alumina ceramic substrate is illustrated.

Figure 21.42 shows the designed configuration of a toroidal RF transformer balun. By using thin-film technology, a transformer is constructed on a ceramic alumina substrate with a high dielectric constant. Its basic components are the microstrip line segments and vias. There are two windings on the substrate, the single-ended and differential

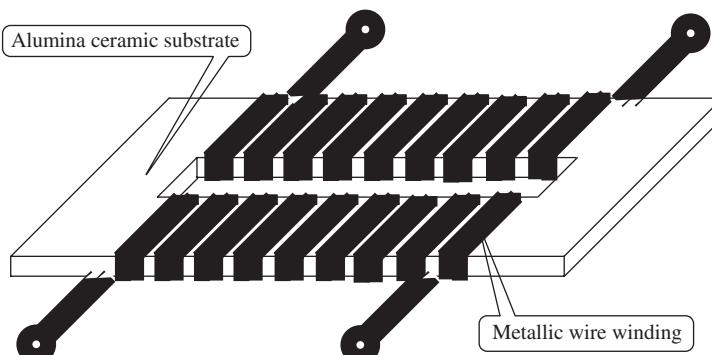
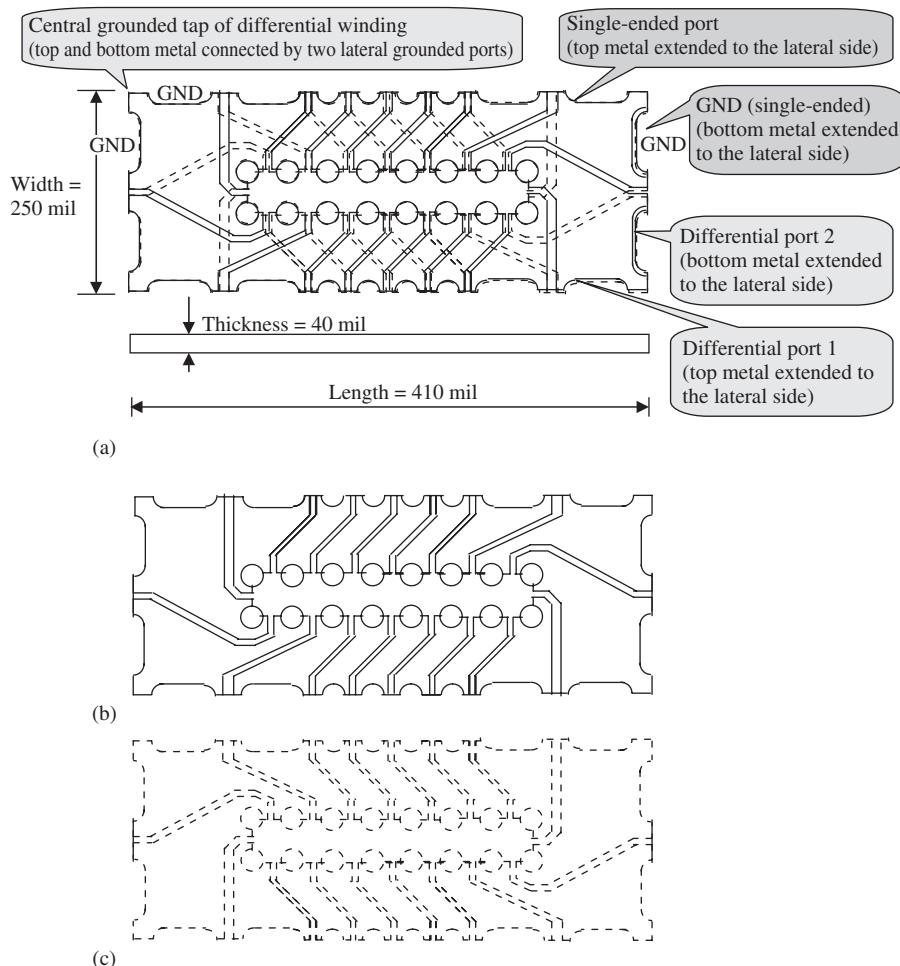


Figure 21.41. Conceptual drawing of a toroidal RF transformer balun.



**Figure 21.42.** Toroidal balun built on ceramic alumina substrate.  $\epsilon_r = 10.5$ , thickness = 40 mil, spacing = 5 mil, diameter of via = 20 mil. (a) Layout of toroidal balun; substrate = ceramic aluminate,  $\epsilon_r = 10.5$  for  $f = 800\text{--}1650$  MHz. (b) Top view of toroidal balun layout. (c) Bottom view of toroidal balun layout.

windings. The single-ended winding starts from the soldering port on the lateral side of the substrate (single-ended port 1 shown in Figure 21.42(a)) and runs between the top and bottom sides connected by conductive vias. Finally, it terminates at another soldering port on the lateral side (GND port in Figure 21.42(a)), which is near the starting single-ended port. The differential winding starts from the lateral side of the substrate (differential port 1 shown in Figure 21.42(a)) and also runs between the top and bottom sides connected by conductive vias. Finally, it terminates at another soldering port on the lateral side (differential port 2 in Figure 21.42(a)), which is near the starting single-ended port. At the middle, there is a central tap by two lateral grounded ports (two GNDs shown in Figure 21.42(a)). The two differential and single-ended windings are interlaced and strongly coupled with each other. No additional parts are needed.

The characteristics, of course, depend on many factors, such as the total length of the windings, the average width of the windings, the spacing between the windings, the

diameter of the vias, the thickness of the substrate, and the dielectric constant of the substrate.

In order to minimize the insertion loss, two aspects have been considered:

1. The toroidal configuration of the transformer confines most of the RF energy within the toroidal windings. Consequently, the leakage of power, and hence insertion loss, can be reduced to a very low level.
2. Therefore, the spacing between the two segments is set to 5 mil only, which is the limit of the size in a thin-film lab.

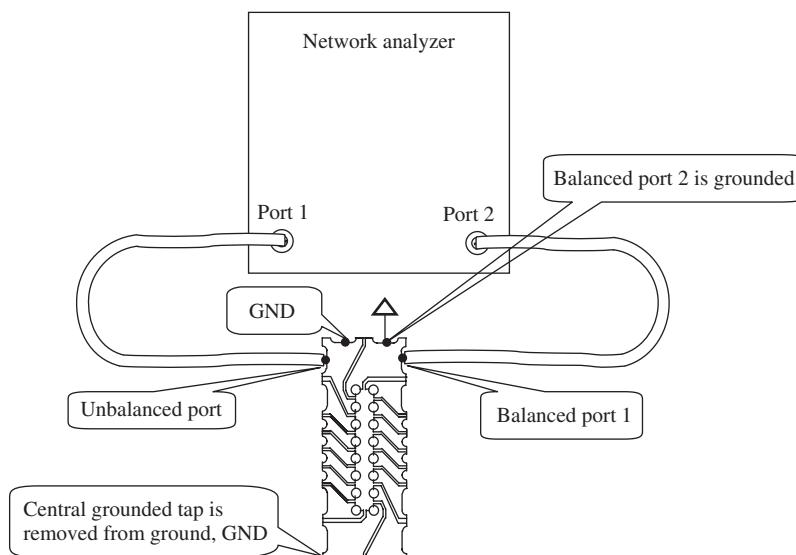
In order to minimize the size, three aspects have been considered:

1. The material of the substrate is chosen to be ceramic alumina with a dielectric constant equal to 10.5. Theoretically, the higher the dielectric constant, the smaller the size of the balun. However, a substrate with a higher dielectric constant will introduce a higher loss. Therefore, it is a trade-off between the size and the insertion loss.
2. Instead of being circular, the shape of the loop is rectangular.
3. The diameter of the vias is 20 mil only, which is a critical limit in the thin-film lab.

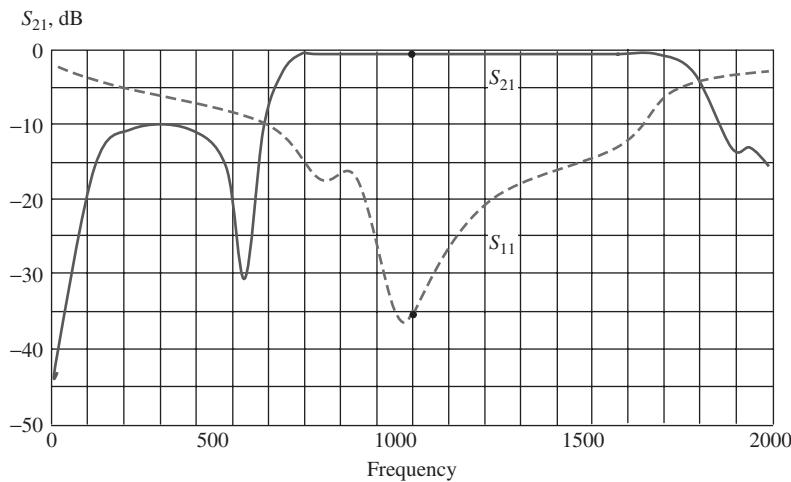
In order to simplify the testing work, the central grounded tap is removed from the ground (GND), and the test setup is as shown in Figure 21.43. This is for the test convenience in the  $50 - \Omega$  system only. It, however, would not lose generality. The test results can be applied in actual circuitry when the impedance of a single-ended port is  $50 \Omega$  while the impedance of each differential port is  $25 \Omega$ , with  $180^\circ$  phase difference.

Figure 21.44 shows one typical test result.

The outstanding performance is due to



**Figure 21.43.** Setup for testing of toroidal RF transformer balun.



**Figure 21.44.** Frequency response of the toroidal RF transformer balun. Balun size:  $L \times W \times T = 410 \times 250 \times 40$  mil. Substrate: ceramic alumina,  $\epsilon_r = 10.5$ . Microstrip line: width = 35 mil; spacing = 5 mil; via (diameter) = 20 mil.

- the extremely low insertion loss, and
- the extremely wide bandwidth.

From Figure 21.44, it can be seen that

$$0 > S_{21} > -0.35 \text{ dB, when } f < 600 \text{ MHz, and } f < 1650 \text{ MHz,} \quad (21.16)$$

$$S_{11} < -10 \text{ dB, when } 600 \text{ MHz} < f < 1650 \text{ MHz,} \quad (21.17)$$

$$S_{11} = -35.3 \text{ dB, when } f = 1000 \text{ MHz.} \quad (21.18)$$

The insertion loss of this module is less than 0.35 dB over a relative bandwidth of 93%!

Its performance is better than the double line mentioned previously. Also, it is much better than the RF balun developed by MLC (multilayer ceramic) technology. Typically, the insertion loss of the MLC balun is about 1.5–2 dB over a relative bandwidth of less than 50%. Other good features of the toroidal RF transformer balun are the following:

- *Small Size.* Usually, MLC baluns are designed with  $\lambda/4$  or  $\lambda/2$  microstrip lines. Their size is larger and their bandwidth much narrower than those of a toroidal RF transformer.
- *Simple and Reliable Configuration.* Indeed, it appears as merely a small piece of ceramic with some metallic pattern.
- *Low Cost.* It contains only a top and bottom metallic layer while the MLC contains more than two metallic layers. Therefore, the former has a lower cost than the latter.

Successful modules have been implemented, and their performances are similar to the module shown above. Table 21.4 lists their configuration and performance.

TABLE 21.4. Configuration and Performance of Double Microstrip Line Balun

	Module 1	Module 2	Module 3	Module 4	Module 5	Module 6	Module 7	Module 8
<i>Substrate</i>								
Material	Alumina ceramic							
$\epsilon_r$	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5
Length, mil	1070	1070	750	750	570	490	410	330
Width, mil	570	570	750	750	250	250	250	250
Thickness, mil	0	40	2 × 40 <sup>a</sup>	2 × 40 <sup>a</sup>	40	40	40	40
<i>Microstrip line</i>								
Width, mil	25	35	35	35	35	35	35	35
Spacing	5	5	5	5	5	5	5	5
Diameter of via., mil	20	20	20	20	20	20	20	20
<i>Performance</i>								
Frequency, MHz	200–600	350–1150	160–500	300–1000	450–850	700–1250	850–1650	1200–2200
BW, MHz	400	800	340	700	400	550	850	1000
Insertion loss, dB	0.31	0.24	0.15	0.15	0.20	0.26	0.25	0.32
Return loss, dB	< -15	< -12	< -18	< -15	< -12	< -13	< -12	< -14

<sup>a</sup>Note: these two modules consists of two identical substrates “piggy-backing” together.

## 21.5 PA WITH TEMPERATURE COMPENSATION

There are many ways to compensate the change of a PA's power due to the variation of temperature. We will introduce a simple and easy means to ensure temperature compensation for a PA (Yamauchi et al., 2001).

In general, the gain of a MOSFET amplifier increases when its gate voltage is increased.

On the other hand, the gain of a MOSFET amplifier can be increased or decreased when the temperature is increased. In order to compensate for the gain variation with temperature of an amplifier, changing the gate voltage of its MOSFET is the right approach.

Figure 21.45 shows the schematic diagram of the temperature compensation circuit at the gate of a MOSFET. The value of the gate voltage  $V_g$  is determined by the reference voltage  $V_{ref}$ , the gate control voltage  $V_c$ , and the current  $I_d$  flowing through the diode  $D_1$ . The current flowing through the diode is dependent on the temperature. If the current flowing through the diode increases as the temperature increases, that is,

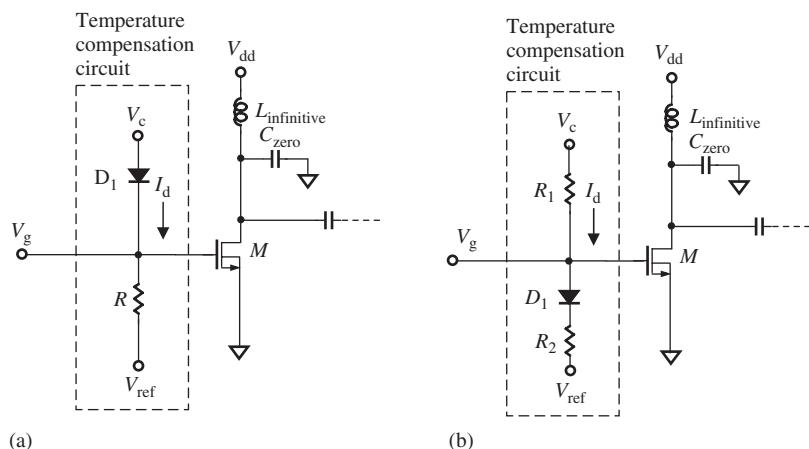
$$\frac{\partial I_d}{\partial T} > 0, \quad (21.19)$$

then the temperature compensation circuit shown in Figure 21.45(a) can be adopted. The gate voltage  $V_g$  is increased as the temperature increases. Consequently, the decreased gain of the MOSFET due to temperature is compensated for by the increase of gain due to the increase of the gate voltage.

On the contrary, if the current flowing through the diode decreases as the temperature increases, that is,

$$\frac{\partial I_d}{\partial T} < 0, \quad (21.20)$$

then the temperature compensation circuit as shown in Figure 21.45(b) can be adopted. The gate voltage  $V_g$  is also increased as the temperature increases. Consequently, the



**Figure 21.45.** Temperature compensation circuit at the gate of a MOSFET. (a)  $\frac{\partial V_g}{\partial T} > 0$ . (b)  $\frac{\partial V_g}{\partial T} < 0$ .

decreased gain of the MOSFET due to temperature is compensated for by the increase of gain due to the increase of the gate voltage.

## 21.6 PA WITH OUTPUT POWER CONTROL

It is desirable to have output power control for power amplifiers in a communication system. Among the power controls, digital control is the most convenient and effective. One such digital control is shown in Figure 21.46.

The power control is set up in the preamplifier stage. The four control terminals can be controlled by the two bits of the multidelexer.

Another type of power control is to add a programmable attenuator before the PA. Figures 21.47–21.49 show this technique (Khannur, 2003).

The programmable attenuator consists of 16 individual attenuators with attenuation from 0 to 7.5 dB in 15 steps. The attenuation is increased by 0.5 dB per step. These 16 individual attenuators are controlled by four bits of the demultiplexer. The total attenuation can be controlled from 0 to 60 dB.

Each individual attenuator has two switched MOSFETs. Their gates are called  $SW_a$  and  $SW_b$ . They are connected together and are controlled by the demultiplexer. The attenuator itself consists of three resistors with a  $\pi$ -type configuration. The attenuation depends on the values of resistors, which can be calculated through a simple mathematical derivation.

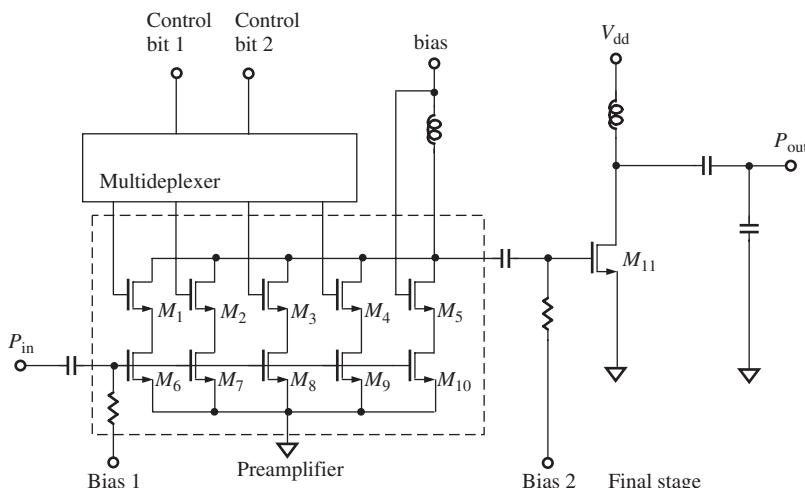


Figure 21.46. Digital power control for MOSFET power amplifier.

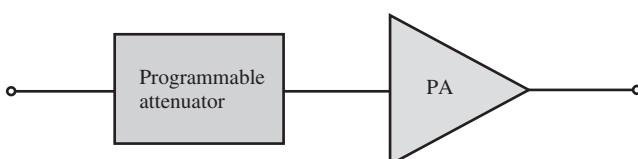


Figure 21.47. PA with programmable attenuator.

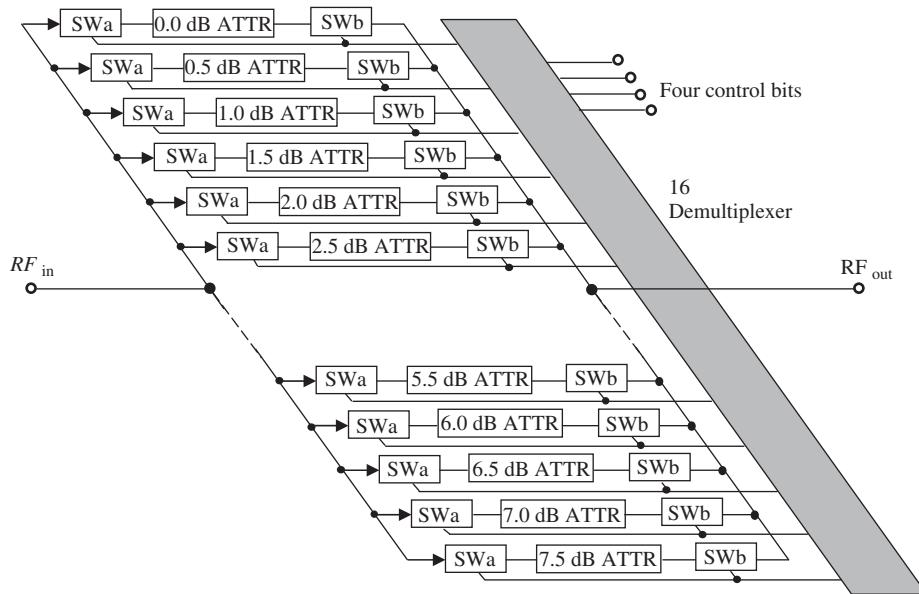


Figure 21.48. Full programmable attenuator.

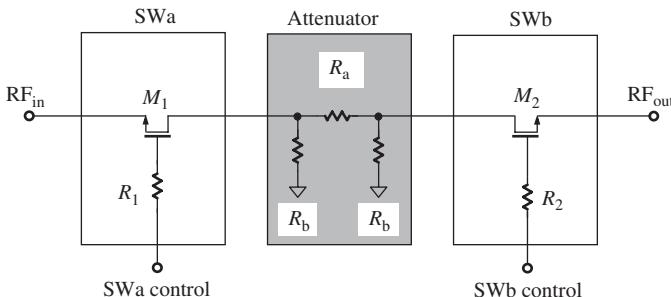


Figure 21.49. Individual attenuator.

## 21.7 LINEAR PA

There are two kinds of nonlinearities:

- The input–output power relationship (AM–AM conversion), and
- The input–output phase relationship (AM–PM conversion).

They are uniquely related through the third-order distortion phase. The nonlinearity of a device can be represented by two components: the in-phase nonlinearity and the quadrature nonlinearity.

There are many different linearization techniques for the PA, such as given below:

- *Predistortion (PD)*. This technique uses a lookup table of premeasured corrections at the desired modulation values to predistort the input. This predistortion technology is sometimes called *the complementary distortion technique*. In order to

achieve the optimal compensation condition, the PD circuit must have independently adjustable amplitude and phase for its third-order distortion output. This type of PD circuit is referred to as a CPL (*cuber predistortion linearizer*). Theoretical investigation indicates that the  $\text{IM}_3$  components of a power amplifier could be suppressed by a CPL.

The PD method is superior to the feed-forward or continuous-feedback error-correction techniques because of its simpler circuit configuration and low power consumption.

- *Feed-Forward Error Cancellation*. This technique compares the output or input to develop an error signal for summing with the output.
- *Continuous-Feedback Error Cancellation*. This technique compares the output or input to develop an error signal for predistorting the input.
- *Cartesian Feedback*. The idea of using Cartesian feedback to linearize power amplifiers has been discussed as early as the 1970s. This technique is called Cartesian feedback because the feedback is based on the Cartesian coordinates of the baseband symbols  $I$  and  $Q$ , as opposed to polar coordinates. Fundamentally, the concept behind this system is negative feedback. Figure 21.50 shows the block diagram of a typical Cartesian feedback system (Dawson and Lee, 2004).
- *Transistor-Level Linearized PA*. System-level methods, such as predistortion, feed-forward error cancellation, and Cartesian feedback, have been proposed to linearize the power amplifier. However, they all require complicated hardware, and may be suitable only for repeaters or base stations.

Now, let us introduce the transistor-level linearization for the PA, which may be more appropriate for power amplifiers in the handset (Tanaka et al., 1997).

It is well known that a power amplifier with a differential configuration rejects the even orders of nonlinearity very effectively, which is widely used in accurate continuous-time analog MOSFET circuits. However, there seems to be little special mention of the third-order and other odd orders of distortion and nonlinearity. The odd-order nonlinearities become the main components of the nonlinearity in a differential PA. The interference

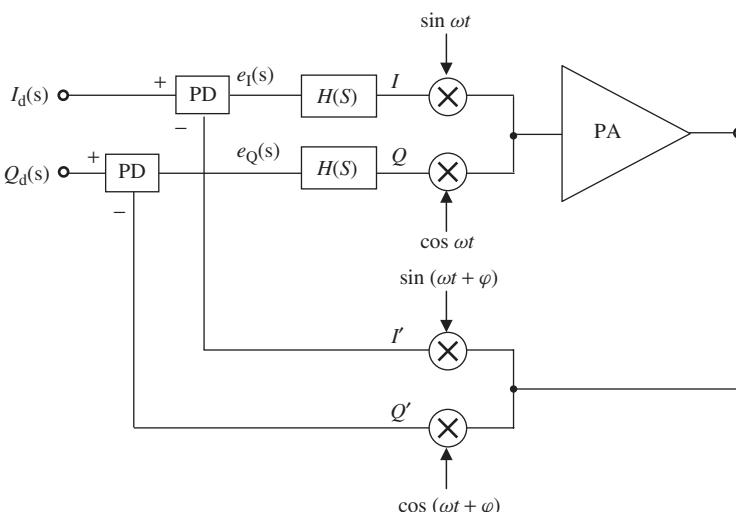


Figure 21.50. A typical Cartesian feedback system.

source of the odd nonlinearity definitely exists because any two signals within the operating bandwidth can be considered as interference sources to the adjacent channels. Figure 21.51 illustrates that the two signals are located within the main channels and their third-order nonlinear products appear in the adjacent channels. Any pair of signals separated by  $\Delta f$  within the main channels is an interference source of the third-order nonlinear products to the adjacent channels.

Instead of linearization of the PA by a complicated circuit loop, Tanaka et al. (1997) suggested the linearization of the PA at the transistor level. Similar to the multi-tanh devices applied to linearize a mixer, a compound MOSFET stage is developed and applied to linearize a PA.

Figure 21.52 shows the characteristics of a MOSFET by two curves.

It is well known that the curve of  $I_d$  versus  $V_i$  can be divided into two regions. One is the saturated region and another is the triode. The third-order nonlinearity coefficient  $a_3$  depends on  $V_i$  with a somewhat complicated relationship. As a matter of fact, the

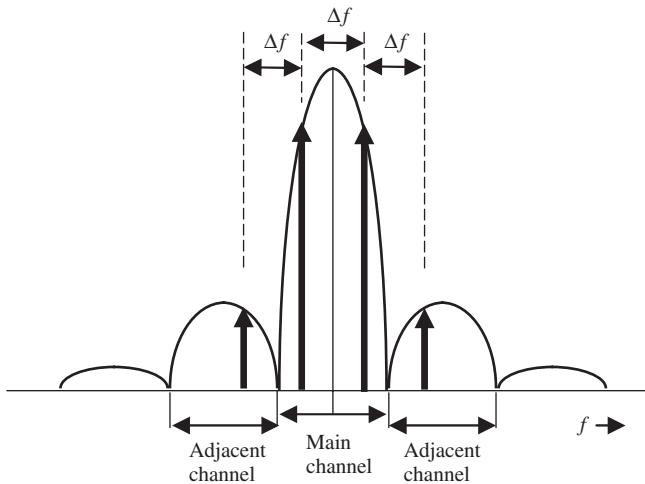


Figure 21.51. Two signals in main channel being the sources of interference due to third-order nonlinearity.

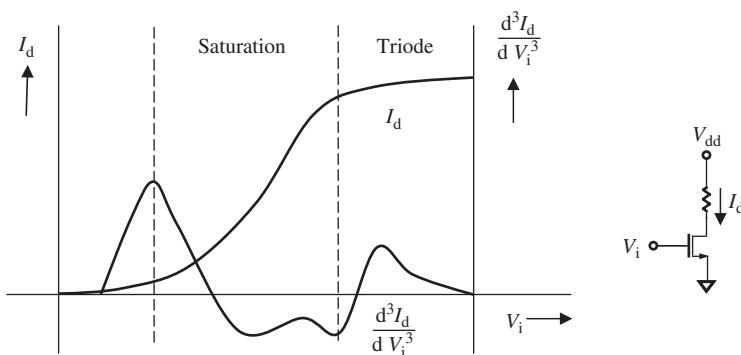


Figure 21.52. Characteristics of MOSFET.  $I_d$  versus  $V_i$  and  $\frac{d^3 I_d}{d V_i^3}$  versus  $V_i$ .

third-order differentiation of  $V_i$  is equal to  $6a_3$ , that is

$$\frac{d^3 I_d}{d V_i^3} = 6a_3, \quad (21.21)$$

if the transfer function of the MOSFET can be expressed as

$$I_d = a_0 + a_1 V_i + a_2 V_i^2 + a_3 V_i^3 + \dots \quad (21.22)$$

Figure 21.53(a) shows a new compound MOSFET in which the bottom MOSFET  $M_1$  is operated in the saturated region while  $M_2$  and  $M_3$  are operated in the triode region with the control voltage  $V_c$ . It is found that the third-order nonlinearity coefficient  $a_3$  of compound MOSFET is very low (which is shown in Figure 21.53(b) and is marked with a rectangular dashed frame), that is

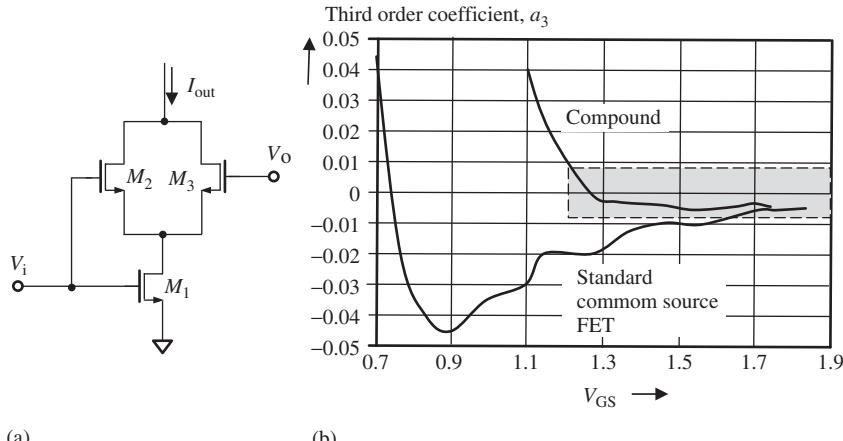
$$|a_3| < 0.01, \quad (21.23)$$

when

$$V_{GS} > 1.2 \text{ V}. \quad (21.24)$$

On the contrary, the third-order nonlinearity coefficient  $a_3$  of a standard common-source MOSFET has a deep negative notch around  $V_{GS} = 0.9 \text{ V}$ .

This is a very prospective work indeed. A linear PA built by a compound FET is undoubtedly better than one built by the standard common-source FET. Of course, there



**Figure 21.53.** Comparison of the measured third-order coefficients between the standard common-source FET and the compound FET. (a) A compound MOSFET stage. (b) The measured third-order coefficients of the standard common-source FET and the compound FET.

is expected to be further work so that condition (21.24) of the third-order nonlinearity coefficient of the compound FET stage can be updated to when

$$V_{GS} > 0.8 \text{ V}, \quad (21.25)$$

or so.

## REFERENCES

- Dawson JL, Lee TH. Cartesian feedback for RF power amplifier linearization. Proceeding of the 2004 American Control Conference; Boston (MA); 2004 June 90 -July 2. 2004. pp. 361–366.
- Khannur PB. A CMOS power amplifier with power control and T/R switch for 2.45-GHz BluetoothWISM band applications. IEEE Radio Frequency Integrated Circuits Symposium, Philadelphia; 2003. pp. 145–148.
- Tanaka S, Behbahani E, Abidi AA. A linearization technique for CMOS RF power amplifier. 4–93081 3-76-X 1997, Symposium on VLSI Circuits Digest of Technical Papers, Kyoto, Japan. 1997. pp. 93–94.
- Yamauchi K, Iyama Y, Yamaguchi M, Ikeda Y, Urasaki S, Takagi T. X-band MMIC power amplifier with an on-chip temperature-compensation circuit. IEEE Trans Microw Theory Tech 2001;49(12):2501–2506.

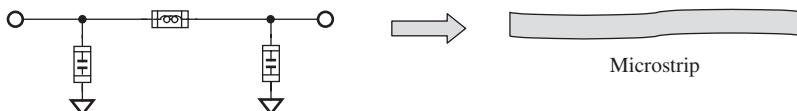
## FURTHER READING

- Ding L, Zhou GT. Effects of even-order nonlinear terms on power amplifier modeling and predistortion linearization. IEEE Trans Veh Technol 2004;53(1):156–162.
- Fager C, Pedro JC, Carvalho NB, Zirath H, Fortes F, Rosário MJ. A comprehensive analysis of IMD behavior in RF CMOS power amplifiers. IEEE J Solid-State Circ 2004;39(1):24–34.
- Gupta R, Ballweber BM, Allstot DJ. Design and optimization of CMOS RF power amplifiers. IEEE J Solid-State Circ 2001;36(2):166–175.
- Isaksson M, Wisell D, Rönnow D. A comparative analysis of behavioral models for RF power amplifiers. IEEE Trans Microw Theory Tech 2006;54(1):348–359.
- Kang J, Yu D, Yang Y, Kim B. Highly linear 0.18-m CMOS power amplifier with deep n-well structure. IEEE J Solid-State Circ 2006;41(5): 1073–1080.
- Lee S-Y, Lee Y-S, Hong S-H, Choi H-S, Jeong Y-H. An adaptive predistortion RF power amplifier with a spectrum monitor for multicarrier WCDMA applications. IEEE Trans Microw Theory Tech 2005;53(2):786–793.
- Lépine F, Ådahl A, Zirath H. L-band LDMOS power amplifiers based on an inverse class-F architecture. IEEE Trans Microw Theory Tech 2005;53(6):2007–2012.
- Mertens KLR, Steyaert MSJ. A 700-MHz 1-W fully differential CMOS class-E power amplifier. IEEE J Solid-State Circ 2002;37(2):137–141.
- Minowa M, Onoda M, Fukuda E, Daido Y. Backoff improvement of an 800 MHz GaAs FET amplifier for a QPSK transmitter using adaptive nonlinear distortion canceller. In: Presented at Vehicular Technology Conference, Orlando, FL; 1990. pp. 542–546.
- Nojima T, Konno T. Cuber predistortion linearizer for relay equipment in 800 MHz band land mobile telephone system. IEEE Trans Veh Technol 1985;VT-34(4):169–177.
- Sahu B, Rincón-Mora GA. A high-efficiency linear RF power amplifier with a power-tracking dynamically adaptive buck-boost supply. IEEE Trans Microw Theory Tech 2004;52(1): 112–120.
- Seidel H. A microwave feedforward experiment. Bell Syst Tech J 1971;50:2879–2916.

- Sowlati T, Leenaerts DMW. A 2.4-GHz 0.18-m CMOS self-biased cascode power amplifier. *IEEE J Solid-State Circ* 2003;38(8):1318–1324.
- Woo W, Miller MD, Kenney JS. A hybrid digital/RF envelope predistortion linearization system for power amplifiers. *IEEE Trans Microw Theory Tech* 2005;53(1):229–237.
- Zhao Y, Metzger AG, Zampardi PJ, Iwamoto M, Asbeck PM. Linearity improvement of HBT-based doherty power amplifiers based on a simple analytical model. *IEEE Trans Microw Theory Tech* 2006;54(12):4479–4488.

## EXERCISES

1. What are the main goals in a PA design?
2. What are the main means to enhance the PAE in the PA design?
3. Specify the power range of a PA with different class.
4. In PA design by means of tuning-on-the-bench, the device is easily burnt before the output impedance of the device is matched. How can this be avoided? Also, the device is easily burnt when the DC power supply is ON or OFF; how can this be avoided?
5. In the development of a balun, it is found that a discrete (chip) parts with special  $\pi$  configuration can be replaced by a microstrip line (Fig. 21.P.1). Compare their advantages and disadvantages.



**Figure 21.P.1.** The discrete parts with special “ $\pi$ ” configuration can be replaced by a microstrip line.

6. In the PA output control, what is the main improvement from the control by an electronic switch directly to the control by a programmable attenuator?
7. Comment on the design of double-line balun.
8. Comment on the design of toroidal RF transformer balun.
9. What is the main technology applied to a transistor-level linearized PA?
10. What is the main goal to be focused on in a PA design?
11. What is the main difficulty in a linear PA design?

## ANSWERS

1. The main goals in a PA design are
  - (a) output power,
  - (b) PAE,
  - (c) harmonics,
  - (d) temperature compensation, and
  - (e) output power control.

2. The main means to enhance the PAE in the PA design is the adjustment of the device's DC operating state. In the history of PA development, the device's DC operating state has been changed from class A to class F. The PAE has been enhanced from about 30% when the PA is operated with classes D, E, and F to more than 90% when the PA is operated with class A.
3. Probably, for the range of output power less than 1 W, PA is designed with class A; For the range of output power between 1 and 10 W, PA is designed with class B or C, or D.  
For the range of output power 10 W and above, PA is designed with class E or F.
4. In the PA design by means of tuning on the bench, the device is easily burnt before the output impedance of the device is matched. In order to avoid the burning of the device, the best way is as follows:
  - (a) At the beginning, tune the input and output tuner by setting up the input power with a tiny amount;
  - (b) After tuning is well done, increase the input power with a small step and tune the input and output tuner;
  - (c) Repeat the procedure (b) until the input power is increased up to the specified value.
 Also, the device is easily burned when the DC power supply is ON or OFF. In order to avoid the burning of the device,
  - (a) when DC power supply is going to be switched ON, the bias voltage must be switched ON first and then the  $V_{dd}$  or  $V_{cc}$  second. On the contrary,
  - (b) when DC power supply is going to be switched OFF, the  $V_{dd}$  or  $V_{cc}$  must be switched OFF first and then the bias voltage.
5. In the development of balun, it is found that a discrete (chip) part balun with "Π" configuration can be replaced by a microstrip line balun. The advantages and disadvantages are given in the table below:

	Microstrip Line Balun	Discrete (Chip) Part Balun with "π" Configuration
Part count	None	3
Reliability	High	Low
Cost	None	For three parts

6. In the PA output control, the design technology is developed from the control by electronic switch directly to the control by programmable attenuator. In the technology of control by electronic switch directly, the input impedance of the PA will be changed as the switches are turned ON or OFF. In other words, the input impedance of the PA will be always in an impedance-unmatched state, so that the power gain and other performance parameters become unstable. In the technique of control by a programmable attenuator, the input impedance of PA will be always in an impedance-matched state, so that the power gain and other performance parameters are kept stable.
7. Double-line balun has the advantages of small size, low insertion loss, allowance of high current, and high reliability. It is a balun suitable to be applied in PA design.

The only drawback is the somewhat high cost due to the ceramic alumina PCB to be adopted.

8. Toroidal RF transformer balun has the advantages of small size, extremely low insertion loss, allowance of high current, and high reliability. It is a balun suitable to be applied in PA design. The only drawback is the somewhat high cost due to the ceramic alumina PCB to be adapted.
9. The main technique in the transistor-level linearized PA design is the application of a compound MOSFET stage, by which the third-order nonlinear coefficient can be reduced down to a negligible level.
10. The main goal to be focused on in the PA design is how to enhance the PAE of the PA.
11. The main difficulty in a linear PA design is how to overcome the conflict between linearity and the PAE.



---

# INDEX

---

- Able–Baker spurs, 457  
AC bypass capacitor, 290  
AC/RF voltages, 328–329, 331–332  
active mixers  
    Gilbert cell, 708–716  
    single-end single device, 706–708  
AC voltage, 401  
“advanced” RFIC (RF integrated circuit) design, 4  
amplification power gain, 432–433  
amplifier, power gain of, 431  
applied regions of specific topology, 81  
arm, defined, 67–68  
artificial electromagnetic interference sources, 15  
attenuation of interference, 402, 406  
avalanche noise, 490
- balun, 503  
    definition, 547  
    double microstrip line, 808–817  
LC, 548–549  
    design, 572  
    features of, 571–572  
    performance of simple, 572–576, 588–602  
    practical, 576–579  
main features of an ideal, 548  
microstrip line  
    ring, 580–582  
    split ring, 582–583  
mixing type of  
    chip inductors and chip capacitors, 585–586  
    microstrip line and chip capacitor, 583–585  
RF, 548  
transformer  
    built by two stacked transformers, 586–588  
    features of, 549–550  
    impedance matching for differential pair by means,  
        568–571  
    in RF circuit design with discrete parts, 550  
    in RFIC design, 550–551  
    for simulation, 551–555  
    between single-ended and differential pair in respect to,  
        555–568
- bandwidth of zero capacitor, 293–295  
baseband blocks, 8  
bipolar CE–CB cascode amplifier, 662–666  
bipolar transistor, 182  
    with CB (common base) configuration  
        input and output impedances, 208–214  
        KCL at node, 210, 212–213  
        open-circuit voltage gain, 204–206  
        short-circuit current gain and frequency response, 206–208  
    with CC (common collector) configuration  
        input and output impedances, 218–221  
        KCL at node, 219  
        open-circuit voltage gain, 214–217  
        short-circuit current gain and frequency response, 217–218  
    with CE (common emitter) configuration, 190–204  
        emitter degeneration, 200–204  
        Miller effect, 197–200  
        open-circuit voltage gain, 190–194  
        primary input and output impedances, 196–197  
        shorted-circuit current gain, 194–196  
        transition frequency, 195  
DC characteristics of, 187  
hybrid-p model, 188  
similarity with MOSFET, 225–235  
small-signal model of a, 187–189  
blocks in transmitter, order of, 8  
BPF (band-pass filter), 386–387  
BPF helical filters, 696  
BPSK (Bi-phase shift keying), 35  
branch, defined, 67–68  
burst noise (popcorn noise), 490  
bypass capacitor, 284–286
- cable balun, 603–604  
*Cal\_En* (calibration enable) input, 517  
calibrated cable, 333  
capacitor in an IC chip, 370–371  
carrier-to-noise ratio (CNR), 430  
cascaded equations  
    of noise figure, 446–448  
    of power gain, 439–441  
CE–CB cascade amplifier, 212

- CG configuration of RF input, 514  
 channel-length modulation, 223  
 chopping block, 515  
 chopping frequency, 514  
 chopping mixer, 511–516  
 circuit block or a system, 281–282  
 circuit testing, 6  
 circulator, 276–277  
 Clapp oscillator, 753–755  
 CMOS (complementary metal–oxide–semiconductor  
     Oxide–Semiconductor) processing, calculation of quarter wavelength in, 420–422  
 CMRR (common mode rejection ratio) of differential pair, 532  
     expression, 533–538  
     in a pseudodifferential pair, 539–541  
     in a single-ended block, 539  
 CNR (carrier-to-noise ratio), 5  
 Colpitts oscillator, 753  
 combiner, 547  
 Combo or Jumbo design, 4  
 common mode voltage gain, 533  
 communication system, RF circuitry *vs* digital circuitry  
     current drain's magnitude, 7  
     demarcation line, 7–9  
     impedance, 6–7  
 controversy between digital and RF circuit designers, 3  
 corner analysis, 378  
 crystal oscillator, 758–759  
 CS–CG cascade LNA (low-noise amplifier), 41  
 current coupling, 287  
 current drains, 7  
 1-dB compression point, 471–472
- $\text{dB}_W$ , 3  
 DC blocking, 85  
     using capacitor, 508–511  
 DC blocking capacitor, 290  
 DC feeding, 85  
 DC offset cancellation  
     by calibration, 516–517  
     difficulties in measuring, 517  
     using chopping mixer, 511–516  
 DC power supply, 281, 284–285  
     connecting the test PCB with, 288–289  
     and current drain in RF design, 480–481  
     method of reducing forward current coupling, 336–338,  
     342–343  
     and RFIC design, 413–414  
     and zero capacitors, 296  
 DC short-circuit, 85  
 demarcation circle of return loss, 133  
 demarcation line in a communication system, 7–9  
 demodulator, 7–8  
 $6\sigma$  design  
     approach, 383–386  
     for a circuit block, 382–383  
     for circuit system, 383  
     implication of, 379–383  
     index  $C_p$  and  $C_{pk}$  applied in, 398  
     and yield return, 378–382  
 design tolerance, 380–381  
 differential pair, 696  
     CMRR (common mode rejection ratio), 532  
         enhancement of, 541–542  
         expression, 533–538  
         in a pseudodifferential pair, 539–541  
         in a single-ended block, 539  
 difference between single-ended and, 501–503  
 fundamentals of  
     DC offset in an ideal, 531–532  
     MOSFET, 528–529  
     small signal approximation, 527–528, 530–531  
     topology and definition of, 521–524  
     transfer characteristics, 524–529  
     nonlinearity, 505–506  
 differential VCO, 778–783  
 digital/analog circuit design, 8  
 double microstrip line balun, 808–817  
 downward impedance transformer, 46  
 dual-conversion portable radio, 696
- electromagnetic radiation, 15  
 electronic circuitry, revolution in, 405–406  
 EMC (electromagnetic compatibility), 15–16  
 emitter degeneration and CE configuration, 200–204  
 emitter follower, 38  
 emitter resistor, 209  
 energy generator portion, 764  
 environmental interference sources, 15  
 equipotentiality on grounded surface  
     approach to achieve  
     using half-wavelength runners, 330–331  
     using quarter-wavelength runners, 331–333  
     using zero capacitors, 328–330  
     forward current magnetic coupling and, 342  
     gold-plated surface of the box, 342  
     and imperfect or inappropriate RF/AC grounding, 335  
     improved grounding path, 343–344  
     indifferent assumptions and ignorances, 335–336  
     multicloset type of system assembly for, 341–344  
     of a PCB (printed circuit board), 326–327  
     in AC/RF grounding for a large PCB, 328  
     problems with large PCB, 327–328  
     reducing forward and return current coupling, method of,  
     336–338  
     in IC die, 338–340  
     in RF blocks, 340–341  
     of a regular RF cable, 325–326  
     return current magnetic coupling and, 342  
     testing for, 333–335  
     extended cable, 333  
     external interference sources, 15
- flicker noise, 491  
 forward current coupling on a PCB, 336–337  
 four-point experimental method of noise figure testing, 445  
 four-pole helical filter, 696  
 free space in layout of a PCB, 371–373

- Gilbert cell, 151–153, 514–515, 697  
bipolar, 712–714  
impedances of, 153–155  
MOSFET, 708–711, 715–716  
grounding, in RF circuit design  
compulsory/enforced/coercing, 305–306  
implications, 281–283  
problems with, 283–284  
    blind selection of bypass capacitor, 285–286  
    connections between grounded surfaces, 289  
    imperfect RF/AC grounding, 286–288  
    improper connection about the ground points, 288–290  
    underestimation or ignoring of bypass capacitor, 284  
    unequipotentiality on the grounded surface, 287–288  
quarter wavelength microstrip line  
    magic open-circuited, 305–307  
    rationale, 304–305  
    runner between parts in a circuitry, 300–304  
    with a specific characteristic impedance, 307  
    testing for, 307–309  
reference ground point, 281–283  
'zero' capacitor applied for RF/AC  
    bandwidth, 293–295  
    chip capacitor as, 296–298  
    combined effect of multi-zero capacitors, 295–296  
    in RFIC design, 298–300  
    selection, 290–292  
    zero capacitor, defined, 290  
group #1, band #2 of UWB (ultrawide band) system, 671–684  
    gain and bandwidth, 673–680  
    noise figure, 680–682  
    nonlinearity, 682–684  
    raw device testing, 671–673
- half IF spurs, 457  
Hartley oscillator, 751–753  
Haus' theory, 443  
helical filter, 696  
HFA3783, 517  
high digital data rates, 9–10  
higher impedance in digital circuitry, 7  
high speed digital circuit design, 9–10
- IC chips, chip capacitor, chip inductor, and chip resistor of, 309–319  
IC packaging, 615–621  
ideal impedance-matched state without reflection, 66  
impedance conjugate matching, 33–42  
    condition of, 34  
    maximum power at the load, 36  
    maximum power transport, 33–35  
power delivered to the load, 37  
significance of, 35  
voltage and power delivered from source to the real part  $R_L$   
    of load, 33
- impedance matching, 4–5, 10, 438  
    of an  $L_S - C_P$  upward and downward resistance transformer, 47–49  
approximate demarcation criterion for, 67
- arms and branches, 67–68  
for drain of the MOSFET device, 177–178  
goals of, 40  
for LO, RF, and IF ports ignoring the bandwidth, 155–158  
necessity of, 40–42  
pumped-up voltage and, 44–45  
relationship between return loss and, 63–68  
stepped-down voltage and, 43–44  
between two basic power transport units, 41  
unchanged voltage and, 43, 45  
voltage delivered from a source to a load through, 43, 46  
voltage pumped up by means of, 42–49  
    when impedance is less than 50 W, 85–93
- impedance matching condition  
    of test equipment, 6
- impedance matching network, 37–40  
    implementation of, 67–68  
    input and output of, 83  
LC, 39  
    of one part, 68–73  
parts in an, 93–94  
passive parts of, 61  
power delivered, 38  
sub-impedance matching loop, 38–40  
of three parts, 84–85  
of two parts, 74–84  
P type or T type impedance matching network, 84, 122–124
- impedance measurement  
    direction of, 263–265  
    low- and high-impedance, 275–276  
    by means of the standard calibration kit, 270–272  
    of parameters, 265–266  
scalar and vector voltage  
    using oscilloscope, 260–261  
    using vector voltmeter, 262–263  
in series and in parallel, 277–278  
 $S$  parameters, 265–270  
using circulator, 276–277  
    using Smith Chart, 272–275
- impedance of chip inductor, 297  
impedance of the load, 16  
impedances of the source, 16  
incident voltage, 433  
inductors in IC die, 369–370  
infinite inductor, 297  
intercept point, 5  
intercept point of the first-order signal, 462  
intermodulation of a RF block, 461–462  
internal interference sources, 15  
isolation in RFIC, definition, 402
- KCL (Kirchhoff's current law), 4  
KVL (Kirchhoff's voltage law), 4
- layout between individual block and system, differences, 349–350
- LC balun, 548–549  
    design, 572  
    features of, 571–572

- LC balun (*Continued*)  
 performance of simple, 572–576, 588–602  
 practical, 576–579
- LC impedance matching networks, 39
- LC passive filter, 4
- LNA (low-noise amplifier), 4  
 with AGC (automatic gain control)  
 control, 684–686  
 dynamic range, 688  
 example, 689–690  
 traditional, 686–687
- for appropriate power level, 628
- design of good  
 1-dB compression point, 654–655  
 gain circles and noise figure circles, 643–645, 648–649  
 impedance matching, 641–643, 646–648  
 nonlinearity, 653–654  
 optimum of source voltage reflection coefficient, 639–641  
 procedures, 655–656  
 stability, 649–653
- group #1, band #2 of UWB (ultrawide band) system, 671–684
- for 800/900 MHz radio, 661–662
- single-ended cascode  
 bipolar CE–CB cascode amplifier, 662–666  
 MOSFET CS–CG cascode amplifier, 666–669  
 need for, 669–671
- single-ended device  
 size, 629–632  
 step and testing, 632–639
- for UHF band, 658–660  
 for VHF band, 656–658
- load-pulling test of VCO, 776–778
- LO (local oscillator) frequency, 458
- low digital data rate, 8
- lower impedance in an RF circuitry, 7
- low-noise amplifier (LNA), 285
- microstrip Lange coupler, 603
- microstrip line, impedance of, 300–309
- microstrip line balun  
 ring, 580–582  
 split ring, 582–583
- Miller effect, 183–187, 197–200, 237  
 defined, 186
- impedance of the equivalent element at the input node, 184  
 impedance of the equivalent element at the output node, 185
- Miller impedance, 186
- Miller's theorem, 184
- mixers  
 active, 695  
 Gilbert cell, 708–716  
 single-end single device, 706–708
- design schemes  
 impedance matching, 717–718  
 multi-tanh technique, 719–722  
 RF input types, 722–723  
 topology with current bleeding, 718–719
- passive, 695
- double-balanced quad-diode, 699–702  
 double-balanced resistive, 702–706  
 simplest, 698–699
- trigonometric and hyperbolic functions, 723–725
- modulation technology, 35
- Monte Carlo analysis  
 BPF (band-pass filter), 386–387  
 sensitivity on parameters of performance, 392  
 simulation with, 387–391
- MOSFET, 41
- MOSFET (metal-oxide-semiconductor field-effect transistor), 260, 264–265, 306, 371–372, 513–514, 634  
 with CD (common drain) configuration  
 input and output impedances, 251–252  
 open-circuit voltage gain, 250  
 short-circuit current gain and frequency response, 250–251  
 with CG (common gate) configuration  
 input and output impedances, 247–248  
 open-circuit voltage gain, 244–245  
 short-circuit current gain and frequency response, 245–247  
 with CS (common source) configuration  
 CS configuration and source degeneration, 240–244  
 input and output impedances, 239–240  
 Miller effect, 237  
 open-circuit voltage gain, 235–237  
 short-circuit current gain, 237–239  
 hybrid-p model, 224  
 similarity with bipolar transistor, 225–235  
 small-signal model of a, 221–225
- MOSFET (metal–oxide–semiconductor field-effect transistor), 182
- input and output impedance testing, 183  
 open-circuit voltage gain, 183  
 short-circuit current drain, 183  
 $m$ th-order intercept point, 462  
 $m$ th-order intermodulation rejection, 463  
 $m$ th-order output intercept point, 462
- multicloset type of system assembly, 341–344  
 multilayered IC chip, 344–346
- n-channel MOSFET, 406
- n-channel MOSFETs, 513
- n-channel transistor, 224
- NF (noise figure), 442–443, 484
- noise power in RF circuit design, 430  
 cascaded equation for noise figure of a system, 446–448  
 20-dB quieting sensitivity of a receiver, 452–453  
 12-dB SINAD sensitivity of receiver, 452  
 noise figure testing, 444  
 in a noisy two-port RF block, 443–444  
 parameters of an RF block, 445–446  
 ratio of signal to noise (S/N), 443  
 sensitivity testing of a receiver, 448–453  
 significance of, 441–443  
 ultrahigh frequency (UHF) range, 444  
 noise sources, in a circuit block, 490–491  
 nonlinearity of device  
 cascaded equation for the intercept point of a system, 472–479

- 1-dB compression point and IP<sub>3</sub>, 471–472  
and distortion of signal, 430–431, 479–480  
with input voltage signal of two frequencies, 455–461  
IP (intercept point) and IMR (intermodulation rejection), 461–471  
    with one frequency signal, 453–455  
nonlinearity of the device, 15–16  
normalized resistances, 65–66  
N-well guard ring, 409–410
- Ohm's law, 209, 219, 401  
one part impedance matching network  
    with an inductor in series, 69  
    with a capacitor in series, 69–70  
one part in parallel, impedance matching network inserted in, 70–73  
one part in series, impedance matching network inserted in, 68–70  
Op-Amp (operating amplifier), 6  
open-circuit voltage gain of the device, 184  
oscilloscope, 260–261
- packaging of IC products, 615–621  
PAE (power added efficiency), 789  
PA (power amplifier)  
    classification  
        class A, 790  
        class B, 790–791  
        class C, 791  
        class D, 791–792  
        class E, 792–793  
        class F (third harmonic frequency), 793–794  
        class S, 794  
    linear, 824–828  
    with output power control, 823–824  
    push–pull PA design, 799–821  
    single-ended  
        PA IC design, 798–799  
        simulation, 796–798  
        tuning-on-the-bench technique and, 795  
    with temperature compensation, 822–823  
part count, 482  
passive mixers  
    double-balanced quad-diode, 699–702  
    double-balanced resistive, 702–706  
    simplest, 698–699  
path of interference in a RF module, 403  
P+ buried pad, 406  
PCB (printed circuit board)  
    attenuation of signal due to imperfect, 351  
    ceramic, 350–351  
    dielectric constant of, 351  
    electromagnetic parameters, 351  
    elimination of coplanar MSL, 372  
    equipotentiality on grounded surface of, 326–327  
    in AC/RF grounding for a large PCB, 328  
    execution of, 373  
    problems with large PCB, 327–328  
    reduction of current coupling, 336–338
- free space in the layout of, 371–373  
metallic layers of, 352  
modeling for vias  
    multivias, 359  
    single, 355–358  
multilayered, 344–346  
operating frequency of the circuit, 351  
parts  
    capacitor, 370–371  
    device, 369  
    inductor, 369–370  
    resistor, 370  
P+ guard ring, 353–354  
plastic, 350  
preferred layout of, 352–355  
reduction of additional radiation in, 372  
reduction of coplanar capacitance in, 372  
runners  
    connected with the load in parallel, 363  
    connected with the load in series, 360–362  
corner of, 365  
    for a differential pair, 368–369  
    multiple/curves in parallel, 363–364  
    in parallel, 367  
    placement of, 367  
    smoothness of, 365–366  
    spacing between adjacent ground surface and, 367–368  
    styles of, 363–369  
size, 351–352  
types, 350–351  
p-channel MOSFETs, 513  
perfect isolation, 403  
P+ guard ring, 409–411  
pinch-off region, 223  
PLL (phase lock loop)  
    implication of VCO, 759–760  
    phase noise from VCO, 764–769  
    transfer function, 760–763  
    white noise from the input of, 763–764  
power delivered from a source to a load, 23–33  
    additional distortion, 28–31  
    additional factor, 25  
    additional interference, 31–33  
    bouncing back and forth and arriving at load, 26  
    desired signal power arriving on the load, 31  
    power instability, 26–27  
    power loss, 27–28  
    power reflection coefficient, 24  
    resultant power, 24  
    signal-to-interference ratio, 31–32  
    voltage reflection coefficients, 24  
power gain in RF circuit design, 429–430  
    cascaded equation for, 439–441  
    concept of reflection, 431–434  
    due to amplification, 432–433  
    transducer, 434–436  
    in a unilateral and impedance-matched case, 438  
    in unilateral case, 437–438  
    and voltage gain, 439

- power measurement in an RF test laboratory, 49–51  
 power (mW) calculated from voltage (mV), 58  
 power reflection coefficient, 66, 485  
 printed inductors, 770–773  
 process capability, 381  
 product designing  
    $6\sigma$  design  
 approach, 383–386  
 for a circuit block, 382–383  
 for circuit system, 383  
 implication of, 379–383  
 index  $C_p$  and  $C_{pb}$  applied in, 398  
 and yield return, 378–382  
 fundamentals of random process, 392–398  
 Monte Carlo analysis  
   BPF (band-pass filter), 386–387  
   sensitivity on parameters of performance, 392  
   simultion with, 387–391  
 R&D stage, 377–378  
 simulation stage, 378  
 tolerance analysis, 378  
 prohibited regions of specific topology, 81  
 pseudodifferential pair, 540  
 PSK (phase-shift keying) technology, 35  
 pumped-up voltage, 44–45  
 push–pull PA design, 799–821  
   block diagram, 799–800  
   double microstrip line balun, 808–817  
   impedance matching, 800–804  
   minimizing insertion loss, 814  
   minimizing size, 815  
   size of PA block, 804–808  
   specifications, 799  
   toroidal RF transformer balun, 817–821
- QAM system, 35  
 QPSK (Quad-phase shift keying), 35  
 QSOP (quad small outline package), 616  
 quality factor, 47
- Radio Frequency Integrated Circuit (RFIC), 697  
 developing, 405  
 grounding rules  
   of circuit branches or blocks in parallel, 412–413  
   for DC power supply point, 413–414  
 interference in circuitry, 401–402  
   in an IC die, 403  
   experiments, 406–407  
   main path of, 403  
   P+ guard ring, 409–411  
   solution for interference from the sky, 411–412  
   trenching of an RF block, 408  
 isolation of circuitry, 402–403  
 metallic shielding for RF module, 403–405  
 problems affecting  
   bonding wire, 419  
   calculation of quarter wavelength in CMOS, 420–422  
   complication of via modeling, 419–420  
   low- $Q$  inductor, 414–419
- zero capacitor, 419  
 SOC (system-on-a-chip) in, 612  
 RC phase-shift oscillator, 755–757  
 real-life RF circuit block designs, 61  
 receive mode current drain, 481  
 reflected power in %, 485  
 reflected powers, 433  
 reflection power gain, 431–434  
   of a two-port block, 437  
 resistor in an IC chip, 370  
 resistor  $R_S$  in parallel, insertion of, 73  
 resistor  $R_S$  in series, insertion of, 73  
 resonant oscillator, 757  
 return loss in dB, 485  
 return loss RL testing, 62–63  
 RF back end, 8  
 RF baluns, 548  
 RF blocks, 5, 8  
 RF circuit design  
   conversion between Watts, Volts, and  $\text{dB}_m$  in a system, 485–487  
   noise power, 430  
   cascaded equation for noise figure of a system, 446–448  
   20-dB quieting sensitivity of a receiver, 452–453  
   12-dB SINAD sensitivity of receiver, 452  
   noise figure testing, 444  
   in a noisy two-port RF block, 443–444  
   parameters of an RF block, 445–446  
   sensitivity testing of a receiver, 448–453  
   significance of, 441–443  
   ultrahigh frequency (UHF) range, 444  
 nonlinearity of device  
   cascaded equation for the intercept point of a system, 472–479  
   1-dB compression point and  $\text{IP}_3$ , 471–472  
   and distortion of signal, 430–431, 479–480  
   with input voltage signal of two frequencies, 455–461  
   IP (intercept point) and IMR (intermodulation rejection), 461–471  
   with one frequency signal, 453–455  
 power gain, 429–430  
 cascaded equation for, 439–441  
 concept of reflection, 431–434  
 transducer, 434–436  
   in a unilateral and impedance-matched case, 438  
   in unilateral case, 437–438  
   and voltage gain, 439  
 system analysis, example, 482–484  
 RF circuit designers, 3  
 circuit testing and test equipment, 6  
 impedance matching and, 4–5  
   parameters used by, 5  
 RF circuit testing, 259  
 RF frequency, 402  
 RF front end, 8  
 RFIC chip, 4  
 RF signal, spectra of, 514–515  
 ring microstrip line balun, 580–582  
 RL (return loss), 51

- RP integrated circuit, 615  
runners for PCB (printed circuit board)  
  connected with the load in parallel, 363  
  connected with the load in series, 360–362  
corner of, 365  
for a differential pair, 368–369  
multiple/curves in parallel, 363–364  
in parallel, 367  
placement of, 367  
smoothness of, 365–366  
spacing between adjacent ground surface and, 367–368  
styles of, 363–369
- SAR (successive approximation register) state control, 517
- SAW (surface acoustic wave) filter, 696
- second-order intercept point, 464–467
- second-order intermodulation, 457
- self-interference, 15–16
- sensitivity testing of a receiver, 448–453
- shielding of RF module, 403–405
- short-channel effect, 224
- shot noise, 490
- simulation for VCO, 773–776
- single-ended stages  
  nonlinearity in, 503–505  
  vs differential pair, 501–503
- single-ended VCO with Clapp configuration, 769
- single-stage amplifier in an RF circuit design, 181
- sinusoidal voltage, 18
- SMA connector, 286–287
- small signal approximation  
  of a bipolar pair, 527–528  
  of a MOSFET differential pair, 530–531
- Smith Chart, 62–63  
accuracy of, 272–275
- Cartesian plot of return loss, 133–134
- center of, 75, 78–80
- critical circle of return loss, 131–136
- evolution of bandwidth at the LO port, 164–172
- fundamentals of, 94–99
- impedance in four regions on, 74
- LNA (low-noise amplifier), 634–639, 642, 645–647, 649–651, 656, 671–677, 680
- original impedance appearing on, wideband case, 136–137
- PA (power amplifier), 801, 806
- quarter wavelength microstrip line, 307–310
- return loss circles on, 63–66
- variation of impedance in, 69–70
- SOC (system-on-a-chip)  
basic concept, 611  
communication system, 613  
development of, 613–615  
in RFIC design, 612
- source degeneration and CS configuration, 240–244
- S parameters, 434  
  measurement of, 265–270
- splitter, 547
- SRF (self-resonant frequency) of a chip inductor, 297
- SRF (self-resonant frequency) of capacitor, 286, 292–293, 513
- SS-DS Connection Mode, 586
- SSOP (shrink small outline package), 616
- standby current drain, 481
- “ $*-\Delta$ ” (star–delta) conversion, 124
- stepped-down voltage, 43–44
- system design of a communication system, 5
- thermal noise, 490–491
- third-order intercept point, 467–471
- third-order intermodulation, 456
- three-part impedance matching network, 84–85  
  topology limitation of, 114–121
- three-point type oscillator, 749–751
- three-pole helical filter, 696
- TITO (tuned input and tuned output) oscillator, 757
- TL (transmission loss), 51
- tolerance analysis, 378
- topologies for an impedance matching network, 81–83
- toroidal RF transformer balun, 817–821
- transducer power gain, 434–436
- transformer balun  
  built by two stacked transformers, 586–588  
  features of, 549–550  
  ideal  
    for simulation, 551–555  
    between single-ended and differential pair in respect to, 555–568
- impedance matching for differential pair by means, 568–571
- in RF circuit design with discrete parts, 550
- in RFIC design, 550–551
- transistor configuration, comparison of, 252–255
- transmission line theory, 17, 300
- transmission loss in dB, 485
- transmit mode current drain, 481
- trenching of an RF block, 408
- p-T transformation, 731, 739
- tunable filter design  
  circuit description, 738–739  
  in a communication system  
    expected constant bandwidth, 732  
    variation of bandwidth, 732–733  
  coupling between tank circuits, 733–738  
  performance, 743–746  
    second coupling, 739–743
- two-part impedance matching network, 74–84  
  formula for, 99–113  
  in wideband cases, 145–151
- two-port block, 488–489
- unchanged voltage, 43, 45
- unilateral case, operating power gain, 436–438
- upward impedance transformer, 47
- UWB (ultra wide band) system  
  bandwidth allocation, 160–161  
  impedance matching using I or Q modulator, 161–174  
  wide bandwidth required in, 159–160
- UWB (ultrawide-band) system, 4

- varactor, 770  
VCO (voltage-controlled oscillator), 429  
vector voltmeter, 262–263  
calibration of, 268–269  
vias, modeling of  
blind, 355  
buried, 355  
multi, 359  
single, 355–358  
through, 355  
voltage delivered from a source to a load, 16–23  
average power across reactance, 17  
delayed time, 18  
distortion or jitter in digital circuit block, 20–23  
impedances, 16  
reactance, 16  
resultant voltage at the load, 18–20  
resultant voltage on resistance, 18  
voltage reflection coefficients, 17  
voltage gain, 439  
voltage measurement  
using oscilloscope, 260–261  
using vector voltmeter, 262–263  
voltage (mV) calculated from power ( $\text{dB}_m$ ), 58  
voltage reflection coefficient, 51, 485  
voltage reflection coefficients at source, 17–18  
voltage standing wave ratio, 485  
VSWR (voltage standing wave ratio), 51  
  
wide-band block, 61  
wide-band RF circuit design, 63  
  
wide bandwidth, methodology of impedance matching for  
with an inductor inserted in parallel, 141–143  
an IQ modulator in a UWB system, 151–174  
with a capacitor inserted in parallel, 143–145  
with a capacitor inserted in series, 139–141  
difference of return loss between narrow and, 131–136  
impedance variation due to the insertion of one part per arm  
or per branch, 136–145  
impedance variation due to the insertion of two parts9 per  
arm or per branch, 145–151  
input impedance matching network, 173  
output impedance matching network, 173  
passive, 174–179  
“rolling” or “squeezing” of impedance, 138–141, 143, 145,  
167  
at a terminal from a narrowband to a wideband response, 135  
topology of, 177–178  
variation of the impedance, 137–138  
Wilkinson coupler, 602  
wireless communication system, 5  
  
zero capacitor applied for RF/AC, 284  
bandwidth, 293–295  
chip capacitor as, 296–298  
combined effect of multi-zero capacitors, 295–296  
in RFIC design, 298–300  
selection, 290–292  
zero capacitor, defined, 290  
zero capacitor for an RFIC, 299  
zero capacitors, 3  
zero IF communication, 507–508