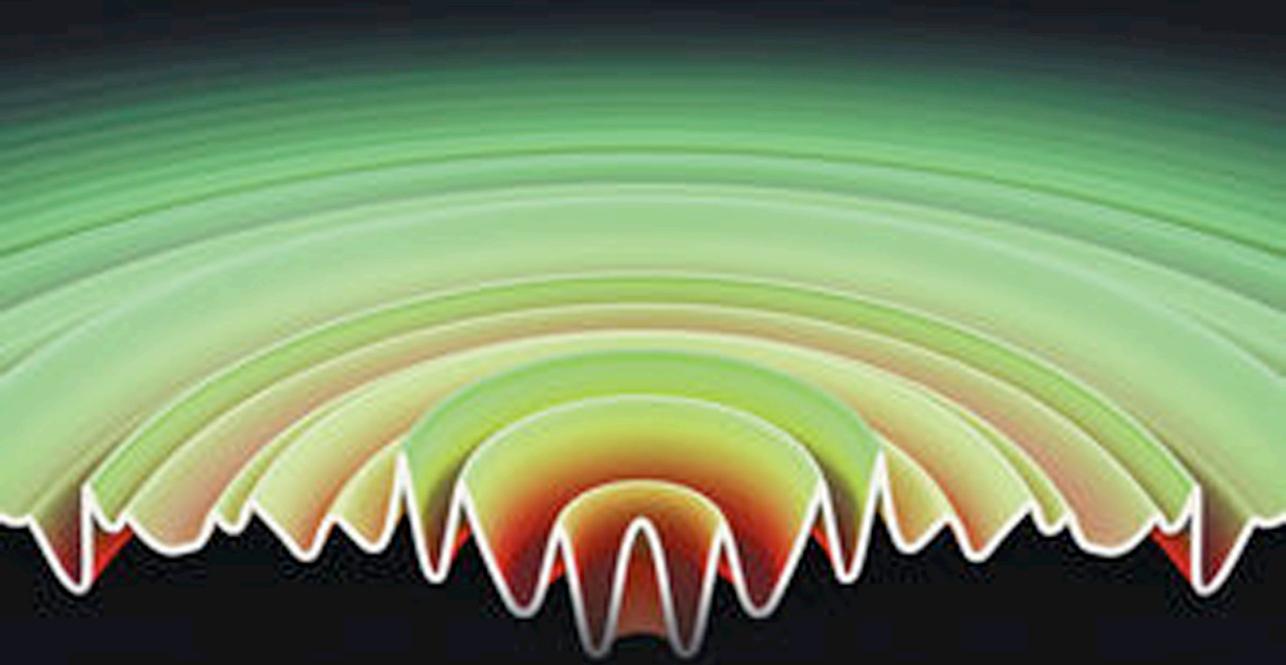


ANTENNA DESIGN FOR MOBILE DEVICES

SECOND EDITION



ZHIJUN ZHANG

IEEE PRESS

WILEY

ANTENNA DESIGN FOR MOBILE DEVICES

ANTENNA DESIGN FOR MOBILE DEVICES

Second Edition

Zhijun Zhang

Tsinghua University, China

WILEY

 **IEEE PRESS**

This edition first published 2017
© 2017 John Wiley & Sons Singapore Pte. Ltd

Registered Office
John Wiley & Sons Singapore Pte. Ltd, 1 Fusionopolis Walk, #07-01 Solaris South Tower, Singapore 138628.

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as expressly permitted by law, without either the prior written permission of the Publisher, or authorization through payment of the appropriate photocopy fee to the Copyright Clearance Center. Requests for permission should be addressed to the Publisher, John Wiley & Sons Singapore Pte. Ltd, 1 Fusionopolis Walk, #07-01 Solaris South Tower, Singapore 138628, tel: 65-66438000, fax: 65-66438008, email: enquiry@wiley.com.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The Publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

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. It is sold on the understanding that the publisher is not engaged in rendering professional services and neither the publisher nor the author shall be liable for damages arising herefrom. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication data applied for

ISBN: 9781119132325

A catalogue record for this book is available from the British Library.

Set in 10/12pt Times by SPi Global, Pondicherry, India

10 9 8 7 6 5 4 3 2 1

To my wife Sheng Li

Contents

About the Author	x
Preface	xi
Acknowledgments	xii
Abbreviations	xiii
1 Introduction	1
1.1 The Evolution of Mobile Antennas	2
1.2 How to Quantitatively Evaluate an Antenna	10
1.3 The Limits of Antenna Designs	12
1.4 The Trade-Offs in Antenna Designs	14
1.5 Mobile Communication and Band Allocations	16
1.6 Quickly Building a Simple Antenna—a Practical Example	18
References	27
2 Antenna Matching	28
2.1 The Smith Chart	29
2.2 Single-Band Matching	33
2.2.1 <i>Matching with Lumped Elements</i>	33
2.2.2 <i>Different Ways to Accomplish a Single-Band Matching</i>	36
2.2.3 <i>Matching with Both Transmission Line and Lumped Elements</i>	39
2.2.4 <i>Bandwidth Consideration</i>	42
2.3 Dual-Band Matching	50
2.4 Reconfigurable Matching	55
2.4.1 <i>Reconfigurable Matching—Varactor-Based</i>	55
2.4.2 <i>Reconfigurable Matching—Switch-Based</i>	60
References	63

3 External Antenna	65
3.1 Stubby Antennas	66
3.1.1 <i>Single-Band Helix Stubby Antenna</i>	67
3.1.2 <i>Multiband Helix Stubby Antenna</i>	86
3.1.3 <i>Ultra-Wideband Stubby Antenna</i>	109
3.2 Whip–Stubby (Retractable) Antenna	117
3.2.1 <i>Decoupled Whip–Stubby Antenna</i>	119
3.2.2 <i>Semi-Decoupled Whip–Stubby Antenna</i>	121
3.3 Meander Line Stubby Antenna	126
3.4 Effect of Ground Plane	131
References	136
4 Internal Antenna	138
4.1 Inverted-F Antenna	141
4.2 Planar IFA	146
4.2.1 <i>Single-Band PIFA</i>	146
4.2.2 <i>Multiband PIFA Antenna with Slits</i>	149
4.2.3 <i>Multiband PIFA with Separate Branches</i>	157
4.2.4 <i>Multiband PIFA with Parasitic Element</i>	158
4.2.5 <i>Manufacturing PIFA Antenna</i>	159
4.3 Folded Monopole Antenna	163
4.4 Loop Antenna	167
4.5 Ceramic Antenna	172
4.5.1 <i>Monopole-Type Ceramic Antenna</i>	173
4.5.2 <i>IFA-Type Ceramic Antenna</i>	176
4.5.3 <i>Loop-Type Ceramic Antenna</i>	177
4.6 Slot Antenna	179
4.7 Design a Hepta-Band Antenna with Multiple Radiators and Multiple Modes	185
4.8 Design a Reconfigurable Hepta-Band Antenna	191
4.9 MIMO Antennas	200
4.9.1 <i>Explaining Capacity Boost Effect Through the Antenna Point of View</i>	200
4.9.2 <i>Antenna Correlation and Antenna Isolation</i>	207
4.9.3 <i>Improve Isolation Between Antennas</i>	209
4.10 Antennas in Recently Released Phones	211
4.10.1 <i>Entry-Level Phone</i>	211
4.10.2 <i>Flagship Phone</i>	221
References	226
5 Antenna Measurement	229
5.1 Passive Antenna Measurement	229
5.1.1 <i>Measurement on a Vector Network Analyzer</i>	229
5.1.2 <i>Fixture</i>	234
5.1.3 <i>Passive Chamber Measurement</i>	246

5.2	Active Antenna (Over the Air) Measurement	253
5.2.1	<i>EIRP, ERP, and TRP</i>	253
5.2.2	<i>EIS and TIS</i>	256
5.2.3	<i>Sensitivity Degradation Due to Interference</i>	259
5.3	Antenna Measurements in the Production Line	262
5.4	Multiple Input and Multiple Output Antenna Test	271
	References	275
6	Regulations Related to Antenna Engineers	276
6.1	Specific Absorption Rate	276
6.1.1	<i>Definition and Measurement Method of SAR</i>	277
6.1.2	<i>SAR Limits in the United States and Europe</i>	283
6.1.3	<i>Controlling SAR</i>	285
6.1.4	<i>Updates on SAR Requirement</i>	294
6.2	Hearing Aid Compatibility	296
6.2.1	<i>HAC Measurement</i>	296
6.2.2	<i>HAC Specification in the United States</i>	299
6.2.3	<i>Updates on HAC Requirement</i>	303
6.3	Electromagnetic Compatibility	304
	References	305
	Appendix: User Manual for ZJ_Antenna_Matching Software	307
	Index	314

About the Author

Dr. Zhijun Zhang is currently a professor at the State Key Lab of Microwave and Communications, the Department of Electronic Engineering at Tsinghua University, China. He received his BS and MS from the University of Electronic Science and Technology of China, in 1992 and 1995, respectively, and his PhD from Tsinghua University, China, in 1999. From 1999, he was a Post-doctoral Fellow with the Department of Electrical Engineering, University of Utah, where he was appointed as a Research Assistant Professor in 2001. In May 2002, he became an Assistant Researcher with the University of Hawaii at Manoa, Honolulu. In November 2002, he joined Amphenol T&M Antennas, Vernon Hills, IL, as a Senior Staff Antenna Development Engineer and was then promoted to the position of Antenna Engineer Manager. In 2004, he joined Nokia Inc., San Diego, CA, as a Senior Antenna Design Engineer. In 2006, he joined Apple Inc., Cupertino, CA, as a Senior Antenna Design Engineer and was then promoted to the position of Principal Antenna Engineer. Since August 2007, Dr. Zhang has been with Tsinghua University. He has published over 100 peer-reviewed journal papers, and has also been awarded 9 US patents and 40+ Chinese patents.

Dr. Zhang is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE). He has served as the Associate Editor of the *IEEE Transactions on Antennas and Propagation* (2010–2014) and *IEEE Antennas and Wireless Propagation Letters* (2009–2015).

Preface

I started my career as an antenna engineer in the mobile phone industry during the early 2000s. Despite over 10 years of researching electromagnetic topics, I felt that I lacked a lot of important aspects of a good antenna engineer, and there was no book available to provide all the necessary information. It took me several years to collect bits and pieces by learning from different sources. Since then, I have worked on both sides of the antenna business, as a vendor when I was with Amphenol Inc.; and then as a client when I was with Nokia Inc. and Apple Inc. My job also changed back and forth between antenna engineer and engineering manager. The experience gave me an opportunity to view antenna design from various angles.

This book presents most of the issues that directly concern mobile antenna design, what I have learned and many ideas I accumulated during the years. I hope the book can help students who want to pursue a career as an antenna engineer in the wireless communication industry. Another goal of the book is to provide a shortcut for electronic engineers who do not know a lot about antennas but want to design some simple antennas without the hassle of learning from various sources. For antenna engineers already working in the field, the challenge facing them is how to squeeze an antenna into a smaller enclosure and still meet all the bandwidth, multiband, and efficiency specifications. The book provides various advanced antenna design techniques, which should be useful to most intermediate antenna engineers.

Zhijun Zhang
Tsinghua University, Beijing, China

Acknowledgments

It is a great pleasure to have a chance to thank the three most important people in my academic life. I would like to thank Professor Zhengde Wu, who was my supervisor when I studied at the University of Electronic Science and Technology of China as a master's student. Professor Wu equipped me with the most essential engineering skills which have benefited me ever since. I would also like to thank Professor Zhenghe Feng and Professor Magdy F. Iskander. Professor Feng was my advisor when I studied for my PhD degree at Tsinghua University. Professor Iskander was my supervisor when I worked as a postdoctoral graduate at the University of Utah. Both of them are my lifetime mentors. With their care and encouragement, I successfully made a mid-life career change, from an engineer to a professor. Since I joined Tsinghua University, I have cooperated successfully with Professor Feng and Professor Iskander, and I cannot thank them enough for all their help and guidance during this period, and hope to enjoy a continuous fruitful collaboration with them.

To the companies and individuals who provided illustrations and the copyright permissions for the illustrations, I am most appreciative. I am especially grateful to Yaobin (Richard) Gu and Xiaodong (Amanda) Jiang from Shanghai *Amphenol* Airwave Inc., who provided most of the photos and drawings of the production antennas in the book. I would also like to express my gratitude to the staff at John Wiley & Sons, especially James Murphy, Renee Lee, and Shelley Chow.

In the second edition of the book, some practical designs of real phones are included. I would like to thank Xiaohui Pi and Dr. Wendong Liu from Xiaomi Inc. for providing phone samples. I would also like to thank Lizhong Li from Shanghai *Amphenol* Airwave Inc. for providing semifinished antenna parts specially prepared for the book.

Abbreviations

3GPP	Third-Generation Partnership Project
AMPS	Advanced Mobile Phone Service
BT	Bluetooth
CDMA	Code division multiple access
CTIA	Cellular Telephone Industries Association
DCS	Digital Cellular Service
DS-MID	Double-shot molded interconnect device
EIRP	Effective isotropic radiated power
EIS	Effective isotropic sensitivity
EMC	Electromagnetic compatibility
ETSI	European Telecommunications Standards Institute
FCC	Federal Communications Commission
FER	Frame error rate
GPS	Global Positioning System
GSM	Global System for Mobile communication
HAC	Hearing aid compatibility
ICs	Integrated circuits
ID	Industrial design
IFA	Inverted-F antenna
LDS	Laser direct structuring
LTCC	Low-temperature co-fired ceramic
MIMO	Multiple input and multiple output
NHPIS	Near-horizon partial isotropic sensitivity
NHPRP	Near-horizon partial radiated power
OTA	Over the air
PCB	Printed circuit board
PCS	Personal Communications Service
PIFA	Planar inverted-F antenna
PTT	Push to talk

RAM	Radiation absorbent material
RL	Return loss
SAR	Specific absorption rate
SNA	Scalar network analyzer
SRF	Self-resonance frequency
TEM	Transverse electromagnetic
TIS	Total isotropic sensitivity
TRP	Total radiate power
UMTS	Universal Mobile Telecommunications System
UWB	Ultrawide band
VNA	Vector network analyzer
VSWR	Voltage standing wave ratio
Wi-Fi	Wireless fidelity
WLAN	Wireless local area network

1

Introduction

The twenty-first century is the wireless century. In the near future, it is very likely that most electronic devices will include some wireless functionality. If we look at the job market, known brands which seem to have nothing to do with antennas, such as Microsoft, Google, Amazon, and so on, are all recruiting engineers with antenna knowledge. On the other hand, there are not that many antenna engineers out there. The root cause of the shortage of antenna engineers can be traced all the way back to the university. The cornerstone of antenna engineering is *electromagnetics* (EMs), which is a quite abstract class and involves a lot of mathematics. The world unveiled by EMs is a four-dimensional one, which includes three spatial dimensions and one temporal dimension. To most students, the many new concepts introduced in the class are counterintuitive and confusing. As a logical consequence of natural selection, the EM major is removed by most students from their list of favorites.

People like to think of antennas as a black box of magic. The explanations given by antenna engineers are always so vague that it seems they never give people a definitive answer. It is easy to come to the conclusion that even designing a simple antenna requires years of experience. The truth is that if there was an appropriate book which presented all the required information, most electronic engineers who have studied some EM theory in university could design antennas. You do not need any mathematics to design an antenna. What you need is an understanding of how an antenna works. Of course, if you want to be an exceptional antenna engineer and design antennas with extreme constraints, a solid knowledge of EM theory and years of experience are still necessary.

This book provides a comprehensive discussion of the state-of-the-art technologies of antenna design for mobile communications. The book covers all the important aspects an engineer might need when designing an antenna, which includes how to make a fixture, how to design various antennas, how to optimize match circuits, and carry out different measurements.

It is recommended that the book is read in its entirety. However, for engineers who only want to design a single-band antenna in the shortest time possible, Section 1.6 will provide enough knowledge to kick-start a simple antenna project.

The book has six chapters, and the chapters are arranged as follows:

Chapter 1 provides an overview of most antenna design technologies used in mobile devices. Before anyone starts to design an antenna, it is very helpful for him or her to understand the following: (1) What can be done? (2) What kind of freedom do we have? Both topics will be briefly discussed here. Based on readers' feedback from the book's first edition, a practical example is added in Section 1.6. The section can also serve as a gamebook which can divert readers to different sections if they want to explore more.

Chapter 2 describes different matching techniques used in antenna design. In real-world engineering, antenna matching circuits are widely used, probably in at least half of all devices. The popularity of the matching network is due to two reasons: (1) it gives the engineer more freedom, one more parameter to play with when making design trade-offs; and (2) the value change of a matching component is quite a quick process, which can be a last-minute change. On the other hand, an antenna modification needs at least several days of lead time. The chapter discusses single-band matching, multiband matching, and advanced matching techniques. Complementary software written by the author will be provided to provide practice matching techniques (see the web address on the back cover).

Chapter 3 introduces different external antennas, including both stubby and whip–stubby antennas. The external antenna dominated the cell phone antenna design. The market share of external antennas has been consistently decreasing in the past decade, but it is still a very important antenna configuration. Many basic techniques used in external antennas, such as multimode single-radiator multiband antennas and multi-radiator multiband antennas, are also used in internal antennas.

Chapter 4 introduces different internal antennas. The internal antenna is the current fashion. Under the internal antenna category, there are several different concepts, such as folded monopole, inverted-F antenna/planar inverted-F antenna (IFA/PIFA), loop, and ceramic antenna. All of these will be discussed in the chapter.

Chapter 5 introduces important issues related to engineering antenna measurement. Besides the passive antenna measurement, which is familiar to most electronic engineers, active measurement will also be discussed. Some details, which are key to accurate measurement, such as how to make fixtures and use a choke, will all be covered in the chapter. Various antenna measurements in the production line are also covered in the chapter.

Chapter 6 is about the various regulations which are important to antenna engineers. These can be split into three topics: (1) specific absorption rate (SAR), which is about the radiation to the head and body; (2) hearing aid compatibility (HAC), which is about electromagnetic compatibility (EMC) with hearing aids; and (3) EMC, which is about the EMC with other devices.

1.1 The Evolution of Mobile Antennas

There is some argument about who invented the first mobile communication system, because for some people mobile communication also means vehicle communication. However, when referring to the first commercial handheld cellular phone, the answer is Motorola DynaTAC 8000X [1], without any doubt, which was introduced in 1983, as shown in Figure 1.1.



Figure 1.1 Sleeve dipole antenna on a Motorola DynaTAC 8000X (1983). (*Source:* Reproduced with permission of Motorola.)

The antenna installed on a DynaTAC 8000X is a sleeve dipole antenna [2], which now is an obsolete design in the mobile phone industry but still widely adopted by various wireless LAN access points, such as the one shown in Figure 1.2. Sleeve dipole antenna is the best performing antenna ever installed on any cellular phone; however, this is also the largest cellular phone antenna. The length of a sleeve dipole is about half the wavelength at its working frequency. At 850MHz, the antenna itself needs a length of 176mm. At the dawn of the personal mobile communication era, those dimensions look quite reasonable when compared to a vintage cellular phone. For instance, the dimensions of a DynaTAC 8000X are 330 mm × 44 mm × 89 mm, without the antenna.

With the significant improvement in cellular technology and the aggressive shrinkage of the size of phones, soon the size of a sleeve dipole was no longer proportional to the phone. Unlike dipole antennas, a monopole antenna [3] on a ground plane has only a length of a quarter of a wavelength, which is 88 mm at 850MHz. Shown in Figure 1.3 is a Motorola MicroTAC 9800X sitting on a charger. The phone is a flip phone and has a microphone located inside the flip. The thin wire on the top of the phone is a monopole whip antenna.

A sleeve dipole, such as the one shown in Figure 1.1, has an integrated choke which retains most radiation current within the antenna; thus, the antenna is insulated from the phone and also from a user's hand on the phone. However, a monopole antenna must use the metal inside a phone as part of the antenna's radiating structure. Some portion of radiating current must



Figure 1.2 Sleeve dipole antennas on a wireless LAN access point. Linksys WAP55AG. (Source: Cisco, Inc.)



Figure 1.3 Whip antenna on a Motorola MicroTAC 9800X (1989). (Source: Reproduced with permission of Motorola.)

flow over the phone. Putting one's hand on the phone absorbs some energy, and thus decreases the overall antenna performance. Although the performance of a whip monopole antenna is inferior to a sleeve dipole, it is still better than all other members of the family of cellular phone antennas. The whip antenna is the second largest one in the family.

In fact, the antenna used on the MicroTAC 9800X is a retractable antenna. A retractable antenna is a combination of a whip antenna and a helix stubby antenna. When the antenna is extended, it functions as a whip monopole and provides good performance. When the antenna is retracted, it functions as a stubby antenna and still has acceptable performance. The retractable antenna has the best of both worlds, as it is a low-profile solution and is still capable of providing good performance when needed.

Obviously, the mechanical structure of a retractable antenna is quite complex, as it involves moving parts and multiple radiators. A stubby antenna, as shown in Figure 1.4, eliminates the whip in a retractable antenna. From the performance point of view, a stubby antenna is not as good as a retractable one. However, stubby antennas dominated the cellular phone market at the end of the past century. The reason for the wide adoption of stubby antennas is the significant improvement in cellular networks. As the number of mobile phone users exploded, the density of base stations also increased dramatically. That means the distance from any user to the nearest base station is much shorter than previously. As the path loss between a cellular phone and a base station tower is directly proportional to the distance between them, a shorter distance means less strain on the antenna's performance. Inside a stubby antenna, the metal radiator can be a helix made of a metal wire, a meander line made of flexible printed circuit board (PCB), or a sheet metal stamping part.



Figure 1.4 Nokia, Inc. Stubby antenna on a Nokia 5110 (1998). (Source: Reproduced with permission of Nokia.)



Figure 1.5 Nokia, Inc. Internal antenna on a Nokia 3210 (1999). (*Source:* Reproduced with permission of Nokia.)

The next antenna to enter the market was the internal antenna. The phone shown in Figure 1.5 is not the first phone to adopt an internal antenna; however, it is one of the most successful phones with an internal antenna. Nokia sold approximately 160 million Nokia 3210 during the phone's whole life span. When tested in free space or next to a phantom head, an internal antenna can achieve performance similar to a stubby antenna. In everyday use, internal antennas are more vulnerable to hand blockage by the user. It is quite a natural gesture for a user to put his or her fingers on top of the antenna and bring the speaker closer to his or her ear.

From the mechanical point of view, the internal antenna is better than the external antenna, as it eliminates the through hole and mating features necessary to accommodate an external antenna. A phone with an internal antenna normally has better performance in drop tests, wearing tests, and various other mechanical tests. Because an internal antenna is totally concealed in the phone, the phone user has little chance to abuse it. Some people have the habit of playing with the item in their hand when they are sitting in meetings or are idle in front of their desks. I have seen some colleagues unconsciously extend and retract their antenna's whip hundred of times in a single meeting.

All traditional internal antennas are located on the upper part of a phone. In a normal talking position, the distance between the top internal antenna and the user's head is quite small. To eliminate the influence of a user's head on the antenna's performance and also decrease the harmful radiation emitted toward the head, a ground layer must be placed beneath the antenna to increase isolation between the user's head and the antenna. However, the ground layer

decreases an antenna's bandwidth. To compensate, the antenna size must be increased. The Motorola Razr V3 was the first phone to adopt a bottom internal antenna. It was a brave act. According to the conventional wisdom of that time, an antenna in the bottom would be held in the center of a user's palm; a bottom antenna might have good performance in the lab but could not provide acceptable performance in real use. That conventional wisdom was proved wrong by the Motorola V3. The Motorola V3 has become another legend in cellular phone history. It sold more than 110 million. By relocating the antenna to the bottom, the antenna is away from the head. The ground layer, which is required by top internal antennas, can be eliminated. Furthermore, the antenna's thickness and volume can both be significantly decreased, as shown in Figure 1.6. The Motorola V3 was the slimmest phone when it was released. Since then, many slim phones have adopted bottom internal antennas, and most big players in the cellular phone market have their own versions of bottom antenna phones. The new wisdom is that whenever you need to design a slim phone, it is better to put the antenna on the bottom.

Shown in Figure 1.7 is the first-generation iPhone, which was released in 2007 by Apple Inc. iPhone is the first phone equipped with a capacitive touch screen. Unlike its predecessor's resistive touch screen, capacitive touch screen does not require a stylus and can be controlled directly by fingers. The detecting layer of capacitive touch screen, which is usually made of a transparent conductor such as indium tin oxide, is embedded under a piece of glass. This configuration gives capacitive touch screen a sleek feeling and almost infinite life span. Companies with the iOS software, iPhone became a disruptive force in the phone market.



Figure 1.6 Bottom internal monopole antenna on a Motorola Razr V3 (2004). (Source: Reproduced with permission of Motorola.)



Figure 1.7 First-generation iPhone (2007). (*Source:* Apple, Inc.)

Since the appearance of the first-generation iPhone, the whole phone industry starts to converge. All companies which insist on including a keypad on their phones didn't end well. In 2015, front portraits of most phones look like they are taken from identical twins. If only looking from the front, one might find it is quite difficult to separate a sub-100 US dollar functional phone from a 600+ US dollar flagship phone. Shown in Figure 1.8 is an iPhone 6s plus. A big screen takes out most of the area of the front surface. Two slices of blank areas, one on the top and one on the bottom, occupy the rest of the area.

The current trends of mobile phone designs are making the screen larger and pushing the device thinner. When the original iPhone came out, it was considered a large-screen phone and also the thinnest. It has a 3.5-inch screen and measured 11.6 mm thick. After 8 years, a 4.5-inch screen is considered small. The iPhone 6 plus has a 5.5-inch screen and its thickness is only 7.1 mm. Some companies, such as Huawei, have even released 6- and 7-inch models. The boundary between phone and tablet device has been blurred.

All these trends have considerable impacts on the antenna designing and manufacturing techniques. As the big screen is actually a liquid crystal display (LCD), which is one of the main noise sources, a piece of metal shielding is always applied on LCD's back. This shielding and the phone's slim form factor make it almost impossible to design antennas in the middle portion of a phone. Thus, pretty much every phone puts its battery and main circuit board in the middle and squeezes all antennas into blank areas on top and bottom of a phone, as illustrated in Figure 1.8.



Figure 1.8 Top and bottom antenna arrangement on an iPhone 6s plus (2015). (Source: Apple, Inc.)

Due to the concern of radiation to human brain, most companies put their main antenna, which is responsible for transmitting second-generation (2G), third-generation (3G), or fourth-generation (4G) signals, on the bottom of the phone. The only exception the author is aware of is iPhone. Before iPhone 4s, similar to others, iPhone only had one bottom main antenna. Then there was the infamous “Antennagate.” The performance of an iPhone 4 antenna could be significantly degraded by tightly holding the phone’s bottom portion, which is nicknamed as “death grip.” The solution Apple came out with is dual main antennas, one on the top and one on the bottom. Since iPhone 4s, all iPhones have two main antennas. It can dynamically switch between these two antennas based on usage. If it detects a head next to phone, it switches to the bottom antenna. If someone is holding the phone at the bottom and the signal strength is too weak, it will switch to the top antenna. As Apple is holding several patents [4, 5] on this switching scheme, it will be difficult for other companies to follow suit.

Thousands of models of cell phones have hit the streets since 1983. It is almost impossible to list them all. To get more comprehensive information, the Internet is a good resource. Some posts [6] show chronicles of cellular phones. Some websites [7] are dedicated to the phones’ news. For more detailed information about certain phones, which are sold in the United States, go to the Federal Communications Commission (FCC) website [8].

1.2 How to Quantitatively Evaluate an Antenna

After designing an antenna, we cannot say whether it is good or bad by simply looking at it. We must find a way to quantitatively evaluate it. In cellular antenna's designs, the frequently used parameters are the reflection coefficient, the voltage standing wave ratio (VSWR), efficiency, gain, and bandwidth. The contents of this section are only a brief review of the frequently used parameters. More comprehensive materials and detailed deductions can be found in some classical textbooks [3, 9–12].

From the circuit point of view, an antenna is a single-port device. A transmission line can be used to feed the antenna, as shown in Figure 1.9. An input signal takes the form of an incident wave traveling along the transmission line. It flows from the signal source toward the antenna, assuming that the amplitude of the incident wave is V_{incident} . At the antenna port, some of the energy carried by the incident wave is radiated by the antenna. In the meantime, the residual energy is reflected at the port and travels back along the transmission line. The amplitude of the reflected wave is $V_{\text{reflected}}$.

The reflection coefficient is given by

$$\Gamma = \frac{V_{\text{reflected}}}{V_{\text{incident}}} \quad (1.1)$$

Clearly, all the reflected energy will be wasted. When designing an antenna, our goal is to minimize the reflection at the antenna port. A perfectly matched antenna can radiate all energy, thus its reflection coefficient is 0. When a device reflects all the energy back, its reflection coefficient is 1.

In microwave theory, the S -parameter matrix is used to quantitatively describe a multiport network. The S stands for scattering. A one-port network is a special type of multiport networks; its S -matrix degenerates to a single element, S_{11} . For an antenna, the definition of S_{11} is identical to the reflection coefficient.

$$S_{11} = \Gamma \quad (1.2)$$

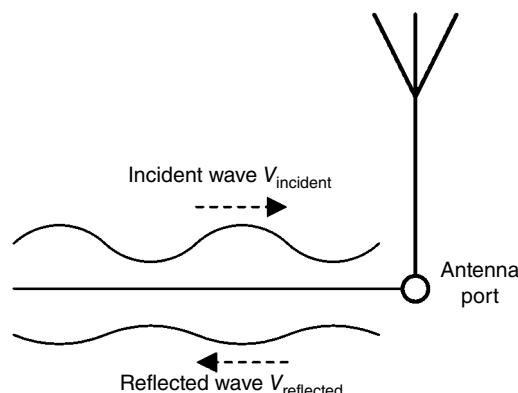


Figure 1.9 Reflection coefficient.

In engineering, the S_{11} is often used in the decibel (dB) scale.

$$S_{11} \text{ (dB)} = 20 \log_{10} (|S_{11}|) \quad (1.3)$$

The S_{11} is defined by the ratio of the voltages of incident and reflected wave, while the S_{11} (dB) is defined by the incident and reflected power. That is the reason why the coefficient in Equation 1.3 is 20. As the $|S_{11}|$ of any antenna is a value less than 1, the S_{11} (dB) is always a negative value. The absolute value of S_{11} (dB) is called the “return loss” (RL):

$$RL = |S_{11} \text{ (dB)}| \quad (1.4)$$

Although the definitions of Γ , S_{11} , S_{11} (dB), and the RL are somehow different, they are all deduced from the incident wave and the reflected wave. The other commonly used parameter, VSWR, is directly defined by the standing wave formed by the superposition of the incident and reflected waves.

$$\text{VSWR} = \frac{|V_{\max}|}{|V_{\min}|} \quad (1.5)$$

The VSWR is the ratio of the amplitude of a partial standing wave at an antinode (maximum voltage) to the amplitude at an adjacent node (minimum voltage) in an electrical transmission line. Although the VSWR’s physical meaning might seem less straightforward than Γ , the VSWR is the only parameter that could be easily measured when the microwave and antenna technology was still in its infancy. Today, the VSWR is still widely used, especially in the antenna business. The correct format of VSWR is X : 1, such as 2 : 1, 3 : 1, and so on. A VSWR 2 : 1 means the maximum voltage is twice as much as the minimum voltage.

As the V_{\max} and V_{\min} are formed when the incident and reflected waves are constructively and destructively superimposed, respectively, Equation 1.5 can be rewritten as follows:

$$\text{VSWR} = \frac{|V_{\text{incident}}| + |V_{\text{reflected}}|}{|V_{\text{incident}}| - |V_{\text{reflected}}|} = \left(1 + \frac{|V_{\text{reflected}}|}{|V_{\text{incident}}|} \right) / \left(1 - \frac{|V_{\text{reflected}}|}{|V_{\text{incident}}|} \right) = \frac{1 + |\Gamma|}{1 - |\Gamma|} \quad (1.6)$$

The relation between VSWR and Γ , or the RL, is a one-to-one correspondence. The RL of 10 dB is a commonly used specification for antennas. The corresponding VSWR is approximately 2 : 1.

Bandwidth is another important parameter used to describe antennas. Whenever we give an antenna’s bandwidth, we must give the criteria that define the bandwidth. As shown in Figure 1.10, the antenna has a -10 dB bandwidth of 70 MHz. However, you can also claim that the antenna’s bandwidth is 132 MHz, if one uses -6 dB as the criteria. Different companies might use different criteria to measure their antennas; it is our responsibility to pay a little more attention to the details.

A well-matched antenna does not necessarily mean it is a good antenna. Efficiency is the parameter which tells us how well an antenna can radiate. The efficiency is given by

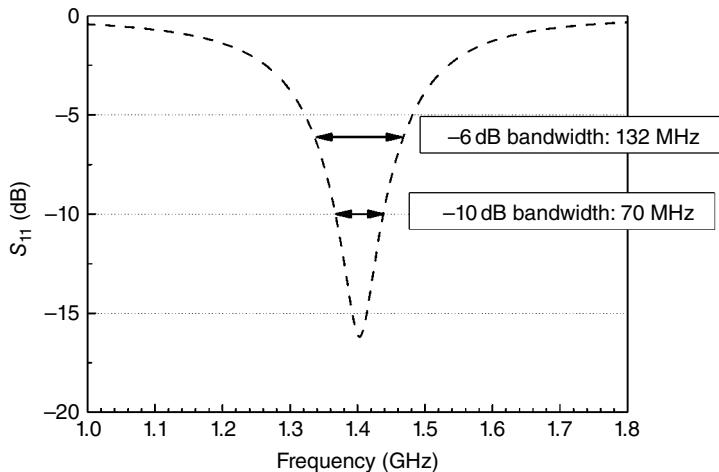


Figure 1.10 Defining an antenna's bandwidth.

$$\text{Efficiency} = \frac{P_{\text{radiated}}}{P_{\text{total available}}} \quad (1.7)$$

where the P_{radiated} is all the power radiated and the $P_{\text{total available}}$ is the total available power from the signal source. Efficiency is a value between 0 and 1. In the antenna business, the efficiency in dB is also commonly used.

$$\text{Efficiency (dB)} = 10 \log_{10} (\text{efficiency}) \quad (1.8)$$

A dB efficiency of -3 dB means 0.5 or 50% efficiency in the linear scale, which is still a pretty good value for real antennas.

In the cellular antenna's world, the gain is not an important parameter, because it is mostly decided by the position in which an antenna is installed and the size of the grounding structure. The antenna element itself does not have too much to do with deciding the gain. The commonly used units for gain measurements are dBi, dBd, and dBic. These are normalized to isotropic linear polarized antenna, dipole antenna, and isotropic circular polarized antenna, respectively. More information about gain can be found in Chapter 5.

1.3 The Limits of Antenna Designs

As antenna engineers, we are under consistent pressure to shrink the size of the antennas and still provide better performance. There is an elegant art to communicating with team members and managers from other disciplines when explaining that a limit in antenna design does exist. For each kind of antenna, there is a boundary, which regulates an antenna's size and its performance. As a new engineer, the easiest way to get a feeling for that boundary is by measuring various phones designed by different companies. Also a much quicker way to learn new design techniques is by reverse engineering using existing antennas on the market.

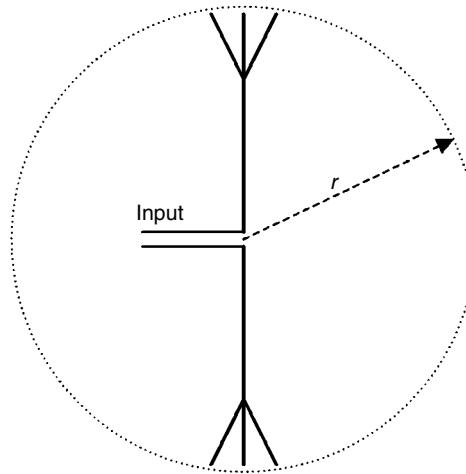


Figure 1.11 The minimal sphere encloses an antenna.

In 1948, L. J. Chu published a paper [13] which quantified the relationship between the lower boundary for the radiation quality Q of an electrically small antenna and its physical size relative to the wavelength. This lower boundary is now known as the “Chu” limit. Shown in Figure 1.11 is a schematic diagram of a vertically polarized omnidirectional antenna. The sphere with radius r is the minimum one which can enclose the antenna. The lower boundary for the radiation Q is decided by r/λ . The boundary given by Chu is based on a simplified model and is considered as the strictest one. Several boundaries based on more realistic scenarios [14–18] have been proposed since then. However, Chu’s limit is still the one that is most referred to.

Bandwidth can be derived from Q by assuming that the antenna is a resonant circuit with fixed values. The normalized bandwidth between the half-power frequencies is [14]

$$\text{Bandwidth} = \frac{f_{\text{upper}} - f_{\text{lower}}}{f_{\text{center}}} = \frac{1}{Q} \quad (1.9)$$

Equation 1.9 is a good approximation when $Q \gg 1$. Otherwise, the representation is no longer accurate. Shown in Figure 1.12 is a figure presented in reference [14]. The x -axis is $kr = 2\pi(r/\lambda)$. The y -axis represents the quality. Different curves are single mode Q for various antenna efficiencies.

With a fixed efficiency value, say, 100%, when the sphere’s radius increases, the radiation Q decreases, which also means that the maximum achievable bandwidth increases. Of course, the bandwidth predicted by the curve can never be achieved in an actual implementation. Various studies [19–22] have been done to approach the limit.

Another thing that can be observed in Figure 1.12 is that a lossy antenna, which has lower efficiency, always has a wider bandwidth. In the real world when the bandwidth of an antenna is abnormally wide, this is not good news, because most of the time it is due to unwanted loss.

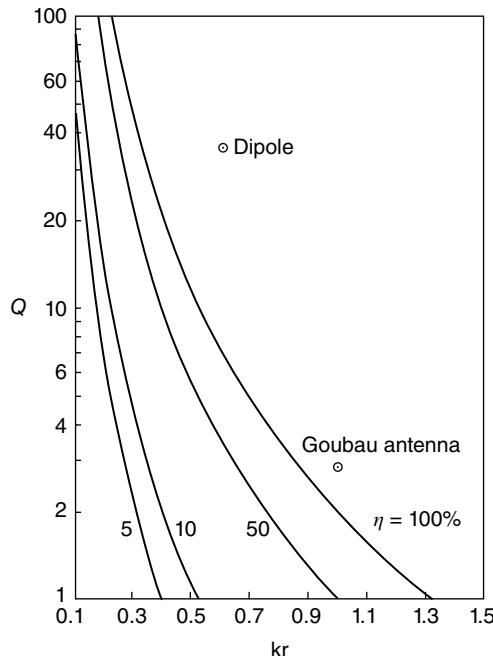


Figure 1.12 Chu–Harrington fundamental limitations for single-mode antenna versus efficiency. (Source: Hansen [14]. Reproduced with permission of IEEE.)

As antenna engineers, we do not really evaluate the achievable bandwidth based on figures and formulas given in references. It is very difficult to define the minimum sphere to enclose the antenna in a cellular phone. It will be demonstrated later that all metal structures, including the ground, in a phone can give off radiation. If we define a sphere that encloses the whole phone, the bandwidth calculated by the Chu limit can be so wide that it is meaningless. From time to time, there are claims that the Chu limit has been surpassed. In most cases, the sphere used in calculations only encloses the antenna element itself. As the ground is also part of the radiator, by excluding the ground from the sphere, the achievable bandwidth is artificially narrowed, and that is why those antennas have wider bandwidth than the theoretical limit.

When designing a cellular antenna, many factors, such as the nearby battery, the speaker under the antenna, the metal bezel on the phone, and so on, all play a role in determining the achievable bandwidth. With the accumulation of experience, eventually one can estimate the achievable bandwidth more accurately.

1.4 The Trade-Offs in Antenna Designs

To be a good antenna engineer not only means designing an antenna with the best performance, but it also means having a profound understanding of the possible trade-offs in antenna designs. Among them, some trade-offs are the same ones which are applicable in all engineering disciplines, such as the trade-offs between the design time and performance. Designing

an antenna is a project with a time constraint instead of an open-ended art creation. The thought of designing a perfect antenna might do more harm than good. In addition to those commonsense trade-offs, there are some that are particular to antennas:

- Bandwidth trade-off. In Section 1.3, the bandwidth limit of a single-band antenna was discussed. Most phone antennas used today are multiband antennas. A similar limit also applies to their combined bandwidth. If an antenna is well designed, whenever the bandwidth of one band increases, the bandwidth of the other bands must shrink. To fully understand the design technique of one kind of antenna, we need to find out how to trade-off bandwidth between different bands. For example, if the specification for an antenna is 50% efficiency across all bands, and the efficiency of the antenna designed is 60% at the lower band and 40% at the high band, your work hasn't finished yet. The unbalanced performance tells everybody that you have not really mastered the design skills of this antenna.
- Trade-off between complexity and performance. By introducing more freedom into an antenna's design, it is possible to achieve better performance. However, the marginal improvement of each incremental variable is regressive. As a new antenna engineer, try to avoid using complex designs in the beginning. It is quite easy to be drowned by a large amount of variables. One should start from simple designs and assess the impact of each design variable. For many applications, an antenna with a handful of variables is good enough.
- Trade-off between manufacture consistency, tooling time, and cost. Better manufacturing consistency means less antenna variation and better antenna performance. However, better consistency also means longer tooling time and higher cost. There is no manufacturing solution that can provide all the benefits; otherwise, it would already have been part of the antenna's manufacturing process. Many different manufacturing processes are available; one should understand the advantages and the disadvantages of each of them. Taking the processing of internal antenna as an example, there is metal-stamping, flex circuit, double-shot molded interconnect device (DS-MID), laser direct structuring (LDS), and so on. The metal-stamping technology is the cheapest and can be adjusted quite quickly if the parameter that needs to be adjusted is known and already included in the tooling design. The flex circuit technologies have better consistency than the metal-stamping; however, it is a little more expensive and it takes a longer time to implement a design change. Both DS-MID and LDS technology have the best consistency, because antennas are part of the plastic structure instead of separate parts. Both of them are based on a technique called "selective metallization." The DS-MID process begins with the application of a shot of plateable thermoplastic resin in an injection mold cavity. Next, the cavity is changed and a second shot of nonplateable thermoplastic resin is molded around the first shot to create a circuit pattern from the plateable material. Depending on the antenna shape, the two resins can be reversed in shot order. After two shots, a part has its intended geometry with select plateable surfaces exposed. These surfaces are then plated with a layer of copper. The DS-MID takes the longest lead time, because any modification to the antenna pattern requires tooling changes. The LDS is a relatively new process. The thermoplastic resin used in LDS process is non-plateable after the molding process and can be transformed to plateable by using a laser beam to activate it. The LDS process literally draws the antenna pattern onto the plastic. The pattern can be adjusted quite easily by uploading a new pattern file to the laser. Similar to the DS-MID, a plating process is required to deposit copper onto the part's surface.

- Trade-off between total radiated power and radiation exposure. Higher total radiated power and lower radiation to the human body are a pair of contradictory requirements. Ideally, we should first design an antenna to have an on-phantom efficiency as high as possible, then choose an appropriate conductive power level to meet the human exposure specification. In reality, there are three constraints: the conductive power, the total radiated power, and the human exposure. An antenna with very good efficiency might give you trouble later on. In the whole design process, one should always keep these three specifications in mind and check their status constantly.

1.5 Mobile Communication and Band Allocations

The radio frequency (RF) EM spectrum is an aspect of the physical world which, like land, water, and air, is subject to usage limitations. The use of RF bands in the EM spectrum is regulated by governments in most countries, in a spectrum management process known as frequency allocation or spectrum allocation [23]. Although countries are working on a universal frequency allocation plan, the existing frequency allocations are still country dependent. In the United States, the spectrum from 0Hz to 1000GHz was allocated by the Federal Communications Commission (FCC) [24]. The US Department of Commerce has a color chart of frequency allocation, which covers 3 kHz to 300 GHz [25]. In most countries, the spectrum allocation plan is not a static one and is being continuously revised.

Most bands used in the design of mobile phones are given in Figure 1.13. However, this is not a complete list.

There are some ambiguities when referring to bands and their respective technology. Only the band names are given in Figure 1.13. Depending on the different countries, the exact frequency range of each band might vary slightly.

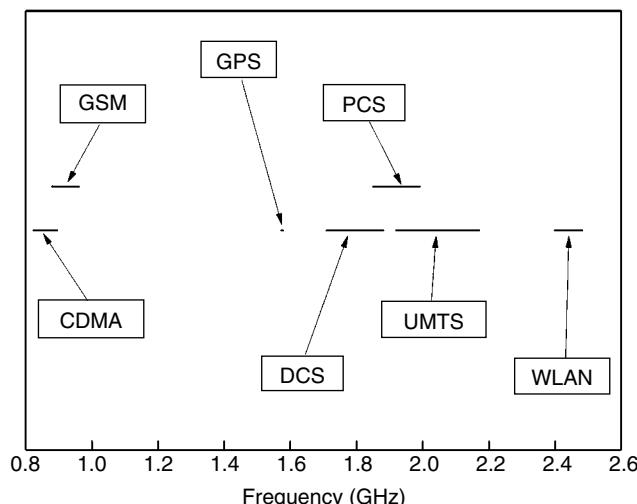


Figure 1.13 Band allocation.

- CDMA band: also known as AMPS band or 850 MHz band, 824–894 MHz
- GSM band: also known as 900 MHz band, 880–960 MHz
- GPS band: 1575 MHz
- DCS band: also known as 1800 MHz band, 1710–1880 MHz
- PCS band: also known as 1900 MHz band, 1850–1990 MHz
- UMTS band: also known as 3G band or 2100 MHz band, 1920–2170 MHz
- LTE band: also known as 4G band, 700–800 MHz and 1710–2700 MHz
- WLAN band: also known as Bluetooth band or 2.4 GHz band, 2400–2480 MHz

Strictly speaking, using those abbreviations to name bands is not appropriate. Some of them are based on specific technologies. The following are brief introductions; more comprehensive information can be found on their respective websites and in other books [26].

- AMPS is the abbreviation for Advanced Mobile Phone Service. It is an analog standard used by the first cellular communication network. Motorola DynaTAC 8000X is based on AMPS technology. It is obsolete in most countries. The United States was one of the last to shut down AMPS services. The final date of use was February 18, 2008.
- CDMA is the abbreviation for code division multiple access [27]. In the cellular business, this means the IS-95 standard or the cdmaOne standard. The technology itself is band independent. In the United States, CDMA systems are deployed in both 850 and 1900 MHz bands.
- GSM is the abbreviation for Global System for Mobile Communications [28]. Both GSM and CDMA are 2G cellular communication standards. In the global market, GSM is the most influential standard. About 80% of the global mobile market used this standard in 2009 [28]. The GSM technology itself is also band independent. GSM systems are deployed in different bands, such as 850, 900, 1800, and 1900 MHz.
- GPS is the abbreviation for Global Positioning System [29]. GPS is a receiver-only technology. It can extract positioning and timing information from signals transmitted by GPS satellites. It is not a mandatory feature for a phone. However, it gradually has become a standard functionality for most middle- to high-tier phones.
- DCS is the abbreviation for Digital Cellular Service. It is the name of the 1800 MHz band.
- PCS is the abbreviation for Personal Communications Service. It is the name of the 1900 MHz band.
- UMTS is the abbreviation for Universal Mobile Telecommunications System [30]. UMTS is one of the 3G mobile telecommunications technologies. The most common form of UMTS is W-CDMA. The Chinese version 3G system, TD-SCDMA, also belongs to the UMTS family. The main competitor of UMTS is CDMA2000, which is another 3G standard. As most countries allocate the 2100 MHz band to 3G systems, UMTS is used as the 2100 MHz band's alternative name. In fact, UMTS has been deployed in different bands, such as 850 and 900 MHz bands.
- LTE is the abbreviation for Long-Term Evolution, commonly marketed as the 4G LTE [31]. LTE has allocated around 40 operating bands and many of them overlap with previous 2G and 3G bands. The allocated bands cover 700, 750, 800, 850, 900, 1800, 1900, 2100, 2300, 2500, 2600, and so on. Similar to what had happened to the 3G standard, most of the country adopts the LTE-FDD standard and China backs TD-LTE standard. However, two smaller carriers in China, which have a combined market share of 35% in 2014, adopted LTE-FDD. LTE bands are omitted in Figure 1.13.

- WLAN is the abbreviation for wireless local area network. WLAN actually involves several standards, such as 802.11b, 802.11g, 802.11a, and so on. WLAN is also known as Wi-Fi [32], which is the acronym for wireless fidelity. From the antenna point of view, the 802.11b/g uses the 2.4GHz band and the 802.11a uses the 5GHz band, which is omitted in Figure 1.13.
- Bluetooth [33] is a standard different from WLAN. It is an open wireless protocol for exchanging data over short distances. However, Bluetooth shares the same 2.4GHz frequency band with WLAN.

Besides the aforementioned bands, some other technologies, such as FM radio, analog TV, and digital TV, are also used in cellular phones. Their band allocations are omitted in Figure 1.13. Another point worth mentioning is that most technologies and band allocations used in Japan are different from the rest of the world. As the Japanese market is the most difficult one to penetrate, related information of that market is also omitted.

In this chapter, only the basic terminologies are introduced. For more in-depth information about different cellular communication technologies, Wikipedia [34] is a good place to start.

1.6 Quickly Building a Simple Antenna—a Practical Example

If an engineer must design an antenna in the shortest time, the antenna is most likely a 2.4GHz one. The 2.4GHz band is one of the industrial, scientific, and medical (ISM) bands. The 2.4GHz band is reserved internationally for short-range communications and it is a license-free band. Many technologies, such as WLAN, Bluetooth, Zigbee, ANT, and cordless phones, use this band.

Design of a 2.4GHz antenna is not a difficult task, because the relative bandwidth of the 2.4GHz band (2.4–2.483GHz) is only 3.7%. Various ceramic chip antennas, which are discussed in Section 4.5, are marketed for these bands. However, for a product which is not space constrained, IFA is a better choice. Unlike chip antennas, an IFA can be integrated into the PCB and thus can be considered as a freebee. Detailed introductions of an IFA can be found in Section 4.1; here, we will carry out the design without going into theoretical details.

A network analyzer is essential to any antenna designing or tuning. Discussions of a network analyzer can be found in Section 5.1.1. If you work in a start-up and does have one on hand, you should convince your manager to buy one. Because you need to check the antenna and debug for problems during the whole life span of a product, renting one is not a good idea.

To be an advanced antenna designer, one should master the Smith chart and use it to design complex antennas. Detailed techniques related to the Smith chart and matching can be found in Chapter 2. Because a 2.4GHz antenna is a simple one, we will design the antenna without the Smith chart and only use reflection coefficient S_{11} .

Shown in Figure 1.14 is a network analyzer. Set the measurement type to [S11] and the display format to [Log Mag]. Set the start frequency to [2GHz] and the stop frequency to [3GHz]. Set the marker frequency to the center of the band [2.44GHz]. Although we only are concerned about 2.4–2.483GHz, it is a good practice to set the frequency span wide enough.

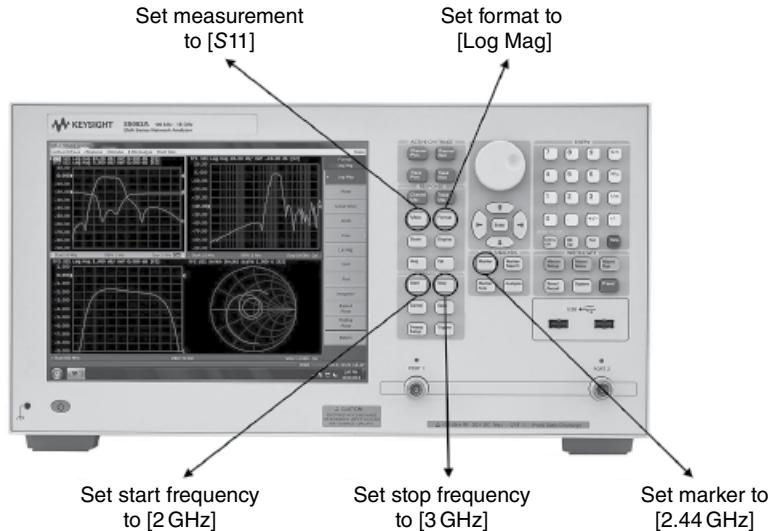


Figure 1.14 Network analyzer.

Thus, we can still observe an antenna's response even it is significantly off the design target, which happens a lot to a fresh engineer.

A network analyzer might have two or four ports. As we have selected [S11], the left-most port is the one to which the test cable should be attached. The coaxial connector on a network analyzer can be N-type, SubMiniature version A (SMA)-type, and so on. You might have to find an adaptor to mate a network analyzer to your cable. If you are not familiar with all these connectors, check a network analyzer's manual to decide which type of adaptor you need. You can order them online from the Digi-key Inc. [35]. There are various grades of coaxial cables, which range from tens of US dollars to thousands of US dollars. For the current project, a half-meter cheap cable (30–60\$) with SMA connectors, which normally has a maximum working frequency up to 18 GHz, will be good enough.

To make accurate measurement, the network analyzer to be used should be calibrated with the testing cable attached. However, in this practice we will proceed without calibration. Ideally, the response of a calibrated analyzer, which has one end of the test cable left open, should be a flat line which overlaps with the 0dB grid line. Shown in Figure 1.15a is the response we should get without calibration. The line should be a slightly inclined curve. The reading at 2 GHz is a little higher than at 3 GHz, that is because a cable always introduces higher loss at the upper band. If the cable is not too long and the network analyzer is not that antique, the measured curve should all be above –1 dB. This will cause about 1 dB error in antenna measurement, which can be tolerated in this practice.

If there are large ripples or nulls in the measured result, as shown in the Figure 1.15b, the connection between the network analyzer and the cable is not secured. If the S_{11} reading is too low, say minus several dB, the cable is either too long or too lossy, and it should be replaced.

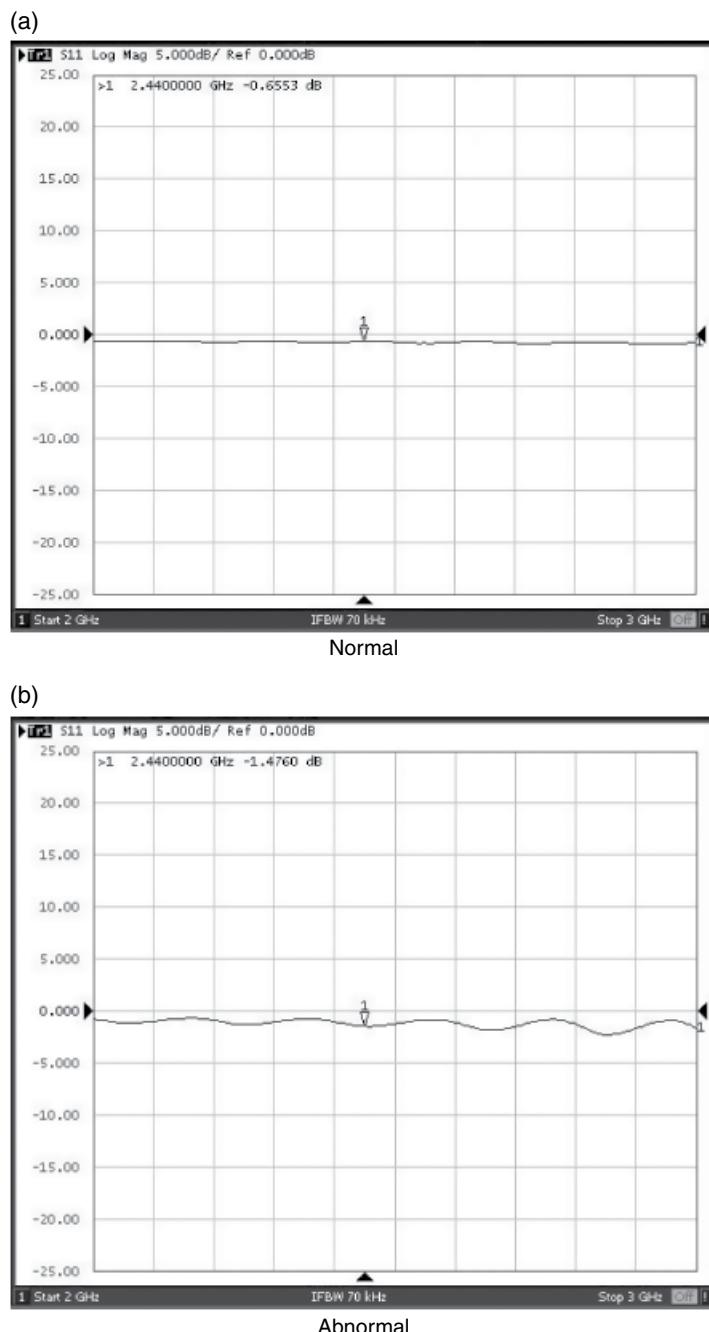


Figure 1.15 Response of an open-ended cable.

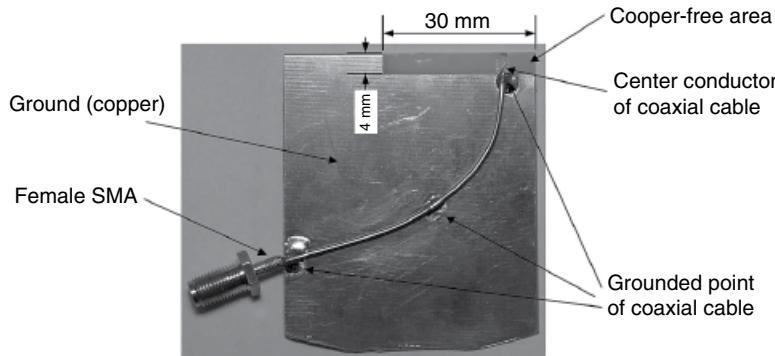


Figure 1.16 Antenna fixture.

Next, we need to find a single-side PCB, which has a solid copper layer only on one side of the board. When making antenna prototypes, try to avoid double side PCB, unless you wanted to spend time to manually make multiple connections between copper layers on top and bottom sides. Without proper connection between both sides, a double-side PCB might generate weird result. A heavy-duty scissor can be used to cut the PCB according to a mechanical design.

To place the antenna, a piece of copper measuring $30\text{mm} \times 4\text{mm}$ should be removed from the PCB. A pigtail cable is then soldered to the PCB. Shown in Figure 1.16 is a prepared antenna fixture. The outer conductor of coaxial cable has been soldered to ground. The center conductor of coaxial cable has been left open. As the preparing and soldering of the pigtail cable is critical to obtain correct and constant results, it is strongly recommended that readers spend some time to go through Section 5.1.2 before continuing this practice.

Now it is time to actually design the antenna. A full wavelength of 2.4 GHz is 125 mm and a quarter of wavelength is 31.25 mm. Because commonly available PCBs use FR4 as its substrate, which has a permittivity around 4.4 and can shift antenna resonant frequency toward lower band, the finished antenna is always shorter than a quarter of wavelength. We can start with 21 mm. The antenna element is cut from a copper tape, which has glue on the bottom side and can be ordered from Digi-Key. Use finger nail to press the copper tape, until it is tightly attached to the PCB. Shown in Figure 1.17a is the half-finished prototype. The antenna has been soldered to the coaxial cable's center conductor. The measured S_{11} of the half-finished antenna has been shown in Figure 1.17b. A ferrite choke was used when measuring the antenna. To make a repeatable measurement, choke is very important. More discussions about choke can be found in Section 5.1.2.2. However, because the bandwidth of this example is quite wide and the PCB is big enough, one can move on without a choke.

As shown in Figure 1.17b, the measured S_{11} barely reaches -10dB , so some matching work is required. There are various matching techniques, and Chapter 2 has been dedicated to this topic. In the following practice, a grounding branch method was used. As shown in Figure 1.18, one end of the branch was soldered to the antenna and the other end is soldered to the ground. Fresh engineers might wonder whether the grounding branch has shorted the antenna. Please relax, it isn't a short circuit, and it functions as a shunt inductor at the feeding point.

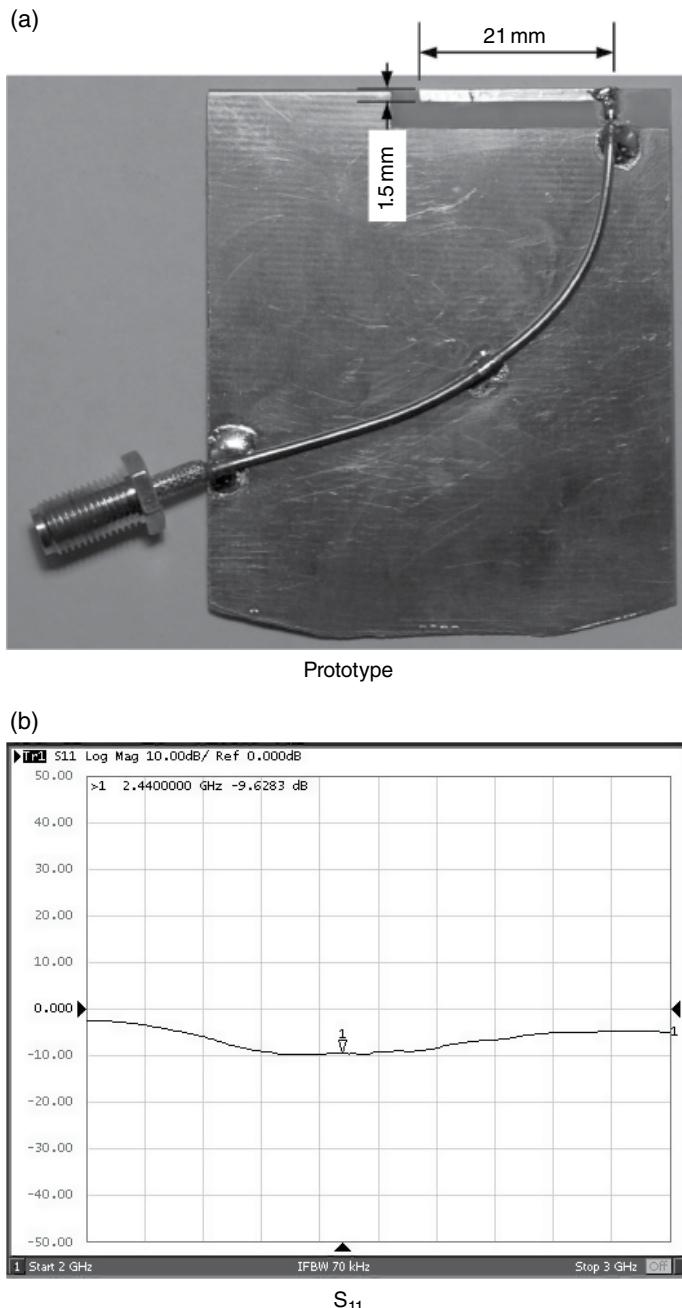


Figure 1.17 Make an antenna with copper tape.

Whenever we apply a new copper tape to an existing antenna structure, always use a solder to guarantee secured contact. Although copper tape is marketed as a conductive tape and can provide direct current (DC) connection, it can cause a lot of headache due to intermittent RF contact.

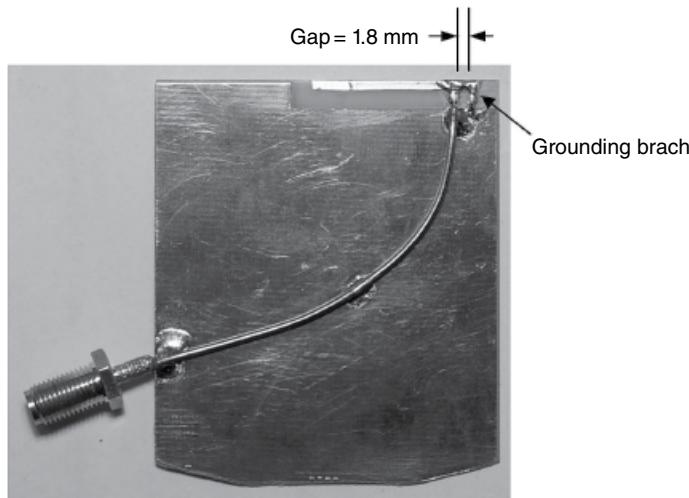


Figure 1.18 Match an IFA by grounding branch.

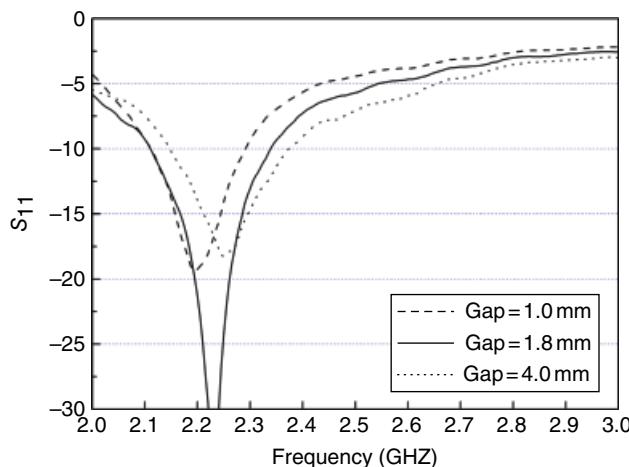


Figure 1.19 Impact of the grounding branch.

As we are not using the Smith chart, the matching effort will be a little less efficient. Fortunately, single-band IFA only has one tuning parameter, which is the gap between feeding strip and grounding strip. There is only one sweat spot for matching. Other than that point, the matching is always degraded no matter whether the gap is too wide or too narrow. Shown in Figure 1.19 are measured results of three different gap values. It is obvious that a 1.8 mm gap gives the best matching.

So far, the IFA has been matched. However, the resonant frequency is not at 2.44 GHz, which is the design target. Because the resonant frequency is lower than 2.44 GHz, we can shift it higher by shorting the antenna trace. The antenna can be tuned by cutting the antenna from 21 to 19 mm. Shown in Figure 1.20 is the measurement result. The center frequency is right at 2.44 GHz; however, the matching has degraded. It normally takes a few iterations to finalize a design.

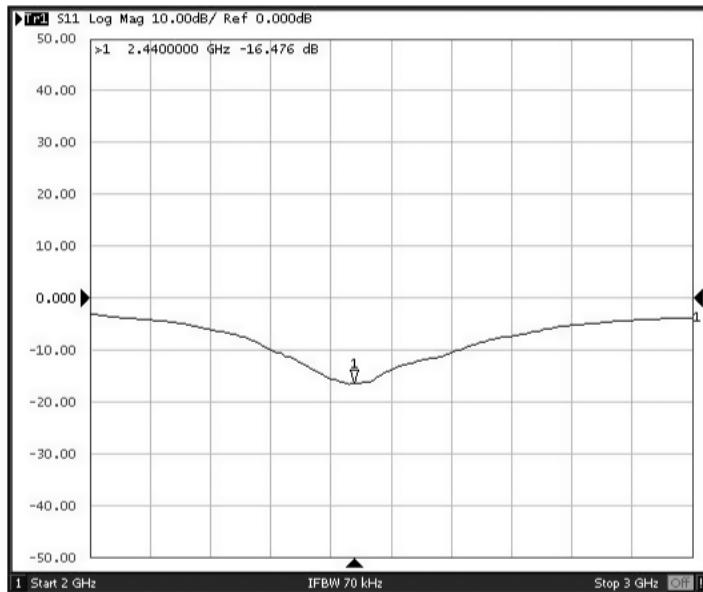


Figure 1.20 Tuning resonant frequency.

In real life, there is always an external plastic case to protect a PCB and components populated on it. Those plastic parts and various types of components around the antenna can have some impact on the antenna, which always requires some further fine-tuning. Shown in Figure 1.21a is a photo of a prototype. The sample antenna is at the top left corner. The plastic cover causes a frequency drift of 100 MHz toward lower frequency. After a few more iterations, the antenna was retuned. The length of the finalized antenna is 18 mm and the gap is 3.5 mm. Shown in Figure 1.21b is the measured S_{11} of the final design. The -10 dB bandwidth is around 500 MHz. Comparing with the measured results shown in Figure 1.19, the bandwidth is significantly increased. The wider bandwidth is due to the dielectric loss caused by the plastic cover. Although the bandwidth seems improved, the antenna efficiency is actually worse.

To make an IFA smaller, the common practice is to use meander line as an antenna element, as shown in Figure 1.22a. The antenna size is 14 mm \times 3 mm and the gap between grounding branch and feed is 1.2 mm. As a rule of thumb, whenever an IFA becomes smaller, its grounding branch always gets closer to the feed. Although the size of a meander antenna is smaller, the length of the metal line is always longer. More discussion can be found in Section 3.1.1.

The measured S_{11} has been given in Figure 1.22b. The -10 dB bandwidth is around 300 MHz. The reason for such a wide bandwidth is because the relative large ground plane. If the antenna was put on a much smaller PCB, the achievable bandwidth would shrink significantly. More discussion about the effect of ground plane can be found in Section 3.4.

The tuning procedure of meander line IFA is pretty much the same as the previous example. However, whenever an antenna is squeezed into a smaller area, the antenna bandwidth always becomes narrower and antenna efficiency gets worse. At some point, it will be a potential problem for an integrated IFA. Although the permittivity of FR4 substrate, which is the most commonly used PCB material, is marked as 4.4, the actual permittivity changes from vendor to vendor and from batch to batch. The permittivity can vary from 4.0 up to 4.8. If the bandwidth of a design is

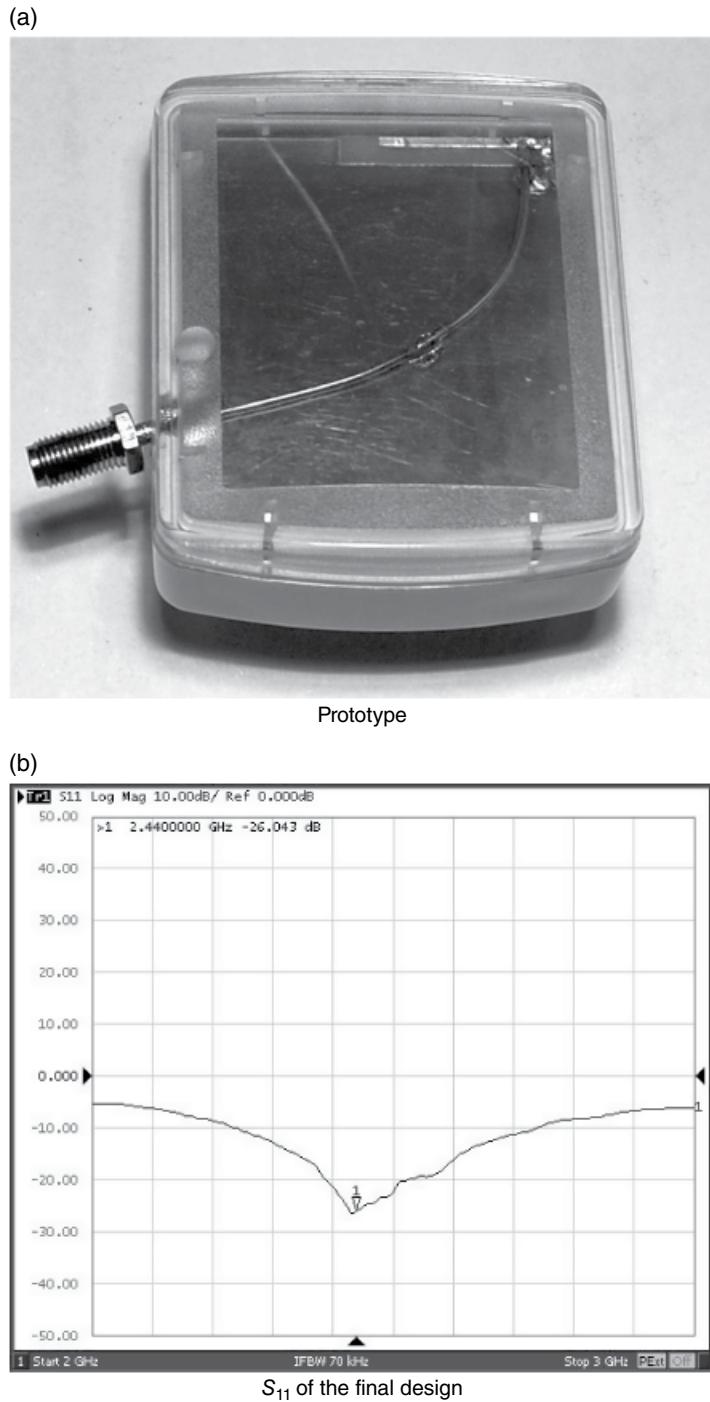


Figure 1.21 Impact of plastic case.

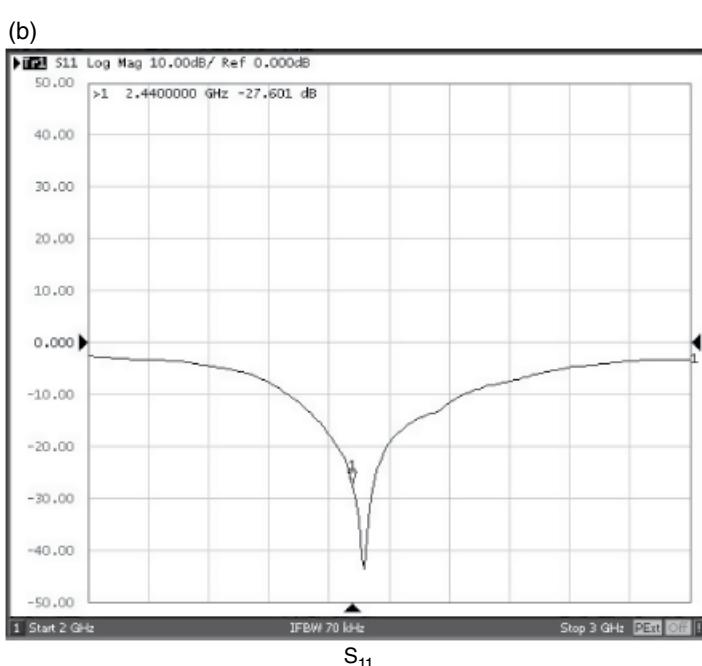
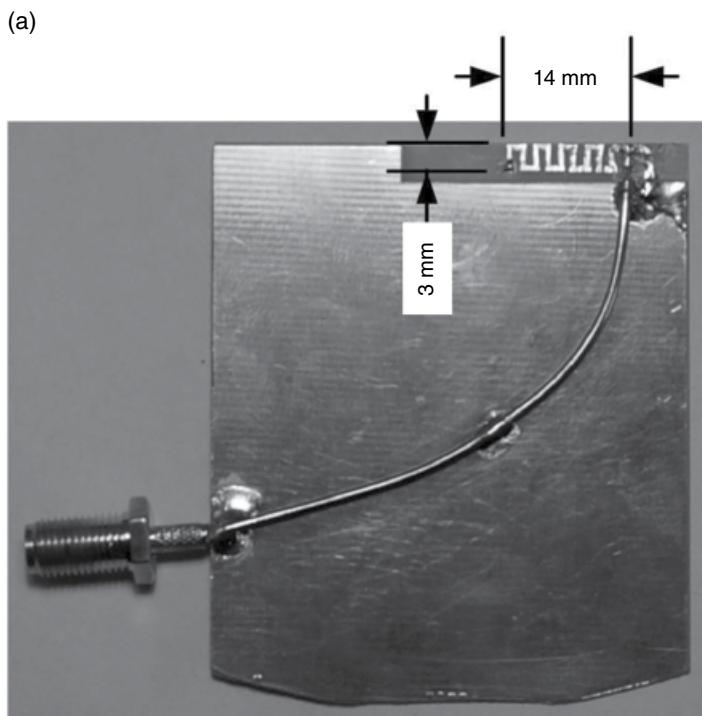


Figure 1.22 Making antenna smaller.

so tight that it can only cover 2.4–2.483 GHz, the yield will be an issue in mass production. When designing a product with very tight space constraints, ceramic antennas are better choices.

As the first practical example in the book, we will stop at here. I hope this example have relaxed you a little bit. After all, designing an antenna isn't as tough as everyone else said.

References

- [1] "Motorola DynaTAC," http://en.wikipedia.org/wiki/Motorola_DynaTAC. Retrieved 25 October 2015.
- [2] Bailey, A.B. (1939) Antenna system. US2184729.
- [3] Balanis, C.A. (2005) *Antenna Theory: Analysis and Design*, 3rd edn, Wiley-Interscience.
- [4] Caballero, R. and Schlub, R. (2013) Electronic device with proximity-based radio power control. US8417296.
- [5] Amm, D., Schlub, R., Leugh, Q. et al. (2015) Electronic devices with capacitive proximity sensors for proximity-based radio-frequency power control. US8947305.
- [6] "The Evolution of Cell Phone Design Between 1983–2009," <http://www.webdesignerdepot.com/2009/05/the-evolution-of-cell-phone-design-between-1983-2009/>. Retrieved 25 October 2015.
- [7] "Phonescoop," <http://www.phonescoop.com/>. Retrieved 25 October 2015.
- [8] "FCC Authorization Search," <https://apps.fcc.gov/oetcf/eas/reports/GenericSearch.cfm>. Retrieved 25 October 2015.
- [9] Iskander, M.F. (2000) *Electromagnetic Fields and Waves*, 1st edn, Waveland Press Inc.
- [10] William Hayt, J.B. (2005) *Engineering Electromagnetics*, 7th edn, McGraw-Hill Science/Engineering/Math.
- [11] Sadiku, M.O. (2009) *Elements of Electromagnetics*, 5th edn, Oxford University Press, USA.
- [12] Ulaby, F.T., Michielssen, E., and Ravaioli, U. (2010) *Fundamentals of Applied Electromagnetics*, 6th edn, Prentice Hall.
- [13] Chu, L.J. (1948) "Physical limitations of omnidirectional antennas," *Journal of Applied Physiology*, **19**, 1163–1175.
- [14] Hansen, R.C. (1981) "Fundamental limitations in antennas." *Proceedings of the IEEE*, **69**, 170–182.
- [15] Wheeler, H.A. (1947) "Fundamental limitations of small antennas," *Proceedings of the IRE*, **35**, 1479–1484.
- [16] Geyi, W. (2003) "Physical limitations of antenna," *IEEE Transactions on Antennas and Propagation*, **51**, 2116–2123.
- [17] Fante, R. (1969) "Quality factor of general ideal antennas," *IEEE Transactions on Antennas and Propagation*, **17**, 151–155.
- [18] McLean, J.S. (1996) "A re-examination of the fundamental limits on the radiation Q of electrically small antennas," *IEEE Transactions on Antennas and Propagation*, **44**, 672.
- [19] Ziolkowski, R.W. and Erentok, A. (2007) "At and below the Chu limit: passive and active broad bandwidth metamaterial-based electrically small antennas," *IET Microwaves, Antennas and Propagation*, **1**, 116–128.
- [20] Best, S.R. (2009) "A low Q electrically small magnetic (TE mode) dipole," *IEEE Antennas and Wireless Propagation Letters*, **8**, 572–575.
- [21] Peng, J. and Ziolkowski, R.W. (2009) "Low-Q, electrically small, efficient near-field resonant parasitic antennas," *IEEE Transactions on Antennas and Propagation*, **57**, 2548–2563.
- [22] Ziolkowski, R.W. and Erentok, A. (2006) "Metamaterial-based efficient electrically small antennas," *IEEE Transactions on Antennas and Propagation*, **54**, 2113–2130.
- [23] "Frequency Allocation," http://en.wikipedia.org/wiki/Frequency_allocation. Retrieved 25 October 2010.
- [24] "FCC Online Table of Frequency Allocations, 47 C.F.R. §2.106," (2010) <http://www.fcc.gov/oet/spectrum/table/fccitable.pdf>. Retrieved 25 October 2010.
- [25] "United States Frequency Allocations, the Radio Spectrum," (2003) <http://www.ntia.doc.gov/osmhome/allochrt.pdf>. Retrieved 25 October 2010.
- [26] Godara, L.C. (2001) *Handbook of Antennas in Wireless Communications*, 1st edn, CRC Press.
- [27] "CDMA Development Group," <http://www.cdg.org/>. Retrieved 25 October 2010.
- [28] "Home of the GSM Association," <http://www.gsmworld.com/>. Retrieved 25 October 2010.
- [29] "Global Positioning System," <http://www.gps.gov/>. Retrieved 25 October 2010.
- [30] "UMTS Forum," <http://www.umts-forum.org/>. Retrieved 25 October 2010.
- [31] "Wikipedia, the Free Encyclopedia," [http://en.wikipedia.org/wiki/LTE_\(telecommunication\)](http://en.wikipedia.org/wiki/LTE_(telecommunication)). Retrieved 25 October 2015.
- [32] "Wi-Fi Alliance," <http://www.wi-fi.org/>. Retrieved 25 October 2010.
- [33] "Bluetooth Special Interest Group," <https://www.bluetooth.org/>. Retrieved 25 October 2010.
- [34] "Wikipedia, the Free Encyclopedia," <http://en.wikipedia.org/>. Retrieved 25 October 2010.
- [35] <http://www.digikey.com>. Retrieved 25 October 2015.

2

Antenna Matching

The most important thing one must always remember when designing a matching circuit for antenna is that only inductors or capacitors can be used. In active microwave circuit designs, such as amplifier designs, resistors are frequently used to improve the port matching and circuit stability [1]. But in antenna designs, one should NEVER use resistors or other lossy components. The whole purpose of an antenna in a cellular device is to transmit or receive power, so the efficiency of the antenna is the most critical parameter. Whenever a resistor is added to a matching circuit, the efficiency always drops and that is not our goal. This mistake is repeatedly made by amateur antenna designers, so it is better to set the record straight before we start this chapter.

In the chapter, familiarity with basic electromagnetic (EM) concepts, such as characteristic impedance, return loss, reflection coefficient, voltage standing wave ratio (VSWR), and transmission line is assumed. If not, then the following textbooks are recommended before continuing [2–5]. An understanding of the Smith chart, which is the essential tool used in any antenna matching, is also recommended.

For those who want to have an in-depth report of matching techniques used in microwave circuit matching, the book by Professor Gonzalez, *Microwave Transistor Amplifiers: Analysis and Design*, [1] is a good reference. A paper by Professor Cripps [6] also provides some good discussions on the matching issue from a different point of view.

Matching is a useful technique in antenna designs. It gives engineers some more design freedom. But one has to remember that matching is not the silver bullet to solve all antenna design problems. Matching circuits are always associated with some degree of loss, which is due to the limited quality factor inherent in all components, no matter whether they are inductors, capacitors, or distributed networks. If an antenna has a reasonable initial resonance, the improvement obtained from the matching circuit will compensate for the loss it introduces. But if the antenna is not well designed, good performance will never be achieved by any matching circuit.

Free software, ZJ_Antenna_Matching, can be found on the book's companion website. Detailed instructions of how to use this software can be found in the Appendix. The data files of all examples of the chapter can also be found on the companion website. To really master the matching techniques, it is strongly recommended that these examples are attempted.

2.1 The Smith Chart

The essential aim of matching is to convert the original antenna impedance to a new one, which is as close to the system's characteristic impedance as possible. In the case of mobile phone, the target impedance is normally 50Ω . The 50Ω is not a magic value existing exclusively in the world; there are other standards, for example, 75 and 300Ω , which are widely used in the broadcasting industry. The 50Ω is only an industry standard for cellular antenna business. Most radio frequency (RF) equipment used by cellular phone companies and cellular phone antenna vendors use 50Ω ports as the standard interface. The 50Ω is chosen as the default characteristic impedance in the book. Nevertheless, the matching techniques discussed here can be used in antenna designs of any characteristic impedance.

To measure the effectiveness of a matching circuit, one needs a quantitative value. The voltage reflection coefficient Γ can serve this purpose. The voltage reflection coefficient Γ is the ratio between the reflected wave and the incidence wave. Γ is defined in Equation 2.1.

$$\Gamma = \frac{V_{\text{reflected}}}{V_{\text{incident}}} = |\Gamma| \angle \theta_\Gamma \quad (2.1)$$

Here, $V_{\text{reflected}}$ and V_{incident} —both of which are complex values and can be represented by amplitude and phase—are voltages of reflected wave and incident wave, respectively. The voltage reflection coefficient Γ is also a complex value. To evaluate the matching of antennas, only the amplitude of Γ , which is denoted as $|\Gamma|$, matters. θ_Γ is the phase angle of Γ . When a signal is fed to a load, in our case the load is an antenna, if the load has the same characteristic impedance as the transmission line, this load is a matched one and only the incidence wave travels from the source to the load on the transmission line, and the minimum reflection coefficient $|\Gamma|$, which is 0, is achieved. When the load is open or short circuit, all of the incidence wave is reflected back and travels from the load to the source, and the reflection coefficient $|\Gamma|$ reaches the maximum value 1.

Figure 2.1 shows the complex plane of Γ . These are a set of concentric circles; the amplitudes of any Γ located in the same circle are all equal; thus, they have the same level of reflection and can achieve the same level of matching. The worst matching happens when Γ falls in the outmost circle, where $|\Gamma|=1$, and it is also the boundary of Γ plane of any passive system. The origin point of the coordinate system represents the best matching, where $|\Gamma|=0$ and there is no reflection at all.

Γ can also be defined as Equation 2.2. The detailed derivation of Equation 2.2 is omitted in the book. For more information on the derivation, refer to the classical textbooks [2–5].

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0} = \frac{Z_L/Z_0 - 1}{Z_L/Z_0 + 1} = \frac{z_L - 1}{z_L + 1} \quad (2.2)$$

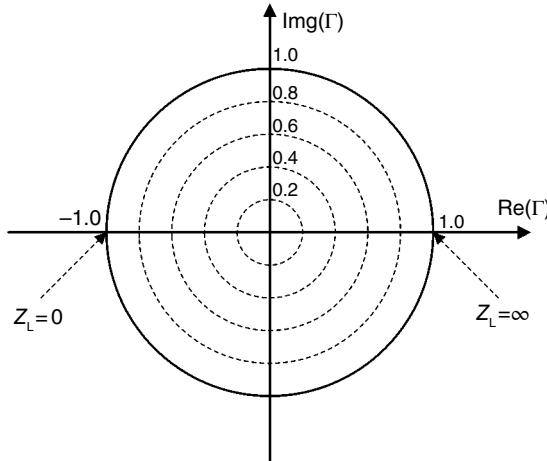


Figure 2.1 Complex plan of Γ .

Here, Z_L is the load impedance, Z_0 is the characteristic impedance of the source, and z_L is the normalized load impedance and is a complex value, which can be presented as a real part and an imaginary part given in Equation 2.3.

$$z_L = r_L + jx_L \quad (2.3)$$

Here, r_L and x_L are normalized load resistance and normalized load reactance, respectively.

As shown in Figure 2.1, when the load is a short circuit ($z_L=0$), the Γ takes the value of -1 , which is the leftmost point on the real axis. When the load is an open circuit ($z_L=\infty$), the Γ equals $+1$, which is the rightmost point on the real axis.

If we keep the r_L constant and change the x_L , the complex value Γ , which can be calculated by Equation 2.2, will generate a curve on the complex Γ plane. If a set of r_L values is used, a family of curves can be generated. Figure 2.2a shows four curves of constant r_L , which are $r_L=0, 0.5, 1$, and 2 , respectively. The $r_L=0$ curve is superimposed on the $|\Gamma|=1$ circle shown in the Figure 2.1, which represents situations when the load is lossless and formed by only the reactance component.

In Figure 2.2b as well as the $|\Gamma|=1$ circle, there are seven constant x_L line/curves, where $x_L=0, \pm 0.5, \pm 1$, and ± 2 , respectively. When $x_L=0$, z_L is always a real value, thus Γ is also a real value, and the corresponding constant x_L trace superimposes on the real axis in the Γ plane.

If we overlay Figure 2.2a and b, what we get is the famous Smith chart, which is shown in Figure 2.3. Compared with Figure 2.2, there are many more curves of constant r_L and x_L in Figure 2.3, which makes the Smith chart look much complicated than it really is.

The correct name for the Smith chart shown in Figure 2.3 is the impedance Smith chart, which represents curves by fixing either the real or the imaginary part of the normalized load impedance.

To get the admittance Smith chart, one needs to calculate curves by fixing either the real part g_L or the imaginary part b_L of the normalized load admittance.

$$y_L = \frac{1}{z_L} = g_L + jb_L \quad (2.4)$$

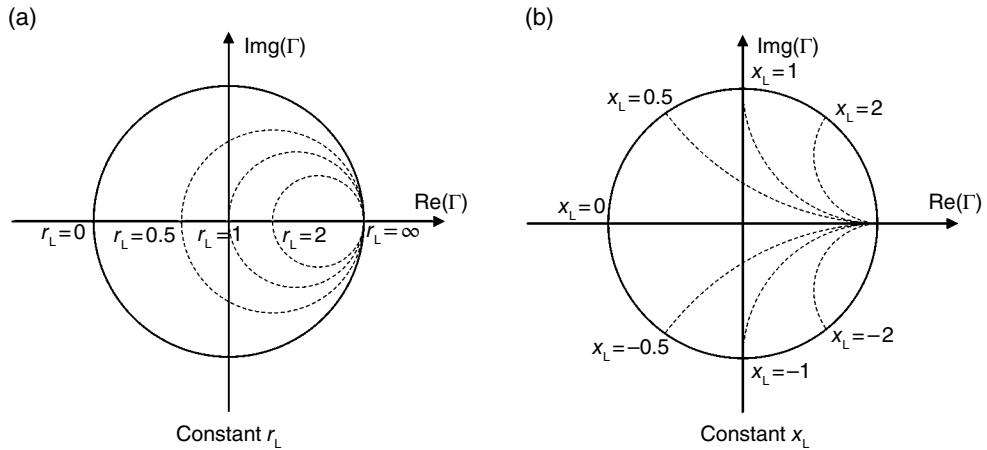


Figure 2.2 Family of curves of constant r_L and x_L .

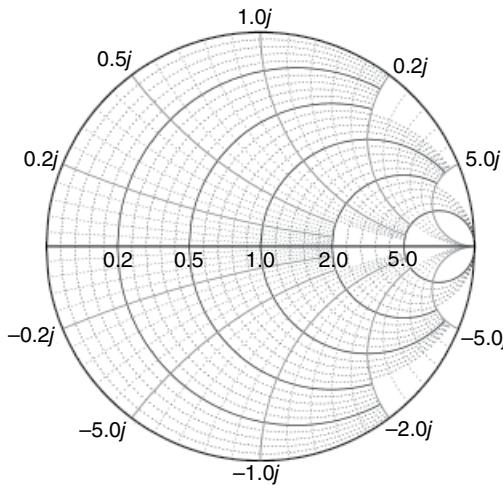


Figure 2.3 The Smith chart.

The curves of the admittance Smith chart can be obtained by simply rotating the impedance Smith chart by 180° .

Figure 2.4 shows the superposition of both the impedance and the admittance Smith charts; we will borrow the name from Gonzalez [1] and refer to these as the *ZY* Smith charts. A color version of the *ZY* Smith chart can be found on the companion website. The file is from ZY-01-N and is colored by J. Colvin [7].

For any complex load impedance, there is only one unique corresponding position in the Smith chart. Because both the impedance and the admittance Smith charts are drawn on the same complex Γ plane, the absolute position of a complex load in the Smith charts is always the same, no matter whether the load is represented by z_L or y_L .

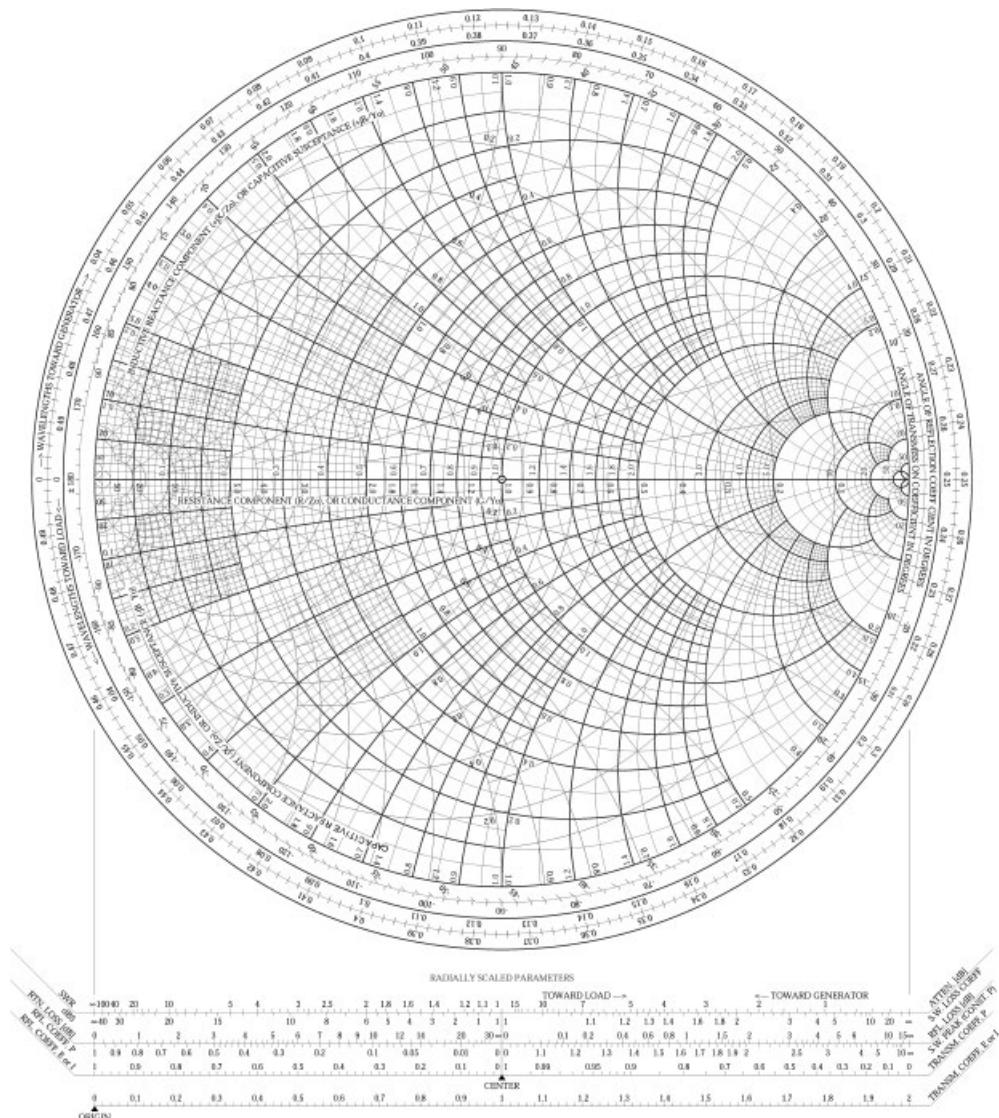


Figure 2.4 The normalized impedance and admittance Smith charts (ZY Smith chart). (Source: Reproduced from J. Colvin, “Color,” *Smith Chart Form ZY-01-N*, University of Florida, 1997.)

In some textbooks [3–5, 8], the admittance Smith chart is identical to the impedance Smith chart. In that case, whenever a conversion between impedance and admittance is needed, the impedance or admittance point must be rotated by 180° on the Γ plane. That method is technically correct; however, it is too cumbersome for hands-on engineering and, therefore, can be omitted.

As a reminder, whenever you see a Smith chart, always remember it is drawn on a complex Γ plane as shown in Figure 2.1. There is a set of hidden concentrical circles of equal $|\Gamma|$. The goal

of matching is to move the load impedance toward the center, where $|\Gamma|$ is 0 and is the perfect matching point.

In most test equipment or simulation software, you can only see the impedance Smith chart or the admittance Smith chart one at a time, so it is much more convenient if you can virtually superimpose a complementary Smith chart in your mind to the one displayed on the screen.

2.2 Single-Band Matching

When the Smith chart was invented in the 1930s, it was a design tool. RF engineers could design a circuit with the Smith chart and a ruler calculator. Of course, there are some hand calculations involved. Some might have done this as homework in universities to practice that skill. The most important part is to decide on the circuit. Calculating the component value by hand is no longer necessary because there is software to help you. There are several commercial software packages available, such as Agilent ADS [9], Microwave Office [10], and so on, which can be used to simulate and optimize a matching network. The companion website to the book includes matching software, ZJ_Antenna_Matching, which is written by the author and distributed as freeware. For most normal antenna matching tasks, ZJ_Antenna_Matching is good enough. It can read a data file saved as TOUCHSTN or in CITIFILE format. A mini-version of the software, which can be executed on Windows®-based network analyzers, is also included on the companion web site. Detailed instructions on how to use this software can be found in the Appendix.

2.2.1 Matching with Lumped Elements

As mentioned at the beginning of the chapter, when designing a matching circuit for antennas, only reactance components can be used. In the lumped-element-only matching method, that means only inductors and capacitors can be used. Thus, a matching component can only move the antenna impedance or admittance along the curves of equal r_L or g_L in the ZY Smith chart, respectively. In such situations, the ZY Smith chart can be simplified like the one shown in Figure 2.5, which has only two sets of circles instead of four sets of curves in a standard ZY Smith chart. All circles crossing the leftmost point are fixed g_L curves, and all circles crossing the rightmost point are equal r_L curves.

For any antenna impedance located on the ZY Smith chart, there are always two circles running through it, as shown in Figure 2.6. The right circle is the constant r_L circle and the left circle is the constant g_L circle. Starting from the antenna impedance, which is marked by a cross in Figure 2.6, there are four directions, a, b, c, and d. Each of them represents one kind of circuit topology as follows:

- When a series inductor is connected to the antenna, which is shown as Figure 2.7a, the combination impedance of the antenna and the series inductor at the output will move in direction a.
- A series capacitor, shown as Figure 2.7b, will move the impedance in direction b.
- A shunt inductor, shown as Figure 2.7c, will move the impedance in direction c.
- A shunt capacitor, shown as Figure 2.7d, will move the impedance in direction d.

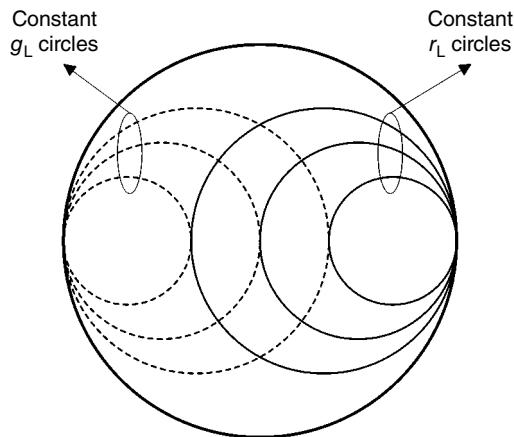


Figure 2.5 Simplified ZY Smith chart for lumped-element-only matching.

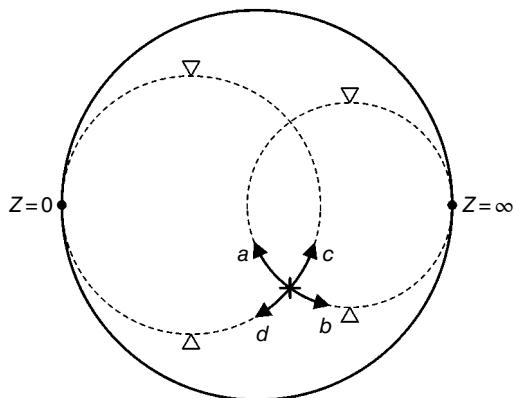
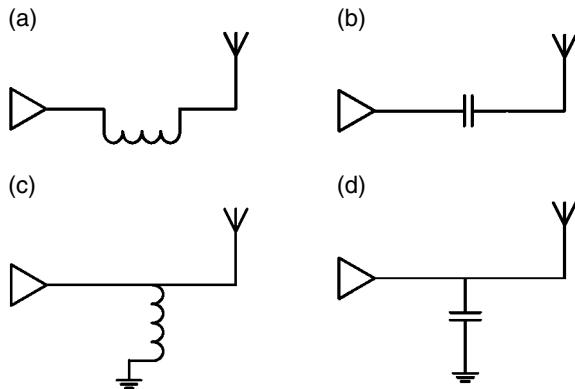


Figure 2.6 Always two circles running through a given impedance on the Smith chart.



(a) Series inductor (b) Series capacitor (c) Shunt inductor (d) Shunt capacitor

Figure 2.7 Four possible connecting methods of matching components.

To design antenna matching network frequently, it is strongly recommended that you MEMORIZE the four aforementioned scenarios. The whole practice of matching is using these four components to move antenna impedances around on the Smith chart. Actually, the memorization process should not be difficult. In most cases when doing antenna matching, you only need to remember two rules. The first rule is that the upper half plane is inductive and the bottom half plane is capacitive; so whenever the impedance needs to be moved up, an inductor is needed, otherwise use a capacitor. The second rule is that the left circle is the shunt circle and the right circle is the series circle, so if the impedance needs to be moved along the left circuit, a shunt component is needed, otherwise use a series component. As an example, assume we want to move the impedance from the cross marker to the “b” direction. Because it is down, a capacitor is required. Because it is on the right circle, a series component is needed. Combining both, a series capacitor is the correct choice.

Note the phrase “in most cases” is used in the previous paragraph. That means in some cases, the rules are not applied. There are four triangle markers in Figure 2.6, which mark the maximum and minimum points of both circles. After those extreme points, the moving direction of impedance, which is either up or down, will flip, but the component selection should not be changed. For instance, if we want to move the impedance toward direction “b” as shown in Figure 2.6, it does not matter whether the final position is above or below the original location, a series capacitor should always be used. Two points, where the impedance equals 0 or infinity, are the stopping points of impedance movement. For instance, a series capacitor can move an impedance along direction “b” but can never move it over the $Z=\infty$ point.

Theoretically, all capacitors and inductors are equal. But in reality, components have different specifications. It is an engineer’s responsibility to check whether a component is suitable for the matching purpose. On the data sheet of high frequency capacitors, the quality factor Q is normally specified. For antenna engineers, a higher Q is better. However, Q is a variable dependent on the measured frequency. The measured Q of a capacitor is lower when measured at a higher frequency. Different manufacturers might measure Q at different frequencies, so be careful when evaluating capacitors from different manufacturers. As a rule of thumb, the Q should be at least around 1000@1 MHz. There are quite a few manufacturers in the RF capacitor business; the brands the author has used include the Murata® GRM series [11] and the Johanson Technology® S series [12].

For high-frequency inductors, there are some more specification needs to be attended to. They are the quality factor Q , the DC resistance, and the self-resonance frequency (SRF). The selection of Q is similar to that of a capacitor; the higher, the better. However, the relation between the Q of an inductor and the measured frequency is reversed. The higher the measured frequency, the higher the Q . With regard to the DC resistance, of course, lower is better. The SRF is the resonance frequency of an inductor. Because in reality, there is always parasitic capacitance in any inductor, the inductance and parasitic capacitance can form a shunt resonator. The SRF tells us at which frequency the inductor starts to resonate. After SRF, the inductor will no longer function as an inductor. When selecting an inductor, the SRF needs to be safely higher than the working frequency. The highest grade inductor is the wire coil chip inductor. The SRF of wire coil inductors is at least twice as high as their film counterpart, and their DC resistance is less than half that of film inductors. The brands of coil inductors the author has used include the Murata LQW series [13] and the Coilcraft® 0402HP series [14]. Coil inductors are the most expensive inductors of all kinds. For antenna applications, whose

performance is not that critical, a high Q film inductor can also be used. The author has used the Murata LQG15H series [13] and the Johanson Technology L series [15]. Coil inductors are suitable for the surface-mount technology (SMT) process, but it is quite difficult to handle manually, so be prepared.

Always remember to check the antenna response whenever the vendor of a component is changed, or the component is replaced by a different type. Although two components are marked with the same inductance or capacitance value, this does not necessarily mean they are identical. Basically, an efficiency measurement in a three-dimensional (3D) chamber is the ultimate evaluation method one can always count on.

2.2.2 Different Ways to Accomplish a Single-Band Matching

By combining the series inductor, the series capacitor, the shunt inductor, and the shunt capacitor, any value of the load impedance in the Smith chart can be matched except those spots located on the $|\Gamma|=1$ circle, where the impedance is purely reactive. The purpose of matching is to convert the real part of the load impedance to the source impedance Z_0 and eliminate the imaginary part. If there is no real part in a load impedance, which is the case of any spot on the $|\Gamma|=1$ circle, there is nothing to be converted from.

Figure 2.8 shows two circles in the Smith chart. The left circle is $g_L = 1$ and the right circle is $r_L = 1$. Any impedance falling in these two circles can be matched by a single component. A spot on the left or right circle can be matched by a shunt or a series component, respectively. Now, it is obvious that any matching task can be broken down into two steps: first, move the impedance to these two circles; second, use one single element to move the impedance to the center of the Smith chart.

At any spot inside $|\Gamma|=1$ circle on the Smith chart, there are up to four ways to achieve matching. For example, if an antenna with impedance Z_L , as shown in Figure 2.8, needs to be matched, four sets of matching circuits, which are illustrated in Figure 2.9, can be used.

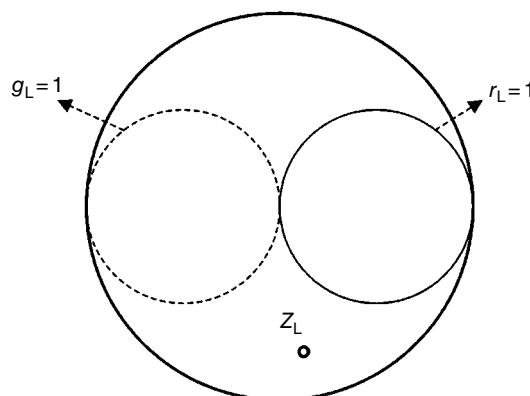


Figure 2.8 Two circles which can be matched by single components.

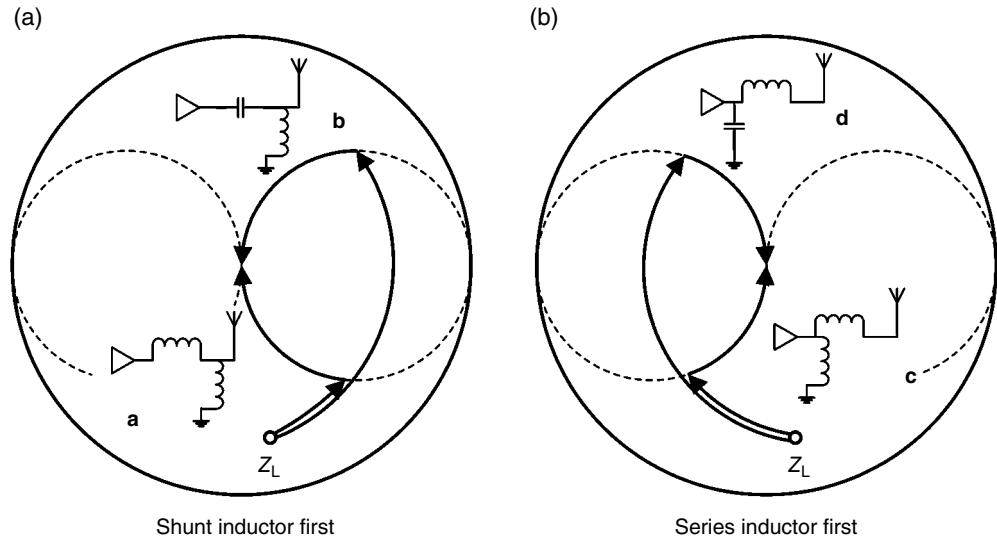


Figure 2.9 Four matching options of single load impedance.

- A shunt inductor cascade by a series inductor, as shown in circuit (a) in Figure 2.9a.
 - A shunt inductor cascade by a series capacitor, as shown in circuit (b) in Figure 2.9a. Compared to circuit (a), both circuits start with the shunt inductor and move the antenna load toward the $r=1$ circle, but the value of the shunt inductor used by circuit (a) and (b) is different. The circuit (a) moves the impedance to hit the bottom part of the $r=1$ circle. The inductor of circuit (b) has a smaller value than circuit (a), thus moving the impedance further up, bypassing the spot circuit (a) and hitting the top part of $r=1$ circle. When used as a shunt component, an inductor with a smaller value has a larger admittance, and thus moves the load further. On the other hand, a capacitor with a smaller value moves the load less.
 - A series inductor cascade by a shunt inductor, as shown in circuit (c) in Figure 2.9b.
 - A series inductor cascade by a shunt capacitor, as shown in circuit (d) in Figure 2.9b. Similar to the cases shown in Figure 2.9a, two inductors with different values are used to move the load to two different spots on the $g=1$ circle. When used as a series component, an inductor with a larger value has a larger impedance, and thus moves the load further. On the other hand, a capacitor with a larger value moves the load less.

The four possible matching schemes give an engineer quite some leeway in actual designs. When the metal surface of an antenna element is exposed to the outside environment, the electrostatic discharge (ESD) is a potential hazard to the internal circuit. In this kind of circumstance, both matching circuits (a) and (b) can be used to eliminate the danger, because there is a shunt inductor which electrically shorts the antenna element to the ground from the DC point of view. As a reminder, the effectiveness of using an inductor as the grounding depends on the inductance value. If the inductance is too large, say, above 10nH , the inductor itself is not enough to take care of the ESD issue; thus, only circuit (b), which has a series capacitor functioning as a DC blocker to provide the extra measure, is a feasible solution.

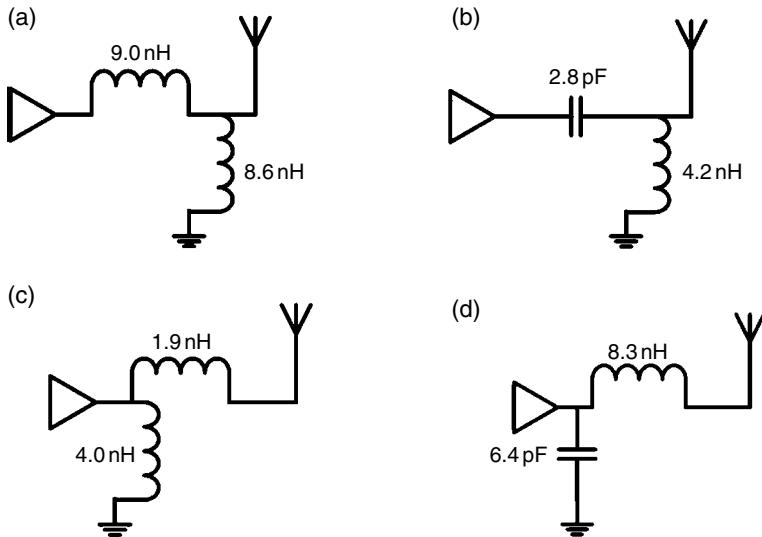


Figure 2.10 Component values of the four circuits used in Figure 2.9.

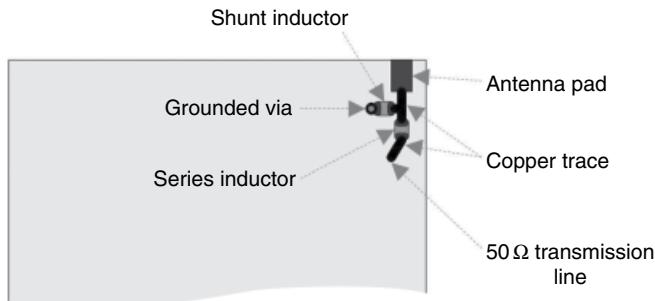


Figure 2.11 Layout of Figure 2.10a on a printed circuit board.

To provide a more intuitive feeling for this event, the component values of circuits shown in Figure 2.9 are illustrated in Figure 2.10. The normalized load impedance is chosen as $0.2 - j0.64$ [Simulation file: Chap2_Fig.10.s1p], and the working frequency is chosen as 1.0 GHz. When using the ZJ_Antenna_Matching as practice in this example, the highlight feature should be used. The impedance in this example is a point impedance. Without the highlight, the impedance dot is too small to be easily distinguished.

Figure 2.11 illustrates the actual layout of Figure 2.10a on a printed circuit board (PCB). The PCB is a double-sided board. On the back of the PCB, there is a whole piece of copper layer functioning as the ground. The signal trace and components are placed on the front of the PCB. The output of the matching circuit is a 50Ω transmission line. When doing a return loss test, this 50Ω transmission line can be connected to a network analyzer by a coaxial cable. The pad on the top-right corner is the launch point of the antenna. The component on

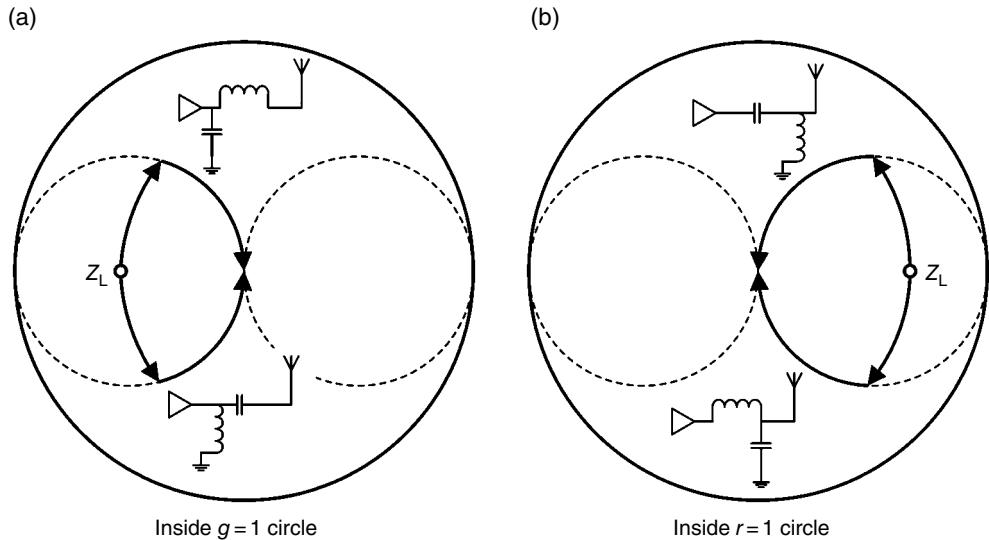


Figure 2.12 Two matching options of single load impedance.

the left side of the antenna pad is the shunt inductor L_1 . A shunt component is always bridged between a transmission line and the ground. The component on the bottom-right side of L_1 is the series inductor L_2 . A series component is always inserted in the transmission line. When laying out a series component, the transmission line is broken into two segments by inserting a gap and then the series component is used to bridge the gap.

When the load impedance is inside either circles $r=1$ or $g=1$, as shown in Figure 2.12a or b, the possible options of matching circuits will drop from four to two. Let's use Figure 2.12a as an example. Because the load is inside the $g=1$ circle, there is no component that can move the load impedance to the $r=1$ circle. Two possible matching circuits, a series inductor cascade by a shunt capacitor or a series capacitor cascade by a shunt inductor, are shown in Figure 2.12a.

For the load impedance shown in Figure 2.12b, these two matching circuits are a shunt capacitor cascade by a series inductor or a shunt inductor cascade by a series capacitor.

2.2.3 Matching with Both Transmission Line and Lumped Elements

Besides using lumped elements, transmission lines can also move the load impedance around on the Smith chart. As shown in Figure 2.13, an impedance Z_L is connected by a transmission line with a characteristic impedance of Z_c and a length of L .

The output impedance Z_{in} can be expressed by Equation 2.5 [1–5, 8].

$$Z_{in} = Z_c \frac{Z_L + jZ_c \operatorname{tg}\left(\frac{2\pi}{\lambda} L\right)}{Z_c + jZ_L \operatorname{tg}\left(\frac{2\pi}{\lambda} L\right)} \quad (2.5)$$

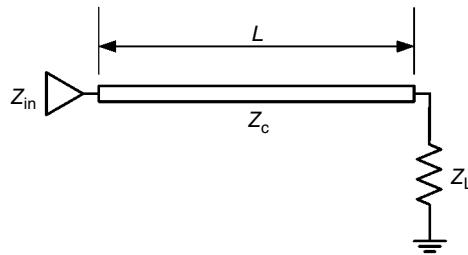


Figure 2.13 Load impedance connected by a segment of transmission line.

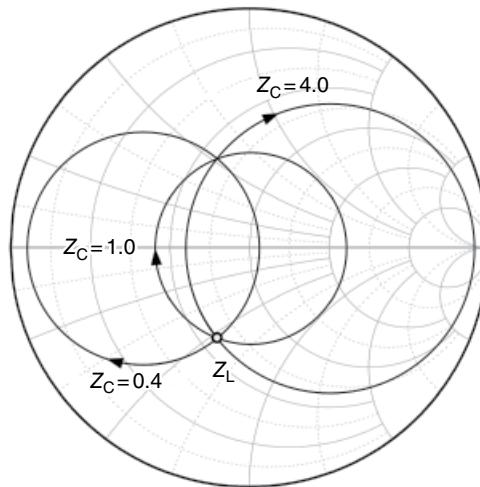


Figure 2.14 The impedance's moving path on the transmission lines.

Here, λ is a guided wavelength in the transmission line. In Equation 2.5, when $L=0$, the $Z_{in} \equiv Z_L$, which means the moving path of a transmission line always starts from the Z_L on the Smith chart. When $L=\lambda/2$, we get $Z_{in} \equiv Z_L$ again, which means the output impedance moves back to the original position when the length of the transmission line is half the guided wavelength.

Figure 2.14 shows the moving paths of three transmission lines with different Z_c . In this example, the load impedance is $Z_L = 0.5 - j0.8$, which is the lower crossing point of all three circles on the Smith chart. The leftmost circle is the moving path of a transmission line with normalized impedance $Z_c = 0.4$. The circle in the center corresponds to a transmission line of $Z_c = 1$, and the rightmost circle corresponds to $Z_c = 4$. It can be seen that, by increasing the characteristic impedance of the transmission line, the circle of the moving path shifts from left to right on the Smith chart. The center of the $Z_c = 1$ circle always is superposed on the center of the Smith chart.

A transmission line always moves the load impedance in a clockwise direction on a circle. Thus, to move the impedance to a nearby position on the circle in the anticlockwise direction, nearly half a wavelength transmission line has to be used.

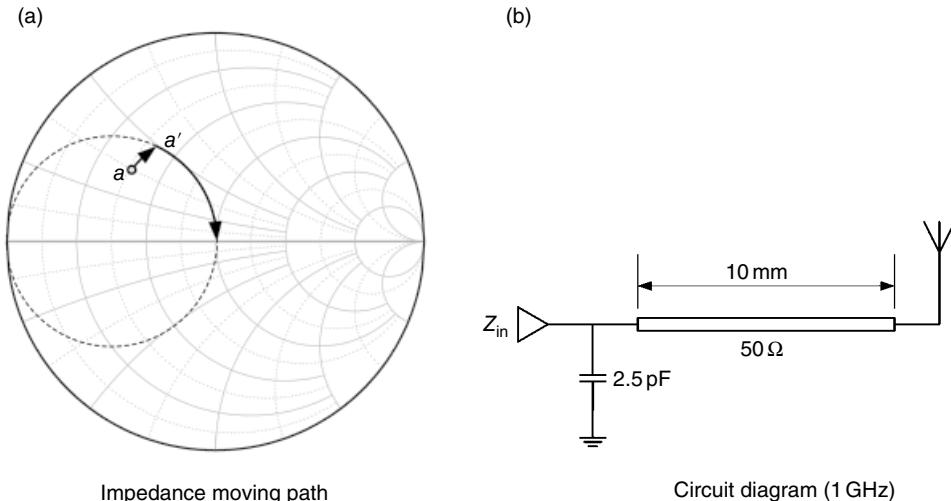


Figure 2.15 Matching with both transmission line and lumped element.

For any antenna impedance inside or on the $r=1$ or $g=1$ circles, a perfect matching can be achieved by a single segment of the transmission line. For any impedance outside those two circles, it is impossible to obtain a perfect match in this way. However, a match can be achieved by a combination of transmission lines or a combination of transmission line and lumped elements. The book only discusses the latter option. More information on matching with only transmission lines can be found in references [3, 4, 8].

As an example, an impedance Z_a , as shown in Figure 2.15a, needs to be matched. The working frequency is 1 GHz. The design procedure can be divided into two steps. First, move the Z_a in a clockwise direction to location a' on the $g=1$ circle by a segment of the transmission line. Next, use a shunt capacitor to move the impedance along the $g=1$ circle to the origin of the Smith chart. The corresponding matching circuit diagram is shown in Figure 2.15b. To match an impedance of $0.3+j0.3$, a 10 mm-long 50Ω transmission line and a 2.5 pF shunt capacitor are needed. As demonstrated in Figure 2.14, to move the impedance in the clockwise direction, the characteristic impedance of the transmission line does not have to be 50Ω . Any transmission line will do. However, the value of the capacitor must be changed accordingly.

Similarly, impedances (b), (c), and (d), shown in Figure 2.16a, can also be matched by a transmission line and a lumped element. The corresponding matching circuits of different loads are shown in Figure 2.16b-d. In the case of load (b), the lumped matching element is a series capacitor. For the case of loads (c) and (d), the lumped matching elements are a series inductor and a shunt inductor, respectively.

In fact, impedances (a), (b), (c), and (d), shown in Figures 2.15 and 2.16, can also be matched by the lumped-element-only matching techniques. Depending on the location of the load impedance on the Smith chart, the transmission line can be replaced by different lumped elements. For the impedances (a) and (d), the transmission line can be replaced by a series inductor. In the case of the impedances (b) and (c), the replacing lumped element is a shunt capacitor.

Although the transmission line can move any load impedance to the $r=1$ circle or $g=1$ circle; thus, in theory, the matching technique introduced in this section can be used in all

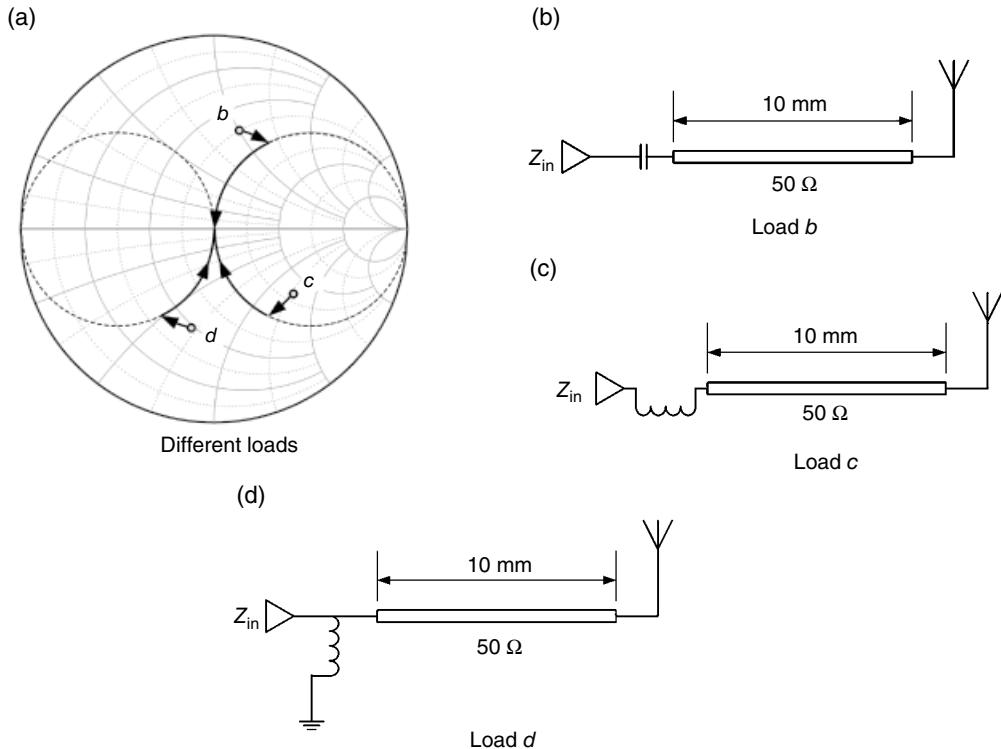


Figure 2.16 Matching circuits of load impedance *b*, *c*, and *d* (1 GHz).

situations. In practice, the transmission line is only used in circumstances when the load impedance is close to the $r=1$ or $g=1$ circle and is on the anticlockwise side of either circle. Unlike lumped element, the transmission line occupies the PCB space. The further it needs to move a load impedance, the longer the transmission line is, thus the larger PCB area it occupies. In addition, the PCB also has inherent loss; a long transmission line degrades the efficiency of an antenna even though it matches the antenna impedance.

2.2.4 Bandwidth Consideration

In the earlier discussion, the antenna impedance is treated as a constant. However, in reality, the antenna impedance varies with frequency. It is always a curve on the Smith chart. Any matching network can only move a limited portion of the impedance curve to the target matching circle on the Smith chart, which means there is a bandwidth limit for any matching network. When using the matching techniques described earlier, the achievable bandwidth does not vary too much, no matter which match circuit is used.

In this section, the emphasis is on various techniques which can expand the antenna bandwidth. Using matching to increase the bandwidth can also be treated as a wide band impedance matching problem. Professor Cripps' article [6] is a very good reference on this topic, and it provides a short list of references to investigate this area further. The section

focuses on techniques frequently used in antenna matching designs. Those complex techniques and theoretical analyses are outside the scope of the book. In the section, three examples are discussed. Example 2.1 antenna has a decent return loss without any matching, so the matching circuit is used to mainly widen the bandwidth. Example 2.2 describes how to achieve both impedance matching and bandwidth enhancement by a π -shaped matching circuit. The techniques discussed in Section 2.2.2 are also applied to Example 2.2 to provide a straightforward comparison between different matching methods. Example 2.3 briefly demonstrates how to use the T-shaped network to achieve matching and bandwidth improvement.

Example 2.1

Let's look at Example 2.1 [Simulation file: Chap2_example1.s1p]. Assuming there is an antenna with decent self-resonance at 1.35 GHz, as shown in Figure 2.17, Figure 2.17a and b are the reflection coefficient and the impedance curve of the antenna. The antenna impedance at a frequency lower or higher than 1.35 GHz is capacitive or inductive, respectively. The solid line curve in Figure 2.17b is complex impedance and the dash line circle is the S_{11} equal to -15 dB circle. To increase the bandwidth, an intuitive thought is to squeeze more impedance curve into the -15 dB circle.

To squeeze the impedance line of higher frequency, a shunt capacitor is needed. However, to squeeze the impedance line at a lower frequency band, a shunt inductor is needed. It seems a contradictory requirement at first glance. Fortunately, a shunt LC resonator has the preferred characteristic. The resonant frequency of a circuit shown in Figure 2.18a is decided by

$$\text{Freq} = \frac{1}{2\pi\sqrt{L \cdot C}} \quad (2.6)$$

where L and C are the inductance and the capacitance, respectively. By selecting the appropriate L and C , the resonant frequency can be allocated to 1.35 GHz; this provides the required reactance condition to bend the impedance curve into the -15 dB circle. From Equation 2.6, it

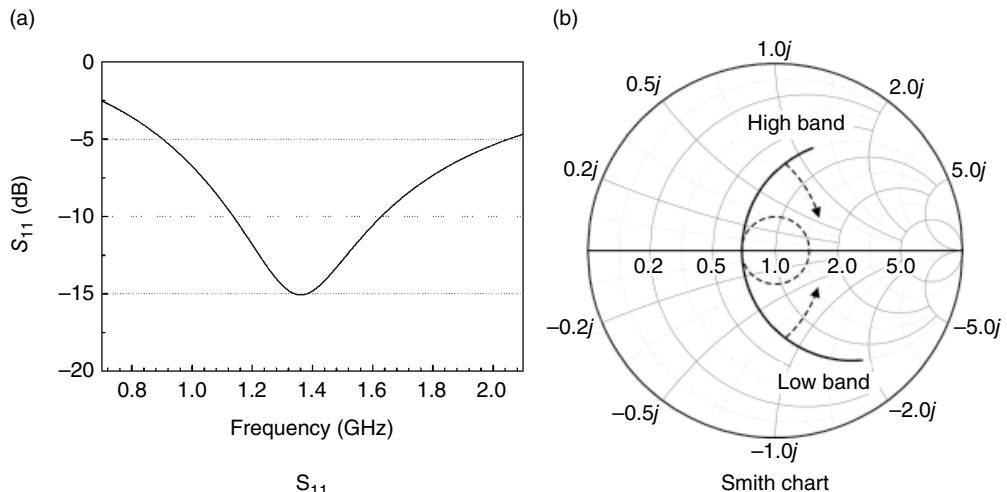


Figure 2.17 Antenna example—bandwidth widened.

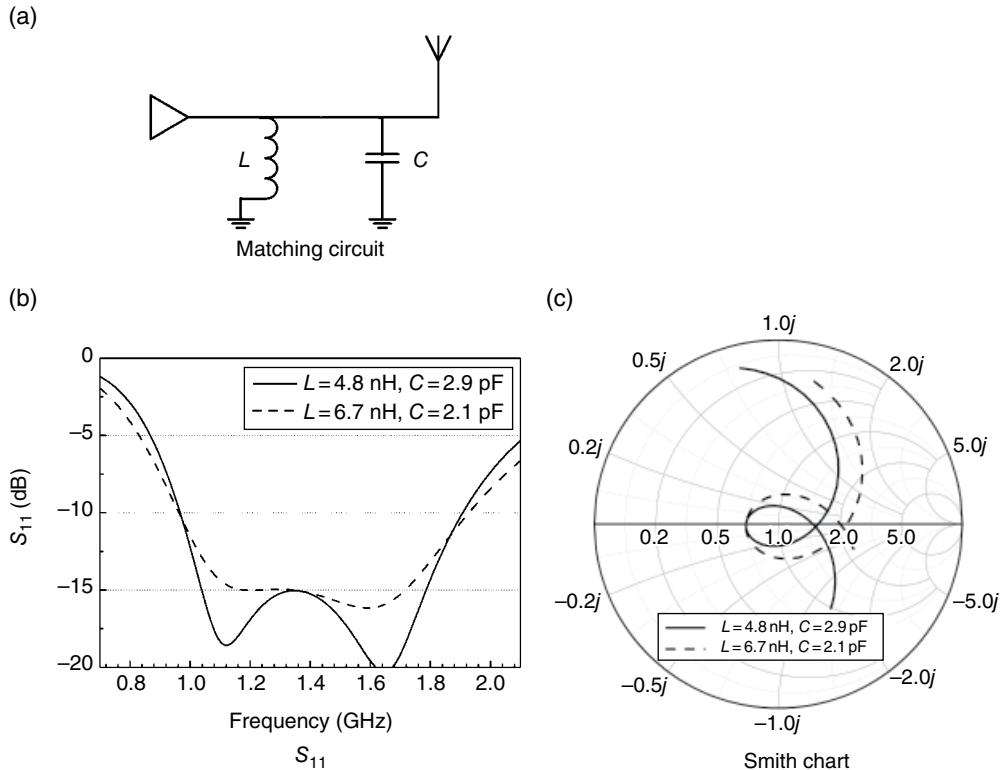


Figure 2.18 Shunt LC resonator as a matching circuit.

is clear that there are infinite combinations of L and C which generate the same resonant frequency. But different L and C combinations bend the impedance curve differently, thus the achieved bandwidth is also different. Figure 2.18b and c shows simulated results. In the figure, the solid line is the result of $L=4.8 \text{ nH}$ and $C=2.9 \text{ pF}$. The dashed line is the result of $L=6.7 \text{ nH}$ and $C=2.1 \text{ pF}$. In this example, the circuit of $L=4.8 \text{ nH}$ and $C=2.9 \text{ pF}$ gives the widest -15 dB bandwidth.

Example 2.2

Figure 2.19 shows the original S_{11} and complex impedance on the Smith chart of Example 2.2 [Simulation file: Chap2_example2.s1p]. The best matching is achieved at 1.35 GHz with a S_{11} of merely -3.5 dB . The specification of return loss used in the example is -10 dB .

For the purposes of comparison, a basic two-element circuit shown in Figure 2.20a is used to achieve the required matching. Looking from the antenna side, the matching circuit is composed of a 2.1 nH series inductor and a 3.7 pF shunt capacitor. As shown in Figure 2.20c, the 2.1 nH inductor moves the impedance at 1.4 GHz along the clockwise direction to hit the $g=1$ circle, then the 3.7 pF capacitor moves the impedance along the $g=1$ circle to the matching area. Figure 2.20b is the S_{11} of the matched antenna. The S_{11} is better than -16 dB at 1.4 GHz , and the -10 dB bandwidth is 70 MHz .

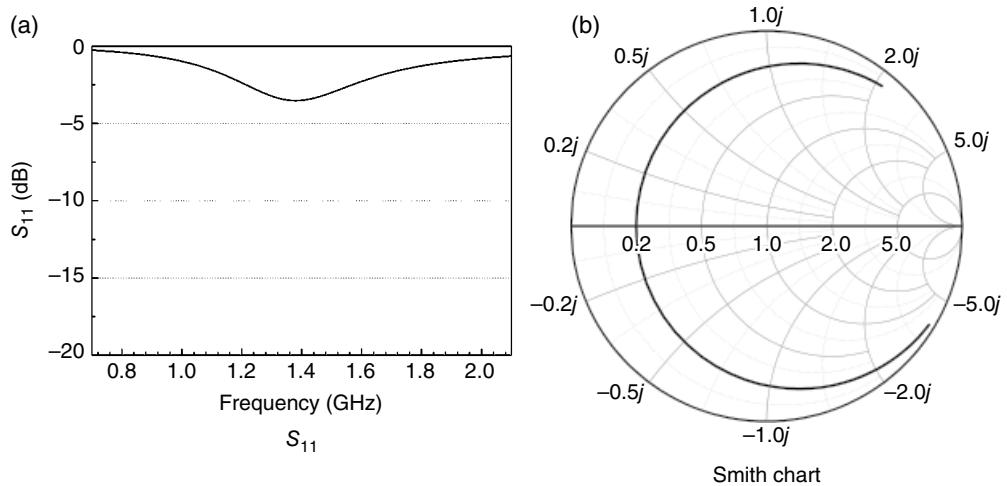


Figure 2.19 Original antenna impedance before matching.

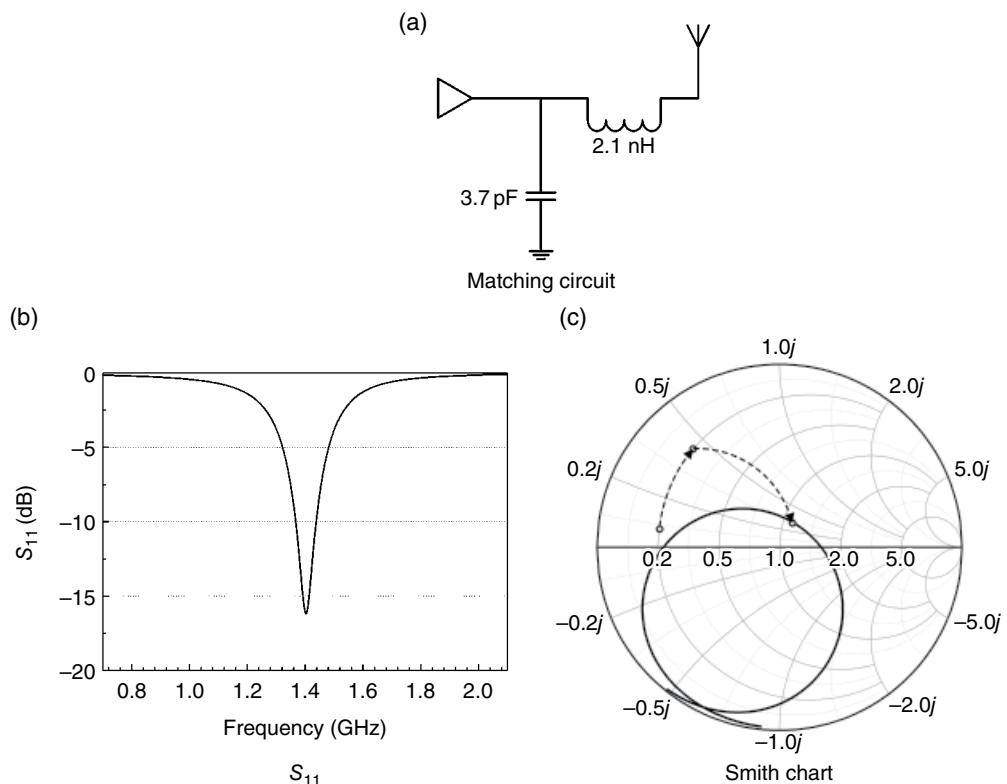


Figure 2.20 Two-element matching circuit.

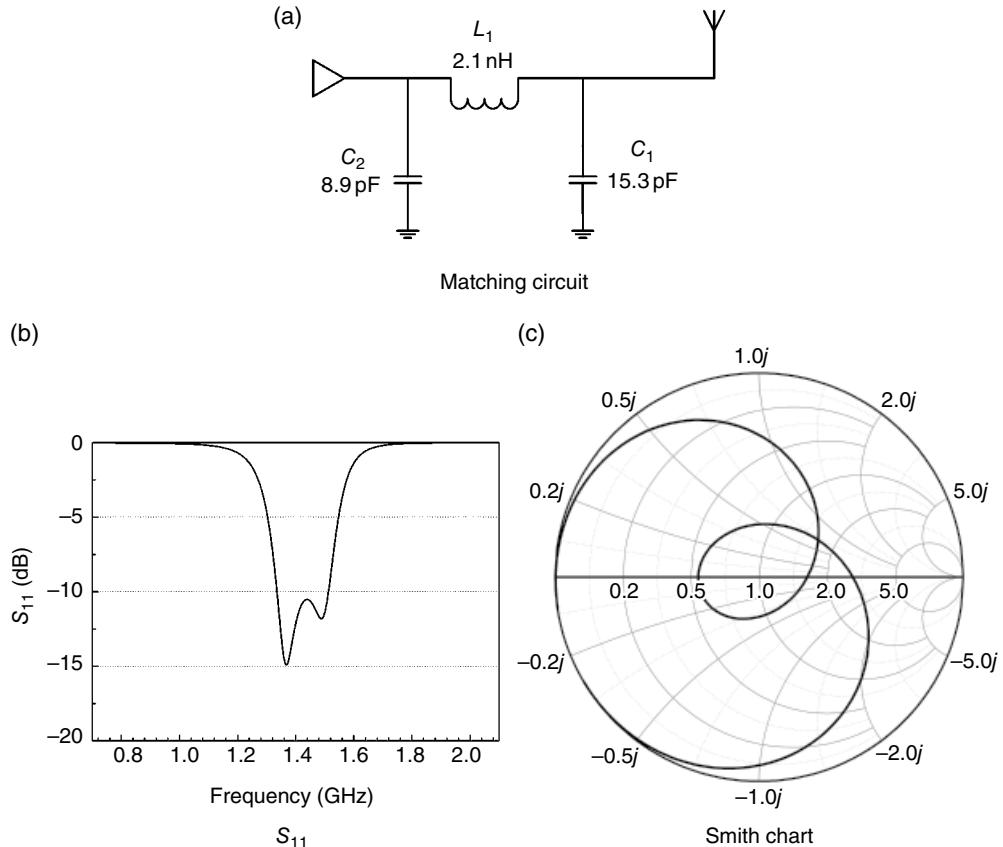


Figure 2.21 π -shaped matching circuit.

Figure 2.21a shows a π -shaped matching network, which achieves both antenna matching and bandwidth expanding. Looking from the antenna side, the matching network is composed of a 15.3 pF shunt capacitor, a 2.1 nH series inductor, and an 8.9 pF shunt capacitor. From Figure 2.21b, it is clear that the -10 dB bandwidth is 180 MHz , which is more than twice as much as the bandwidth achieved by the matching network shown in Figure 2.20. Comparing Figures 2.20c and 2.21c, it is clear that the impedance curve of the two-element matching simply passes by the origin point of Smith chart, meanwhile the impedance curve of the π -shaped matching circles around origin point, which generates the double resonance phenomena on S_{11} results shown in Figure 2.21b.

The working principle of π -shaped matching can be explained by mathematically deduced formulas, which can be found in various books on filter designs. A more intuitive explanation is given in this book. There are three components in the π -shaped matching circuit. Each component serves a distinct purpose. To separate the contribution of each matching component, components are added to the circuit one by one. Figure 2.22a shows the effect of the first capacitor if looking from the antenna side. The dashed line curve in Figure 2.22a is the original impedance curve without matching. The solid line curve is the impedance with the first

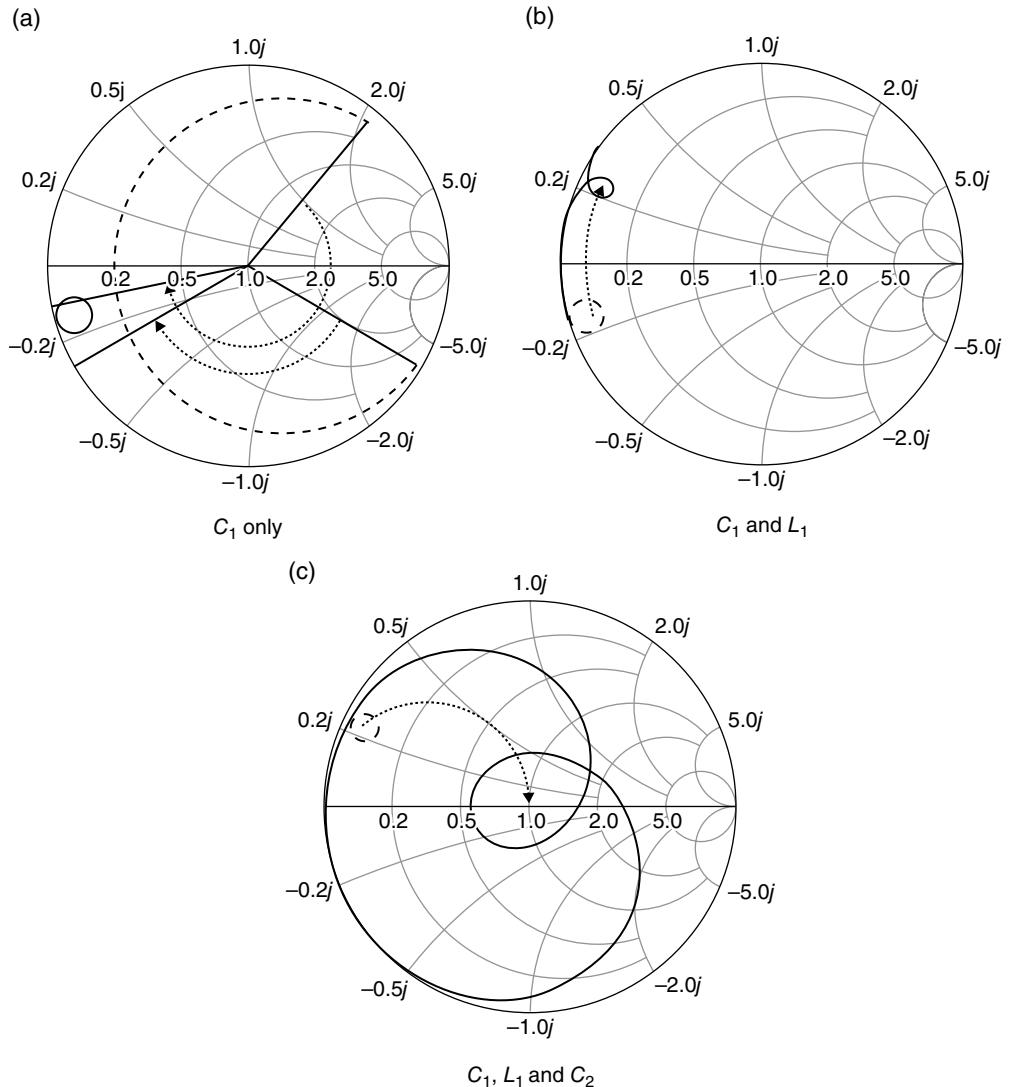


Figure 2.22 Contribution of each component.

capacitor added. Compared with the shunt capacitor used in Figure 2.20, which is 3.7 pF, the capacitance of the first capacitor is much larger, up to 15.3 pF. For a shunt capacitor, a higher value means it moves the impedance further on the Smith Chart. The other property of a shunt capacitor is that it moves the impedance at the higher frequency band faster than that at the lower band. In this example, the impedances at lower and the higher frequencies are moved around 110° and 200°, respectively. It is the first shunt capacitor which tightens up the impedance curve to generate a loop.

Figure 2.22b shows the impedance curve when both C_1 and L_1 are added. The effect of L_1 is quite clear, which is moving the impedance along in a clockwise direction to hit the $g=1$

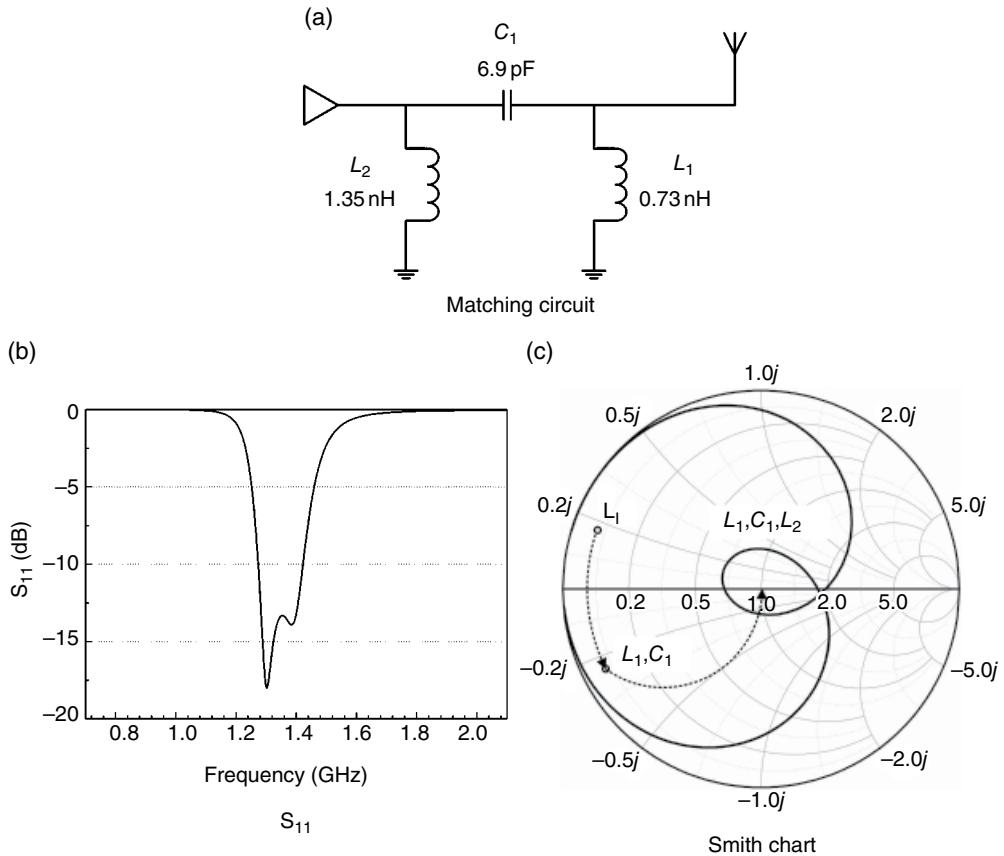


Figure 2.23 Another version of the π -shaped matching circuit.

circle. A series inductor has a similar frequency dependence property as a shunt capacitor, thus besides the impact of moving impedance, it also makes the loop tighter. Figure 2.22c shows the effect of the third matching component, the 8.9 pF shunt capacitor. It moves the impedance on the $g=1$ circle to the center of the Smith chart. Compared with the basic matching circuit shown in Figure 2.20, the C_2 capacitor needs to move the impedance further, thus its capacitance is also much larger.

Besides using a shunt capacitor as the first component to generate the required loop as shown in Figure 2.22a, a shunt inductor also has a similar effect. Figure 2.23a shows another version of the π -shaped matching circuit. Starting from the antenna side, the first matching component is a 0.73 nH shunt inductor, followed by a 6.9 pF series capacitor and a 1.35 nH shunt inductor. The 0.73 nH shunt inductor L_1 generates the required loop. The capacitor C_2 moves the impedance along the anticlockwise direction to hit the bottom part of $g=1$ circle. The L_2 moves the impedance on the $g=1$ circle to the center of the Smith chart. When an inductor is used as a shunt component, the smaller the value, the larger the impact. Compared with inductor values used in a basic two-element matching circuit, both inductors shown in Figure 2.23 have quite small values. That means both of them move the impedance further.

The -10 dB bandwidth obtained by Figure 2.23 is 180 MHz, which is same as the one achieved by the matching circuit shown in Figure 2.21.

Example 2.3

For an antenna with an impedance similar to the one shown in Figure 2.19, only a π -shaped matching circuit can serve the purpose of expanding the antenna bandwidth. In Example 2.3, the original antenna impedance is shown in Figure 2.24a [Simulation file: Chap2_example3.s1p]. For this type of antenna impedance, a T-shaped matching network shown in Figure 2.24b is the appropriate choice. Starting from the antenna side, the first matching component is a

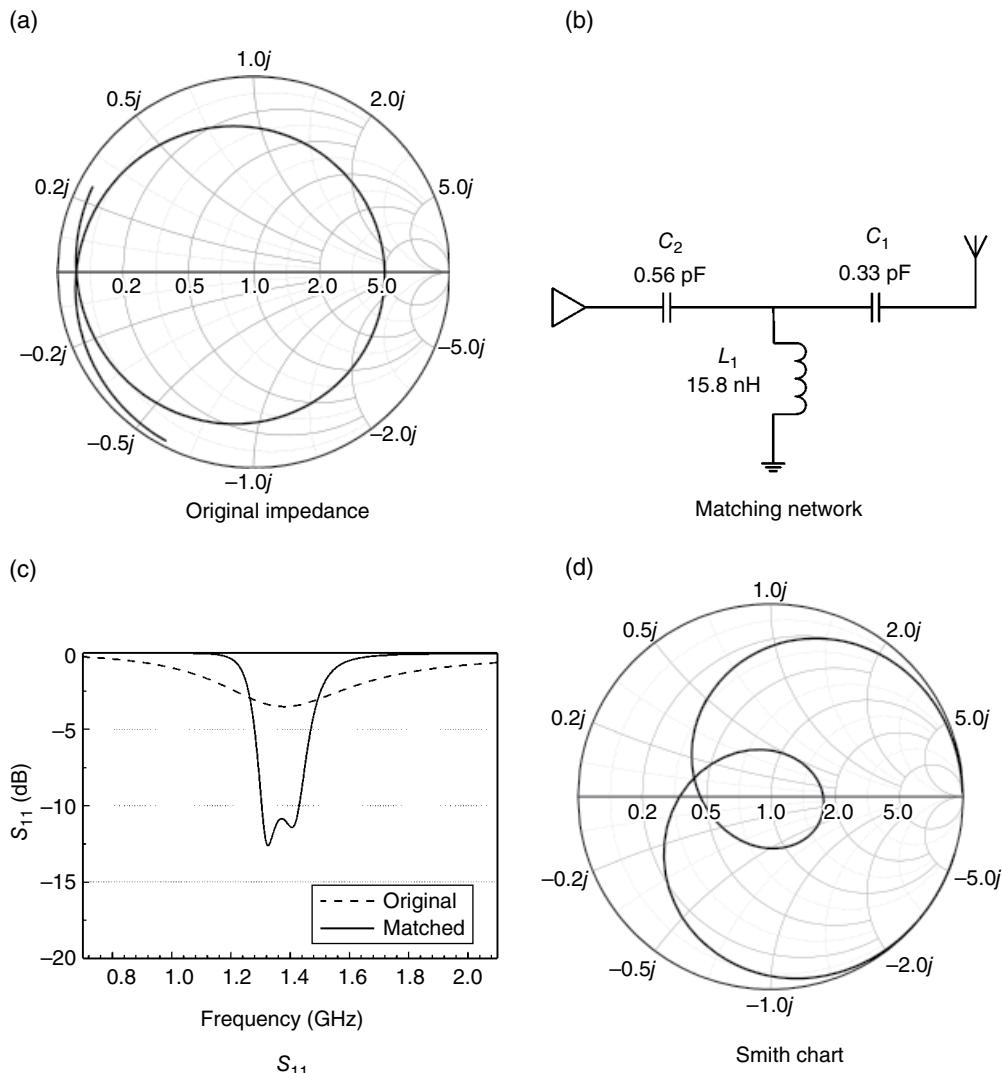


Figure 2.24 T-shaped matching circuit.

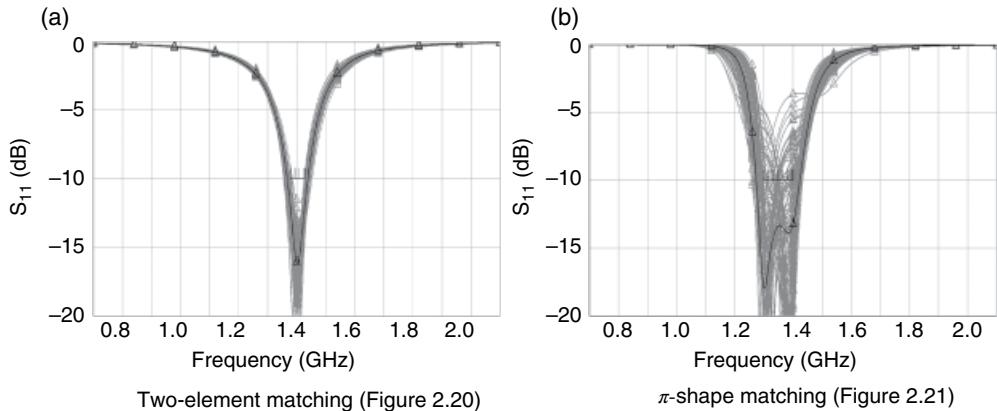


Figure 2.25 Tolerance analysis (5% of component value).

0.33 pF series capacitor, followed by a 15.8 nH shunt inductor and a 0.56 pF series capacitor. A detailed explanation is omitted here, because the design consideration is similar to those demonstrated in Figures 2.21 and 2.23.

When compared with a two-element matching, a π -shaped matching or a T-shaped matching can significantly improve the antenna bandwidth; however, there is a price to pay. Because the components used in a π -shaped matching or a T-shaped matching must move the impedance further, they are more sensitive to manufacturing variation. Figure 2.25 is the yield analysis results of circuits shown in Figures 2.20 and 2.21, respectively. The tolerance of all components used in the simulation is assumed to be $\pm 5\%$ uniformly distributed. Figure 2.25a is the yield results of a two-element matching. The results of 200 simulations are superimposed on Figure 2.25 by light gray lines. The black line is the optimal response when there is no tolerance. The designed -10 dB bandwidth of the two-element matching is 70 MHz. If selecting a 50 MHz bandwidth as the specification, the yield rate is 96%. On the contrary, although the design bandwidth of the π -shaped matching is 180 MHz, the yield is only 81%, even using the same 50 MHz bandwidth as the specification. The yield analysis of the π -shaped matching is carried out again using a 100 MHz bandwidth as the specification. As expected, the yield drops to 74% this time.

It is now clear that the bandwidth expanding techniques are not a free lunch. They have some inherent risks. If using a complex matching circuit, one must select components with tighter tolerance, manufacture antenna more cautiously, and always remember to do a thorough yield analysis.

2.3 Dual-Band Matching

The bandwidth of an antenna is constrained by the physical volume it occupies. If an antenna has enough volume, it is not difficult to design an antenna to cover all the required bands, even without the help of any matching network. But if the antenna size is limited, the matching circuit can give the designer much more freedom. When designing a dual band antenna in a limited space, it is normal practice to optimize one band, then use the matching circuit to take

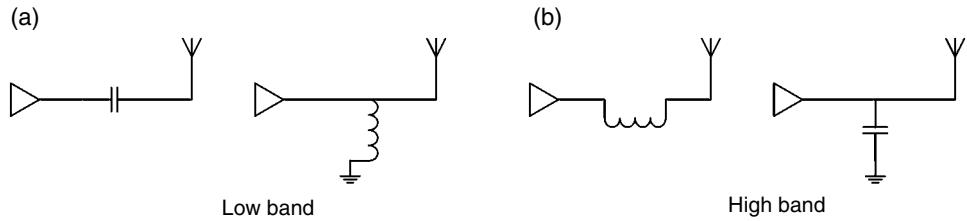


Figure 2.26 Matching components for low band and high band, respectively.

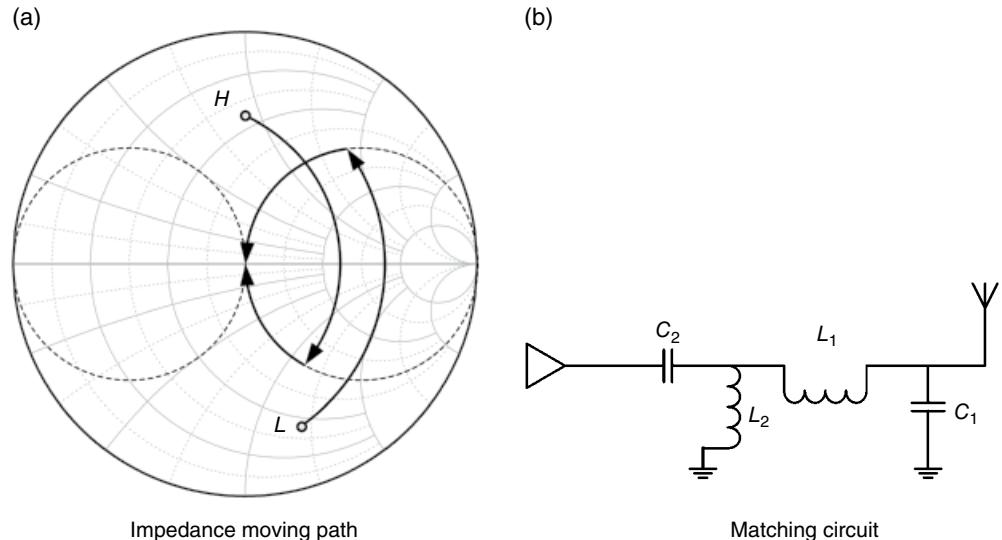


Figure 2.27 Dual-band matching Example 2.1.

care of the other band. Some devices have more than one position, such as the flip open and the flip close positions on a clam shell phone, the dual-band matching techniques are also very handy when a balanced performance needs to be achieved in all bands and all positions.

As shown in Figure 2.26, the building blocks of dual-band matching, which are the series capacitors, the shunt inductors, the series inductors, and the shunt capacitors, are the same as those used in the single-band matching. But the components that can be used at each band are limited to two. Using the series inductor as an example, at very low frequencies the complex impedance of a series inductor is close to 0, thus it has no effect on a circuit. At very high frequencies, a series inductor is equivalent to an open circuit, which totally breaks the circuit. As a rule of thumb, when selecting a matching component, the component should have more impact in the target band and less effect in the other band. Thus, a series inductor should only be used to match the higher band. Similarly, shunt capacitors can also be used in the higher band matching. When the band that needs to be matched is the lower band, only the shunt inductor and the series capacitor can be used.

As an example, a dual-band matching scenario is shown in Figure 2.27a. The impedance at low band and high band are marked as L and H , respectively on the Smith chart. In an

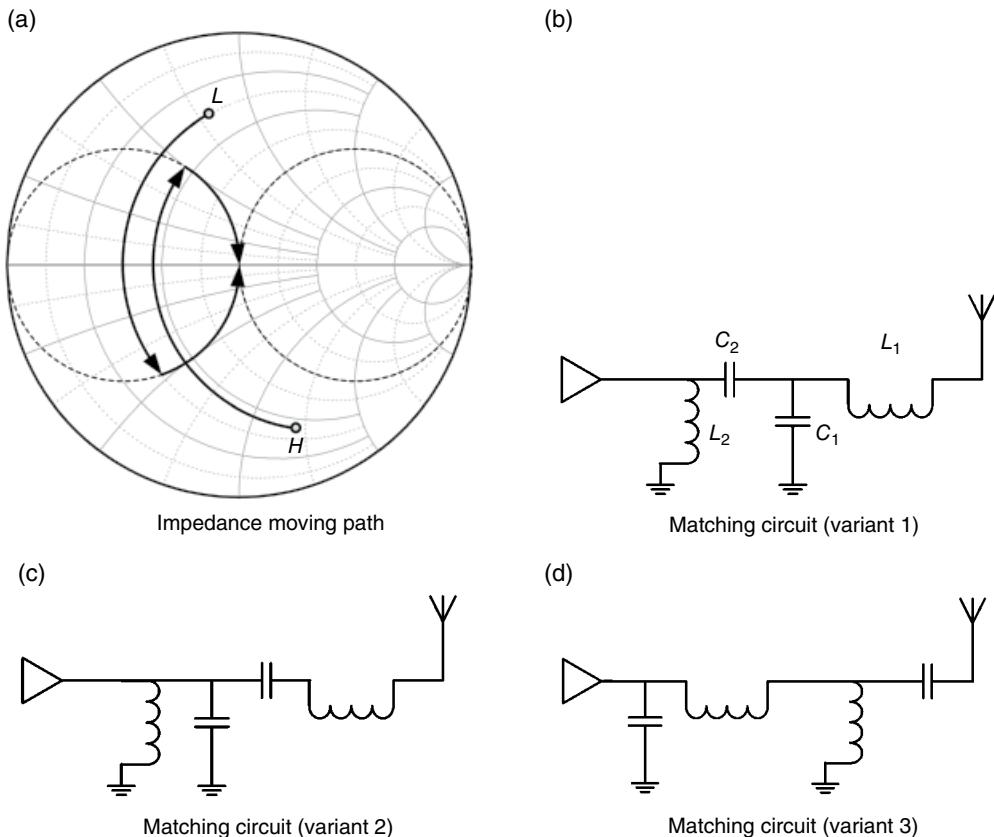


Figure 2.28 Dual-band matching Example 2.2.

ideal world, where components used for high band only impact the high band, the circuit shown in Figure 2.27b is one of many possible choices to match both bands. The shunt capacitor C_1 moves the high band impedance to the bottom part of $r=1$ circle; then the series inductor L_1 moves it to the center of the Smith chart. Similarly, the shunt inductor L_2 moves the lower band impedance to the top part of $r=1$ circle; then the series capacitor C_2 moves it to the center.

Figure 2.28 shows an example where the load impedances are the same as shown in Figure 2.27a; however, the corresponding band of each load impedance is reversed. The impedance on the top half is now the low band one. The matching circuit is shown in Figure 2.28b. It is still assumed that each component only impacts its corresponding low or high bands. In an ideal world, if the relative sequence between L_1/C_1 and C_2/L_2 can be kept, any variants in Figure 2.28b should have the same frequency response. Two equivalent variants under the ideal assumption are shown in Figure 2.28c and d.

In reality, a component cannot exclusively influence one specific band. It always influences both bands simultaneously. Figure 2.28 only shows a few variants; there are tens of possible permutations and combinations. When optimizing the component sequence of a matching

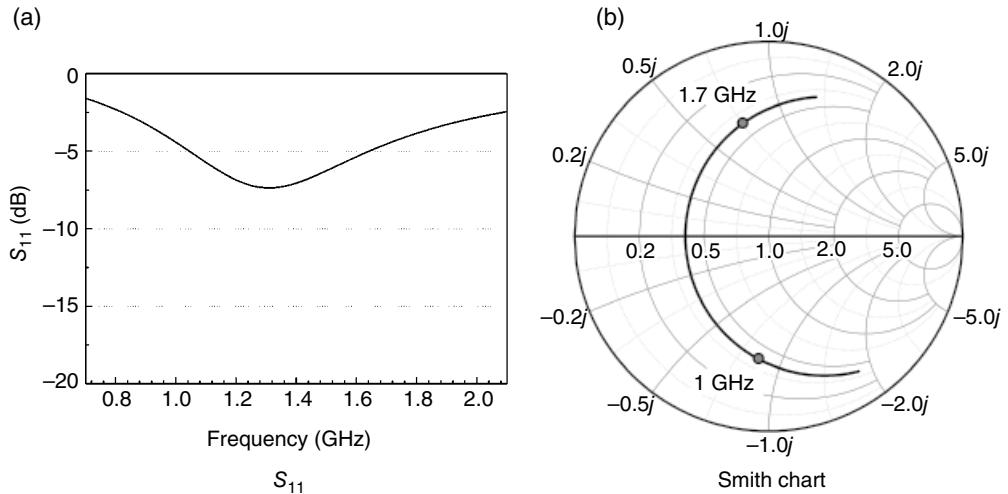


Figure 2.29 Dual-band matching Example 2.3.

circuit, both experience and trial and error play a role. The achievable bandwidth and value of matching component depend on the circuit layout. As a demonstration, an antenna [Simulation file: Chap2_Fig.29.s1p] with S_{11} and impedance curves shown in Figure 2.29a and b is used. The optimization goal of matching is to achieve a S_{11} better than -20 dB at both $0.98\text{--}1.02 \text{ GHz}$ and $1.67\text{--}1.73 \text{ GHz}$.

Figure 2.30a shows a variant of matching circuits, which matches the high band first. In this example, when using the shunt capacitor and the series inductor to match the high band impedance, these components also favorably shift the low band impedance; thus, only a shunt inductor is needed for the low band. In the design process, it is an engineer's responsibility to decide on the circuit layout; the optimizing work can be "subcontracted" to any circuit simulation software. Figure 2.30b is a matching circuit which takes care of the low band first. Similarly, the shunt inductor and the series capacitor have a positive impact on the high band impedance; thus, only a shunt capacitor is needed for the high band. Comparing Figure 2.30a and b, these two circuits have different component values and even different inductor/capacitor counts. But both circuits achieve a similar bandwidth. Figure 2.30c and d shows final impedance curves of circuits (a) and (b), respectively. The simulated S_{11} of circuit (b) is shown in Figure 2.30g as the solid line.

Figure 2.30e is a demonstration of how a component sequence can change the achievable bandwidth. By keeping the component sequence of the low band matching block of Figure 2.30b, and only flipping the matching blocks of low and high band, the achievable bandwidth is shrunk.

It has been mentioned that the component sequence in each band cannot be flipped. Figure 2.30f shows an example which has flipped the component sequence of low band. The simulated result, which is far from the optimization goal, is also shown in Figure 2.30g. The result intuitively demonstrates that the component sequence in each band is critical.

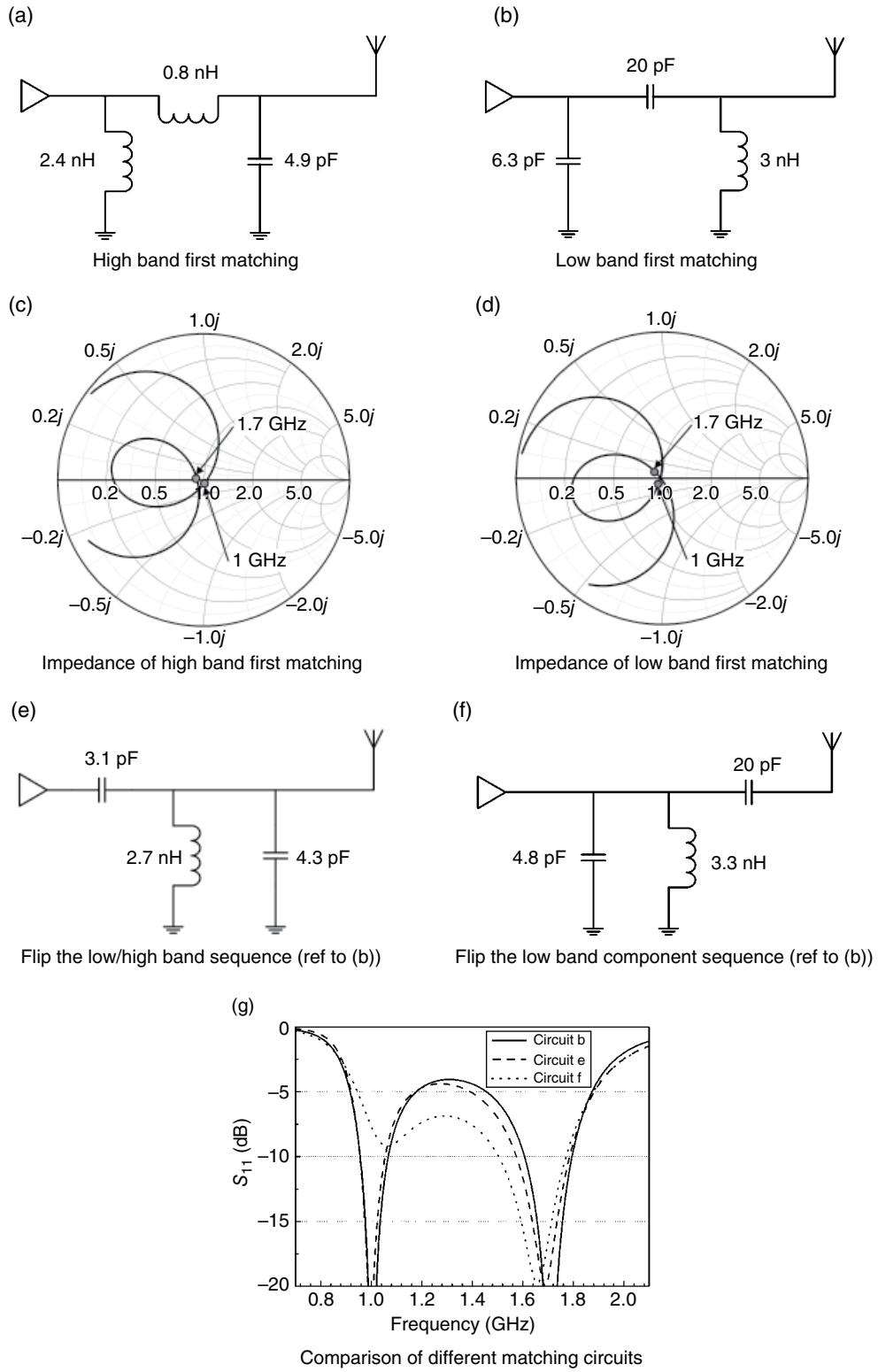


Figure 2.30 Selection of matching circuits.

2.4 Reconfigurable Matching

So far, all the techniques discussed belong to the category of fixed matching, which means whenever an antenna design is finished, the antenna matching network is fixed and the antenna bandwidth is also fixed. It is well known that the bandwidth of an antenna is constrained by the physical volume it occupies. The static bandwidth of an antenna has a physical limit, which cannot be passed even with the help of matching networks. Note the word “static” in the last sentence. Actually that is the key to breaking the physical limit of antenna bandwidth. Based on the earlier discussion, it is clear that an antenna can be matched to a wide range of frequency bands by using different matching components and circuit topology. The function of a reconfigurable circuit is to provide various matching statuses based on the control signal, thus achieving good matching in a much wider frequency range.

Although the total working bandwidth of an antenna with reconfigurable matching can surpass the bandwidth limit of the antenna with fixed matching, it does not actually break the physical law. At any given time, the static bandwidth of the antenna is at best equal to the bandwidth achievable by a fixed matching. The total bandwidth of a reconfigurable matching antenna is defined as the summary of all static bandwidths at different status.

The reconfigurable matching is suitable for a system which needs to cover a wide frequency band, but at any given time the system only uses a small portion of that frequency band. For example, the frequency band allocated to the Integrated Services Digital Broadcasting-Terrestrial (ISDB-T) system is 470–770 MHz, but each ISDB-T channel is 6 MHz wide. The reconfigurable solution is a perfect fit in such a system. On the other hand, if a system, such as a device based on the ultra-wideband (UWB) standard, requires a working bandwidth that exceeds the static bandwidth that an antenna can provide, a reconfigurable solution cannot really help.

Another issue that needs to be mentioned is the efficiency degradation due to the loss of reconfigurable circuits. All passive matching components, such as inductors and capacitors, have inherent losses. Similarly, all active components used in reconfigurable circuits, such as a varactor, the PIN diodes, the field-effect transistor (FET) switches, or the microelectromechanical system (MEMS) switches, also have losses. The loss from those switchable devices is accumulated with the existing loss introduced by the passive components, thus decreasing the overall antenna efficiency. If comparing at the same frequency point, the efficiency of an antenna using reconfigurable matching is always less than its counterpart with fixed matching. The reconfigurable matching does sacrifice some performance at the peak of the efficiency curve of an antenna; in exchange, it obtains overall better efficiency in a much wider frequency range.

2.4.1 *Reconfigurable Matching—Varactor-Based*

The varactor is a diode, when reverse biased, whose capacitance is sensitive to the applied voltage. The higher the reverse bias voltage on a varactor, the lower the capacitance. Due to this unique character, a varactor is widely used in electrical tuning devices, such as a voltage controlled oscillator (VCO). In a VCO, the varactor is a part of a resonator which decides the resonant frequency. By adjusting the bias voltage, the capacitance changes, thus causing the change of the working frequency.

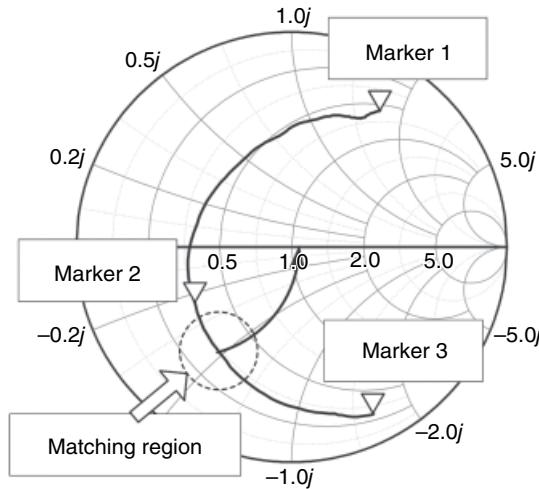


Figure 2.31 Impedance curve of the original antenna.

As an example, a varactor-based reconfigurable matching is designed for ISDB-T service. The ISDB-T operates at the frequency range of 470–770 MHz, which covers a relative bandwidth of 48.4%. The typical dimensions of a mobile device are quite small in comparison to a quarter of wavelength at 470 MHz, so designing a passive internal ISDB-T antenna to cover the whole band and provide good performance is always a challenge. With the help of reconfigurable matching, a better than 10 dB measured S_{11} across the 470–770 MHz bandwidth can be achieved.

Without loss of the generality, a meander line antenna is used as the design example. The detailed procedure of how to design a meander line antenna will be discussed in Chapter 3. For now, we only need to use the impedance curve of the meander line antenna. In fact, the antenna element can be replaced by an antenna of any form factor, as long as its impedance has a similar frequency response.

Figure 2.31 shows the impedance curve of the original antenna on the Smith chart. Three markers on the curve, 1, 2, and 3 denote the highest, median, and lowest working frequency, respectively. There are two considerations in designing a varactor-based antenna matching network. First, the capacitance ratio between the maximum and minimum capacitance which a varactor needs to provide must fall within a reasonable range. Second, the topology of the matching circuit should not be overly complex.

For the impedance curve shown in Figure 2.31, a range of impedance can be matched if a matching network is as shown in Figure 2.32. In this configuration, the antenna is connected by a series varactor C_v , and then followed by a shunt inductor L . The shunt inductor can move any impedance inside the dash line circle area toward the matching point, where the origin of the Smith chart is. The series varactor can move the other impedance between markers 1 and 2 into the solid line circle. Through the combined effects of the series varactor and shunt inductor, all impedance between markers 1 and 2 can be matched.

Parametric studies and optimizations lead to the conclusion that a meander line antenna element and a varactor with a capacitance ratio of 27.9 (25.1 pF/0.9 pF) are required to achieve

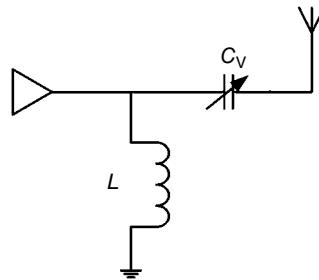


Figure 2.32 Matching circuit of the original antenna.

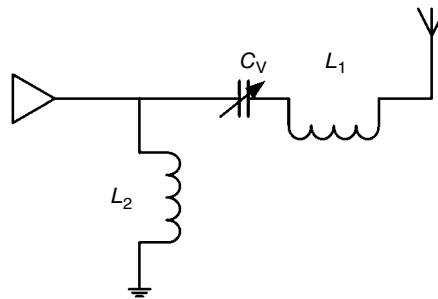


Figure 2.33 Diagram of matching circuit.

the 470–770 MHz bandwidth. To decrease the capacitance ratio required by the varactor, a series inductor L_1 is added to the matching network. The final matching circuit layout is shown in Figure 2.33.

To explain the function of L_1 , let's reinvestigate the impedance curve shown in Figure 2.31. Besides the impedances between markers 1 and 2 which have been properly matched by a circuit with a series varactor and a shunt inductor, the impedances between markers 2 and 3 also have reasonable return loss and its reactance varies smoothly when the frequency changes. However, this segment of impedance is beyond the reachable tuning range of a matching circuit composed only of a series varactor and a shunt inductor. With the addition of a series inductor, this segment of the impedance line is moved from the bottom to the top of the solid line circle, and thus into the tunable range of the matching circuit. Simulations indicate that the addition of the extra series inductor reduces the required capacitance ratio from 27.9 to 8.2 (7.7 pF/0.94 pF).

To verify the concept of the varactor tuning matching network, a prototype is fabricated and measured. The specific antenna dimensions can be found in Figure 2.34. The antenna element is made of a flat metallic meander line. The element is folded and attached to the outer side of a foam support, whose dimensions are 80 mm × 10 mm × 10 mm. The antenna is installed on a single-sided FR4 board. The size of the board is 140 mm × 80 mm, with a thickness of 1.6 mm. The impedance of the antenna without matching has been shown in Figure 2.31.

The matching circuit is fabricated on a 10 mm × 10 mm double-sided FR4 board shown in Figure 2.35. Most of the components labeled in Figure 2.35 correspond to Figure 2.33, with the exception of R_b and C_p , which are a bias resistor and a bypass capacitor supplying bias

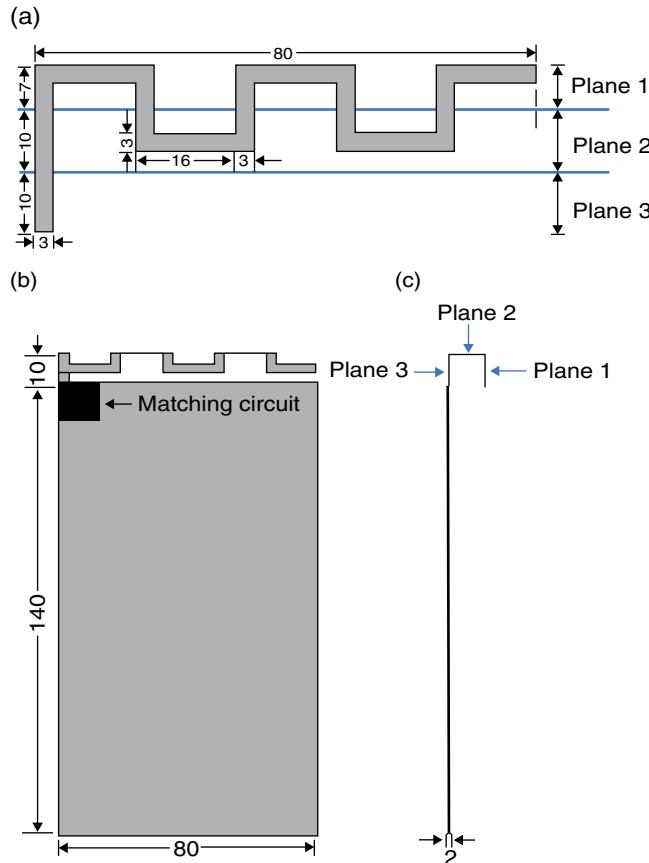


Figure 2.34 Geometry of meander line antenna: (a) antenna element, (b) front view, and (c) side view. (Source: Li et al. [16]. Reproduced with permission of John Wiley & Sons, Inc.)

voltage to the varactor. The varactor used in the experiment is a Skyworks SMV1247-079. The series inductor L_1 and shunt inductor L_2 used in the experiment are 15 and 22 nH, respectively. The back of the matching circuit, a solid piece of copper, which also functions as the ground, is soldered to the big board. Voltage is supplied by three AA batteries. A variable resistor is used to adjust the bias voltage.

Figure 2.36 shows the measured S_{11} of the antenna with the varactor bias voltage changed from 0 to 3.92V. Clearly, the operating frequency can be continuously tuned and cover the whole ISDB-T band with S_{11} better than -10 dB . Based on the specification sheet, the capacitance values are 8.86 and 0.78 pF at 0 and 3.92V bias, respectively. This represents a capacitance ratio of 11.3 rather than the 8.2 obtained in the simulation. It is believed that this discrepancy results from the imperfection of the varactor and parasitic reactance from lumped inductors and circuit board.

Because the capacitance of a varactor is sensitive to the bias voltage, a varactor-based reconfigurable matching circuit is suitable for a receiver-only device or a transceiver with a

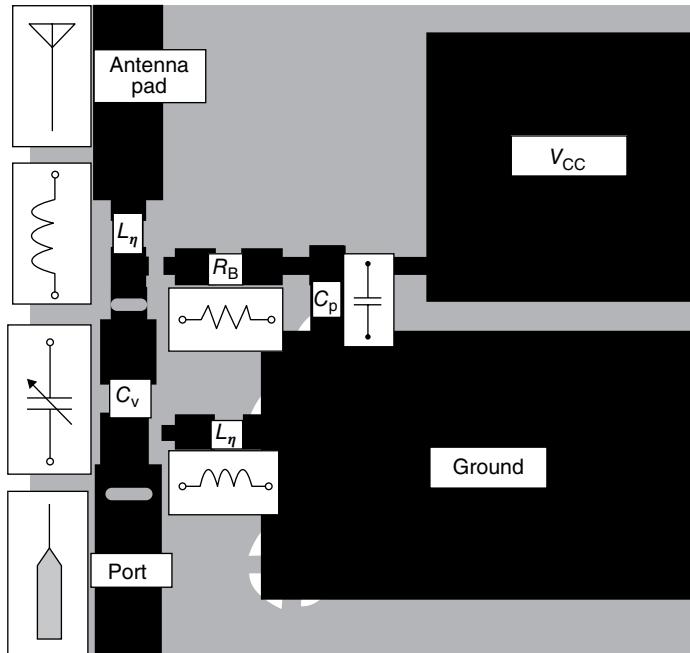


Figure 2.35 Layout of matching circuit. (Source: Li et al. [16]. Reproduced with permission of John Wiley & Sons, Inc.)

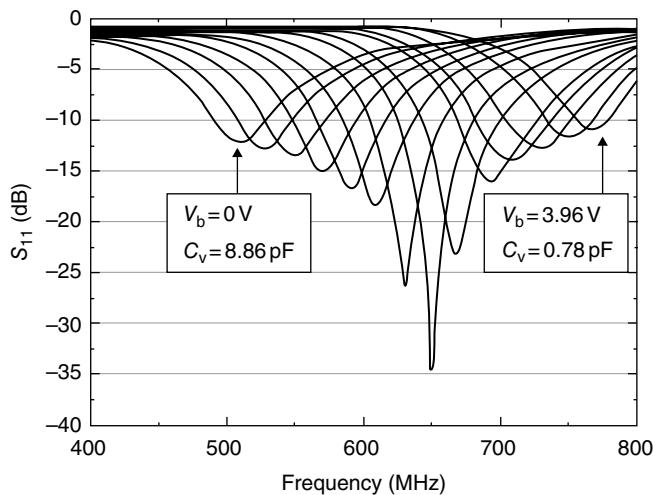


Figure 2.36 Measured S_{11} . (Source: Li et al. [16]. Reproduced with permission of John Wiley & Sons, Inc.)

relatively low transmitting power. If it is used in a high power transmitter, the voltage of the high power input signal is superimposed on the bias voltage, thus, the combined effective bias voltage becomes time-variable and the matching circuit also becomes time-variable and non-linear. As a consequence, it generates unwanted modulation in the original input signal.

2.4.2 Reconfigurable Matching—Switch-Based

In a varactor-based reconfigurable matching, both the control/bias voltage and the signal voltage are superimposed on the varactor, thus causing a nonlinear problem when the signal is strong. A straightforward solution to this problem is to separate the control voltage and the signal. A switch-based reconfigurable matching is the implementation of such an idea. Unlike the continuously varied bias voltage applied to a varactor, the control voltage of a switch is a discrete digital variable which is either high or low. If a switch is based on a MEMS technology, there is no nonlinear problem at all and the power handling capability is the dominant constraint. If a switch is based on FETs or PIN diode technology, the maximum power it can handle depends on the DC voltage applied to the component.

As an example, a switch-based matching circuit is used to design an ISDB-T antenna. ISDB-T is a receive-only standard, and a varactor-based matching circuit is enough to take care of it. The reason for selecting an ISDB-T band as an example is for the purpose of comparison. By using two different technologies to achieve the same goal, the pros and cons of both technologies can easily be compared.

The antenna's radiating element used here is similar to the one shown in Figure 2.34. For the convenience of comparison, the most specific dimensions of two devices, such as the PCB board size and antenna size, are identical. The only difference is the total length of radiator elements. To obtain the best coverage, both elements are co-optimized with the corresponding matching circuits and, as a consequence, they are slightly different.

Without losing the generality, the impedance curve shown in Figure 2.37 is divided into four regions. To match all four regions, the matching circuit needs to be reconfigurable among

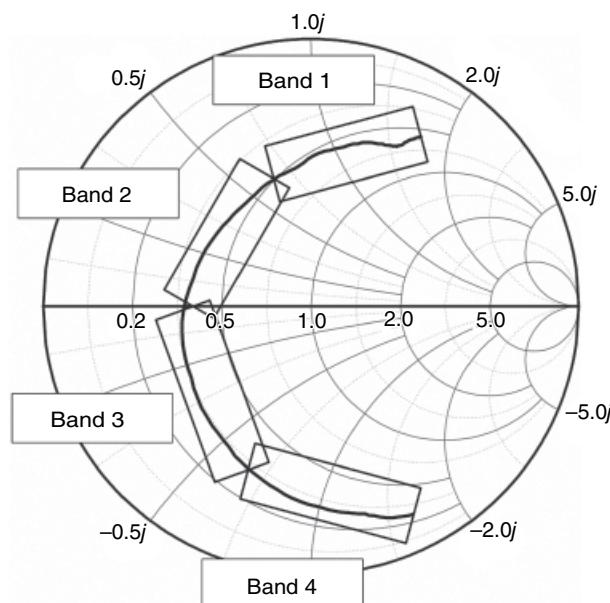


Figure 2.37 Dividing the impedance curve into four regions.

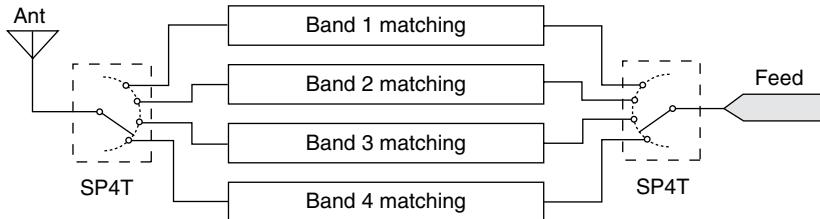


Figure 2.38 Reconfigurable matching circuits with four positions: SP4T architecture.

the four positions. Of course, the impedance curve can be divided into fewer or more regions, which means the matching circuit must have less or more working positions.

In Figure 2.37, band (1) represents the highest frequency range which covers 690–770 MHz; band (4) is the lowest range which covers 479–550 MHz. As has been discussed in Section 2.2.2, there are up to four different approaches which can match an antenna with two components. Considering there are four regions, the total possible variants of matching circuit can be 16.

One way to implement four matching positions is shown in Figure 2.38. There are two single-pole four-throw (SP4T) switches. The antenna is connected to the single-pole terminal of one switch. The 50Ω transmission line is connected to the single-pole terminal of the other switch. All four matching circuits are connected between corresponding four-throw terminals of both switches. Two SP4T switches are synchronously controlled to select one from four available matching circuits. By using this architecture, all four matching circuits are isolated from one another, and thus can be designed separately. Assuming each matching network uses two components, the matching circuit shown in Figure 2.38 requires eight components and two SP4T switches.

There is another way to obtain the required four positions. In this approach, four positions are designed and optimized simultaneously. When selecting matching components and the circuit layout, reuse is the primary consideration. As shown in Figure 2.39, for matching bands (1)–(4), four different matching layouts are selected. Looking from the antenna side, a series capacitor and a shunt capacitor are used to match the band (1); a series capacitor and a shunt inductor are used for band (2); a series inductor and a shunt capacitor are used for band (3); and a series inductor and a shunt inductor are used for band (4).

The four positions of matching circuits shown earlier can be integrated into one circuit as shown in Figure 2.40. This circuit is composed of two single-pole double-throw (SPDT) switches. There are four components: a series inductor L_1 , a series capacitor C_1 , a shunt inductor L_2 , and a shunt capacitor C_2 . One SPDT is connected between the antenna port and two series components, and the other SPDT is connected between the 50Ω output transmission line and two shunt components. There are two switches and each switch has two positions, so the total sets of the circuit are 4 (2^2). Each combination of switch positions in Figure 2.40 corresponds to a circuit status in Figure 2.39.

Unlike the matching network shown in Figure 2.38, where components in each matching branch are independent. Any capacitor/inductor shown in Figure 2.40 has to be used in matching sets. Thus, the simultaneous optimization of all four components across all four frequency regions is necessary. A detailed description of the optimization procedure is beyond the scope of the book and can be found in Li *et al.* [17]. Based on the antenna impedance shown in Figure 2.37, the final optimized values of four components are $L_1=5.1\text{ nH}$, $L_2=15\text{ nH}$,

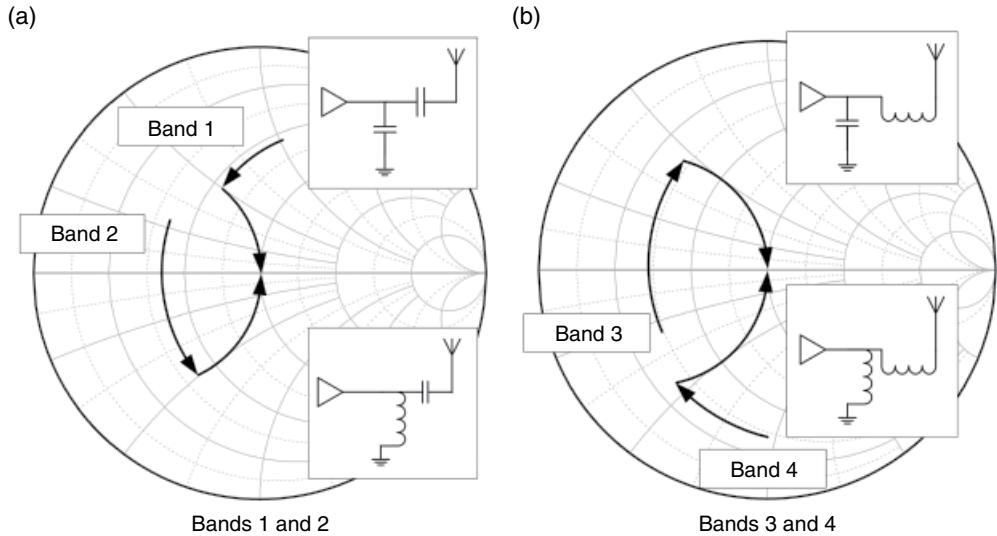


Figure 2.39 Corresponding matching circuits of four regions.

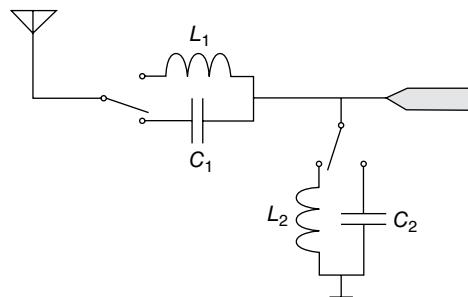


Figure 2.40 Reconfigurable matching circuit with four states: SPDT architecture.

$C_1=4.7\text{ pF}$, and $C_2=5.6\text{ pF}$. Lines (a)–(d) shown in Figure 2.41 correspond to bands (4)–(1) shown in Figure 2.37.

When tuning the matching circuit, each component has its own effect on the antenna's frequency response. The series inductor L_1 shifts the resonant frequency of lines (a) and (b), which corresponds to bands (4) and (3). Increasing the value of L_1 can lower the resonant frequencies of both. The series capacitor C_1 shifts the frequency of lines (c) and (d), which corresponds to bands (2) and (1). Increasing the value of C_1 can lower their resonant frequencies. Tuning L_2 can simultaneously change the matching of lines (a) and (c). C_2 can impact the matching of both lines (b) and (d).

As shown in Figure 2.41, the final achieved S_{11} by the SPDT type reconfigurable matching is -5 dB across the $470\text{--}770\text{MHz}$. If adding one more SPDT switch to the circuit shown in Figure 2.40, the total circuit status can reach 8 (2^3); therefore, the total impedance curve can be divided into eight regions and better matching is achievable. Of course, all switches have inherent loss. The efficiency degradation due to the matching circuit eventually will surpass

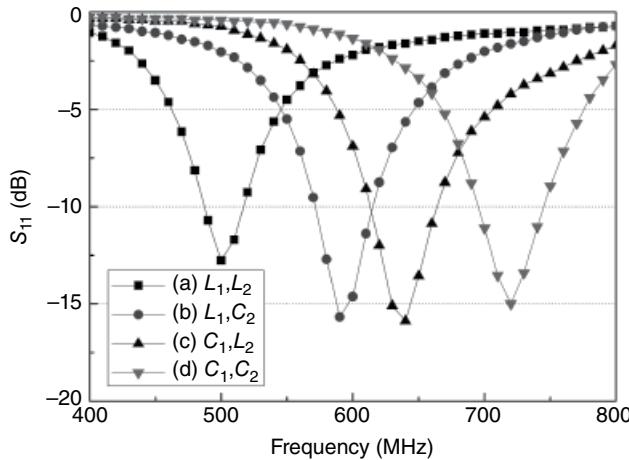


Figure 2.41 Simulated S_{11} .

the efficiency gained from improved matching. There is a balance between better matching and higher efficiency.

Compared with the varactor-based matching circuit, which can achieve a better return loss, -10 dB , across the whole band, the switch-based matching technique needs more components and is more complex to design. In exchange for all its complexity, the switch-based matching can handle higher transmitting power and can be controlled directly by digital signals instead of analog bias voltages required by varactors. With the progress of MEMS switches, in the future the loss of MEMS switch circuits might be less than the loss introduced by PIN diodes; therefore, switch-based solutions might be able to achieve better overall antenna efficiency.

References

- [1] Gonzalez, G. (1996) *Microwave Transistor Amplifiers: Analysis and Design*, 2nd edn, Prentice Hall.
- [2] Balanis, C.A. (2005) *Antenna Theory: Analysis and Design*, 3rd edn, Wiley-Interscience.
- [3] Iskander, M.F. (2000) *Electromagnetic Fields and Waves*, 1st edn, Waveland Press Inc.
- [4] Sadiku, M.O. (2009) *Elements of Electromagnetics*, 5th edn, Oxford University Press, USA.
- [5] William Hayt, J.B. (2005) *Engineering Electromagnetics*, 7th edn, McGraw-Hill Science/Engineering/Math.
- [6] Cripps, S.C. (2007) "Chasing Chebyshev," *IEEE Microwave Magazine*, **8**, 34–44.
- [7] "Normalized Impedance and Admittance Coordinates, from ZY-01-N. Color by J. Colvin, University of Florida," (1997) http://rfic.ucsd.edu/files/smith_chart.pdf. Retrieved 9 July 2016.
- [8] Ulaby, F.T., Michielssen, E., and Ravaioli, U. (2010) *Fundamentals of Applied Electromagnetics*, 6th edn, Prentice Hall.
- [9] "Advanced Design System (ADS), Agilent Technologies Inc.," <http://www.home.agilent.com/agilent/product.jspx?nid=-34346.0.00>. Retrieved 25 October 2010.
- [10] "Microwave Office, AWR Corporation," <http://web.awrcorp.com/Usa/Products/Microwave-Office/>. Retrieved 25 October 2010.
- [11] "Murata Monolithic Ceramic Capacitors," <http://www.murata.com/~media/webrenewal/support/library/catalog/products/capacitor/mlcc/c02e.ashx?la=en>. Retrieved 9 July 2016.
- [12] "Johanson Technology: Multi-Layer High-Q Capacitors, HiQ MLCC, SMD, Low ESR Capacitors," <http://www.johansontechnology.com/downloads/jti-catalog.pdf>. pp. 7–18. Retrieved 9 July 2016.

- [13] "Murata Chip Inductors Selection Guides," http://www.murata.com/products/inductor/selection_guide/chip-inductor/index.html. Retrieved 25 October 2010.
- [14] "Coilcraft 0402HP High Performance Wirewound Ceramic Chip Inductors," <http://www.coilcraft.com/0402hp.cfm>. Retrieved 25 October 2010.
- [15] "Johanson Technology: RF Ceramic Chip Inductors," <http://www.johansontechnology.com/downloads/jti-catalog.pdf>. pp. 19–27. Retrieved 9 July 2016.
- [16] Li, Y., Zhang, Z., Chen, W. *et al.* "A compact DVB-H antenna with varactor-tuned matching circuit," *Microwave and Optical Technology Letters*, **52**(8), 1786–1789.
- [17] Li, Y., Zhang, Z., Chen, W. *et al.* (2010) "Using switchable matching circuits to design compact wideband antennas," *IEEE Transactions on Antennas and Propagation*, **58**, 2450–3457.

3

External Antenna

The external antenna is the first member of the antenna family in the world of mobile phones. It appeared in the first commercial mobile phone, the Motorola DynaTAC 8000X, in 1983. External antennas were the only players in the mobile phone antenna market for more than 15 years. In the late 1990s, phones with internal antennas started to emerge. Candy-bar phones, also called single-piece phones, with internal antennas rapidly expanded their market share in Europe; in the meantime, external antennas in clam shell phones kept dominating the North American market for another 5 or more years. Eventually, internal antennas took over the global antenna market.

The driving forces behind the fading of external antennas are both fashion and technical progress. Of the two, technical progress is the necessary condition. Even 10 years after mobile phones came onto the market, there were just over one million users in the world at the end of 1993. In comparison, mobile subscriptions were more than six billion at the end of 2014. It is easy to imagine the user density was quite low after those limited users spread out. At that time, the most urgent issue was coverage. All mobile phone service providers wanted to expand their network coverage as much as possible with all the money they could spare. The coverage area of each base station depends on transmitting power, receivers' sensitivity, the path loss of electromagnetic wave, the antenna gain of base stations, and the antenna gain of mobile phones. The received signal power P_{RX} in dBm can be expressed by Equation 3.1:

$$P_{RX}(\text{dBm}) = P_{TX}(\text{dBm}) + G_t(\text{dBi}) + G_r(\text{dBi}) - \alpha \log_{10} R + C \quad (3.1)$$

where P_{TX} is the transmitter power in dBm. G_t and G_r are the transmitter and receiver antenna gains, respectively, in dBi, relative to isotropic. R is the distance between a base station and a mobile phone. C is a constant related to the frequency. α is the propagation attenuation factor, and normally it is a number between three and four in a city environment.

To get a phone to work properly, the received signal power must be higher than its sensitivity threshold. It is obvious that higher gain means larger coverage area. For example, assuming α equals three and the antenna gain of mobile phones is improved by 3 dB, the coverage area of the same base station will increase about 58%. When a phone is held by hand next to the head, the efficiency of a phone with an external whip antenna is likely to be 3 dB or more higher than one with an internal antenna. That made a huge difference when the mobile industry was still in its infancy.

There are two constraints on the mobile phone network planning. The first is how large an area each base station can cover. The second is how many users each station can support. In the beginning of the mobile phone's evolution, the primary challenge was how to design the whole system, including both base stations and mobile phones, to cover a huge area.

With the explosive adoption of mobile phones by the mass population, the atmosphere started to change. With more and more people enrolled into a mobile network, eventually the network reached a saturated status. All the available frequency resources or channels which could support users were exhausted. Now the second constraint, how to support more users, became the primary challenge. Fortunately, due to the cellular architecture that mobile networks adopted, there is a solution to this challenge. The total number of users a network can support depends on both the maximum capacity of each base station and the total number of base stations in the network. By adding more and more base stations to the network, the mobile service providers are able to keep up with the growth of mobile users. As a consequence, the distance between a mobile phone and the nearest base station is continuously shrinking. Shorter distance means less path loss, which in return reduces the requirement to improve antenna gain or antenna efficiency.

Compared to the United States, Europe has a much higher population density and started to shrink the cell size of each base station earlier. That partially explains why the European market was the first to successfully popularize the internal antenna. Actually, the population density in Japan is even higher than Europe; however, candy-bar phones with internal antennas were not popularly accepted there. That can only be explained by fashion or cultural difference.

In the United States, one of the nationwide mobile providers, Nextel, merged with Sprint in 2003. Before the merger, Nextel ranked fifth or sixth in the US market, which was pretty close to the bottom of the list of nationwide providers. Nextel had a unique business model and some attractive services, such as push-to-talk (PTT) which is also marketed as "walkie-talkie from coast to coast." This uniqueness gave Nextel the possibility of keeping its core customers and a considerable profit margin, but it also kept Nextel from expanding its customer base. Just before the merger, Nextel still had to cover the United States with limited base stations. All the phones that Nextel carried had external antenna, and most of them even had a whip in the antenna assembly.

To some, the external antenna is like a dinosaur and belongs to the past century. However, the reason for not writing off the external antenna is that there is a lot of mature and valuable knowledge in the designing of external antennas. Much of this knowledge can be applied to or referred to by designs of state-of-the-art internal monopole and ceramic antennas.

3.1 Stubby Antennas

There are mainly two kinds of external antennas: stubby antenna and whip-stubby antenna. A stubby antenna extrudes from a phone and does not have any movable parts. Earlier in the mobile phone's history, all stubby antennas were helix antennas made of copper wire or



Figure 3.1 Stubby antennas. (Source: Reproduced with permission of Shanghai Amphenol Airwave.)

copper-coated steel wire (music wire). From the mechanical point of view, a helix antenna is the same as a spring coil and can be cheaply made by the same machines and tools. When mass-produced, spring coils cost a few pennies a piece, so the helix antenna is a considerable cost-effective solution. A spring coil has a round cross section, so a helix antenna normally possesses a column shape. That is why it has the name of “stubby.”

Latterly, stubby antennas have all kinds of cross sections, many of them are not made of helix and they are meander line antennas made of flex circuits. From an electrical point of view, a helix antenna is similar to a meander line antenna. It is not too difficult to master meander line antennas when you know helix antennas well.

Figure 3.1 shows some production stubby antennas. We will discuss the design of some of them in this section.

3.1.1 Single-Band Helix Stubby Antenna

Before discussing the single-band stubby antenna, let's take one step back. Let's start from the radiation of an infinitesimal current element. A detailed exposition of this is found in the classic textbooks [1–4]. The conclusion drawn in those books is directly borrowed here. In a word, any current can generate radiation.

Figure 3.2 is a matched transmission line system. On the left side is a signal generator with a source impedance of Z_0 . On the right side is a load whose impedance is also Z_0 . A transmission line of parallel wires with characteristic impedance Z_0 connects the signal generator and

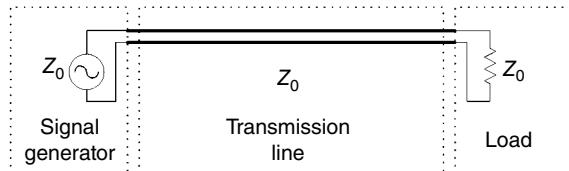


Figure 3.2 A terminated transmission line.

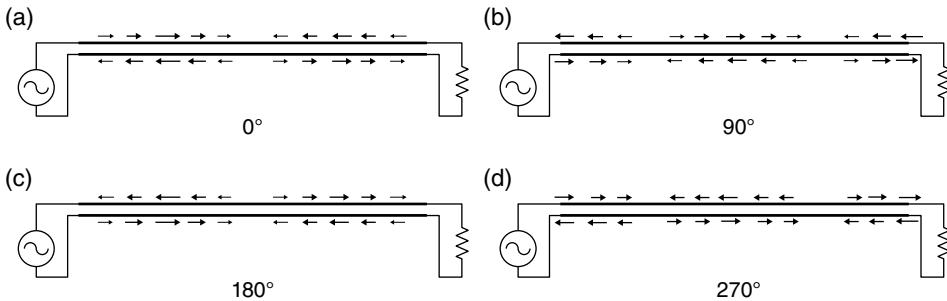


Figure 3.3 Transient current distribution of a traveling wave on a terminated transmission line.

the load. In a system like this, impedances at both interfaces, between the signal generator and the transmission line and between the transmission line to the load, are perfectly matched. The electromagnetic energy transfers from the generator to the load through a pure form of traveling wave. Whenever energy travels in a transmission line, it induces currents. Can we draw the conclusion that a transmission line radiates? The answer is no. A transmission line does not radiate in its designated working frequency band, which is from direct current (DC) to a certain frequency point. To explain that, we need to investigate the current distribution in a matched transmission line system.

Figure 3.3a-d shows transient current distributions at four different phase angles, 0° , 90° , 180° , and 270° , respectively. The current distribution varies with a period of 360° . The length of the transmission lines shown in all four figures is one wavelength. The arrows in the figures represent the direction of currents. The length of each arrow represents the current's amplitude at corresponding location. Figure 3.3a shows the transient current distribution when the phase angle is 0° . It can be seen from Figure 3.3a that there are currents in both wires of the transmission line. The current in the same wire reverses its direction every half a wavelength. At any given location, the currents in both wires are always the same in amplitude and the opposite in direction.

In theory, each current can generate a radiating electromagnetic field in the whole space. However, if the working frequency is low enough, the two currents in opposite directions generate two radiating electromagnetic field with complete reversed polarities. Based on the superimposition principle, the actual electromagnetic field induced by a transmission line is zero everywhere. That means a transmission line cannot radiate at low frequency.

Figure 3.3b shows the transient current distribution when the phase angle is 90° . Compared to Figure 3.3a, it is clear that the electromagnetic wave is propagating along the transmission line in a purely traveling wave mode and the propagating direction is from left to right.

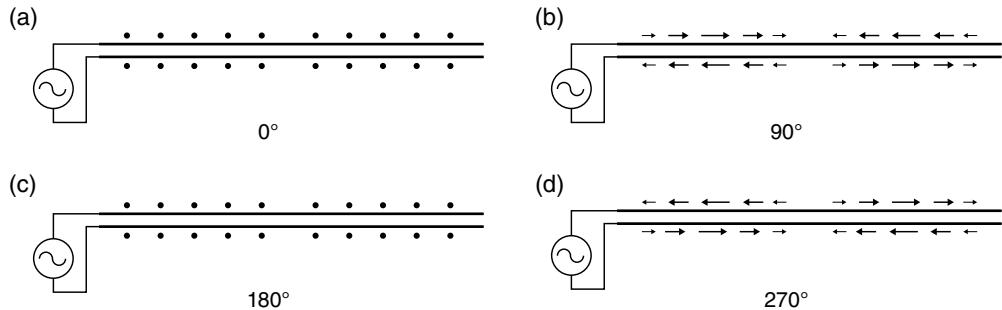


Figure 3.4 Transient current distribution of standing wave on an open-circuit transmission line.

In Figure 3.3b, currents in both wires are also the same in amplitude and the opposite in direction. So the transmission line still cannot radiate at low frequency. Similarly, currents in Figure 3.3c and d cannot radiate.

Based on Figure 3.3a–d, let's try an interesting exercise. Try to imagine an animated film; in the film, the current distribution starts as shown in Figure 3.3a, then the current continuously spreads to the right until it appears to be same as the current shown in Figure 3.3b. Keep the current moving to pass the status shown in Figure 3.3c, Figure 3.3d, and then Figure 3.3a again to finish a full cycle. This practice can help us better understand the process of energy propagation in a transmission line.

There is always a distance difference between any given location to these two wires. As is well known, the electromagnetic wave travels with a limited speed and the phase of an electromagnetic wave varies 360° every wavelength. When the frequency is so high that the signal's wavelength is comparable to the gap between two wires, the phase difference caused by the gap cannot be ignored, then that 180° relationship is also broken and the transmission line starts to radiate.

Next, let's look at a transmission line without a load. Using a water pipe as an analogy, similar to a water pipe which can constrain the water's flow without leaking, a transmission line can constrain energy's flowing without radiating. In our real-life experience, one can let the water out by turning on the faucet. Is it possible to radiate electromagnetic energy from the transmission line by simply removing the load or terminator shown in Figure 3.2? The answer is still no. Unlike water which can flow through a faucet, electromagnetic wave spreading to the right is completely reflected back at the location of the open circuit. The reason for the reflection is due to the mismatch between the characteristic impedance Z_0 and the impedance of the open circuit, which is $+\infty$. A detailed explanation is omitted here and can be found in references [1–4].

The incident wave propagating to the right and the reflected wave spreading to the left match in the transmission line to form a standing wave. Figure 3.4 shows the transient current distributions of a standing wave on an open-circuit transmission line. Figure 3.4a–d shows transient current distributions at four different phase angles. At the right end of the transmission line of all four figures is the open circuit, and the current is always 0. In the middle of the transmission line which is half a wavelength away from the open circuit there is another spot where the transient current is always 0. In fact, in all locations where the distance to the open circuit is an integral multiple of half a wavelength, the transient

currents there are always 0. The terminology of all these points is the wave node. At all locations where the distance to the open circuit is an odd multiple of a quarter of wavelength, the absolute amplitude of current is always the maximum. These points are called the “wave crest.” Because a standing wave is formed by superposition of both an incident and a reflected wave, its maximum amplitude is twice as much as the maximum amplitude of the incident wave.

Now we can look at the transient current of a standing wave. In Figure 3.4a when the phase angle is 0° , the current at all places along the transmission line is 0. Then amplitudes of current at all places start to increase proportionally. However, the wave is not traveling and always stays in the same place. The absolute amplitudes of current at all locations reach their extreme when the phase angle is 90° , which is shown in Figure 3.4b. The maximum current appears at all wave crests. The current is twice as much as the current shown in Figure 3.3 when the signal generator injects the same amount of power.

Similar to Figure 3.3, the current in the same wire reverses its direction every half a wavelength. At any given location, the currents in both wires are always the same in amplitude and the opposite in direction. It is obvious that those currents cannot generate a radiating field. After the phase angle passes 90° , amplitudes of current at all places start to decrease proportionally. They become 0 again when the phase angle is 180° , as shown in Figure 3.4c. They reach another extreme at 270° , as shown in Figure 3.4d. Comparing Figure 3.4b and d, it can be seen that although currents in both status reach their extreme, the directions of the current are all reversed.

So far, we have stated that a matching transmission line does not radiate and neither does an open-circuit transmission line. It is easy to prove that a short-circuit transmission line or a transmission line with any kind of loading could not radiate.

To design something to radiate, we have to find some structures which can carry currents, and far fields generated by these currents do not totally cancel each other. Figure 3.5 shows a dipole antenna fed by a parallel wire. Actually, this structure can also be converted from an open-circuit transmission line. By bending wires at the open end of the transmission line by 90° and keeping them apart, we get the structure shown in Figure 3.5. To decide the current on the dipole antenna, we need Kirchhoff's current law, which states that in an electric circuit the algebraic sum of all the currents flowing out of a junction is zero. Without loss generality, we can assume the current on the top wire is going toward the right. Based on Kirchhoff's current law, the current direction on the top arm of the dipole is up. Because the currents are always in the reverse direction on wire pairs of a transmission line, the current on the bottom wire is to the left. Then the current direction on the bottom arm of the dipole can also be decided and this is up.

It is now clear that the current directions at any given location on both arms of the dipole are all the same. Instead of canceling each other out, fields generated by currents enforce each other in the outer space. By bending the end of transmission line, we obtain an antenna. To be able to radiate effectively, a basic dipole antenna has to meet another requirement, which is that the total length, L , needs to be half a wavelength.

Figure 3.6 shows a simulation result of a dipole antenna. The antenna's length, L , is designed to be half a wavelength at f_0 . In the simulation, a 50Ω is used as the source impedance of signal generator. The simulated frequency range is from 0 to $8 \times f_0$. The Y -axis in Figure 3.6 is the reflection coefficient, for which the lower the better. There are two X -axes. The upper X -axis is the relative frequency normalized by f_0 , which is from 0 to 8. Because the frequency is

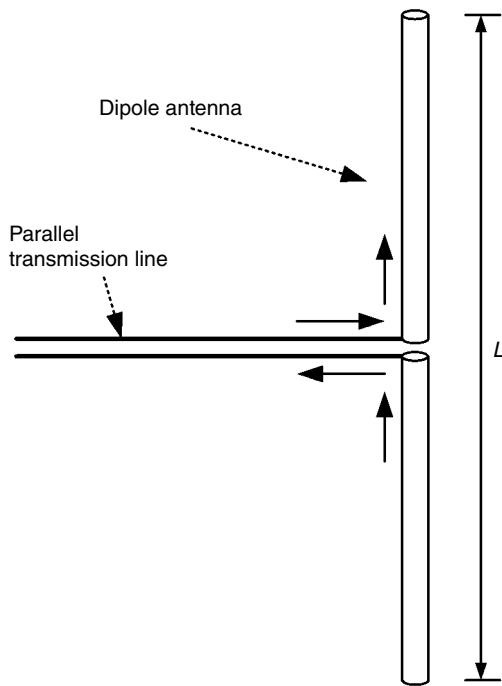


Figure 3.5 Dipole antenna.

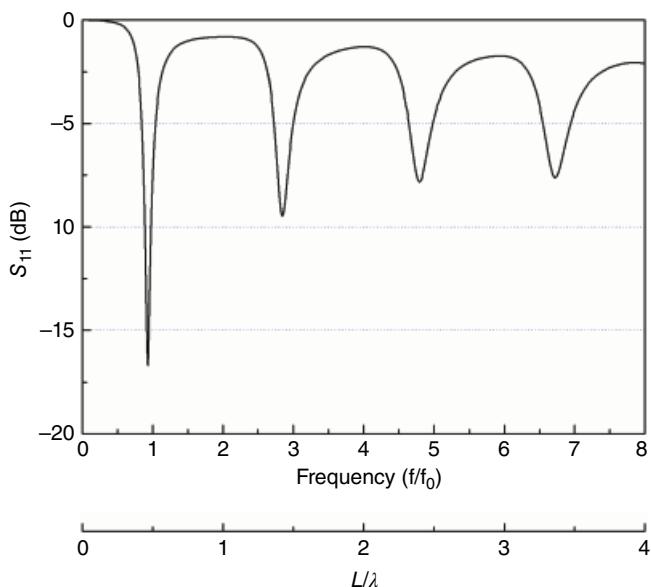


Figure 3.6 Reflection coefficient of a dipole antenna.

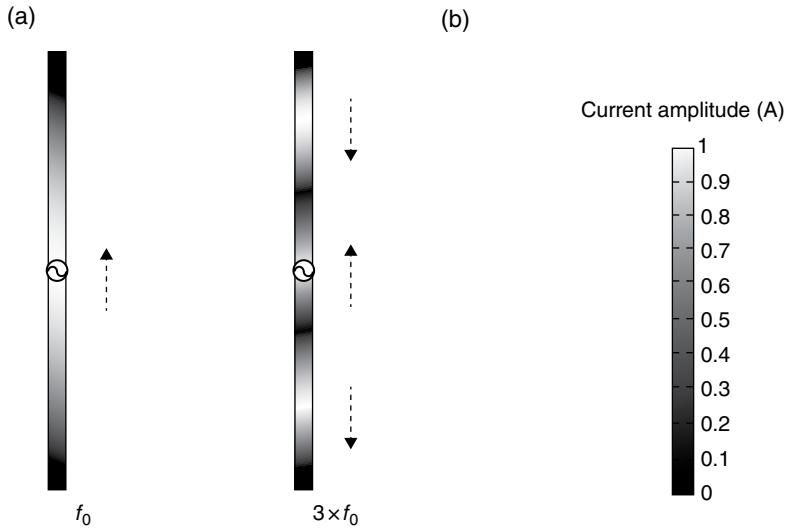


Figure 3.7 Current distributions.

inversely proportional to the wavelength, the X -axis can also be represented by the wavelength. The lower X -axis is the ratio between the antenna length L and wavelength λ .

It can be seen from Figure 3.6 that a dipole antenna cannot cover all frequency bands. It only periodically achieves good matching. Here, a reflection coefficient lower than -5 dB is considered good matching. Based on the theory with some simplifications, besides the base frequency f_0 , a half-wavelength dipole also radiates well at $3 \times f_0$, $5 \times f_0$, and any odd multiple of the f_0 . That can be verified by the simulated results shown in Figure 3.6. One might note that the resonant frequencies of simulated result are around but not exactly at the odd multiple of f_0 . That is the usual difference between theory and practice. In theory, everything is perfect. In reality, the end of the dipole antenna always has some distributed capacitance which effectively loads the antenna and makes the resonant frequency lower. However, in practice, it is not really an issue at all; we can just tune the antenna length to get the right frequency.

Figure 3.6 can also be described from the wavelength point of view. When the dipole length is half a wavelength, it is well matched. When the length of a dipole is 1.5λ , 2.5λ , and any odd multiple of half a wavelength, it can also effectively radiate. Shown in Figure 3.7 are current distributions of a dipole antenna at f_0 and $3 \times f_0$. The rightmost graph is the color bar of current amplitude. At frequency f_0 , there is only one standing wave. The current's amplitude is 0 at both ends of the antenna and it reaches the maximum in the middle, where the feed point is. At frequency $3 \times f_0$, there are three standing waves. The current's amplitude reaches the maximum at three locations. As shown in Figure 3.7b, the current on the antenna alternatively changes directions. Based on these two current distributions, it should be easy to work out distributions of other higher order modes.

The phenomenon of multiple resonances of a single antenna is very important and is widely used in designing multiband antennas. This topic shall be discussed in later sections. For now, we'll focus on the lowest resonant frequency f_0 .

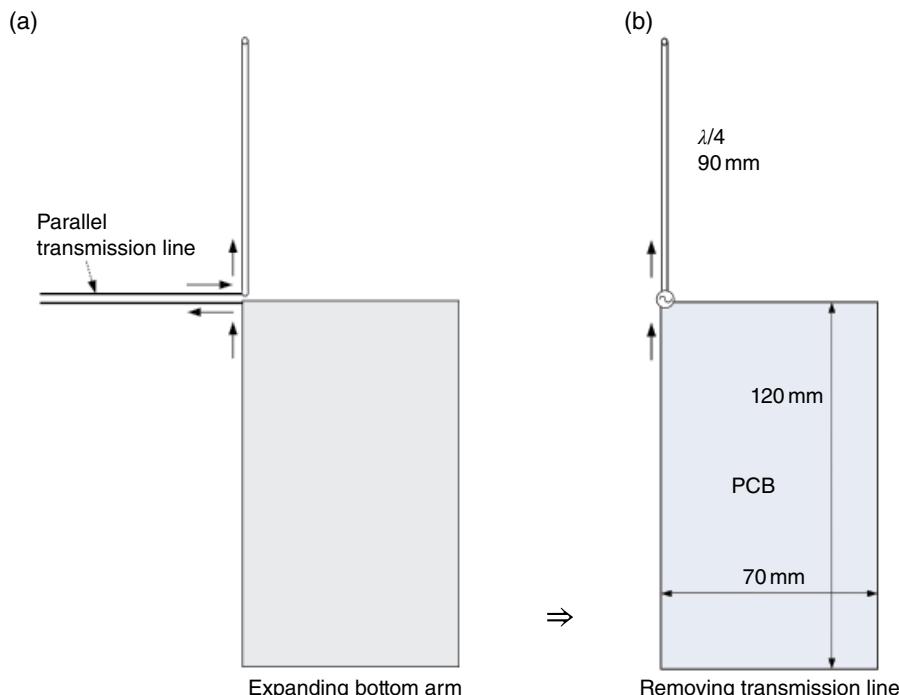


Figure 3.8 Transformation to a monopole antenna.

Although a dipole used to be a widely adopted antenna form factor in the wireless communication industry, it is too cumbersome to be used in today's mobile phones. As a quantitative example, a dipole needs to be 182 mm long at 824 MHz, which is the band edge of the US cellular band. This length is much longer than any phone currently available on the market, so we definitely need something shorter.

Using the dipole antenna shown in Figure 3.5 as the starting point, we can get the structure shown in Figure 3.8a by expanding the bottom arm of a dipole antenna. Then we can remove the transmission line and replace it with a signal source directly placed between the top "normal" arm and the bottom "fat" arm. The final transformed new antenna structure is shown in Figure 3.8b. It may still look like a dipole antenna, but the name of the antenna changes to monopole antenna now, because from the point of view of a system engineer or a project manager, the bottom arm does not belong to antennas. In a real phone, the bottom arm is actually a piece of the printed circuit board (PCB), where all the circuit components, such as all kinds of integrated circuits (ICs), resistors, capacitors, and inductors are mounted. Except for a few matching components, an antenna engineer can change practically nothing on that board. The board's specific dimensions, such as length, width, and thickness, are all decided by engineers from other disciplines.

A PCB board normally has a multilayer structure. It is composed of layers of insulating sheet and copper traces laminated by pressing processes. In the mobile phone industry, four-layer, six-layer, and eight-layer boards are widely used. The number of layers shows how many layers of copper traces it has. To eliminate noise and electromagnetic compatibility

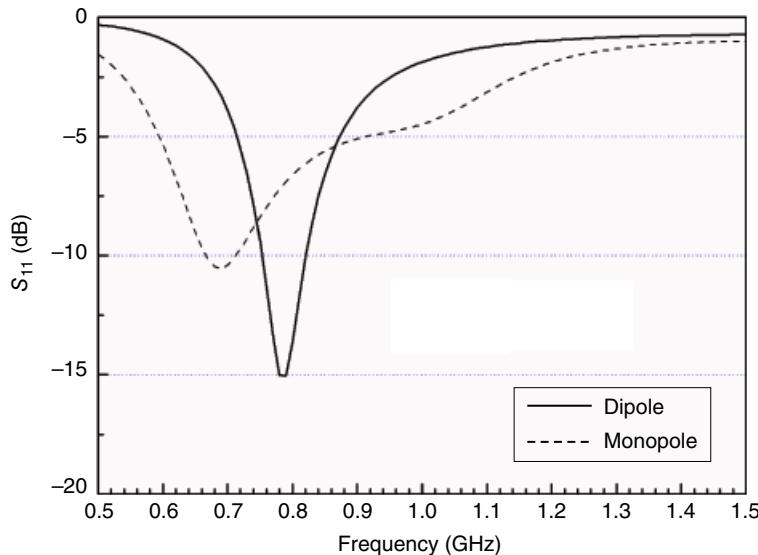


Figure 3.9 Reflection coefficients of a 180 mm-long dipole and a 90 mm-long monopole.

issues, normally there is a ground layer inside the PCB lamination. From the antenna's point of view, the ground layer is the dominant radiator and the whole board can be simplified as one piece of metal.

In a structure as shown in Figure 3.8b, only the top arm is considered an antenna, so the total length of a monopole antenna is a quarter of a wavelength instead of the half a wavelength required by a dipole antenna. But we must always remember that the PCB is the other half of an antenna. Without a PCB, any monopole antenna would not work. The size of a PCB should be larger than an antenna; otherwise, the antenna might have problems at low frequency.

For the purpose of comparison, the simulated results of a dipole and a monopole antenna are shown in Figure 3.9. The total length of the half a wavelength dipole is 180 mm, and each arm is 90 mm. The resonant frequency of the dipole antenna is 790 MHz, which is lower than the 830 MHz predicted by the half a wavelength assumption. The length of the quarter of a wavelength monopole antenna is 90 mm. The PCB on which the whip monopole is installed is simulated by a perfect conductive sheet with the dimensions of 70 mm × 120 mm as shown in Figure 3.8b. Although both the dipole and the monopole have identical top arms, the resonant frequency of the monopole antenna is 690 MHz, which is lower than the dipole's. That is due to the effect of the PCB. Compared with a dipole antenna which has a 90 mm-long bottom arm, the PCB is 120 mm long and it makes the whole radiating structure bigger and drives the resonant frequency lower.

Although a whip monopole antenna is only half the size of a dipole antenna, it is still too big to be used as a fixed antenna on a mobile phone. Now it is the time for helix monopole antennas to enter the stage. As we know that the resonant frequency of a monopole antenna is roughly decided by the length of wire, it is quite natural to think about coiling the wire round to make a helix. Figure 3.10 shows a helix. The critical dimensions to define a helix include

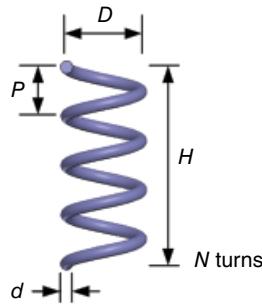


Figure 3.10 Critical dimensions of a helix.

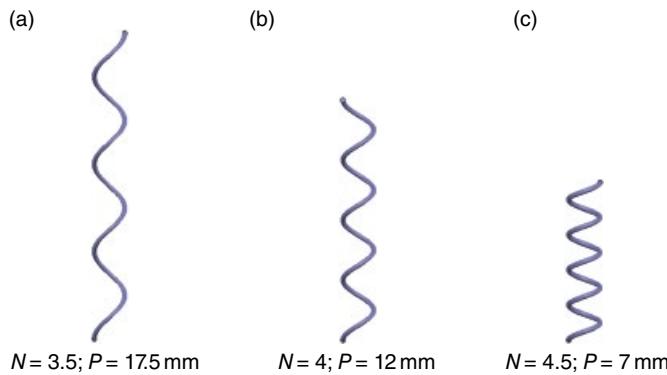


Figure 3.11 Three 90 mm-long helices with the same diameter of 6 mm.

the pitch P , the diameter of the helix D , the diameter of the metal wire d , and the total number of turns N .

The total wire length L_{total} of a helix can be expressed by Equation 3.2:

$$L_{\text{total}} = N \times \sqrt{(\pi \times D)^2 + (P)^2} \quad (3.2)$$

The total height of a helix H can be expressed by Equation 3.3:

$$H = N \times P \quad (3.3)$$

By adjusting D , P , and N , any required length can be obtained. However, even with the same wire length, different helices have different frequency response. Some parametrical studies will be presented in Figures 3.11, 3.12, 3.13, 3.14, 3.15, 3.16, and 3.17 to demonstrate the effect of different parameters.

Figure 3.11 shows three 90 mm-long helices with the same helix diameter of 6 mm and wire diameter of 1 mm. It can be seen from Equation 3.2 that if the diameter of the helix is fixed, only the pitch and number of turns can be adjusted. The helix (a) has 3.5 turns and a pitch of 15.5 mm. The helix (b) has 4 turns and a pitch of 12 mm. The helix (c) has 4.5 turns and a pitch of 7 mm. The total height of helices (a), (b), and (c) is 61.3, 48, and 31.5 mm, respectively.

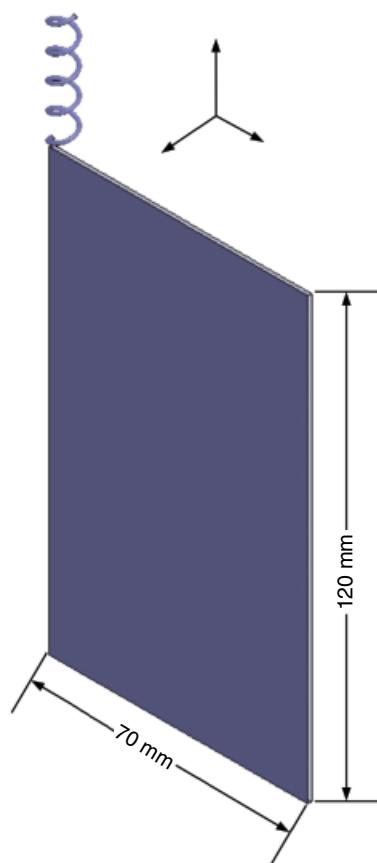


Figure 3.12 Helix (c) installed on a PCB.

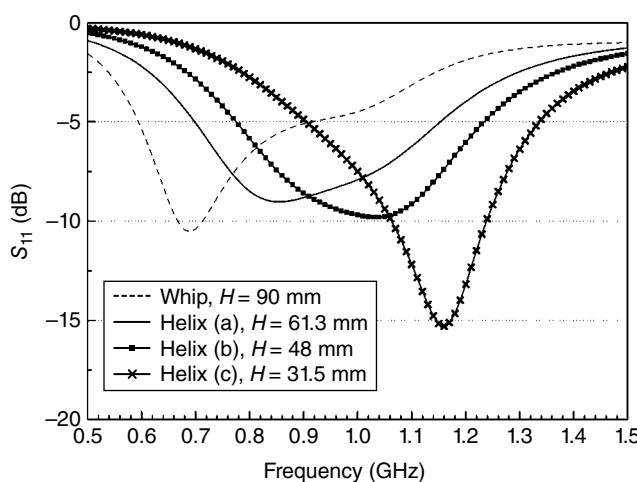


Figure 3.13 Reflection coefficients of a whip monopole and three helixes, $D=6$ mm.

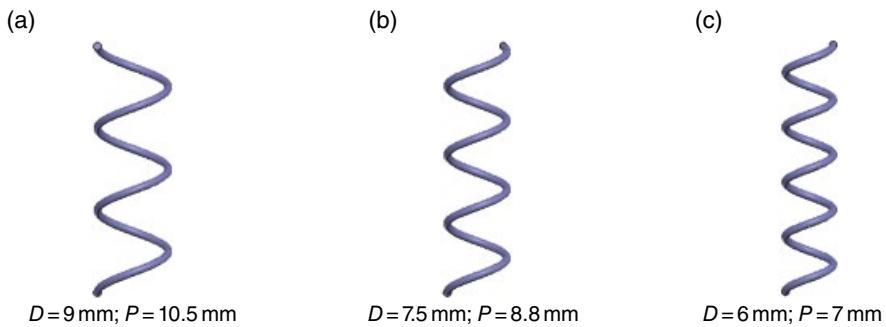


Figure 3.14 Three 90 mm-long helices with the same height of 31.5 mm.

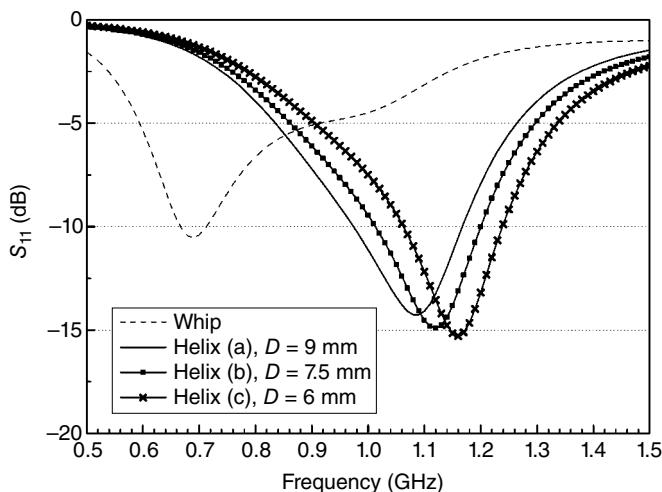


Figure 3.15 Reflection coefficients of a whip monopole and three helices, $H=31.5 \text{ mm}$.

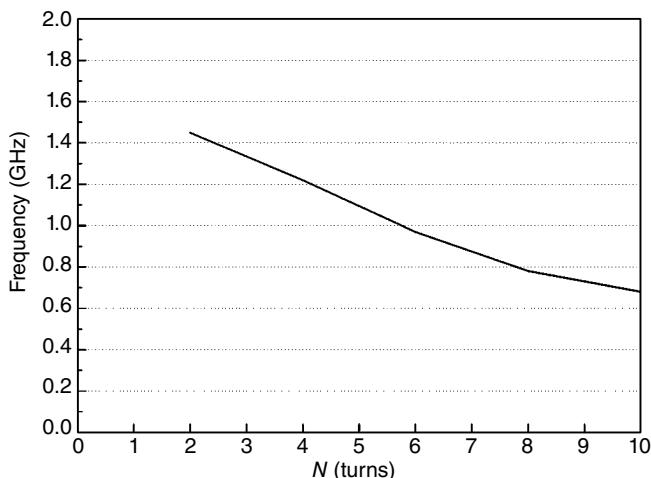


Figure 3.16 Resonant frequency versus number of turns, $H=31.5 \text{ mm}$ and $D=6 \text{ mm}$.

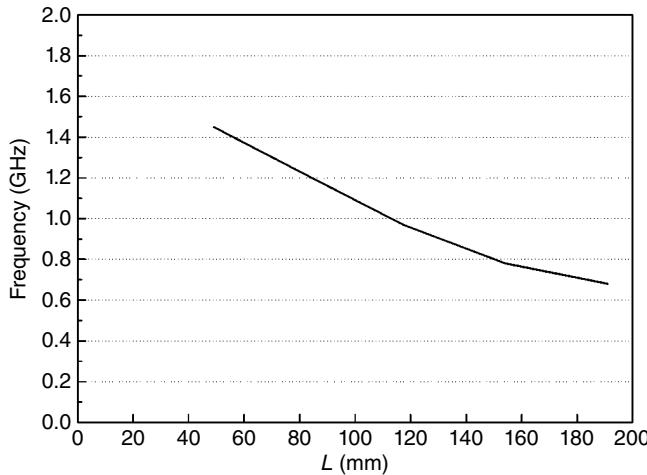


Figure 3.17 Resonant frequency versus length of helix, $H=31.5$ mm and $D=6$ mm.

Figure 3.12 shows the drawing of the helix (c) with a PCB, which is identical to the one shown in Figure 3.8b. All three helices are simulated with the same PCB.

Figure 3.13 is the simulated results of a whip monopole and three helixes when installed on the same PCB. The result of whip monopole is the same as shown in Figure 3.9. Although all four antennas have the identical wire length, 90 mm, their resonant frequencies are different. With the decrease of an antenna's total height, the resonant frequency of the antenna shifts higher. However, the resonant frequency of a helix antenna is still much lower than a whip monopole with the same height. For example, the helix (c) resonates at 1.16 GHz and a 31.5 mm tall whip monopole resonates at 1.6 GHz.

Figure 3.14 shows three 90 mm-long helixes with the same height of 31.5 mm and wire diameter of 1 mm. The diameters of helixes (a), (b), and (c) shown in Figure 3.14 are 9, 7.5, and 6 mm, respectively. By fixing the height and the diameter of the helix, there is only one combination of pitch and number of turns which can produce the required total length. In Figure 3.14, the pitch of the helix (a) is 10.5 mm and the number of turns is 3. The pitch of the helix (b) is 8.8 mm and the number of turns is 3.58. The pitch of the helix (c) is 7 mm and number of turns is 4.5. The helixes (c) shown in both Figures 3.12 and 3.14 are identical.

Figure 3.15 is the simulated results of four antennas. The results of the whip antenna and helix (c) are identical to those shown in Figure 3.13. They are included for the convenience of comparison. Once again, four antennas with the same length have different resonant frequencies. When the height is fixed, an antenna with larger helix diameter has lower resonant frequency. However, the effect of helix diameter is much less than the effect of helix height. By increasing the helix diameter from 6 to 9 mm, the total volume of the helix is more than doubled. The frequency variation due to that amount of volume change is only 6.5%. When the helix height changes from 31.5 to 61.5 mm, the volume change is less than two times, but the frequency varies 36%.

So far, we have studied several different helixes under one constraint, which is that the total length of helix is always 90 mm. As a summary, whenever we try to coil a fixed-length wire up to make a helix, the smaller the helix, the higher the resonant frequency. If the volumes of two

same-length helixes are also the same, the resonant frequency of the taller and thinner one is always lower.

The ultimate goal when designing a helix antenna is to replace a whip monopole antenna, so a helix antenna must have the same resonant frequency as the one it needs to replace. To compensate for the frequency shift that any helix antenna inheres, the total wire length of a helix should be longer than the whip one it replaced. When the height and diameter of a helix are fixed, its resonant frequency can be tuned by adjusting the helix's turns. Based on Equation 3.3, its pitch also needs to be inversely proportionally adjusted.

As an example, a helix antenna with a height of 31.5 mm and a diameter of 6 mm is parametrically studied. By proportionally adjusting the number of turns and the pitch, different resonant frequencies can be obtained. Shown in Figure 3.16 is the simulated result of resonant frequencies versus number of turns. By adjusting the helix's turns from 2 to 10, the resonant frequency changes from 1.45 to 0.68 GHz. To replace a 90 mm whip monopole antenna, which is resonant at 0.69 GHz, a helix antenna needs roughly 10 turns when $H=31.5$ mm and $D=6$ mm.

Figure 3.17 shows the simulated result of resonant frequencies versus wire's length of helixes. The simulated cases used in Figure 3.17 are identical to those used in Figure 3.16. However, the X-axis in Figure 3.17 is replaced by the helix's length calculated by Equation 3.2. It is clear in Figure 3.17 that to obtain the 0.69 GHz resonant frequency, the total wire length of the helix is 190 mm, which is more than twice the length of a 90 mm whip monopole.

In the real world, to design a whip monopole or a helix monopole antenna, one really does not need any parametrical study. To design a whip monopole, just cut a quarter-wavelength copper wire and solder it to the device. The resonant frequency should be lower than the theoretical one. That is what we want. We always start the tuning process of a monopole antenna at a lower frequency. Lower frequency means a longer radiator. It is much easier to cut copper wire off to shift the frequency up than to put wire back in.

When designing a helix antenna, the height and the diameter of the helix are normally already decided by industrial design (ID) engineers. For those not familiar with various disciplines of engineering, ID is a combination of applied art and applied science, whereby the aesthetics and usability of mass-produced products may be improved for marketability and production. They are the number one "enemy" who make antenna engineers suffer. They always want to remove any cosmetic trace of the existence of an antenna, and they are the ones who want to wrap the whole antenna with a shiny stainless steel cover. The "enemy" number two is mechanical engineers. They are responsible for creating the cosmetic design generated by ID engineers. Due to the consistent pressure from the marketing department to make mobile phones smaller and slimmer, mechanical engineers always find themselves in an embarrassing situation; it seems that there is never enough space to squeeze all the stuff in. Naturally, mechanical engineers will conspire to "steal" space allocated to antennas. To be an antenna engineer, part of the job's responsibility is to convince all the others that there is an undisputed reason for the reserved antenna volume and to communicate the minimum volume that must be retained in order to achieve the required antenna performance.

Similar to a whip antenna, to design a helix antenna, we can start with a half a wavelength wire, and then coil it to the wanted diameter and stretch it to the required height. This should generate a resonant frequency lower than what we want. The tuning process is cutting a little of the helix and stretching it to the correct height. Repeat the process until the correct frequency response is obtained.

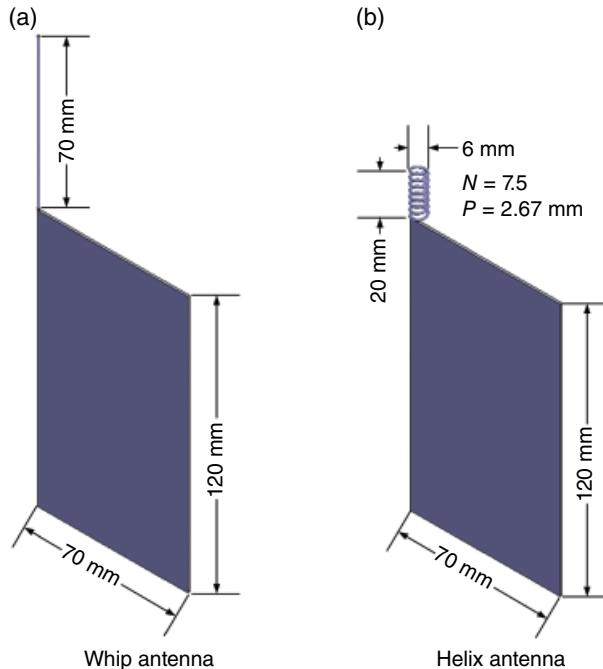


Figure 3.18 Dimensions of two antennas designed to cover 824–894 MHz.

As a demonstration, we will design two single-band antennas, a whip and a helix, for phones used in the US market. That means both antennas need to have a working frequency band covering 824–894 MHz. A reflection coefficient less than -10 dB is used here as the specification to calculate the bandwidth. Through the cutting and measuring processes, the dimensions of the whip and the helix can be decided. The final dimensions of both antennas are shown in Figure 3.18. Both antennas are made of 1 mm copper wire. The whip antenna is 70 mm tall. The dimensions of the helix antenna are $H=20$ mm, $D=6$ mm, $P=2.67$ mm, and $N=7.5$ turns. The total wire length of the helix is 142 mm.

Shown in Figure 3.19 are the reflection coefficient results of both the whip antenna and the helix antenna. The reflection coefficient of the whip clearly shows a double resonance. The explanation for this phenomenon can be found later in Section 3.5. The gray line segment is the specification line, which is -10 dB from 824 to 894 MHz. The return loss of the helix meets the specification. All return loss value of the whip antenna is above -10 dB, which means the original whip antenna cannot meet the specification even at a single frequency point.

Does Figure 3.19 give us enough information to make the judgment on whether the helix antenna has a better performance than the whip antenna? Not really. When we talk about an antenna's performance, we actually refer to both its efficiency and its bandwidth. To evaluate an antenna's efficiency, an anechoic chamber must be used. To predict an antenna's bandwidth, the locus of impedance shown in a Smith chart is required. Shown in Figure 3.20 are the impedance's loci of both antennas. Two band edges, 824 and 894 MHz, are marked in Figure 3.20a and b. It is clear that the bandwidth of both antennas can be improved by adding matching circuits.

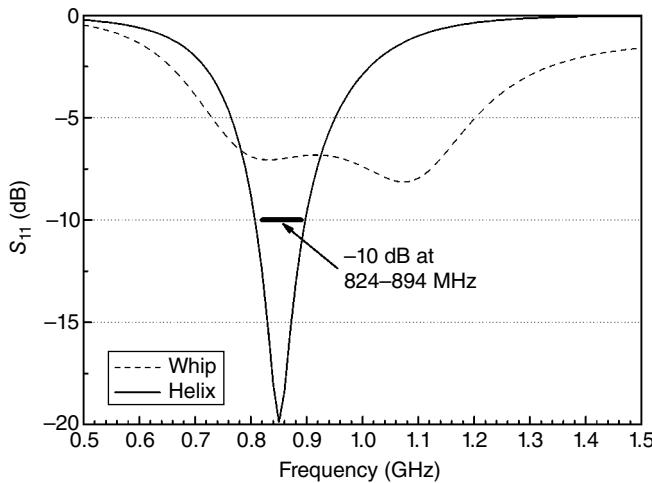


Figure 3.19 Original reflection coefficients of whip and helix antennas.

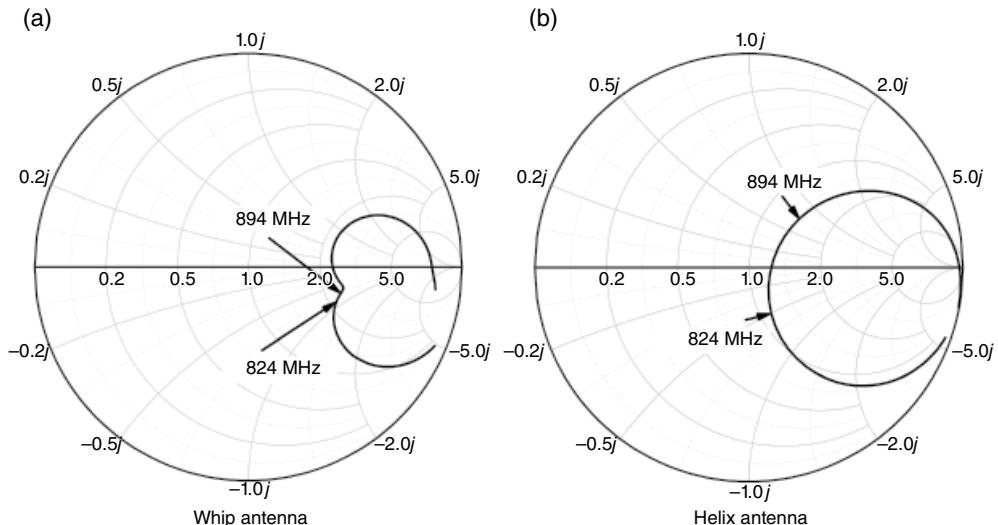


Figure 3.20 Impedances of whip and helix antennas.

Among the many factors which decide an antenna's achievable bandwidth, the following two are the most important and have the most influence:

1. The length of impedance locus between two band edges. The shorter the length, the easier it is to achieve wide bandwidth. What a normal matching circuit does is move the locus of impedance around on the Smith chart. If an antenna has long impedance's locus, whenever one of the band edges is matched, the other band edge moves out of the center of the Smith chart. Of course, the bandwidth expanding techniques introduced in Chapter 2 can be used

to bend an impedance locus, thus achieving a wider bandwidth. As shown in Figure 3.20a, the locus length of the whip antenna between 824 and 894 MHz is quite short, so it is quite easy to significantly improve the whip's bandwidth by a matching network. On the contrary, the locus length of a helix antenna is relatively long; thus, only limited improvement can be achieved through matching.

2. The distance from impedance locus to the center of the Smith chart. This distance actually is the amplitude of reflection coefficient Γ introduced in Chapter 2. The center of the Smith chart corresponds to $|\Gamma|=0$. The outermost boundary of the Smith chart corresponds to $|\Gamma|=1$, which means total reflection. It is obvious that the lower the Γ , the easier it is to achieve wide bandwidth. To move the impedance further on a Smith chart, the frequency response of the matching circuit is steeper; as a consequence, the antenna bandwidth is narrower. It must be remembered that the distance to the center is not linearly related to the achievable bandwidth. Any impedance on the circle of $|\Gamma|=0.5$, which has a reflection coefficient of 6 dB, is still pretty easy to match. As a comparison, the achievable bandwidth of an impedance on the circle of $|\Gamma|=0.9$ is extremely narrow.

Combining the impact of both factors, it is safe to say that the whip antenna has a much wider bandwidth than the helix antenna.

Two circuits shown in Figure 3.21 are used to match the whip antenna and the helix antenna, respectively. Both of them are simple two-element matching. For fairness of comparison, both circuits use the same topology.

Shown in Figure 3.22 are the reflection coefficient results of the matched whip and the matched helix. With a matching circuit, the -10 dB bandwidth of the helix antenna increases from 80 to 110 MHz. Meanwhile, the bandwidth of the matched whip increases to 560 MHz, and it covers 0.74–1.3 GHz.

It is now clear that it is not a free lunch to compress an antenna into a smaller volume. With the decrease of the antenna's volume, the bandwidth shrinks rapidly. For a helix antenna, decreasing the diameter D or the height H shrinks its bandwidth. However, decreasing a helix's height has more impact on its bandwidth.

Some might have noted that there are dual resonances in the whip's frequency response. This is due to both its unsymmetrical structure and its wideband characteristic. The whip and the PCB are two equivalent arms of a dipole antenna. They are different in dimension, so each of them has their own resonant frequency. When both resonances have a wideband

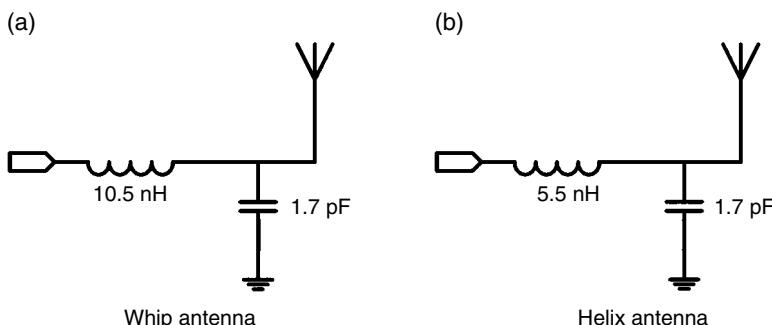


Figure 3.21 Matching circuits of whip and helix antennas.

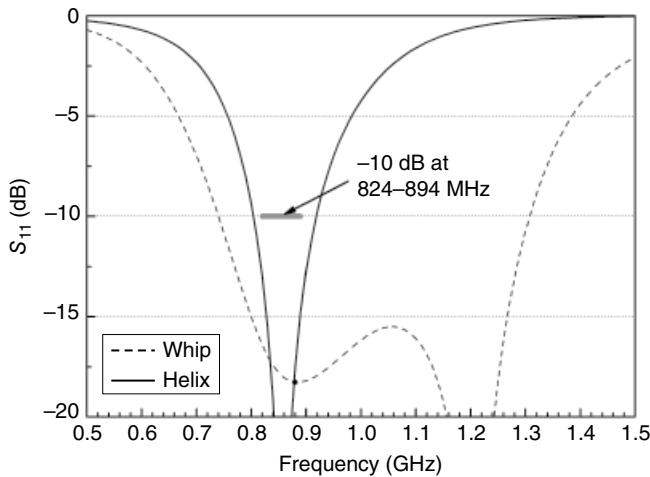


Figure 3.22 Reflection coefficients of whip and helix antennas after matching.

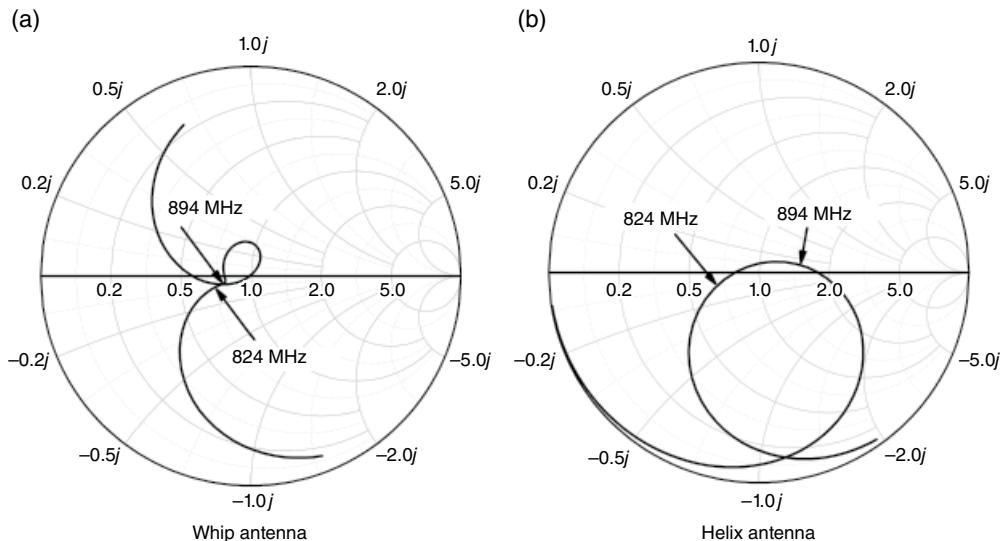


Figure 3.23 Impedances of whip and helix antennas after matching.

characteristic, their combined frequency response has dual resonances. That happens to be the case in a whip antenna. In comparison, a helix antenna is also asymmetric, but the bandwidth of a helix is so narrow that its frequency response conceals the resonance of a PCB. This is the reason why a small helix antenna always has a clean impedance locus.

Shown in Figure 3.23 are the impedance's loci of the matched whip and the matched helix. It can be seen that the loci between 824 and 894 MHz of both antennas are well positioned in the center of the Smith chart. As a comparison, the original impedance in Figure 3.20b passes by the center from the right side of the Smith chart. The locus of the

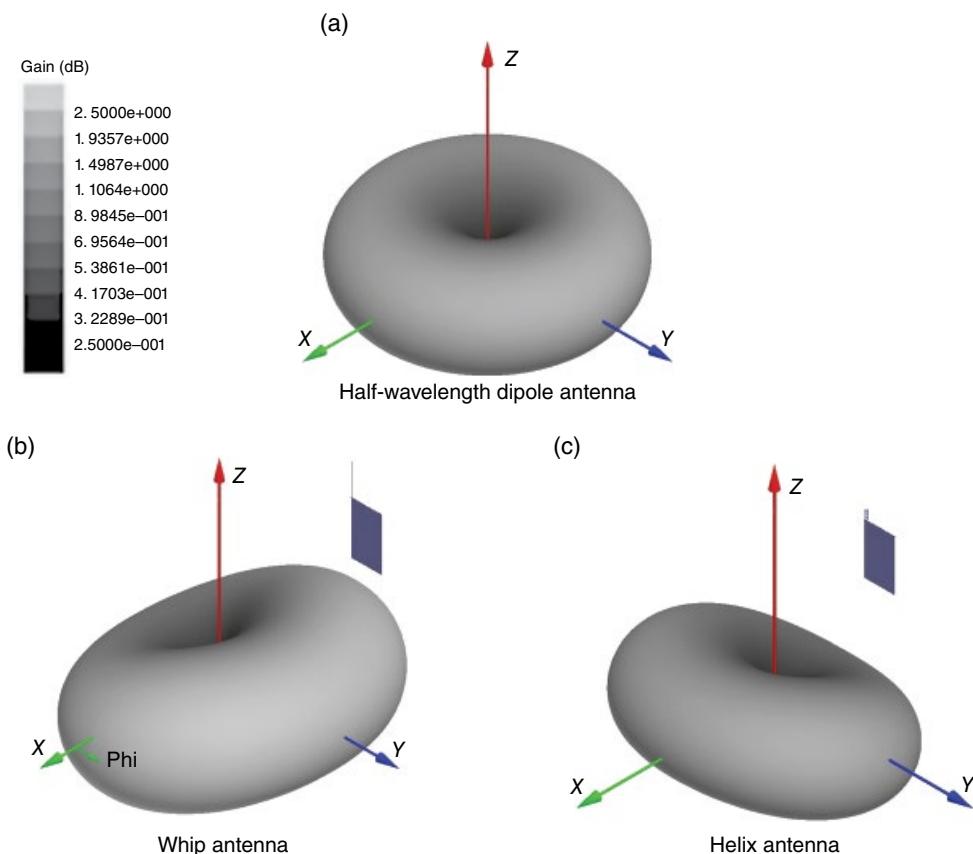


Figure 3.24 Three-dimensional radiation patterns at 850 MHz.

matched helix moves around the center of the Smith chart. That brings the whole locus closer to the center and thus provides wider bandwidth. Due to the wide frequency scale, it may not seem clear in Figure 3.22 that a matching network still improves the helix antenna's bandwidth by about 30%.

Figure 3.24a–c shows three-dimensional (3D) radiation patterns of an ideal half-wavelength dipole antenna, a whip antenna, and a helix antenna, respectively. The axis of the dipole antenna is aligned to the Z-axis. The 3D pattern of the dipole looks like a donut. As shown in Figure 3.24a, due to its axial symmetrical property, a dipole's pattern is also axial symmetrical and has its peak located in the XY-plane. In Figure 3.24b and c, the axis of the whip and the helix are also aligned to the Z-axis. The PCBs of both the whip and the helix are placed in the YZ-plane. Figure 3.24b and c has the same angle of view as Figure 3.24a. It is clear that radiation patterns of both still have a null around Z-axis, but patterns are no longer symmetrical and are slightly tilted to different directions.

For the simplicity of studies, most antennas simulated in the book do not include plastic covers or supports. In practice, these features are necessary to ensure the antennas'

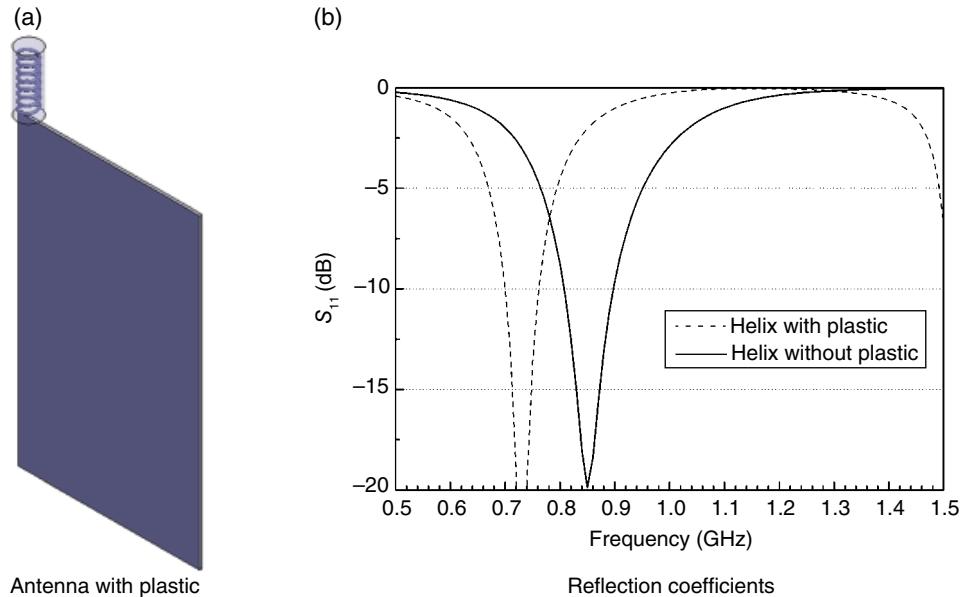


Figure 3.25 Comparison between antennas with and without plastic.

mechanical robustness. The dielectric constant or permittivity of commonly used plastics is around 2.6–4.5. Due to the loading effect of plastics, the resonant frequency of an antenna decreases. This frequency shift can easily be compensated for by making the antenna radiator shorter. The more important issue related to plastic is loss. Different plastics have different loss properties, which has the potential to degrade the antennas' efficiency significantly. If the difference in efficiency of an antenna with plastic and without plastic is more than 1 dB, we need to investigate the plastic. Of course, the antenna in both cases needs to be well tuned.

Shown in Figure 3.25 are the simulation results of a helix antenna with and without a plastic cover. The permittivity of the plastic is 3.2. The diameter of the plastic part is 9 mm and the height is 23 mm. For the purposes of comparison, the result of the antenna shown in Figure 3.18b is also included. In Figure 3.25b, the solid line is the result of the original helix and the dashed line is the response after the helix is insert-molded by plastic. Due to the loading effect, the resonant frequency shifted from 850 to 730 MHz.

The following guidelines are given as the summary of helix antennas' design:

- Whip or helix antennas must be installed on a PCB to work properly.
- The height of a helix antenna is the most important parameter, which decides the antenna's performance. The taller, the better.
- The diameter of a helix antenna also influences its performance. The larger, the better. However, the diameter of the helix's wire has limited impact on its performance. Larger diameter can slightly improve the bandwidth.
- When a plastic support or cover is used, the antenna's resonant frequency shifts lower, and it must be compensated for.

3.1.2 Multiband Helix Stubby Antenna

When the penetration rate of mobile phones started to take off in the early 1990s, it seemed that the allocated frequency bands, 824–894 MHz in the United States and 880–960 MHz in Europe, would soon run out of channels. Countries around the world began allocating new frequency bands to accommodate the rapidly increased demand. Around 1995, Europe allocated 1710–1880 MHz and the United States allocated 1850–1990 MHz as the new personal mobile communication bands, respectively.

As a consequence, single-band antennas finally reached the end of their life span. Any antenna needs to support at least two frequency bands. In the United States, that means 824–894 MHz and 1850–1990 MHz. In Europe, that means 880–960 MHz and 1710–1880 MHz. A world phone, also called a tri-band or quad-band phone, needs to support bands in both continents.

3.1.2.1 Multi-Branch Multiband Stubby Antenna

It has been demonstrated that a whip or a helix can function as an antenna. By putting two such radiating elements together, what we get is a multiband antenna.

Figure 3.26 shows a two-branch stubby antenna. As shown in Figure 3.26b, the antenna is composed of an outer helix and an inner straight wire. Two radiating elements are connected together at the bottom. The heights of the whip and the helix are identical; both are 30 mm long. The helix's diameter is 6 mm, its pitch is 6 mm, and it has 5 turns. Both elements are made of 1 mm-diameter copper wire.

Shown in Figure 3.27 is the simulated reflection coefficient of the antenna. Around 0.9 GHz, it has a clean sharp resonance, which is marked as (1). From 1.7 to 2.9 GHz, the antenna has a

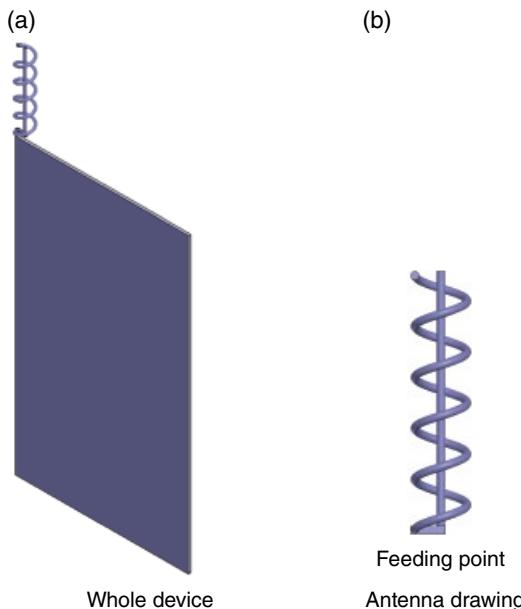


Figure 3.26 Two-branch multiband stubby antennas.

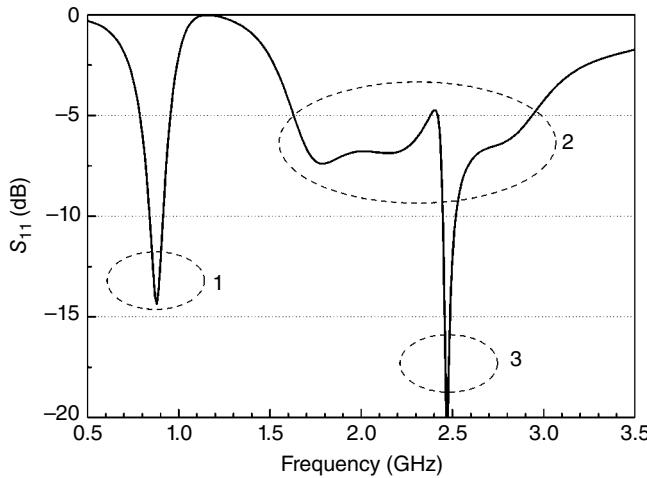


Figure 3.27 Reflection coefficient of a two-branch stubby antenna.

wide response with several resonances, which is marked as (2). At 2.49 GHz, there is another sharp resonance, which is marked as (3).

Shown in Figure 3.28 are radiation patterns of the two-branch stubby antennas. At the low band, which is 0.88 GHz in this case, the radiation pattern is similar to a dipole's. The radiation is mostly generated by the vertical current along the ground plane. The vertical polarized electrical field E_θ is the dominant field component. At the higher band, which is 1.80 GHz, the E_θ and E_ϕ have the same order of amplitude. The field of E_θ is mostly generated by the current along the longer edge of the ground plane and the current on the antenna element. The field of E_ϕ is mostly generated by the current along the top horizontal edge of the ground plane, which explains why the radiation pattern of E_ϕ has a null along $\pm Y$ directions. Shown in Figure 3.28e is the radiation pattern of the total E field at 1.80 GHz. At the high band, the ground plane is relatively large compared with a quarter of a wavelength, thus some portion of current flows down along the vertical edge of the ground in a traveling wave's form factor, which in turn generates a downward-tilt radiation pattern. For cellular phones, it is quite safe to assume that at high frequency bands the peak of radiation pattern always tilts downward, away from where an antenna is installed.

The frequency response of a two-branch stubby has the contributions of both radiating elements. To distinguish the effects of two individual radiators and their mutual impact, the helix and the whip are separately simulated. Shown in Figure 3.29 are simulation results of a whip-only antenna and a helix-only antenna. Both antennas have identical specific dimensions like the antenna shown in Figure 3.26. In the figure, the helix-only antenna has two resonances, which are at 1.1 and 2.9 GHz, respectively. These two resonances correspond to modes of f_0 and $3 \times f_0$. The whip-only antenna has a resonance between 1.5 and 2.8 GHz. Comparing Figures 3.27 and 3.29, we can tell that resonances (1) and (3) in Figure 3.27 are generated by the helix and the resonance (2) is the contribution of the whip. When we put the two elements together, the helix is loaded by the whip, so resonant frequencies of both resonances (1) and (3) are 10% lower than its standalone version. The helix's bandwidth in both bands is also narrowed.

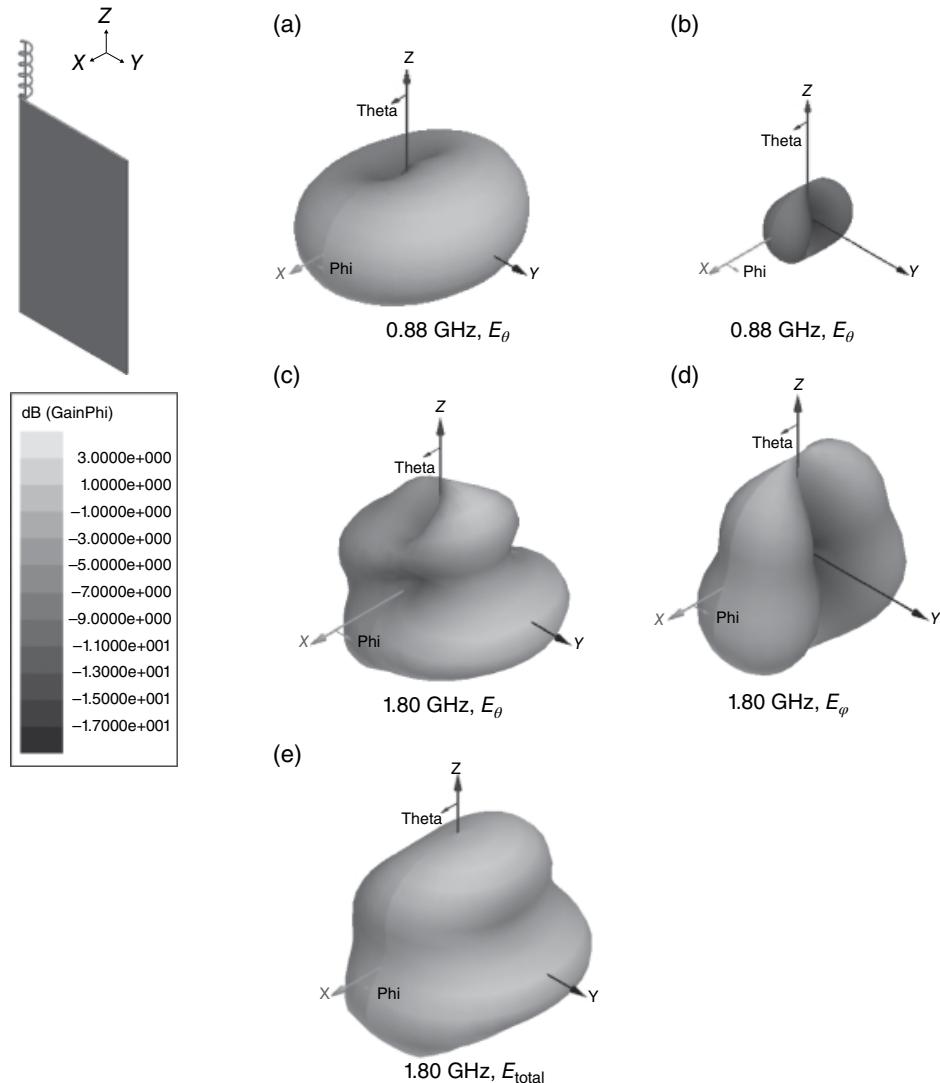


Figure 3.28 Radiation patterns of a two-branch stubby antenna.

Now let's study a little bit more about the resonance (3) shown in Figure 3.27. From the return loss point of view, the helix resonance at 2.49 GHz, where the resonance (3) is found, has the best matching in the 1.7–2.9 GHz band. One might assume that the antenna also has the best efficiency at this frequency point. Unfortunately, this assumption is not correct. At 2.49 GHz, the antenna has the worst efficiency in the band.

Shown in Figure 3.30 is the radiation efficiency of the dual-branch antenna around 2.49 GHz. The simulated frequencies are zoomed in on the 2.0–2.7 GHz range. The radiation efficiency is defined as the ratio between radiated power and accepted power by the antenna. Thus, the

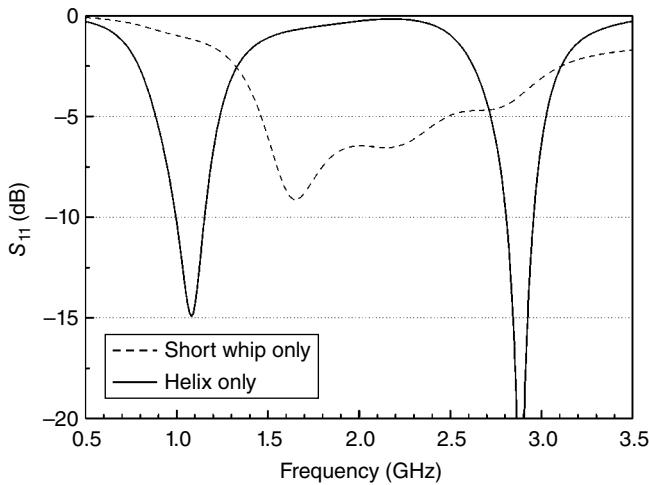


Figure 3.29 Reflection coefficients of a whip-only antenna and a helix-only antenna.

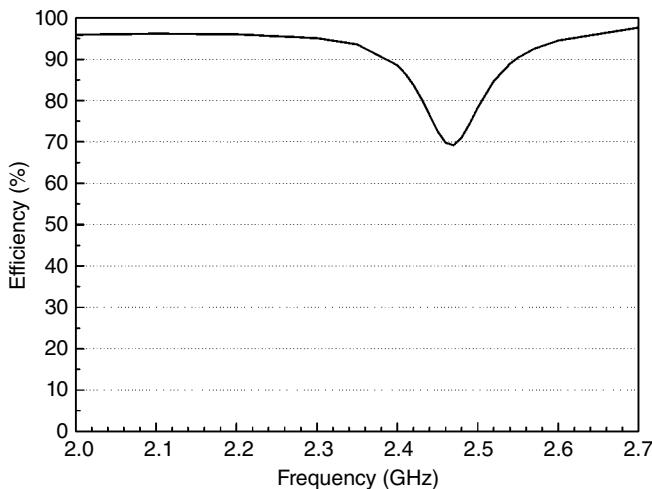


Figure 3.30 Antenna radiation efficiency.

radiation efficiency does not count in the mismatch loss. To simulate the loss effect, a block of lossy material ($20\text{ mm} \times 20\text{ mm} \times 35\text{ mm}$) is used to enclose the antenna. The permittivity of the material is 1 and its loss tangent is 0.01. The worst efficiency occurs at 2.49 GHz. Between 2.45 and 2.55 GHz, the radiation efficiency is less than 80% and the reflection coefficient shown in Figure 3.27 is better than -10 dB . When the frequency is far from the 2.49 GHz, although the return loss value is around -7 dB , the radiation efficiency is better than 95%.

The reason for the low efficiency is not due to the helix element itself. The antenna efficiency is around 95% at 0.9 GHz, which is the first resonant frequency of the helix. The efficiency drop is due to the mutual impact between whip and helix. As shown in Figure 3.29, the whip

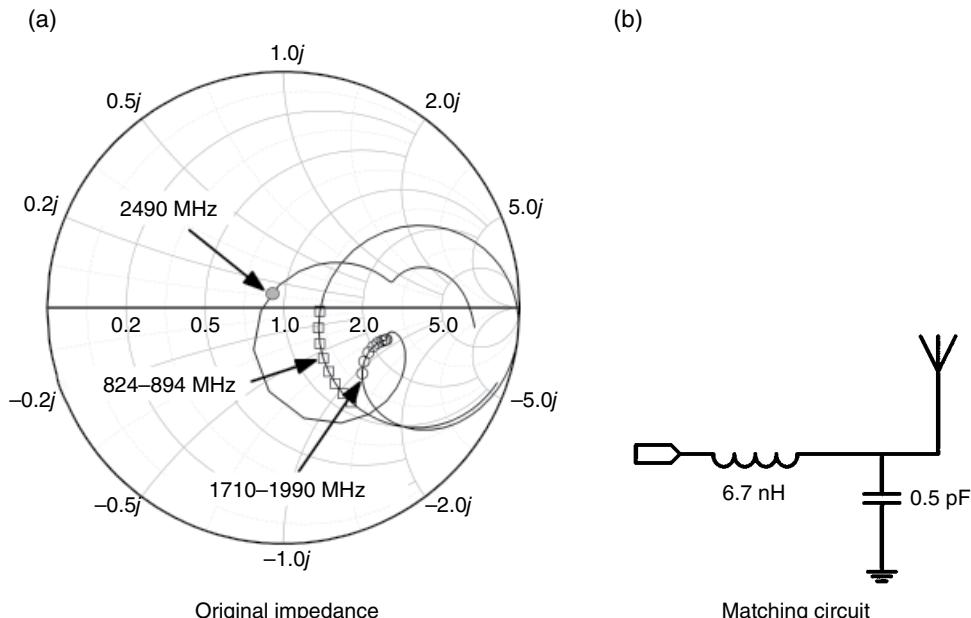


Figure 3.31 Matching a dual-branch stubby antenna.

itself has a quite wide resonance between 1.5 and 2.8 GHz. Outside the close range of 2.49 GHz, the resonance of the two-branch antenna is mostly contributed by the whip. At 2.49 GHz, both the whip and the helix can resonate, which induces a self-resonance between themselves. In a whip and helix structure, the helix has a narrower bandwidth, and it is outside the wideband whip that makes its self-resonance stronger. When self-resonance happens, substantial amounts of energy are stored in the approximate area around the antenna. If an antenna is made of lossy material, which is always the case, this kind of self-resonance dissipates the signal energy on the lossy material.

When designing an antenna, we need to avoid the self-resonance frequency. It will be shown in Section 3.1.2.2 that the second resonance of a helix does not have to be $3 \times f_0$ and we have the means to move the helix's higher order resonance out of the working bands.

Now, let's finalize the design of the dual-branch antenna to provide tri-band coverage. The antenna needs to cover 824–894 MHz, 1710–1880 MHz, and 1850–1990 MHz. Shown in Figure 3.31a is the original impedance locus. The locus marked by hollow rectangular markers is the low band segment and the one marked by hollow circular markers is the high band segment. The filled circle is 2490 MHz, which has the best matching and worst efficiency. The matching circuit for this antenna is shown in Figure 3.31b. The original antenna already has a decent reflection coefficient at the low band, so the matching circuit is mostly for the high band.

Shown in Figure 3.32a is the impedance of the matched antenna. It is clear that the matching circuit moved both the high and low band. However, it moves the high band further. When designing an antenna, we always want to use the simplest matching circuit. That requires an integrated design procedure including both matching circuits and radiating elements. A matching circuit might shift an antenna's resonant frequency, so it is not necessary to

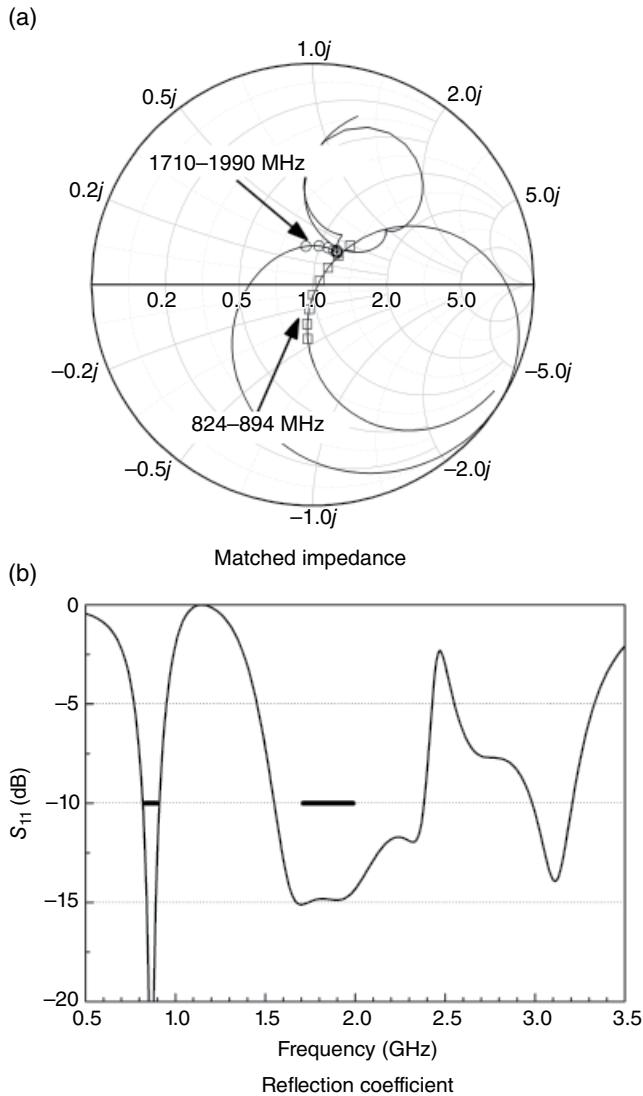


Figure 3.32 Matched dual-branch stubby antenna.

fine tune the original antenna's frequency before a matching circuit is designed. After making the antenna roughly resonate at the right frequency, we can start to design the matching circuit. Then it is time to tune the antenna again. It normally takes a few iterations to get a satisfying design.

The results shown in Figure 3.32a are those obtained after a few iterations of tuning procedures. The 824–894 MHz band is well positioned near the origin of the Smith chart. The best matched frequency in the lower band is 860 MHz. As a comparison, the best matched frequency in the lower band of the original antenna, as shown in Figure 3.31a, is 900 MHz and its low band locus is offset toward the lower half of the Smith chart.

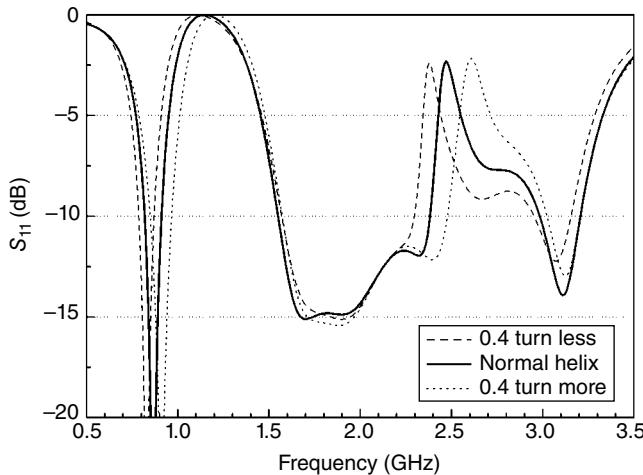


Figure 3.33 Tuning the lower resonance by adjusting the helix.

Figure 3.32b is the simulated reflection coefficient result. The two thick lines mark the -10 dB of 824–894 MHz and 1710–1990 MHz bands. The antenna barely meets the -10 dB specification at the low band and has large margin at the high band.

To tune the lower resonant frequency, we can adjust the total length of the helix. By decreasing the total turns of the helix and increasing the pitch accordingly, we can decrease the total length of the helix while maintaining its height. Shown in Figure 3.33 is a parametric study of the impact of helix length. All three antennas have the same height and use the same matching network. The solid line represents the normal antenna. The dashed line and dotted line represent antennas with helices 0.4 turns shorter and 0.4 turns longer, respectively. It is clear that the helix's length mostly affects the lower band and it does not have too much impact on the 1.6–2.2 GHz band.

One might already have figured out that by adjusting the length of the whip, the higher band can also be tuned. In fact, the effect of the whip is a little bit more complex. As already shown in Figures 3.27 and 3.29, the whip also has a loading effect on the helix. When adjusting the whip, the lower band also shifts, so iterative tuning procedures of both bands have to be performed.

In the following example, we will design a quad-band antenna to cover 824–894 MHz, 880–960 MHz, 1710–1880 MHz, and 1850–1990 MHz bands. The antenna has the same outside dimensions as the tri-band antenna shown in Figure 3.26b. By simply looking at the reflection coefficient of the tri-band antenna shown in Figure 3.32b, we can tell it is not difficult to design a quad-band antenna. To understand why, we must first understand one of the most important trade-offs in antenna design, which is the trade-off between different bands.

To master the art of designing any type of multiband antennas, we must at least learn two techniques. The first is how to generate multiple resonances and how to tune their frequency separately, which we have already discussed. The second is how to trade off the bandwidth between different bands.

Sometimes, one might find an antenna does not meet the specifications in one band and has a huge margin in the other band. This most likely means the antenna's design engineer does

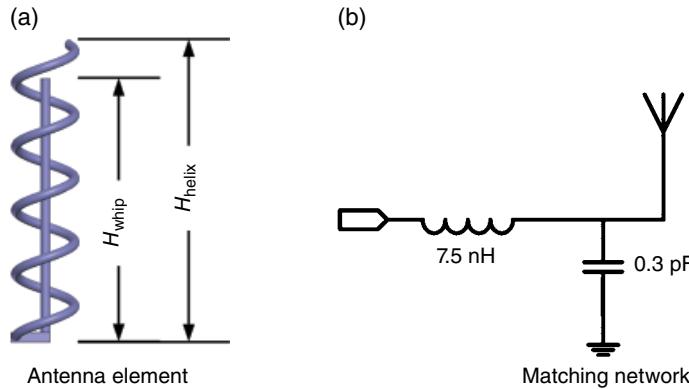


Figure 3.34 Quad-band antenna and its matching network.

not understand the bandwidth trade-off of that kind of antennas. For most antennas, such a trade-off is always possible.

For a dual-branch antenna, the critical parameter when exchanging bandwidth is the height ratio between helix and whip. The height ratio between a whip and a helix is in direct proportion to their bandwidth ratio, as shown in Equation 3.4:

$$\frac{H_{\text{whip}}}{H_{\text{helix}}} \propto \frac{\text{Bandwidth}_{\text{whip}}}{\text{Bandwidth}_{\text{helix}}} \quad (3.4)$$

Shown in Figure 3.34a is the optimized quad-band antenna. The height of the antenna, which is now decided by the height of the helix, is still 30 mm. The length of the whip has decreased to 26 mm. After iteratively adjusting the helix and the matching circuit, a quad-band coverage is achieved. The final helix has 4.6 turns and its pitch is 6.52 mm. The matching network is shown in Figure 3.34b.

Shown in Figure 3.35 is the reflection coefficient of the quad-band antenna. The reflection coefficient of the tri-band antenna is also included for comparison. By decreasing the whip's length, the -10 dB bandwidth of the lower band increases from 70 to 136 MHz. In the meantime, the -10 dB bandwidth of the high band decreases from 830 to 470 MHz.

In fact, the quad-band antenna we have just designed still does not fully exploit the potential of the antenna's volume. There are different ways to expand the working bandwidth of both bands simultaneously. Shown in Figure 3.36a is one such example, it is an antenna with a metal base. The metal base is 10 mm long. The height of the helix is 20 mm long. This antenna has the same outer dimension as both antennas discussed earlier. Shown in Figure 3.36b is the antenna's matching network. Compared with the quad-band antenna, the antenna with a metal base has an even wider bandwidth at both low and high bands.

The following guidelines are given as the summary of multi-branch helix antennas' design:

- The higher and lower resonances of a dual-branch antenna are relatively independent. The whip decides the frequency of higher resonance and the helix decides the lower one.

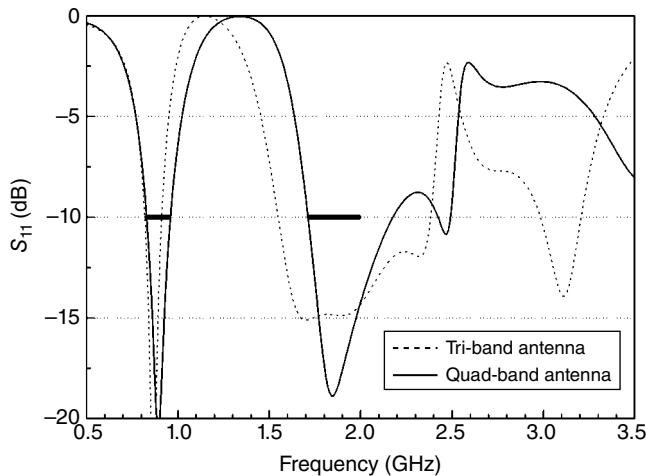


Figure 3.35 Reflection coefficient of the quad-band antenna.

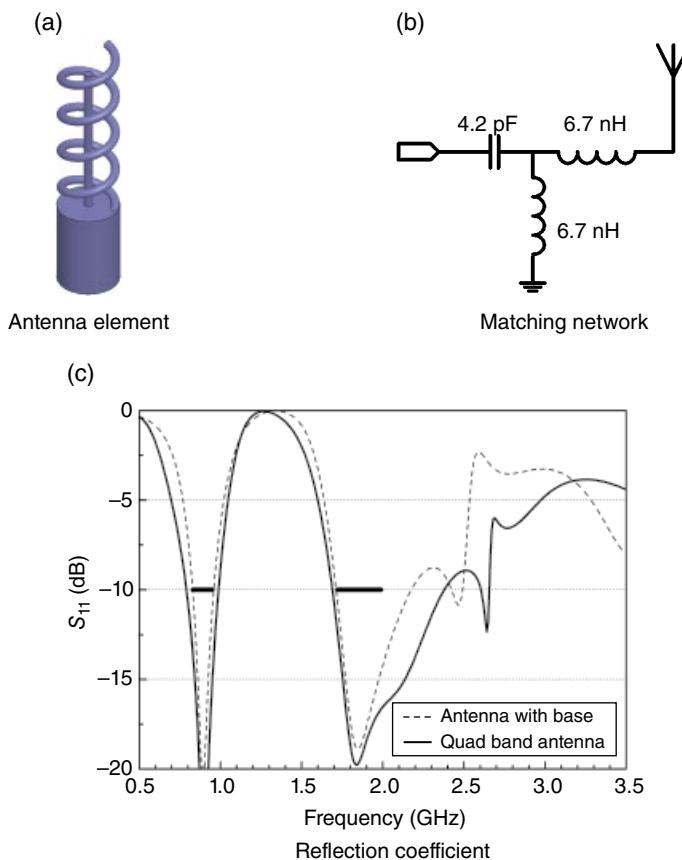


Figure 3.36 Antenna with a metal base.

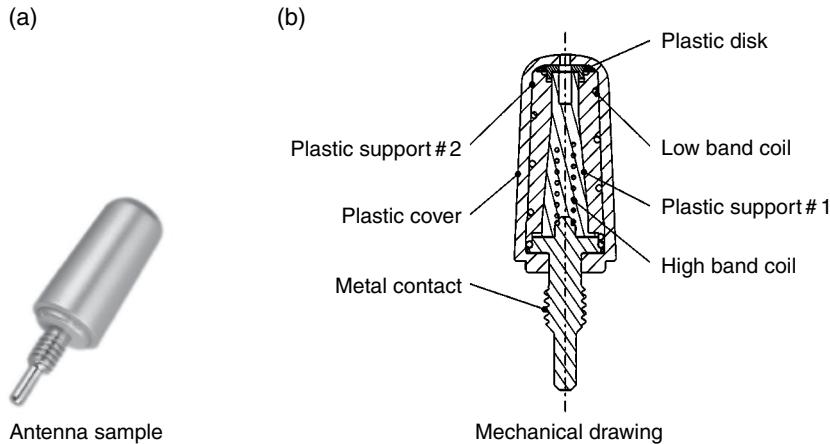


Figure 3.37 Dual-coil helix antenna. (Source: Reproduced with permission of Shanghai Amphenol Airwave.)

- The bandwidth between higher and lower band can be exchanged by adjusting the relative height between the whip and helix. Increasing the height of one can expand its corresponding bandwidth
- The metal base can be used to expand the bandwidth of both bands.

Shown in Figure 3.37 is a dual-coil helix antenna. It is also a multi-branch antenna. The center wire is replaced by a coil. As shown in the mechanical drawing, two coils and the metal contact are effective radiators. All plastic parts are used for the purpose of mechanical durability. By looking at the design, it is easy to tell that more consideration was given to the low band.

3.1.2.2 Single-Branch Multiband Stubby Antennas

In this section, we are going to introduce multiband stubby antennas composed of only one piece of wire. To generate multiple resonances on a single radiator, the higher order mode of that radiator must be used. The concept of a dipole antenna's higher order modes has been detailed in Figure 3.6. The conclusion is that if the resonant frequency of a half wavelength dipole antenna is f_0 , the dipole can also be resonant at $3 \times f_0$, $5 \times f_0$ and all other odd multiples of f_0 .

The higher order mode resonances are not an exclusive property of dipole antennas. A helical antenna also has the similar characteristic. In fact, higher order modes are nothing unique and they exist in all kinds of antennas.

Let's start from dual-pitch antennas. As it is a widely used antenna configuration, several companies hold related patents [5–7]. Shown in Figure 3.38 are three variants of six-turn helices. All of them have same diameter (6mm), total height (30mm), and number of turns. They are all fed from the bottom. Shown in Figure 3.38a is a uniform pitch helix. Shown in Figure 3.38b and c are two dual pitch helices. The dual-pitch 1 helix has 2 turns of large pitch

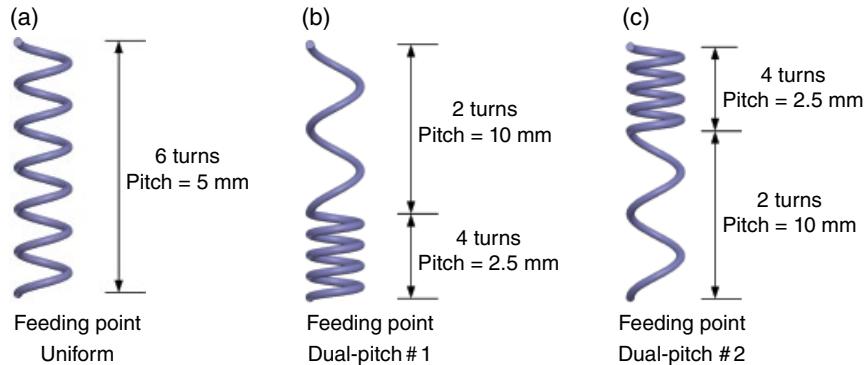


Figure 3.38 Three variants of six-turn helices.

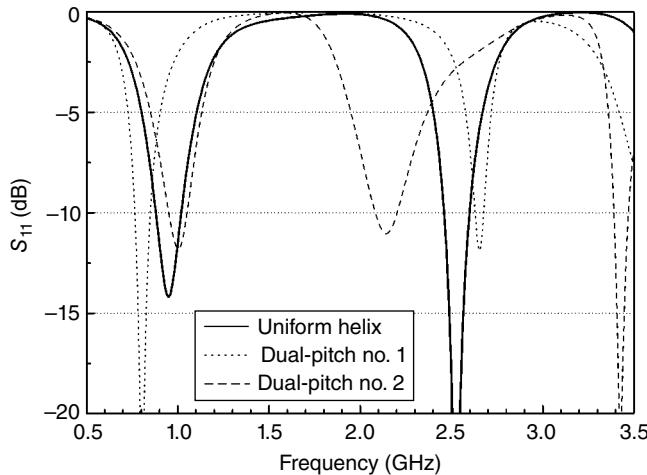


Figure 3.39 Simulated reflection coefficients.

(10mm) and 4 turns of small pitch (2.5 mm). The height of the large pitch and the small pitch segments are 20 and 10mm, respectively. The 2 helix is similar to the 1 except its large pitch segment is at the bottom.

Shown in Figure 3.39 are results of three aforementioned helices. Let's use the uniform helix as the base line. The first resonance of the uniform helix is at 0.95 GHz and the second one is at 2.5 GHz. That means the second resonance is at $2.6 \times f_0$, instead of $3 \times f_0$. This is due to the asymmetric characteristic of the structure, which is composed of both a helix and a PCB. By compressing the bottom portion of a uniform helix and stretching its top portion, we have the 1 antenna. Compared with the uniform helix, the 1 antenna resonates at a lower frequency at the low band and at a higher frequency at the higher band. On the contrary, the 2 antenna has a reverse relationship to the uniform helix. It has a higher resonant frequency at the low band and a lower resonant frequency at the higher band.

Table 3.1 Frequency ratio of different dual-pitch antennas

Antenna type	Ratio between second and first resonant frequency
Uniform	2.6
Dual-pitch 1	3.3
Dual-pitch 2	2.1

Listed in Table 3.1 are the ratios between the second and the first resonant frequencies of all three antennas. It is clear that we can separate the first and the second resonant frequencies by squeezing the bottom portion of a helix. If we squeeze the top portion of a helix, we can make two resonant frequencies closer.

In Figure 3.30, we discussed the efficiency drop of a whip–helix antenna due to the helix’s second resonance. It is obvious that by using a dual-pitch helix, the second resonance can be moved around to avoid any negative impact inside the working band.

In the North American market, a dual-band antenna needs to cover 824–894 MHz and 1850–1990 MHz, which has a relative ratio of 2.2. In the European market, an antenna covers 880–960 MHz and 1710–1880 MHz, the relative ratio is 2.0. Both standards require a ratio less than 2.6, which is the ratio of a uniform pitch helix. So for mobile phone antennas, what we need are dual-pitch helices with a squeezed top portion.

There are many parameters which can be adjusted in the design of dual-pitch antenna. The effect of some parameters is the same for both uniform and dual-pitch helices. Such parameters include the helix’s height, the total wire’s length, the number of turns, and the helix’s diameter. All of them have been investigated elsewhere and are not discussed here.

Now let’s focus on the two most important issues of antenna designing: (1) how to tune frequencies and (2) how to exchange bandwidth. Let’s start with the first one. As we already know, by adjusting the total length of a helix’s wire, the lowest frequency can be changed. To be able to tune both the low and high bands simultaneously, what we need is a way to adjust the ratio between the first and the second resonance. This can be achieved by changing the pitch ratio between two segments.

To illustrate the effect of pitch ratio, three antennas as shown in Figure 3.40 are simulated. All three helices have the dimensions except the pitch ratio. They have 2 turns of large pitch helices and 4 turns of small pitch helices; the total height is 30 mm. There are slight differences in wire length (≈ 117 mm), which can be calculated using Equation 3.2, among three helices. The effect of this length difference can be ignored.

The antenna shown in Figure 3.40b is the same one shown in Figure 3.38c, and it is used as the normal sample. The pitch ratio between the large and small pitches are 8, 4, and 1.75, respectively, for three antennas.

Results of three helices are shown in Figure 3.41. By adjusting the pitch ratio, we are able to change the ratio between the first and the second resonance. The frequency ratios between the second and the first resonances of three helices are 1.7, 2.1, and 2.5, respectively. When the pitch ratio changes, the frequency of the first resonance is relatively stable and the second resonance drifts in a wide range.

Now let’s discuss how to exchange bandwidth between bands. The critical parameter here is the ratio of number of turns between large-pitch and small-pitch portions. All three helices

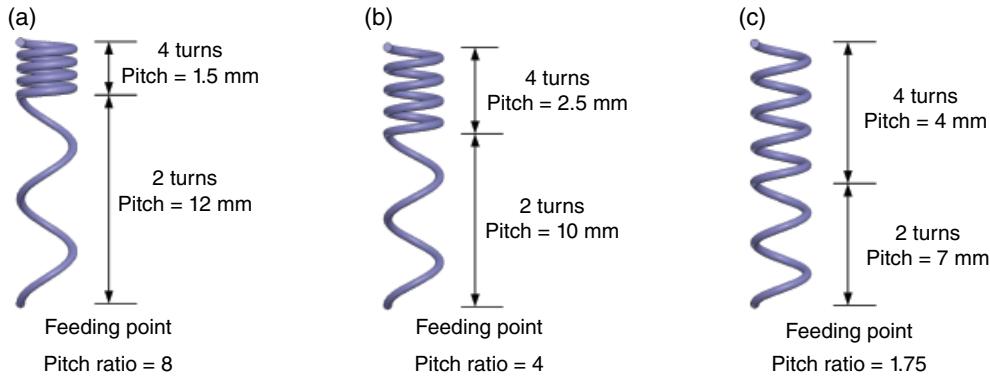


Figure 3.40 Effect of different pitches.

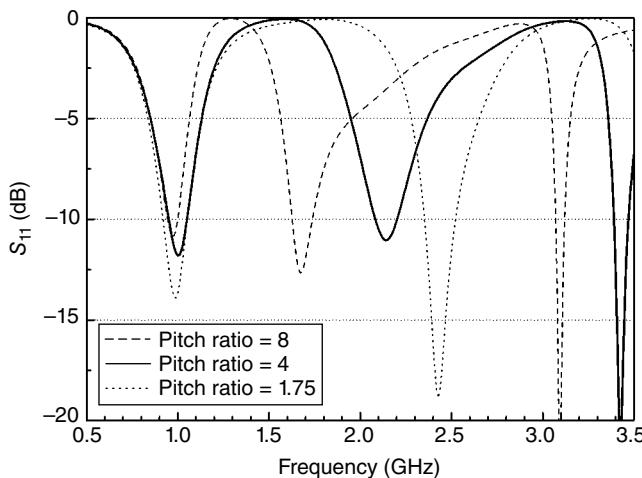


Figure 3.41 Simulation reflection coefficients.

have the same height, diameter, and number of turns. The antenna shown in Figure 3.42b is the same as the one shown in Figure 3.38c, and it is again used as the normal sample. In this comparison, the heights of both small-pitch and large-pitch portions are kept unchanged, which are 10 and 20 mm, respectively. What is changed is how many turns of the helix are put in each portion. In the top portion, three antennas have 3, 4, and 5 turns of helix, respectively. One might have noted that when the turn ratio changes the pitch ratio also changes. As a result, when the bandwidth is exchanged between low and high bands, their frequency ratio is also altered. It is obvious that an iterative procedure must be included to compensate for this effect.

As shown in Figure 3.43, by decreasing the number of turns at the large-pitch segment, we are able to decrease the bandwidth at the lower band in exchange for wider bandwidth at higher band. The antenna shown in Figure 3.42a has the widest low band and the narrowest high band. On the contrary, the antenna shown in Figure 3.42c has the narrowest low band and

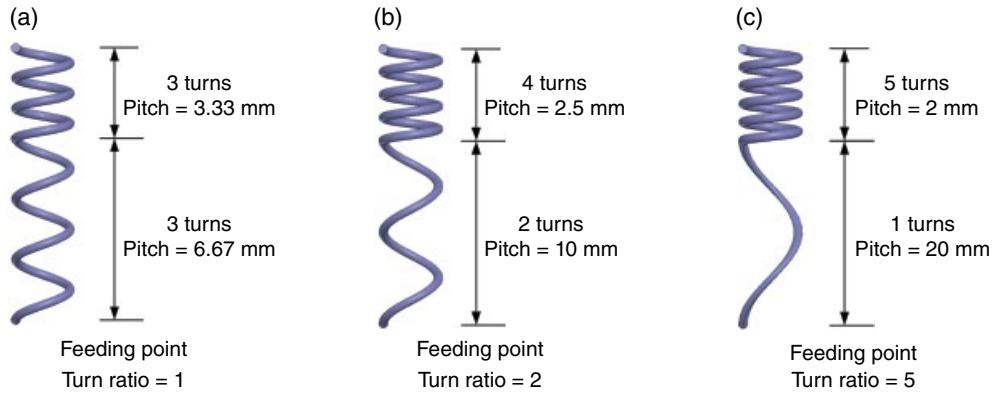


Figure 3.42 Effect of different turns on small pitch portion.

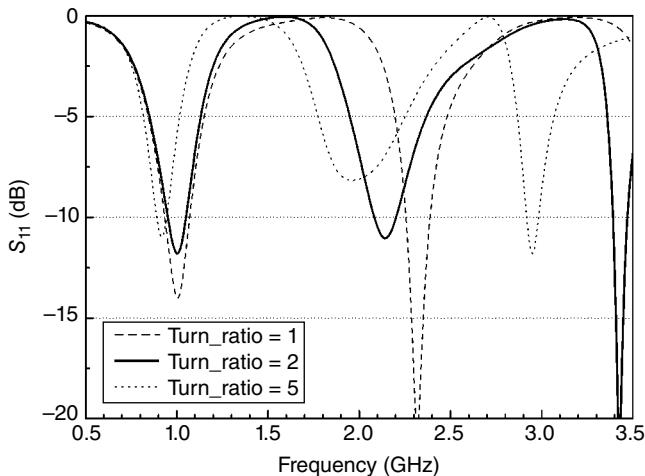


Figure 3.43 Simulated reflection coefficients.

the widest high band. As the median sample, the Figure 3.42b antenna's bandwidth at both bands is moderate among all three.

Now let's use the helix shown in Figure 3.40a as the radiator element to finalize a quad-band antenna design. The antenna also needs to cover the 824–894 MHz, 880–960 MHz, 1710–1880 MHz, and 1850–1990 MHz bands. Shown in Figure 3.44 is the original impedance locus of the Figure 3.40a antenna. The locus marked by hollow rectangular markers is the low band segment and the one marked by hollow circular markers is the high band segment. Compared with a dual-branch antenna, as shown in Figure 3.31a, the impedance locus of a single-branch helix is more regular. At a very low frequency, the antenna is too small to have any effect, so the antenna likes an open circuit and its impedance is close to infinity. At each resonant frequency, the antenna's impedance moves through a circle. There are three circles in Figure 3.44, and they correspond to resonance at 0.95, 1.7, and 3.1 GHz respectively, as shown in Figure 3.41.

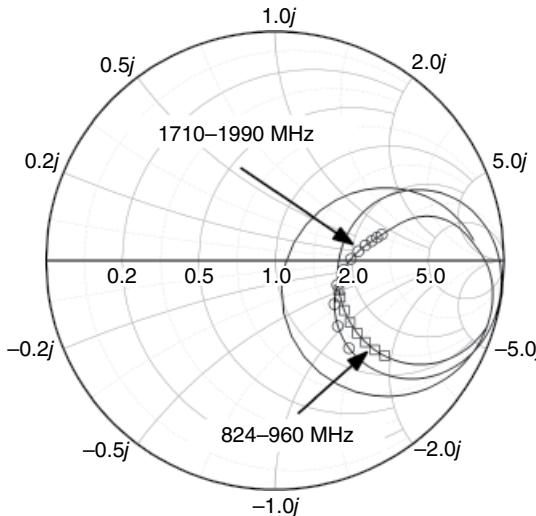


Figure 3.44 Impedance of the original Figure 3.40a antenna.

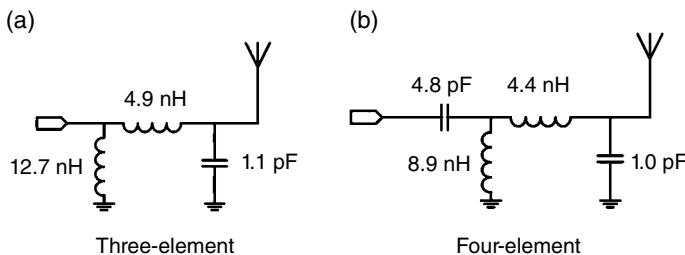


Figure 3.45 Matching circuits for the Figure 3.40a antenna.

Shown in Figure 3.45a is a three-element matching circuit and Figure 3.46a is the matched impedance with this circuit. Shown in Figure 3.45b is a four-element matching circuit and Figure 3.46b is the matched impedance with this circuit.

Shown in Figure 3.47 are results of the matched Figure 3.40a antenna. Both matching circuits can meet the -10dB quad-band specification. The bandwidth of the low band of a three-element matching marginally meets the specification, which can also be observed from Figure 3.46a, as both bands are not centered in the Smith chart. By increasing one more matching component, the antenna's matching is improved. In reality, to decide which matching is better, some chamber efficiency measurements are necessary. Because the extra component also introduces extra losses, only chamber measurements can evaluate these two counteracting effects.

So far, we have discussed dual-pitch helices. Of course, you can make a multi-pitch helix, which introduces more design variables and makes the design procedure more complex. In my opinion, it is not necessary to use multi-pitch helices in a dual-band design. Only when you need to cover multiple separated frequencies, say, 1, 2, and 5 GHz, can multi-pitch helices be considered. You can also use a whip–helix antenna introduced in Section 3.1.2.1. By replacing

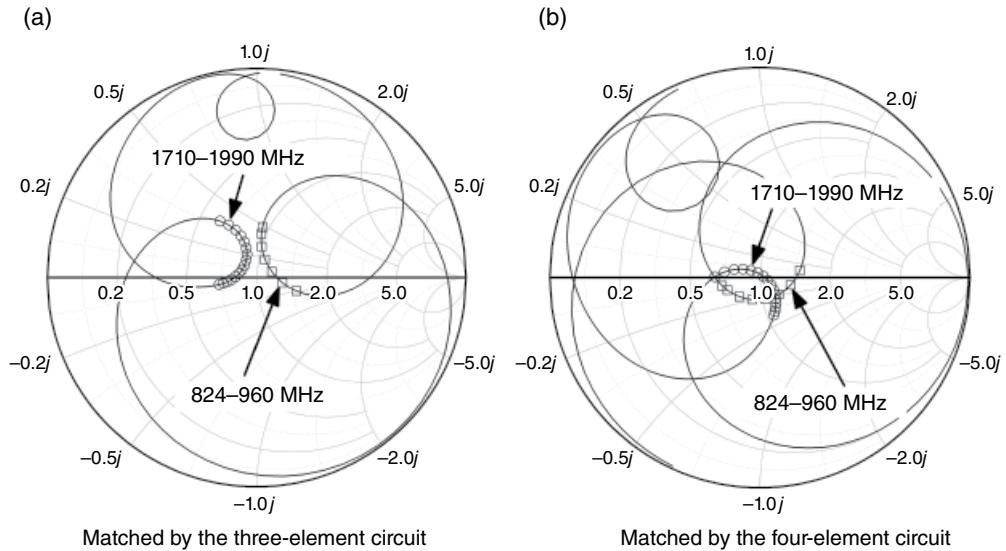


Figure 3.46 Impedance of the matched Figure 3.40a antenna.

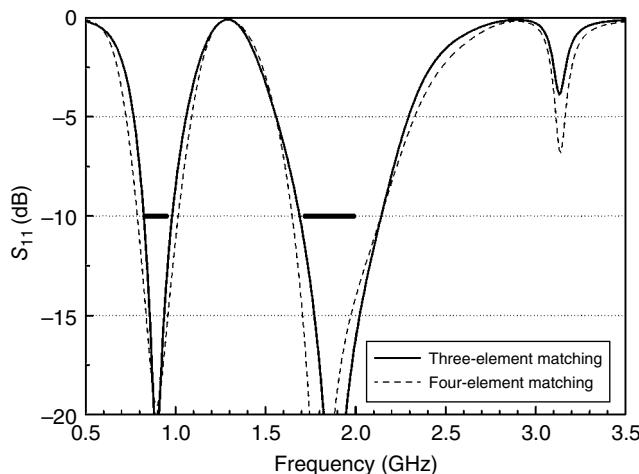


Figure 3.47 Simulated reflection coefficients of the matched Figure 3.40a antenna.

the uniform helix with a dual-pitch helix, the lowest and highest band can be covered. By adjusting the whip's length, the middle band can also be covered.

The following guidelines are given as the summary of dual-pitch antenna's design:

- Although the parametrical studies are omitted on the effect of the height and diameter of dual-pitch antennas, they are all critical parameters which affect bandwidth. Increasing a helix's height can significantly expand its bandwidth; the only issue here is nobody likes a big antenna. Increasing diameter can moderately increase bandwidth.

- The frequency of the first resonance is mostly decided by the total length of the wire. The second resonance can be adjusted by changing the pitch ratio between small-pitch and large-pitch segments.
- To exchange the bandwidth between low and high bands, the ratio between the number of turns at small-pitch and large-pitch segments needs to be changed.
- Similar to the structure shown in Figure 3.36a, a metal base can increase an antenna's bandwidth.

Besides the dual-pitch helix, there are other ways to achieve dual resonance. The helix-loaded whip antenna, as shown in Figure 3.48, is one such example. For a traditional helix-loaded whip, the helix is located on the end of a whip and extends upward. The total height of such an antenna equals the sum of the heights of helix and whip. The helix-loaded whip introduced here is different, and the helix goes downward. That makes the antenna smaller; the antenna's height equals the height of the whip.

For this antenna, we will focus on the frequency tuning technique. To exchange bandwidth between bands, the distance between whip and helix needs to be changed. If the diameter of the helix is fixed, the whip must be biased from the axial line of the helix. An asymmetrical part like that is not easy to manufacture, so we will not discuss how to design it.

Shown in Figure 3.48 are three helix-loaded whips. All of them have the same height (30 mm); the helix's diameter is 6 mm and the number of turns is 4. They are all made of 1 mm diameter copper wire. The difference between them is the pitch of helix, which is 3, 5, and 7 mm for the 1, 2, and 3 antennas, respectively. The total wire lengths of three antennas are almost same, which are 109, 111, and 113 mm, respectively.

Shown in Figure 3.49 are the reflection coefficient results of the three antennas. By stretching the pitch of the helix from 3 to 7 mm, the antenna's low band resonant frequency shifts from 1.02 to 0.87 GHz. In the meantime, the high band resonance maintains a relative stable location. The overall effect of changing the helix pitch is to adjust the frequency ratio between low and high bands. The other tuning parameter is the total wire length. By decreasing or increasing the wire length, the high and low bands can be simultaneously shifted higher or lower. Combining both parameters, we are able to allocate low and high bands to any required frequencies.

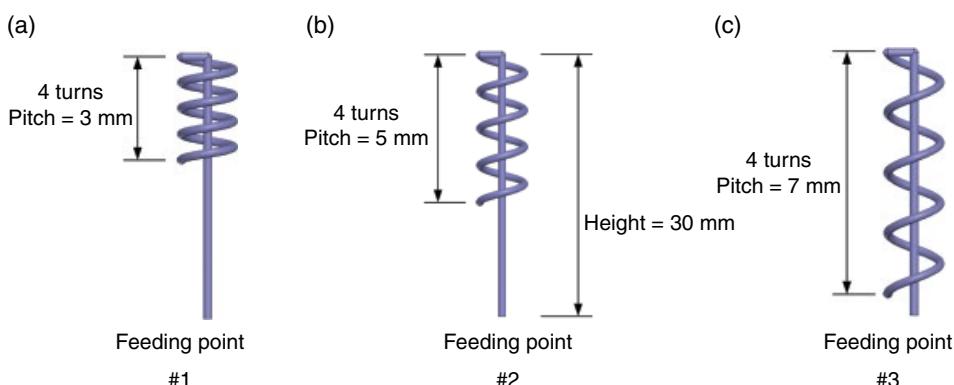


Figure 3.48 Three variants of helix-loaded whips.

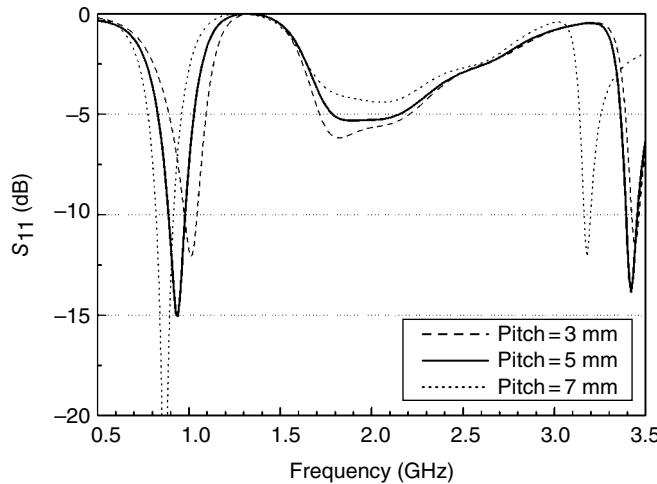


Figure 3.49 Simulated reflection coefficients.

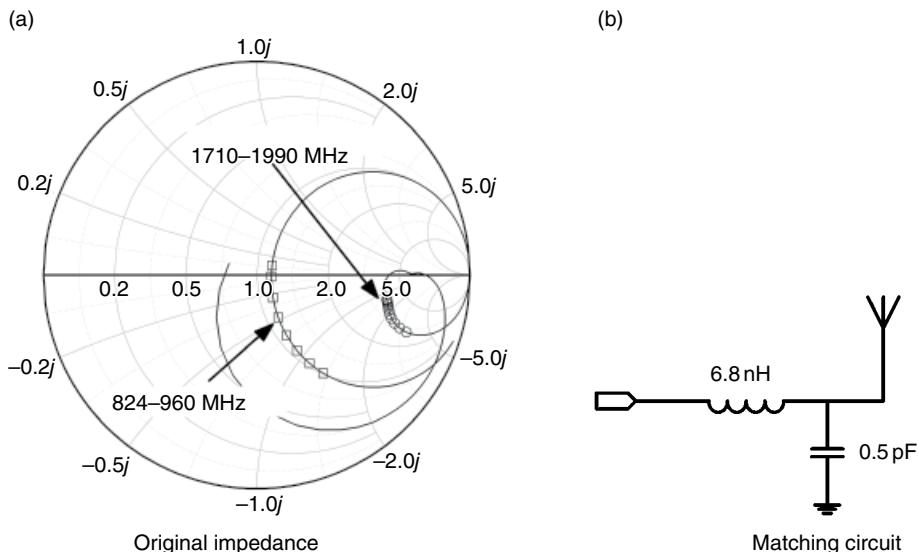


Figure 3.50 Matching a helix-loaded whip.

Now let's use the helix-loaded whip structure to design a quad-band antenna. The antenna also needs to cover 824–894 MHz, 880–960 MHz, 1710–1880 MHz, and 1850–1990 MHz bands. The antenna used here is 30 mm tall. It has a 3.5 turn and its pitch is 8 mm. Shown in Figure 3.50a is the original impedance locus. The locus marked by hollow rectangular markers is the low band segment and the one marked by hollow circular markers is the high band segment. Shown in Figure 3.50b is a two-element matching network. The network is mostly for

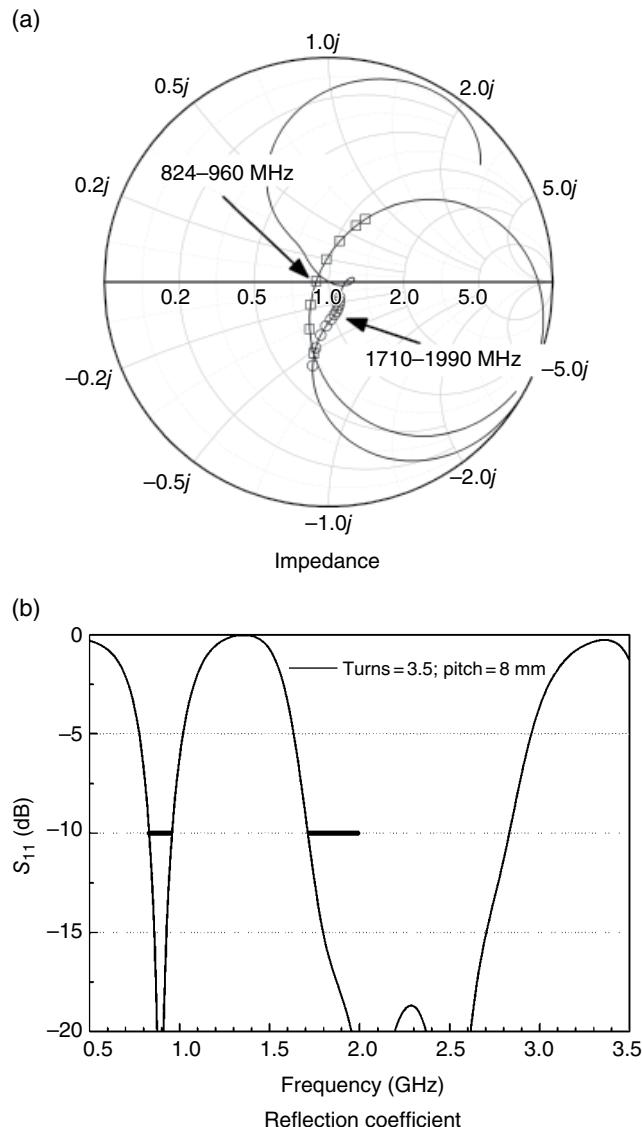


Figure 3.51 Matched helix-loaded whip antenna.

the high band's matching. However, it still shifts the resonant frequency at a low band even lower. To compensate for the shift, the original antenna is tuned to resonate at 940 MHz.

Shown in Figure 3.51a is the impedance of the matched antenna. Figure 3.51b is the reflection coefficient of the matched antenna. The antenna's reflection coefficient is -9.5 dB at the band edge of the low band, and it does not meet the -10 dB specification. However, it has a huge margin at high band. The -10 dB bandwidth of the high band is more than 1 GHz; it covers from 1.71 to 2.83 GHz.

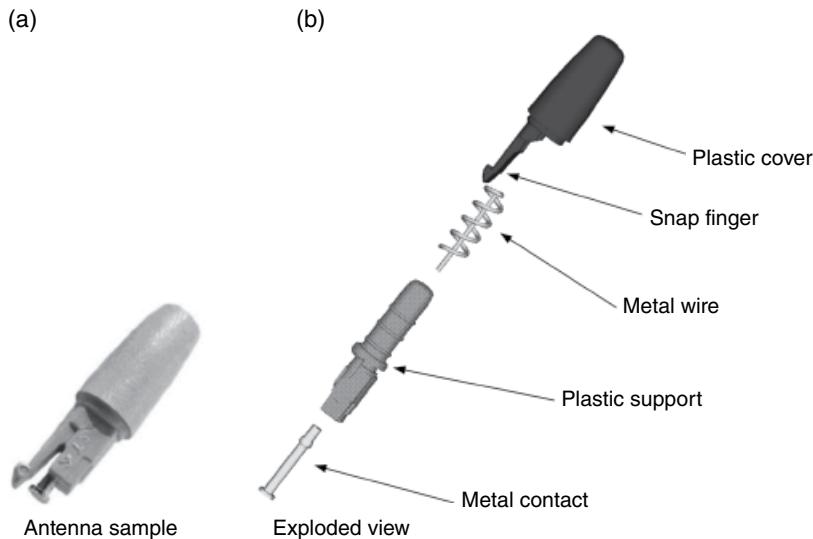


Figure 3.52 Single-branch stubby antenna. (Source: Reproduced with permission of Shanghai Amphenol Airwave.)

The high band's reflection coefficient of an unmatched helix-loaded whip antenna is only around -5 dB , as shown in Figure 3.49. If we only look at the return loss result, it is very easy to get a false impression that this structure has limited bandwidth at high band. The wide band property at the high band can be easily observed on the Smith chart, because the locus is concentrated in a very small area. This example reminds us whenever we design an antenna; we need to use two legs, both reflection coefficient results and impedance locus on the Smith chart should be utilized.

Shown in Figure 3.52 is a production single-branch stubby antenna. The coil and the metal contact are effective radiators. All the plastic parts are used for mechanical purposes. The snap finger is a feature to mechanically mate an antenna to a phone. In this design, the snap finger is integrated with the plastic cover and molded onto the antenna in one shoot.

Now let's check out one more way to design a dual-band antenna. Shown in Figure 3.53 are three whip-loaded helices. The antenna is changed from a uniform-pitch helix by attaching a straight vertical segment to the end of helix. All three elements have the same dimensions apart from the whip length. They are 30 mm tall. The diameter of helixes is 6 mm. The pitch is 6 mm and the number of turns is 5. The whip length is 15, 20, and 25 mm for the loaded helix 1, 2, and 3 antenna, respectively. The total wire length is 117, 122, and 127 mm, respectively. For the same reason as the helix-loaded whip antennas, only frequency tuning techniques are discussed here and the means of exchanging bandwidth are omitted.

Shown in Figure 3.54 are the reflect coefficient results of the three antennas. By increasing the whip's length, the resonant frequencies of both low and high band become lower. Between the two bands, a whip has more impact on the high band. When the whip grows 10 mm (from 15 to 25 mm), the resonant frequency at the high band shifts 430 MHz (from 2.10 to 1.67 GHz). And the ratio between the two bands changes from 2.0 to 1.83. The other tuning parameter is the total wire length. By decreasing or increasing the wire length, the high and low band can

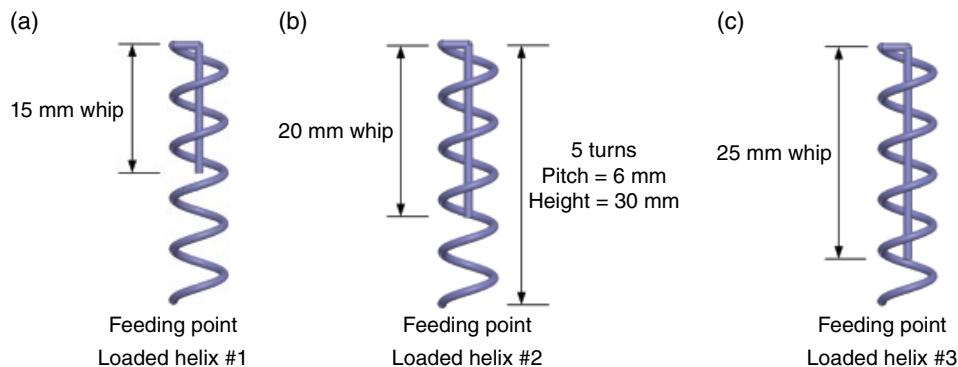


Figure 3.53 Three variants of whip-loaded helices.

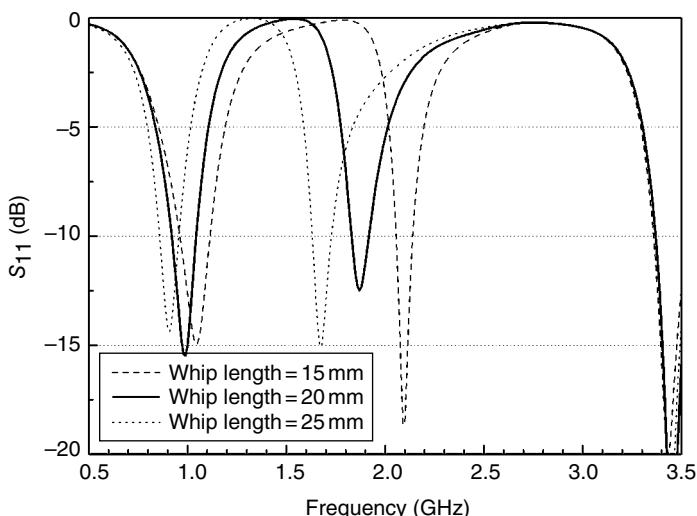


Figure 3.54 Simulated reflection coefficients.

be simultaneously shifted higher or lower. Combining both parameters, we are able to allocate low and high bands to any required frequencies.

As a comparison, the whip-loaded helix in Figure 3.53c is used to design a quad-band antenna. Shown in Figure 3.55a is the original impedance locus. Shown in Figure 3.55b is a two-element matching network.

Shown in Figure 3.56a is the matched antenna's impedance. The Figure 3.56b is the reflection coefficient of the matched antenna. The antenna meets the -10 dB reflection coefficient specification of both bands with some margin.

So far, three single-branch multiband antennas, the dual-pitch helix, the helix-loaded whip, and the whip-loaded helix have been introduced. A rational question one may ask is which we should select. For the convenience of comparison, the results from Figures 3.47, 3.51, and 3.56 are combined into Figure 3.57. It is clear that the bandwidth of the helix-loaded whip

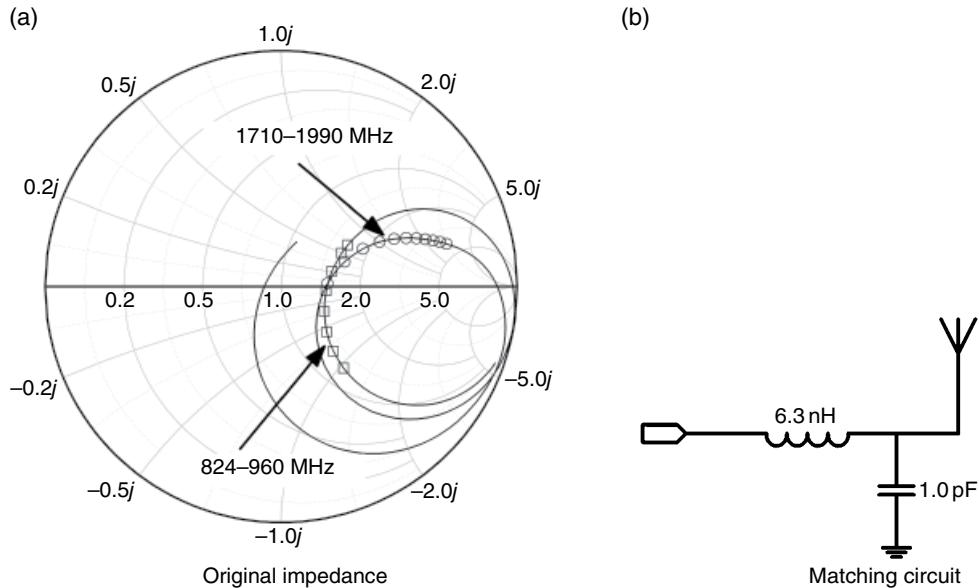


Figure 3.55 Matching a whip-loaded helix.

is the narrowest at the low band and the widest at the high band. On the contrary, the whip-loaded helix has the widest and narrowest bandwidth at the low and high bands, respectively. The bandwidths' difference of three antennas at low band is limited, but their difference at high band is significant.

The bandwidth of a dual-pitch antenna can be adjusted by many parameters, so it is quite easy to design an antenna with well-balanced low and high bands. The potential issue with this kind of antenna is in the production. The antenna is sensitive to the pitch variation, especially at the small-pitch segment. For example, if an antenna has a pitch of 1.5 mm and its wire diameter is 1 mm, then the smallest gap between the wires is only 0.5 mm. A 0.1 mm variation means 20% change on the gap. When designing a dual-pitch antenna, remember to give it enough margin on bandwidth.

The helix-loaded whip has an impressive wide high band. With enough height, it is easy to design an antenna whose high band is able to cover 1.575–2.5 GHz, which includes GPS (1575 MHz), DCS (1710–1880 MHz), PCS (1850–1990 MHz), UMTS (1920–2170 MHz), and Wireless LAN/Bluetooth (2400–2484 MHz). The shape of a helix-loaded whip may look difficult to make; in fact, it only costs a few pennies (US\$) if an experienced vendor can be found. The disadvantage of this antenna is its narrow bandwidth at low bands. If the bandwidth at low bands is critical, choose one of the others.

The whip-loaded helix has the narrowest high band. That is not always a disadvantage. Due to concerns about electromagnetic compatibility, sometimes a limited bandwidth is a good thing. A wideband antenna can pick up many out-band interferences and might radiate undesired noise. From a mechanical point of view, a whip-loaded helix antenna has a large pitch, so it is insensitive to manufacturer's tolerance. In a whip-loaded helix, the longest part is the helix, so it is possible to design an antenna which is easy to assemble. By making the helix

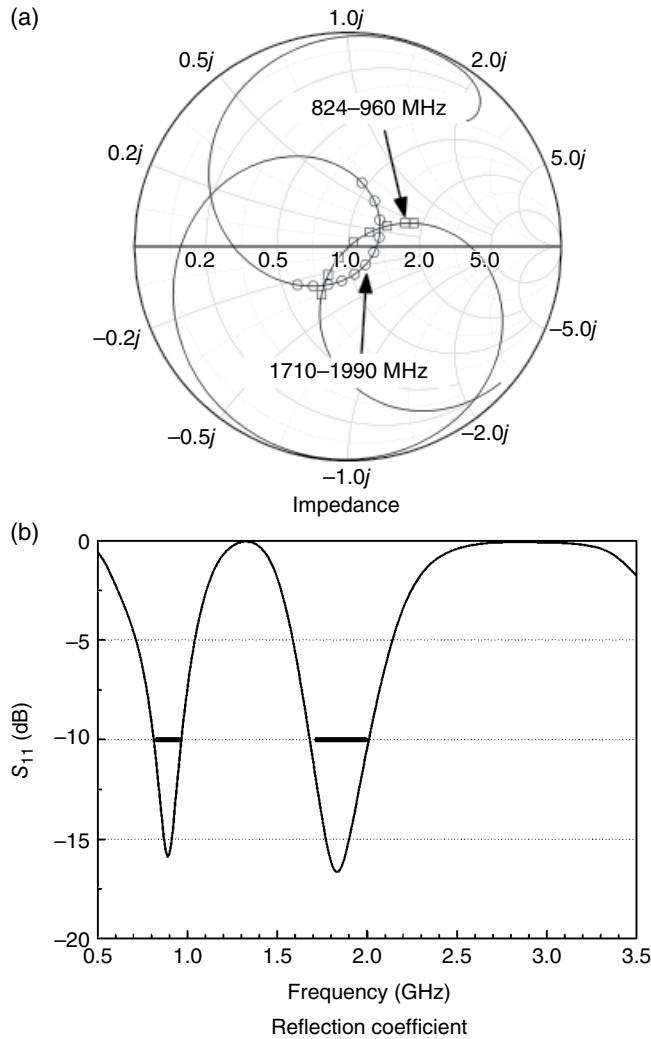


Figure 3.56 Matched whip-loaded helix antenna.

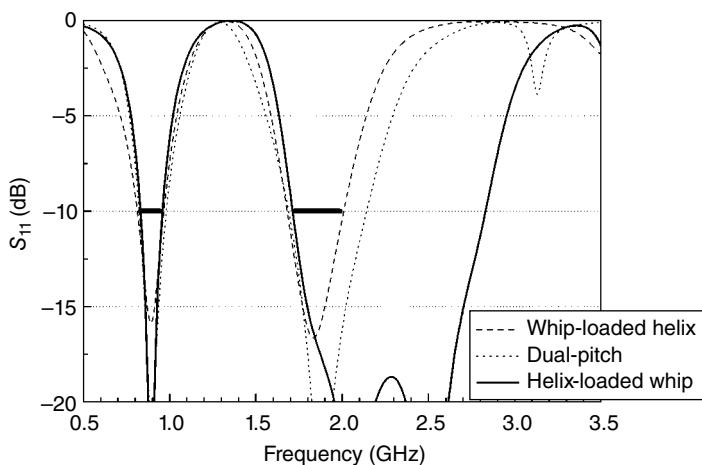


Figure 3.57 Comparison among three single-branch multiband antennas.

longer and then pushing it into an assembly, the helix's spring force can guarantee a good contact between itself and other metal parts. As a counterexample, a helix-loaded whip requires some crimping processes to assemble the metal wire part into the whole assembly, because the whip, which is the longest bit, cannot provide any spring force.

3.1.3 Ultra-Wideband Stubby Antenna

In 2002, the Federal Communications Commission (FCC) allocated the spectrum from 3.1 to 10.6 GHz for unlicensed ultra-wideband (UWB) measurement and medical and communication applications [8]. Many papers [9–22] have reported various different UWB antennas. At the lowest frequency of the UWB, a quarter of a wavelength is less than 25 mm. In this application height is not a critical issue, variations of discone antennas or planer disk antennas have been proposed.

Shown in Figure 3.58 is a wideband cylindrical monopole antenna proposed by Professor Kin-Lu Wong [19]. The antenna consists of two major portions: an upper hollow cylinder and a lower hollow cone. Shown in Figure 3.59 is the measured and simulated return loss of this antenna.

In the following, we will introduce an UWB stubby antenna that covers all current cell phone communication bands from 824 MHz to 6 GHz [23, 24]. The antenna has a volume of

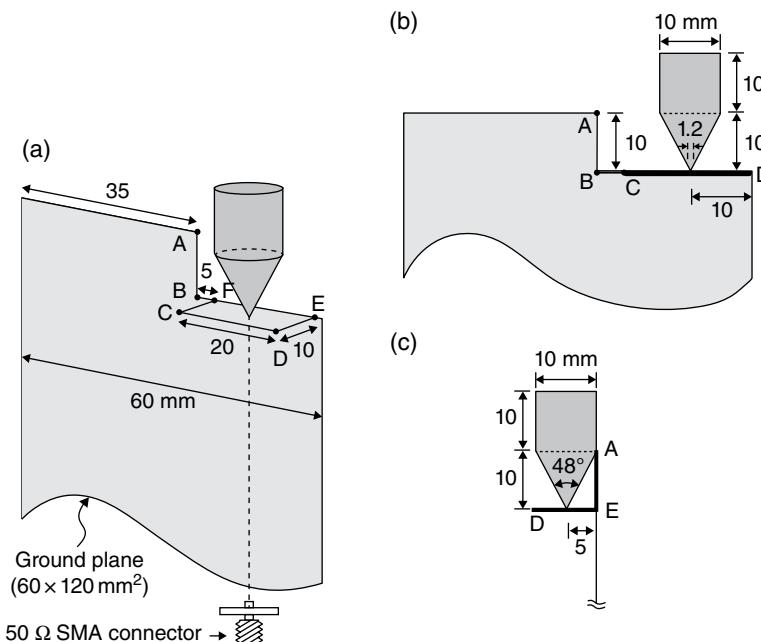


Figure 3.58 Wideband cylindrical monopole antenna. (a) Geometry of the proposed wideband cylindrical monopole antenna. (b) Front view. (c) Side view. (*Source:* Wong and Chien [19]. Reproduced with permission of IEEE.)

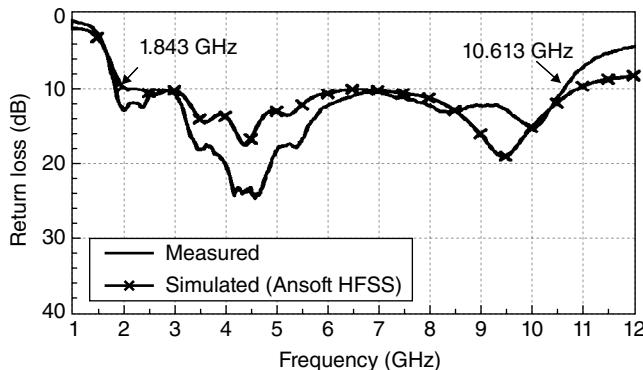


Figure 3.59 Measured and simulated return loss. (Source: Wong and Chien [19]. Reproduced with permission of IEEE.)

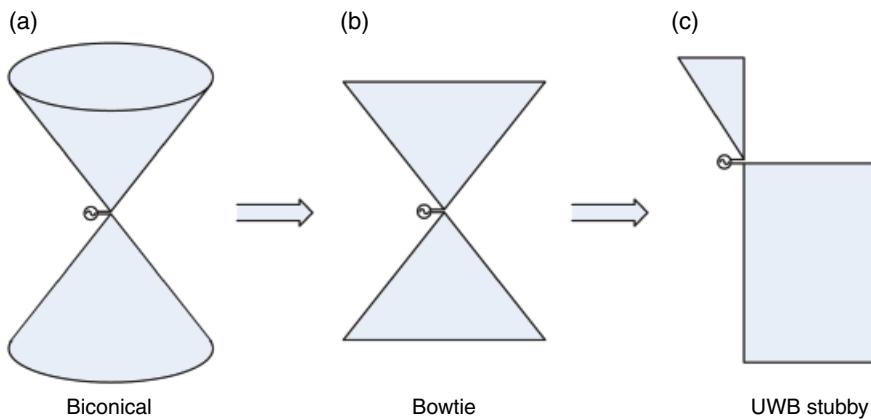


Figure 3.60 Transformation from a biconical antenna to a UWB stubby antenna.

5 mm × 8 mm × 30 mm. The antenna is made of a folded V-shaped stamped sheet metal, which can be easily manufactured and folded to form a 3D shape.

It is well known that tapered structures can exhibit wideband properties. Biconical, disc-conical, and bow-tie antennas [25], as shown in Figure 3.60a and b, can easily achieve UWB coverage. For these kinds of antennas, however, the total length of antenna determines the lowest working frequency and the precision of the tapered feature near the feeding point determines the highest operating frequency.

For a discone antenna with the lowest working frequency of 800 MHz, the length and the diameter of the antenna are all about 160 mm. The size of this antenna is, therefore, too large to be integrated into modern mobile devices that operate in the GSM850 and GSM900 bands. One way to reduce the size of the antenna is to use a matching circuit to compensate for the high capacitance of the antenna at the lower frequency band. This technique enables a good impedance matching over extended frequency band. The antennae design described here, as shown in Figure 3.60c, utilizes this approach.

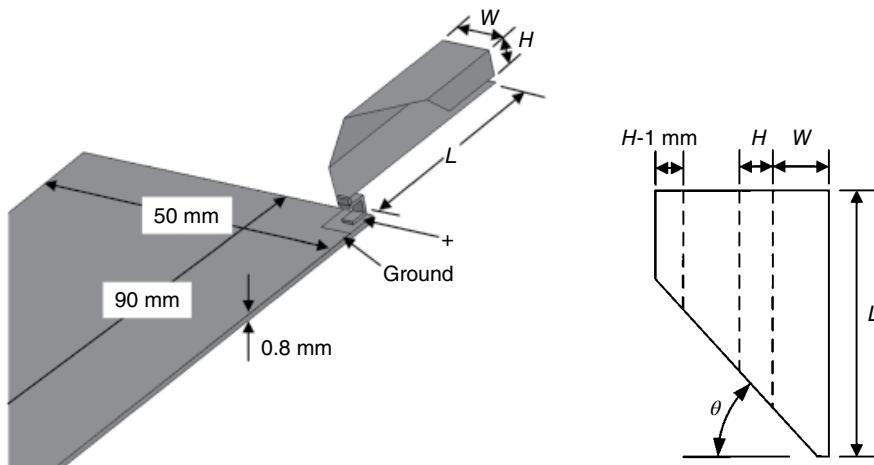


Figure 3.61 Mechanical drawing. (Source: Zhang *et al.* [24]. Reproduced with permission of IEEE.)

Figure 3.61 is the mechanical drawing of the antenna. The antenna consists of a piece of stamped sheet metal as shown on the right side of Figure 3.61 and folded into a 3D shape. The length of the antenna is L , the width is W , the height is H , and the taper angle is θ . The dimension of the FR4 PCB is $0.8\text{ mm} \times 50\text{ mm} \times 90\text{ mm}$. A C-clip is used to make contact between the antenna and the feed point on the PCB. The C-clip is 3 mm tall and has a $2\text{ mm} \times 3\text{ mm}$ footprint.

When designing an antenna for a real phone, the PCB size, C-clip height, and antenna size are mostly determined by ID engineers. In this example, those parameters that are important to antenna performance were parametrically studied. Data presented in this example can be used as a guideline to predict the antenna performance and also as a support claim when negotiating antenna dimensions and design alternatives with industry design engineers.

Although a matching network can be used to obtain good impedance matching, the antenna efficiency at the low band is still highly correlated to the antenna's voltage standing wave ratio (VSWR) or return loss when no matching network is used. An antenna cannot achieve good performance if the original VSWR is too high. When designing the proposed antenna, the goal is to first optimize the antenna dimensions to obtain a reasonably low VSWR response in the lower band and also acceptable VSWR values in the 1.5–3 GHz and 5–6 GHz bands without a matching network. Matching networks are then used to improve and bring the VSWR values in the lower band into the acceptable range.

Figure 3.62 shows four antenna variants of the proposed UWB stubby antenna. All four variants have the same W , H , L , and θ . The difference between the four variants is the number of sides of the antenna. Antenna 1 includes all four sides, and the number of sides in antenna 2, 3, and 4 are reduced to three sides, two sides, and one side, respectively.

Figure 3.63 shows the simulation result of VSWR values for the four antenna variants. It is important to mention at this point that in Figures 3.63, 3.64, 3.65, and 3.66, all data presented are simulation data which do not include the use of a matching network. After obtaining an optimized antenna design based on these simulations, a matching network will be designed and used to obtain good impedance matching across all bands. The design procedure includes

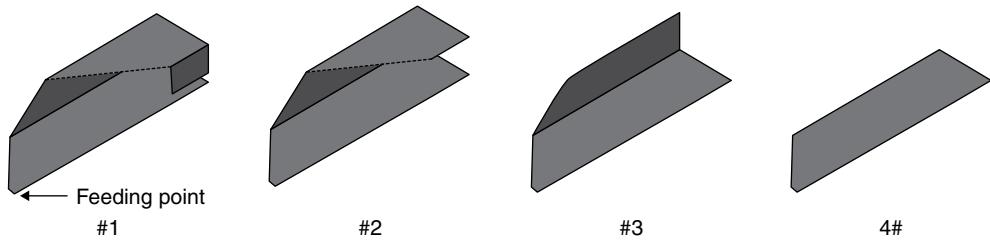


Figure 3.62 Four antenna variants.

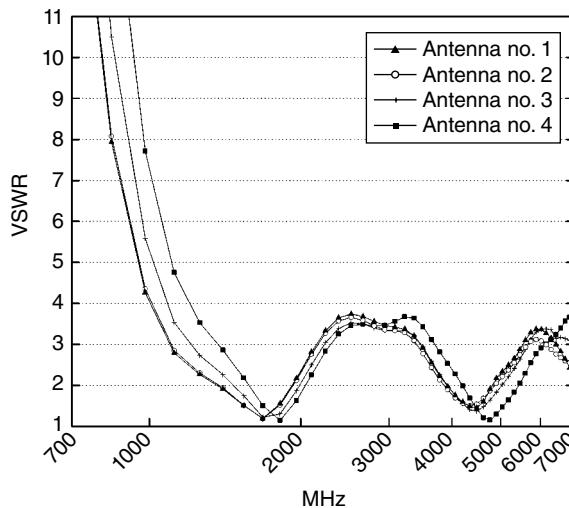


Figure 3.63 VSWR versus antenna variants shown in Figure 3.62 $W=8\text{ mm}$, $H=5\text{ mm}$, $L=30\text{ mm}$, C-clip height = 3 mm, and $\theta=30^\circ$.

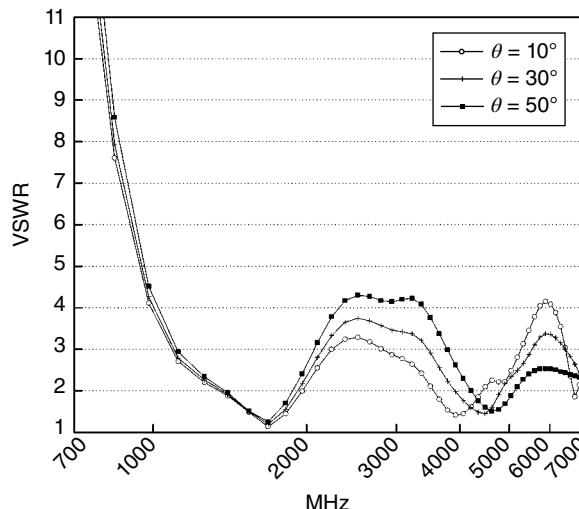


Figure 3.64 Antenna VSWR versus different taper angle $W=8\text{ mm}$, $H=5\text{ mm}$, $L=30\text{ mm}$, and C-clip height = 3 mm. (Source: Zhang *et al.* [24]. Reproduced with permission of IEEE.)

changing the antenna dimensions W , H , L , and the taper angle θ . To better distinguish the VSWR difference of various designs at the lower band, a logarithmic scale was used on the frequency axis (X -axis) in all VSWR versus frequency plots.

For the results shown in Figure 3.63, all four antennas have the same dimensions of $W=8\text{ mm}$, $H=5\text{ mm}$, $L=30\text{ mm}$, C-clip height = 3 mm, and $\theta=30^\circ$. As may be seen from

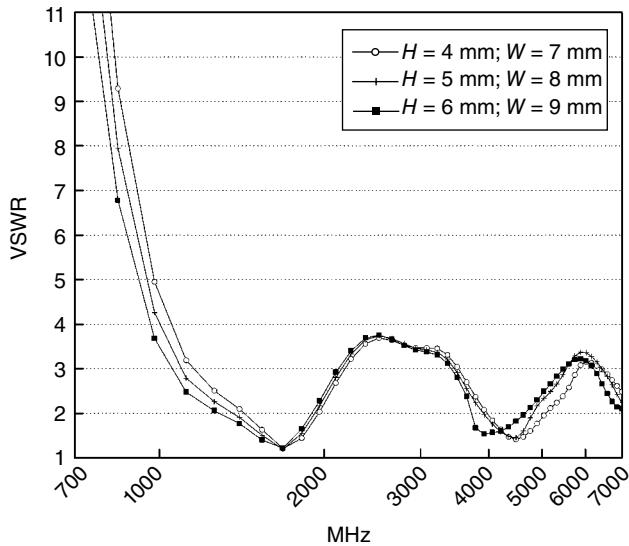


Figure 3.65 VSWR versus antenna width and height $L=30\text{ mm}$, C-clip height = 3 mm, and $\theta=30^\circ$. (Source: Zhang *et al.* [24]. Reproduced with permission of IEEE.)

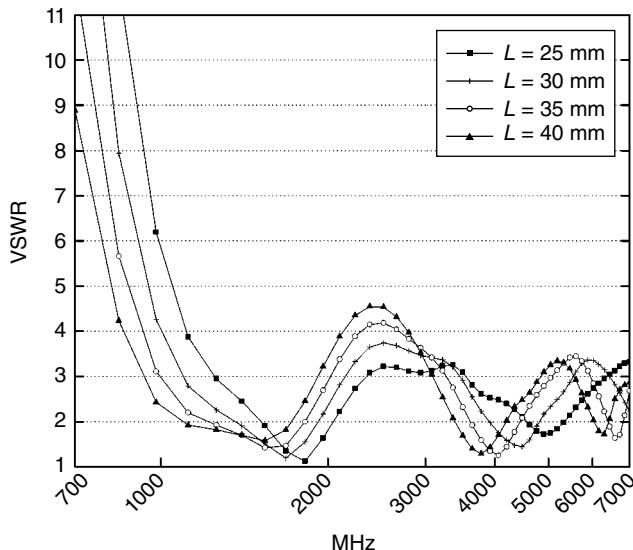


Figure 3.66 VSWR versus antenna length $W=8\text{ mm}$, $H=5\text{ mm}$, C-clip height = 3 mm, and $\theta=30^\circ$. (Source: Zhang *et al.* [24]. Reproduced with permission of IEEE.)

Figure 3.63, VSWR values above 1.8 GHz of all four variants are similar; however, VSWR values below 1.8 GHz have significant differences. At the edge of the lower band, that is, at 824 MHz, the VSWR values of antennas 1, 2, 3, and 4 are 8.7, 8.9, 11.6, and 16.1 to 1, respectively. Therefore, the structure of antenna 1 has the best electrical performance at the lower band and will be used to continue this parametric study.

Figure 3.64 shows the simulation results illustrating the impact of the taper angle θ on the antenna's VSWR. All three antennas have the same dimensions of $W=8\text{ mm}$, $H=5\text{ mm}$, and $L=30\text{ mm}$ with a C-clip height = 3 mm, except for the antenna taper angle, which varies between 10° and 50° . From these results, it may be seen that the taper angle has little impact on the lower band, but it does change the response in the higher band. Figure 3.64 also shows that the taper angle can be used to balance the VSWR values between the 2.5–3 GHz and 5–6 GHz bands. When increasing the antenna taper angle from 10° to 50° , the VSWR from 5 to 6 GHz is improved from 4.1 : 1 to 2.5 : 1, but the VSWR from 2.5 to 3 GHz is degraded from 3.2 : 1 to 4.2 : 1. The taper angle of 30° has a balanced VSWR of 3.5 : 1 in both bands; thus it is used as the optimal angle to carry out the rest of the optimization simulations.

Figure 3.65 shows the simulation results of three antennas with different widths and heights. The three antennas have different widths and heights, which are $4\text{ mm} \times 7\text{ mm}$, $5\text{ mm} \times 8\text{ mm}$, and $6\text{ mm} \times 9\text{ mm}$ respectively. All three antennas have the same length $L=30\text{ mm}$, C-clip height = 3 mm, and $\theta=30^\circ$. When the height and width of the antenna increase, the higher band performance does not significantly change while the lower band performance of the antenna improves. At 824 MHz, the VSWR of antennas with dimensions of $4\text{ mm} \times 7\text{ mm}$, $5\text{ mm} \times 8\text{ mm}$, and $6\text{ mm} \times 9\text{ mm}$ is 10.3, 8.7, and 7.4 to 1, respectively. Based on these results, an antenna with dimensions of $5\text{ mm} \times 8\text{ mm}$ is selected as a good compromise.

Shown in Figure 3.66 are the simulation results of four antennas with different antenna lengths. The antenna length varies from 25 to 40 mm. All four antennas have the same dimensions of $H=5\text{ mm}$, $W=8\text{ mm}$, C-clip height = 3 mm, and $\theta=30^\circ$. As the antenna length increases, the VSWR of the antenna at the lower band improves significantly. At 824 MHz, the VSWR values of antennas of lengths 25, 30, 35, and 40 mm are 12.9, 8.7, 6.2, and 4.6 to 1, respectively. It is observed that the higher band VSWR does degrade with the change in length, while the lower band's performance improves, as expected, with the length increase. As a compromise between industry design and antenna performance, a length of $L=30\text{ mm}$ is selected for the proposed design. It needs to be mentioned, when the L changes, the optimized θ might be different from 30° . The 30° is used here just to show the general trends.

A prototype antenna with following dimensions, $H=5\text{ mm}$, $W=8\text{ mm}$, $L=30\text{ mm}$, C-clip height = 3 mm, and $\theta=30^\circ$, was manufactured and tested. Generally speaking, the larger an antenna, the better the performance at the lower frequencies. The lower band efficiency of an antenna is expected to improve with the increase in the antenna size.

Figure 3.67 shows the measured impedance of the prototype antenna before applying the matching network. The dashed line circle inside Figure 3.67 is the 3 : 1 VSWR circle. There are two markers that indicate the start frequency, 824 MHz, and the stop frequency, 6 GHz, of the desired operating band of the prototype antenna.

To match the lower band of a wideband antenna or a multiband antenna, only two types of components can be used, which are shunt inductors and series capacitors. The order of those two components is decided by the impedance moving path on the Smith chart; in this example, a shunt inductor must be placed before a series capacitor when looking from the antenna side. Two extra matching components are used to improve the high band performance. When

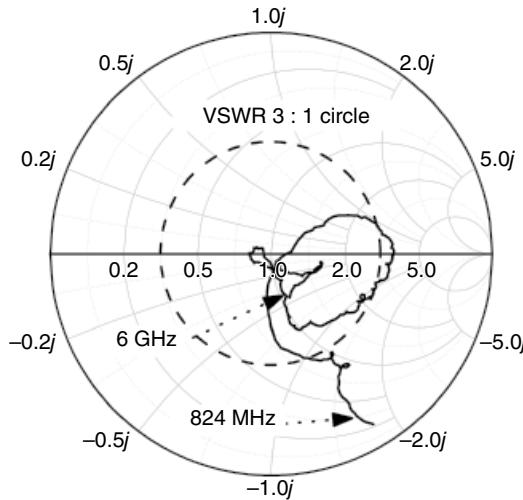


Figure 3.67 Measured impedance of the prototype antenna without matching.

looking from the antenna side, the higher band matching networking includes, first, a 0.3 pF shunt capacitor and then a 1.0 nH series inductor. Figure 3.68a shows the final matching network implemented to achieve a VSWR better than $2.7:1$ across the entire band. All four elements are 0402 ($1\text{ mm} \times 0.5\text{ mm}$) surface mount chip components. Shown in Figure 3.68b is the impedance of the matched antenna. There is some measurement phase error (negative reactance slope) approaching 6 GHz . The reason is when the network analyzer was calibrated, it was calibrated to the SMA connector of the fixture pigtail instead of the matching circuit on the board. The standard port extension function of the network analyzer was used to simply shift the phase reference point to the matching circuit. That causes the larger phase error in the higher frequency band. To be more accurate, some means of calibration needs to be done instead of using the port extension method.

Figure 3.69 shows the measured VSWR data of the prototype antenna with the matching network. The measured VSWR is below $2.7:1$ across the entire band. At the GSM850\&900 ($824\text{--}960\text{ MHz}$) band, the VSWR is below $2.1:1$. At the $1575\text{--}1675\text{ MHz}$ (GPS\&DVB) band, the VSWR is below $2.0:1$. At the $1710\text{--}1990\text{ MHz}$ (DCS\&PCS) band, the VSWR is below $1.8:1$. At the $1920\text{--}2175\text{ MHz}$ (UMTS) band, the VSWR is below $1.8:1$. At the $2400\text{--}2484\text{ MHz}$ (802.11b/g) band, the VSWR is below $2.5:1$. At the $5150\text{--}5875\text{ MHz}$ (802.11a) band, the VSWR is below $1.8:1$.

Figure 3.70 shows the measured efficiency data. A Satimo® Stargate-64 3D near-field chamber was used to measure the efficiency and patterns of the prototype antenna. The efficiency is defined as the ratio of radiated power versus total available power from power source. Thus, the efficiency value includes all the impacts from mismatch loss, dielectric loss, conductor loss, and matching component loss.

From the results shown in Figures 3.69 and 3.70, it may be noted that at the $824\text{--}960\text{ MHz}$ band the efficiency is less than 65% when the VSWR values are better than $2.1:1$. The lowest efficiency is actually 55% at 824 MHz , and this is where the VSWR value is the highest without the matching network. If better than 55% efficiency is required, the highest VSWR

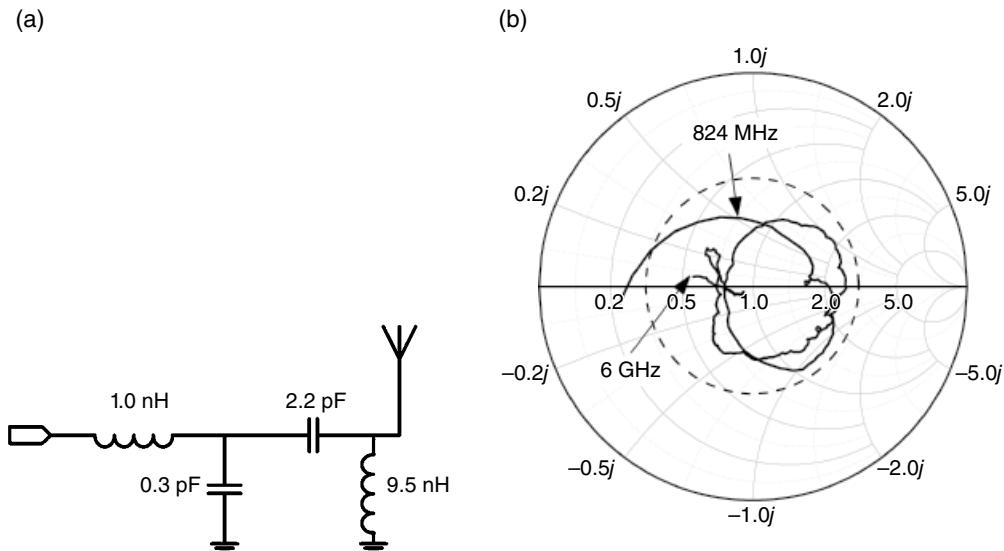


Figure 3.68 Matched UWB stubby. (a) Matching circuit. (b) Measured impedance.

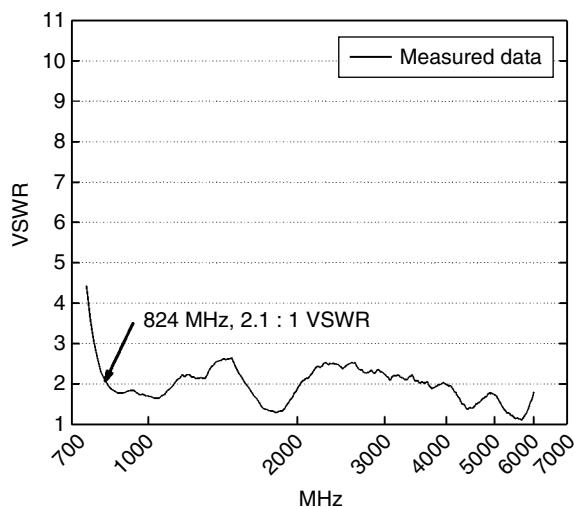


Figure 3.69 Measured VSWR data of the prototype antenna with matching network. (*Source:* Zhang et al. [24]. Reproduced with permission of IEEE.)

without a matching network at the band edge of 824 MHz will have to be improved in the original design, say, by increasing the antenna height, width and/or length as demonstrated in Figures 3.65 and 3.66.

At the higher band, the overall efficiency is better than that at the lower band because the antenna VSWR at the higher band is quite low even without a matching network. From 1575 MHz to 6 GHz, an efficiency of 67–88% was achieved. At the 1575–1675 MHz (GPS&DVB-H US)

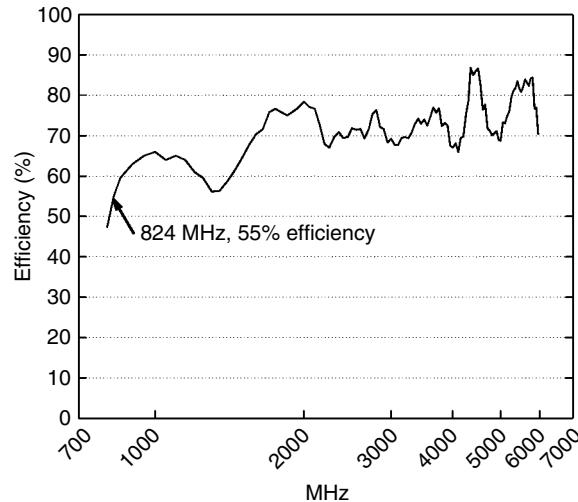


Figure 3.70 Measured efficiency data of the prototype antenna with matching network. (*Source:* Zhang *et al.* [24]. Reproduced with permission of IEEE.)

band, the efficiency is 67–72%, while at the 1710–1990 MHz (DCS&PCS) band, the 1920–2175 MHz (UMTS) band, the 2400–2484 MHz (802.11b/g) band, and the 5150–5875 MHz (802.11a) band, the efficiencies are 75–79%, 70–78%, 69–72%, and 70–83%, respectively.

Figure 3.71 shows the measured radiation patterns at 900 and 5500 MHz. At 900 MHz, the antenna is similar to a dipole, which has a dominant polarized pattern E_θ . At higher frequency, because the current can be excited at both vertical and horizontal edges of PCB, the patterns are less polarized.

Shown in Figure 3.72 is a production UWB stubby antenna [23]. The only radiating part is the metal stamping. The snap finger, which is a mechanical feature to mate an antenna with a phone, is integrated with the plastic support. The contact feature is integrated with the metal stamping.

3.2 Whip–Stubby (Retractable) Antenna

In general, compared to a stubby antenna or an internal antenna, a whip antenna always has the widest bandwidth and the best performance. However, whip antennas are too cumbersome to be installed on most phone platforms, except on those phones that existed in the mid-1980s and had an appropriate nickname “brick.” To obtain the benefit of both stubby antennas and whip antennas, engineers invented the whip–stubby antenna, which is also called “retractable antenna.” From the name, it is obvious that a whip–stubby antenna is a cross-breed between two different antennas. A whip–stubby usually has a retractable whip. When the whip is extended, because the whip is much longer than the stubby, it dominates the radiation and can achieve better performance. When the whip is retracted, the stubby becomes the primary radiator and the performance will degrade to a certain extent. Two different whip–stubby antennas are introduced in later Sections 3.2.1 and 3.2.2, respectively. There are many more variants of whip–stubby antennas besides them. If using “retractable antenna” as the keywords, nearly 100 patents can be found in the US patent database.

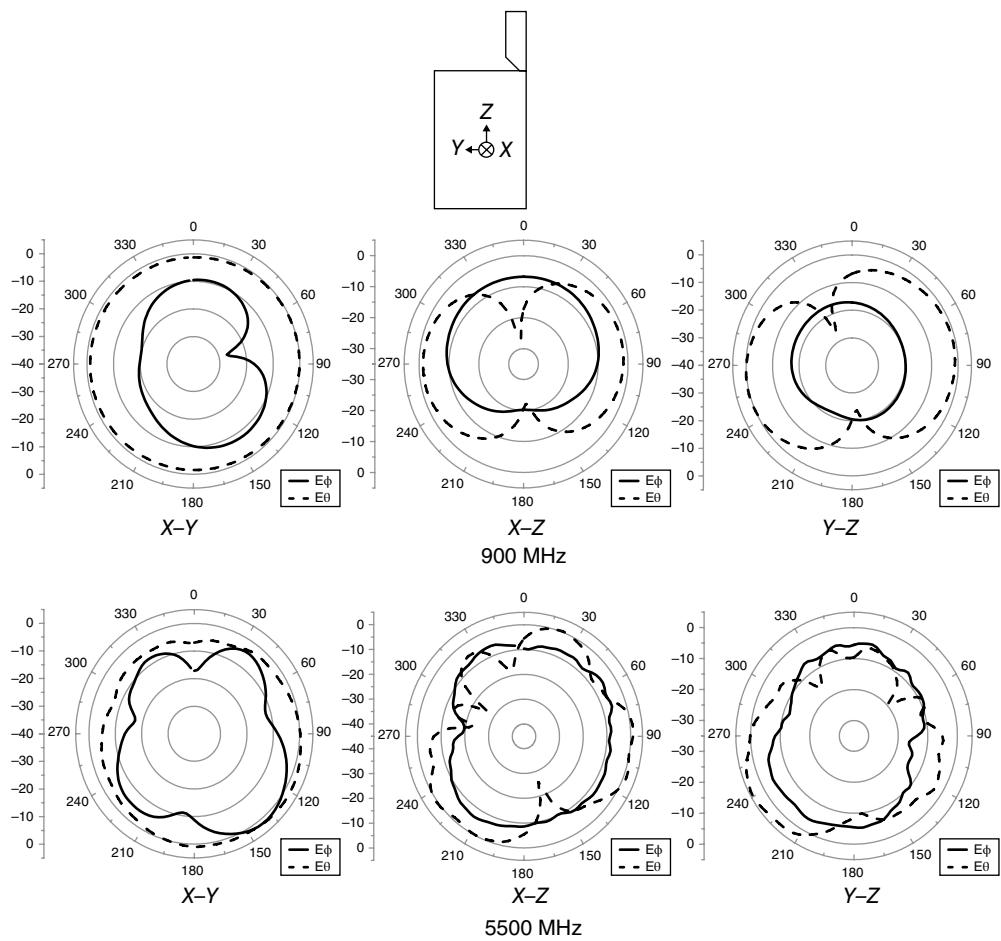


Figure 3.71 Measured antenna patterns. (Source: Zhang *et al.* [24]. Reproduced with permission of IEEE.)

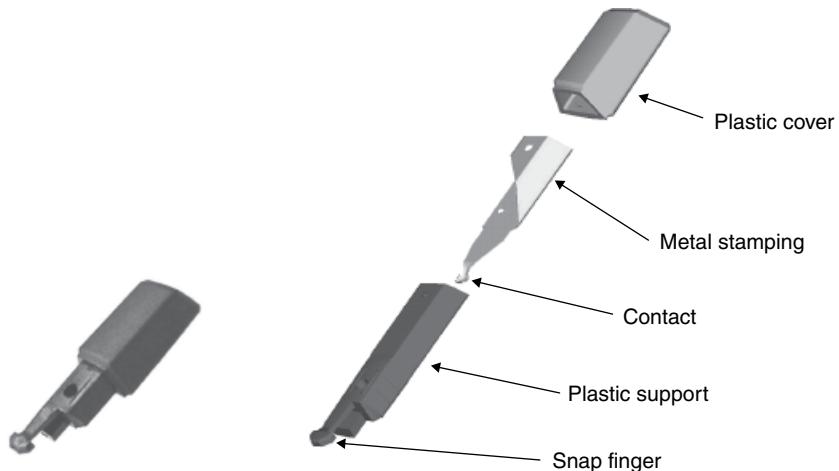


Figure 3.72 Production UWB stubby antenna. (Source: Reproduced with permission of Shanghai Amphenol Airwave.)

3.2.1 Decoupled Whip–Stubby Antenna

The decoupled whip–stubby antenna is the most straightforward combination of a whip and a stubby. Figure 3.73 depicts such an antenna. The appearance and mechanical construction of decoupled antennas can vary significantly; however, they are all the same from the electrical point of view, as the whip and the stubby never work simultaneously. In Figure 3.73, all the transparent parts are made of plastic and all the solid parts are made of metal. The whole antenna assembly is composed of two independent radiators, a stubby in the top and a whip in the bottom. To ensure the electrical isolation between two radiators, the length of the plastic separator between radiators must be 5 mm or more. The metal whip is normally made of nickel–titanium alloy.

Figure 3.74 illustrates a decoupled antenna when it is installed on a phone. All the components inside the phone are omitted to better illustrate the antenna. Some mechanical features, such as the spring finger, the C-shape, or S-shape clip, are normally used to make an electrical connection between the metal connector and the feeding pad on a PCB. In Figure 3.74a, it is obvious that in the retracted position the whip is pulled all the way back and away from the metal connector inside the phone. The helix antenna is the only active radiator. In the extended position shown in Figure 3.74b, the whip makes connection with the metal connector and the helix is electrically floating on the tip of the whip. The whip is the only active radiator in this alternative position.

Designing a decoupled antenna is relative easy because the two radiators are totally independent, so they can be tuned separately. However, because most phones only have one matching circuit, some extra attention needs to be paid to the design of both antenna and matching network.

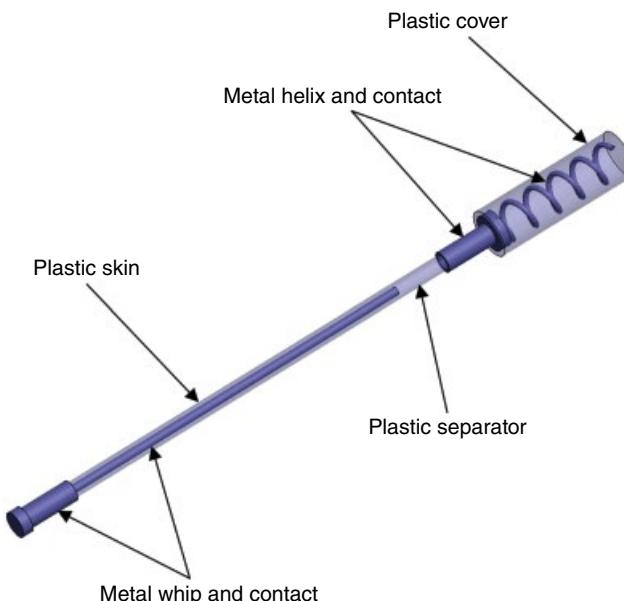


Figure 3.73 Decoupled whip–stubby antenna.

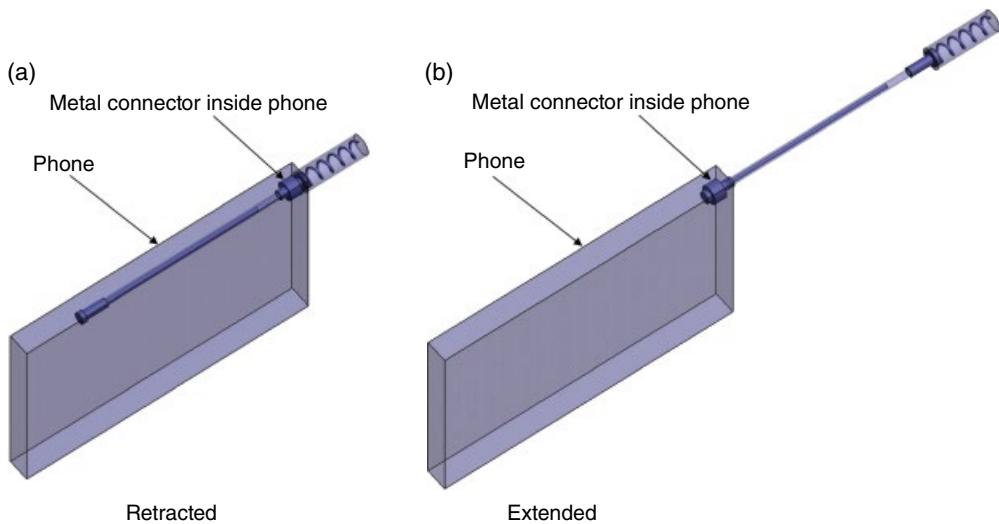


Figure 3.74 A decoupled whip–stubby antenna on a phone.

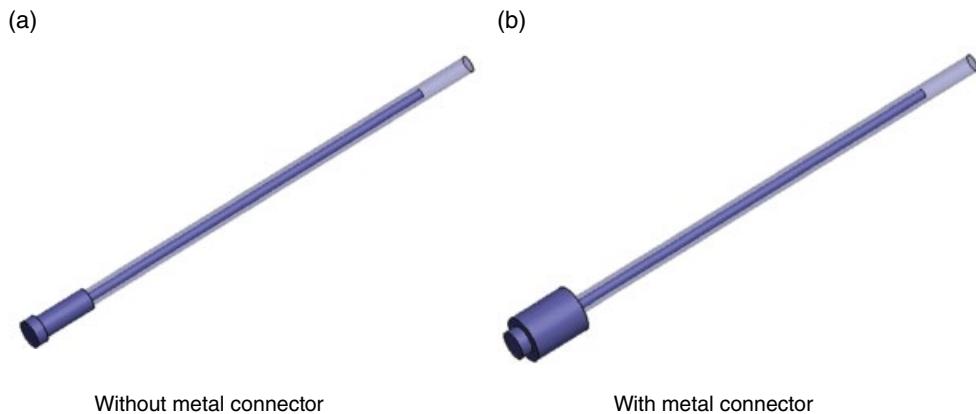


Figure 3.75 Two different whip antennas.

The metal connector, which serves as the interface between an antenna and a circuit board, is the critical part in all whip–stubby antennas’ design. To demonstrate the effect of the metal connector, two cases are simulated for comparison. Shown in Figure 3.75 are two whip antennas. In the first case as shown in Figure 3.75a, the whip is fed directly. In the second case as shown in Figure 3.75b, the whip is fed indirectly through a metal connector which is 10 mm long and 8 mm in diameter. Both whips are 80 mm long. Similar to the configuration shown in Figure 3.8a, a 70 mm × 120 mm metal board is used as the ground plane.

Shown in Figure 3.76 is the simulated reflection coefficients of both cases. It is clear that the metal connector slightly improves the lower band and significantly improves the higher band. Unlike stubby antennas, whips do not have too many tuning parameters besides the whip length and diameter. The length of a whip decides its lowest working frequency.

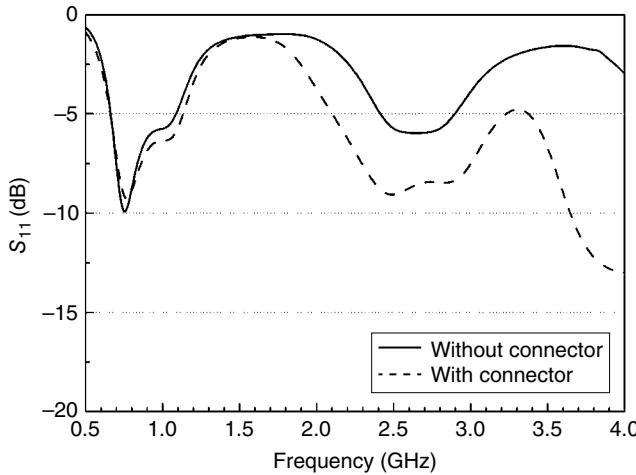


Figure 3.76 Metal connector's impact on whip antennas.

Without the design freedom of metal connectors, we are probably stuck with the higher band response whatever a whip has. Fortunately, there are metal connectors to come to our assistance. Not a lot of skill is required to design a metal connector; the larger, the better. As has been demonstrated in earlier sections, a metal connector can also improve the performance of stubby antennas.

3.2.2 *Semi-Decoupled Whip–Stubby Antenna*

Shown in Figure 3.77 is an example of a semi-decoupled whip–stubby antenna. All transparent parts are made of plastic and all solid parts are made of metal. All components inside the phone, including features to make the electrical connection between the metal connector and the feeding pad on a PCB, are omitted to better illustrate the antenna. The whole antenna assembly is composed of two radiators, a metal whip, and a metal helix. The helix portion of the antenna is permanently attached to the phone. The whip can be extended and retracted. In the retracted position, as shown in Figure 3.77a, only the plastic part of whip sits inside the stubby and the metal whip is decoupled from the helix completely. In this position, the helix is the only active radiator. In the extended position, the whip extends all the way out of the phone and the end of whip slides into the metal connector. In this position, both the helix and the metal whip are electrically connected together, so both of them are activated.

From an electrical point of view, the design procedure of a semi-decoupled antenna is more complex than a decoupled one and its performance is a little worse. The merit of a semi-decoupled antenna is its robustness. As the stubby of a semi-decoupled antenna is permanently attached to the phone, the antenna assembly has better performance in drop tests and wear-out tests. Usually, the whip portion of an antenna is the weakest point in the whole structure. For a decoupled antenna, if the whip is broken, the antenna might get lost. For a semi-decoupled antenna, there is still a helix in the phone. Of course, different companies

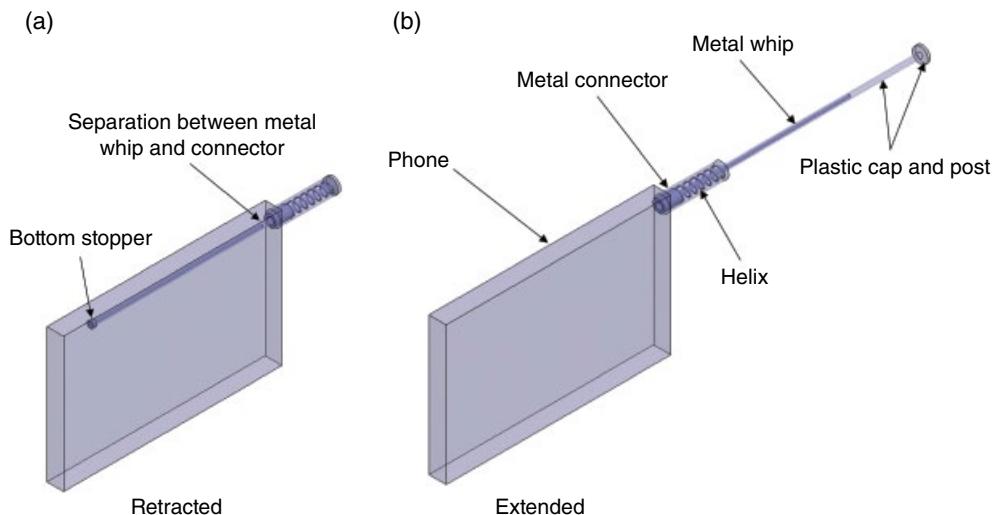


Figure 3.77 A semi-decoupled whip–stubby antenna on a phone.

have different opinions on this matter, so some of them like decoupled solutions and the others like semi-coupled approaches.

In the retracted position, the helix of a semi-coupled antenna is the only active radiator. In the following example, the dimensions of the PCB, which functions as the ground, are $120\text{ mm} \times 60\text{ mm}$. The diameter and height of the metal connector are 8 and 10 mm, respectively. The total height of the helix is 26 mm. The diameter of the helix is 7 mm. The pitch and number of turns of the helix are 4 and 6.5 mm, respectively.

Shown in Figure 3.78 is the simulated response of the helix by itself. The thick solid lines in Figure 3.78a and b illustrate the 824–894 MHz and 1850–1990 MHz, which are the presumed low and high bands. It is clear that the resonant frequency at low bands is offcentered, which is purposely designed.

Shown in Figure 3.79 is the antenna's response when the whip is extended. The metal whip length is 90 mm if measuring from the top of the metal connector. Two narrow resonances, which are marked in Figure 3.79 a and b, are generated by the helix. As we have discussed in Figure 3.30, at these two frequencies there is strong current on the helix; however, it cannot effectively radiate due to the blocking effect of the center whip. As a result, the radiating efficiency at these frequencies is pretty low and needs to be avoided.

Shown in Figure 3.80 are the responses of a 90 mm whip by itself. Unlike Figures 3.78 and 3.79, the whip by itself is not a valid working state and the purpose of showing Figure 3.80 is to demonstrate the mutual impact between a whip and a helix. By comparing Figures 3.79 and 3.80, it is clear that due to the existence of the helix, the bandwidth of the whip is shrunk.

It might seem cumbersome to design a semi-coupled antenna, because so many elements are playing a role here. However, there is a programmatic design procedure which can be followed. When tuning a semi-coupled antenna, we have two goals. The primary goal is to adjust the impedances of both retracted and extended positions so they are in the same area on the Smith chart. The location of the impedance at different bands can be different, as long as

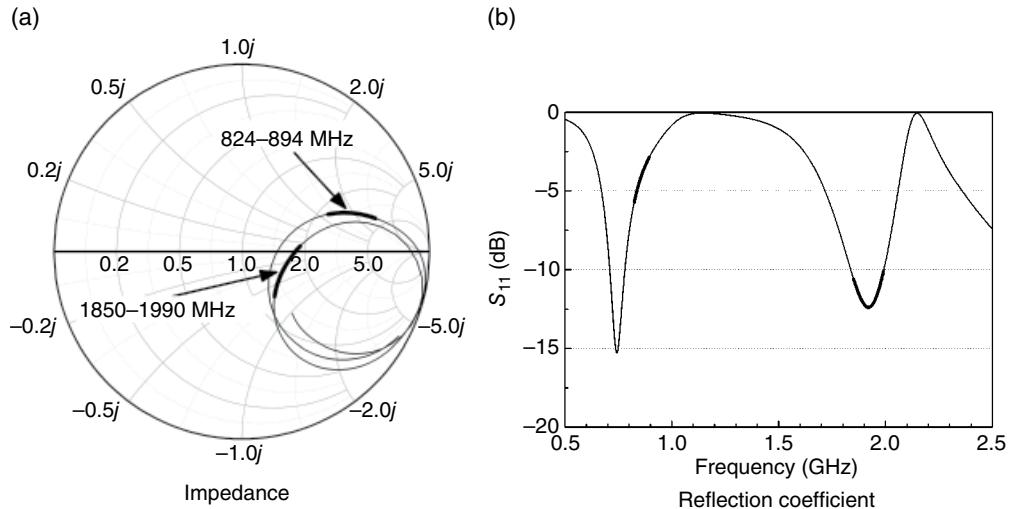


Figure 3.78 Design of the helix (retracted position).

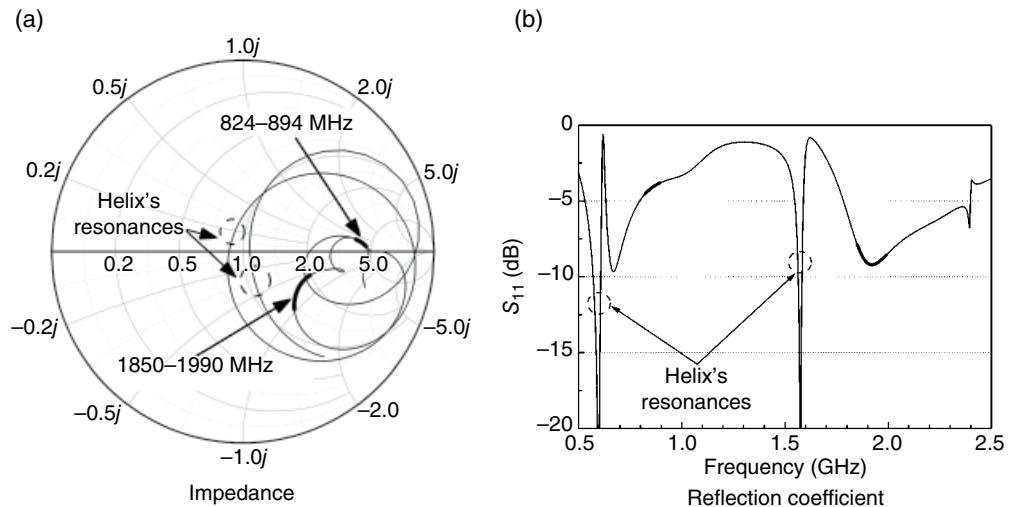


Figure 3.79 Design of the helix-whip (extended position).

the impedances of the same band stay together. Then a matching circuit can be used to achieve a good impedance matching over all bands. If we tried to deal with the antenna's two positions separately, most likely it would be that whenever the antenna's performance in one position was tuned to the best, the performance in the other position would be significantly worse. To achieve a balanced performance, we must treat the impedance of both positions as a whole.

With the priority of the first goal, we still want to design the antenna to obtain a better matching, which is the secondary goal. Because the smaller an antenna's return loss is, the

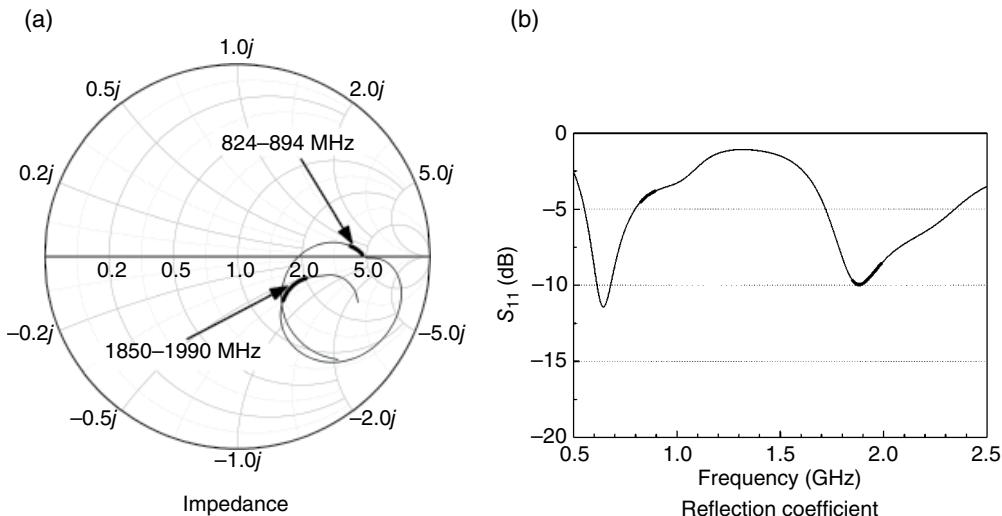


Figure 3.80 The whip by itself.

easier it is to design a matching circuit and the better its efficiency is. With the above two goals in mind, the actual design procedure has the following four steps:

1. The first step is designing the whip. There is only one parameter, whip length, that can be adjusted. If a whip's basic resonant frequency is f_0 , its next resonance is around $3 \times f_0$. However, the ratio between the 850 and 1900 MHz bands is about 2.2, so it is not practical to design a whip which intrinsically resonates at both bands. In the example shown in Figure 3.80, the whip is tuned at the high band, where its return loss is better than -8 dB. At the low band, the return loss is around -4 dB.
2. The second step is designing the helix. A helix by itself corresponds to the antenna's retracted position. Our goal is to match the response of the helix to that of the whip at both bands. Techniques for designing a dual-pitch helix can be used to adjust the helix's response to the desired locations on the Smith chart. In this example, a uniform helix is used. Because with the help of the metal connector, a uniform helix's response at both bands already meets our requirement, it is unnecessary to use a dual-pitch design.
3. The third step is to put the helix and the whip together, fine-tune them, then remove the whip, tune the helix alone, then tune both again. This step might take a few iterations. Some extra attention needs to be paid to those unwanted helix's resonances shown in Figure 3.79. The helix's resonances always move lower at the extended position. For a narrow band antenna, as in this example, no special treatment is needed, because the unwanted resonances will shift out of bands as shown in Figure 3.79. For a wideband antenna, the unwanted resonances might still be inside the low end of working bands. In this kind of circumstance, the helix's resonance must be tuned to a lower frequency and the whip length also needs to be tuned accordingly.
4. The fourth step is to design a matching network. We must use the same circuit to match both the retracted and extended position simultaneously. This step is quite straightforward if the impedances of both positions are close enough on the Smith chart.

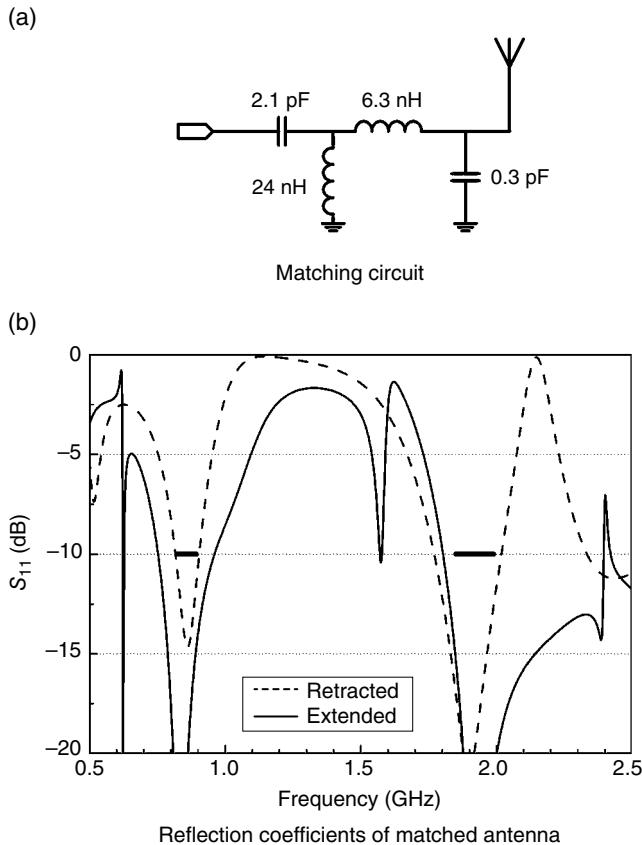


Figure 3.81 Matching both retracted and extended positions.

Shown in Figure 3.81a is the four-element matching circuit of the final antenna. Shown in Figure 3.81b is the simulated reflection coefficients of both positions. The return losses of all bands at all positions meet the -10 dB specification. Beside the bandwidth of the low band in the retracted position, which is marginal, there is large margin in all other situations. The bandwidth in the extended position is more than twice as much as required. In practice, the specifications for retracted and extended position are often different. The performance of the extended position is more critical, because this is the normal usage position. In the meantime, the specification for the retracted position can be relaxed, because all users are aware that the whip should be extended if the call quality starts to degrade. If -6 dB is used as the specification for the retracted position, an antenna of the same dimension can cover penta-bands, 850 MHz, 900 MHz, DCS (1800 MHz), PCS (1900 MHz), and UMTS (2100 MHz), with some tweaks.

Shown in Figure 3.82 is a production retractable antenna. The antenna is a semi-decoupled whip-stubby. A dual pitch coil is used as the radiator when the antenna is in retracted position. In the extended position, a metal whip functions as the radiator. As the whip needs to be fully retracted into the phone, the total whip length must be less than its host's length. In this

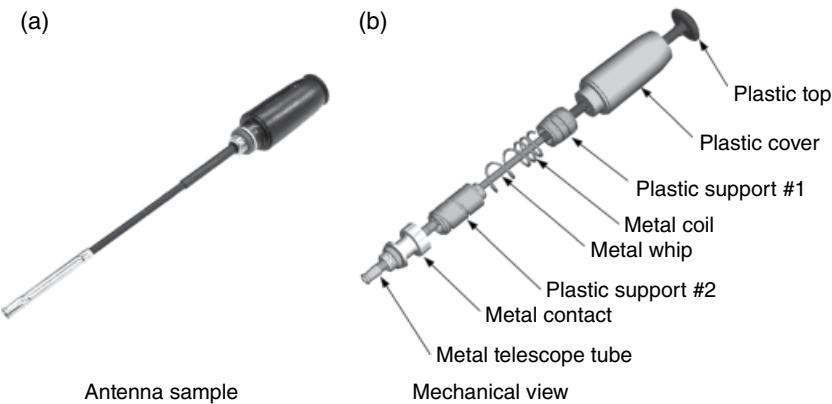


Figure 3.82 Retractable antenna. (Source: Reproduced with permission of Shanghai Amphenol Airwave.)

specific design, the phone's length is not long enough for the required band. A telescope tube is added to the design to provide the extra length. The whip slides in and out the telescope tube in the retracted and extended positions, respectively.

3.3 Meander Line Stubby Antenna

With the progress of flex circuit technology, industry design engineers started to adopt the stubby antenna with the irregular cross section. The radiator elements of some of those antennas are made of flex instead of coil. To accompany the flex, meander lines are introduced to replace the coil. The principle of a meander line antenna design is similar to a helix. Figure 3.83 demonstrates the transformation of a multiple branch meander line antenna.

When using the flex circuit technology, it is easier to implement multiple branches than use coils. Many variants of this type of antennas have been reported. Shown in Figure 3.84 is a dual band antenna for WLAN application [26]. Shown in Figure 3.85 is the simulated and measured reflection coefficients.

Shown in Figure 3.86 is another dual-branch multiband monopole antenna [27]. Its simulated and measured reflection coefficients are shown in Figure 3.87. Although it may look like a single branch structure, the paths A and B are actually two independent branches.

Shown in Figure 3.88 is a stubby antenna made of flex. The square pattern with a hole on the flex is the feeding point, which connects to a metal spring finger. A metal screw is used to secure the connection between them. Two traces come out of the feeding point. The short trace on the left side is for the high band and the longer one on the right is for the low band. Compared with a coil, a flex provides much more freedom on the trace layout. However, the basic rule still applies. Larger trace area means wider bandwidth. By diverting the available flex area to traces of different bands, the trade-off between bands can be made.

Shown in Figure 3.89 are some variants of multi-pitch meander line antennas. The left variant is quite straightforward. The right one is less conceivable. Both variants shown in Figure 3.89 are on a 2D surface. This type design can be extended to 3D surface.

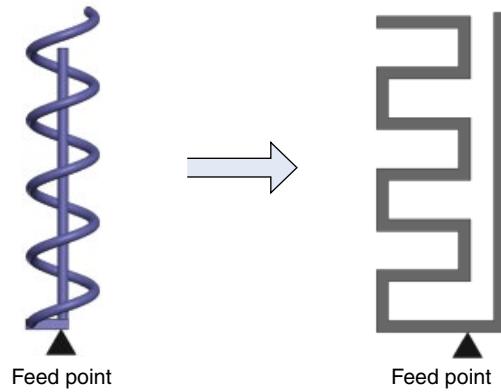


Figure 3.83 Multiple branch meander line antennas.

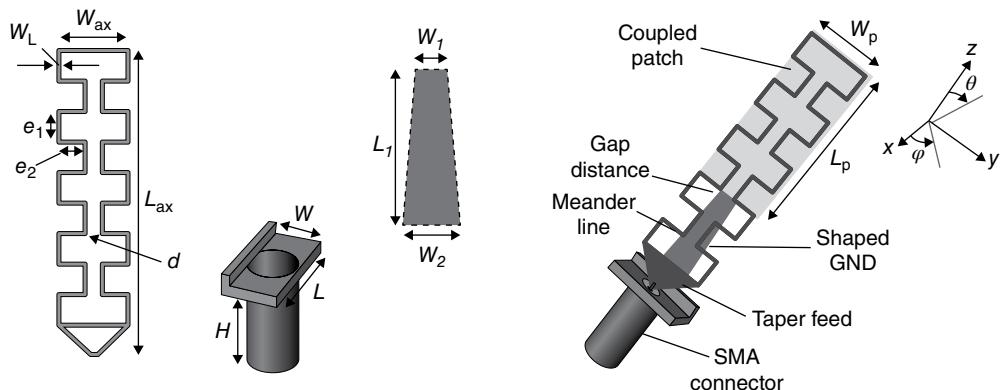


Figure 3.84 Dual-band antenna for WLAN application. (Source: Khaleghi [26]. Reproduced with permission of IEEE.)

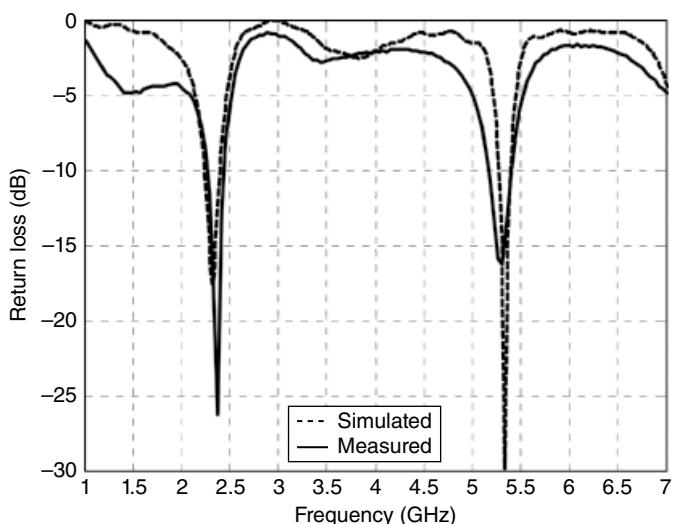


Figure 3.85 Reflection coefficients. (Source: Khaleghi [26]. Reproduced with permission of IEEE.)

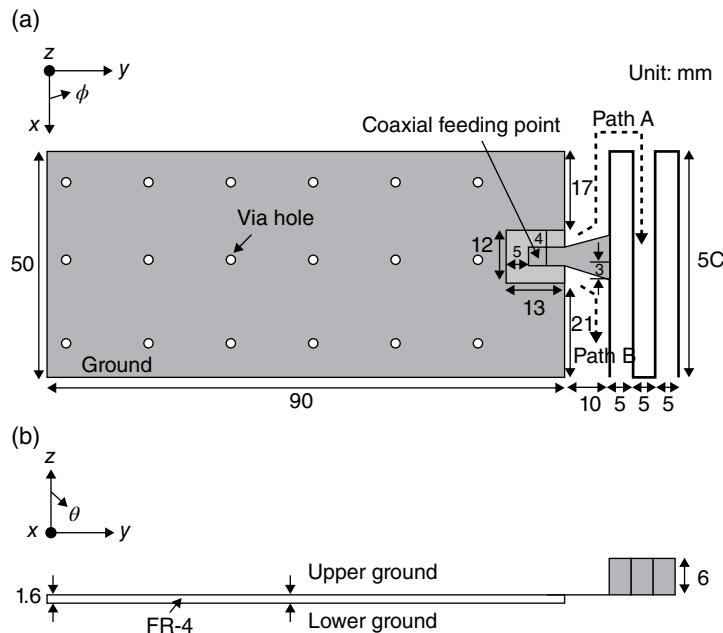


Figure 3.86 Multiband monopole antenna. (Source: Kim *et al.* [27]. Reproduced with permission of IET.)

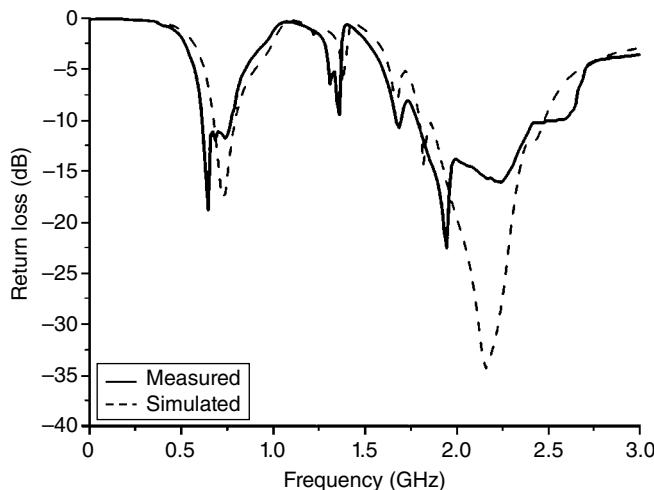


Figure 3.87 Reflection coefficients. (Source: Kim *et al.* [27]. Reproduced with permission of IET.)

Shown in Figure 3.90 is a multiband antenna [28]. It is a 3D design and its design seems quite complex. However, this design is a combination of multi-branch and multi-pitch. This antenna gives a good demonstration of how much design freedom the flex technology

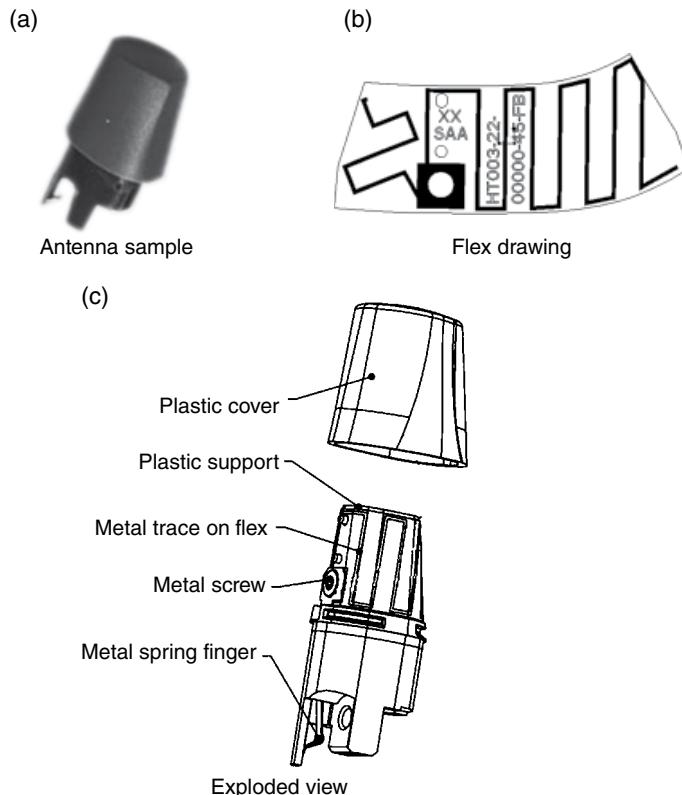


Figure 3.88 Stubby antenna made of flex. (Source: Reproduced with permission of Shanghai Amphenol Airwave.)

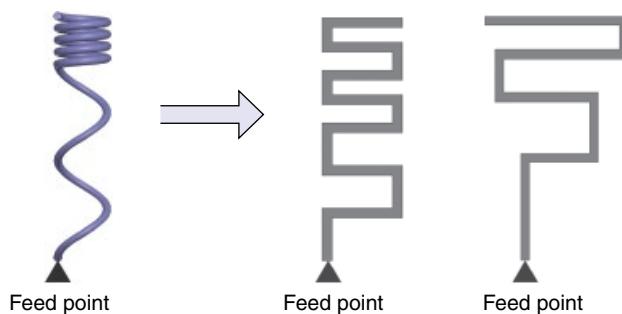


Figure 3.89 Variants of multi-pitch-type meander line antennas.

introduces. Just remember, never overdesign an antenna. An over-complex antenna will cause trouble in manufacturing and fine tuning later on.

The flex technology is also used in designs of retractable antennas. Shown in Figure 3.91 is a flex-based external antenna [29]. The antenna used a single-branch multi-band flex design. The flex design is similar to the whip-loaded helix antenna as shown in Figure 3.53. However,

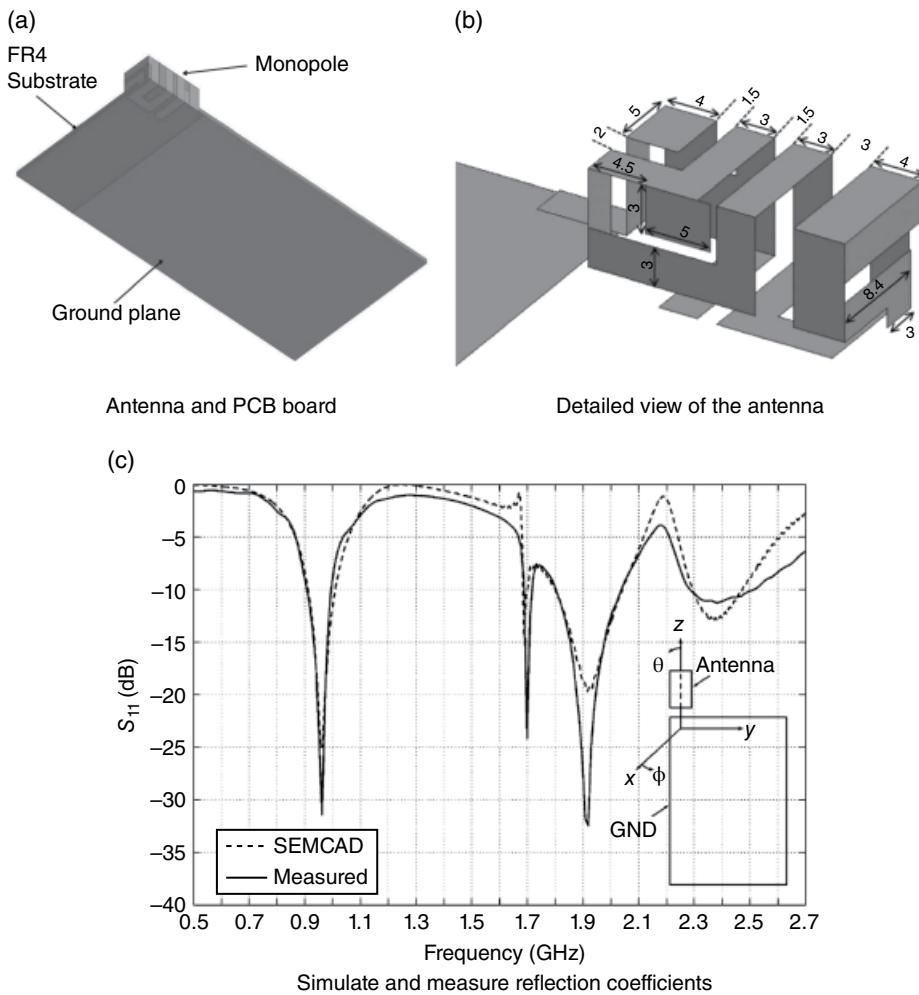


Figure 3.90 A three-dimensional multiband flex antenna. (Source: Kanj and Ali [28]. Reproduced with permission of IEEE.)

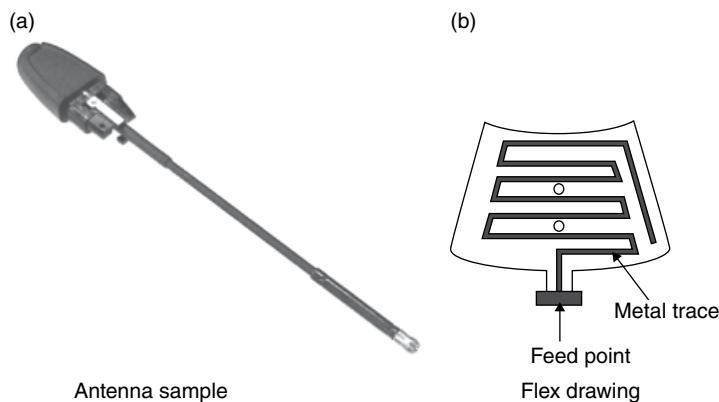


Figure 3.91 Retractable antenna made of flex. (Source: Reproduced with permission of Shanghai Amphenol Airwave.)

in a coil-based design, it is quite difficult to change the distance between the center whip and the coil. When adopting the flex technology, this parameter can easily be adjusted. Thus bandwidth between low and high band can also be easily exchanged.

3.4 Effect of Ground Plane

So far, we have always accepted a ground as it is and then designed antennas around it. On the other hand, it has been demonstrated in Figure 3.8 that a ground isn't really a ground but one of the two radiation arms of an asymmetric dipole. Thus, it should come as no surprise to know that the shape and size of a ground plane have prominent effects on the antenna built on it. Some published articles [30–33] have investigated the ground effect on a PIFA antenna or an IFA antenna. A straight-wire monopole antenna is used here to study the effect. However, one should not count the dimension of a ground plane as a design freedom, because most of the time that dimension is solely decided by a product's industry design team. Some parameter studies presented here provide quantitative examples to help explain why some antennas work well on one phone but have problems when migrated to other phones. Other examples in this section illustrate some extreme cases when the ground is too small to function well.

The reason for using a monopole instead of a PIFA is to simplify the parameter studies. As shown in Figure 3.92, there are only three significant dimensions: monopole length L_{ant} , ground width W , and ground length L .

The antenna length, L_{ant} , used in all simulations shown in Figure 3.93 is 70mm. In Figure 3.93a and b, the width of the ground plane, W , is 30mm. The width is 70mm for Figure 3.93c and d. For each ground width, three different ground lengths are compared.

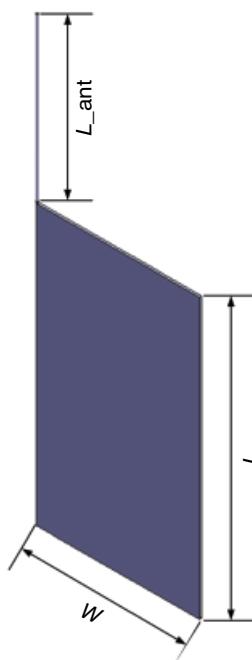


Figure 3.92 Three critical dimensions.

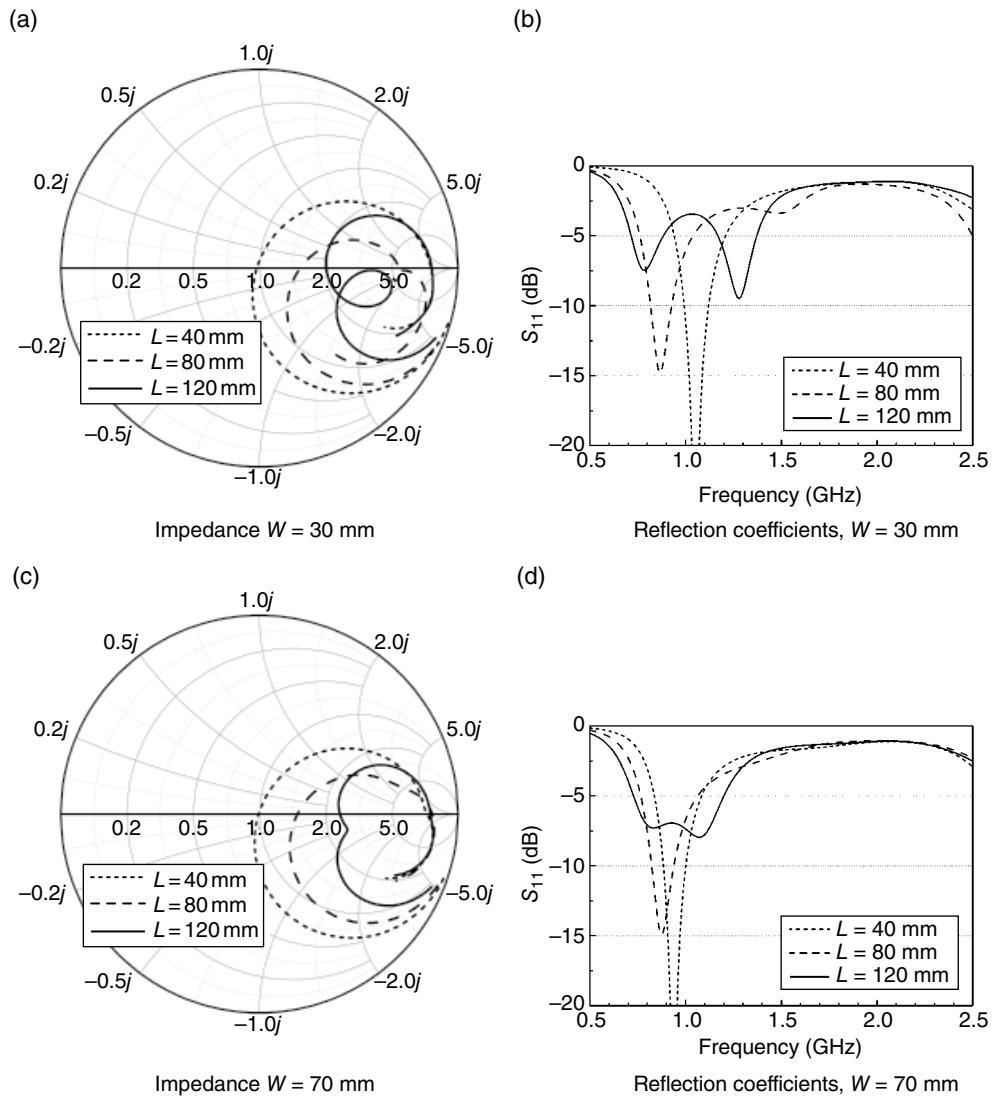


Figure 3.93 Effect of ground width and length, $L_{\text{ant}}=70 \text{ mm}$.

In an asymmetric antenna, the resonant frequency is mostly decided by the smaller arm. The larger arm has a significant impact on the impedance, bandwidth and matching, but less effect on the frequency. In practice, sometimes it is quite tricky to decide which one is the short arm. As an example, the ground's length used in the simulation which generates the short dash curve shown in Figure 3.93b is 40mm, which is shorter than the monopole's length 70mm, but the ground cannot be considered the short arm. When a ground is small, its effective length can be approximated by the sum of two adjacent edges. In this case, the ground's effective length is the sum of 40 and 30mm, which is about the same as the monopole. This

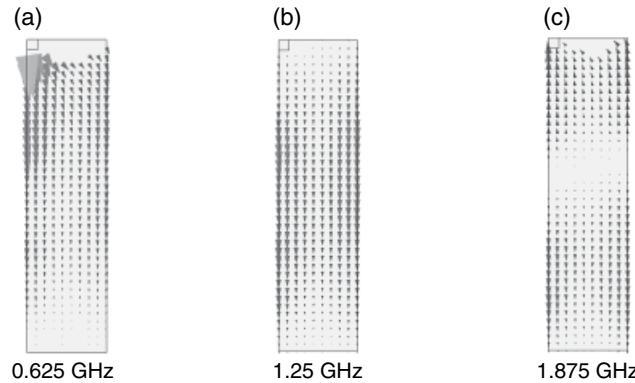


Figure 3.94 Ground's current of different modes, $W=30\text{ mm}$.

can also be confirmed by Figure 3.93b. The resonant frequency of the monopole is 1.06 GHz, while a balance dipole with unilateral arm length of 70 mm resonates at 1.07 GHz.

The ground lengths of three curves in Figure 3.93a and b are 40, 80, and 120 mm respectively. When the ground length increase, the antenna's response changes from single resonance to dual resonances. In the meantime, the impedance locus shifts toward the left side of the Smith Chart. It is obvious that a ground's length does have a significant impact on an antenna's response.

To illustrate the effect of the ground's width, all three cases shown in Figure 3.93a are simulated again with ground width increased to 70 mm. The results are shown in Figure 3.93c and d. It is clear that an antenna on a wider ground is less sensitive to the ground's length change.

The impact of ground on an antenna can be explained by the intrinsic modes of a ground itself. As shown in Figure 3.94, a $30\text{ mm} \times 120\text{ mm}$ ground can support various mode at different frequencies. At 0.625 GHz, 120 mm equals to a quarter wavelength and the current on the board represents half a standing wave as shown in Figure 3.94a. In this figure, the size of arrows represents the current amplitude. To obtain the current distributions shown in Figure 3.94a–c, a 70 mm-long monopole is used to excite the ground. The square on the top-left corner is the feeding spot. The current on monopole antennas is omitted in all plots to conserve space. The current shown in Figure 3.94b represents one standing wave. In fact, at 1.25 GHz the current on the ground is composed of both a standing wave and a downward traveling wave. Figure 3.94b is a snapshot of the time-varying current distribution on the board. This snapshot illustrates the idea of current mode quite well. However, it is not an accurate representation of the actual board current at all times. At 1.875 GHz, there is one and a half standing waves on the ground. Similar to the current at 1.25 GHz, there are both a standing wave and a downward traveling wave on the board.

When the ground becomes wider, a strong current can also be excited along the shorter edge. That spreads the current more evenly on the ground and thus makes the antenna's characteristic less sensitive to the ground length variation.

When an antenna is installed on a ground, the final antenna characteristic shows a combined effect of both elements. For example, a dipole with 70 mm unilateral arm length should resonate at 0.93 GHz. By combining a 70 mm-long monopole with a 120 mm-long ground, the antenna resonates at 0.78 and 1.28 GHz, respectively, as shown in Figure 3.93b.

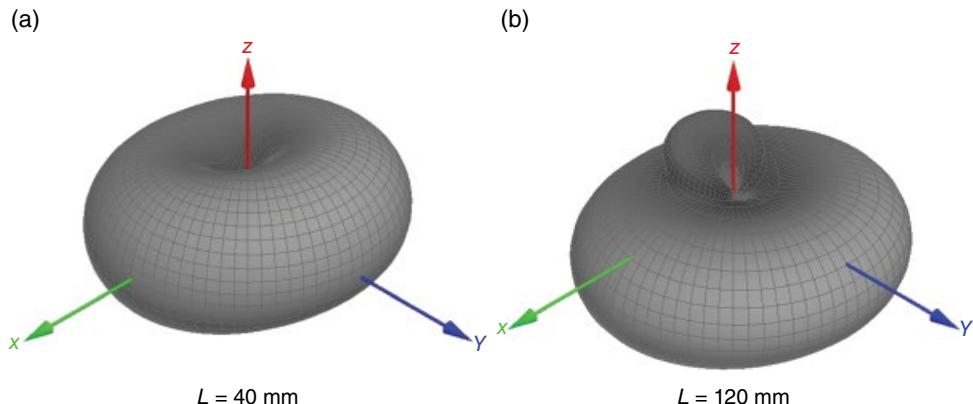


Figure 3.95 Three-dimensional radiation patterns. $L_{\text{ant}}=70 \text{ mm}$, $W=30 \text{ mm}$ at 1.2 GHz.

Besides impedance change, antennas' radiation patterns also vary when the ground's dimension is changed. Shown in Figure 3.95a is the radiation pattern of a 70 mm monopole when it is installed on a $30 \text{ mm} \times 40 \text{ mm}$ ground. The radiation pattern is quite similar to the "donut" shape, which is the trademark of a half-wavelength dipole antenna. The antenna can provide omnidirectional coverage in the azimuth plane and also has its peak gain in that plane.

When fixing other parameters and only increasing the ground length from 40 to 120 mm, the antenna's pattern tilts downward as shown in Figure 3.95b. The tilting is due to the radiation of the downward-traveling wave. The mathematical explanation can be found in many antenna books [25]; the fact we need to remember here is the radiation pattern of a traveling wave structure always tilts toward the direction in which the wave travels. In the case of monopole antennas, when the ground starts to be significantly larger than a quarter of a wavelength, a larger part of the current on the ground starts to take the form of a traveling wave, which tilts the radiation pattern down. On a cellular phone, it means an antenna's pattern at high frequency bands always tilts downward when the antenna is installed on the upper portion of the phone.

Shown in Figure 3.96 are the results of parameter studies of ground length. The monopole length is 30 mm and its approximate resonant frequency is 2 GHz. Three different board lengths, 40, 80, and 120 mm, are simulated. Shown in Figure 3.96a and b are results of narrow grounds, whose widths are 30 mm. Shown in Figure 3.96c and d are the results of 70 mm-wide grounds. Although the absolute ground dimensions used in Figure 3.96 are identical to Figure 3.95, the monopole length is different, thus the working frequency is different. At 2 GHz, the wavelength is less than half of that at 0.93 GHz, thus from the electrical point of view, the ground's relative dimension is more than doubled at 2 GHz. That explains why all the results shown in Figure 3.96 are less fluctuating. Comparing Figure 3.96b and d, it is clear that an antenna's impedance is less sensitive to ground length change when the working frequency is higher.

So far, we have discussed the ground's impact on antennas' impedance when the ground grows bigger. Logically one will ask what will happen if a ground becomes smaller. Shown

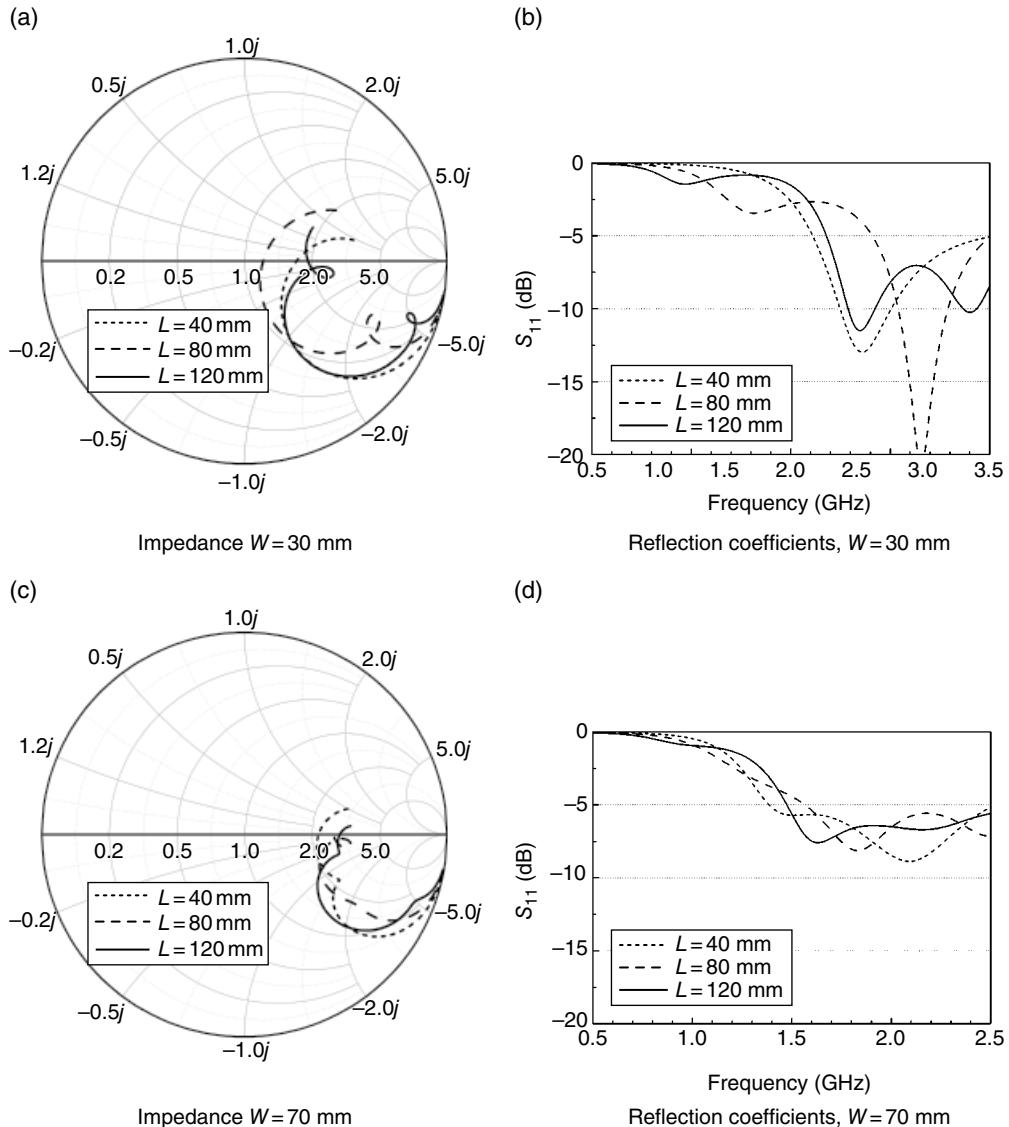


Figure 3.96 Effect of ground width and length, $L_{\text{ant}}=30 \text{ mm}$.

in Figure 3.97 are results of a 70 mm-long monopole antenna installed on three small grounds, respectively. All three lines in Figure 3.97 illustrate responses of single resonance. With the ground becomes smaller, the resonant frequency shifts higher. One can shift the resonant lower by making the monopole longer, but only to a certain degree. It is quite difficult to achieve the required specification when designing a multiband antenna on a small ground.

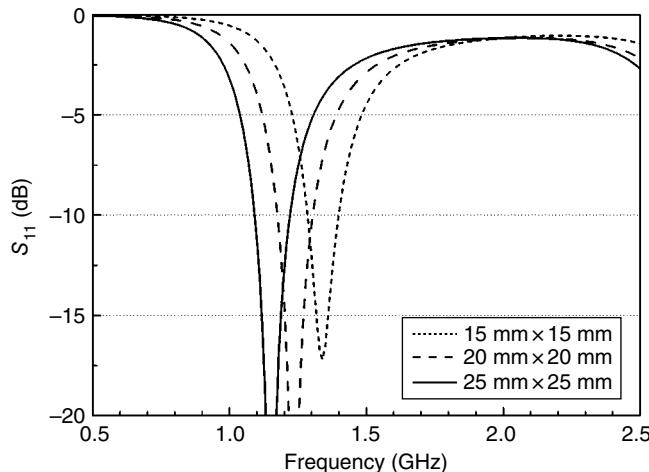


Figure 3.97 Effect of small ground, $L_{\text{ant}} = 70 \text{ mm}$.

In summary, although antenna engineers do not have too much freedom in choosing a ground's dimension, we must be aware that the ground is one of the decisive factors influencing an antenna's bandwidth, radiation pattern, and efficiency. Generally speaking, a larger ground is always better.

References

- [1] Iskander, M.F. (2000) *Electromagnetic Fields and Waves*, 1st edn, Waveland Press Inc.
- [2] Sadiku, M.O. (2009) *Elements of Electromagnetics*, 5th edn, Oxford University Press, USA.
- [3] Ulaby, F.T., Michielssen, E., and Ravaioli, U. (2010) *Fundamentals of Applied Electromagnetics*, 6th edn, Prentice Hall.
- [4] Buck, J. and Hayt, W. (2005) *Engineering Electromagnetics*, 7th edn, McGraw-Hill Science/Engineering/Math.
- [5] Moore, T.G., Borisov, E., Derdzinski, A., and Repplinger, D. (2002) Antenna assembly and multiband stubby antenna. US6369775B1.
- [6] Sutter, R.W., Weaver, W.B., and Dou, W. (2005) Multiple pitch antenna assembly. US6922178B2.
- [7] Zhinong, Y. (2000) Multi-band non-uniform helical antennas. US6112102.
- [8] "The Federal Communications Commission of United States, First Report and Order, Revision of Part 15 of the Commission's Rule Regarding Ultra Wide Band Transmission Systems," http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-02-48A1.pdf. Retrieved 25 October 2010.
- [9] Abbosh, A.M., Bialkowski, M.E., Mazierska, J., and Jacob, M.V. (2006) "A planar UWB antenna with signal rejection capability in the 4–6 GHz band," *IEEE Microwave and Wireless Components Letters*, **16**, 278–280.
- [10] Jianxin, L., Chiau, C.C., Chen, X., and Parini, C.G. (2005) "Study of a printed circular disc monopole antenna for UWB systems," *IEEE Transactions on Antennas and Propagation*, **53**, 3500–3504.
- [11] Kiminami, K., Hirata, A., and Shiozawa, T. (2004) "Double-sided printed bow-tie antenna for UWB communications," *IEEE Antennas and Wireless Propagation Letters*, **3**, 152–153.
- [12] Lin, C.C., Kan, Y.C., Kuo, L.C., and Chuang, H.R. (2005) "A planar triangular monopole antenna for UWB communication," *IEEE Microwave and Wireless Components Letters*, **15**, 624–626.
- [13] Low, Z.N., Cheong, J.H., and Law, C.L. (2005) "Low-cost PCB antenna for UWB applications," *IEEE Antennas and Wireless Propagation Letters*, **4**, 237–239.
- [14] See, T.S.P. and Chen, Z.N. (2008) "An electromagnetically coupled UWB plate antenna," *IEEE Transactions on Antennas and Propagation*, **56**, 1476–1479.

- [15] Verbiest, J.R. and Vandenbosch, G.A.E. (2006) "Small-size planar triangular monopole antenna for UWB WBAN applications," *Electronics Letters*, **42**, 566–567.
- [16] Verbiest, J.R. and Vandenbosch, G.A.E. (2006) "A novel small-size printed tapered monopole antenna for UWB WBAN," *IEEE Antennas and Wireless Propagation Letters*, **5**, 377–379.
- [17] Toh, W.K., Chen, Z.N., and Qing, X. (2009) "A planar UWB antenna with a broadband feeding structure," *IEEE Transactions on Antennas and Propagation*, **57**, 2172–2175.
- [18] Toh, W.K., Chen, Z.N., Qing, X., and See, T. (2009) "A planar UWB diversity antenna," *IEEE Transactions on Antennas and Propagation*, **57**, 3467–3473.
- [19] Wong, K.L. and Chien, S.L. (2005) "Wide-band cylindrical monopole antenna for mobile phone," *IEEE Transactions on Antennas and Propagation*, **53**, 2756–2758.
- [20] Low, X.N., Chen, Z.N., and See, T.S.P. (2009) "A UWB dipole antenna with enhanced impedance and gain performance," *IEEE Transactions on Antennas and Propagation*, **57**, 2959–2966.
- [21] Cho, Y.J., Kim, K.H., Choi, D.H. et al. (2006) "A miniature UWB planar monopole antenna with 5-GHz band-rejection filter and the time-domain characteristics," *IEEE Transactions on Antennas and Propagation*, **54**, 1453–1460.
- [22] Ren, Y.-J. and Chang, K. (2006) "An annual ring antenna for UWB communications," *IEEE Antennas and Wireless Propagation Letters*, **5**, 274–276.
- [23] Zhang, Z., Langer, J.C., Sutter, R., and Kfouri, T. (2007) Multiple band antenna and antenna assembly. US7161538B2.
- [24] Zhang, Z., Langer, J.C., Li, K., and Iskander, M.F. (2008) "Design of ultrawideband mobile phone stubby antenna (824 MHz–6GHz)," *IEEE Transactions on Antennas and Propagation*, **56**, 2107–2111.
- [25] Balanis, C.A. (2005) *Antenna Theory: Analysis and Design*, 3rd edn, Wiley-Interscience.
- [26] Khaleghi, A. (2007) "Dual band meander line antenna for wireless LAN communication," *IEEE Transactions on Antennas and Propagation*, **55**, 1004–1009.
- [27] Kim, S.C., Lee, S.H., and Kim, Y.S. (2008) "Multi-band monopole antenna using meander structure for hand-held terminals," *Electronics Letters*, **44**, 331–332.
- [28] Kanj, H. and Ali, S.M. (2009) "Compact multiband folded 3-D monopole antenna," *IEEE Antennas and Wireless Propagation Letters*, **8**, 185–188.
- [29] Kfouri, T. and Mathews, J. (2006) Clipped contact whip and flex antenna assembly for a device. US7095375.
- [30] Arkko, A.T. (2003) Effect of ground plane size on the free-space performance of a mobile handset PIFA antenna. Twelfth International Conference on Antennas and Propagation, 2003. (ICAP 2003). (Conf. Publ. No. 491). vol. 1, pp. 316–319.
- [31] Chan, K.H., Fung, L.C., Leung, S.W., and Siu, Y.M. (2006) Effect of internal patch antenna ground plane on SAR. 17th International Zurich Symposium on Electromagnetic Compatibility, 2006. EMC-Zurich 2006, pp. 513–516.
- [32] Fujio, S. (2006) Effect of ground size on plate inverted-F antenna. International Workshop on Antenna Technology Small Antennas and Novel Metamaterials, 2006 IEEE, pp. 269–272.
- [33] Urban, R. and Peixeiro, C. (2004) Ground plane size effects on a microstrip patch antenna for small handsets. 15th International Conference on Microwaves, Radar and Wireless Communications, 2004. MIKON-2004, vol. 2, pp. 521–524.

4

Internal Antenna

The internal antenna emerged around the end of the twentieth century and gained its popularity in merely a few years. Nokia was the first mainstream cell phone company who successfully marketed phones with internal antennas. As shown in Figure 4.1, the Nokia 3210, which came onto the market in 1999, was a huge hit and more than 300 million of them were eventually sold around the world. “Candy-bar phones,” a nickname for those phones with internal antennas strongly resonates with the European fashion taste, which appreciates minimalism and sleek curves. With that momentum, Nokia successfully eroded Motorola’s market share in Europe.

After that, there were a few years of *de facto* draw between Nokia and Motorola, who shared the first and second places on the global chart of cell phone sales. Nokia claimed that it sold the most phones and Motorola insisted that its revenue of cell phones was the highest. Both claims were correct, because Motorola was still holding onto the high-end market. What happened next was quite theatrical. Nokia wanted to use only candy-bar phones to conquer the US market but kept failing. US customers didn’t want to give up their love of flip phones, which had been invented by Motorola. Nokia did not want to develop flip phones and hoped, through advertising and customer education, that Americans’ attitude toward candy-bar phones could be converted in a relatively short period. Sadly, history demonstrates again how strong the persistence of people’s habits is.

Nokia’s love of internal antenna is not really emotional; it is because there are some technical and financial merits inherent in internal antennas. A candy-bar phone with an internal antenna has no moving parts and only one printed circuit board (PCB). A handful of screws can securely assemble all the core parts. The front and back covers, which are made of acrylonitrile butadiene styrene (ABS plastic), are snapped onto the core without any screws. When a candy-bar phone is accidentally dropped on the ground, the covers take most of the force. Because there are no screws attached to the covers, there is no extreme stress point. When the force is too strong to bear, the snap feature just breaks loose and absorbs all energy in the



Figure 4.1 Nokia 3210. (Source: Reproduced with permission of Nokia.)

process. After the drop, a candy-bar phone might fall apart; however, nothing is really broken; one can put it all back together and the phone is ready to work again. The robustness of a candy-bar phone is one of critical factors that made Nokia so successful in the markets of the developing countries. Due to the cut-throat competition, the margin of low-tier phones is razor thin, and an increase of field failure rate (FFR) by a small percentage might totally wipe out the profits.

Unlike flip phones, candy-bar phones do not have features such as flip, hinge, board-to-board connector, and so on; the shorter build of material (BOM) of candy-bar phones also means less manufacturing cost. Combining the cost-saving on the stages of both manufacturing and customer service, Nokia can sell a low-tier phone at a price which is lower than Motorola's cost and still make decent money. For Nokia, moving into the field of flip phones with external antennas meant fighting a war in its enemy's territory while losing most of its competitive advantage. Nokia's own wishful thinking was that it could corner Motorola in the US market with candy-bar phones with internal antennas combined with a price war, but apparently this didn't work.

On the other hand, Motorola did not want to fight in the field of candy-bar phones either. The reason for that is because of Nokia's fearful logistic capability and its sophisticated

global sourcing system. For most components used in Nokia, there is more than one vendor in the supplier pool. By doing this, Nokia can hammer vendors and demand a lower price and mitigate the possible part shortage due to any single vendor's failure. In the meantime, the possible vendor of any component in the supplier pool is limited to a handful. There is a strict process in place to qualify a potential vendor, only heavyweight players in the according discipline have the possibility of being qualified. The strict screening process ensures the quality of parts used in phones, thus assuring the reliability in phone manufacturing. All these issues might sound quite simple, but in fact it is quite a challenge to implement them successfully.

Finally, both sides moved into their counterpart's territory. It was a fierce fight, which eventually expanded to new fields with the emergence of new form of factors, such as slide phones and swivel phones. It now all seems a distant history. When the first version of this book was written, in 2009, Nokia claimed more than 40% of the global phone market and Motorola's market share was wandering in the territory of single digit. With the blow of the ongoing economic crisis, which was the worst since the 1930s, Motorola's fate did not look bright.

When the second version of this book was written, in 2015, a lot has happened. The mobile division of Motorola had been sold to Google for \$12.5 billion in 2012. Google then resold the Motorola Mobility to Lenovo for \$2.91 billion in 2014. Google still retains ownership of the vast majority of Motorola Mobility's patents. Nokia Mobile's market share had collapsed from the above 40% in 2009 to less than 10% in 2012. Microsoft bought Nokia Mobile in 2013 for \$7.2 billion. Microsoft failed to turn around Nokia Mobile and layoff off 12,500 previous Nokia's employees in 2014.

The disruptive force behind all that has happened is Apple. In 1997, everything seemed against Apple, the inventor of the personal computer, and bankruptcy seemed its only fate. Then suddenly, with the return of the expelled founder Steve Jobs, everything turned around. The first iPhone, iPhone 2G, was released in 2007 and sold more than 10 million in the next year. In 2011, Apple sold 71 million iPhone and Steve Jobs passed away in the same year. Many people were wondering whether Apple might lose its direction and luckily it didn't. Apple has sold 169 million iPhone in 2014. Since iPhone 2G, every generation of iPhones adopted internal antennas. Although Google has failed to revive Motorola Mobile, Google's phone operation system, Android, has dominated smartphone market. Except iPhone, which uses its own iOS system, everyone else pretty much all use Android. In 2015, the first-tier smartphone vendors are Apple and Samsung, each claims around 20% worldwide market share. Three Chinese companies, Lenovo, Huawei, and Xiaomi, are distant followers. Each takes around 5–6% market share. In 2015, all phone models released by those companies use internal antennas.

Although to a customer, all internal antennas might have the same appearance, that is, no visible evidence of an antenna's existence, there is more than one kind of internal antennas if we categorize them from a technical point of view. We are going to discuss some of them in this chapter. The aim of the book is to introduce the world of antenna designing. With the knowledge shared in the book, one should be able to design an antenna for most applications. In some tightly constrained applications, such as designing a multiband antenna in a tiny volume, the antennas introduced here may not be enough to fulfill the requirements. In those circumstances, a book by Professor Wong [1], *Planar Antennas for Wireless Communications*, provides more advanced techniques to meet such challenges.

4.1 Inverted-F Antenna

Let's start with the inverted-F antenna (IFA). We know that a monopole antenna works pretty well for mobile applications. However, the size of a monopole is too big to be adopted in most state-of-the-art phones. Naturally, L-shaped antennas as shown in Figure 4.2 were proposed as a measure to decrease the total height of an antenna.

While keeping the total antenna's length, say, 70 mm, the antenna whip can be bent into an L shape in different positions, thus providing a different separation, G , from the ground plane. The dimension G is a critical parameter which decides the antenna's impedance. Shown in Figure 4.3 are simulated results of a straight whip antenna and two L-shaped antennas. When the horizontal arm approaches the ground plane, the real part of the antenna's impedance becomes smaller. That means the impedance locus sweeps a larger circle on the Smith chart. As the impedance of a straight whip is around 100Ω , when the whip is bent and approaching the ground, the matching of the antenna is actually improved at first. In this example, the antenna with the G equaling 14 mm is the best matched one of all three. However, if the separation between whip and ground is less than 14 mm, the matching starts to degrade.

Using the matching techniques discussed in Chapter 2, good matching can be easily achieved. Either a shunt inductor or a shunt capacitor can be used to match the antenna. Shown in Figure 4.4 are simulated results. A lumped inductor is used to match the antenna with $G=6\text{ mm}$. The dashed line, solid line, and dotted line correspond to three different matching values. For a shunt inductor, a smaller value has more impact on the port impedance. A 6.8 nH inductor provides the best matching.

By replacing the lumped inductor with a grounding metal strip, what we get is an IFA as shown in Figure 4.5. The shape of the antenna element is like the letter "F" on its side; that is why it is called "inverted F." Shown in Figure 4.6 are the results of two IFAs. For the sake of consistency, the total whip length of both antennas is kept to 70 mm if measured from the feeding point to the end of the whip. The separations between the feeding and grounding strips are chosen to be 2 and 5 mm, respectively. Comparing Figures 4.4 and 4.6, it is clear that the grounding strip indeed functions as a shunt inductor. As the inductance of the grounding strip

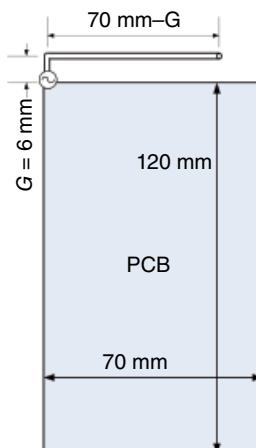


Figure 4.2 L-shaped antenna.

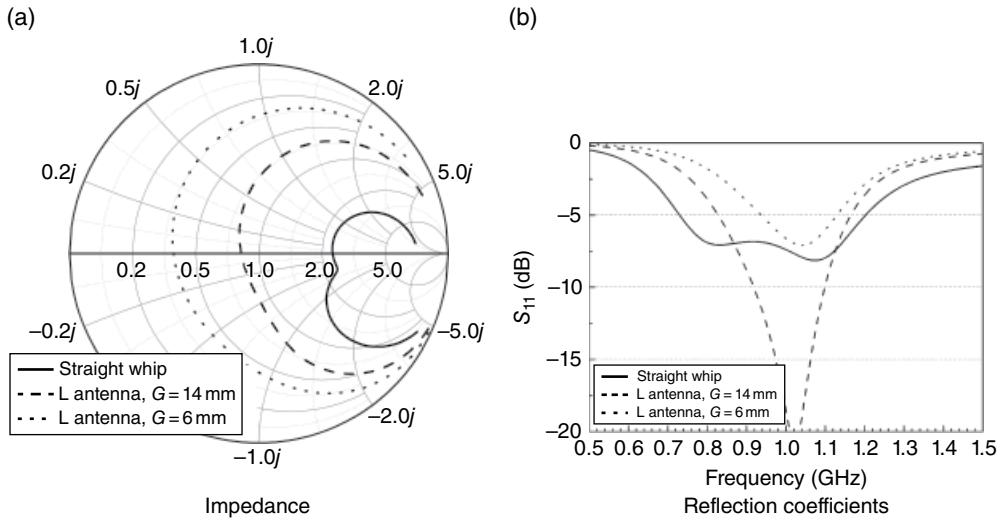


Figure 4.3 Impact of separation G on an L-shaped antenna.

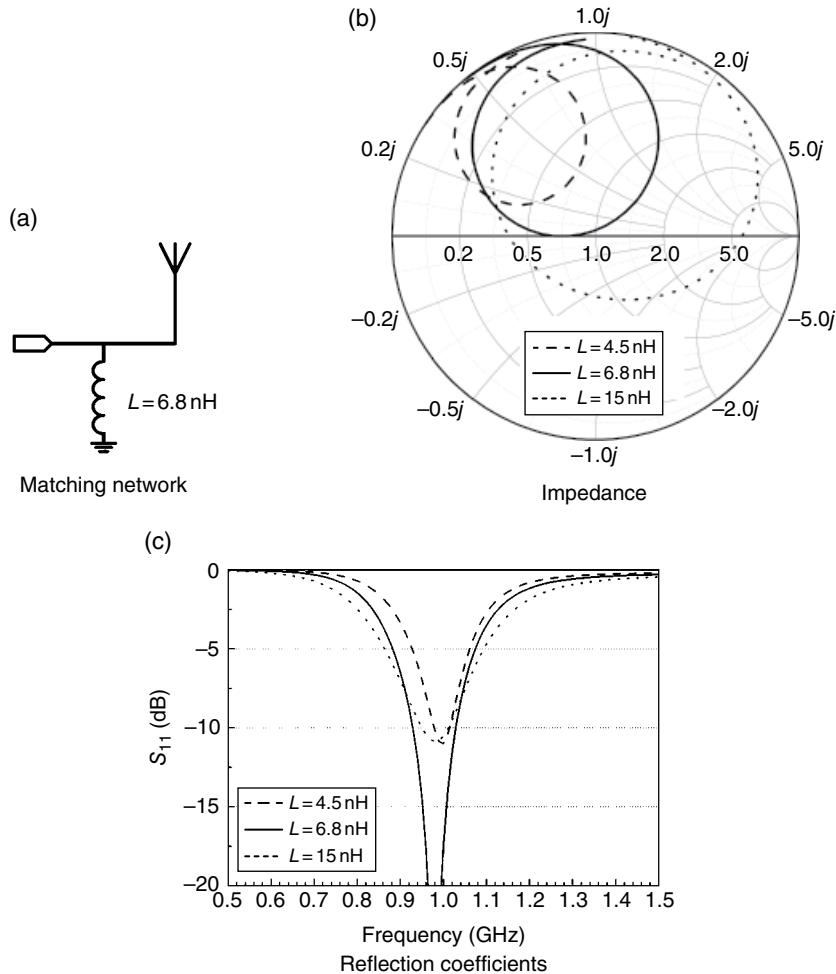


Figure 4.4 Using a shunt inductor to match the antenna with $G = 6 \text{ mm}$.

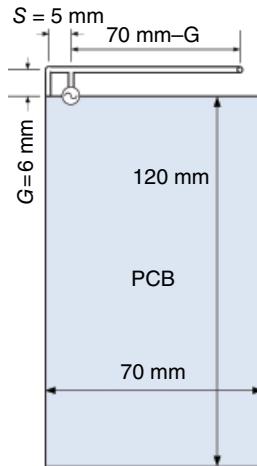


Figure 4.5 Inverted-F antenna (IFA).

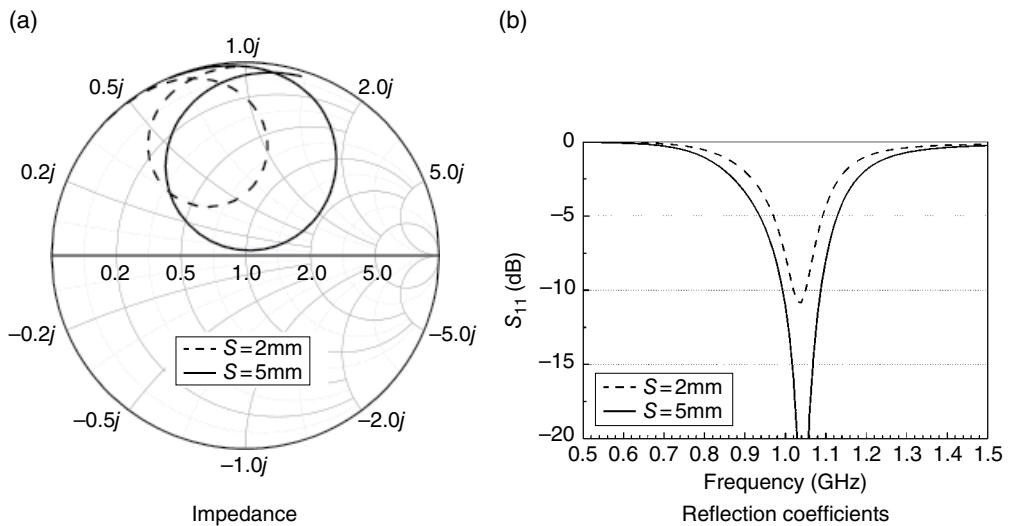


Figure 4.6 Impact of separation S on an IFA.

is decided by its total length, when the separation between the feeding and grounding strips becomes shorter, the distributed inductance become smaller, thus the grounding strip has more impact.

As a rule of thumb, whenever the separation between the arm of an IFA and ground gets smaller, the grounding strip must be placed closer to the feeding strip to compensate for the impedance change.

In practice, designing a single-band IFA is quite easy. The following are some design steps:

- Start with a strip a little longer than a quarter of wavelength. A 4 mm gap between grounding and feeding strips can be used as the starting point.
- Tune the antenna length to get a resonance, which is slightly lower than the target frequency.
- Check the impedance on the Smith chart. The impedance must look similar to one of the three curves shown in Figure 4.4. If it resembles the dashed line, increase the gap S . If it resembles the dotted line, decrease S . The solid line is the preferred option.
- Fine-tune the antenna length again until the correct resonant frequency is obtained.

Looking at Figures 4.4 and 4.6 more closely, the difference between resonant frequencies can be observed. When the antenna lengths are identical, an L-shaped antenna can work at a lower frequency. To qualitatively explain this, a simplified equivalent circuit of an IFA is given in Figure 4.7. Unlike an L-shaped antenna, whose resonant frequency is decided by its whole whip, only a part of IFA, which is marked by the dark color in Figure 4.7, is the active element. The effect of grounding and feeding strips can be approximated by a shunt inductor and a series inductor, respectively. Of course, this is only a coarse simplification, when the distance between feeding and grounding strip become too short, the distributed shunt capacitance between two strips can also play a role in the antenna performance.

As we know that E fields excited by a whip antenna are always parallel to the whip itself as shown in Figure 4.8a, one might assume that the E field generated by an IFA should rotate 90°, as the whip is rotated 90°. In fact, the rotation of an IFA's whip has little impact on the polarization of radiation pattern. Similar to monopole antennas, an IFA cannot work by itself. It must be installed on a ground and it is the ground that decides the E -field polarization. As shown in Figure 4.8b, when the ground is vertically placed, the E field is also vertically polarized.

Shown in Figure 4.9 are simulated three-dimensional (3D) radiation patterns of an IFA. The antenna is the same one simulated in Figure 4.6 with $S=5$ mm. Shown in Figure 4.9a is the 3D

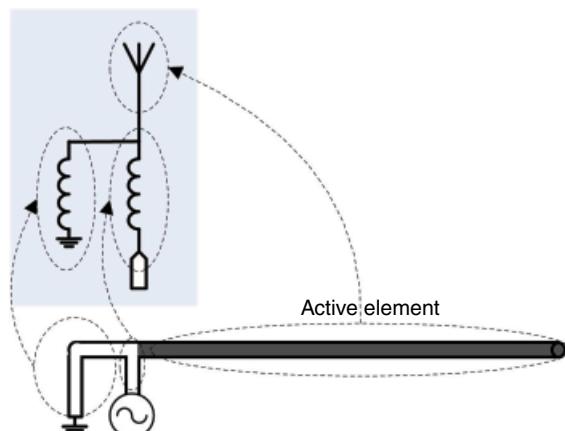


Figure 4.7 Simplified equivalent circuit of an IFA.

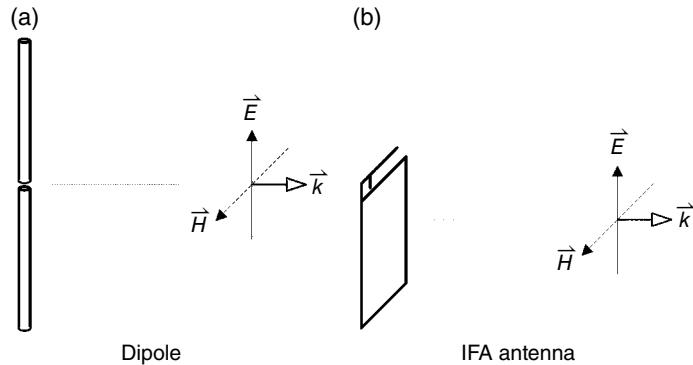


Figure 4.8 E field excited by antennas.

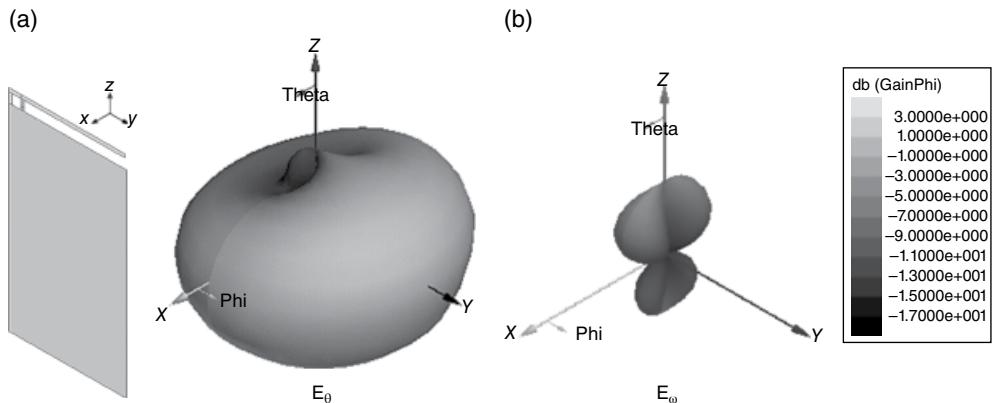


Figure 4.9 Three-dimensional radiation patterns, IFA, $S=5$ mm, 1.05 GHz.

pattern of E_θ , which is the vertical polarization part of total E field. Figure 4.9b is the E_φ , the horizontal part of the total E field. The amplitude unit used in both plots is dB. The scale they used is also the same. It is clear that the vertical E field is the dominant radiation component. The peak gain of the vertical E field is 3.4 dBi. That of the horizontal E field is only -5.8 dBi. At a low frequency, such as 1.05 GHz, the radiation pattern of the ground is similar to a dipole, which is donut-shaped and approximately omnidirectional in the azimuth plane. At a higher frequency, because the ground is much larger than one wavelength, the radiation pattern will split. This behavior is similar to that of a monopole antenna. Detailed studies of the ground effect can be found in Section 3.4 and published papers [2–5].

An IFA does not provide a lot of design freedom, so it is mostly used in single-band applications, such as GPS, BT, or WiFi. Although an IFA is similar to an L antenna with a shunt matching inductor, there is a minor difference. Because the active length of an L antenna starts from the feeding point, it can be designed to be smaller than an IFA. An L antenna also requires a smaller area on the PCB, because it only needs one contact pad. The drawback of an L antenna is that the lumped matching inductor it requires has some inherent losses, thus

the efficiency of an L antenna is normally a few tenths of dB lower than an IFA's. If there is enough board area for two pads, one for grounding and one for feeding, an IFA is a better choice, as it provides better efficiency and also saves a matching component, which costs a few pennies.

4.2 Planar IFA

4.2.1 Single-Band PIFA

Shown in Figure 4.10 is a planar IFA (PIFA). PIFAs can be thought of as a mutation of IFAs. Both IFA and PIFA have a ground strip and a feeding strip. By replacing the radiating strip of an IFA with a patch, we get a PIFA. In most cases, the patch of a PIFA is above the ground plane.

In all the following samples in this section, the width of both feeding and grounding strip is 1 mm; the dimension of PCB, which is used as the ground, is 50 mm \times 120 mm. If the PCB size is changed, the antenna response will change accordingly. The ground's impact on a PIFA is the same as that on an external antenna. Refer to Section 3.4 for further details.

In Figure 4.10, the dimension H is the distance between a patch and a ground. The S is the edge-to-edge distance between feeding and grounding strips. Similar to IFAs, the matching of antennas can be tuned by adjusting S .

Let's start with single-band PIFAs. Shown in Figure 4.11 are simulated results of three PIFAs. The dimension L of all three antennas is 50 mm, which is the same as the PCB width. The dimension W of three antennas is 5, 15, and 25 mm, respectively. To best match individual antennas, the dimension S , the distance between feeding and ground strips, is tuned case by

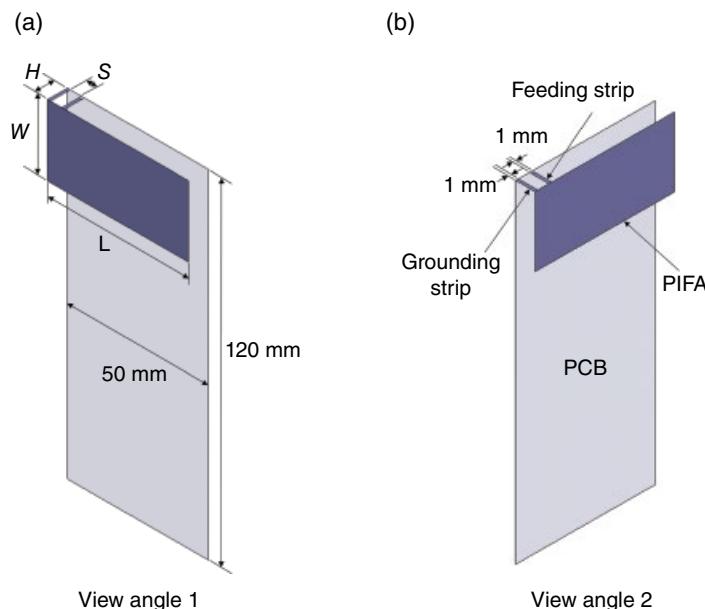


Figure 4.10 Planar inverted-F antenna (PIFA).

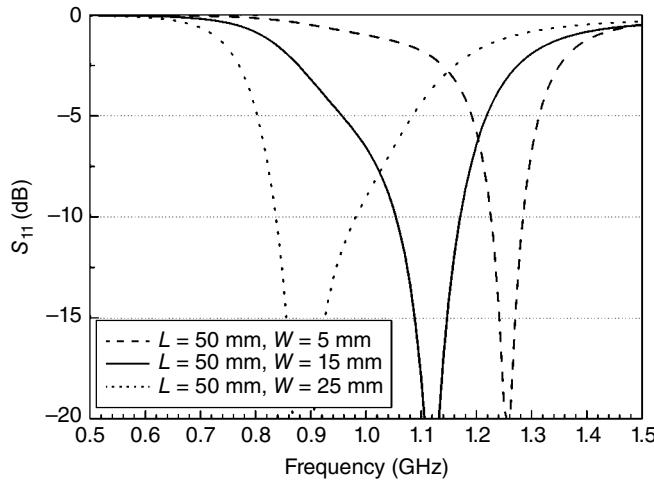


Figure 4.11 Three PIFAs with different widths $H=7$ mm.

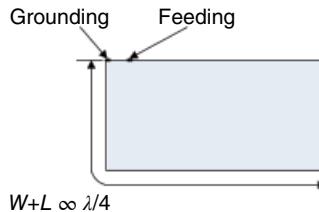


Figure 4.12 Estimating the resonant frequency of a PIFA.

case. The optimal S for 5, 15, and 25 mm patches are 1.5, 4, and 4 mm, respectively. When the W is 5 mm, the antenna is more like an IFA than a PIFA. With the increase in W , the resonant frequency of a PIFA decreases. This phenomenon gives us a hint that, unlike an IFA, the resonant frequency of a PIFA is not solely decided by the patch's length.

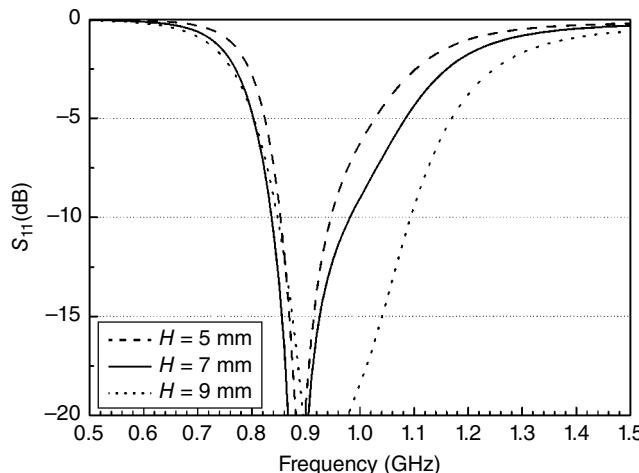
For a PIFA, the resonant frequency is proportional to the summary of its length and width. As shown in Figure 4.12, a PIFA's width plus its length is roughly a quarter of a wavelength at its resonant frequency.

When discussing a single-band PIFA, most books prefer to give a current distribution which shows two clear current flows. Both of them start from the upper-left corner and end at the bottom-right corner. One of them passes the bottom-left corner and the other one passes the opposite top-right corner. That is a quite intuitive but coarse simplification. The real current distribution is a little more complex than that. In many cases, it is difficult to see a clear current pattern when you simulate a PIFA.

Shown in Table 4.1 is a comparison between three antennas. The second column in Table 4.1 is the calculated resonant frequencies based on the assumption that $W+L$ equals to a quarter of wavelength. The third column is the actual frequency. The fourth column is the ratio between the calculated frequencies and the actual ones. The actual resonant frequency of a patch

Table 4.1 Accuracy of simplified frequency calculation formula

	Calculated frequency (GHz) $\lambda/4 = (W+L)$	Actual frequency (GHz)	Actual frequency/ calculated frequency
$W=5 \text{ mm}, L=50 \text{ mm}$	1.36	1.26	0.92
$W=15 \text{ mm}, L=50 \text{ mm}$	1.15	1.12	0.97
$W=25 \text{ mm}, L=50 \text{ mm}$	1.00	0.89	0.89

**Figure 4.13** PIFA's bandwidth vs. height. $L=50 \text{ mm}$, $W=25 \text{ mm}$.

is normally lower than the theoretical one. That is due to the loading effect of the edge capacitance between patch and ground.

The critical parameter, which controls the bandwidth of a PIFA, is H , the distance between a patch and a ground. Shown in Figure 4.13 are the reflection coefficients of three antennas with different heights. All patches have the same L and W , which are 50 and 25 mm, respectively. The H of three patches is 5, 7, and 9 mm, respectively. All three antennas are individually tuned, their S are 3, 4, and 7 mm accordingly. By increasing the H from 5 to 9 mm, an antenna's -10 dB bandwidth is almost tripled.

As a summary, there are three design guidelines for single-band PIFA:

1. The dimension H is the most critical parameter which decides an antenna's overall performance. However, because H is directly related to a device's thickness, it is a tough task to bargain about it with industry design engineers or mechanical engineers. If there is only a 5 mm height clearance and a quad-band GSM antenna is required, then PIFA is not an appropriate candidate. A folded monopole antenna, discussed in Section 4.3, is a better choice.
2. By adjusting the total length $W+L$, the antenna's working frequency can be tuned. In practice, the L should be similar to the PCB's width that results in wider bandwidth.

3. The dimension S is the tuning parameter which optimizes an antenna's matching. One can either use a wider strip with a larger S or a narrower strip with a smaller S to achieve the same matching. A larger S is better for manufacturing, because a fixed process tolerance causes less electrical variance in the antenna. From a mechanical point of view, a narrower strip is better for spring finger type of designs, because it is more flexible and it is easier to achieve the required spring force.

4.2.2 Multiband PIFA Antenna with Slits

If you have already discovered PIFAs, then you might find that the following explanation is somewhat different from other books. All versions of PIFA explanations are partially correct, including the following version. As a PIFA is a distributed radiating system, too many parameters play a role in deciding the antenna's characteristic. To explain such a complex matter, one has to substantially simplify it and that is the reason for the different explanations.

There are several ways to design a multiband PIFA. The most popular way is to cut slits on the patch, as shown in Figure 4.14. Similar to a single-band PIFA, a dual-band PIFA also has a grounding strip and a feeding strip. Two critical dimensions, D and C , are used here to describe the slit. The dimension D is the distance between the patch's corner and the opening of the slit. The dimension C is the slit's length. The dimension P is the distance between the horizontal part of slit and the edge of patch. The P is not really a critical dimension; however, it decides the shape of the slit.

Shown in Figure 4.15 are the decisive paths of different bands. Figure 4.15 only serves the purpose of predicting the trends of frequency shifting when different dimensions are adjusted. One should not try to use the illustrated critical path shown in Figure 4.15 as a way of

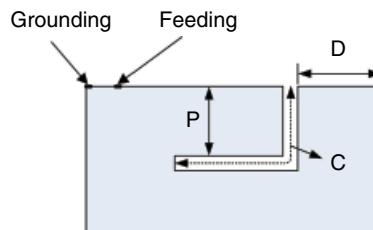


Figure 4.14 Dual-band PIFA.

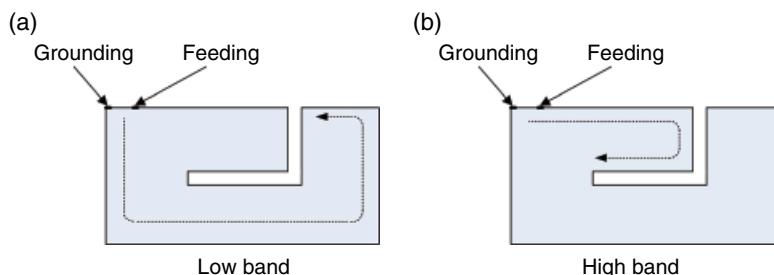


Figure 4.15 Decisive paths of different bands.

calculating a patch's resonant frequency. In reality, trying to predict a patch's resonant frequency with a calculator is not really an efficient design approach. There are always so many unknown parameters, such as the permittivity of both the antenna support and the phone cover, the electrical property of nearby objects, and so on. Usually, all these unknowns can make the theoretical calculation far removed from the correct answer. A better approach is to memorize the effect of each parameter first, then make a PIFA with a piece of copper tape, install it on a mock-up phone, cut a slit on the PIFA, and tune it by trial and error.

The impacts of the two critical dimensions are listed below:

1. The distance D can affect the resonant frequency of both the lower and the higher bands. It always shifts two bands in the opposite directions. For example, if we decrease D , the critical path of the lower band will be shorter, thus increasing its resonant frequency. In the meantime, the critical path of the higher band is increased, thus inducing a lower resonant frequency.
2. The slit's length C only influences the higher band. The slit is about half the total critical path of the higher band. Increase C can lower the high band.

Of course, in the real world, the effect of any dimension change cannot be exclusively constrained in a single band. In fact, it is safer to claim that it has more impact on some bands than others. To help understand the design rules, the following design examples are provided. In all the following examples, the area of patch is $50\text{ mm} \times 25\text{ mm}$; the area of ground is $120\text{ mm} \times 50\text{ mm}$. H , which is the distance between ground and patch as shown in Figure 4.10, is 7 mm. The impact of D is illustrated in Figure 4.16. By adjusting D , both low and high resonant frequencies can be tuned simultaneously and in opposite directions.

Illustrated in Figure 4.17 are the impacts of slot length C . With the increase of a slot's length, the resonant frequency of the higher band decreases. It can be seen that the resonant frequency at the lower band still drifts a little bit. For all three cases shown in Figure 4.17, the

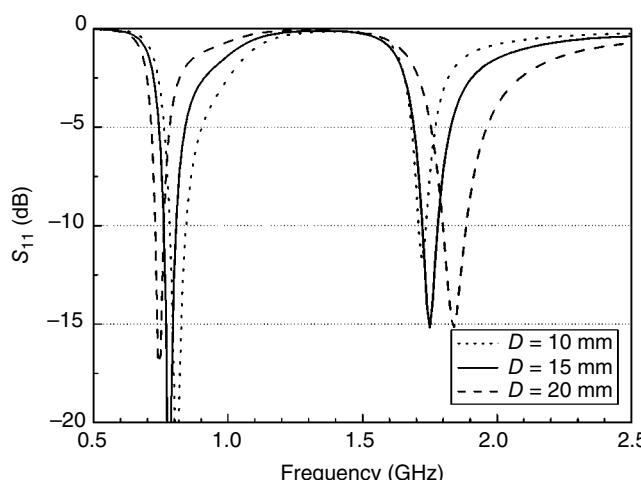


Figure 4.16 Impact of dimension D . $50\text{ mm} \times 25\text{ mm}$ Patch, $P=15\text{ mm}$, $C=40\text{ mm}$.

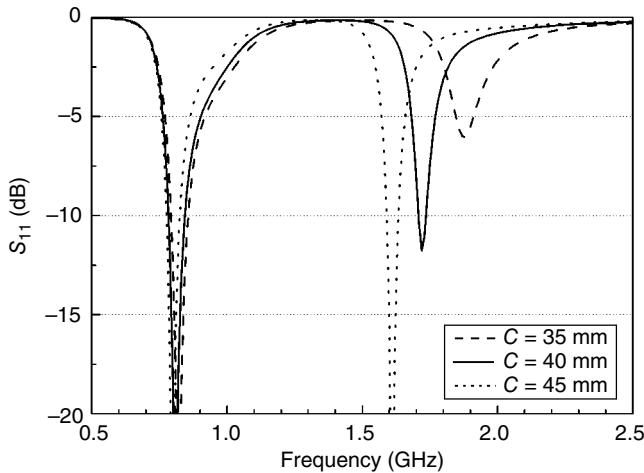


Figure 4.17 Impact of dimension C . $50\text{ mm} \times 25\text{ mm}$ Patch, $P = 15\text{ mm}$, $D = 10\text{ mm}$.

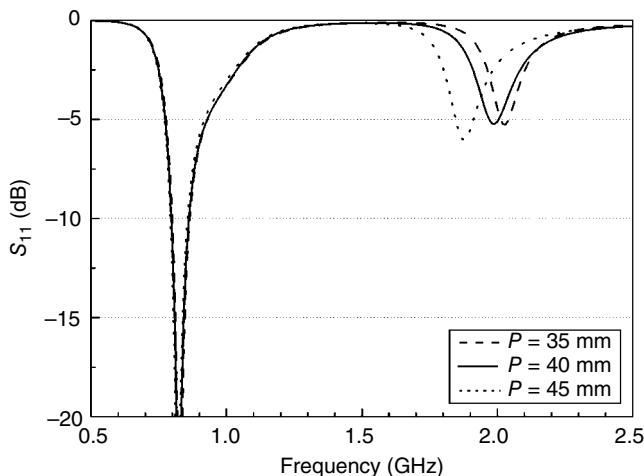


Figure 4.18 Impact of dimension P . $50\text{ mm} \times 25\text{ mm}$ Patch, $C = 35\text{ mm}$, $D = 10\text{ mm}$.

dimension P is fixed, the modification of C is realized by only increasing or decreasing the length of the horizontal portion of the slit.

Shown in Figure 4.18 are the results when the dimension P is changed while keeping the slit length C fixed. Although not significantly, the dimension P also has some impact on the higher band. It can be used in some circumstance as a means to provide the “extra mile” for tuning.

Shown in Figure 4.19 are the radiation patterns of a PIFA. At the low band, which is 0.78 GHz in this case, the radiation pattern is similar to a dipole’s. The radiation is mostly generated by the vertical current along the edge of the ground plane. The vertical polarized field E_θ is the dominant field component. At the higher band, which is 1.75 GHz, the amplitudes of E_θ and E_ϕ are in the same order of magnitude. Shown in Figure 4.19e is the radiation

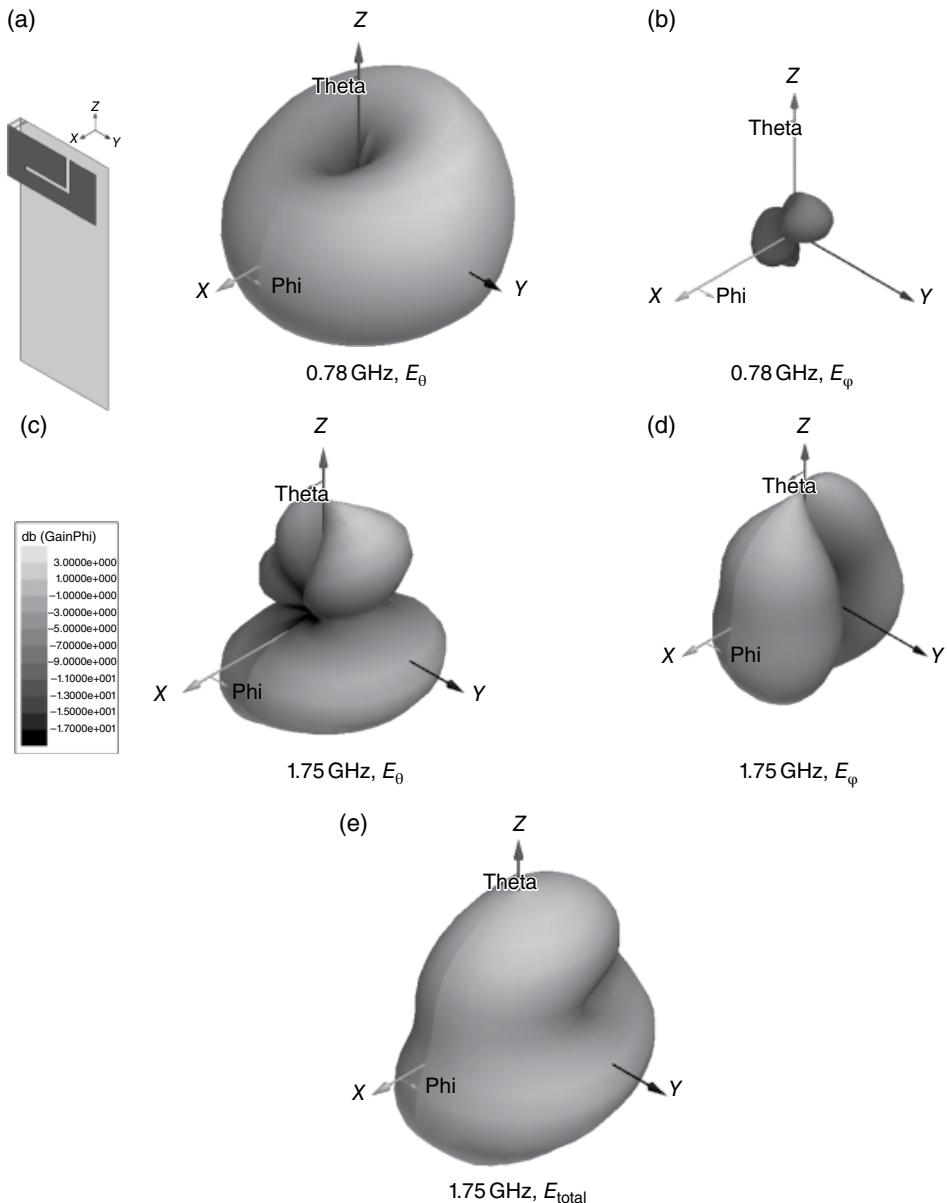


Figure 4.19 Radiation patterns of PIFA, $P=15$ mm, $D=15$ mm, $C=40$ mm.

pattern of the total E field at 1.75 GHz. It confirms again that it is the ground plane, instead of the antenna element itself, which definitively decides the radiation patterns.

When designing any kind of antenna, antenna engineers may find themselves in a situation where one band has more than enough bandwidth but the other band cannot meet the specification. Knowing how to exchange bandwidth between bands is an essential technique.

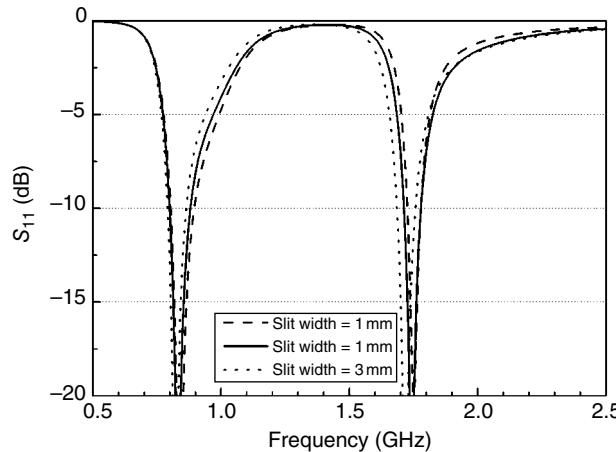


Figure 4.20 Bandwidth trade-off between low and high bands.

For a PIFA, cutting the slit shape differently can have some effect on the bandwidth ratio between bands. The other way to change the ratio is by adjusting the slit's width. Shown in Figure 4.20 are the simulated results of three PIFAs with the same dimensions but different slit widths. With the decrease in the slit's width, the bandwidth increases at the lower band and decreases at the higher band.

If an antenna still cannot meet the specification after the performance of both bands have been balanced, the only way out is to attempt to increase the patch's height, which is the dimension H shown in Figure 4.10a. This definitely is an uphill battle.

So far, we have shown how to adjust both bands simultaneously or modify the high band independently. By combining both methods, we can tune the resonances freely in a quite wide range. Now let's discuss some matching techniques. The easiest way to match an antenna is to design a matching circuit. However, a matching circuit always has some inherent loss. For a patch antenna, some of the dimensions can be modified, so we should always try to achieve a good match by tweaking the patch itself first. By comparing Figures 4.16 and 4.17, we can see that there are different ways to cut the slit to achieve two required resonant frequencies, and their matching conditions are also different.

Besides the patch itself, another useful tuning mechanism is the feeding and grounding structure. Shown in Figure 4.21 is the equivalent circuit of feeding and grounding strips. The grounding strip is equivalent to a shunt inductor, so it has more impact on the lower band. The feeding strip is equivalent to a series inductor, and it can be used to tune the high band. To increase the inductance of grounding strip, the following methods can be used:

- Increase the length of grounding strip by bending the strip.
- Decrease the width of grounding strip, t_2 .
- Increase the distance between feeding strip and grounding strip, S .

Similarly, the inductance of feeding strip can also be increased by increasing the length or decreasing the width of the strip. Shown in Figure 4.22 are two other ways of matching. Figure 4.22a shows how to use the slit to increase the shunt inductance. Figure 4.22b shows

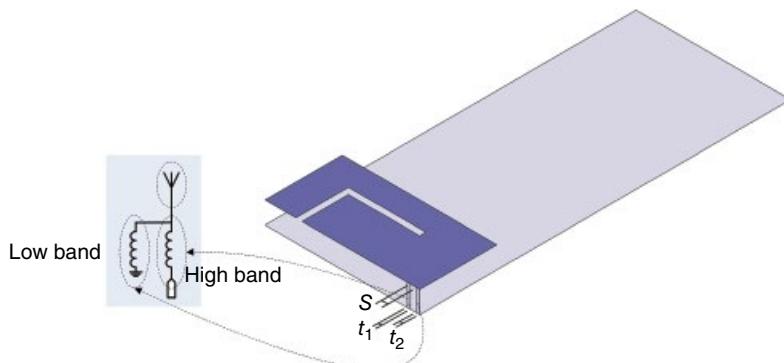


Figure 4.21 Equivalent circuit of feeding and grounding strips.

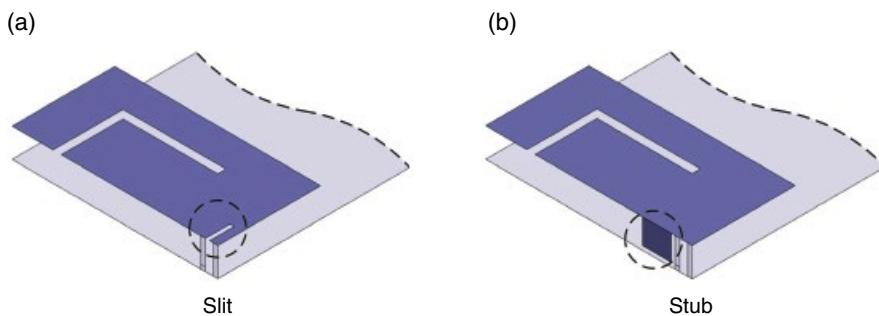


Figure 4.22 Other ways of matching.

how to add some distributed shunt capacitance. Another effect of the stub shown in Figure 4.22b is to increase the total current length, thus decreasing the overall resonant frequency. If you have a brainstorming session on this topic, lots of matching structures can be invented. Where should we stop and hand over the remaining tasks to lumped matching element? There is no right or wrong answer to this question. Normally, it is a purely personal choice. Some engineers prefer a self-matched design without requiring any lumped components. Many of those engineers treat antenna design as an art. Some engineers like to reuse previous designs and want to push dimension changes to the minimum. For them, a matching circuit is the favorite choice.

Putting the personal preference aside, there are some objective criteria that can be used to evaluate any antenna design. How sensitive is a design to manufacturing tolerance? What kind of manufacturing process can it use? Is a design easy to be tuned when frequency drifts? Does frequency tuning require expensive and time-consuming tooling modifications? What is the efficiency of an antenna?

It is not difficult to imagine that the weight of each criterion varies from project to project. For a low-end phone, cost might be the first consideration. For a high-end phone, it is most likely that all physical constraints have been pushed to the limit; the only goal for an antenna engineer is to meet the specification at any cost. The duration of a project is also an important

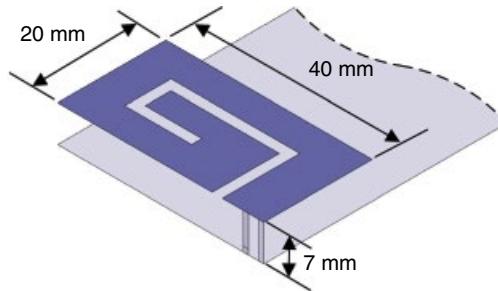


Figure 4.23 A small PIFA.

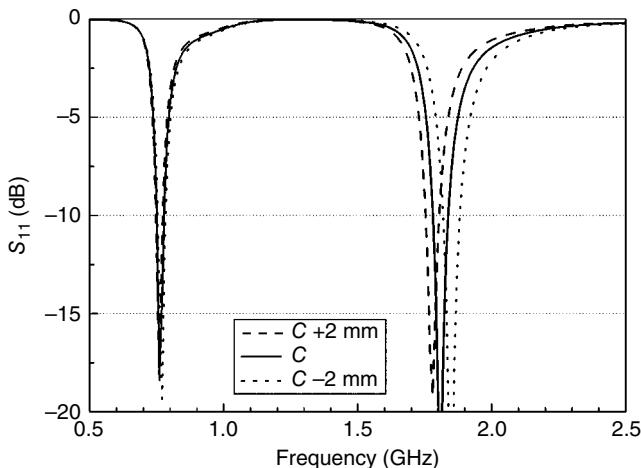


Figure 4.24 A small PIFA: slot length variations.

factor when selecting different approaches. In fact, most issues mentioned here are not academic at all, they are purely practical matters. Section 4.2.5 is dedicated to this topic.

Referring to Figure 4.15, if the patch size gets smaller, to maintain the resonant frequency at the lower band, the opening of the slit needs to be moved closer to the feeding point, and the slit must be routed through the other direction. Shown in Figure 4.23 is an example of such a small PIFA. The rules for tuning the small PIFA shown in Figure 4.23 are pretty much the same as those shown in Figure 4.15. The critical dimensions are the slit's length and the opening position of a slit. The length of the slit has more impact on the higher band, while the slit's open position affects both bands.

Shown in Figure 4.24 are the simulated results of a small PIFA when the slot length varies. By decreasing or increasing the slot length by 2 mm, the resonant frequency at a higher band is shifted up or down, respectively, by around 35 MHz. Meanwhile, the lower band response is relatively immune to those changes.

So far, we have discussed how to design and tune a PIFA. The topic of current modes on a patch has deliberately been avoided in order to minimize confusion in the first stage of learning. Although the current distribution on a patch can provide a very useful insight on how

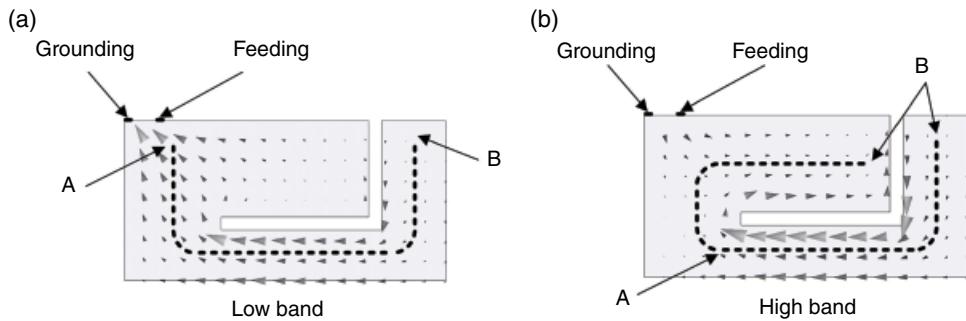


Figure 4.25 Current distributions of a “normal” PIFA.

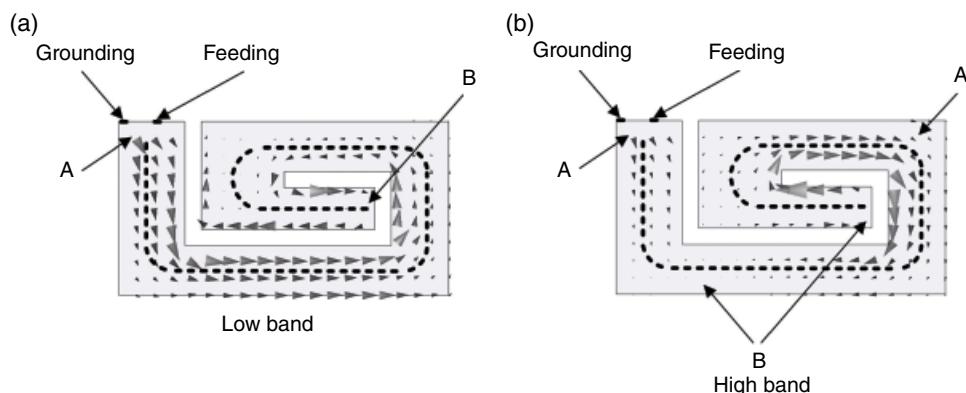


Figure 4.26 Current distributions of a “small” PIFA.

different patches work, it frequently makes a new engineer confused about how to tune a PIFA. Of course, if we want to make some innovation in the field of antennas, current modes is a topic we must discuss.

The current distributions of a “normal” PIFA at both low and high bands are illustrated in Figure 4.25. The dashed lines in Figure 4.25 are the significant paths. At the low band, the current distribution along the critical path is more like a quarter-wavelength monopole. The current reaches its maximum at point A, which is around the feeding point. At point B, the current decays to zero.

In the high band, the current is more like a dipole along the critical path. It has the maximum in the middle, which is marked by point A. At both ends, which are marked as point B, the current decays to zero. As we know, a monopole antenna is not able to resonate when the antenna length is half of an effective wavelength, because its port impedance is too high to be matched. On the other hand, unlike a monopole, which is fed from one end of a radiator, a high-band patch mode is accessed from the middle, which explains why it has a decent impedance matching.

Although with regard to antenna tuning, a “normal” PIFA and a “small” PIFA are quite similar, their current distribution is totally different. Shown in Figure 4.26 are current distributions of a “small” PIFA. The critical current paths of a “small” PIFA at both bands are

the same. At the lower band, the current along the critical mode is similar to a monopole, which has the peak current at point A and zero at point B. At the higher band, the total critical path corresponds to the third-order mode, which is three-quarter wavelength. There are two peak current spots and two zero current spots. One might wonder whether the resonant frequency of the third-order mode is three times its basic mode. To explain that, we need to recall the technique we discussed in Section 3.1.2.2. There, by adjusting the pitch of a helix, the resonant frequency of the third-order mode can be tuned in a wide range. Similarly, the third-order mode of a “small” PIFA can also be tuned to where we want.

Although mathematical formulas are given in many papers or books to calculate resonant frequencies of different bands, sometimes we do not need them. There are so many unknowns when designing an antenna for a real device; no formula can give an accurate prediction. The best way is to find out what the effect of each parameter is, then tune the antenna accordingly.

4.2.3 Multiband PIFA with Separate Branches

When the antenna area is big enough, a better design approach is to use two independent branches to cover low and high bands separately. The antenna shown in Figure 4.27a is such an example. The antenna's size is 60 mm × 25 mm. The ground's size is 60 mm × 120 mm. The antenna is divided into two branches by the grounding strip. In Figure 4.27a, the feeding strip is located on the side of longer branch. This arrangement is more favorable to the performance of the lower band.

As the two branches are far apart, the responses at the low and high band are decoupled. The tuning process of this antenna is quite easy. When designing a large antenna, the first step is to adjust the length of both branches to get the resonances at the expected bands. Then by tweaking the distance between the feeding strip and the grounding strip, the matching at the lower band can be tuned. The final step is to design a matching circuit as shown in Figure 4.27b to obtain a good matching at the higher band.

Shown in Figure 4.28 is the simulated reflection coefficient of the large PIFA. By comparing this result with results shown in Figures 4.16 and 4.24, it is clear that the antenna has a wider bandwidth at both bands. As a rule of thumb, a larger antenna area is always better, which gives more design options, and it is easier to achieve better performance.

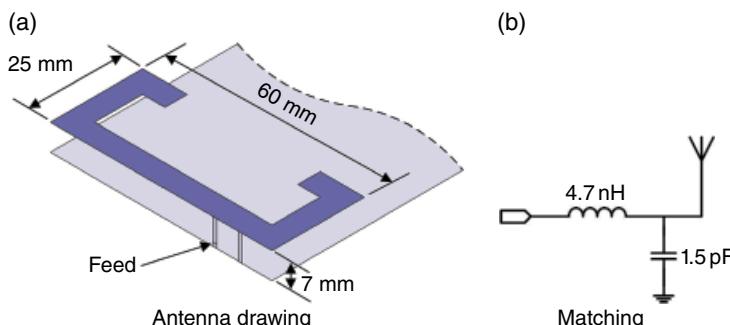


Figure 4.27 A large PIFA.

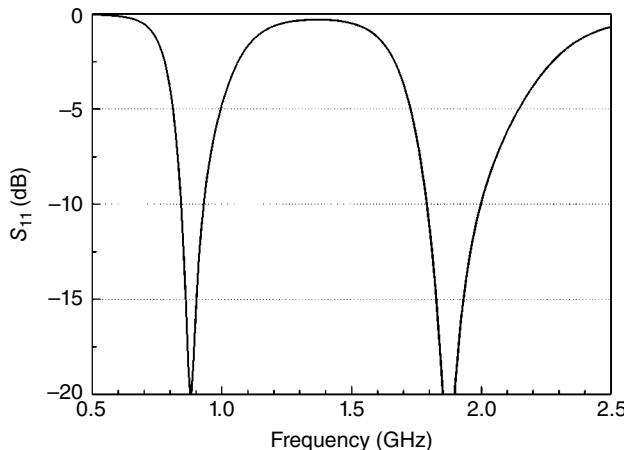


Figure 4.28 Reflection coefficient of the large PIFA.

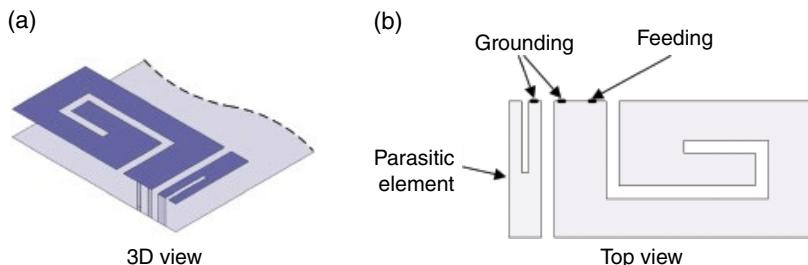


Figure 4.29 PIFA with parasitic element.

4.2.4 Multiband PIFA with Parasitic Element

There are many ways to expand PIFA's bandwidth, which include using one slit with branches or multiple slits. PIFA variants using those techniques are too many to be enumerated in the book. Interested readers can find some enlightening examples in Professor Wong's book [1]. The technique we are going to discuss here uses the parasitic element.

Shown in Figure 4.29 is a PIFA with a parasitic element. The parasitic element is electrically connected to the ground through a metal strip. There is no direct connection between the main radiator and the parasitic radiator. However, they are electromagnetically coupled.

Shown in Figure 4.30 is the simulation result of a PIFA with a parasitic element. Similar to a normal PIFA, the main radiator generates two resonances, one at the lower band and one at the higher band. The parasitic element only resonates at the higher band. The parasitic elements can frequently be found in penta-band 3G phones, which normally require frequency coverage over 1710–2170 MHz at the higher band. A quad-band 2G phone only needs to cover 1710–1990 MHz.

In practice, the parasitic element is normally used to cover the highest band. As higher frequency means smaller parasitic radiator, this makes antenna design easier. The other reason for using parasitic elements at the highest band is an antenna's overall performance. With the

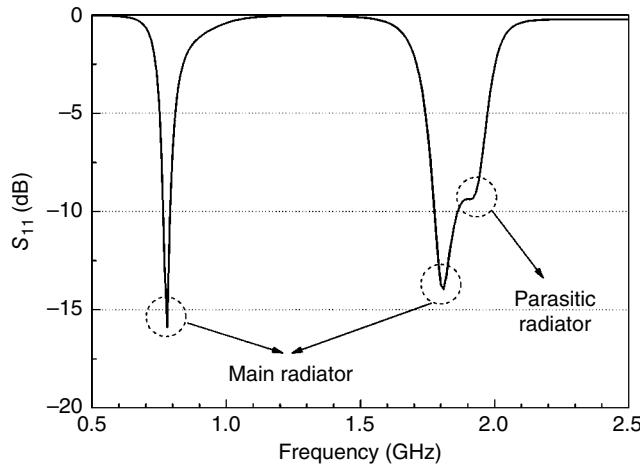


Figure 4.30 Reflection coefficient of a PIFA with a parasitic element.

existence of parasitic element, the efficiency of the main radiator might degrade from one-tenth of a dB to a few dB. By assigning the resonance of a parasitic element to the highest frequency, the adverse effect on the lower band can be mitigated.

4.2.5 Manufacturing PIFA Antenna

Regarding designing techniques for PIFAs, this is as far as the book will go. The aim of the book is to aid in jump-starting an antenna project. The book mostly focuses on the core principles of antenna designing. However, there are so many phones on the market, that if you do some reverse engineering, you can discover various elegant tricks. The book does not touch upon those advanced techniques. On the academic side, Professor Kin-Lu Wong is one of the most innovative scholars in the mobile antenna world. Professor Wong has written two books [1, 6] that cover advanced techniques.

In this section, we are going to discuss the various processes of manufacturing PIFAs. The cheapest technology used to make an internal antenna is metal stamping. In production, a metal stamping antenna can be tuned quite quickly if the parameter needing to be adjusted is known and already included in the tooling design. The minimum number of parts necessary for a metal stamping antenna is two: one plastic carrier and one metal radiator. Multiple heat stakes are used to assemble the two parts together. As heat stakes are designed as a part of the plastic carrier, this feature is actually free. An assembly fixture is required to melt all the heat stakes and deform them into a mushroom shape. Shown in Figure 4.31 are some production metal stamping antennas. Most of the time, the volume available for antenna designing is irregular. To optimize the performance, an antenna must take full advantage of an irregular space. However, it is not difficult to see that those antennas still follow the basic design principles. Spring fingers are frequently used in metal stamping antennas as the contact feature. As spring fingers are formed from the same metal sheet using which antennas are made of, they are also free. To meet the humidity and other environment requirements, spring fingers

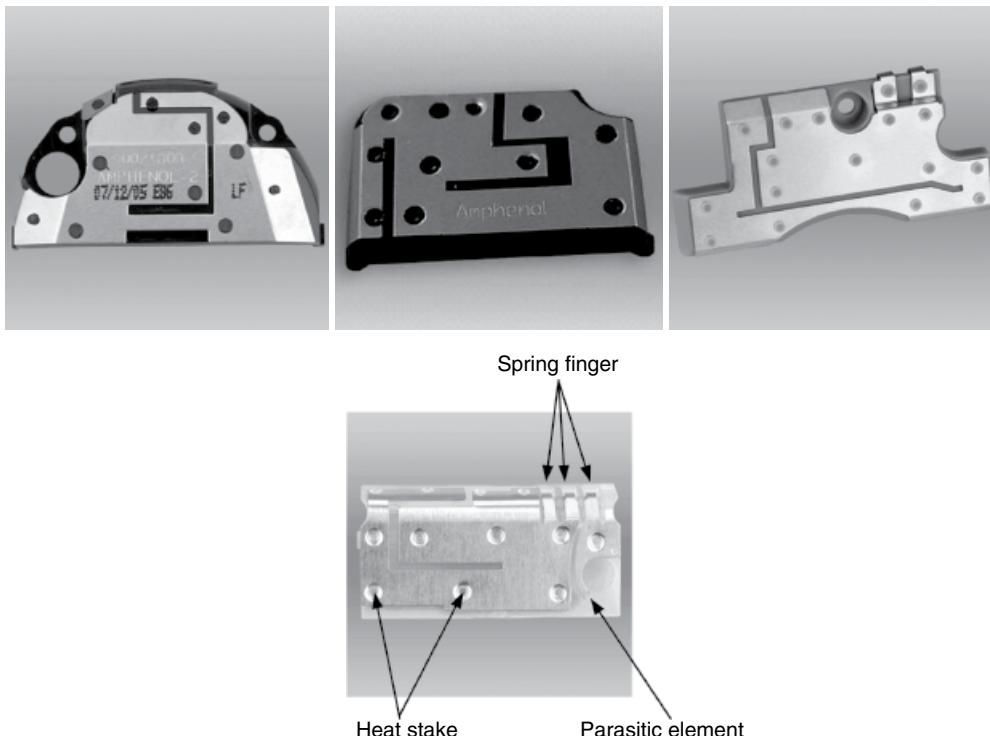


Figure 4.31 Metal stamping technology. (Source: Reproduced with permission of Shanghai Amphenol Airwave.)

must be gold plated. To cut down on the manufacturing costs, a selective gold plating process can be used to minimize the plated area.

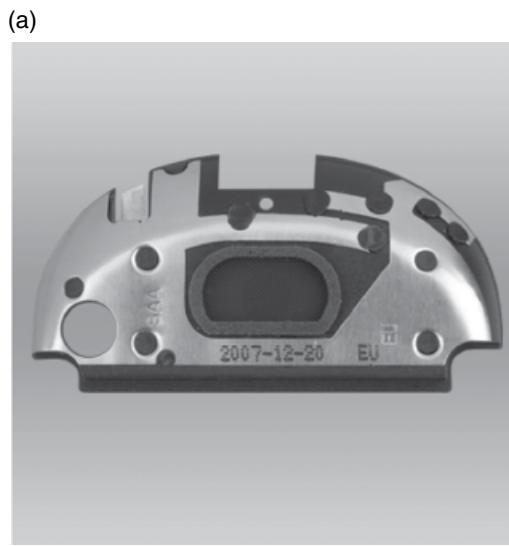
At the dawn of internal antennas, dedicated volumes were reserved for antenna designing. With the continuous shrinking of mobile devices, eventually antennas have had to coexist with other components, such as the microphone, the speaker, the camera, and so on. Shown in Figure 4.32 are some samples of integrated antennas. If one is putting other components underneath an antenna, they must be RF isolated. Take the speaker as an example. There are two lead lines out of a speaker. Both of them must be isolated by inserting series inductors into signal paths.

Normal metal stamping process can only bend metal sheets, so all the antennas shown in the figure are not truly 3D. They are either composed of multiple flat surfaces, or combinations of flat and cylindrical surfaces. The deep draw process can manufacture a true 3D metal radiator. Shown in Figure 4.33 is an antenna made by the deep draw process. This antenna is also an integrated antenna. The exploded drawing is shown in Figure 4.33b. Only parts 1 and 2, which are the metal radiator and the plastic carrier respectively, are destined for the antenna functionality. All the other parts are intended for the functionality of the speaker.

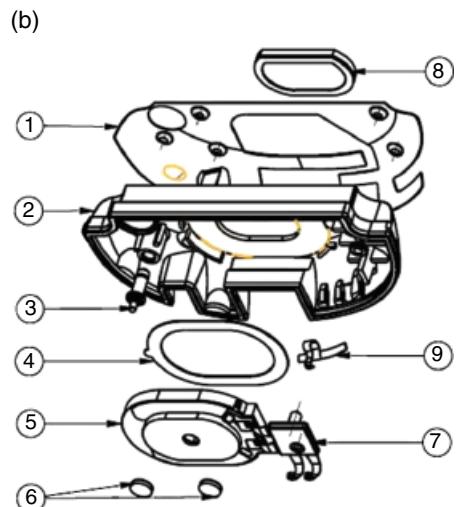
Compared with metal stamping technologies, the flex circuit technology has better consistency, but it is little more expensive. In many state-of-the-art designs, the antenna volume is quite tight, which means the performance margin is relatively slim. Because flex antennas



Figure 4.32 Integrated antenna with other components. (*Source:* Reproduced with permission of Shanghai Amphenol Airwave.)



Photo



Exploded drawing

Figure 4.33 Three-dimensional metal stamping integrated antenna. (*Source:* Reproduced with permission of Shanghai Amphenol Airwave.)

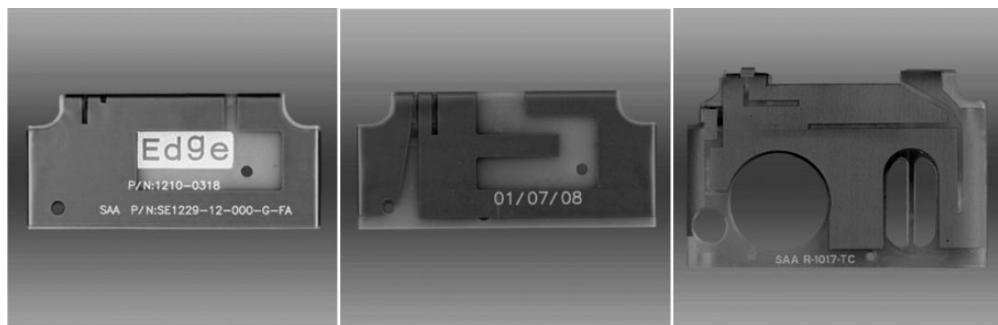


Figure 4.34 Flex technology. (Source: Reproduced with permission of Shanghai Amphenol Airwave.)



Figure 4.35 Antenna made by DS-MID technology. (Source: Reproduced with permission of Shanghai Amphenol Airwave.)

have better consistency, they also have better yield rate. The flex itself cannot provide connecting features, some extra parts, such as metal spring fingers or pogo pins, are needed.

The flex technology is not a true 3D technology (see Figure 4.34). A flex can only wrap around a combination of 2D surfaces. The most advanced and also the most expensive technologies in antenna manufacturing are double-shot molded interconnect device (DS-MID) and/or laser direct structuring (LDS). They have the best consistency, because the antennas are part of the plastic structure instead of separate parts. Both of them are based on a technique called “selective metallization.” The DS-MID process begins with the application of a shot of plateable thermoplastic resin into an injection mold cavity. Next, the cavity is changed and a second shot of nonplateable thermoplastic resin is molded around the first shot to create a circuit pattern from the plateable material. These parts are then plated with a layer of copper. The DS-MID takes the longest lead time, because any modification to the antenna pattern requires tooling changes.

The LDS is now the standard process for high-end phones. The thermoplastic resin used in LDS process is nonplateable after the molding process and can be transformed to plateable by using a laser beam to activate it. The LDS process literally draws the antenna pattern on to the plastic. The pattern can be adjusted quite easily by uploading a new pattern file to the laser. Similar to the DS-MID, a plating process is required to deposit copper onto the part’s surface. Shown in Figure 4.35 is an antenna made by the DS-MID process.

When choosing from different antenna technologies, an engineer needs to take into consideration cost, consistency, and lead time. It is always a good idea to consult experienced engineers and antenna manufacturers.

4.3 Folded Monopole Antenna

A PIFA implies that there is a ground plane underneath the antenna element. When a phone with a PIFA is used in a talking position, the ground is sandwiched between the head and the antenna. The ground functions as a shield, which can direct some of the near-field energy away from the head, thus decreasing the radiation to the user's brain. However, from an antenna design point of view, the antenna bandwidth can be significantly expanded if the ground can be removed. When the ground is removed, the antenna is no longer a PIFA, it is called a "folded monopole antenna."

For all phones with PIFAs, in my experience, the antennas are on the top of the phone. The "ancient" wisdom is that by putting an antenna on the top of a phone, the hand effect on the antenna can be mitigated. When the phones became thinner and smaller, the volume left for antenna also become too small to fulfill the bandwidth requirement. An obvious way to ease the pressure on antennas is by using monopole antennas. However, a monopole antenna cannot be placed on the top of a phone, because as there is no ground acting as the shield between the user's head and the antenna, the radiation to user's brain would be too high. A logical way of working around this is to put the antenna at the bottom of a phone. As the bottom of a phone never comes into contact with people's head, this solution allows the phone to pass the radiation safety regulations. In my experience, I have proposed this kind of solution in two projects and been rejected twice. The reason for rejection is that if we put the antenna at the bottom, the user's hand will always cover the antenna, thus degrading the antenna's efficiency. I am sure that due to the pressure on antenna designing, many antenna engineers were starting to think about bottom-installed monopole antennas at that time.

Everything changed in 2004, when the Motorola Razr was revealed, as shown in Figure 4.36. The Motorola Razr is a *de facto* grand slam. More than 20 million Razrs and its derivates have been sold worldwide. I have to admit that I had never imagined that by putting an antenna at the bottom, industry engineers could have so much design freedom. The success of the Razr started the new trend for the ultra-slim phone. Nokia, Samsung, and so on, all released their own models which looked just like Razr's cousins, and all of them have a bottom-installed internal monopole antenna.

Figure 4.37 depicts an internal monopole antenna. The antenna protrudes over the edge of the ground plane. The antenna is composed of two pieces of metal wires. The thicker wire makes the connection to the feeding spring finger on the PCB board. The thinner wire is attached to the thicker wire by a welding process. The thinner wire is divided into two branches by the thicker wire. The shorter branch resonates at the higher band and the longer branch resonates at the lower band. If comparing an internal monopole with the dual-branch multiband stubby antenna discussed in Section 3.1.2.1, we can see that although they have totally different form factors, they are pretty much the same from the electrical point of view.

Shown in Figure 4.38 is a simulation model of a folded monopole antenna. The area of ground is $100\text{ mm} \times 50\text{ mm}$. The dimension D , which is the length of the thick wire and also is



Figure 4.36 Motorola Razr. (Source: Reproduced with permission of Motorola.)

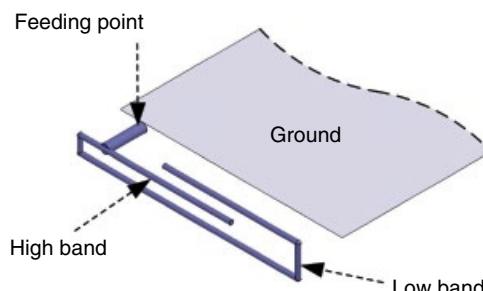


Figure 4.37 Internal folded monopole antenna.

the distance between the antenna element and the edge of the ground, is the decisive factor in an antenna's bandwidth.

Shown in Figure 4.39 are simulation results of folded monopoles with different dimensions D . When the D changes, the resonant frequencies at different bands will drift. The branches' lengths need to be tweaked accordingly to compensate for the frequency drift. In Figure 4.39, when the D is 5 mm, the antenna element is too close to the ground, thus inducing a strong current on the horizontal edge of the ground. This induced current is always in the opposite direction to the current on the antenna element. The radiations of both currents cancel each

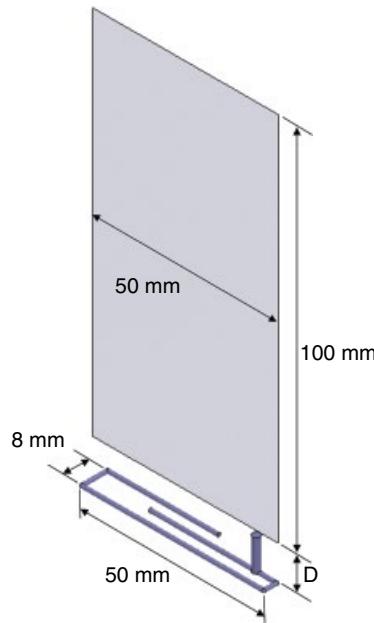


Figure 4.38 Dimensions of a sample internal monopole antenna.

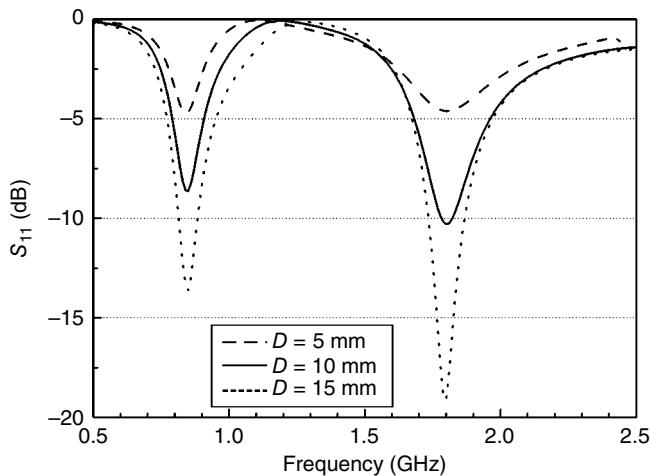


Figure 4.39 Internal folded monopole with various D .

other out, so the antenna cannot radiate well. When D increases, the bandwidths of both low and high bands increase. When D increases from 10 to 15 mm, the bandwidth increases a little at the higher band but quite significantly at the lower band. This is because the impact of the gap is measured by how much wavelength it is instead of the absolute dimension. At the lower

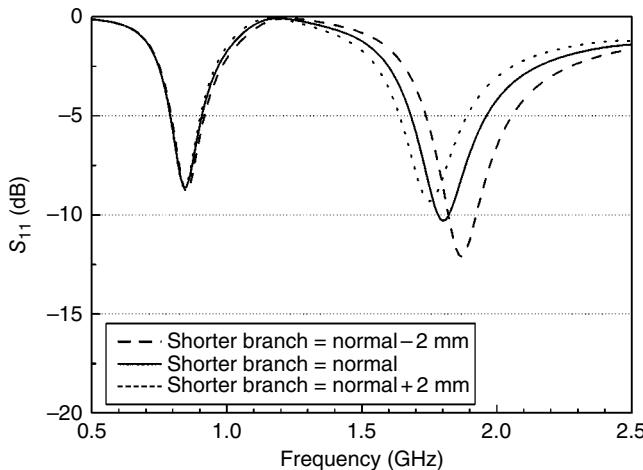


Figure 4.40 Internal folded monopole, variation in the shorter branch, $D=10$ mm.

band, the wavelength is longer, so it is always more difficult to achieve good performance at the lower band in a constrained space.

Shown in Figure 4.40 is the impact of the shorter branch. The length of the shorter branch mostly affects the higher band. By increasing or decreasing the shorter branch, the resonant frequency of the higher band can be tuned lower or higher. For the example shown in Figure 4.40, when adjusting the branch length by 2 mm, the frequency changes roughly 60 MHz.

The length of the longer branch affects the resonant frequencies of both low and high band. As shown in Figure 4.41, when decreasing the longer branch, both resonant frequencies shift higher.

Before tuning any antenna, it is important to have a clear understanding of what the impact of each dimension is. With that knowledge, it is easy to decide the best tuning procedure and whether an iterative tuning is necessary. For the antenna shown in Figure 4.38, if we tune the higher band first, then an iterative procedure has to be used. The appropriate tuning sequence should be tuning the lower band first. We can start with an antenna whose both branches are a little bit longer, then gradually trim off the longer branch. After correctly tuning the lower band, double check the higher band's resonance to make sure it is lower than what we need. The final step is to gradually trim off the shorter branch until the higher band is well tuned.

The antenna shown in Figure 4.38 is a simplified model. In reality, when there is a hands-free speaker, microphone, shielding box, battery, and other components around the antenna, you may find that both branches have an effect on both bands simultaneously, and an iterative tuning process has to be used in such circumstances.

Shown in Figure 4.42 are some production internal monopole antennas. The manufacturing techniques used are a little bit different. However, they are pretty much the same from the electrical point of view.

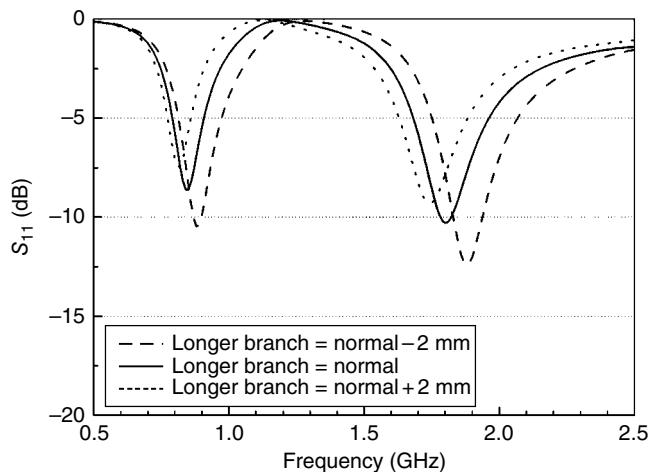


Figure 4.41 Internal folded monopole, variation in the longer branch, $D=10\text{ mm}$.



Figure 4.42 Production internal monopole antennas. (Source: Reproduced with permission of Shanghai Amphenol Airwave.)

4.4 Loop Antenna

Loop antennas are one of the oldest antennas. In 1886, H. Hertz demonstrated the first wireless (spark) communication system which verified the existence of the electromagnetic wave. In the experiment setup, a dipole antenna was used as the transmitter and a loop antenna was

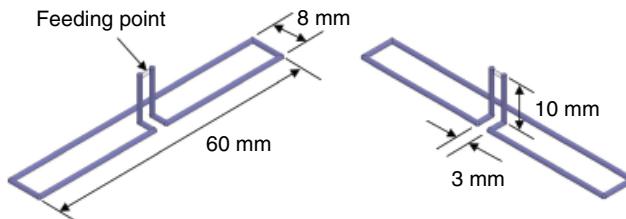


Figure 4.43 Oblique views of a loop antenna.

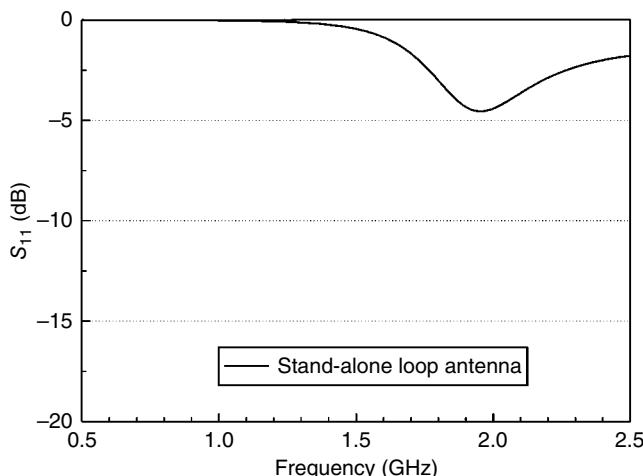


Figure 4.44 Reflection coefficient of a stand-alone loop antenna.

used as the receiver. Loop antennas are classical antennas and are explained in most antenna textbooks [7–11]. There are electrical small loop antennas and electrical large loop antennas, and a detailed theoretical analysis of both antennas can be found in textbooks [7]. Only electrical large loop antennas can be used in the cellular communication. Similar to a folded monopole antenna, a loop antenna also needs to be installed on the bottom of a device to avoid high radiation to a human brain.

Shown in Figure 4.43 is a loop antenna. The shape of this loop antenna is a little bit abnormal, because we want to use it in the following studies. The shape of the loop antenna slightly affects the resonant frequency; other than that, it is a good example without losing generality. The loop of the antenna shown in Figure 4.43 is quite narrow, so it is sometimes also categorized as a folded dipole antenna. We should obtain the same result no matter which analytical method is used.

The loop antenna shown in Figure 4.43 has a dimension of $60\text{ mm} \times 8\text{ mm}$, which is the same as the folded monopole antenna discussed Section 4.3. The antenna is fed by a parallel transmission line, which is perpendicular to the antenna plane. The feeding point is at the end of the transmission line.

Shown in Figure 4.44 is the simulated reflection coefficient of the stand-alone loop antenna. The first resonance appears around 2 GHz, which is about two times higher than the folded

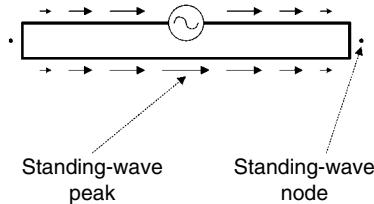


Figure 4.45 Current distribution of a one-wavelength stand-alone loop antenna.

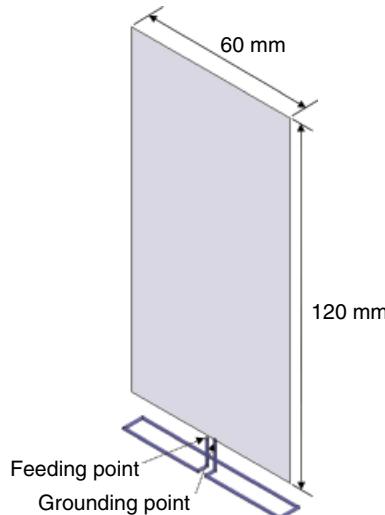


Figure 4.46 Bottom-installed loop antenna.

monopole antenna whose longest branch is even shorter than the loop antenna. The reason is that the first resonance of a stand-alone loop antenna appears when the loop length equals one wavelength. A folded monopole can resonate when it is a quarter of a wavelength.

Shown in Figure 4.45 is the current distribution of a one-wavelength loop antenna. There are two half-wavelength standing-wave currents on the loop. The standing-wave nodes are on both sides of the loop. On a straight line, the currents at both side of a standing-wave node are in opposite directions. A loop antenna is bent at the node, so two standing-wave currents are actually flowing in the same direction and can effectively radiate. A loop antenna must be excited differentially, which means at the feeding point, the currents on both sides are always in the opposite direction. As shown in Figure 4.45, the currents on the left and right sides flow into and out the source, respectively.

The reflection coefficient of a stand-alone loop antenna is not very good. The impedance of a stand-alone loop antenna is around 300Ω , about four times a half-wavelength dipole antenna. Detailed analysis is omitted here and can be found in textbooks [7]. Beside the base frequency f_0 , a stand-alone loop antenna also resonates at integer multiples of f_0 . If the first resonance is at 2 GHz, the second and third resonances should appear around 4 and 6 GHz.

Shown in Figure 4.46 is a bottom-installed loop antenna. The dimensions of the antenna were illustrated in Figure 4.43. The ground plane has a dimension of $120\text{ mm} \times 60\text{ mm}$. Unlike

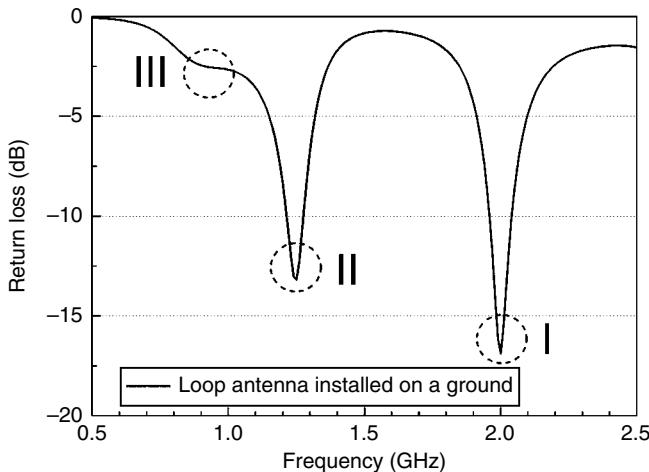


Figure 4.47 Reflection coefficient of a bottom-installed loop antenna.

a stand-alone loop antenna, the loop antenna shown in Figure 4.46 is unbalanced fed. One end of the loop antenna is connected to the ground plane and the other end serves as the feeding point.

Shown in Figure 4.47 is the simulated reflection coefficient of a bottom-installed loop antenna. By comparing Figures 4.44 and 4.47, significant difference can be observed. The resonance (I) shown in Figure 4.47 can be explained by the one-wavelength loop antenna mode. Resonances (II) and (III) do not even exist on a stand-alone loop antenna, so they must have something to do with the ground plane.

Shown in Figure 4.48 is a bottom-installed monopole antenna. The monopole antenna is made by splitting the loop antenna and removing half of it. Apart from the antenna, everything else in the fixture is identical to the one shown in Figure 4.46. As we know, a monopole antenna is only half of the radiator and the ground functions as the other half.

Shown in Figure 4.49 is the simulated reflection coefficient of a bottom-installed monopole antenna. Compared with results shown in Figure 4.47, the resonance (I) has disappeared, but resonances (II) and (III) are still preserved. For a monopole antenna, the resonance (II) is introduced by the monopole itself and resonance (III) is the contribution of the ground plane. If you want to verify the root cause of those two resonances, you can vary the length of either the monopole or the ground plane separately and the resonant frequency will shift accordingly.

Based on Figures 4.47 and 4.49, it is quite safe to assume that a loop antenna functions as a monopole antenna when it is half a wavelength long and installed on a ground plane. Shown in Figure 4.50 are the current distributions of a quarter-wavelength monopole antenna and a half-wavelength loop antenna when installed on a ground plane. For both antennas, the current distributions on the ground plane are pretty much the same. Some people might categorize the current on the half-wavelength loop as a common mode. It is fine as long as we remember such a common mode cannot exist without a ground plane. For any signal source, the sum of all currents flowing in and out is always zero. For a half-wavelength dipole antenna or a one-wavelength loop antenna, the currents on both sides are always in opposite directions, so they

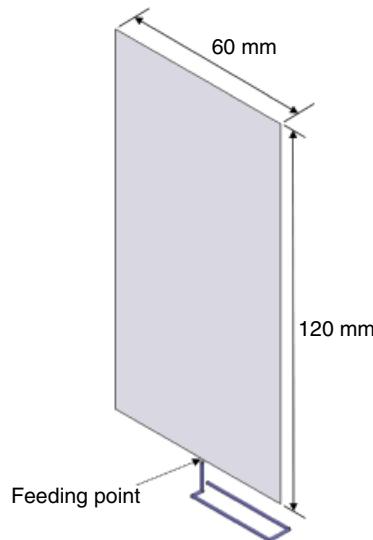


Figure 4.48 Bottom-installed monopole antenna.

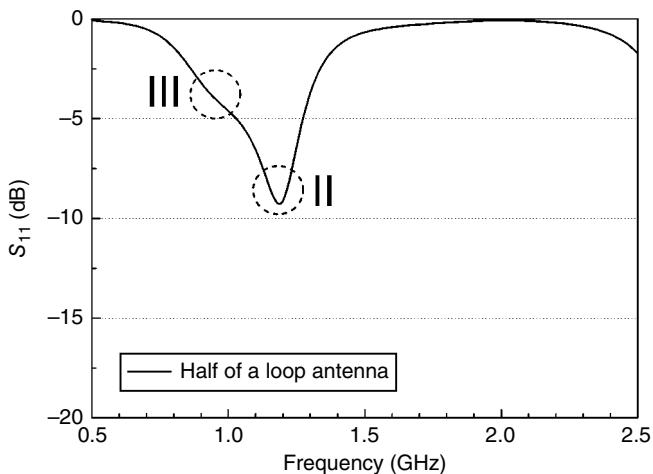


Figure 4.49 Reflection coefficient of a bottom-installed monopole antenna.

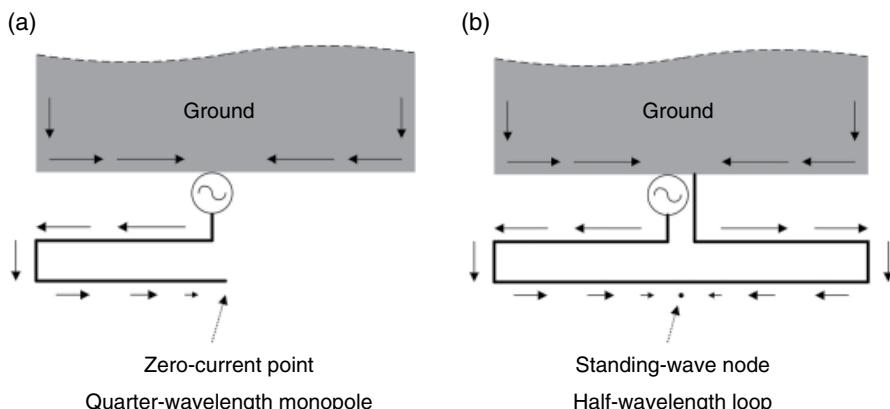


Figure 4.50 Current distributions of a monopole and a loop.

satisfy the condition of zero sum current. For a quarter-wavelength monopole antenna or a half-wavelength loop antenna, because the sum of the currents on the antenna element is not zero, there must be a ground which provides a path for the opposite current. With the assistance of a ground, a loop antenna can also resonate when it is odd multiples of half a wavelength.

By utilizing the half-wavelength, one-wavelength and/or one and a half wavelength modes, multiband antennas have been designed and reported [12–15]. So far, in my experience, the best one is reported in [13], which achieved a VSWR better than 3:1 at 822–964 MHz and 1570–2260 MHz. The antenna volume is 60 mm × 10 mm × 4 mm.

Some articles or companies might claim that a loop antenna has a localized current distribution, thus has much less radiation to human brains and is immune to the loading effect when the ground plane is held by hand. This claim is only partially true. As we have demonstrated, when the antenna is working in the one-wavelength mode, the antenna is differentially fed, thus has limited current on the ground. In this circumstance, the claim holds well. However, when the antenna is working at the half-wavelength mode, which normally corresponds to the 900 MHz band, the ground plane becomes the main radiator and a loop antenna is really no better than its monopole counterpart from the radiation safety point of view. Similarly, for one and a half wavelength mode, the claim is also invalid.

4.5 Ceramic Antenna

Just as its name implies, any ceramic antenna is made of ceramic. There are mainly three kinds of antennas which are made of ceramic. The first is often referred to as the dielectric resonator antenna [16–21]. A dielectric resonator antenna normally has a cylinder or a cuboid form factor. The ceramic block becomes a good radiator when resonant modes, such as TM_{10δ}, and so on, can be effectively excited. The air–dielectric boundary of an antenna can sustain the resonant standing wave in the ceramic, so no metal is needed. However, at the cellular communication frequency, a dielectric antenna is too chunky to be used in hand-held devices.

The second kind is the ceramic patch antennas [22–25]. Those antennas can generate circular polarized radiation patterns and are normally used in satellite applications, such as global positioning system (GPS) and XM satellite radio. Although ceramic patch antennas are widely used in hand-held GPS navigators, they are seldom used in cellular phones. Ceramic patch antennas do give better performance than their counterparts. However, due to the size consideration, most cellular phones still use other types of antennas, such as IFAs, to cover the GPS frequency band.

The third kind of ceramic antenna is the one widely adopted in the mobile industry. From an electromagnetic technology point of view, these ceramic antennas are not really utilizing any brand new technology. All of them can be labeled as monopole, IFA, and so on. All antennas introduced in this section belong to the third category.

The main reason for the introduction of ceramics into antenna design is size considerations. Most material used in ceramic antenna design has high relative permittivity, which ranges from a few tens to more than 100. With the higher relative permittivity, an antenna's size gets smaller and its bandwidth also shrinks accordingly. Thus the tolerance of ceramic's relative permittivity must be well controlled. Fortunately, good tolerance is one of the advantages ceramic material provides. The ceramic used in antenna design also has very little loss, which is a critical merit to guarantee good efficiency when an antenna is exceedingly small. Unlike

the plastic support used in other types of antenna, ceramic can sustain the temperature in the soldering oven. Most ceramic antennas are designed using surface-mount technology (SMT) and are compatible with the standard SMT soldering process. Unlike other antennas, ceramic antennas do not require a dedicated manual assembly step; they can be handled like any other components, such as capacitors, resistors, or integrated circuits.

4.5.1 *Monopole-Type Ceramic Antenna*

In a state-of-the-art cellular phone, there is most likely more than one antenna. The antenna used for the voice communication is normally referred to as the primary antenna. All other antennas are referred to as auxiliary antennas, such as GPS, Bluetooth, or wireless LAN (WLAN) antennas. Almost all ceramic antennas are narrow-band antennas. Due to their limited bandwidth, they are mostly used as the auxiliary antennas. Some companies have proposed using ceramic antenna as the main antenna, but this has not been popular due to the performance issue.

For the primary antenna, the performance is critical, because users perceive the quality of a phone mostly by its sound's quality. If all other phones can make a call in one spot but one phone cannot, the reputation of the phone will be badly damaged. On the other hand, auxiliary antennas are used in applications which are not critical. People do not expect seamless coverage for those applications in the same way they do for voice communication. Thus, the performance requirements for auxiliary antennas are not that stringent. A primary antenna must be custom-made as optimized for each phone model to squeeze the last drop of the performance from the available physical space. On the other hand, ceramic antennas are designed and built as off-the-shelf components. The main consideration is reuse and standardization.

Shown in Figure 4.51 are two types of monopole ceramic antennas. One is made by Panasonic, which is a coil-type antenna. The other one is made by Johanson Technology, which adopts the multilayer technology. The plating and laser etching process are involved in making coil-type ceramic antennas. A small piece of ceramic block is plated with metal first. Then a procedure, which is analogous to making a screw, is used to laser etch a spiral slot on the metal layer. The end product is a metal coil rounding around a ceramic block. The axis of the coil aligns with the longest axis of the ceramic block. This process is quite similar to the one used in making metal film resistors. The advantage of this process is that the resonant frequency can be adjusted quite easily. The number of turns of a coil, which corresponds to the resonant frequency, can be conveniently adjusted by reprogramming the laser fixture.

Multilayer-type ceramic antennas can be made by either the standard chip capacitor process or low-temperature co-fired ceramic (LTCC) technology. From the electromagnetic point of view, both processes are the same. They stack multiple layers of meander lines and make an interconnection between layers. The end product is a 3D meander line. The relation between a coil ceramic antenna and a multilayer ceramic antenna is analogous to a helix-stubby antenna and a meander line stubby antenna.

A monopole ceramic antenna is still a monopole, which means its relative position to a ground plane must be correctly arranged. Shown in Figure 4.52 are two incorrect ways of placing monopole-type ceramic antennas. Why they do not work well can be explained by the

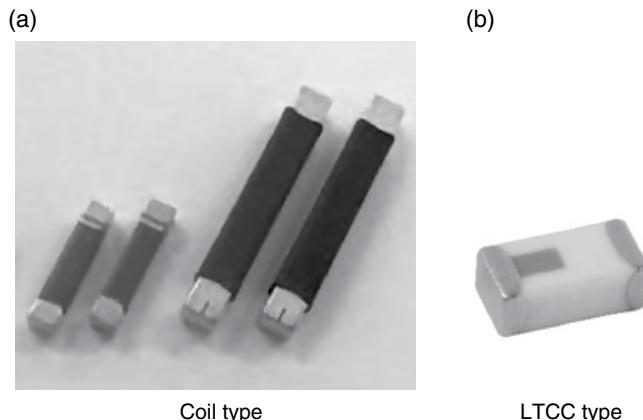


Figure 4.51 Monopole-type ceramic antennas (not to scale). (a) (Source: Panasonic Corporation.) and (b) (Source: Reproduced with permission of Johanson Technology.)

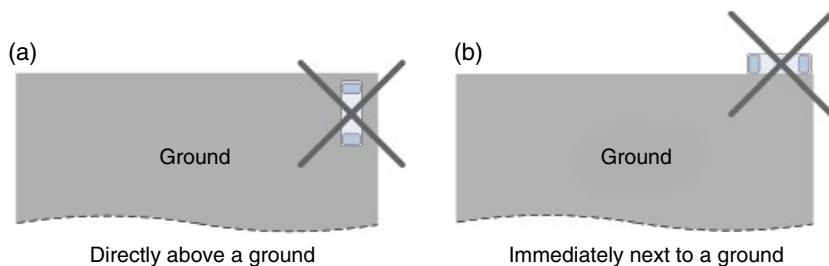


Figure 4.52 Wrong ways of placing monopole-type ceramic antennas.

image theory. For both placements, the image current on the ground is in the direction opposite to the current on the antenna, and the separation between the two currents is too small. The radiation forms of the respective currents cancel each other in the far field, and that prevents the antenna from radiating. If measuring the reflection coefficient of either antenna shown in Figure 4.52a or b, we might not observe any resonance at all.

Shown in Figure 4.53 are three correct ways of placing monopole-type ceramic antennas. The one shown in Figure 4.53a is the best scenario, where the antenna is vertically placed on the edge of ground plane. All the copper layers on the top portion of PCB are clearly outside. When using this kind of layout, the antenna is identical to an external stubby antenna except it is concealed from the cosmetic point of view. Many companies use this kind of antenna placement when measuring the antenna's performance.

In Figure 4.53a, the keep-out area means no components or traces can be routed in that area, and that is quite a waste, especially for a cellular phone, where it is always a struggle for space to fit in all the components. Shown in Figure 4.53b is a real scenario. The keep-out area has significantly shrunk. In this kind of layout, the dimension D , which is the distance between the antenna and the vertical edge of the ground, is the critical parameter. By decreasing D , the keep-out area can be smaller; however, the antenna's bandwidth and

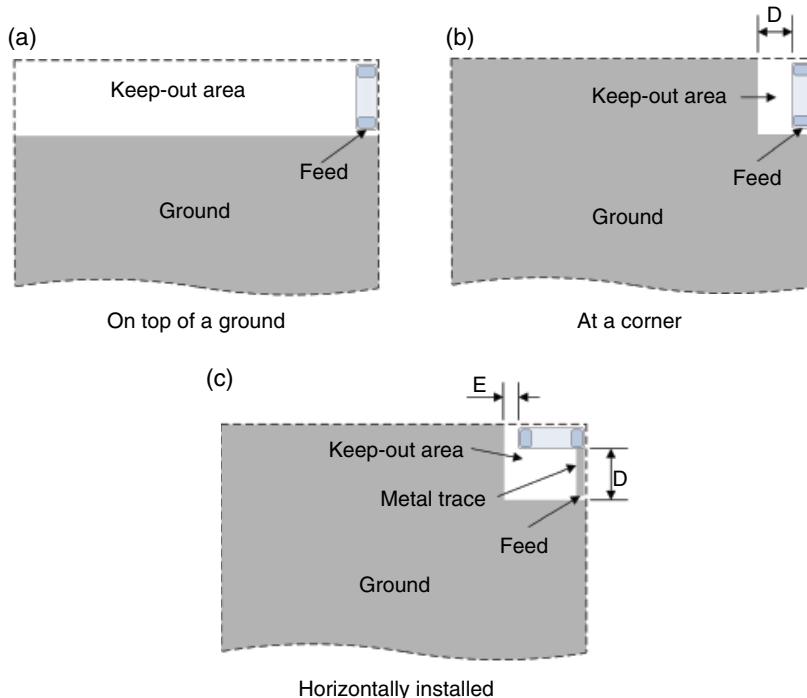


Figure 4.53 Correct ways of placing monopole-type ceramic antennas.

efficiency are sacrificed. So it is an engineering trade-off again: you have to pay something to gain something.

Shown in Figure 4.53c is another frequently used layout. The antenna is horizontally placed and a segment of metal trace is used to feed the antenna. In this arrangement, both dimensions D and E are critical parameters. The larger the gap, the better the performance. When using a ceramic antenna, it is always easier to shift an antenna's resonant frequency down than shift it up. So some ceramic antennas are purposely made with a resonant frequency a little bit higher than the target band. When using this kind of antenna, a small segment metal trace on a PCB is required to tune the frequency down.

Shown in Figure 4.54 is the measure response of a typical WLAN chip antenna. The antenna's model number is 2450AT43A100, and it is made by Johanson Technology. It has a 10dB reflection coefficient bandwidth of 300MHz. The antenna is 1.2mm tall, 7mm long, and 2mm wide.

For all monopole ceramic antennas, the ground around and underneath the antenna must be removed. A monopole ceramic antenna itself can be quite small. It is normally around 1mm tall, a few millimeters wide, and less than 10mm long. Most companies are often proposing those antennas as small footprint solutions. As antenna designers, we need to be very careful here. If an antenna claims a footprint of $2\text{ mm} \times 10\text{ mm}$ and we request an antenna area accordingly, we will soon find ourselves stuck in a bad situation. We must not forget to include the keep-out area. As a rule of thumb, it is always much easier to get antenna space at the start of a project rather than at the end.

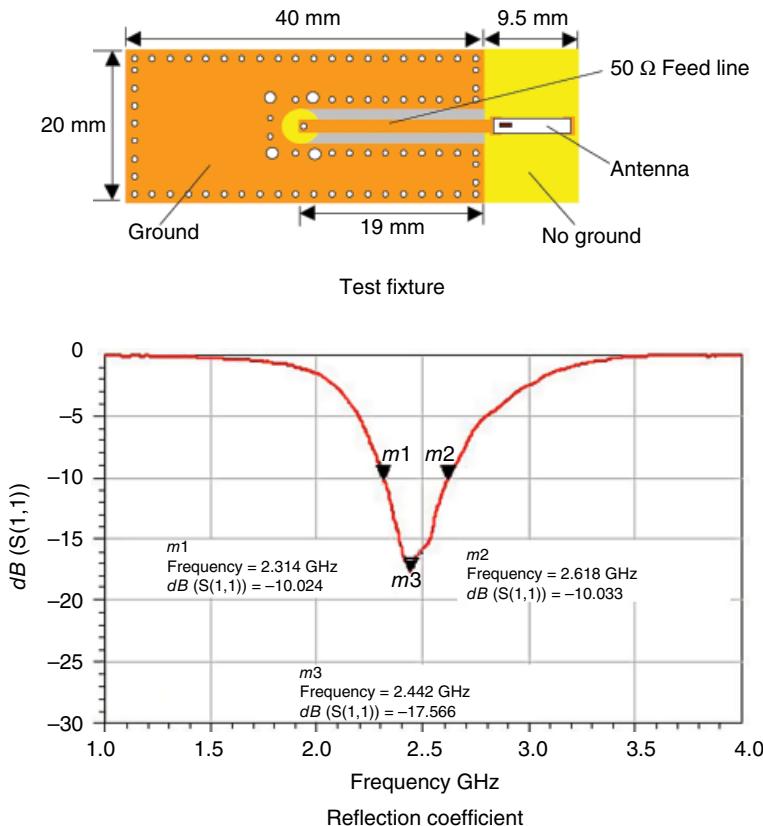


Figure 4.54 Performance of Johanson Technology 2450AT43A100. (Source: Reproduced with permission of Johanson Technology.)

4.5.2 IFA-Type Ceramic Antenna

Shown in Figure 4.55 is an IFA ceramic GPS antenna made by Murata Inc.©. To increase the effective length of the antenna, the metal trace is routed into a U shape at the top surface of the ceramic block. On the bottom of the antenna, there are three solder pads or terminals. The antenna is fed from terminal (1). Terminal (1) on the bottom is connected to the antenna on the top through the sidewall. However, there is no galvanic contact between terminal (1) and the antenna. There is a gap between them on the sidewall. The gap functions as a series capacitor which is a part of the matching network. The other matching component required is a shunt capacitor installed on a PCB. If we remove these two capacitors, it is easy to see that without any matching, the antenna impedance is inductive and is on top side of the Smith chart. Terminal (2) is the grounding pad, through which the antenna is connected to the system ground. Terminal (3) of the antenna is a tuning pad. This pad is DC floating but electromagnetically coupled to the antenna. By connecting a capacitor or an inductor, the effective loading of an antenna can be changed, thus fine-tuning of the antenna can be achieved.

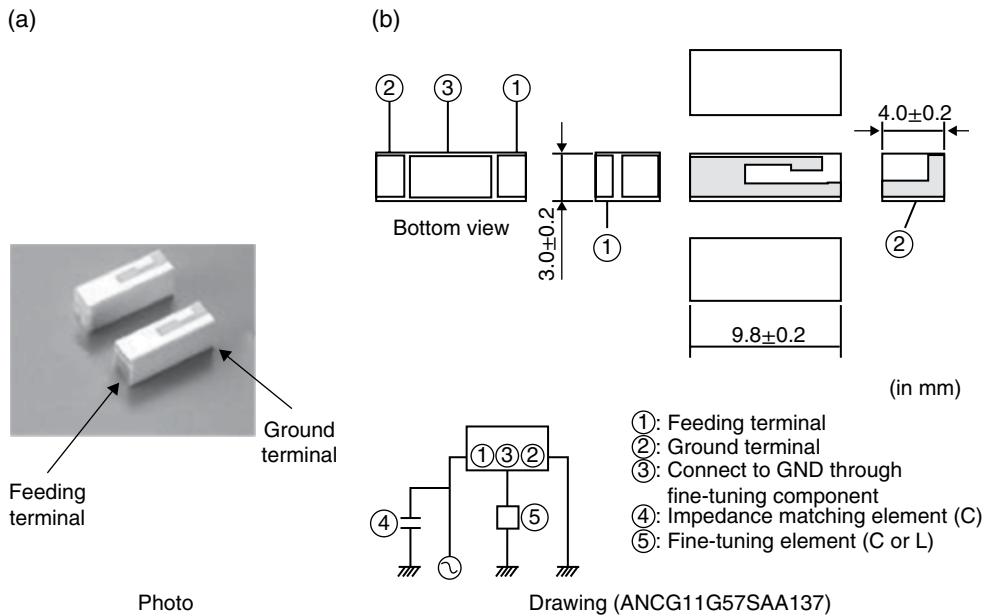


Figure 4.55 IFA ceramic GPS antenna. (Source: Murata Manufacturing Co. Ltd.)

An IFA ceramic antenna is also marketed as an over-ground chip antenna solution. Based on the basic principle of image theory, we know that an antenna must be adequately separated from the ground to radiate effectively. The antenna shown in Figure 4.55 is 4 mm tall, which is about four times taller than its counterparts of monopole ceramic antennas. As an example of such antennas, a Murata WLAN IFA ceramic antenna (ANCG12G44SAA145) has a 4.5 dB reflection coefficient bandwidth of 84 MHz. The antenna is 3.8 mm tall, 9.8 mm long, and 2 mm wide. If comparing with a monopole antenna by the absolute size of the ceramic block, an IFA ceramic antenna is significantly bigger and its bandwidth is a lot narrower. However, an IFA ceramic antenna does not require a keep-out area on the PCB, so the effective antenna space it requires is actually less than a monopole.

In practice, the PCB area around antenna can be safely used to lay out small components. However, be cautious when placing large metal components, such as the shield box, battery, and so on, around the antenna. The antenna's performance will be significantly degraded when it is immediately adjacent to a metal object.

4.5.3 Loop-Type Ceramic Antenna

Shown in Figure 4.56 is a loop ceramic antenna. The antenna is 1.85 mm tall, 4 mm long, and 1.5 mm wide. As illustrated in Figure 4.56b, all gray areas are metallized surface. The two long sidewalls are cleared of metal. At the bottom of the antenna, two slits are cut to form two solder pads. On the top of the antenna, there is one slit. It might seem small, but be careful because the “antenna” is actually only a component of a loop antenna. Shown in Figure 4.56c is a PCB layout of this antenna. The effective antenna is the loop around the

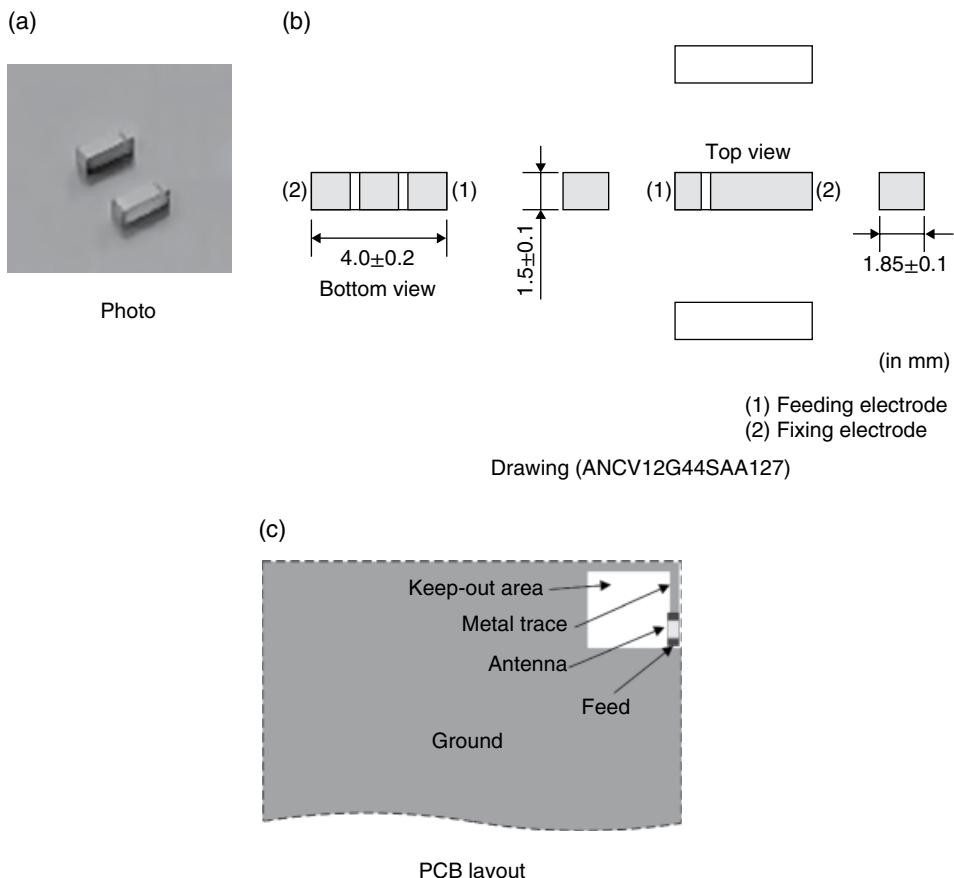


Figure 4.56 Loop ceramic WLAN antenna. (Source: Murata Manufacturing Co. Ltd.)

keep-out area. The ceramic chip antenna functions more or less like a series capacitor. However, the ceramic chip is comparably big, so it does play a role in the antenna's radiation. The ceramic chip is made of high dielectric material that effectively decreases the overall size of the loop antenna. The chip is composed of low loss dielectric and has a high concentration of electromagnetic field, which improves the radiation efficiency. By cutting slits on a ceramic block, the capacitance of the chip can be better controlled than a multilayer ceramic capacitor, which improves frequency stability.

The impedance of a loop-type ceramic antenna is affected by the loop shape and position. In most circumstances, a shunt matching component is needed to achieve good matching. For the WLAN antenna chip (ANCV12G44SAA127), the -5 dB reflection coefficient bandwidth given in the datasheet is 84 MHz.

In conclusion, of monopole, IFA and loop ceramic antennas, the one that occupies the least PCB area is ceramic IFAs. For monopole and loop ceramic antennas, there is a design trade-off between required PCB area and the antenna's performance. There are different variants of

ceramic antennas in the market, but in reality they can be analyzed by the way we are familiar with. After all, the ceramic antenna is not a startling new innovation, but it is only a different way of manufacturing antennas.

4.6 Slot Antenna

In recent years, slot antennas have been adopted widely in published papers as an alternative of mobile antenna designs. The theoretical result of a slot antenna has been available since 1946 [26]. In most antenna textbooks [7], slot antennas are always introduced as a complementary structure of metal antennas, as shown in Figure 4.57. It is assumed that the slot antenna is etched on an infinite perfect electric conductor, which is often referred to as the ground plane.

The power radiation patterns of both a slot antenna and a dipole antenna are identical, as shown in Figure 4.58. Both antennas have omnidirectional radiation patterns and the maximum gain appears in the azimuth plane. However, the orientations of electric and magnetic fields are swapped. If a dipole antenna is placed along Z-axis, its vector of electric field in the azimuth plane is all perpendicular to the plane. If it is a slot antenna, all vectors of electric field in the azimuth plane are parallel to the plane.

For the slot antenna, most textbooks stop at here. Pretty much all students get a wrong impression that they can design a real-world slot antenna and get an omnidirectional radiation pattern as shown in Figure 4.58b. Shown in Figure 4.59a is a slot antenna on a finite ground. The metal plate has a square shape and the length of each side is one wavelength. The slot length is half of one wavelength. Shown in Figure 4.59b is its radiation pattern. Along the Y-axis, where the maximum gain supposes to appear, there is a deep null. On the other hand, there is some radiation along the Z-axis, which is the direction where nulls should appear according to classical textbooks [7].

Both radiation patterns are correct. The discrepancy is generated by the different ground condition where slot antennas are etched on. Shown in Figure 4.60 is the electric fluxlines in the XY plane, which is the symmetric plane cutting through the center of the slot. Two horizontal thick solid lines are cross sections of the ground. The gray curves are electric flux-lines. When the ground plane is finite, electric fields on the top and bottom surface will meet each other at the ends of the ground plane, which are marked by dashed-line circles in Figure 4.60. Because electric fields on the top and bottom surface have the same amplitude, but point to opposite directions, the combined radiation along the Y-axis is always zero.

If the ground is infinite, it is a different story. Because the electric field on the top surface will never meet its counterpart on the bottom, they can peacefully coexist and radiate outwardly. In the real world, one can never find an infinite ground, so radiation patterns always like Figure 4.59b instead of the textbook's results shown in Figure 4.58b.

The next phenomena need to be explained is why there is radiation along Z-axis. Shown in Figure 4.61 is instantaneous current distribution on a finite ground of a slot antenna. It is obvious that the current is not constrained around the slot, the whole ground has been excited. Horizontal currents, which are on top and bottom edges along the Y-direction, function as an electric dipole antenna along the Y-axis. Those currents are responsible to radiation along the Z-axis.

In published papers, there are two kinds of slot antennas. One kind is etched in the middle of a ground and the slot are surrounded by metal, as shown in Figure 4.62a. The basic

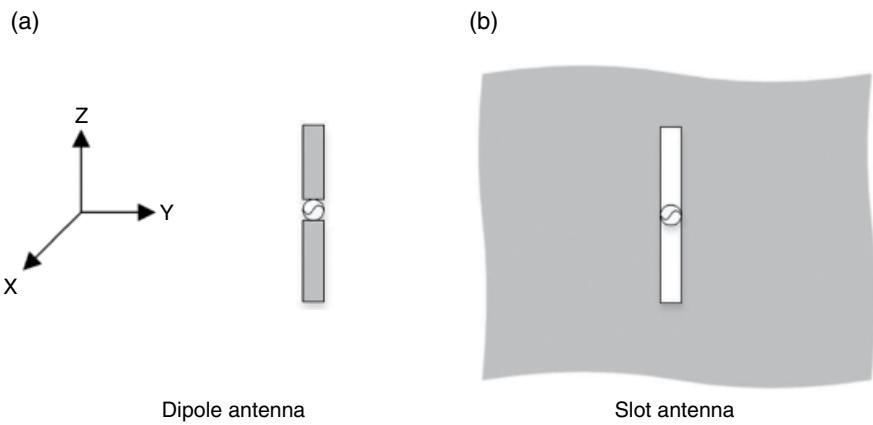


Figure 4.57 Slot antenna and its complimentary dipole.

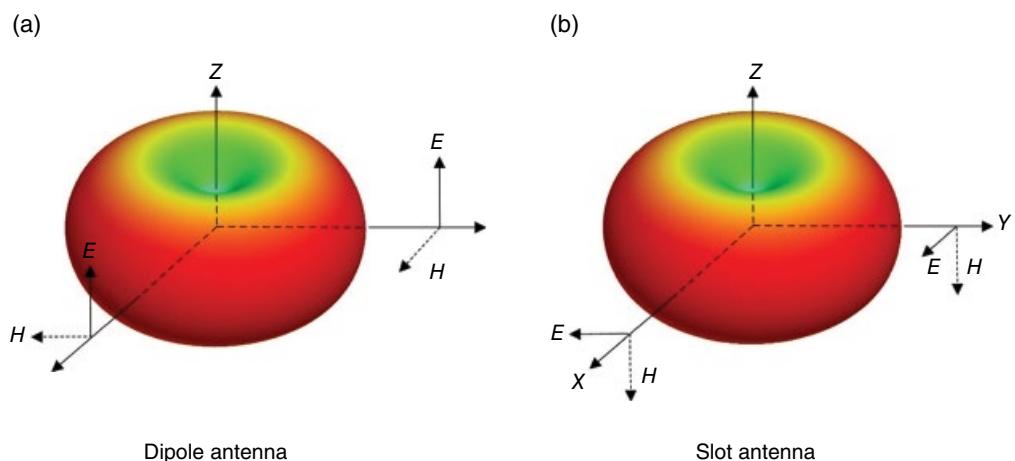


Figure 4.58 Three-dimensional radiation patterns of dipole and slot antenna.

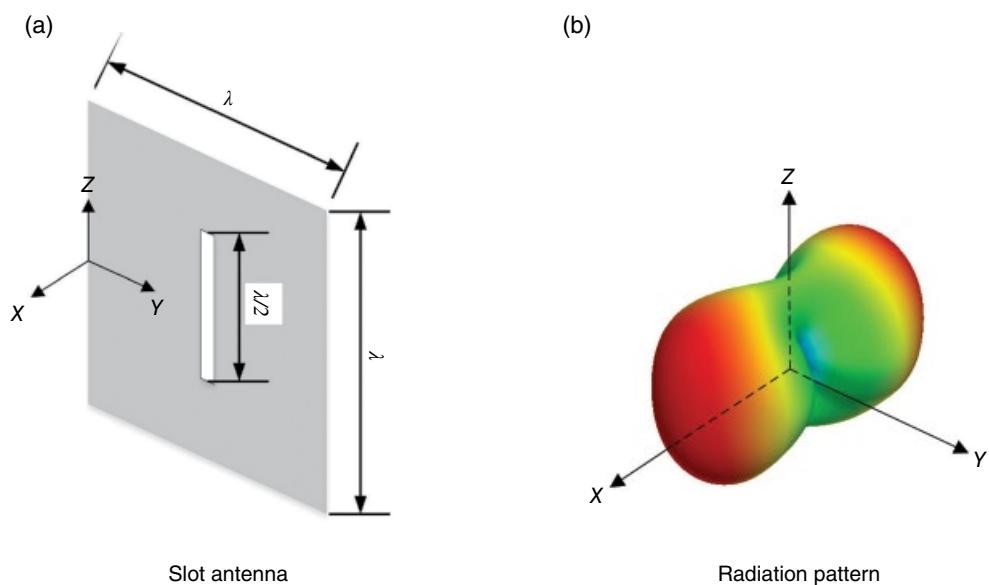


Figure 4.59 A real-world slot antenna and its radiation pattern.

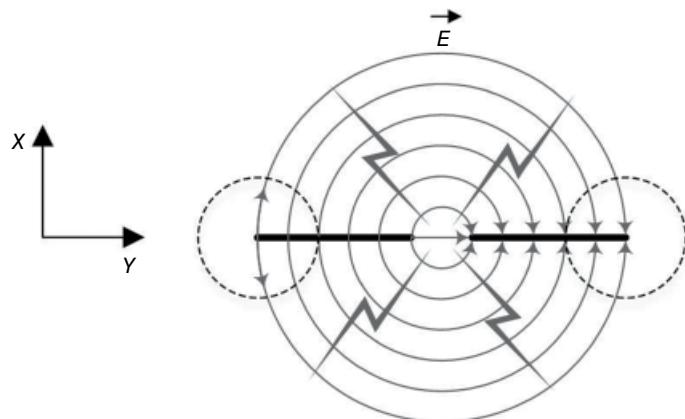


Figure 4.60 Explanation of the radiation null along the Y -axis.

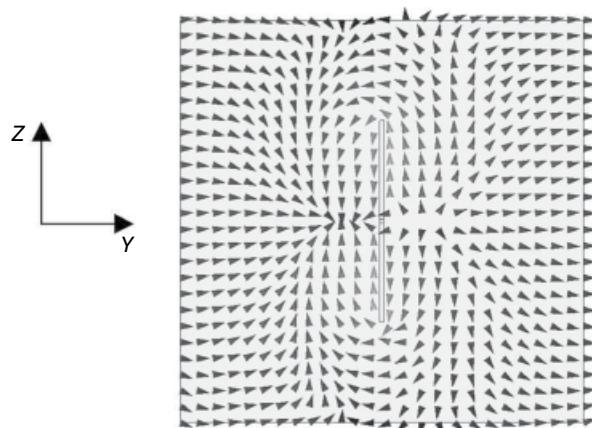


Figure 4.61 Instantaneous current distribution on a finite ground.

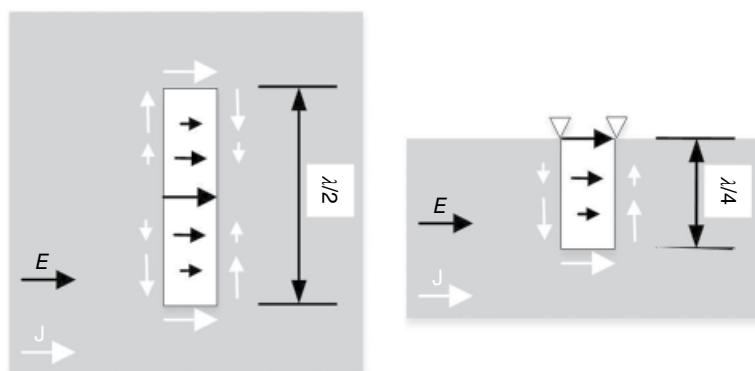


Figure 4.62 Different modes of a slot antenna

radiation mode of this type of slot is half-wavelength mode, which means the slot length is half of one wavelength. Both electric field inside the slot and current on the metal are given in Figure 4.62a. The electric field, which is illustrated as black arrows, reaches its maximum strength at the center of the slot. The electric field decays until it is zero at the end of the slot. On the contrary, the current density, marked as white arrows, has the maximum value at the both end of the slot. The minimum current spot is at the center.

The other kind of slot antenna is shown in Figure 4.62b. The slot is etched on the edge of a metal. In most published papers, people like to give the explanation as following. As the one end of the slot is opened and the other end is shorted, the structure can support a quarter of a standing wave. Thus, it is given the name of a quarter-wavelength mode slot antenna. Comparing Figure 4.62a and b, the quarter mode can be considered as a half of the half-wavelength mode. Because the quarter mode slot requires less board to achieve a desired frequency, it is more popular than its half-wavelength counterpart.

In the real world, the zero current spots, which are marked by two inverted hollow triangles in Figure 4.62b, can't stop current from overflowing to adjacent conductor. Depending on where a quarter-wavelength slot is placed on a ground, the antenna's response can be significantly different. Shown in Figure 4.63a is a top placed quarter mode slot antenna. For the convenience of comparison, the ground plane has the same dimension as the IFA shown in Figure 4.5. The slot length is 65 mm. The feeding point is 7 mm away from the shorted end of the slot. Theoretically, the resonant frequency of this slot antenna should be 1.15 GHz. As shown in Figure 4.63b, the simulated result (1.11 GHz) is quite close to the theoretical one.

However, if we look at the radiation pattern as shown in Figure 4.63c, it is more like a vertically placed dipole antenna than a horizontal slot antenna. Current distribution shown in Figure 4.63d further confirms the observation. Strong currents are induced along vertical edges of the ground and those currents are the dominant contributor of radiated power. Now, let's compare this slot antenna's results with IFA's given in Figure 4.6 and Figure 4.9. It is obvious that they are pretty much the same. In fact, from electromagnetic point of view this type of quarter-wavelength slot antenna is almost identical to the IFA discussed in Section 4.1.

If the quarter slot antenna shown in Figure 4.63a is relocated to the middle portion of a ground, as shown in Figure 4.64a, the antenna is no longer matched. By tweaking the position of the feeding point, which is moving the feed away from the end of the slot and increasing the distance from 7 to 12 mm, the antenna can be rematched. However, the resonant frequency is 0.84 GHz, which is much lower than the theoretical one (1.15 GHz) based on a quarter-wavelength estimation.

In fact, the slot is not a quarter-wavelength mode slot antenna at all. By looking at the current distribution shown in Figure 4.64c, strong current are excited on the whole ground plane. The dominant radiator is the metal ground instead of the open ended slot.

As the ground is a huge metal structure, it is possible to design a wideband antenna by properly exciting the ground. By using matching techniques discussed in Chapter 2, a "slot" antenna was tweaked to cover a band from 0.9 to 2.2 GHz with better than 10 dB return loss. The simulation results are shown in Figure 4.65b. The ground dimensions shown in Figure 4.65a are identical to the ones in Figure 4.64a. However, the slot is 48 mm long and the feeding position is 13 mm away from the shorted end of the slot. A 2.5 pF series capacitor is used to achieve better matching.

By looking at Figure 4.65b, one might think that a silver bullet has been found. Please don't be overexcited, the design will not work well in a real phone. Most state of the art phones have

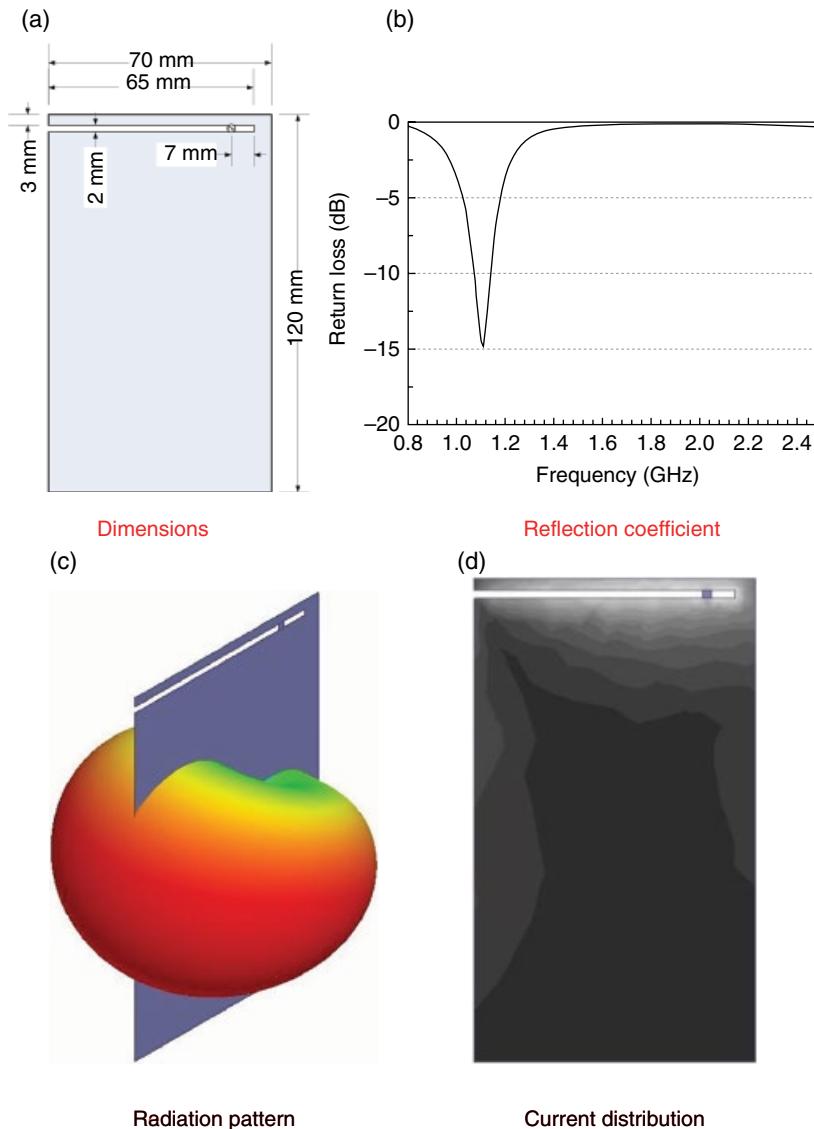


Figure 4.63 A quarter-wavelength slot on the top.

large screens. For electromagnetic compatible reason, a metal shield is applied underneath an LCD screen. This shield destroys current distribution of the wideband antenna shown in Figure 4.65. Shown in Figure 4.66a is a wideband antenna placed next to a metal shield. The metal shield is 10mm, which is thicker than most current phones, away from the antenna. Shown in Figure 4.66b is the simulation result. Obviously, the antenna is no longer a wideband antenna anymore.

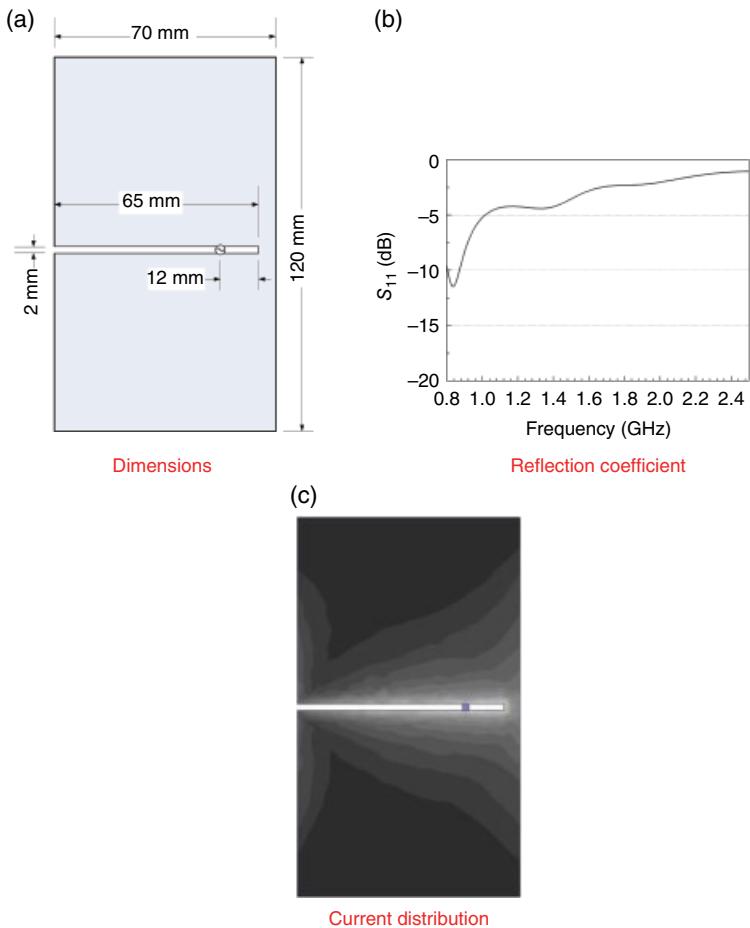


Figure 4.64 A quarter-wavelength slot in the middle.

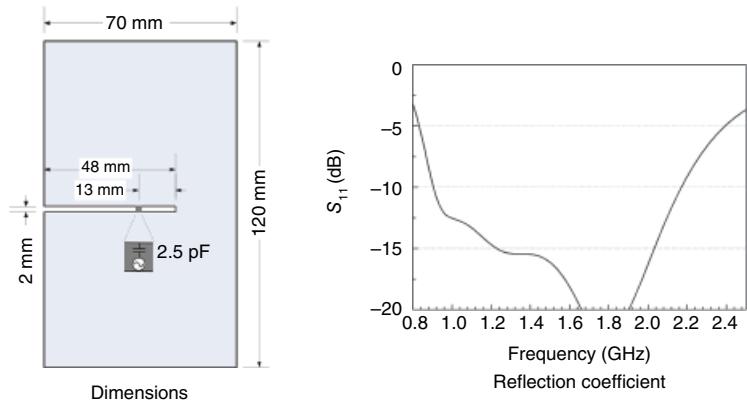


Figure 4.65 A wideband antenna which is excited by a slot.

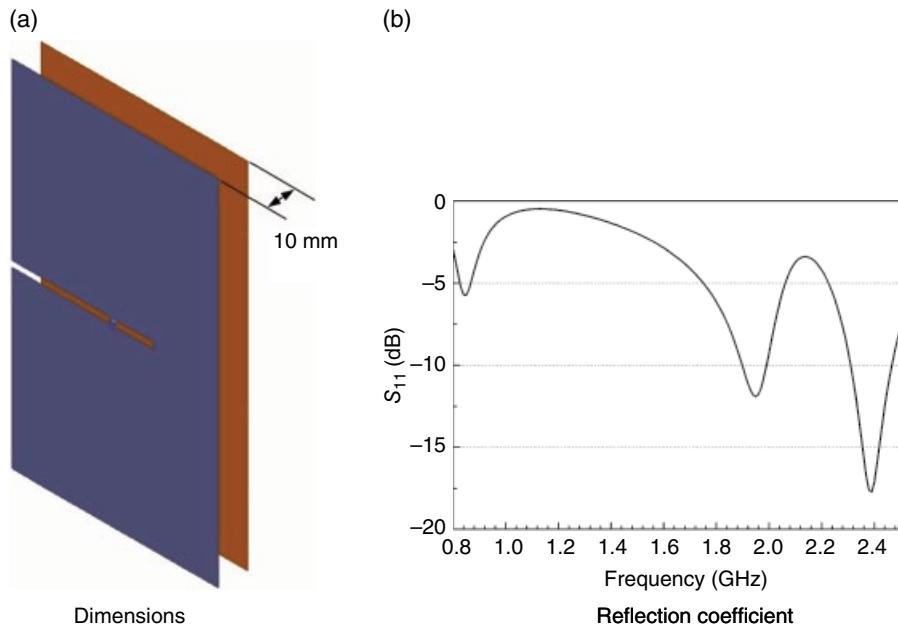


Figure 4.66 Impact of LCD screen shield.

4.7 Design a Hepta-Band Antenna with Multiple Radiators and Multiple Modes

As discussed in Section 3.1, techniques used in multiband external antennas are multiple radiators and multiple modes for each radiator. When design an internal antenna, the basic principle is pretty much the same. The challenge facing internal antenna designs is its limited space. Radiators used in internal antenna designing can be IFA, loop, slot, and their various mutations.

In the following example [27], monopole and open slot antennas are integrated in a planar structure. Two 0.25λ modes radiators, one monopole and one open slot, are combined to cover the lower band. Two 0.75λ modes of the same radiators are also adopted to provide bandwidth for the upper band. To cover the whole upper band and have more freedom in tuning the bandwidth of upper band, two more radiators are added.

Figure 4.67a shows the geometry of the planar handset antenna. Detailed dimensions of the antenna are given in Figure 4.67b. A 0.8-mm-thick FR4 substrate ($\epsilon_r = 4.4$, $\tan\delta = 0.02$) is used as the system circuit board. Three monopole-type branches are printed on the front side of the substrate. Branch 1 is a U-shaped strip, branch 2 is a straight strip, and branch 3 is an L-shaped strip. A tuning pad is connected to the longer arm of branch 1. The three branches are directly fed by a $50\text{-}\Omega$ microstrip line. The ground plane is printed on the back side of the substrate. An open slot is etched on the ground plane with a length of 59 mm. The main ground, namely the part at one side of the slot, has a size of $100\text{ mm} \times 60\text{ mm}$. Another parasitic ground branch (branch 4) is placed at the other side of the slot.

The tuning process of the lower band is shown in Figure 4.68. In the beginning, a folded monopole (branch 1) antenna is shown in Type I. The resonance of the monopole in the lower

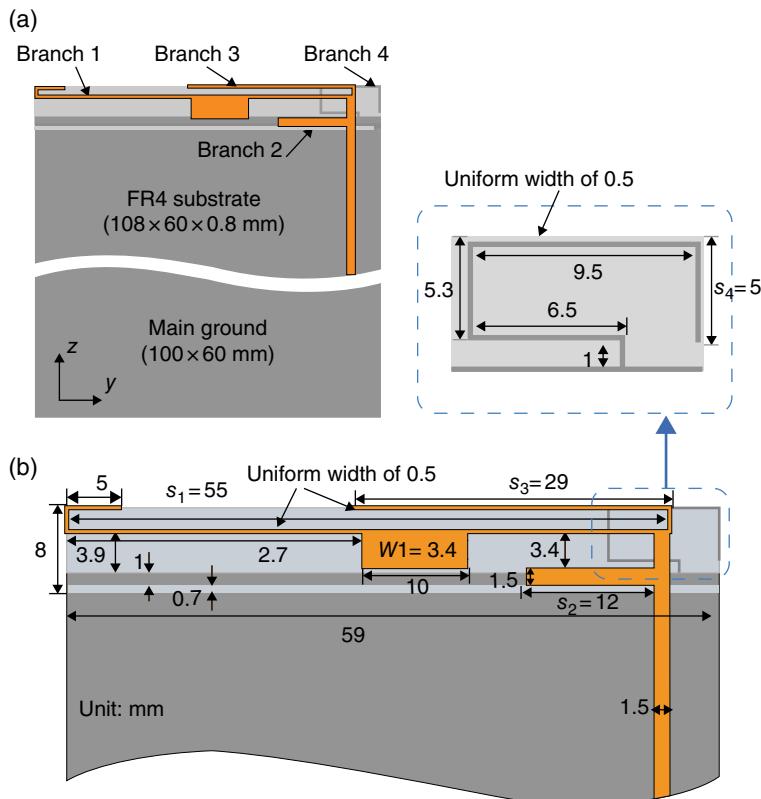


Figure 4.67 Geometry and dimensions of the antenna. (Source: Deng *et al.* [27]. Reproduced with permission of IEEE.)

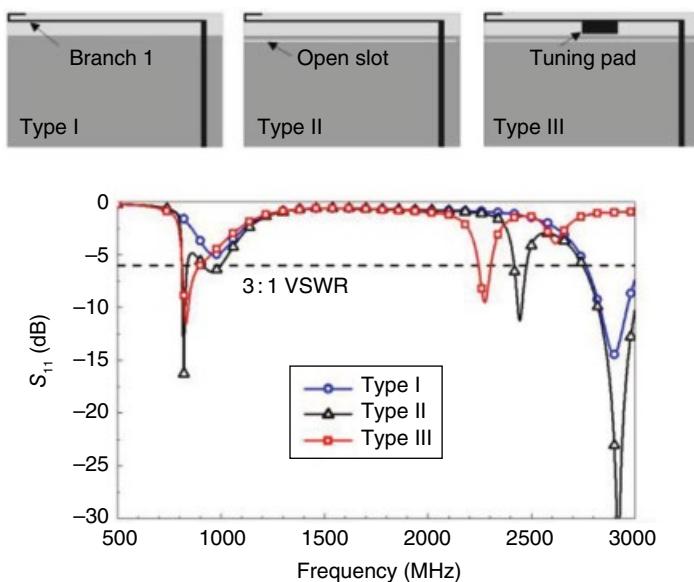


Figure 4.68 Comparison of simulated S_{11} for different antenna types. Type I: only branch 1, Type II: Type I+open slot, and Type III: Type II+tuning pad. (Source: Deng *et al.* [27]. Reproduced with permission of IEEE.)

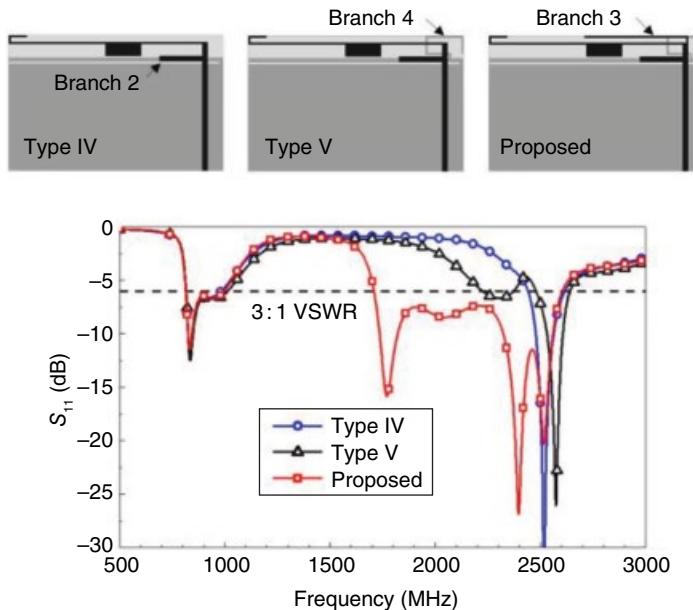


Figure 4.69 Comparison of simulated S_{11} for different antenna types. Type IV: Type III+branch 2, Type V: Type IV+branch 4, proposed: Type V+ branch 3. (Source: Deng *et al.* [27]. Reproduced with permission of IEEE.)

band is at about 990 MHz, corresponding to the 0.25λ mode of branch 1. Then, an open slot is etched on the ground plane in Type II. It is shown that another resonance in the lower band is generated at about 850 MHz. The resonance is corresponding to the 0.25λ mode of the open slot. Then, a tuning pad is added in Type III. Comparing Type III with Type II, it is shown that the 0.25λ mode of branch 1 is deceased. It can be explained that the tuning pad increases the current path length of the monopole. By merging the two modes, the coverage of GSM850/900 operation is achieved. It is worth mentioning that LTE700 operation can also be covered if the slot width or the monopole profile is increased, but this improvement will also increase the profile of the antenna.

The tuning process of the upper band is shown in Figure 4.69. In Type III, there are two resonances in the upper band, which correspond to the two 0.75λ modes of the open slot and the monopole. Then, branch 2 is added in Type IV. The main purpose of branch 2 is to tune the two 0.75λ modes. Considering that the bandwidth provided by the two 0.75λ modes is limited, additional resonances are required to cover the whole bandwidth in the upper band. Then, branch 4 is added in Type V to provide a new resonance. The branch operates at its 0.25λ mode and generates a resonance at about 2300 MHz. Finally, branch 3 is added in the antenna to provide another resonance. The branch operates at its 0.25λ mode and generates a resonance at about 1800 MHz. It is shown that branch 3 can also tune the whole upper band effectively. By merging the four modes, the coverage of the upper band is achieved.

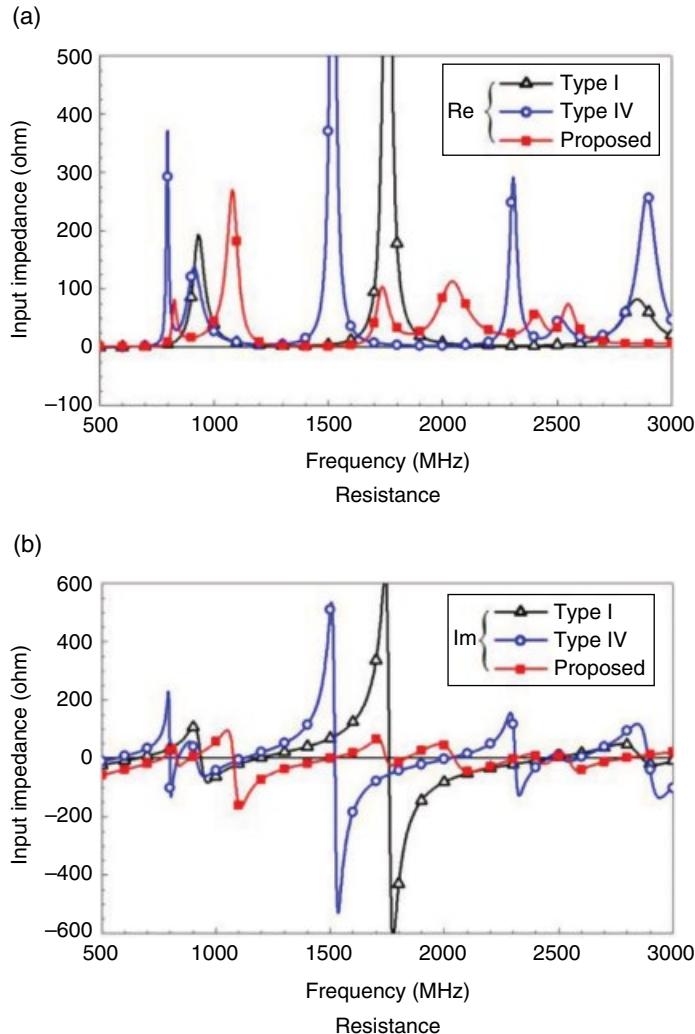


Figure 4.70 Simulated input impedance of the antenna. (*Source:* Deng *et al.* [27]. Reproduced with permission of IEEE.)

Figure 4.70 compares the simulated input impedance of three typical antennas. It is observed that the monopole in Type I generates a resonance at about 990 MHz in the lower band. By applying the coupling feed, the open slot is excited and an additional resonance is provided at about 850 MHz in Type IV. The comparison of the two antennas indicates the mechanism of feeding in the lower band. Based on Type IV, the proposed antenna has an additional monopole branch and ground strip. It is shown that two more resonances in the upper band are excited in the proposed antenna. The comparison of the two antennas indicates the mechanism of feeding in the upper band.

The simulated distribution of the electric field and the surface current at different resonant frequencies are shown in Figure 4.71. The corresponding radiating antenna part for each

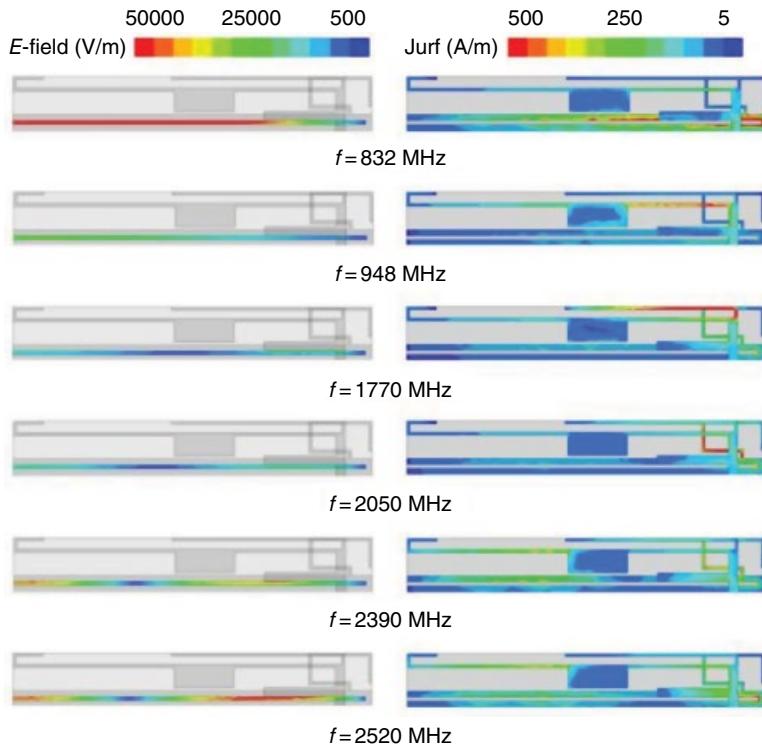


Figure 4.71 Simulated distribution of the electric field in the open slot and the surface current of the antenna. (Source: Deng *et al.* [27]. Reproduced with permission of IEEE.)

resonant mode is clearly shown. To be specific, the resonance at 832 MHz is generated by the open slot, the resonance at 948 MHz is generated by branch 1, the resonances at 1770 and 2050 MHz are generated by branch 3 and branch 4, and the resonances at 2390 and 2520 MHz are generated by the third-order modes of branch 1 and the open slot.

Following are some parametric studies. As the lower band is difficult to cover, the key parameters for the lower band are studied first. There are two resonances in the lower band, which are generated by branch 1 and the open slot. Therefore, the lengths of branch 1 and the open slot are the key factors to tune the two modes. Figure 4.72a shows the simulated S_{11} with different lengths of branch 1. It is shown that increasing s_1 can both decrease the 0.25λ mode and 0.75λ mode of branch 1. However, the length of branch 1 is limited by the width of the ground plane. Therefore, a tuning pad is added to further decrease the 0.25λ mode of branch 1. It is shown in Figure 4.72b that the 0.25λ mode of branch 1 decreases with the increase of the tuning pad width w_1 . Besides, the two 0.75λ modes of branch 1 and the open slot also decrease with the increase of w_1 . It can be explained that the tuning pad works as a shunt capacitor. These results clearly indicate that the lower band can be effectively controlled by tuning the length of branch 1 and the tuning pad width.

The key parameters for the upper band are also studied. There are four resonances in the upper band, namely the 0.75λ mode of the open slot, the 0.75λ mode of branch 1, the 0.25λ

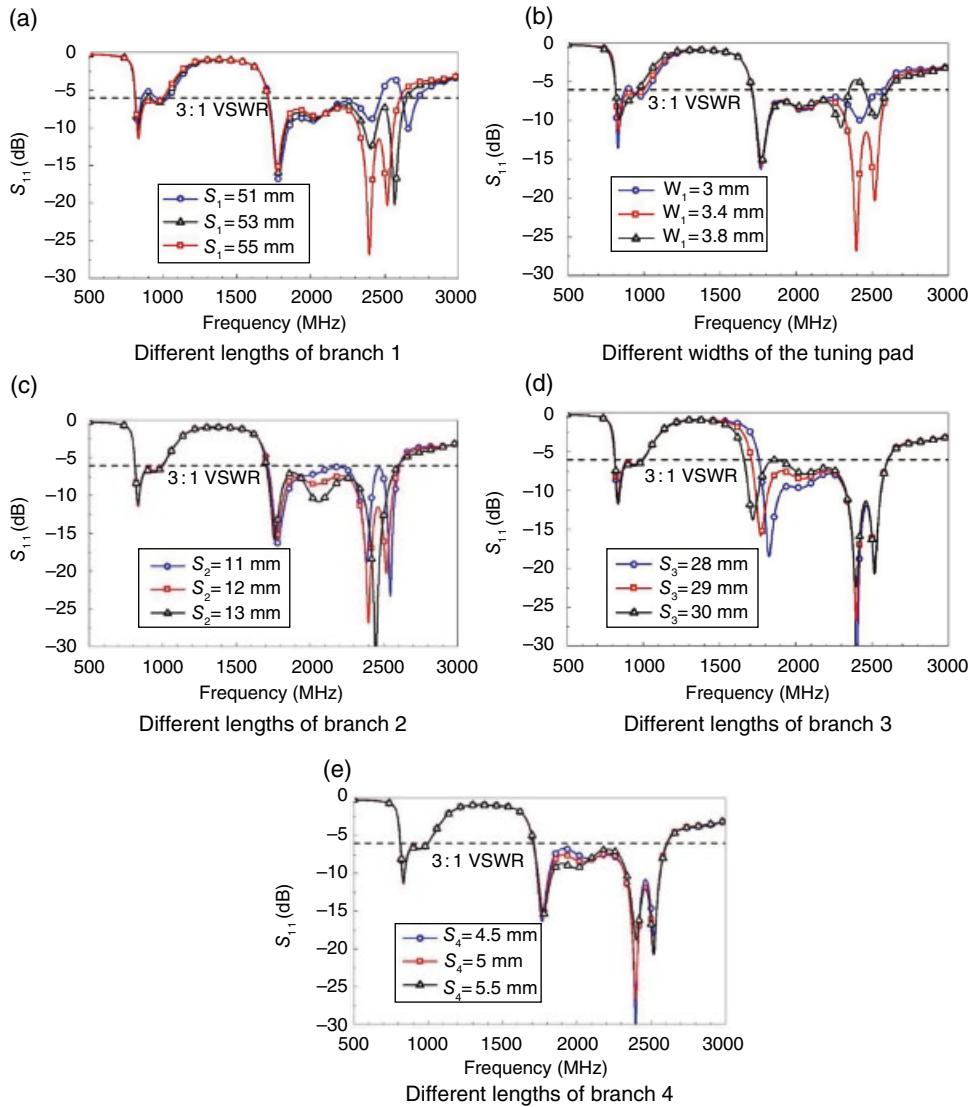


Figure 4.72 Parametric studies. (Source: Deng *et al.* [27]. Reproduced with permission of IEEE.)

mode of branch 3, and the 0.25λ mode of branch 4. As the lengths of branch 1 and the open slot are used to tune the 0.25λ modes, an alternative parameter is found to tune the two 0.75λ modes. Figure 4.72c shows the effect of branch 2. When s_2 increases, the third-order mode of branch 1 decreases, but the third-order mode of the open slot increases. Figure 4.72d shows that branch 3 only affects the resonance at 1800 MHz. A similar phenomenon can

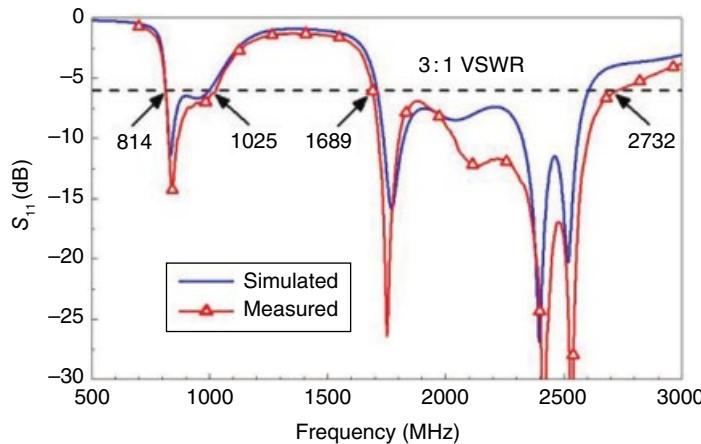


Figure 4.73 Simulated and measured S_{11} of the antenna. (Source: Deng *et al.* [27]. Reproduced with permission of IEEE.)

be found in Figure 4.72e, where the resonance is determined by branch 4. It is worth mentioning that the two resonances in the lower band keep almost unchanged during the parameter tuning of the upper band. These results indicate that the four modes in the upper band can be tuned easily.

Figure 4.73 shows the simulated and measured reflection coefficients of the antenna. The difference between simulation and measurement is mainly caused by fabrication error and substrate property. Two resonances are observed in the lower band, and a bandwidth of 205 MHz (815–1020 MHz) is achieved, which covers the GSM850, GSM900 operations. Four resonances are observed in the upper band, and a bandwidth of 1040 MHz (1690–2730 MHz) is achieved, which covers DCS, PCS, UMTS, LTE2300, and LTE2500 operations.

The normalized radiation patterns of the antenna are shown in Figure 4.74. For the lower frequency 900 MHz, a dipole-like radiation pattern can be observed. For the upper frequencies 1900 and 2400 MHz, more variations and nulls appear in the patterns when compared with that at 900 MHz. The simulated and measured gain and efficiency in the lower and upper bands are presented in Figure 4.75a and b, respectively. For the lower band, the radiation efficiency is better than 40% and the antenna gain varies from -2 to 1 dBi. For the upper band, the radiation efficiency is about 44–70%, and the antenna gain varies from -2 to 2 dBi. The results deteriorate at the boundary of the concerned band, but are acceptable in practical mobile applications.

4.8 Design a Reconfigurable Hepta-Band Antenna

In Section 2.4, it has been demonstrated that reconfigurable matching can be used to expand an antenna's working bandwidth. The idea of reconfigure can also be applied to antenna design. Most Reconfigurable antennas can be categorized into two groups. One group uses

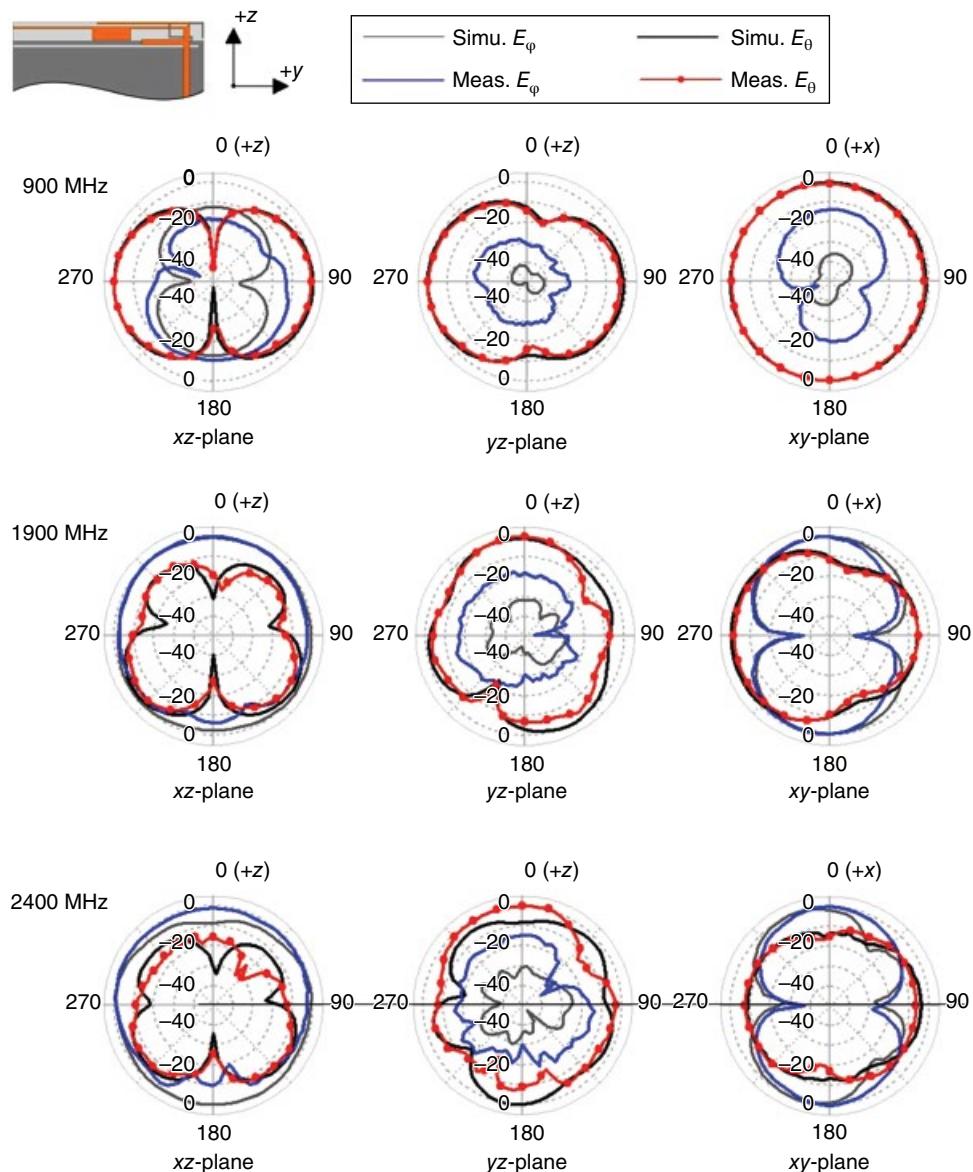


Figure 4.74 Simulated and measured radiation patterns. (Source: Deng *et al.* [27]. Reproduced with permission of IEEE.)

PIN diodes, which function as switches when different bias voltage is applied, or as MEMS switches. By inserting switches into radiators, effective length of radiators can be dynamically changed depending on the status (ON/OFF) of switches [28, 29]. Antenna's resonant frequency will shift accordingly to cover different bands. Switches can also be used to alter radiators between different modes [13, 30], such as IFA mode and loop mode. Because different modes are resonant at different frequencies, an antenna's frequency bands are expanded accordingly.

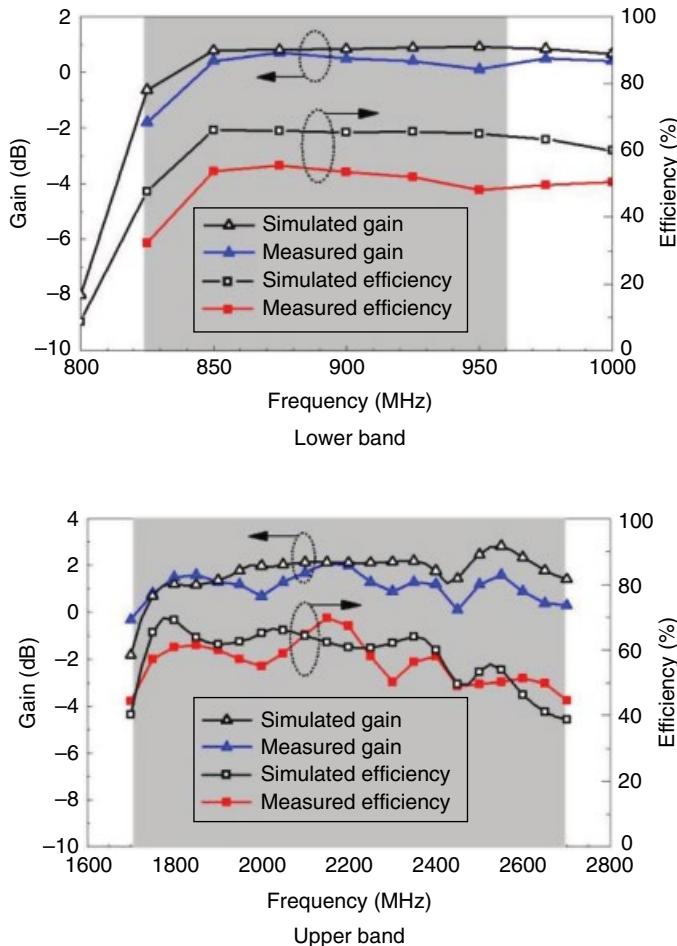


Figure 4.75 Simulated and measured gain and efficiency. (Source: Deng *et al.* [27]. Reproduced with permission of IEEE.)

When designing a reconfigurable antenna, one important guideline is to use as less switches as possible. In theory, by using a large amount of switches an antenna can cover as much frequency bands as one desires. However, all switches have inherent losses. More switches mean higher loss. On the other hand, each switch requires a dedicated bias circuit. A handful switches will make the bias circuit design more dreadful than the antenna design itself. Bias circuit also occupies space on a circuit board. If we had that much space, we should use it to design antenna instead of bias circuit.

The following is an example [31] of a folded loop antenna and can be reconfigured to IFA by breaking the loop, using only one PIN diode and with a simple bias circuit. Besides adopting the tuning pad proposed in [13, 14], a matching bridge is also utilized in both loop and IFA modes of the proposed antenna to achieve wider bandwidth. Hepta-band coverage is realized in a compact volume of $60\text{ mm} \times 5\text{ mm} \times 5\text{ mm}$, including GSM850, GSM900, GPS, DCS, PCS, UMTS, and WLAN.

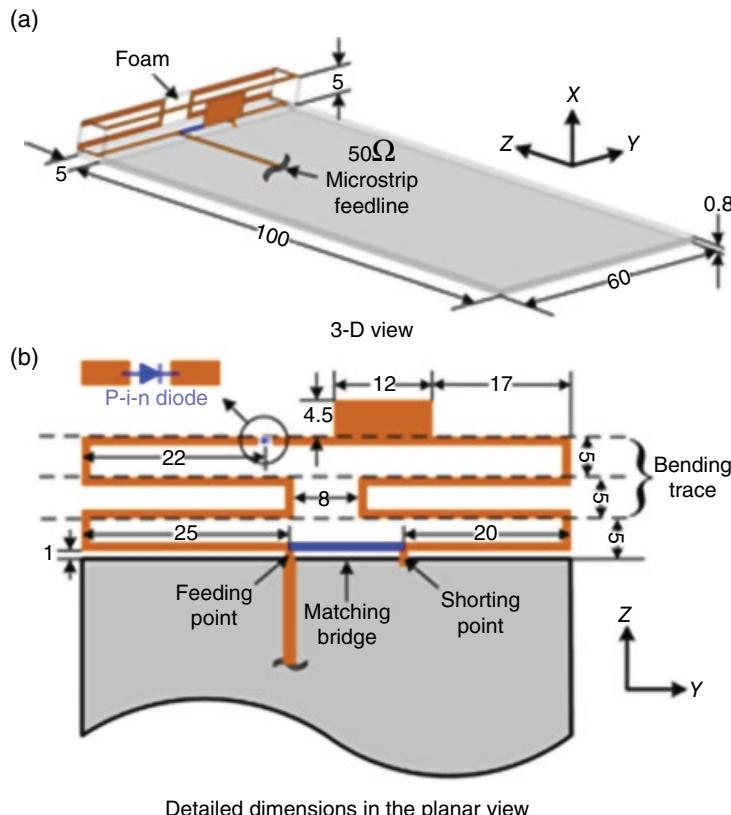


Figure 4.76 Geometry and dimensions of the proposed antenna. (Source: Li *et al.* [31]. Reproduced with permission of IEEE.)

Figure 4.76 shows the geometry of the proposed reconfigurable antenna, and the overall dimensions are $60\text{ mm} \times 5\text{ mm} \times 5\text{ mm}$. The folded loop antenna (uniform width of 1 mm) is supported by foam with the permittivity close to that of air. The antenna structure is arranged just outside the ground plane area and above the main board. The main board is made of FR4 substrate ($\epsilon_r=4.4$, $\tan\delta=0.02$), with a thickness of 0.8 mm. A $100\text{ mm} \times 60\text{ mm}$ metal ground is printed on the backside of the main board, connecting the shorting point of the loop through a via hole. A 50Ω microstrip line is arranged on the front side of main board and is connected to the feeding point of the loop. A shorting bridge, which has the same width as the loop (1 mm), connects the feeding point with the shorting point. The position of PIN diode is illustrated in Figure 4.76b. On the one hand, when the PIN diode is “ON,” the antenna works in a typical loop mode. On the other hand, when the PIN diode is “OFF,” the loop will be broken into two IFAs, and the left part (the one without tuning pad) is the main working branch. As a result, the operating modes of the proposed antenna can be switched by the state of PIN diode.

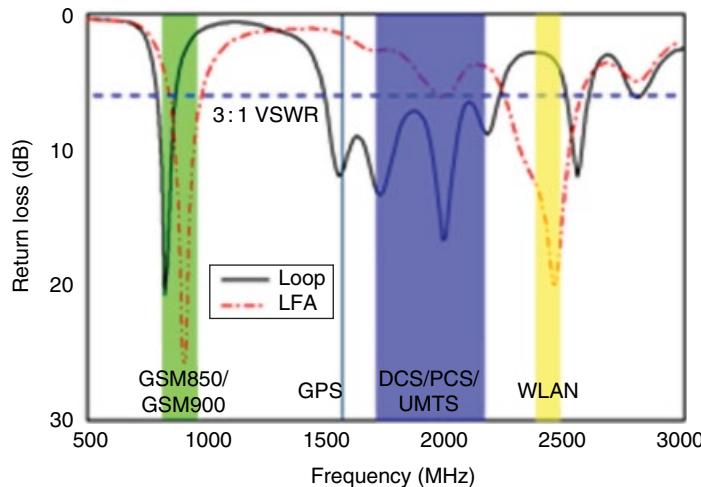


Figure 4.77 Measured return loss of loop and IFA modes of the proposed antenna. (Source: Li *et al.* [31]. Reproduced with permission of IEEE.)

The measured return losses of both working modes are shown in Figure 4.77. These data agree well with simulation results as will be shown in parametric studies later. Hepta-band is covered with -6 dB ($\text{VSWR} = 3:1$) by combining the bandwidths of two modes. For the loop mode (solid line), the achieved bands are 790–870 MHz and 1490–2225 MHz, covering GSM850, GPS, DCS, PCS, and UMTS bands. For the IFA mode (dash dotted line), the achieved bands are 845–980 MHz and 2240–2565 MHz, covering GSM900 and WLAN bands.

Typical loop antennas and their applications in the mobile phones area are systematically discussed in Section 4.4. Three resonant modes, including 0.5-wavelength mode, one-wavelength mode, and 1.5-wavelength mode, are usually utilized. The tuning pad is a good method for impedance matching and with different effects to different modes. The 0.5-wavelength mode is tuned for lower band, and the one-wavelength and 1.5-wavelength modes are tuned together to cover the higher bands. In order to achieve wider bandwidth, a matching bridge is added between the feeding and shorting points of the loop as shown in Figure 4.76. The simulated return loss of the loop mode with and without the matching bridge is illustrated in Figure 4.78. Simulations were made using the Ansoft High Frequency Structure Simulator (HFSS) software. As it may be seen, the inclusion of the matching bridge resulted in improving the bandwidth, thus covering the GSM850, GPS, DCS, PCS, and UMTS bands.

The matching method is shown on the Smith chart of Figure 4.79. The matching bridge works as a shunt inductor at the feeding point of loop. A shunt inductor is able to move the impedance curve along the equal admittance circle. The susceptance introduced by the shunt inductor is $1/j\omega L$, where ω is the angular frequency and L is the equivalent inductance of the matching bridge. For the lower band, the susceptance is larger than that of the high band. As a result, the impedance curve moves further away from the matching center and toward the

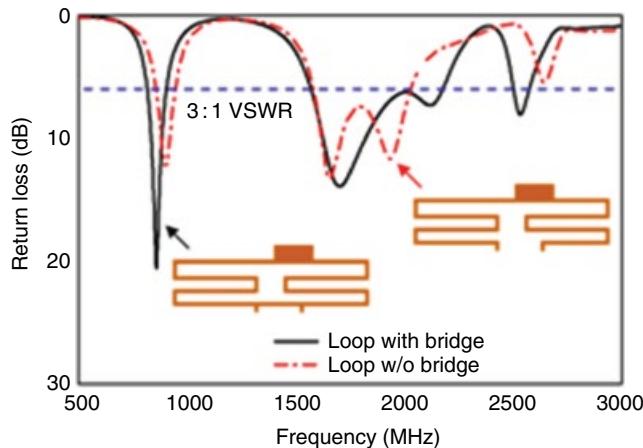


Figure 4.78 Simulated return loss of loop mode of the proposed antenna with or without the matching bridge. (Source: Li *et al.* [31]. Reproduced with permission of IEEE.)

lower frequencies. Similar to the tuning pad, the matching bridge also has different effect that depends on the frequency. For the lower band, shown in Figure 4.79a, the frequency shifts to the lower band with similar bandwidth. However, more band moves into the VSWR 3 : 1 circle as shown in Figure 4.79b, and as a result the bandwidth is enhanced.

In order to cover more bands by the internal antenna in the same volume, another mode is developed based on the loop structure. By cutting the loop, the antenna can be divided into two IFAs with the matching bridge, or two monopoles without the matching bridge. The comparison of return loss between two IFAs and two monopoles is illustrated in Figure 4.80. For the dash dotted line, we can see two resonant frequencies appearing in both lower band and high band, but with unacceptable bandwidth. The narrow bandwidths for two monopoles are mainly due to the compact volume without extra matching structure. For the two IFA results shown in the solid line, one of the two former resonant frequencies has been matched by adding the matching bridge. Therefore, two more bands of GSM950 and WLAN have been achieved based on the matched loop mode.

To demonstrate the validity of the presented matching strategy, the proposed antenna with PIN diode was fabricated and tested, as shown in Figure 4.81a. A detailed diagram of bias circuit of PIN diode is shown in Figure 4.81b. The selected PIN diode is Philips BAP64-03 silicon PIN diode, with good performance up to 3 GHz. When the PIN diode is forward-biased, it works as a series resistance. In the frequency band of 0.5–2.5 GHz, the insertion loss introduced by PIN diode is 0.1–0.2 dB at its typical bias current of 10–100 mA. When the PIN diode is reverse-biased, it is equivalent to a series capacitance of approximately 0.45 pF; the isolation in the required band is better than -15 dB. Therefore, less number of PIN diodes will reduce the insertion loss and improve the performance of the systems. One is the minimum number of diode for two switchable states. In the bias circuit, a capacitor (C_{b1}) is used between the port and the loop antenna for DC blocking; another DC block capacitor (C_{b2}) is used between the feeding and shorting points; an inductor (L_b) is used for RF choking; another capacitor (C_s) is used between V_{contr} and the ground in order to short the

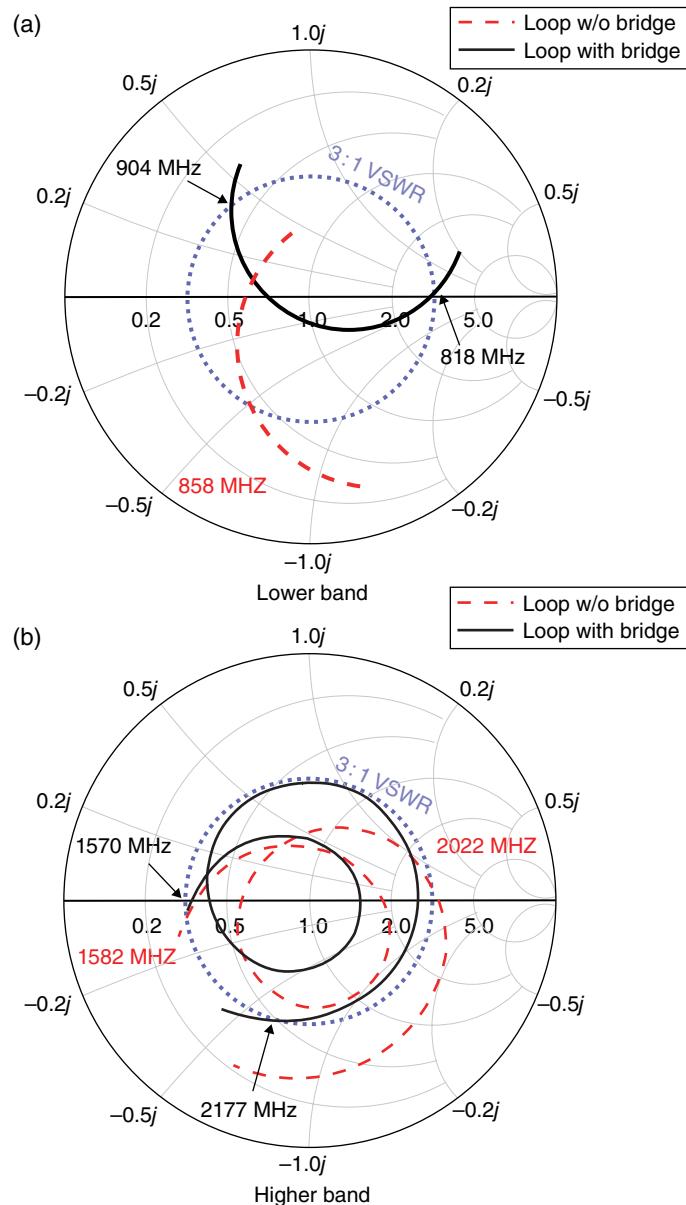


Figure 4.79 Smith chart of the loop mode with or without the matching bridge. (Source: Li et al. [31]. Reproduced with permission of IEEE.)

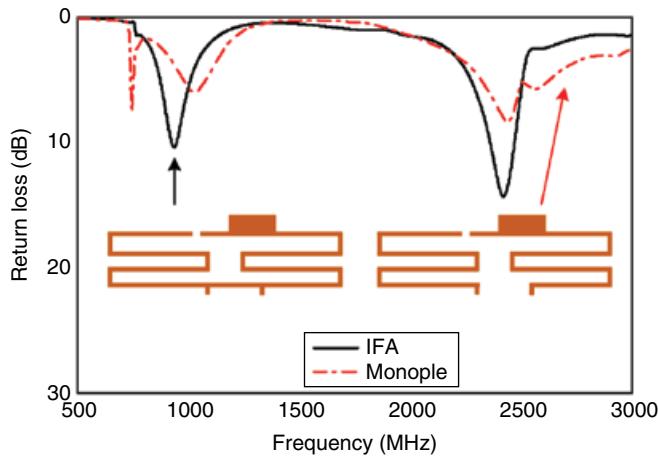


Figure 4.80 Simulated return loss of IFA mode of the proposed antenna with or without the matching bridge. (Source: Li *et al.* [31]. Reproduced with permission of IEEE.)

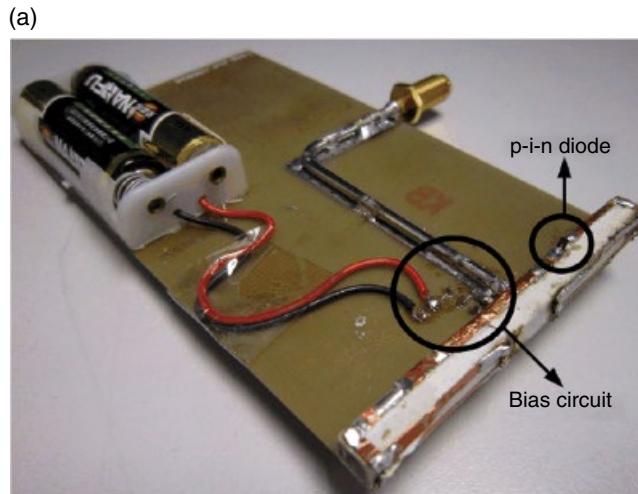


Photo of the proposed antenna with bias circuit

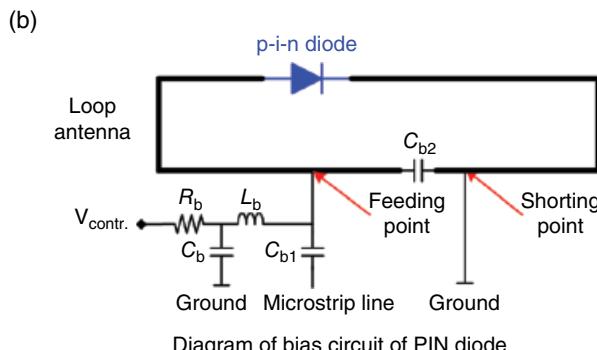


Diagram of bias circuit of PIN diode

Figure 4.81 Prototype. (Source: Li *et al.* [31]. Reproduced with permission of IEEE.)

RF signal leaked from L_b ; a resistor (R_b) is used to control the bias current. The bias voltage for V_{cont} is 3 V, supplied by two AA batteries. The values of each component in the bias circuit are $C_{\text{bl}} = 120 \text{ pF}$, $C_{\text{b2}} = 120 \text{ pF}$, $L_b = 120 \text{ nH}$, $C_s = 470 \text{ pF}$, and $R_b = 46 \Omega$ with bias current is 65 mA. The bias voltage is controlled by a single-pole two-throw switch on the backside of the ground plane.

The measured radiation efficiencies for loop and IFA modes are shown in Figure 4.82a. By combining the curves of two modes, the improvement of efficiency is clearly observed. For the GSM band, the efficiency is better than 64.7%; for the GPS, DCS, PCS, and UMTS bands, the efficiency is better than 47.4%; and for the WLAN band, 62.8% efficiency is achieved. The efficiency can be improved by using high-quality diodes in the practical applications. The gain is also measured as shown in Figure 4.82b. For the GSM band, the gain varies in the range of 0.22–1.10 dBi; for the GPS, DCS, PCS, and UMTS bands, the gain ranges from 0.43 to 3.13 dBi; and for the WLAN band, gain of 2.12–2.41 dBi is achieved. The results indicate the performance improvement by adopting switching mechanism.

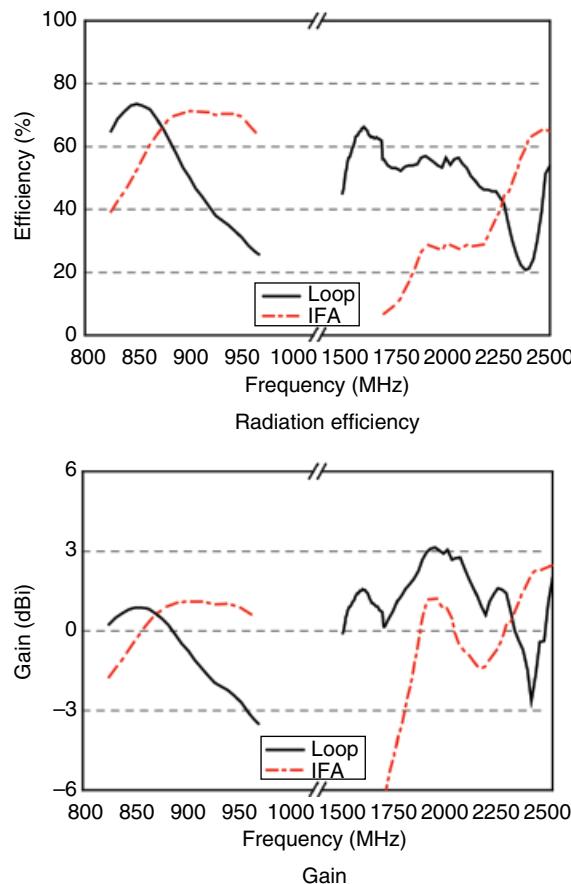


Figure 4.82 Measured results of loop and IFA modes of the proposed antenna. (Source: Li *et al.* [31]. Reproduced with permission of IEEE.)

4.9 MIMO Antennas

4.9.1 Explaining Capacity Boost Effect Through the Antenna Point of View

The term “MIMO” means multiple input and multiple output. It refers to a system which has multiple antennas at the transmitter side and also multiple antennas at the receiver side. Of course, most communication systems are bidirectional, which means that each side has its own transmitters and receivers. In most mobile communication system, the base station has much more space to deploy antennas than mobile terminals; thus, the number of antennas at the base station side is normally more than or at least equal to the number of antennas at the terminal side.

Assume there are N antennas at the base station side and M antennas at the terminal side; this forms an $N \times M$ MIMO system. Based on communication theory, given fixed frequency bandwidth, if a system with one transmitter and one receiver can achieve a communication capacity of C , the maximum capacity an $N \times M$ ($N \geq M$) MIMO system can achieve is $M \times C$. At first glance, this conclusion is contradictory to our instinct. With a given frequency bandwidth, how can we transmit M times information than the channel can support? Why don't those different information streams in the same frequency band interfere with each other? Communication society likes to explain the principle of MIMO system from the point of view of signal processing; the tools they use are the Shannon limit, singular value decomposition (SVD), eigenvector, and so on. If you want to read more, online resource [32] is a very good one; you can also refer to some classic journal papers [33].

In this book, we will try to look at the same problem from a different angle. Let's first look at an $M \times M$ system. For such a system, if we can think out of box, it is not difficult at all to construct a scenario which M transmitters send M timers information in a given frequency bandwidth, as long as there is no constrain on the center frequency of the carrier.

Shown in Figure 4.83 is a 3×3 wireless communication system which can easily triple the communication capacity. It uses three lasers as transmitters and three photodetectors as receivers. We all know that visible or infrared light is also electromagnetic wave, which is essentially the same as radio waves used in mobile communications except the center frequency.

When we look at Figure 4.83, it is so obvious that the 3×3 system can transmit three independent data streams, thus has three times capacities. However, when we think about mobile communications, we instinctively worry about in-band interference between different data streams. The reason is that in cellular frequency bands, which is around GHz, there is no way we can focus a beam as tight as a laser.

Shown in Figure 4.84 is a 3×3 MIMO antenna system in free space. There are three transmitting antennas. The distance between adjacent antennas is half wavelength. There are also three receiving antennas. The distance between transmitting array and receiving array is 10λ . Looking from the transmitting array side, the view angle between two adjacent receivers is 3° . If the distance is longer, which is always true in real-world scenarios, the view angle is even smaller.

Communication society has drawn the conclusion that a MIMO system in free space doesn't work. Let's first introduce their explanation. For a 3×3 MIMO system, the input–output relation can be represented as Equation 4.1.

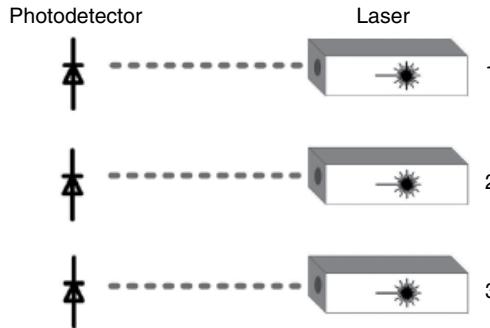


Figure 4.83 A 3×3 MIMO laser system.

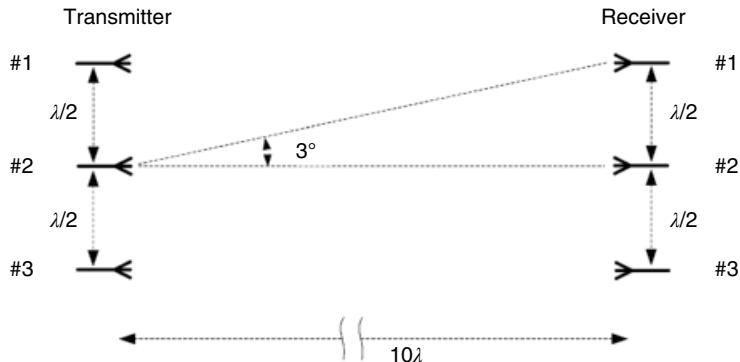


Figure 4.84 A 3×3 MIMO antenna system in free space.

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad (4.1)$$

where x_1 , x_2 , and x_3 are signals transmitted by transmitters 1, 2, and 3, respectively. h_{ij} ($i=1, 2, 3$; $j=1, 2, 3$) is transfer coefficient from transmitter j to receiver i . For each transmitter j , its signal can reach all three receivers through transfer paths h_{1j} , h_{2j} , and h_{3j} , respectively. For each receiver i , it can receive signals from all three transmitters through transfer paths h_{i1} , h_{i2} , and h_{i3} , respectively.

Equation 4.1 can also be rewritten as follows:

$$Y = H \cdot X \quad (4.2)$$

Here, the H matrix is the famous channel matrix. Communication society has proven that the condition of the H matrix decides the channel capacities. For the 3×3 MIMO system in free space, as shown in Figure 4.84, the H matrix is an ill-conditioned one and the total system capacity has little improvement.

Now, let's look at the problem from the antenna point of view. Based on antenna theory, an N -element antenna can generate $N - 1$ nulls. In theory [34], the three-element transmitting array of a 3×3 MIMO system can form a radiation pattern with two nulls. Each null can point to a different receiver. Thus, two receivers out of three will not receive any signals; only the last one in the receiving array can acquire transmitted signal. Similarly, two more radiation patterns with different nulls can be generated. In theory, by using three nulled patterns, three individual signals can be sent to three receivers independently.

Using the scenario shown in Figure 4.84 as an example, to only send data to receiver 2 without interfere with receivers 1 and 3, a radiation pattern with two nulls at $\pm 3^\circ$ is required. The solid line shown in Figure 4.85 is the normalized radiation pattern generated by algorithm given in reference [34]. The radiation pattern has two deep nulls at $\pm 3^\circ$. However, although the radiation pattern can deliver signal toward 0° , the most transmitted power is sent toward $\pm 90^\circ$. The normalized gain at 0° is only -44 dB.

For comparison, the normalized pattern of a 3×1 system, which is optimized to achieve the highest gain at 0° , is also given in Figure 4.85 as the dashed line. Its normalized gain at 0° is 0 dB.

Using the nulled pattern method, although we can send independent information to individual receivers, the overall system capacity is not increased. To understand that, let's look at the maximum channel capacity or the Shannon limit given in Equation 4.3,

$$C = B \cdot \log_2 (1 + \text{SNR}) \quad (4.3)$$

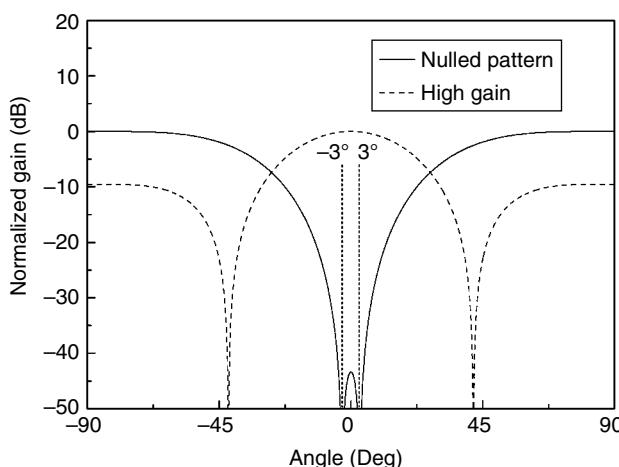


Figure 4.85 Send independent signal to individual receiver by using nulled pattern.

where C is the maximum capacity, B is the occupied bandwidth, and SNR is signal-to-noise ratio (in linear scale, not in dB).

As shown in Figure 4.85, to generate two nulls, the normalized gain at 0° drops from 0 to -44 dB. Because the signal strength is linear to antenna gain, the SNR in Equation 4.4 will decrease accordingly. If compared with a 3×1 communication system, which doesn't have nulling, the capacity of a nulled system is actually much less. Even considering the capability of sending three parallel data streams in a nulled system, the total capacity has little advantage over a 3×1 system.

Communication society has also proven that in a multipath-rich environment, the H matrix of MIMO system is well conditioned and a 3×3 MIMO can improve the overall capacity close to three folds. Next, we will explain that from the antenna point of view. Shown in Figure 4.86 is a multipath environment. There are two reflecting objects, one at the top and the other at the bottom. There are three independent propagation paths. The line-of-sight (LOS) path is a direct path from transmitting array to receiver array. Multipath 1 leaves transmitting array with a direction-of-departure (DOD) angle of -30° . It goes by the top reflecting object and arrives at the receiving array with a direction-of-arrival (DOA) angle of -30° . Multipath 2 has a DOD angle of 30° and DOA angle of 30° . In this scenario, the distance between the transmitting array and the receiver array is not important, because DODs and ODAs are decided by the position of reflecting objects.

Nulled patterns are used in both transmitting and receiving sides. To independently send information through multipath 1, a nulled pattern is generated at the transmitting side. Two nulls correspond to the DOD angle of 0° and 30° , respectively. The nulled pattern used at the receiving side also has two nulls, which correspond to the DOA angle of 0° and 30° , respectively.

Shown in Figure 4.87 are transmitting and receiving patterns for multipath 1. The patterns are shown in linear polar coordinates. It is assumed that the radiation pattern of each antenna element is omnidirectional. Both the transmitting array and the receiving array are linear array, so both radiation patterns are symmetrical along the array axis. Due to the existence of reflecting objects, the DOD and DOA angles of the desired travel path can be well separated

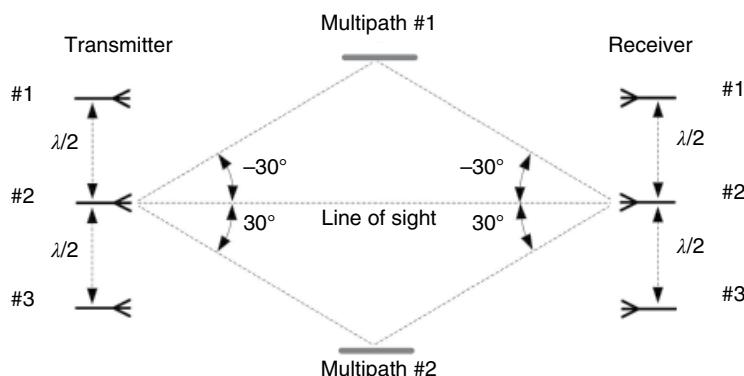


Figure 4.86 MIMO system in a multipath environment.

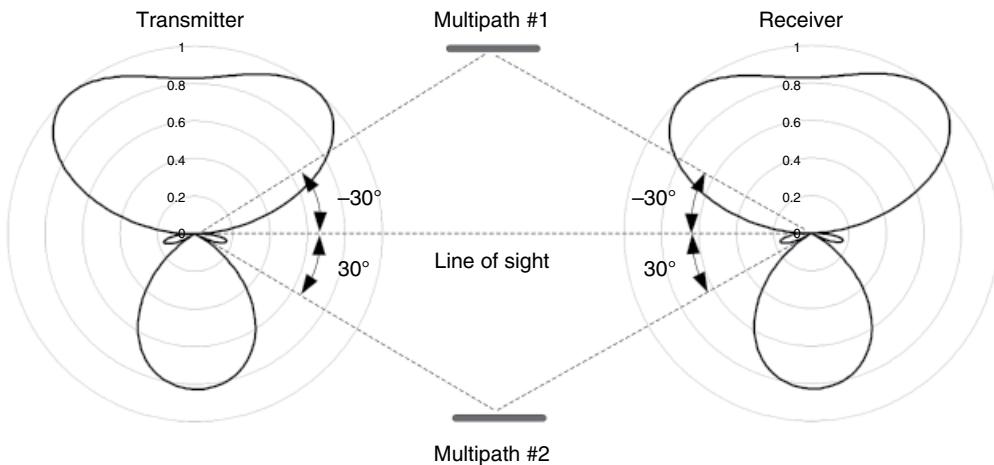


Figure 4.87 Nulled patterns corresponding to the multipath 1.

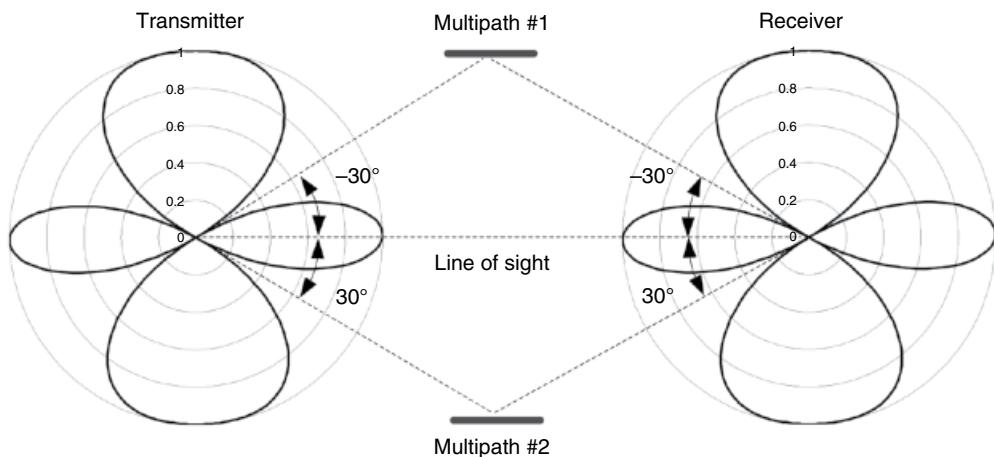


Figure 4.88 Nulled patterns corresponding to the LOS path.

from other two interfering paths. The nulled pattern still has reasonable gain at signal's direction while forms nulls to suppress other two interferences.

Shown in Figure 4.88 are transmitting and receiving patterns for the line of sight path. Similarly, because the wide separation between desired path and interfering paths, acceptable gain has been achieved at the desired angle.

Now it should be clear what the multipath-rich environment means to the communication society and to the antenna society. In a multipath-rich environment, there must be multiple widely separated traveling paths. From the signal processing point of view, these paths are statistical independent; thus, the H matrix is well conditioned and the overall capacity is

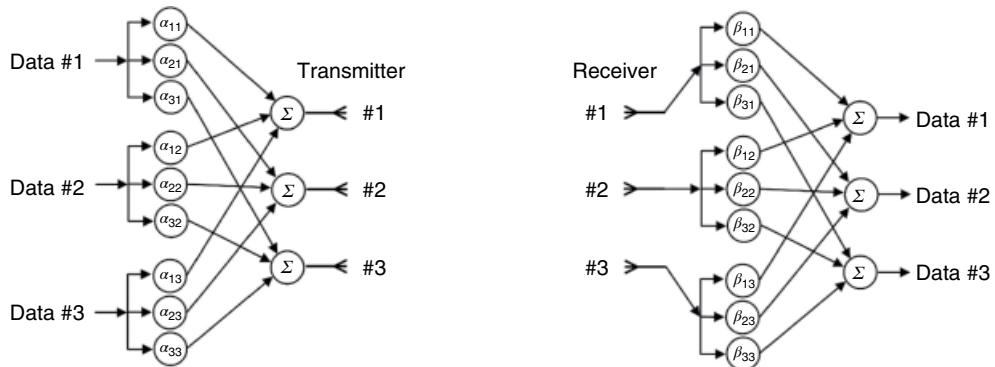


Figure 4.89 Explain MIMO system through the view of phased array.

improved. For our antenna society, widely separated paths allow antenna array to form nulled patterns to suppress unwanted paths while still achieving reasonable gain at the desired path. By using nulled patterns, several data stream can be independently transmitted simultaneously, thus significantly increasing the overall capacity.

One might wonder how one antenna array can transfer several different data streams simultaneously. Shown in Figure 4.89 is a conceptual explanation through the view of phased array. It is a 3×3 system. On the left side is the transmitting array. Data stream 1 is multiplied by weightings α_{11} , α_{21} , and α_{31} , and sent to transmitting antennas 1, 2, and 3, respectively. The weightings used here are the same as those used in phased arrays. Each weighting can modify both amplitude and phase of the original signal. With weightings α_{11} , α_{21} , and α_{31} , the transmitting array generates a radiation pattern as shown in Figure 4.87. Data stream 1 exists in the direction of multipath 1 and is suppressed in other two directions. Data streams 2 and 3 have their own sets of weights; two more radiation patterns can be generated.

On the receiver side, different sets of weightings are applied to signals received on antennas 1, 2, and 3. With weightings β_{11} , β_{12} , and β_{13} , a receiving pattern as shown in Figure 4.87 can be generated. Only data stream 1 in multipath 1 is received. Signals in the LOS and multipath 2 are suppressed. Similarly, data streams 2 and 3 can be individually picked up.

In traditional phased arrays, the waveforms in all transmitting antennas are almost identical. The only difference among signals on transmitting antennas is their amplitude and phase. However in a MIMO system, signal waveforms on different antennas are different. Each waveform is a combination of three distinct data streams, each with a different weighting.

On the receiver side of traditional phased arrays, each antenna receives multiple copies of the original signal. The waveform of combined signals is pretty much the same as the original one, except for its amplitude and phase. For a MIMO system, each receiving antenna can collect all three transmitted waveforms. On each receiving antenna, these three waveforms are combined again with different phase and amplitude. It is obvious that waveforms on all receiving antennas are also different.

Finally, let's discuss how communication society sends multiple data streams still using a 3×3 system as an example. Based on Equation 4.1, it is clear that even if three data streams are separately sent to dedicated antennas, three data streams will all mix together at the receiver side. The way communication society handles this is through SVD method. Assuming the transfer matrix H is known, the H can be decomposed into three matrixes as follows:

$$H = U \cdot D \cdot V^H \quad (4.4)$$

Here U , D , and V^H are all 3×3 matrixes. $(\cdot)^H$ is the conjugate transpose operator. Both U and V^H are unity matrixes. D is a diagonal matrix,

$$D = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \quad (4.5)$$

To successfully send three discrete data streams, two steps are required. First, before sending three data streams to antennas, they are left multiplied by matrix V . Second, before processing output data streams, all signals received by three antennas are left multiplied by matrix U^H . The process can be expressed as follows:

$$y = U^H \cdot (H \cdot (V \cdot x)) \quad (4.6)$$

Substituting Equation 4.4 into Equation 4.6, we get

$$y = U^H \cdot U \cdot \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \cdot V^H \cdot V \cdot x \quad (4.7)$$

Because $U^H U$ and $V^H V$ are both identity matrixes, we have

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad (4.8)$$

Equation 4.8 means each data stream x_i can be independently sent to the receiver i . By comparing Figure 4.89 and Equation 4.6, it is clear that V and U^H in Equation 4.6 are in fact the transmitting and the receiving weighting matrixes of the 3×3 phased array shown in Figure 4.89.

$$V = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{bmatrix} \quad (4.9)$$

$$U^H = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{bmatrix} \quad (4.10)$$

4.9.2 Antenna Correlation and Antenna Isolation

In Section 4.9.1, we have discussed why MIMO system can boost system capacity. In fact, the H matrix shown in Equation 4.1 includes contributions from two parts. Besides propagation environment, there are also contributions from antenna array itself.

Let's examine one extreme case, when three identical transmitting antennas are placed in the same spot. In this scenario, the H matrix degenerates into Equation 4.11. The 3×3 system degenerates into a 1×3 system.

$$H = \begin{bmatrix} h_{11} & h_{11} & h_{11} \\ h_{21} & h_{21} & h_{21} \\ h_{31} & h_{31} & h_{31} \end{bmatrix} \quad (4.11)$$

To quantitatively evaluate the performance of an antenna array in a MIMO system, antenna correlation was introduced [35, 36]. In the following discussion, to simplify the problem it is assumed that multipath environment is isotropic in the sense of both power density and polarizations. Under this assumption, the envelope correlation between antenna 1 and antenna 2 is given in Equation 4.12.

$$\rho_e^{12} = |\rho|^2 = \frac{R_{12}^2}{\sigma_1^2 \cdot \sigma_2^2} \quad (4.12)$$

where R_{12} , σ_1^2 , and σ_2^2 are given in Equations 4.13–4.15.

$$R_{12} = \int_0^{2\pi} \int_0^\pi \left(E_{\theta 1}(\theta, \varphi) \cdot E_{\theta 2}^*(\theta, \varphi) + E_{\varphi 1}(\theta, \varphi) \cdot E_{\varphi 2}^*(\theta, \varphi) \right) \sin \theta \cdot d\theta \cdot d\varphi \quad (4.13)$$

$$\sigma_1^2 = \int_0^{2\pi} \int_0^\pi \left(E_{\theta 1}(\theta, \varphi) \cdot E_{\theta 1}^*(\theta, \varphi) + E_{\varphi 1}(\theta, \varphi) \cdot E_{\varphi 1}^*(\theta, \varphi) \right) \sin \theta \cdot d\theta \cdot d\varphi \quad (4.14)$$

$$\sigma_2^2 = \int_0^{2\pi} \int_0^\pi \left(E_{\theta 2}(\theta, \varphi) \cdot E_{\theta 2}^*(\theta, \varphi) + E_{\varphi 2}(\theta, \varphi) \cdot E_{\varphi 2}^*(\theta, \varphi) \right) \sin \theta \cdot d\theta \cdot d\varphi \quad (4.15)$$

Here, $(\cdot)^*$ is the conjugate operator. In a spherical coordinate system, there are three orthogonal vectors, \hat{r} , $\hat{\theta}$, and $\hat{\varphi}$ respectively. In the far field, all antennas radiate spherical electromagnetic wave, which travels along \hat{r} direction. Based on the Poynting theorem [8, 9], the electronic field can only have $\hat{\theta}$ and $\hat{\varphi}$ components. In Equations 4.13–4.15, $E_{\theta i}(\theta, \varphi)$ and $E_{\varphi i}(\theta, \varphi)$ ($i=1, 2$) are the 3D radiation patterns of the $\hat{\theta}$ and $\hat{\varphi}$ components of the i th antenna. In practice, 3D radiation patterns are measured by a 3D antenna anechoic chamber. Detailed information about 3D chamber can be found in Section 5.1.3.

For the extreme case described in Equation 4.11, correlations among three antennas, ρ_e^{12} , ρ_e^{13} , and ρ_e^{23} , are all equal to 1. That means three antennas are totally correlated. One important goal of MIMO array design is to minimize correlations among antenna elements in an array.

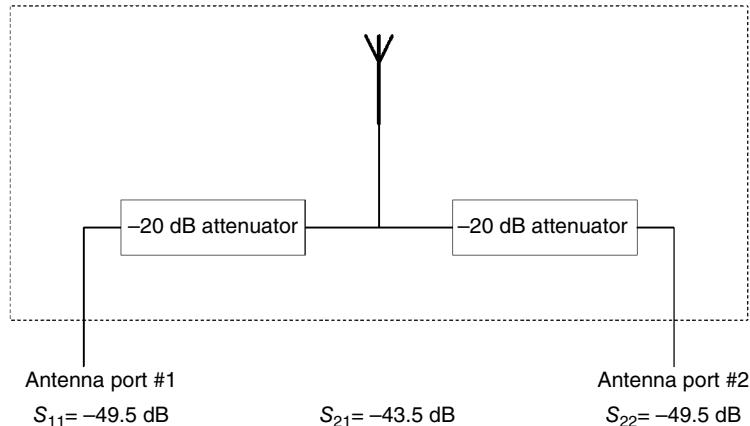


Figure 4.90 A “perfect” dual-port MIMO antenna.

In some published papers, the envelope correlation between any two antennas is calculated through S parameters, as given in Equation 4.16.

$$\rho_e = \frac{\left| S_{11}^* S_{12} + S_{21}^* S_{22} \right|^2}{\left(1 - \left(|S_{11}|^2 + |S_{21}|^2 \right) \right) \left(1 - \left(|S_{22}|^2 + |S_{12}|^2 \right) \right)} \quad (4.16)$$

Detailed deduction can be found in the reference [37]. When an MIMO system is **lossless**, Equations 4.12 and 4.16 are equivalent. Based on Equation 4.16, it is clear that the better the isolation, the lower the envelope correlation. When S_{12} and S_{21} are 0, the ρ_e is also 0.

When discussing an MIMO array, people often don't distinguish between antenna correlation and antenna isolation. However, these two are not always proportional related. Shown in Figure 4.90 is an extreme example.

In Figure 4.90, inside the dashed-line rectangle is a “well-matched and high isolated dual-port antenna.” If measuring the antenna with a network analyzer and using Equation 4.16 to calculate its correlation, the antenna seems a perfect dual-port MIMO antenna. Both S_{11} and S_{22} are -49.5 dB . The isolation between two ports is -43.5 dB . Using Equation 4.16, the correlation ρ_e between two ports is almost 0.

However, if we look into the dashed-line rectangle, there is only one antenna. The antenna is connected to port 1 and port 2 through two 20 dB attenuators, respectively. Because there is only one antenna, signals received from port 1 and port 2 are identical. Obviously, 3D radiation patterns of both ports, which are measured in an antenna chamber, are also identical. Based on Equation 4.12, the correlation ρ_e between two ports is 1.

In this example, the wrong result given by Equation 4.16 is not its own fault. The precondition of using Equation 4.16 is that a multiport system under evaluation must be lossless. In antenna designs for real products, antenna efficiency is normally less than 50% or -3 dB . That means Equation 4.16 is always unreliable. We should measure radiation patterns in a 3D chamber and use Equation 4.12 to evaluate correlations between antennas.

4.9.3 Improve Isolation Between Antennas

In a MIMO system, there are mainly two problems which are related to poor isolation. First, as has been discussed in Section 4.9.2, poor isolation means strong correlation between antennas, which decreases the overall system capacity. The whole purpose of putting multiple antennas in a MIMO system is to increase capacity. Second, poor isolation also means low antenna efficiency, because some portion of output power from an antenna will be absorbed by nearby antennas.

The easiest way to improve isolation is increasing distance between antennas. However, it isn't always practical when designing mobile devices. To achieve high isolation between antennas in devices with constrained space, several methods have been proposed. Some of them can be used when there is enough freedom on select antennas' form factor. Antennas with different radiation patterns can achieve very good isolation even when they are placed together [38]. By exciting orthogonal modes on a single radiator, two antennas with high isolation can be realized [39]. Electric dipole and magnetic dipoles can also be placed together to generate independent omnidirectional patterns with orthogonal polarization [40, 41].

In many mobile devices, due to thickness constraints and chipset arrangement, there are little freedom on selecting antenna locations. Thus we need some handy technique to suppress mutual coupling between antennas. Most isolation improvement ideas fall into two categories. One is blocking leaking signal and the other is canceling out leaking signal by introducing a new signal path.

Shown in Figure 4.91a is a device with two closely placed monopole antennas. Both antennas are identical and resonate at around 3 GHz. The antenna length is 20 mm, which is shorter than a quarter wavelength; this phenomenon has been discussed in Section 3.1.1.

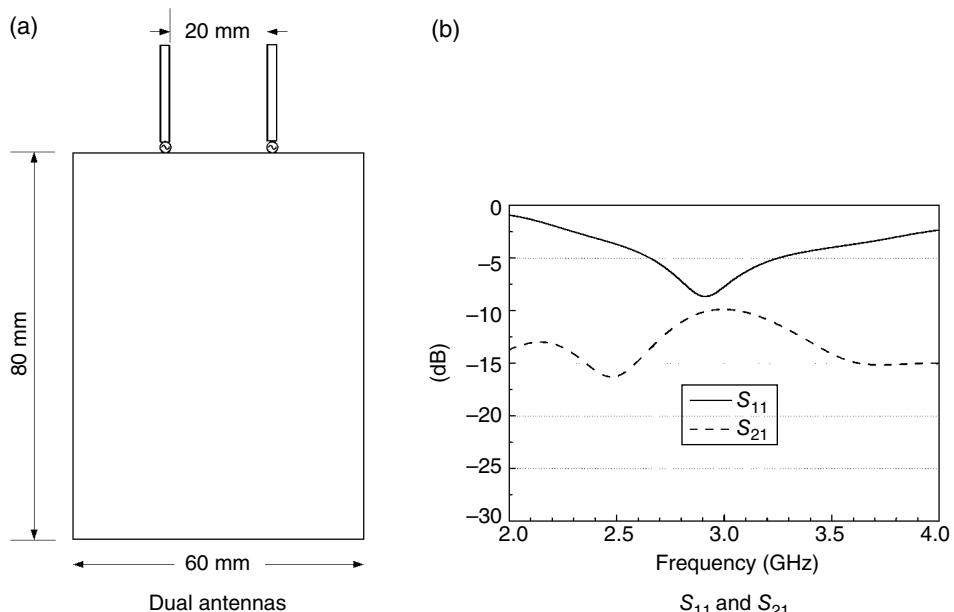


Figure 4.91 Two closely placed monopole antennas.

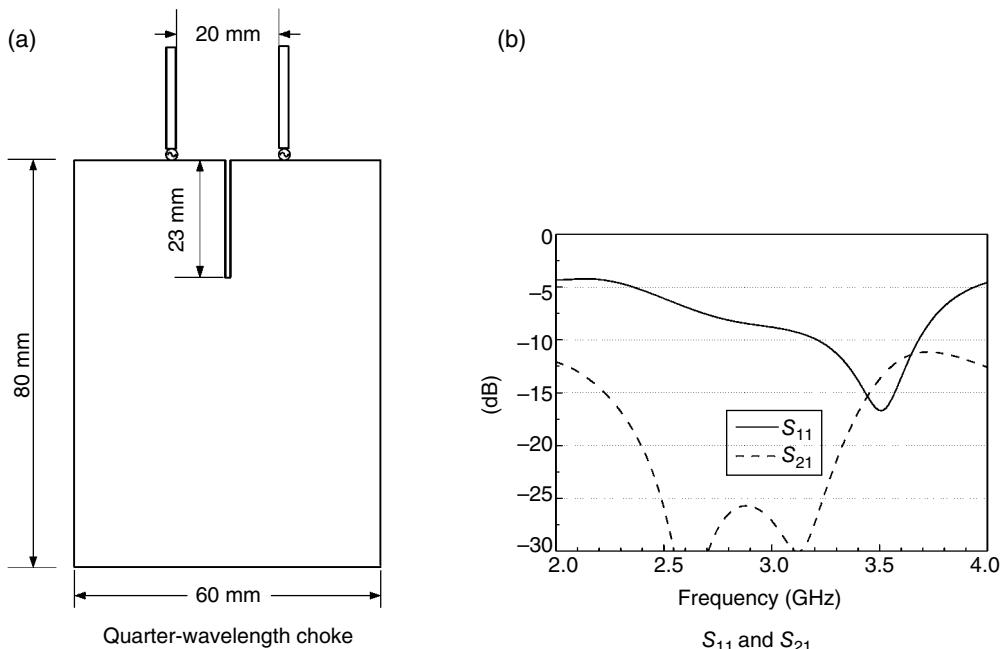


Figure 4.92 Using choke to improve isolation.

Because isolation is the focus here, no matching network is applied to antennas. The simulated S parameters have been shown in Figure 4.91b. The isolation between antennas, which is S_{21} , is around -10 dB .

One way to improve isolation is using choke. By adding a $1\text{ mm} \times 23\text{ mm}$ slot, as shown in Figure 4.92a, the isolation between two antennas, as shown in Figure 4.92b, has been improved from -10 dB to better than -25 dB . The slot functions as a quarter-wavelength choke, which stops current leaking from one to another. Again, the slot's length is actually 23 mm instead of the theoretical 25 mm . As has been discussed in Section 3.1.1, the ground plane is also part of the radiation structure. The added slot in the ground impacts not only isolation but also antenna matching. This impact can be easily compensated by various antenna matching techniques discussed in Chapter 2.

The other way to improve isolation is to use a neutral line. The physics behind this is out-of-phase cancellation. Instead of blocking leaking current, which is what a choke does, a neutral line introduces a new transmission path. Shown in Figure 4.93a is a conceptual diagram. Inside the dashed-line rectangle is the original two-port antenna and there is around -10 dB mutual coupling between two ports. To mitigate mutual coupling, two 50Ω transmission lines, each has an electrical length of 26° , are added to the antenna port. Another 123Ω “neutral” transmission line with electrical length of 241° is used to connect ports 1 and 2.

Although the characteristic impedance of two 26° long lines is 50Ω , they do have impact on both isolation and matching. Signal passing through the new path has the same amplitude as the original mutual coupling signal, but it is out of phase. As shown in Figure 4.93b, due to the cancellation effect of the out-of-phase signal, the isolation between two ports has been significantly improved.

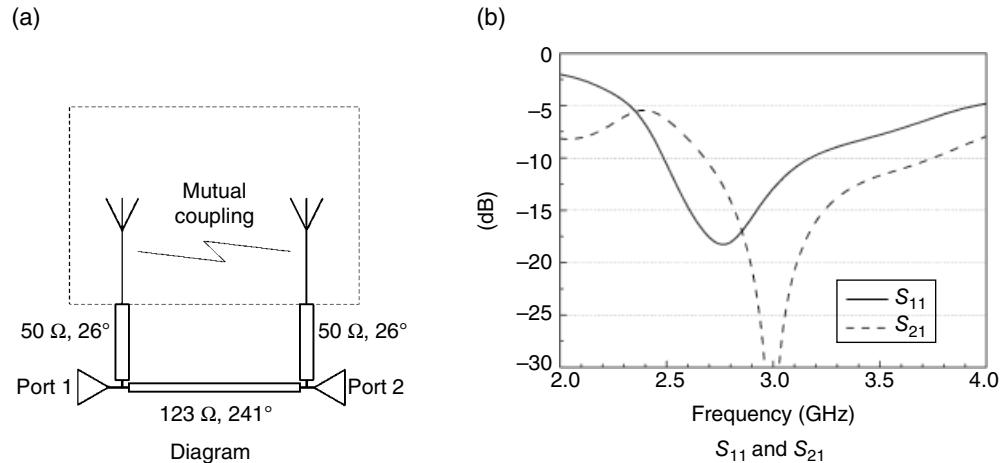


Figure 4.93 Using neutral line to improve isolation.

Figure 4.93 is only a simplified demonstration. In a real design, the neutral line can take any form factor. Shown in Figure 4.94a is such an example [42]. The reverse U-shaped line between two antennas functions as a neutral line. By adopting the neutral line, the isolation in 1.6–2.3 GHz band is improved by around 5 dB.

4.10 Antennas in Recently Released Phones

When writing the first edition of the book in 2010, I did want to tear down some phones to explain how antennas look like in the real world. Sadly, I didn't get approval from any phone manufacture. Since then, 5 years has passed and the world's phone ecosystem has evolved a great deal. Most dominating players then, such as Nokia, Black Berry, and so on, have passed their prime time.

Fortunately, this time Xiaomi Inc. has approved our request to use their phones as examples. Xiaomi Inc. was funded in April 2010 and released their first phone in August 2011. In the second quarter of 2015, Xiaomi ranked the first in China market. It took 18% market share. In comparison, Huawei and Apple claimed 16 and 12%, respectively, of the Chinese market.

Most companies have several models on the market simultaneously. Some of phones are targeting entry-level users. Others are flagship phones, which are marketing to less price-sensitive buyers. From antenna point of view, those phones do have some different features, which can have impact on antenna designing. Two phones, Hongmi 2A and Xiaomi 4, are tore down in the following sections.

4.10.1 Entry-Level Phone

Shown in Figure 4.95 are front, side, and back views of Hongmi 2A. The price of this phone is 549 RMB, which is around 87 USD, when purchased without a contract. Hongmi 2A was released in March, 2015. It has a 4.7 inch 1280×720 IPS screen, 2GB ram, and 16GB flash memory.

The phone supports multiple cellular standards, such as TD-LTE (4G), TD-SCDMA (3G), and GSM (2G). TD-LTE and TD-SCDMA are 4G and 3G standards, respectively. China Mobile,

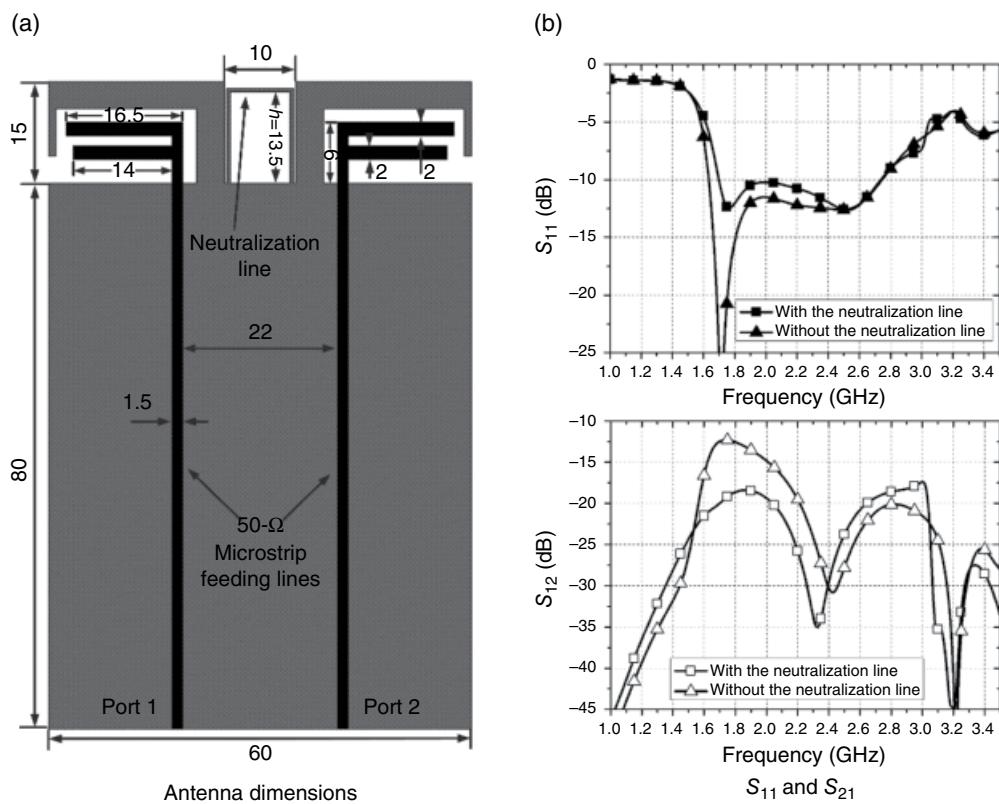


Figure 4.94 A wideband dual antenna. (*Source:* Wang and Du [42]. Reproduced with permission of IEEE.)



Figure 4.95 Hongmi 2A. (*Source:* Reproduced with permission of Xiaomi, Inc.)



Figure 4.96 Hongmi 2A with back cover detached. (Source: Reproduced with permission of Xiaomi, Inc.)

which has 816 million users and around 70% Chinese market, is the leading adopter of those two standards. The phone also supports 2.4GHz WLAN (802.11b/g/n), Bluetooth, and GPS.

As a low-cost phone, the back cover and main frame of Hongmi 2A are all made of plastic. The back cover of Hongmi 2A is removable and its battery is exchangeable. There is a strange phenomenon. Before the birth of iPhone in 2007, every phone had a removable back cover and an exchangeable battery. When iPhone came out, unchangeable battery is a “big” bug people liked to laugh at. I am not sure since when most high-end phones all adopted unchangeable battery. In 2015, only low-cost phones had exchangeable batteries.

Because structure parts of low-cost phones are mostly made of plastic, it is actually easier to design antennas for those phones. High-end phones like to use metal frames or even metal back covers, which make antenna designs more challenging.

Shown in Figure 4.96 is a Hongmi 2A with back cover detached. Because users have access to the main plastic frame, to disguise funny parts from curious users, all three antennas are covered with black solder mask. The positions of three antennas are marked by dashed arrow lines. However, the black solder mask does a pretty good job, not too much can be seen from this photo.

There are three antennas on a Hongmi 2A. All three antennas are manufactured by using flexible print circuit (FPC) technologies, which has been discussed in Sections 3.3 and Section 4.2.5. The Hongmi 2A supports the following cellular bands:

4G TDD-LTE	B38 (2570–2620 MHz) B39 (1880–1920 MHz) B40 (2300–2400 MHz) B41 (2555–2655 MHz)
3G TD-SCDMA	B34 (2010–2025 MHz) B39 (1880–1920 MHz)
2G GSM bands	B2 (1850–1990 MHz) B3 (1710–1880 MHz) B8 (880–960 MHz)

The primary cellular antenna supports all 2G, 3G, and 4G bands. The secondary cellular antenna supports only 3G and 4G bands. As the 2G B8 band is the lowest frequency band, the primary cellular antenna is the biggest one among three antennas. At 4G bands, the primary and secondary antennas work together to form a MIMO system. Based on 3GPP specification 36.306 [43], devices belong to category 2–4 have two antennas. The Hongmi 2A is a category 4 device.

Shown in Figure 4.97a is the bottom view of the antennas. There are three sets of gold plated contacts. When a phone is put together, there is always some assemble variations. To mitigate that, three sets of spring fingers, as shown in Figure 4.97b, are used in Hongmi 2A. The primary antenna locates at the bottom of the phone. The cellular chipsets are at the top portion of the phone. A thin coaxial cable is used to transfer signal from the top board to the bottom board.

Shown in Figure 4.98a is a production primary antenna. It has not been attached to a phone's main frame yet, so it is still a flat piece. Only two gold plated contacts can be clearly identified in Figure 4.98a. Most other features are masked by a layer of black paint.

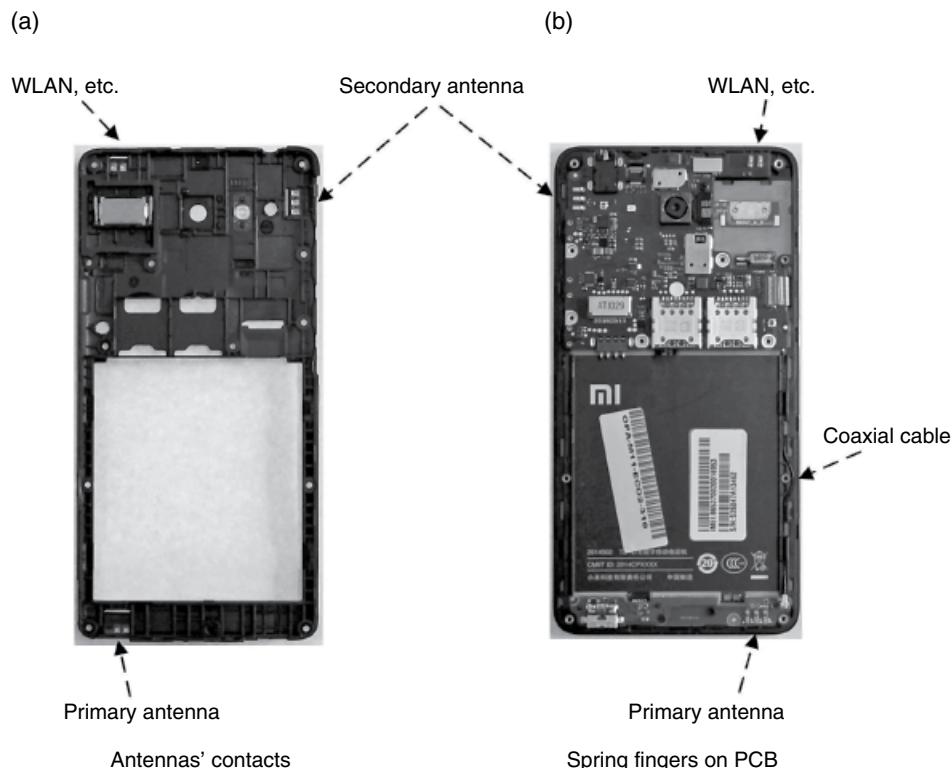


Figure 4.97 Antennas contacts on the back and spring fingers on PCB. (Source: Reproduced with permission of Xiaomi, Inc.)

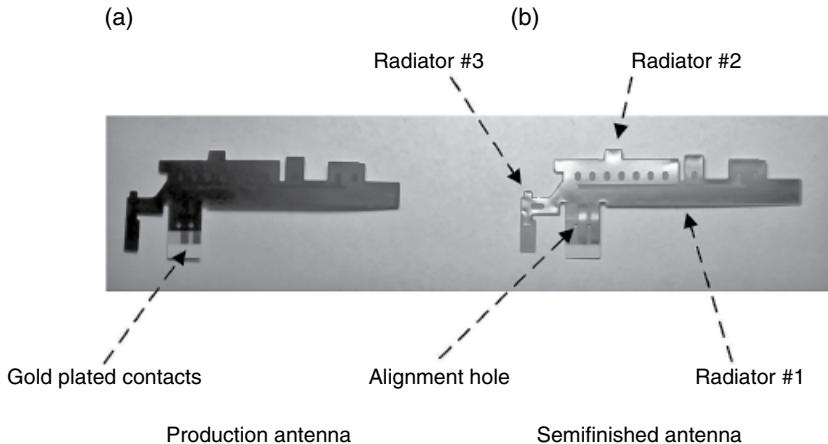


Figure 4.98 Hongmi 2A's primary antenna. (*Source:* Reproduced with permission of Xiaomi, Inc. and Shanghai Amphenol Airwave, Inc.)

To give us a better look at the antenna, a semifinished sample, which is kindly provided by friends at Amphenol, has been shown in Figure 4.98b. It is clear that the antenna adopts a multiband IFA design. There are three radiators. Radiator 1 has a folded trace and is the longest. Radiator 1 surely radiates at the 880–960 MHz and can also radiate at higher band with its higher order modes. Radiators 2 and 3 correspond to higher bands.

There are two alignment holes next to gold plated contacts. As contacts are where an antenna makes connection to spring fingers, assembly tolerance must be well controlled. Alignment posts on phone frames and alignment holes on flexes are commonly used feature.

Shown in Figure 4.99 is a semifinished primary antenna, which has been installed on a plastic frame. Radiators 1, 2, and 3 are all marked out. There are several eclipse-shaped holes along the folded edges of the antenna. They are stress-release holes. They have little impact on antenna performance but make bending antenna much easier. At the end of radiator 3, some portion of the flex doesn't have any copper. It is called "adhesive tab." As radiator 3 is quite small, this part of flex tends to raise up even with a stress-release hole. By adding some extra adhesive area, this portion can be better secured.

It is always a good practice to leave some copper-free area on a flex antenna. Because whenever the resonant frequency of an antenna needs to be tuned lower, those reserved area will be much appreciated. Asking for more antenna area in late stage of any project is always a huge headache for everyone.

Although the antenna occupies a 3D volume, it is not a true 3D antenna. To be manufactured by flex technologies, it must be able to be spread into a 2D surface. To accommodate this requirement, the plastic frame is recessed and the antenna is also designed as an irregular shape. As the volume of an antenna is proportional to its performance, a flex antenna has to scarify in some degree.

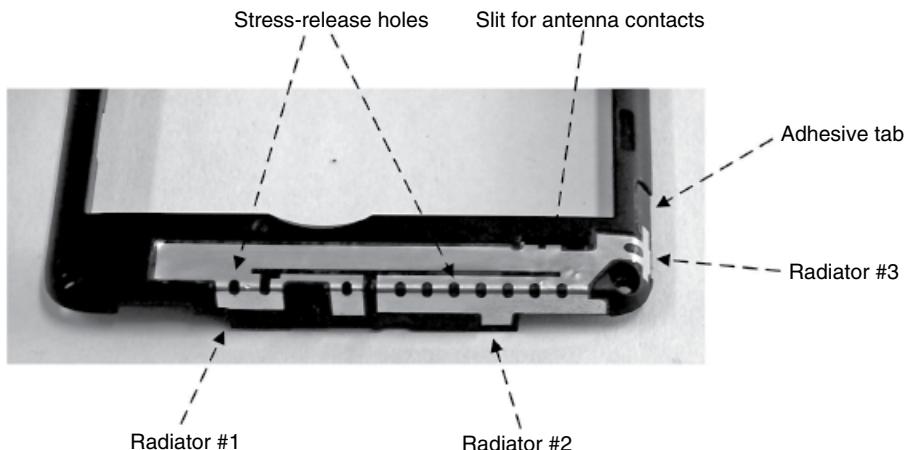


Figure 4.99 Semifinished primary antenna installed on a phone. (Source: Reproduced with permission of Xiaomi, Inc. and Shanghai Amphenol Airwave, Inc.)

The antenna makes galvanic contact with the PCB on the back of the plastic frame. The portion of flex, where antenna contacts are, passes through a slit and folds back to stick itself onto the back side of the plastic frame.

Shown in Figure 4.100a is a zoom-in photo of the matching circuit of the primary antenna. There are three spring fingers on the PCB. Spring fingers 1 and 3 are all connected to ground directly. Spring finger 2 is connected to the matching circuit.

When looking at Figure 4.98 or 4.99, it is clear that there are only two contacts on the antenna. Actually, spring finger 3 is a redundant one. One possible guess is that in some stages in the development, a parasitic element has been considered and later abandoned. The slit on the frame and contact area on the flex are all reserved some space for this dummy finger. It is also a good practice to leave the redundant area as it is. One never knows whether those features will be needed again later on.

The matching network adopts a dual-band matching. Four matching components are used. C1 and L1 take care of the lower band. C2 and L2 handle the higher band. Detailed information of various considerations on dual-band matching can be found in Section 2.3.

Shown in Figure 4.101 are photos of the secondary antenna. Figure 4.101a is a production antenna. Figure 4.101b is a semifinished antenna, which is provided for a better view of different radiators. The antenna is also a multiband IFA. The main IFA has two radiators: 1 and 2. Radiator 3 is a parasitic element, which is shorted to ground directly. A detailed discussion of parasitic element can be found in Section 4.2.4. The secondary antenna only needs to cover frequency bands between 1880 and 2655 MHz; thus, its dimension is much smaller than the primary antenna.

Shown in Figure 4.102 is a semifinished secondary antenna, which has been installed on the top-left corner of a plastic frame. Some portion of the flex, which has three gold plated contacts, is passed through the slit on the plastic frame and wrapped to its back.

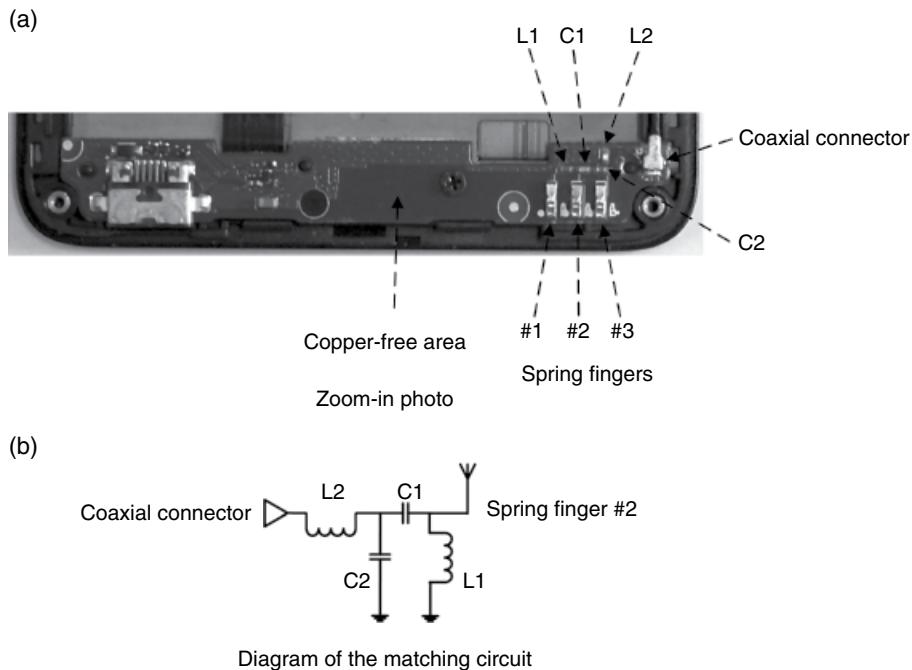


Figure 4.100 Matching circuit of the primary antenna. (Source: Reproduced with permission of Xiaomi, Inc.)

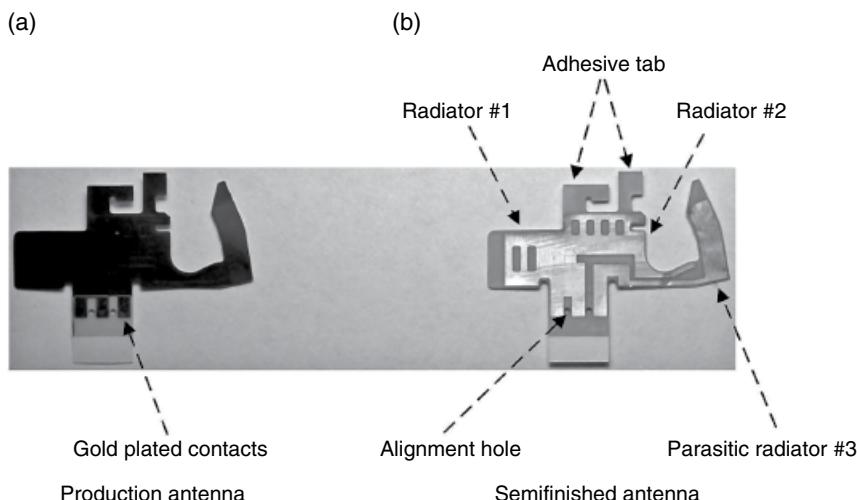


Figure 4.101 Hongmi 2A's secondary antenna. (Source: Reproduced with permission of Xiaomi, Inc. and Shanghai Amphenol Airwave, Inc.)

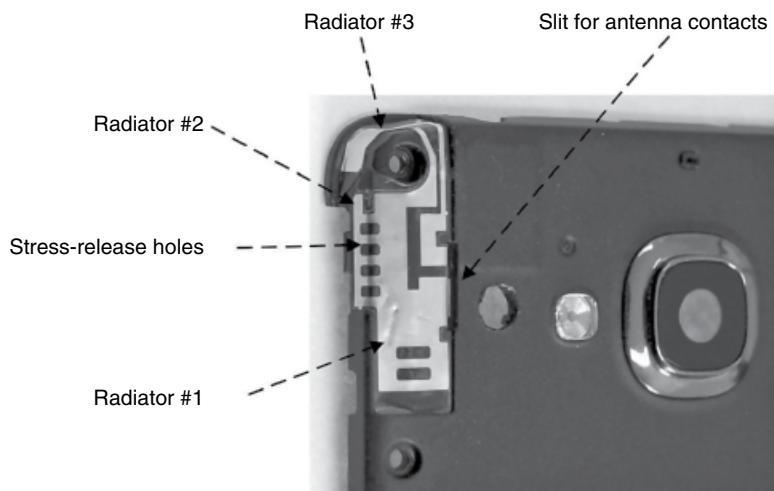


Figure 4.102 Semifinished secondary antenna installed on a phone. (Source: Reproduced with permission of Xiaomi, Inc. and Shanghai Amphenol Airwave, Inc.)

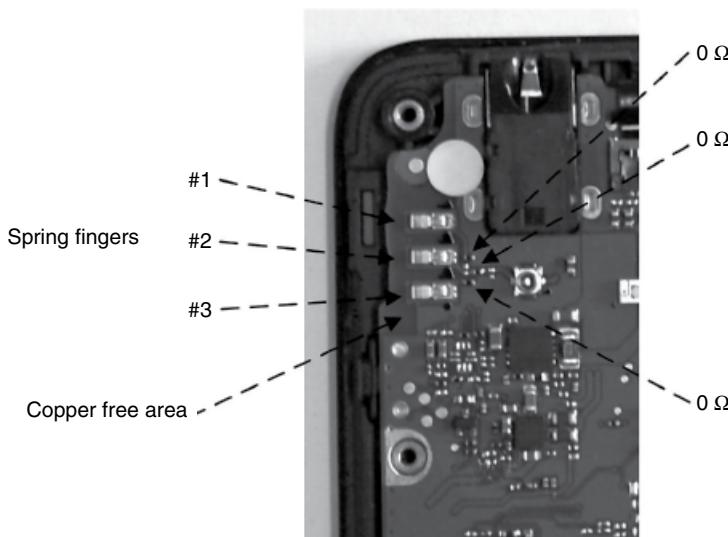


Figure 4.103 Matching circuit of the secondary antenna. (Source: Reproduced with permission of Xiaomi, Inc.)

Shown in Figure 4.103 is a zoom-in photo of the matching circuit of the secondary antenna. There are three spring fingers on the PCB. Spring finger 1 connects to the parasitic element. Spring fingers 2 and 3 connect to the IFA. Both spring fingers 1 and 3 are shorted to ground through $0\ \Omega$ resistors. Those two $0\ \Omega$ resistors are not necessary in the design and spring fingers can be grounded directly. However, those resistors can be replaced by inductors or capacitors to fine-tune antenna, which might be helpful. Using the parasitic element as an example,

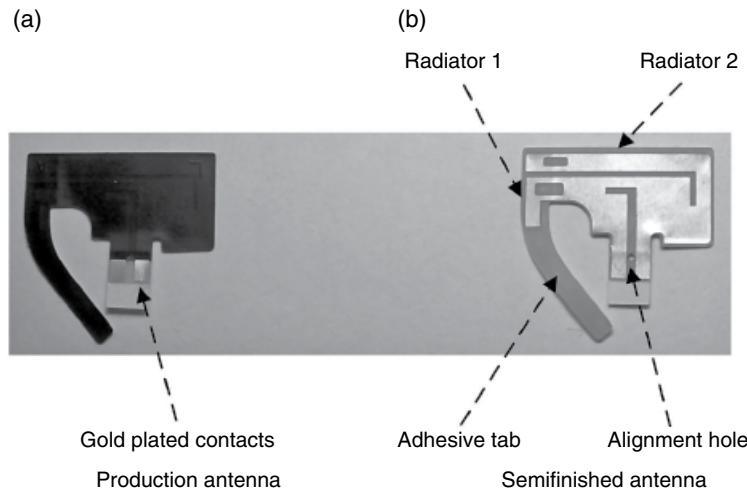


Figure 4.104 Hongmi 2A's WLAN and GPS antenna. (Source: Reproduced with permission of Xiaomi, Inc. and Shanghai Amphenol Airwave, Inc.)

by replacing the 0Ω resistor connected to spring finger 1 with an inductor, the element's resonant frequency can be tuned lower. An inductor with larger value shifts frequency more. On the other hand, a capacitor can shift frequency higher. A capacitor with smaller value shifts frequency more.

Unless it is an emergency, try to avoid using this tuning method. Either inductors or capacitors have inherent loss, which decreases the overall antenna efficiency. It is a better practice to tune antenna through adjusting radiator length.

Spring finger 2 connects to a π -shaped matching network. There are three pairs of solder pads (one shunt, one series, and one shunt) on the PCB. It seems that the antenna is tuned pretty well and no matching is used in the final design. A 0Ω resistor is placed on the series pad to simply make galvanic connection.

Shown in Figure 4.104 are photos of the WLAN and GPS antenna. Figure 4.104a is a production antenna. Figure 4.104b is a semifinished antenna, which is provided for a better view of different radiators. This antenna covers two frequency bands: 1.575 GHz (GPS) and 2.4–2.482 GHz (802.11b/g/n and Bluetooth). It adopts a dual-band IFA design. Radiator 2 is the longer trace and responsible for the GPS band. Radiator 1 takes care of the higher band.

Shown in Figure 4.105 is a semifinished WLAN and GPS antenna, which has been installed on the top-right corner of a plastic frame. Some portion of the flex, which has two gold plated contacts, is passed through the slit on the plastic frame and wrapped to its back.

Shown in Figure 4.106 is a zoom-in photo of the matching circuit of the WLAN and GPS antenna. There are two spring fingers on the PCB. The spring finger 1 connects to ground directly. The spring finger 2 connects to a π -shaped matching network, which is similar to the one used in the secondary antenna. It seems that the WLAN and GPS antenna is also pretty well tuned. No matching component is used and only a 0Ω resistor is placed on the series pad to make galvanic connection.

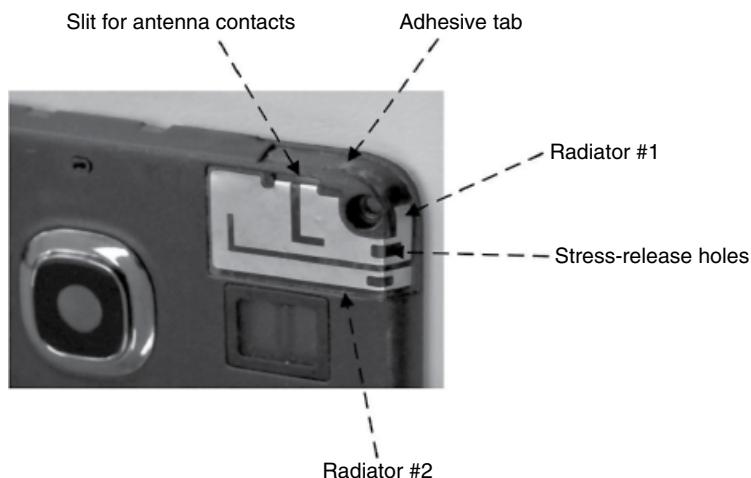


Figure 4.105 Semifinished WLAN and GPS antenna installed on a phone. (Source: Reproduced with permission of Xiaomi, Inc. and Shanghai Amphenol Airwave, Inc.)

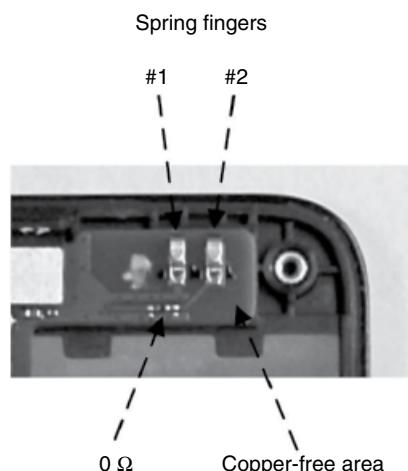


Figure 4.106 Matching circuit of the WLAN and GPS antenna. (Source: Reproduced with permission of Xiaomi, Inc.)

Shown in Figure 4.107 are switch connectors of the primary and the secondary antennas. Detailed discussion about switch connector can be found in Section 5.1.2.1. Because the impedance of a switch connector is always 50Ω , it is an ideal place to tap in a coaxial cable for passive antenna measurement. No matter whether it is a 2G, 3G, or 4G phone; all phones have switch connectors on the signal paths of cellular antennas. The switch connector is used to measure a phone's RF circuit; it can't be used to measure antenna directly.

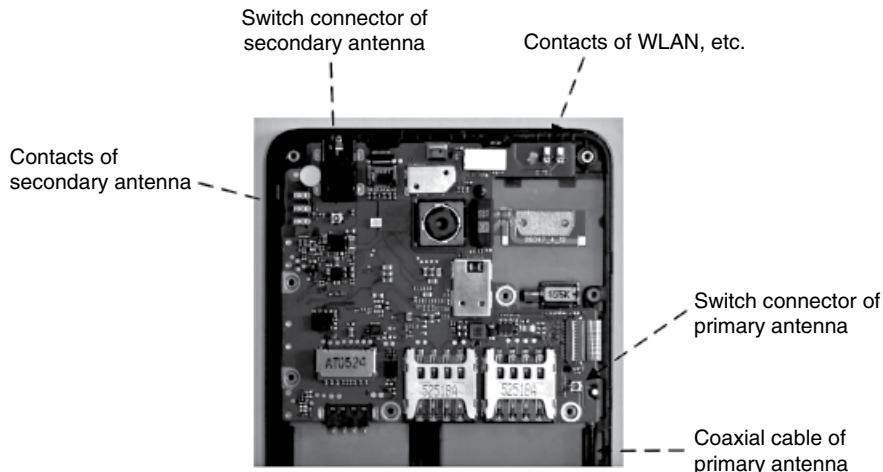


Figure 4.107 Switch connectors on the PCB. (*Source:* Reproduced with permission of Xiaomi, Inc.)

4.10.2 Flagship Phone

Shown in Figure 4.108 are front, side, and back views of Xiaomi 4. The starting price of this phone is 1999 RMB, which is around 317 USD, when purchased without a contract. Xiaomi 4 was released in June, 2014. It has a 5 inch 1920×1080 IPS screen, 3GB ram, and 16GB flash memory.

To convince a customer to buy a high-end phone, which costs several times more than an entry-level phone, faster CPU and GPU, larger memory, higher resolution LCD, and better camera are all essential. However, the most important aspect above all is its appearance, because most customers can't even tell what GPU stands for.

A high-end phone must look like a luxury item. One iconic feature of an high-end phone is its metallic frame. For antenna engineers, metal frames make designing job much more challenging. Apple has patented hybrid slot antenna solutions [44–46] by using the gap between a whole metal frame and PCB inside a phone. This technique has been used in iPhone 3G and 3Gs. Since the advent of iPhone 4, Apple adopted a new approach which is cutting slits on the metal frame. By cutting several slits on the phone frame, from antenna point of view, the frame is actually several isolated metal pieces, which gives antenna designing much more freedom. Recently, big screen phones become the new normal, which actually makes available area for antenna designing much larger. People start to revisit the idea of using the gap between a whole metal frame and inner PCB to designing antennas [47, 48].

Xiaomi 4 adopts the approach of cutting slits on the metal frame. It is obvious there are four slits: two on the top edge and two on the bottom edge.

Shown in Figure 4.109 is the front and back photos of Xiaomi 4's antennas. Because the battery of Xiaomi 4 is unexchangeable, its back cover is also not supposed to be removed. Unlike Hongmi 2A, all three antennas are not masked by black paint. Xiaomi 4's antennas are manufactured by using LDS technologies, which has been discussed in Section 4.2.5.



Figure 4.108 Xiaomi 4. (Source: Reproduced with permission of Xiaomi, Inc.)

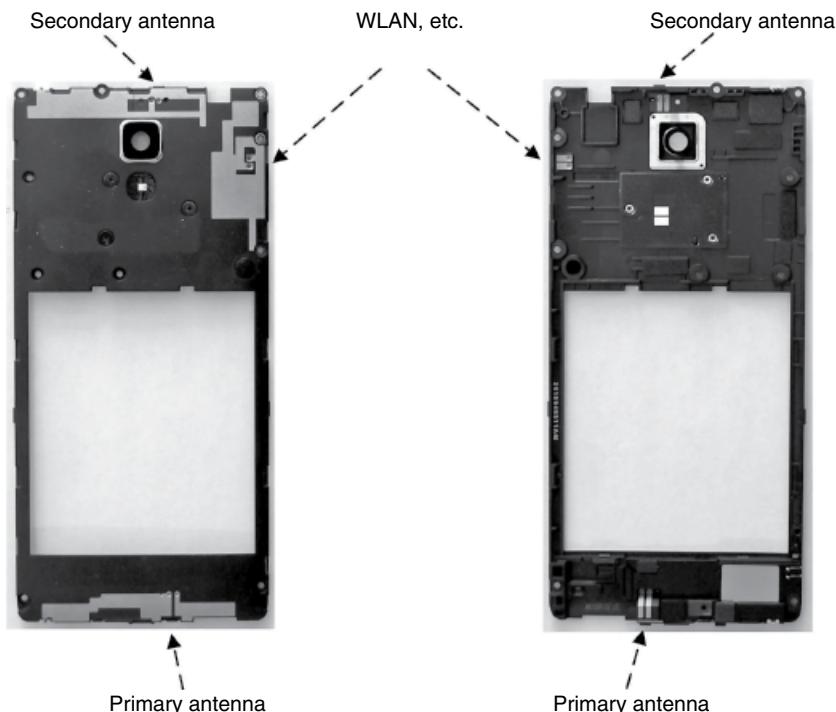


Figure 4.109 Xiaomi 4's antennas. (Source: Reproduced with permission of Xiaomi, Inc.)

The Xiaomi 4 supports both 2.4 and 5 GHz WLAN band. The cellular bands which are supported by Xiaomi 4 are listed in the following:

4G TDD-LTE	B38 (2570–2620 MHz)
	B39 (1880–1920 MHz)
	B40 (2300–2400 MHz)
3G TD-SCDMA	B34 (2010–2025 MHz)
	B39 (1880–1920 MHz)
2G GSM bands	B2 (1850–1990 MHz)
	B3 (1710–1880 MHz)
	B8 (880–960 MHz)

Shown in Figure 4.110 is a PCB of Xiaomi 4. Similar to Hongmi 2A, there are three sets of spring fingers for the primary, secondary, and WLAN antennas, respectively. There are also three switch connectors. Different from Hongmi 2A, Xiaomi 4's WLAN antenna has its own switch connector. More discussion can be found in Section 6.1.4, the short version explanation

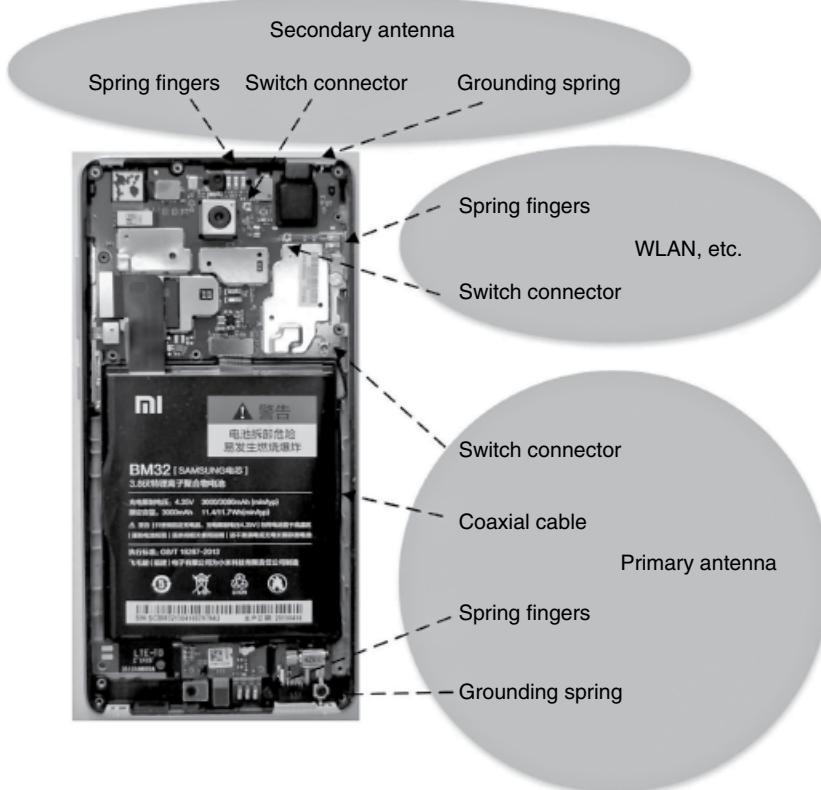


Figure 4.110 Xiaomi 4's PCB. (Source: Reproduced with permission of Xiaomi, Inc.)

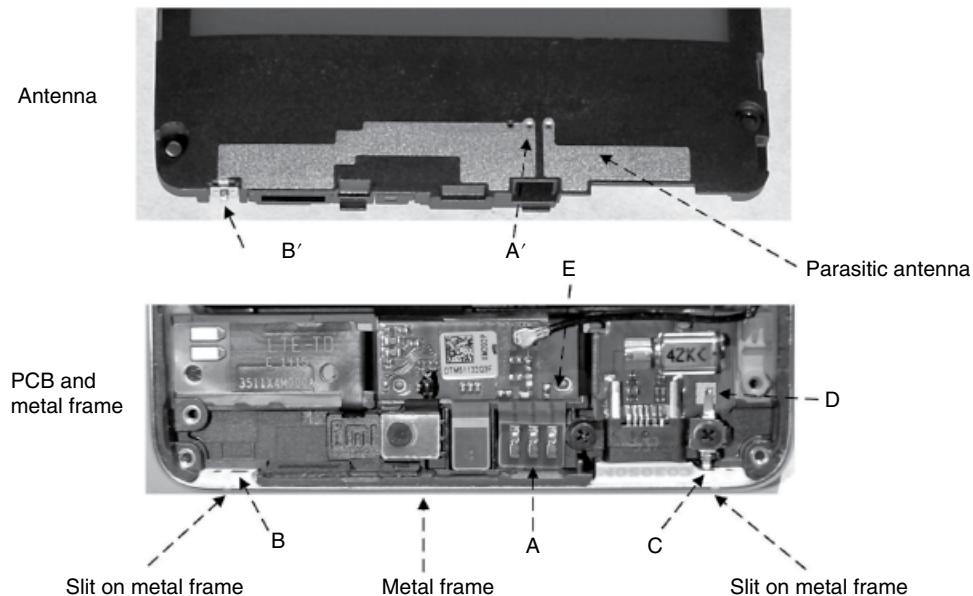


Figure 4.111 Xiaomi 4’s primary antenna. (Source: Reproduced with permission of Xiaomi, Inc.)

is that the new SAR regulation now requires measurements on WLAN bands. To carry out a SAR evaluation, the conducted power, which is measured through a switch connector, must be recorded first. Xiaomi 4 also adopts the design of bottom primary antenna. A coaxial cable is used to connect the antenna to the main PCB on the top. Both the primary and secondary antennas have grounding springs, which are part of antenna structure and different from Hongmi 2A.

Shown in Figure 4.111 is the primary antenna. This antenna is a variation of loop antenna, which has been discussed in Section 4.4. Actually, the antenna on the plastic cover can’t function by itself. It might be difficult to observe, but there is a closed loop embedded in the structure. The loop path, marked as A-A’-B’-B-Metal Frame-C-D-E in Figure 4.111, passes through the antenna, metal frame, and PCB.

Following a detailed description of the loop path. The bottom portion of Xiaomi 4’s metal frame is broken into three pieces by two slits. One slit is next to point B and the other is next to point C. Because of these two slits, the middle portion is isolated from the rest of the metal frame. Points B and C are exposed inner surfaces of the floating metal frame. Point B’ on the antenna connects to point B on the metal frame. Because the LDS antenna can’t provide required spring force, a separate spring fingers is installed at point B’.

Point D on the PCB connects to point C on the metal frame through a grounding spring. The grounding spring was designed to provide required spring force on both points C and D. Both points D and E are on the system ground and automatically connected.

Point A is the center spring fingers on the PCB. It is also the feeding point of the antenna. Point A connects the back of the antenna. There are two metallized strips on the back of the antenna, as shown in Figure 4.109. The right strip connects to point A’ through a metallized hole, which is similar to a via on PCB.

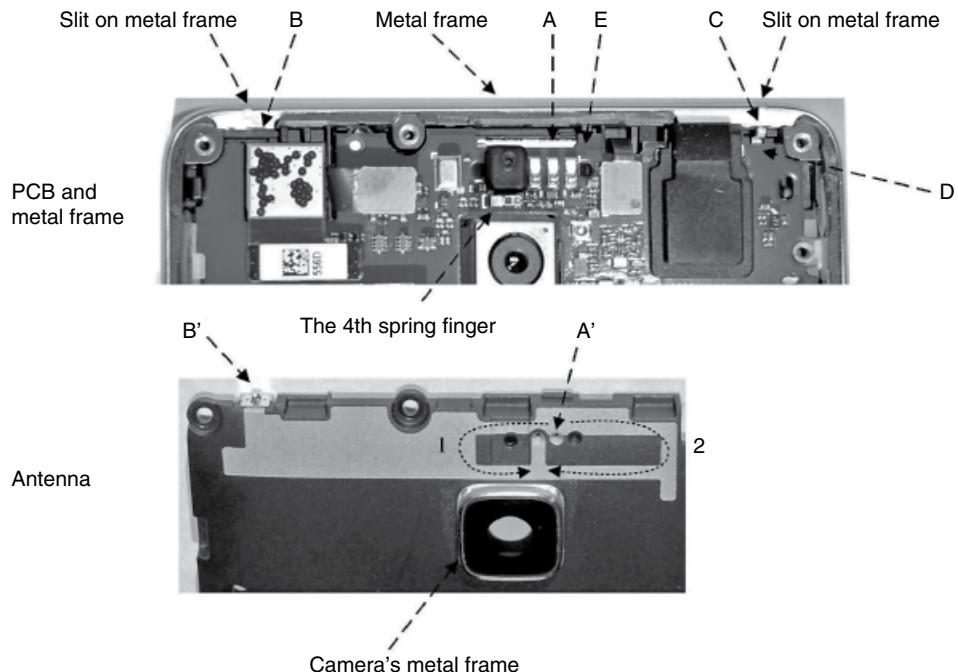


Figure 4.112 Xiaomi 4's secondary antenna. (Source: Reproduced with permission of Xiaomi, Inc.)

Besides the loop antenna, there is a parasitic antenna. The parasitic antenna connects to the right spring finger on the PCB through another metallized hole. The right spring finger is connected to system ground through a 0Ω resistor. It can be seen that there are three though holes on the structure. The left one is not metallized and there is also no metal strip on the back. As has been discussed in Hongmi 2A, this is a redundant feature and is left there just in case.

The antenna has a five-element matching network, but only has three components been populated. Starting from the center spring finger, there is a shunt inductor, a series inductor, and a 0Ω series resistor. The coaxial cable that connects to the matching network is the cable shown in Figure 4.109. The left spring finger is also connected to system ground through a 0Ω resistor; however, it is a floating finger and does not connect to anything.

Shown in Figure 4.112 is the secondary antenna. This antenna is also a variation of loop antenna. Similar to the main antenna, there is a loop path, which is marked as A-A'-B'-B-Metal Frame-C-D-E in Figure 4.112, passes through the antenna, metal frame, and PCB. This loop is almost identical to the main antenna. Two slits on both sides separate the metal frame to three pieces. The middle portion of the metal frame becomes part of the loop.

Similar to the main antenna, there are also three spring fingers beneath the secondary antenna. This time, the right finger is the redundant one. The left finger is connected to system ground through a 0Ω resistor. This generates two extra loops, marked as loop 1 and loop 2 in Figure 4.112, between feeding point and system ground. Loop 1 is relatively small, so it likely provides matching in the lower band, which is similar to what IFA's grounding branch does.

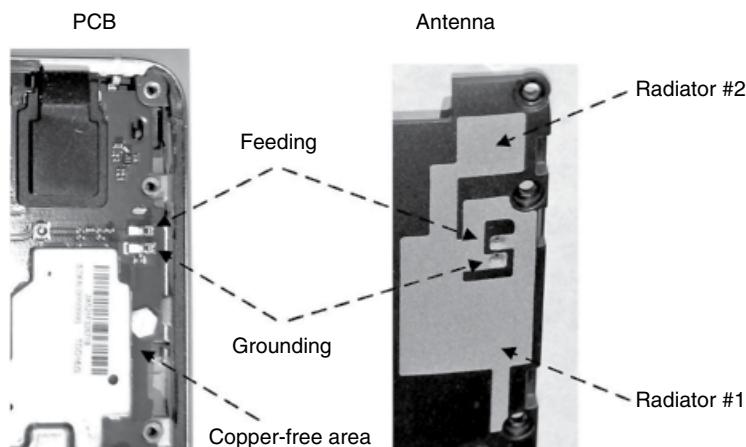


Figure 4.113 Xiaomi 4's WLAN, etc. antenna. (Source: Reproduced with permission of Xiaomi, Inc.)

Loop 2 is reasonably large and might function as an extra loop antenna at the higher frequency band.

There is one interesting feature, which is marked as the 4th spring finger in Figure 4.112. This 4th finger is also connected to system ground. The finger connects to the metal frame of the back camera. If we look at Figure 4.109, there is significant overlap between antenna and camera's metal frame. It seems that the metal frame has been used as a parasitic antenna.

Shown in Figure 4.113 is the WLAN, etc. antenna. Xiaomi 4 supports GPS, Bluetooth, 2.4, and 5 GHz WLAN. The third antenna needs to cover all three bands. This antenna is a variant of IFA design. There are two spring fingers on the PCB. The top spring finger is the feeding point and the bottom one is the grounding point. Radiator 1 can generate the lowest band, GPS band. Radiator 2 and higher order mode from radiator 1 can take care of 2.4 and 5 GHz bands. There are only two through hole on the antenna. Both of them are metallized. Although there is a preserved matching network on the PCB, no matching component is populated. Only two series $0\ \Omega$ resistors are used to make the connection.

Contents in Section 4.10 are the author's observations. Xiaomi Inc. has kindly provided phones, and Shanghai Amphenol Airwave Inc. has kindly provided antenna parts. To avoid conflict of interest or unnecessary leaking of trade secrets, all materials are provided as it is. None of the observation has been confirmed by these two companies. On the other hand, the author has not verified any observation by experiment or simulation. It is recommended that readers only use this section as a general reference.

References

- [1] Wong, K.-L. (2003) *Planar Antennas for Wireless Communications*, Wiley-Interscience.
- [2] Arkko, A.T. (2003) Effect of ground plane size on the free-space performance of a mobile handset PIFA antenna. 12th International Conference on Antennas and Propagation, 2003. (ICAP 2003). (Conf. Publ. No. 491), vol. 1, pp. 316–319.
- [3] Chan, K.H., Fung, L.C., Leung, S.W., and Siu, Y.M. (2006) Effect of internal patch antenna ground plane on SAR. 17th International Zurich Symposium on Electromagnetic Compatibility, 2006. EMC-Zurich 2006, pp. 513–516.

- [4] Fujio, S. (2006) Effect of ground size on plate inverted-F antenna. IEEE International Workshop on Antenna Technology Small Antennas and Novel Metamaterials, 2006, pp. 269–272.
- [5] Urban, R. and Peixeiro, C. (2004) Ground plane size effects on a microstrip patch antenna for small handsets. 15th International Conference on Microwaves, Radar and Wireless Communications, 2004. MIKON-2004., vol. 2, pp. 521–524.
- [6] Wong, K.-L. (2002) *Compact and Broadband Microstrip Antennas*, Wiley-Interscience.
- [7] Balanis, C.A. (2005) *Antenna Theory: Analysis and Design*, 3rd edn, Wiley-Interscience.
- [8] Iskander, M.F. (2000) *Electromagnetic Fields and Waves*, 1st edn, Waveland Press Inc.
- [9] Sadiku, M.O. (2009) *Elements of Electromagnetics*, 5th edn, Oxford University Press, USA.
- [10] Ulaby, F.T., Michielssen, E., and Ravaioli, U. (2010) *Fundamentals of Applied Electromagnetics*, 6th edn, Prentice Hall.
- [11] Buck, J. and Hayt, W. (2005) *Engineering Electromagnetics*, 7th edn, McGraw-Hill Science/Engineering/Math.
- [12] Chi, Y.W. and Wong, K.-L. (2007) “Internal compact dual-band printed loop antenna for mobile phone application,” *IEEE Transactions on Antennas and Propagation*, **55**, 1457–1462.
- [13] Wong, K.-L. and Huang, C.-H. (2008) “Printed loop antenna with a perpendicular feed for pentaband mobile phone application,” *IEEE Transactions on Antennas and Propagation*, **56**, 2138–2141.
- [14] Chi, Y.-W. and Wong, K.-L. (2008) “Compact multiband folded loop chip antenna for small-size mobile phone,” *IEEE Transactions on Antennas and Propagation*, **56**, 3797–3803.
- [15] Chi, Y.-W. and Wong, K.-L. (2009) “Quarter-wavelength printed loop antenna with an internal printed matching circuit for GSM/DCS/PCS/UMTS operation in the mobile phone,” *IEEE Transactions on Antennas and Propagation*, **57**, 2541–2547.
- [16] Almanpis, G., Fumeaux, C., and Vahldieck, R. (2006) “Offset cross-slot-coupled dielectric resonator antenna for circular polarization,” *IEEE Microwave and Wireless Components Letters*, **16**, 461–463.
- [17] Guha, D. and Antar, Y.M.M. (2006) “New half-hemispherical dielectric resonator antenna for broadband monopole-type radiation,” *IEEE Transactions on Antennas and Propagation*, **54**, 3621–3628.
- [18] Leung, K.W., Tse, K.K., Luk, K.M., and Yung, E.K.N. (1999) “Cross-polarization characteristics of a probe-fed hemispherical dielectric resonator antenna,” *IEEE Transactions on Antennas and Propagation*, **47**, 1228–1230.
- [19] Chang, T.-H. and Kiang, J.-F. (2007) “Dualband split dielectric resonator antenna,” *IEEE Transactions on Antennas and Propagation*, **55**, 3155–3162.
- [20] Chang, T.-H. and Kiang, J.-F. (2009) “Bandwidth broadening of dielectric resonator antenna by merging adjacent bands,” *IEEE Transactions on Antennas and Propagation*, **57**, 3316–3320.
- [21] Wong, K.L., Chen, N.C., and Chen, H.T. (1993) “Analysis of a hemispherical dielectric resonator antenna with an airgap,” *IEEE Microwave and Guided Wave Letters*, **3**, 355–357.
- [22] Chen, Y.C., Lin, J.Y., and Tsao, S.M. (2007) Planar patch antenna using temperature stable high-permittivity ceramics. IEEE Conference on Electron Devices and Solid-State Circuits, 2007. EDSSC 2007, pp. 761–764.
- [23] Hauser, R., Fachberger, R., Bruckner, G. et al. (2005) Ceramic patch antenna for high temperature applications. 28th International Spring Seminar on Electronics Technology: Meeting the Challenges of Electronics Technology Progress, 2005, pp. 173–178.
- [24] Natashah, N.M., Muammar, M.I., Soh, P.J. et al. (2008) Design of a microstrip patch antenna using low temperature co-fired ceramic technology. International Conference on Electronic Design, 2008, pp. 1–3.
- [25] Kula, J.S., Psychoudakis, D., Liao, W.J. et al. (2006) “Patch-antenna miniaturization using recently available ceramic substrates,” *IEEE Antennas and Propagation Magazine*, **48**, 13–20.
- [26] Booker, H.G. (1946) “Slot aerials and their relation to complementary wire aerials,” *Journal of the Institution of Electrical Engineers*, Part III A, 620–626.
- [27] Deng, C., Li, Y., Zhang, Z., and Feng, Z. (2015) “A novel low-profile Hepta-band handset antenna using modes controlling method,” *IEEE Transactions on Antennas and Propagation*, **63**, 799–804.
- [28] Lee, S., Jung, H., and Sung, Y. (2015) “A reconfigurable antenna for LTE/WWAN mobile handset applications,” *IEEE Antennas and Wireless Propagation Letters*, **14**, 48–51.
- [29] Mun, B., Jung, C., Park, M., and Lee, B. (2014) “A compact frequency-reconfigurable multiband LTE MIMO antenna for laptop applications,” *IEEE Antennas and Wireless Propagation Letters*, **13**, 1389–1392.
- [30] Sung, Y., (2012) “Compact quad-band reconfigurable antenna for mobile phone applications,” *Electronics Letters*, **48**, 977–979.
- [31] Li, Y., Zhang, Z., Zheng, J. et al. (2012) “A compact Hepta-band loop-inverted F reconfigurable antenna for mobile phone. *IEEE Transactions on Antennas and Propagation*, **60**, 389–392.

- [32] "MIMO (Multiple Input Multiple Output)," http://www.sharetechnote.com/html/BasicProcedure_LTE_MIMO.html. Retrieved 25 October 2015.
- [33] Goldsmith, A., Jafar, S., Jindal, N., and Vishwanath, S. (2003) "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, **21**, 684–702.
- [34] Applebaum, S. (1976) "Adaptive array," *IEEE Transactions on Antennas and Propagation*, **24**, 585–598.
- [35] Vaughan, R.G. and Andersen, J. (1987) "Antenna diversity in mobile communications," *IEEE Transactions on Vehicular Technology*, **36**, 149–172.
- [36] Fujimoto, K. and James, R. (2001) *Mobile Antenna Systems Handbook*, 2rd edn, Artech House.
- [37] Blansh, S., Romeu, J., and Corbella, I. (2003) "Exact representation of antenna system diversity performance from input parameter description," *Electronics Letters*, **39**, 705–707.
- [38] Zhong, H., Zhang, Z., Chen, W. *et al.* (2009) "A tri-polarization antenna fed by proximity coupling and probe," *IEEE Antennas and Wireless Propagation Letters*, **8**, 465–467.
- [39] Li, Y., Zhang, Z., Chen, W. *et al.* (2010) "A dual-polarization slot antenna using a compact CPW feeding structure," *IEEE Antennas and Wireless Propagation Letters*, **9**, 191–194.
- [40] Li, Y., Zhang, Z., and Feng, Z. (2011) "Dual-polarised monopole-slot co-located MIMO antenna for small-volume terminals," *Electronics Letters*, **47**, 1259–1260.
- [41] Li, Y., Zhang, Z., Zheng, J., and Feng, Z. (2012) "Compact azimuthal omnidirectional dual-polarized antenna using highly isolated colocated slots," *IEEE Transactions on Antennas and Propagation*, **60**, 4037–4045.
- [42] Wang, Y. and Du, Z. (2013) "A wideband printed dual-antenna system with a novel neutralization line for mobile terminals," *IEEE Antennas and Wireless Propagation Letters*, **12**, 1428–1431.
- [43] "3GPP Specification Detail 36.306," <http://www.3gpp.org/DynaReport/36306.htm>. Retrieved 25 October 2015.
- [44] Zhang, Z., Schlub, R., Hill, R., and Caballeo, R. (2009) Multiband antenna for handheld electronic devices. US7768462B2.
- [45] Zhang, Z., Hill, R., and Schlub, R. *et al.* (2010) Hybrid antennas with directly fed antenna slots for handheld electronic devices. US7551142B1.
- [46] Zhang, Z., Chen, W., Feng, Z., and Iskander, M. (2008) "Integrated dual-band antenna system design incorporating cell phone bezel," *IEEE Antennas and Wireless Propagation Letters*, **7**, 585–587.
- [47] Ban, Y., Qiang, Y., Chen, Z. *et al.* (2015) "A dual-loop antenna design for Hepta-band WWAN/LTE metal-rimmed smartphone applications," *IEEE Transactions on Antennas and Propagation*, **63**, 48–58.
- [48] Hsu, C. and Chung, S. (2015) "Compact multiband antenna for handsets with a conducting edge," *IEEE Transactions on Antennas and Propagation*, **63**, 5102–5107.

5

Antenna Measurement

So far, we have discussed how to design various external and internal antennas. After making a prototype, the next thing to do is to evaluate the antenna's performance. In the process of designing an antenna, there are two kinds of antenna measurements: passive and active. The passive measurements evaluate the performance of an antenna when the host device is inactive, so some means of external signal generators must be used. The passive measurements include voltage standing wave ratio (VSWR) or reflection coefficient, efficiency, gain, and so on. The active measurements evaluate the performance of an antenna when it is installed in a working device. Measurements, such as total radiate power (TRP), effective isotropic radiated power (EIRP), total isotropic sensitivity (TIS), and effective isotropic sensitivity (EIS), all belong to this category. After an antenna design is rolled out to an assembly line, various production tests are carried out. Passive production tests are carried out by antenna vendors and active production tests are implemented by phone manufacturers.

5.1 Passive Antenna Measurement

5.1.1 Measurement on a Vector Network Analyzer

The network analyzer is a piece of equipment which can measure a wide variety of parameters, such as reflection coefficient and transmission loss. The scalar network analyzers (SNAs) can only measure the amplitude of various parameters. SNAs are relatively cheaper and are widely used in production lines. The vector network analyzers (VNAs) can measure both amplitude and phase, which means they can display impedance on the Smith chart. For an antenna engineer, a VNA, such as the one shown in Figure 5.1, is a necessity of life. When we tune an antenna, what we are actually doing is moving the antenna's impedance around on the Smith chart by tweaking different parameters. With the help of a VNA, the tuning process can be conveniently visualized.

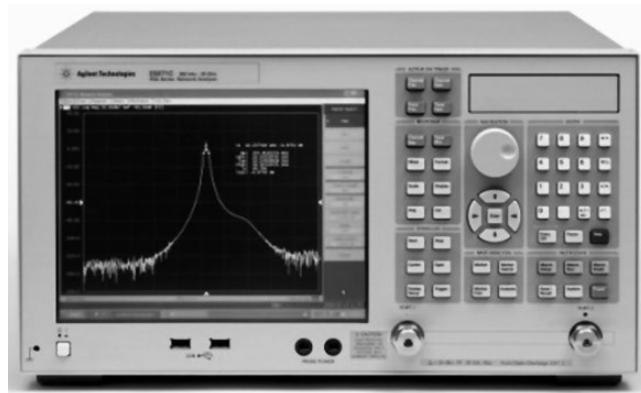


Figure 5.1 An entry-level VNA, E5071C. (Source: Reproduced with permission of Agilent Technologies.)

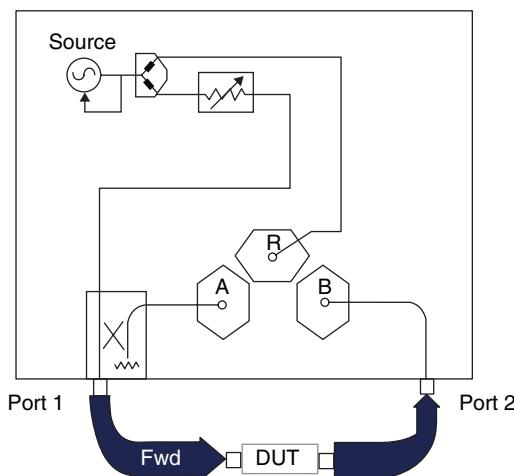


Figure 5.2 Schematic of a simplified VNA. (Source: Reproduced with permission of Agilent Technologies.)

For more in-depth knowledge on VNA, refer to the “Agilent Network Analyzer Basics” [1]. Shown in Figure 5.2 is the schematic of a simplified VNA. The simplified VNA has two ports, and is only able to measure S_{11} and S_{21} . There are three vector receivers inside the VNA. Each of them can measure amplitude and phase simultaneously. R is a reference receiver. It is connected to the signal source through a power divider. Therefore, what it measures is proportional to the power flows out of port 1, which is connected to the other branch of the power divider. A is a reflection receiver, which is connected to the port 1 through a directional coupler. The coupler is designed to only couple the power reflected from the device under test (DUT) and disregard the power from port 1. B is a transmission receiver, which measures how much power passes through the DUT and enters port 2. Based on the measured complex values from receivers A and R, S_{11} can be calculated. With values from receivers B and R, S_{21} can also be obtained.

The schematic of a real two-port VNA, which can measure all four S parameters (S_{11} , S_{21} , S_{12} , and S_{22}), will be more complex. However, the principles of measurement are the same. All other parameters, such as complex impedance and group delay, can all be deduced from the S parameters.

To evaluate the matching of an antenna, both the reflection coefficient and the VSWR can be used. The reflection coefficient, another name for either S_{11} or S_{22} , calculates the ratio between reflected power and total incident power. The VSWR calculates the ratio between the maximum and minimum voltage of a standing wave on the transmission line. Although their definitions are different, those two measurements are actually equivalent. The conversion formulas between them are shown in Equation 5.1.

$$\text{Reflection coefficient} = 20 * \log_{10} \left(\frac{\text{VSWR} - 1}{\text{VSWR} + 1} \right) \text{ (dB)} \quad (5.1)$$

$$\text{VSWR} = \frac{1 + 10^{\text{RL/20}}}{1 - 10^{\text{RL/20}}} \quad (:1)$$

Shown in Table 5.1 is the corresponding list between reflection coefficient and VSWR. When setting an antenna's specification, some companies like to use reflection coefficient while others prefer VSWR, so conversion is commonly done. It does not look very professional if you have to pull out a calculator every time a conversion is needed. However, it is also a little bit too much to ask somebody to memorize the whole table.

As an easy way out, only the following two approximate relations need to be memorized:

1. Reflection coefficient -10 dB approximately equals VSWR $2:1$.
2. Reflection coefficient -6 dB approximately equals VSWR $3:1$.

For a cellular antenna design, -10 dB is great. It is hard to find a reflection coefficient better than that. An antenna with -6 dB reflection coefficient is still a good one if it is an internal

Table 5.1 Conversion table between VSWR and the reflection coefficient

VSWR	Reflection coefficient (dB)
$1:1$	$-\infty$
$1.22:1$	-20.0
$2:1$	-9.54
$3:1$	-6.02
$4:1$	-4.44
$5:1$	-3.52
$6:1$	-2.92
$7:1$	-2.50
$8:1$	-2.18
$9:1$	-1.94
$10:1$	-1.74

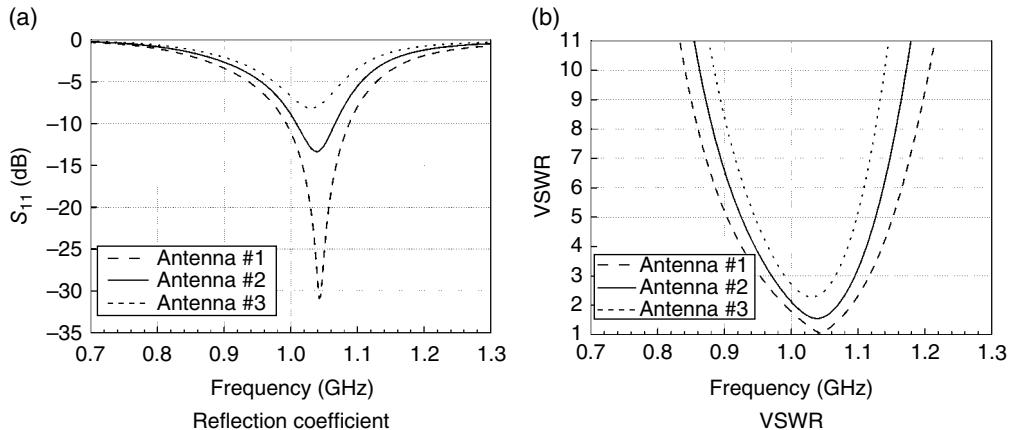


Figure 5.3 Return loss vs. VSWR.

multiband antenna. Depending on the project, a reflection coefficient of -4.5 dB at band edges might be acceptable. As a comparison, for big antennas, such as base station sector antennas and reflector antennas, a -20 dB reflection coefficient, which approximately equals VSWR $1.2:1$, is normally chosen as the specification.

Although theoretically the reflection coefficient and the VSWR are equivalent, when designing cellular antennas, VSWR is better for illustrating results. The reflection coefficient puts more emphasis on the range of small reflection. For a cellular antenna, an antenna with return loss of -30 dB is not much better than a -10 dB one. However, a 6 dB antenna is notably different from a -4.5 dB one. Contrary to the reflection coefficient, the VSWR puts more emphasis on the range of large reflection. Any VSWR between $1:1$ and $2:1$ means great matching, which corresponds to a reflection coefficient range from $-\infty$ to -10 dB . A VSWR from $2:1$ to $\infty:1$ corresponds to a reflection coefficient from -10 to 0 dB . To give a more intuitive example, both the reflection coefficient and the VSWR of the same antenna are illustrated in Figure 5.3.

The VSWR can effectively represent an antenna's matching condition in the frequency domain, and the Smith chart is more useful in tuning an antenna. It is very useful to be able to look at both of them simultaneously and most modern VNAs support that. There is more than one way to display both the VSWR and the Smith chart simultaneously. Using Agilent E5071C as an example, you can use either multichannel mode or multiple trace mode. However, multichannel mode is overkill and is inconvenient. The multichannel mode requires a dedicated calibration and port extension for each channel. In this mode, to display both the VSWR and the Smith chart, two separate calibrations are required. The multiple trace mode is the one which suits antenna designing better. By selecting two traces and a horizontal-tile arrangement, the display of VNA should look like Figure 5.4.

An antenna is a radiator, which means any nearby object can affect the antenna's response. So a phone fixture needs to be far from the bench's surface and be supported by a piece of foam, whose permittivity is close to 1. Shown in Figure 5.5 is such a testing setup. To ensure a fixture is immune from nearby objects, vary the distance between a fixture and a bench by adding more foam, the antenna's response should always be stable.

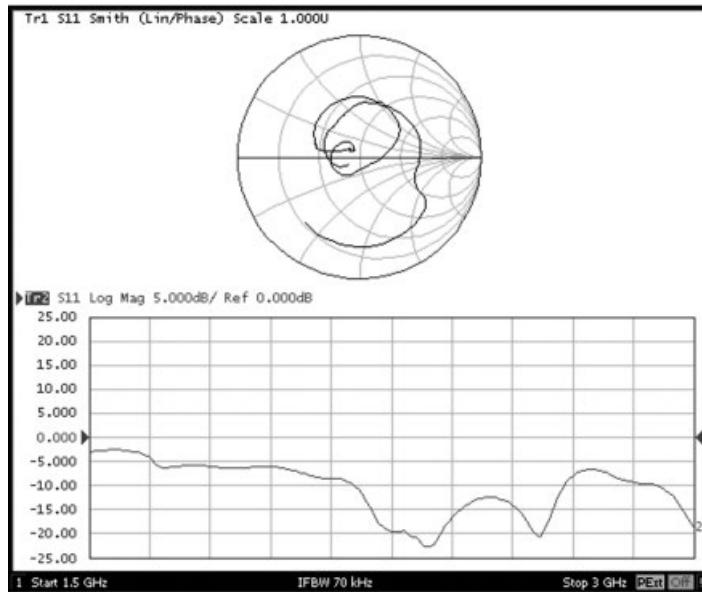


Figure 5.4 Displaying both the VSWR and the Smith chart on a VNA.



Figure 5.5 A fixture connected to a VNA.

After designing an antenna, there is a trick which can give a rough idea whether the antenna works well. If you sweep your hand several centimeters above the fixture, the antenna's response should synchronously fluctuate with the hand's movement. More fluctuation normally means better performance. If there is a resonance which is very stable, no matter how the hand moves, it is most likely that the resonance is an internal parasitic one, which normally does not radiate well and has very poor efficiency. As shown in Figure 5.6, the antenna's response varies with the hand's movement.

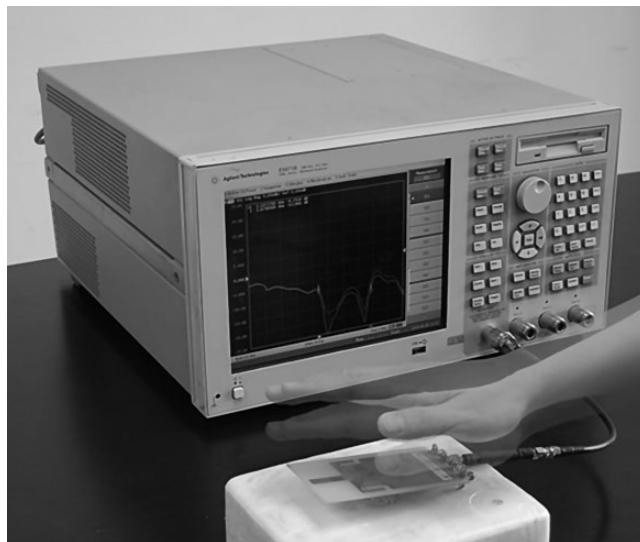


Figure 5.6 Use hand to quickly check an antenna's performance.

5.1.2 *Fixture*

Making a good fixture is an essential skill to ensure success in all antenna projects. The ultimate goal for a fixture is to duplicate an environment which is electromagnetically identical to the real phone, for an antenna under test. As has been demonstrated in the previous chapters, usually the body of a phone is also part of the radiator. Any modification required to make a fixture might move the resonant frequency or affect the matching condition. Thus, it is quite critical to make a fixture which will disturb the radiation current on a phone as little as possible.

Normally, when we begin to design an antenna for a phone project, we start with a blank slate. It is important to solicit some prototype parts from mechanical engineers to start with. The following two parts are the most important ones:

1. Printed circuit board, PCB. The trace layout on the PCB is not critical. The ground pattern on the PCB is more important, because the ground is part of the radiator. All the metal in the keep-out area should be removed. At the early stage, a single-sided board is good enough for antenna designs. Extra caution must be paid if a double-sided PCB is used. The metal on both sides of the PCB must be well connected; otherwise, some undesired parasitic resonance might occur.
2. Antenna support. An antenna's bandwidth and performance are highly correlated with the volume an antenna occupies. Without antenna support, it is very difficult to control an antenna's volume. A handful of samples from a machine shop should be enough at the beginning. Try to use the same plastic which the antenna is planning to use in the mass production.

Other parts, such as the phone housing, the battery, and the LCD, are also important. The more parts you have, the more accurate you will be able to predict the antenna's performance; however, it can still be done without them.

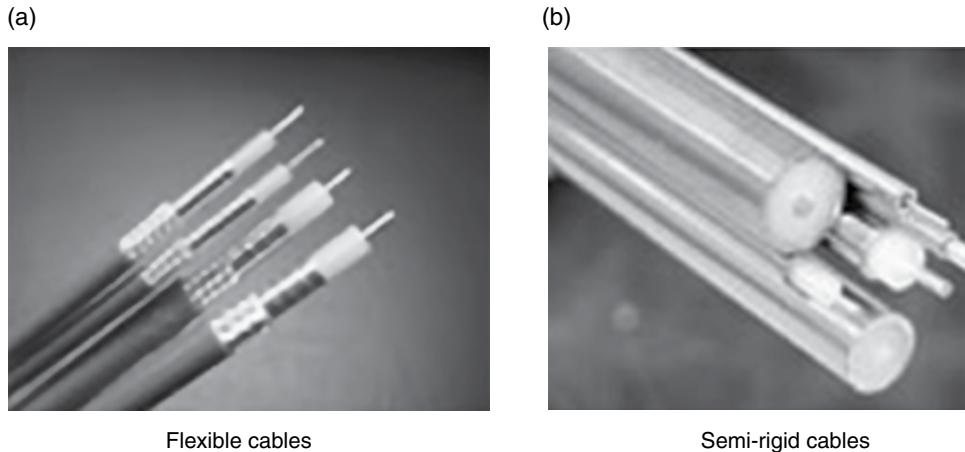


Figure 5.7 Coaxial cables.

If you find the first prototype antenna has 80% efficiency, don't be too excited. The efficiency will drop when you get all the real parts. If the efficiency of your first prototype is the same as the project's specification, that's not the right state to be in.

5.1.2.1 Location of Testing Cable

Because a phone is inactive when doing passive measurement, an external source must be used to carry out various measurements. To connect equipment to an antenna, some kind of transmission line is needed. At the frequency range on which the phone operates, coaxial cables are the commonly used transmission lines. Shown in Figure 5.7 are various coaxial cables.

Figure 5.7a shows several high-grade flexible cables with different dimensions. The inner conductor is a solid silver plated copper wire. Outside the inner conductor is the insulation layer, which is made of polytetrafluoroethylene or Teflon. The outer conductor of the insulation layer is composed of two layers of metal shield. The inner shield layer is a silver plated copper tape, helically wrapped with overlap between layers. This layer is the main shielding layer. The outer shield layer is a silver plated copper wire, tightly braided over the inner shield. The braids are primarily a strength member that also adds additional radio frequency (RF) shielding. Outside the shield layer is the jacket layer. For some cheaper coaxial cables, the outer conductor layer is composed of only silver plated copper wire.

A semirigid cable, as shown in Figure 5.7b, also has inner and outer conductors separated by an insulation layer. What makes a semirigid cable different from a flexible cable is that it has a solid copper outer sheath. Several cables with different diameters are shown in Figure 5.7b. Semirigid cables offer superior performance compared to flexible cables, especially at higher frequencies. For some applications, the major disadvantage is that the cable, as its name implies, is not very flexible, and is not intended to be flexed after initial forming. However, when making a fixture, a semirigid cable is the best choice.

For the convenience of testing, do not make a fixture with a long cable attached. Instead, put a short pigtail cable on the fixture and use a long flexible cable to make the interconnection between

equipment to the short pigtail cable. To maintain the continuity of the coaxial structure, a coaxial connector is required as the interface between a pigtail cable and a flexible cable. SubMiniature version A (SMA) connectors are the *de facto* industry standard for mobile antenna industry. Either buy a semirigid cable and a suited connector separately, then assemble them yourself, or order semirigid cable assemblies. The SMA connectors shown in Figure 5.8 are PCB board mount-type connectors. They are supposed to be soldered to a board directly. The one with a pin at the center is the male type. The other one with a hole in the middle is the female type.

Shown in Figure 5.9a is an inverted-F antenna (IFA) model used in simulations. All gray areas are metal surface. The small area, which is marked by a dash circuit and magnified in the insert, is a lumped port. A voltage source is used at the port to feed the antenna. Shown in Figure 5.9b is a prototype of the IFA. A semirigid coaxial cable is used as the feeding transmission line. The detail of the feeding structure is magnified in the insert. The end of the

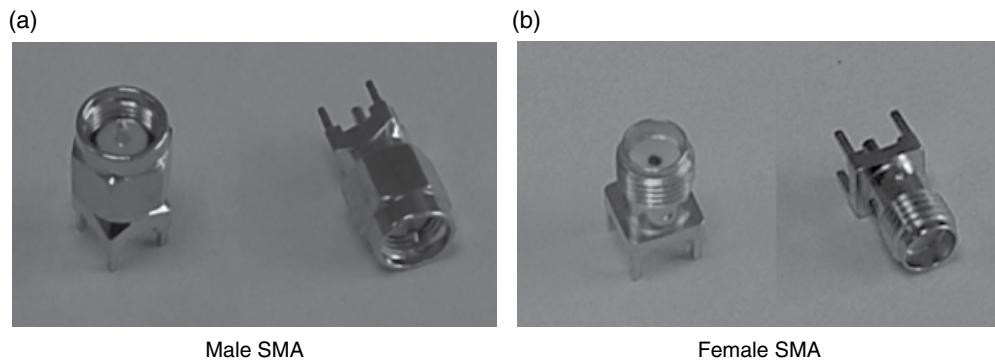


Figure 5.8 SMA connectors.

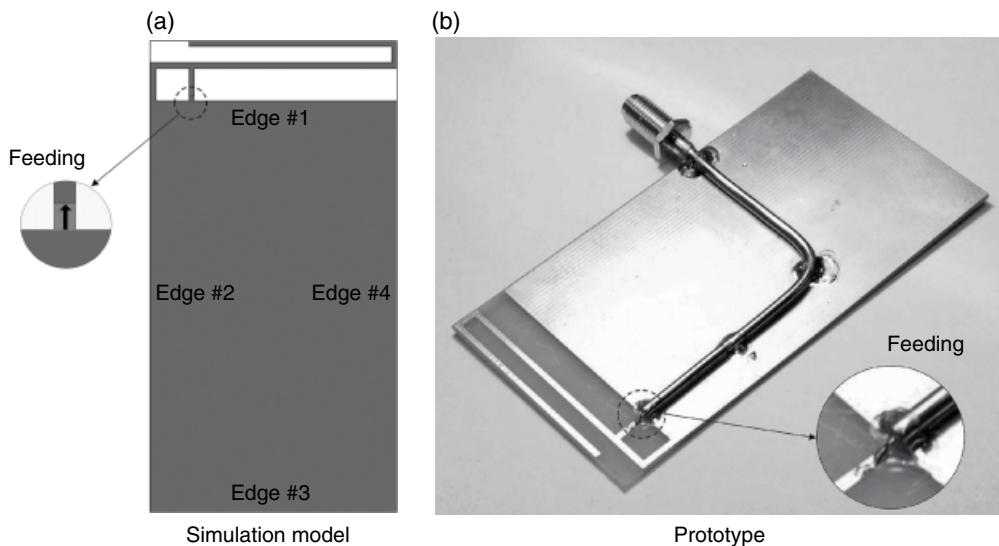


Figure 5.9 Making a prototype.

coaxial cable's outer conductor is soldered to the edge of the ground plane and the inner conductor is soldered to the antenna. The whole coaxial cable is placed on top of the ground plane and soldered to the ground at several points. When a signal travels along the cable, all the currents and electromagnetic field are constrained inside the cable. There is no current on the outside surface of the outer conductor and thus no radiation. At the open end of coaxial cable, the current on the inner conductor flows to the antenna and the current on the outer conductor flows to the ground. As the end of coaxial cable is connected across the gap, it is equivalent to a voltage source over the gap, just like the voltage source used in Figure 5.9a.

In Figure 5.9b, at the opposite end of the semirigid cable there is an SMA connector. The connector is placed on the right side. Actually, the connector's position is not selected arbitrarily; it is optimized to minimize the intrusion effect of a testing cable.

For most cellular antennas, the ground is part of the radiator. There are strong electromagnetic fields and current distributions on the ground. If the position of the pigtail connector is not well chosen, currents on the ground can leak to the testing cable. As a testing cable is much longer than the ground, the cable becomes a radiator by itself.

Let's use the dual-band antenna shown in Figure 5.9 as an example to explain how to select a position for a pigtail connector. There are four edges on a ground plane. Edge 1 is adjacent to the antenna element, putting anything here will strongly affect the antenna's response. Edge 3 is the furthest edge of the antenna, so some people might think this edge is the best choice. However, it isn't. The length of ground can affect an antenna's response, especially at the lowest band. More discussion on the impact of a ground on an antenna's response can be found in Section 3.4. On edge 2, there are strong currents at both low and high bands. Of all four edges, edge 4 is the best location for the connector.

In practice, the exact location of the connector is decided by trial and error. You can put the connector at a location of your best guess, connect the fixture to a network analyzer through a flexible cable, use your hand to touch the connector and different spots along the PCB's edge, and check the antenna's response at the same time. The ideal location for the connector is where the antenna's response does not change at all when your hand touches it. That ideal location may not exist, so just pick a location which provides the most stable response across all interested bands. A choke, which will be discussed later, can be used to take care of the residual effect.

Shown in Figure 5.10 is a fixture which has some improper details. Spot A, which is the end of coaxial cable, must be soldered to the ground. Without that solder point, the current has to flow on the outer surface of the cable until it reaches the first solder point. That is equivalent to adding a series inductor into the feeding line and can distort the antenna's real response, especially at the higher band. The inner conductor marked at spot B is too long and it also adds parasitic series inductance to the antenna's response. The length of the pigtail, which extrudes out of the PCB, at spot C is a little longer than necessary. If an ideal sweet spot for the connector can be found, in theory, the length of the pigtail should not affect the antenna's response. In reality, more or less, the connector still has some impact, so it is good practice to keep that length as short as possible.

When making a fixture out of a real phone, it is important to figure out where to disconnect the transceiver and tap the semirigid cable in. For the primary antennas, there is always a connector; usually, there is a switch connector between an antenna and a transceiver. Due to production variations when making a phone, a phone must go through a tuning and calibration process to be compliant with various requirements on safety and performance. Shown in

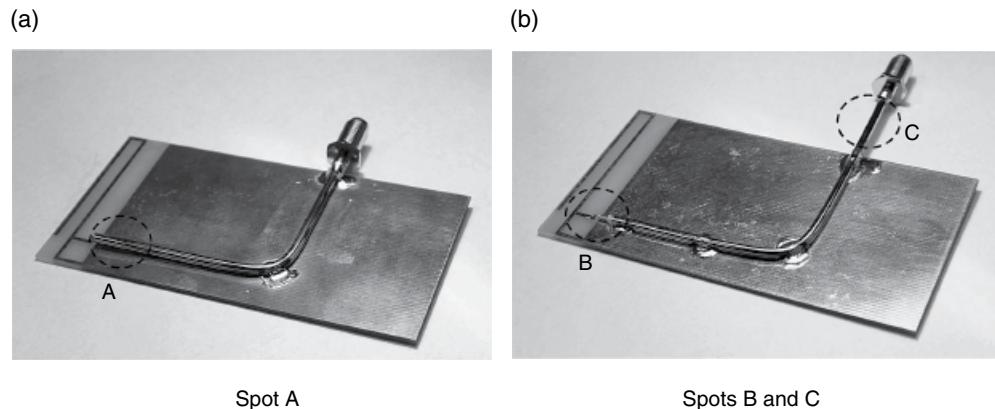


Figure 5.10 A fixture with improper details.

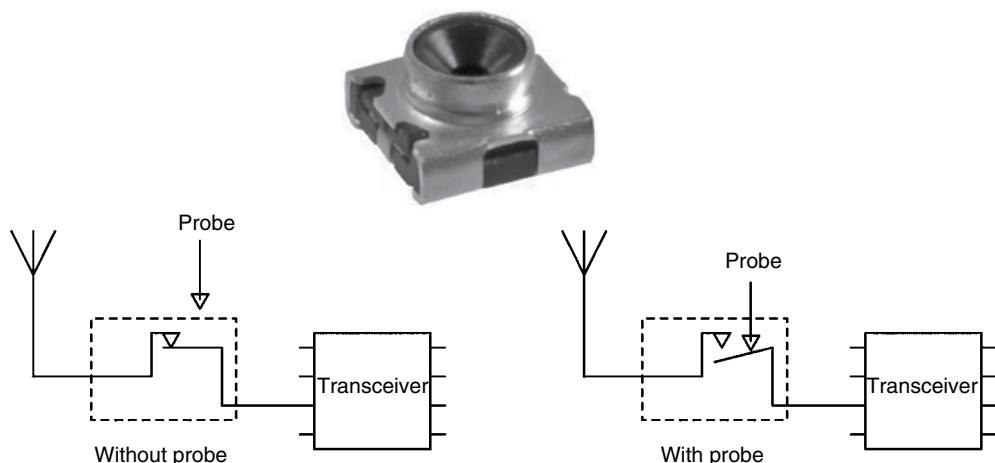


Figure 5.11 Switch connector. (Source: Murata Manufacturing Co. Ltd.)

Figure 5.11 is a switch connector. In a normal operating mode, the signal just passes through the connector. When a probe is plugged in, it disconnects one side and makes contact with the other side. In a phone, the connector disconnects an antenna, so various pieces of equipment can make accurate measurements on the transceiver.

The PCB area of a switch connector is an ideal tapping point for a semirigid cable. Switch connectors are coaxial connectors, so they have both inner and outer conductors. Naturally, the solder pads of a switch connector on a PCB are always readily routed to both signal and ground. All ports of a switch connector are designed as 50Ω , which means all transmission lines connected to a connector have a characteristic impedance of 50Ω . When making a fixture, we just remove the connector with a hot gun and put the cable in, and then we are all set.

For some auxiliary antennas, such as GPS antennas or WLAN/BT antennas, there is no switch connector on the PCB. As the specifications for those applications are quite loose, no

tuning process is needed. However, even in such circumstances, the antenna is still connected to the chipset by a 50Ω transmission line. In some cases, the 50Ω line terminates at the transceiver chip, which can be removed to make space for a pigtail cable. In other cases, a transceiver has a balanced RF port and a balun is used to convert a single-end signal to a balanced one. In such cases, the balun needs to be removed and the pigtail cable must be tapped to the single end 50Ω line.

5.1.2.2 Choke and Ferrite Beads

When making a fixture, it is not always possible to put the pigtail cable at the ideal location. A phone's PCB is populated with various integrated circuits, shielding boxes, board-to-board connectors, and so on. It is not good practice to significantly modify the PCB, as it might change the radiation property of the PCB. When a testing cable is attached to the pigtail of a fixture, some residual current will flow on the surface of the cable. That may affect an antenna's frequency response, radiation pattern, and efficiency.

To mitigate the current on the cable, some kind of RF choke should be used. Shown in Figure 5.12 is a sleeve choke. The choke is formed by several series of arranged metal cylinders. One end of each cylinder is soldered to the outer conductor of a coaxial cable. This kind of choke is not reciprocal and can only be connected to a fixture in one direction.

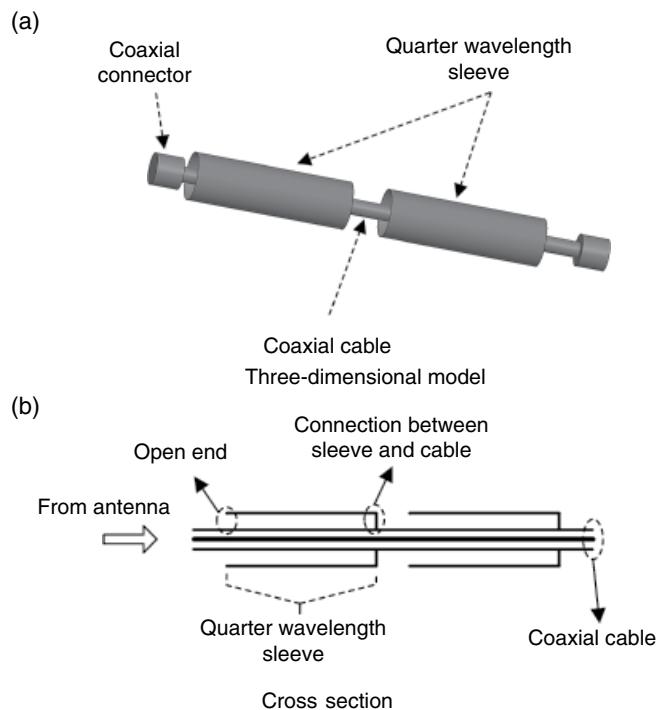


Figure 5.12 Sleeve choke.

Shown in Figure 5.12b is the cross section of a sleeve choke. The metal sleeve and the outer conductor of coaxial cable form another coaxial structure. For current flows on the cable's outer surface, this coaxial structure functions as a transmission line. The further end of the sleeve, which is referred to the antenna side, is connected to the outer conductor of coaxial cable. This connection is equivalent to a short circuit for the sleeve structure. The length of sleeve is one quarter of a wavelength. Based on the transmission line theory, a short circuit becomes an open circuit after one quarter of a wavelength. At the open end of the sleeve, it becomes an open circuit and the current here should be 0. In reality, although most currents are choked at the first sleeve, some residual currents still pass through. A few more sleeves are used to further reduce the surface current on the cable.

The sleeve choke shown in Figure 5.12 is a single-band choke. As the effectiveness of current suppression is related to the quarter wavelength, the frequency band of the choke is quite narrow. To cover different bands, chokes of different dimension are required. Some studies [2–4] on how to make multiband chokes have been done, but they are quite complex and seldom used in practice.

Nowadays, many antennas are multiband, so it is preferable to use a wideband choke instead of several narrowband ones. Shown in Figure 5.13 are some ferrite bead chokes. Ferrite chokes are produced mainly for electromagnetic interference (EMI) purposes. As the cable radiation in antenna measurements is basically the same as the EMI radiation, some means of suppression can be used. Just as its name implies, ferrite chokes are made of ferrite material. There are two commonly used ferrite material: manganese-zinc (Mn-Zn) and nickel-zinc (Ni-Zn). Mn-Zn ferrite chokes are used in 1–30 MHz. The marked working frequency for Ni-Zn chokes can be up to 1 GHz. Mobile communication starts at hundreds of megahertz, so only Ni-Zn chokes can be used.

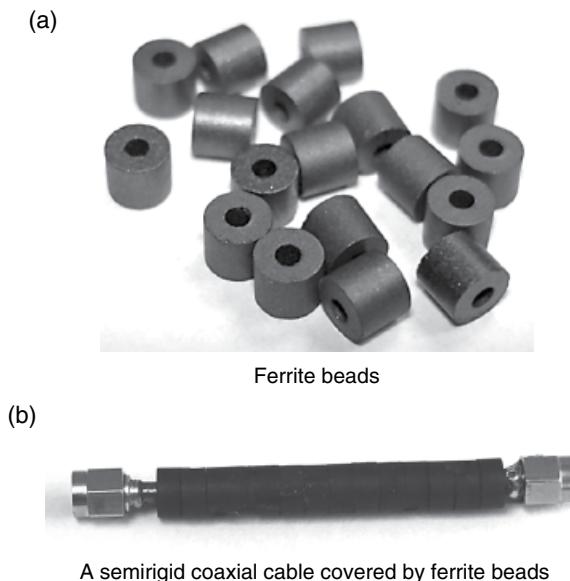


Figure 5.13 Ferrite bead choke.

Ferrite material has very high permeability (μ), which ranges from hundreds to thousands. The inductance L of a piece of wire is directly correlated to the permeability of the surrounding material, and the surface impedance is defined by $j2\pi fL$, where f is the working frequency. Based on these relations, a ferrite choke should have a better suppression effect at a higher frequency. That is only half correct, because the permeability of ferrite material is not a constant but a variable of frequency.

Shown in Figure 5.14 is the measured data of Round Cable EMI Suppression Core 2661023801 made by Fair-Rite, Inc. Below tens of megahertz, a wire with ferrite choke appears to be pure reactance, which is lossless. Between the tens of megahertz and hundreds of megahertz frequency range, the wire becomes lossy, so it functions as an inductor in series to a resistor. In the middle of this frequency range, while the resistance keeps increasing, the reactance starts to decrease. Around 600MHz, the reactance part of a ferrite choke becomes so small that it functions mostly as a resistor. At the highest frequency, the resistance will also start to decrease.

To function as an ideal choke, the choke should be pure reactance. When there is resistance portion in the total impedance, the choke becomes lossy and consumes power, thus making the measured antenna's efficiency worse than it actually is. In some engineers' opinion, as ferrite chokes are specified as 1GHz components, they should not be used above that frequency. Other engineers think differently. They think this kind of ferrite chokes can be used in all mobile communication bands, which is up to 2.17GHz. The author supports the later opinion. The following are several arguments.

Although the ferrite chokes become quite lossy at high frequency, the impedance is still relatively high. The power consumed by a ferrite choke is proportional to V^2/R . The E field at the connector spot is less than the E field at the antenna feeding. As the R is also much larger than the antenna's impedance, the total impact on the antenna's efficiency is reasonable. As discussed in Section 5.1.2.1, the connector position must be carefully selected so the current leak to a cable is as little as possible. The ferrite choke only serves as an auxiliary means to

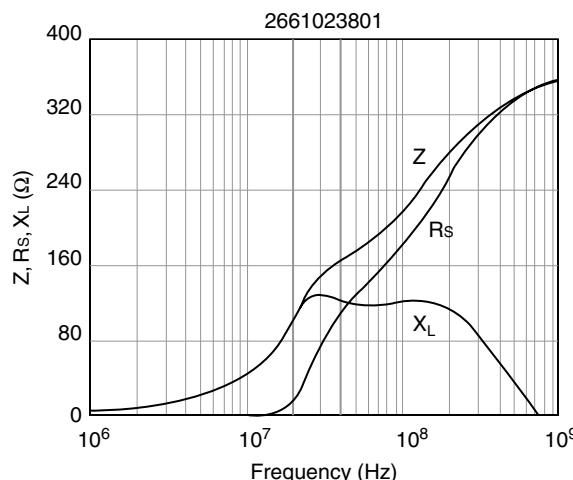
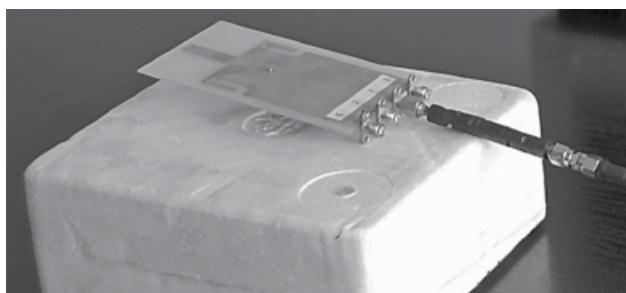


Figure 5.14 Impedance, reactance, and resistance vs. frequency (EMI Core 2661023801). (Source: Fair-Rite Products, Corp.)

mitigate residual current on a cable. So even if the ferrite chokes absorb all the currents instead of stopping them, it shouldn't degrade the antenna efficiency too significantly. On the other hand, lossy chokes can still provide a stable condition for antenna measurement that is much better than shooting a moving target. As the chokes always make the antenna efficiency worse, it is safe to use the value measured with chokes, which inherently provide a safety cushion, and we can be certain that the real value must be better. A lossy choke might change the frequency response of an antenna, so at the later stages when doing active antenna measurements, when the testing cable is no longer an issue, a batch of slightly differently tuned antennas should be measured to capture the optimum parameters for the final production antenna.

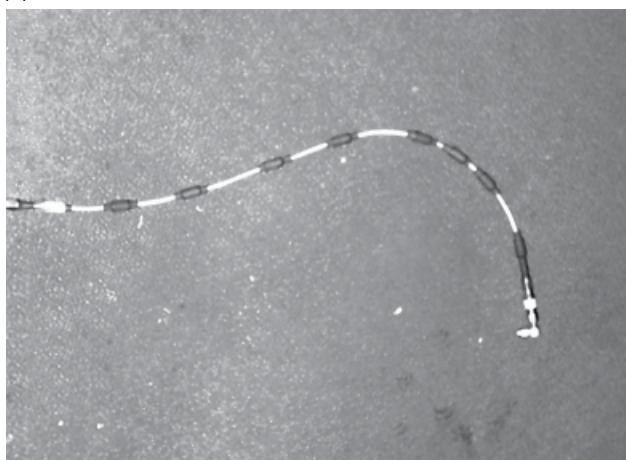
Shown in Figure 5.15a is a fixture with some ferrite chokes attached to effectively suppress the surface current on the cable. A 10cm long tightly packed ferrite beads should be enough to isolate the measurement cable. You can check the effectiveness of the choke by touching the measurement cable with your hand. With or without a hand, the measurement value should always be stable.

(a)



A cable used with a fixture

(b)



A cable used in an anechoic chamber

Figure 5.15 Cables with chokes.

Shown in Figure 5.15b is a flexible cable with some ferrite chokes sparsely attached. This kind of cable is used in anechoic chambers. For a chamber used for measuring high gain antennas, an untreated cable is fine. Because the cable can be routed on the back of the antenna and placed in the shadow of sidelobes, the measurement error introduced by the cable is negligible. On the other hand, all mobile communication antennas are low-gain antennas. That means these antennas radiate in almost all directions. If a long untreated cable is used in the chamber, some electromagnetic energy will be picked up by the cable and re-radiate back to the space, which causes measurement error. By putting ferrite chokes on a cable, the long metal object is divided into discrete small segments, if considered from the electromagnetic point of view. Each segment is much shorter than a wavelength, thus it is transparent to the electromagnetic wave. As the cable is not immediately adjacent to a phone fixture, the coupled current on the cable is quite weak, so it is not necessary to use tightly packed ferrite chokes as shown in Figure 5.15a. To suppress secondary radiation from a test cable, whether a ferrite choke is lossy or not is no longer a concern. As long as a ferrite choke can kill the surface current, it is qualified.

In conclusion, traditional sleeve chokes have pretty good characteristics in a relatively narrow bandwidth. They were popular when most antennas were single-band ones. Nowadays, if you need to use a choke, the best option is a ferrite choke. As different ferrite materials have different specifications, you might have to use two or three ferrite chokes with different dimensions and frequency characteristics to cover an ultrawide bandwidth, say, from 800MHz to 6GHz.

However, those who are not convinced that ferrite chokes can be used at high frequency can live without a choke. Just make sure that the pigtail connector is located at an optimal location which can minimize the impact of a measurement cable.

5.1.2.3 Port Extension

Before we can use a VNA, normally we need to calibrate it first. The calibration process used to be quite cumbersome and required multiple standards, such as OPEN, SHORT, and LOAD. For more on the principle of calibration, some application notes [5, 6] from Agilent, Inc. are quite helpful. Nowadays, the calibration can be done with an E-Cal module, shown in Figure 5.16, which is pretty much foolproof. After the calibration, the reference plane of the VNA is set to the end of the flexible cable. If we only want to know whether an antenna works well, a calibrated VNA can provide an accurate measurement. If we need to tune an antenna or design a matching network, we need one more step: port extension.

Shown in Figure 5.17a is the schematic of a test fixture. The impedances shown in Figure 5.17b and c are results of the same dual-band antenna using different reference planes, one at the calibration plane and the other at the antenna feed. The complex impedances of two measurements are totally different. As discussed in Chapter 2, when designing a matching network, the topology of a matching circuit depends on its original impedance, whether it is capacitive or inductive. If the measured complex impedance is wrong, it is impossible to design the matching circuit correctly. The difference between two results is caused by the pigtail cable between the two reference planes. As we know, a transmission line can clockwise rotate any impedance on the Smith chart. The impedance moves a full circle on the Smith chart for every half wavelength. For a wideband or a dual-band antenna, the cable's electrical length varies significantly at different frequencies, thus an antenna's impedance curve is spread widely on the Smith chart.

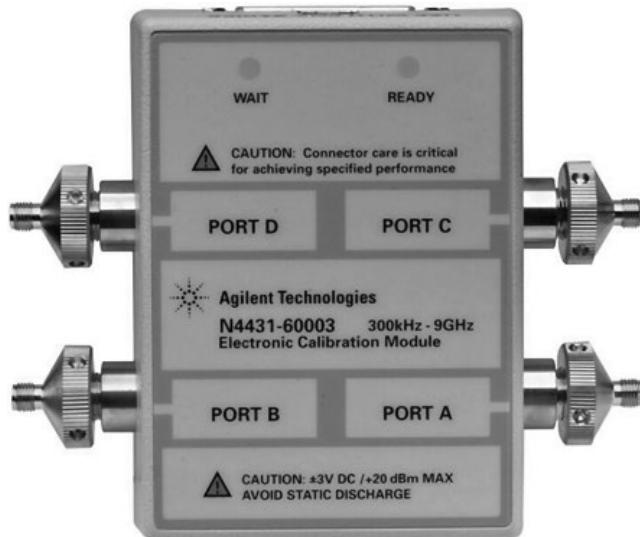


Figure 5.16 E-Cal Module N4431-60003. (Source: Reproduced with permission of Agilent Technologies.)

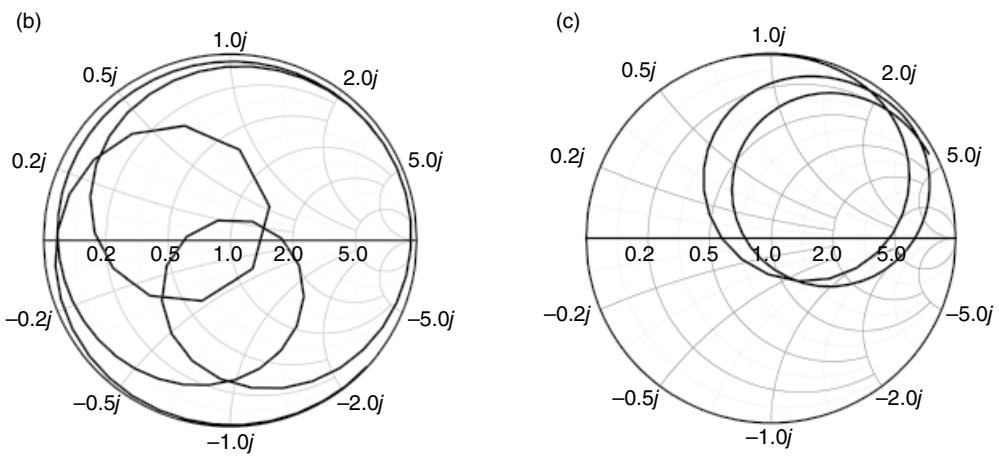
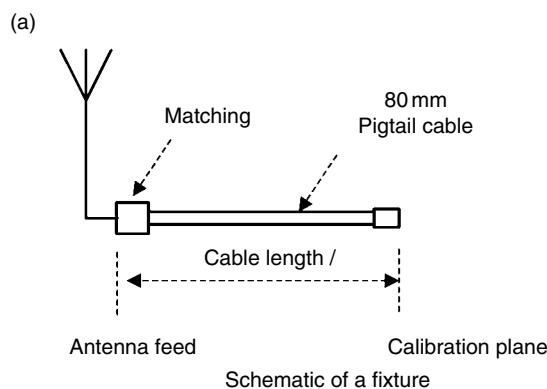


Figure 5.17 Impedance of a test fixture.

When we design an antenna, what we want to know is the complex impedance at the feeding point or at the matching circuit. To move the reference plane from the calibration plane to the antenna feed, a port extension function is required. Fortunately, most VNAs have a port extension as a built-in function, which moves the calibration reference plane by specifying the electrical delay. Mathematically, the effect of a pigtail cable on an antenna's impedance can be expressed by Equation 5.2:

$$\Gamma_{\text{Cal}} = \Gamma_{\text{ant}} \cdot e^{-j2\pi(l/\lambda)} \quad (5.2)$$

Here, l is the length of the pigtail cable, λ is the wavelength, Γ_{ant} is the actual antenna reflection coefficient, and Γ_{Cal} is the measured reflection coefficient. What a port extension does is multiply Γ_{Cal} by an $e^{-j2\pi(l/\lambda)}$ factor, so Γ_{ant} can be restored.

Shown in Figure 5.18 is the screenshot of the port extension menu of an Agilent VNA E5071B. It is under\Cal\Port Extension. A port extension can be carried out separately on each port. VNAs from different manufacturers may have different menu arrangements. Some VNAs even have an automatic port extension function. You should refer to the respective user manual accordingly. As shown in Figure 5.18, five steps are required to properly set the port extension on an E5071B.

The unit used in port extension can be either length or time, which means how long the pigtail cable is or how much time an electromagnetic wave needs to travel through the cable. You should never measure the length of a cable by a ruler and manually input the value. The inherent assumption used here is that the cable is air-filled and a signal travels at the speed of light in vacuum. As the semirigid cable used in a fixture is filled with dielectric material, the assumption is not valid. In practice, a port extension has two steps. First, we disconnect the antenna and make a short circuit at the feed or the matching circuit. Second, we adjust the port extension value until the whole impedance curve is concentrated into a tiny area around the zero impedance point on the Smith chart. If the frequency band is quite wide, say, from 800MHz to 6GHz, the impedance curve may spread a little bit. That is normal.

If you have several fixtures, each fixture must have their own port extension value. A good way to avoid making mistakes is to save the setting of each fixture in a separate file. Each time before you start testing, load the setting file for the specific fixture. If a choke is used with a fixture, the choke should be attached to the flexible cable when doing the calibration and port extension, as the choke also has a certain length. It is good practice for each engineer to have a dedicated choke. That avoids the potential error caused by different choke lengths.

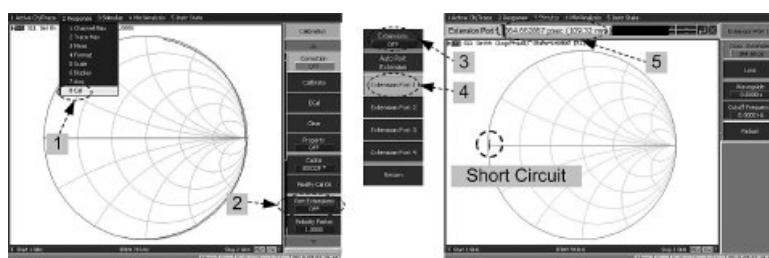


Figure 5.18 Port extension interface.

5.1.3 Passive Chamber Measurement

The best environment to measure an antenna's efficiency is in free space. However, this is not practical, so the anechoic chamber was introduced. "Anechoic" means nothing bounces back from the chamber wall. In an ideal anechoic chamber, all electromagnetic wave travels outward and nothing is reflected back, which is just what will happen in free space. An anechoic chamber is a metal shielding room with all the internal surfaces covered with radiation absorbent material (RAM). The RAM can absorb most incident electromagnetic waves. Detailed information of anechoic chamber technology can be found in books [7–9]. Only the essential knowledge is introduced here.

Shown in Figure 5.19 are a rectangular and a tapered chamber. Both of them are widely used in the mobile phone industry. In most chambers, the transmitting horn antenna and the DUT are located along the axis of the longest dimension. In Figure 5.19b, the cone on the end of tapered section is the transmitting horn and the antenna under test is at the center of the rectangular section. Compared with the tapered chamber, the rectangular chamber is more compact and its construction cost is less. However, its working frequency range is narrower and its overall electrical performance is also inferior.

The RAM is designed and shaped to absorb incident RF radiation, as effectively as possible, from as many incident directions as possible. Based on the principle of physics, at an interface of two different dielectrics, some portion of an incident electromagnetic wave will be reflected and some of it will pass through. The reflection coefficient depends on the difference between the dielectric properties of two dielectrics. The larger the difference, the more reflected energy there is. As the whole purpose of RAM is to absorb incident energy instead of reflecting it, we must minimize the difference between two dielectrics. The dielectric on one side is air, which has a permittivity of 1. On the other side, the absorbing material is made of rubberized foam, which is composed of a large amount of air and its permittivity is close to 1. Some people might think that the absorber is made of solid plastic or rubber, but now it is obvious why it cannot be.

To function as an absorber, low permittivity is not the only requirement. The RAM must have appropriate lossy characteristics. The RAM needs to adequately attenuate the electromagnetic wave before the wave reaches the metal shielding wall behind it, where the wave gets reflected back. By doping rubberized foam with controlled mixtures of carbon and/or iron, the desired absorbing property can be achieved.

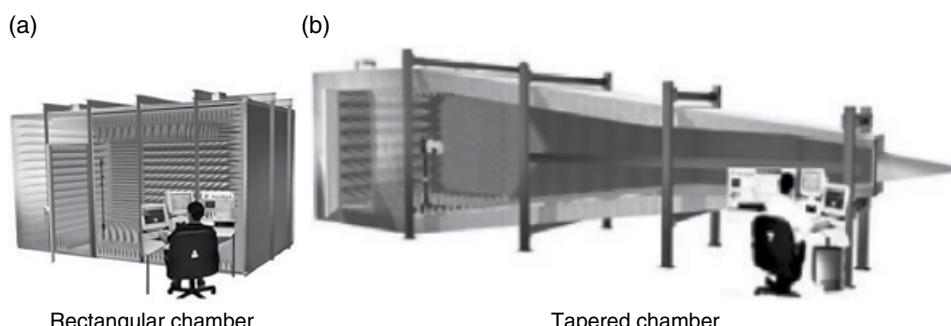


Figure 5.19 Two kinds of chambers. (Source: ETS-Lindgren Inc.)

Shown in Figure 5.20 are two commonly used absorbers. One is pyramid shaped and the other one is wedge shaped. The wedge absorber has been used in the tapered chamber shown in Figure 5.19b. You should be able to see them in the tapered section of that chamber.

Now let's explain why the absorber has such a weird form factor. Although the RAM is designed to have as little reflection as possible, there is still some residual reflection. For a normal anechoic chamber, the reflection from the wall should be less than -40 dB , which is well below a flat RAM's reflection level. By forming the RAM into a wedge shape as shown in Figure 5.21, most reflected energy bounces into the surface of the adjacent cone or ridge. Each time the electromagnetic wave hits an interface, most of the energy will penetrate into the RAM and be attenuated. After several bounces, only a tiny portion of the incident energy is still left.

A pyramid absorber can be used in any location in a chamber. A wedge absorber can only be used when the incident plane of wave approximately aligns to the wedge's edge. If the incident wave is orthogonal to the edge, the reflection from the absorber will increase due to the diffraction from the edge. An absorber has the best absorbing capability when the wave

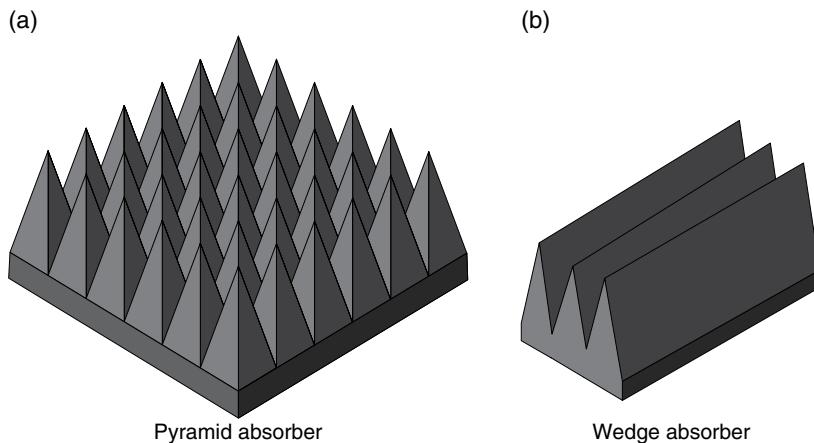


Figure 5.20 Radiation absorbent material. (Source: Reproduced with permission of Microwave Vision.)

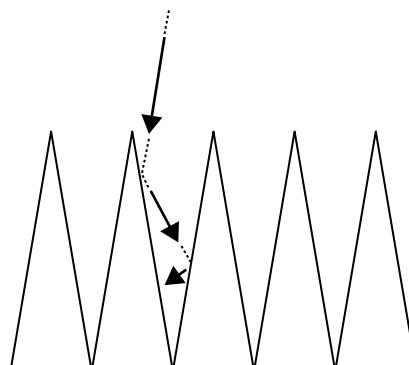


Figure 5.21 How an absorber works.

incidents come from the normal direction. The case shown in Figure 5.21 is such a good working condition, which has large incident angle. The effectiveness of an absorber starts to decrease when the incident angle gets smaller. At small incident angle, a wedge absorber always has better performance than a pyramid one. If we revisit the tapered chamber shown in Figure 5.19b, it should be clear why the wedge absorber is used at the tapered section.

The thickness, which is from the tip to the base, of an absorber is decided by the wavelength of its lowest working frequency. For an absorber working in a frequency as low as 800 MHz, the height is around 30–40 cm range. Of course, the bigger absorber is always better, but it also costs more.

Because the absorber is doped with carbon, its natural color is black. A layer of paint, normally blue, is sprayed on the surface of absorbers to increase their light reflection. Without that layer of paint, an anechoic chamber would be too dark to work in.

Shown in Figure 5.22 is a standard test setup in a two-dimensional (2D) chamber. A dual-polarization horn antenna serves as the transmitting antenna. A phone fixture, which is installed on a rotation table, serves as the receiver. As the system is reciprocal, so the measured value should be the same if the phone serves as the transmitter and the horn as the receiver. A network analyzer, which measures the transmission loss between the horn and the phone fixture, is at the heart of the measurement system.

To evaluate an antenna, radiation patterns of both horizontal and vertical polarization must be measured. For most chambers used in the phone industry, the dual polarization horn is the standard. A single-pole double-throw (SPDT) switch can be used to change between two polarizations. For chambers which require wider bandwidth and better polarization purity, single-polarization horns are normally used. In such a system, the phone fixture must be measured twice with the horn rotated 90° between measurements. In a dual-polarization horn, the pair of vertical ridges generates the vertical polarization. The horizontal ridges correspond to the horizontal polarization.

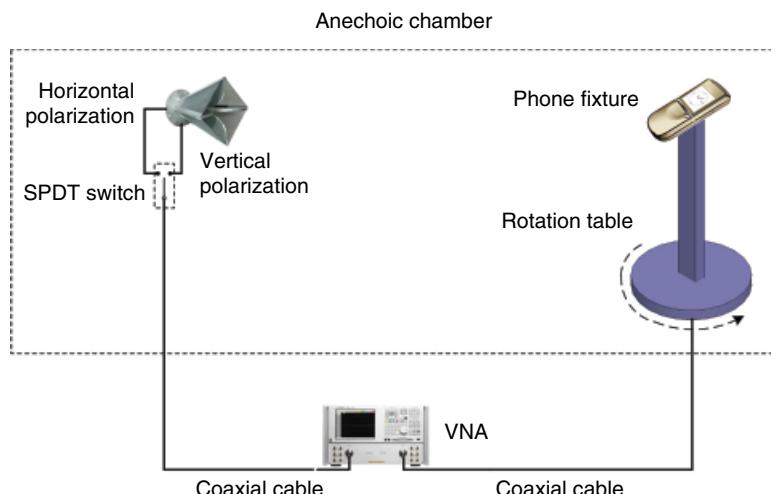


Figure 5.22 Simplified block diagram of a 2D test setup.

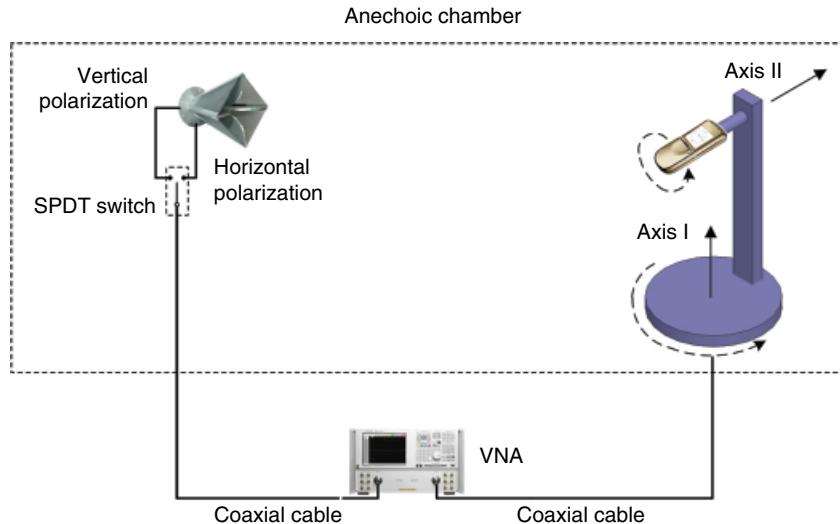


Figure 5.23 Simplified block diagram of a 3D test setup.

A 2D chamber is very useful for measuring high-gain antennas, such as reflector antenna and array antenna. However, in the phone industry, the three-dimensional (3D) chamber is the standard. Several measurements, which are required in the process of certification, must be carried out in a 3D chamber. Shown in Figure 5.23 is a 3D chamber. The only difference between a 2D chamber and a 3D chamber is the rotation table. On a 3D rotation table, a phone fixture can rotate in the azimuth plane and it can also revolve around itself.

The coordinates' definition in a 3D chamber might be very confusing. For most far-field 3D chambers on the market, the Axis II is the Z-axis. Rotation around Axis I corresponds to the elevation angle θ and around Axis II corresponds to the azimuth angle ϕ . When measuring a 3D radiation pattern, the rotation table rotates around Axis I from 0° to 180° and the rotation arm revolves around Axis II between 0° and 360° . The moving speed of the whole rotating table and the rotating arm is quite different. By distributing most movement to Axis II, the total measurement time can be minimized. In a 3D chamber, the definition of a horn's polarizations also changes. The vertical E field actually corresponds to the horizontal polarization and vice versa.

Shown in Figure 5.24 is the measurement sphere of the phone shown in Figure 5.23. The sphere is also called the “conical” cut sphere. The moving step of Figure 5.24 is 15° . The circles around the sphere are the measurement paths. For each turn around Axis II, the measurement path is a full circle on the sphere. On each circle, there are evenly distributed test locations, which are also 15° apart from each other and are not marked in the figure. At each test location, an antenna's responses to both horizontal and vertical polarizations are measured.

Both 2D and 3D chambers discussed before are far-field chambers. It means that the distance between the horn and the phone fixture is far enough, and the sphere wave at the fixture can be approximated by a planer wave. Shown in Figure 5.25 is a near-field chamber [10], which is normally known as the Satimo chamber. The raw data measured by a Satimo chamber is near-field data. A computer code is used to transform near-field data to a far-field result.



Figure 5.24 Measurement 3D sphere (15° step).

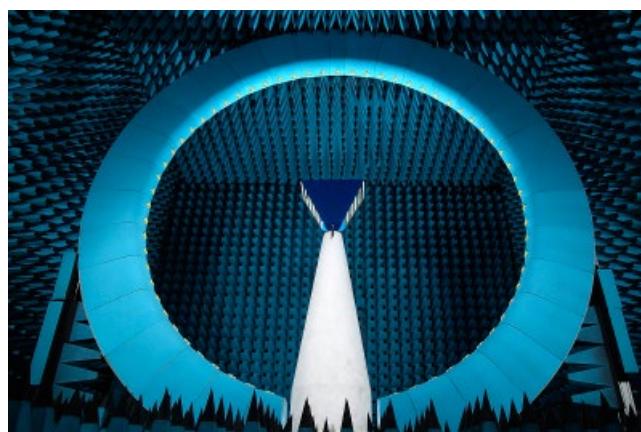


Figure 5.25 Near-field chamber. (*Source:* Reproduced with permission of Microwave Vision.)

One unique feature inside a Satimo chamber is the arch. The arch is composed of an array of probes, which are evenly distributed on the circumference and wrapped by absorbing material. Each probe can detect two orthogonal polarizations. The white column in the middle, which is made of foam, is to support phone fixtures. In Figure 5.25, a horn antenna was placed on the support as the DUT. A testing cable runs through the middle of the column to connect fixtures. In a Satimo chamber, the horizontal plane is the azimuth one and the vertical plane is the elevation one. To get a pattern of elevation cut, the system only needs to electronically scan the array of probes. To get a full 3D pattern, the system rotates the foam column from 0° to 180° and scans the probes accordingly. As electronic scanning is faster than a mechanical scanning, the primary scanning plane is the elevation plane.



Figure 5.26 Measurement 3D sphere of a Satimo chamber (15° step).

Shown in Figure 5.26 is the measurement sphere of a Satimo chamber. The moving step is 15°. The sphere is also called the “great circle” cut sphere. This measurement sphere is more intuitive. The 3D spheres shown in Figures 5.24 and 5.26 may seem different; the actual test locations on both spheres are identical. The only difference is the measurement path which connects all test locations.

Both kinds of chambers can be qualified to carry on CTIA certification tests [11], which means they have adequate accuracy. In normal operation, a multiprobe system is faster when doing measurement. However, a multiprobe system must be calibrated more frequently, as it has more probes and is more sensitive to environment variation, such as temperature fluctuation. On the other hand, a standard far-field chamber is more stable, but it also takes more time to do the measurements.

A 3D passive anechoic chamber can measure various antenna parameters, such as efficiency, gain, axial ratio, and beamwidth. All those parameters are deduced from the same set of measured raw data. Of all the parameters, efficiency is the most important one for a cellular antenna. The efficiency definition used in the cellular antenna industry is shown in Equation 5.3.

$$\begin{aligned}
 \text{Eff} &= \frac{P_{\text{radiated}}}{P_{\text{available}}} \cdot 100(\%) \\
 &= 10 * \log_{10} \left(\frac{P_{\text{radiated}}}{P_{\text{available}}} \right) (\text{dB}) \\
 &= 1 - \text{LOSS}_{\text{mismatch}} - \text{LOSS}_{\text{matching circuit}} - \text{LOSS}_{\text{antenna}} \text{ (normalized)}
 \end{aligned} \tag{5.3}$$

Efficiency is defined by the ratio between the TRP and the available power at the mating connector of the testing cable. The efficiency value includes the contribution of

Table 5.2 Conversion table of efficiency between percentage and dB

Antenna efficiency	
%	dB
25	-6
50	-3
~80	-1

the impedance mismatch loss, the matching circuit loss, and the conductive and dielectric loss from the antenna and the phone structure. Besides using percentage (%) as the unit when referring to efficiency, the decibel (dB) is also widely used. Shown in Table 5.2 is the corresponding list.

In the cellular phone business, 50% or -3 dB is the gold standard for the antenna efficiency. When a project is accomplished, if the antenna has efficiency better than -3 dB across all bands, you have reason to smile. In some literatures, you might see authors claim antennas' efficiency better than 95%. The efficiency definition they used, which does not include mismatch loss, is a legitimate one but is different from Equation 5.3.

Besides efficiency, gain is the other frequently used parameter to describe an antenna's electrical characteristic. However, gain is not really important when designing cellular antennas, as it is not solely decided by antennas. Comparing the antenna gain of different phones is not meaningful. In the cellular application, as a phone's orientation and base stations' locations are all arbitrary, high-gain antennas do not necessarily mean better performance. When an antenna has higher gain in a certain direction, it must have deeper null in other directions. As there is no way to continuously point the peak of antenna's radiation pattern toward a base station, the overall system performance of a high-gain antenna is actually worse. As a rule of thumb, as the radiation pattern at lower band is more uniform, the lower band's gain is always less.

When both a phone's form factor and its antenna's shape have been determined, the gain of the antenna is highly correlated to its efficiency. To measure an antenna's efficiency, a 3D pattern data must be collected, which is quite time-consuming. As an alternative, measuring gain is a better way when you need to assess several antenna samples. For an antenna with known radiation peak, you can use chamber software to point the peak toward the testing horn antenna. The peak gain of most cellular antennas is quite low, which means the pattern's beamwidth is wide, which in turn means the gain measurement is not sensitive to slight positioning variations. To assess 20 antenna samples, the efficiency measurement needs a few hours to half a day, the gain measurement can be accomplished in several minutes. There is a precondition when using the gain method to substitute efficiency measurements, which is that normalized patterns of all antenna samples must be consistent. If a phone has intermittent connection between metal objects, the radiation patterns might change between tests. In such circumstances, the efficiency measurement may be the only reliable method.

5.2 Active Antenna (Over the Air) Measurement

In a passive antenna measurement, an antenna is isolated from the active system, which gives us a clean environment to design, debug, and optimize the antenna. The passive measurement is a component-level test. As a product, cellular phones are always evaluated by customers as a whole. A customer is not really interested in how well an antenna works. Similarly, cellular phone service providers only evaluate each phone as a whole system. The active antenna test, which is often referred to as the over-the-air (OTA) performance testing, is the system-level chamber test accepted by service providers. OTA tests determine how a specific network will influence the connectivity performance of a mobile handset. The tests reduce the need to perform field tests for connectivity and also allow operators to rapidly evaluate new products.

Different standard organizations have released several versions of OTA test guidelines. Of them, the US-based CTIA—The Wireless Association, was the first to set up the test framework [12]. CTIA stands for Cellular Telephone Industries Association, it changed its name to “CTIA—The Wireless Association” in 2004 to better represent the expanding wireless industry. Europe’s 3GPP/ETSI also has a version of the OTA test standard [13]. The 3GPP stands for Third-Generation Partnership Project. ETSI stands for European Telecommunications Standards Institute. As most 3GPP’s standards are identical to ETSIs, they are normally counted as one. Both CTIA and 3GPP/ETSI are international organizations. In some countries, domestic versions of the OTA test standard are used. Using China as an example, the “YD/T 1484-2006 Measurement Method for Radiated RF Power and Receiver Performance of Mobile Stations” is the official standard.

From the technical point of view, all those standards are pretty much the same. Some differences, which are technically insignificant, are deliberately introduced for political reasons. In this section, the principle of measurements is the emphasis. For more information, refer to the corresponding documents [12–14].

5.2.1 EIRP, ERP, and TRP

Effective isotropic radiated power (EIRP), effective radiated power (ERP), and total radiated power (TRP), are all parameters used to evaluate a cellular phone’s performance as a transmitter. The unit normally used is dBm, which is an abbreviation for the power ratio in decibels of the measured power referenced to 1 mW.

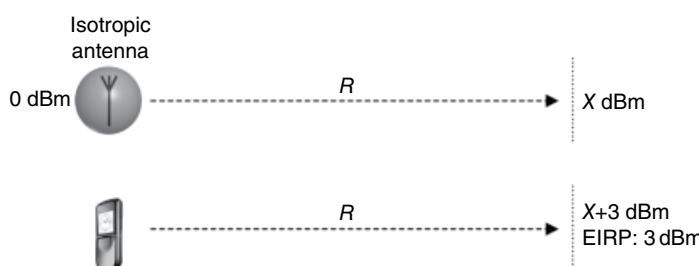


Figure 5.27 Illustration of EIRP.

Shown in Figure 5.27 is an illustration of EIRP. Assume there is a 0dBm signal radiated from an isotropic antenna and it travels outward in an ideal free space. Obviously, the isotropic assumption is a theoretical one, as there is no such antenna in the real world. At a given distance R , the theoretical received signal from the isotropic antenna is X dBm. Now replace the isotropic antenna with a phone, if the power of the received signal increases by 3dB, the EIRP of the phone at that angular direction is 3dB above 0dBm, which equals 3dBm. The EIRP is a relative measurement, so no matter what distance R is, the EIRP of a phone is always the same.

As a phone is not isotropic, the power it radiates toward different directions is also different. As a result, a phone's EIRP is a variable which depends on the test direction and a phone's orientation. However, usually when people refer to EIRP, it actually means peak EIRP. The relation between a device's peak EIRP, conductive power, and gain can be written as follows:

$$\text{EIRP}_{\text{peak}} (\text{dBm}) = P_{\text{cond}} (\text{dBm}) + \text{Gain} (\text{dB}) \quad (5.4)$$

The conductive power is the power available at the switch connector. Based only on a phone's peak EIRP, we are not able to tell how good the antenna is. It is possible to get the same peak EIRP by increasing the conductive power of a phone which has a low-efficiency antenna.

ERP is pretty much the same as EIRP, except that it uses a half-wavelength dipole as the reference antenna instead of an isotropic one. ERP is numerically 2.14dB, which is the gain value of a dipole antenna, less than EIRP. Some people might wonder why we need both ERP and EIRP. This is for historical reasons. In the 2D chamber era, dipoles were normally used as the calibration standard, calculating ERP from the difference between an antenna and a dipole is a little bit easier than calculating EIRP. After the wide adoption of 3D chambers, the EIRP is more frequently used. It is a strange phenomenon that some service providers use ERP as the 850MHz band specification and EIRP as the 1900MHz band one. Their aim is to make the specification back-compatible. The history between ERP and EIRP is analogous to that between dBd and dBi, which are normalized to a dipole and an isotropical radiator, respectively.

Today, the most frequently used parameter for evaluating transmitter is neither ERP nor EIRP, but it is TRP. Just as its name implies, the TRP represents how much power a phone radiates into free space. It is the integral of the EIRP over the surface area of a hemisphere.

$$\text{TRP} = \frac{1}{4\pi} \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} [\text{EIRP}_{\theta}(\theta, \phi) + \text{EIRP}_{\phi}(\theta, \phi)] \sin(\theta) d\phi d\theta \quad (5.5)$$

Here, EIRP_{θ} and EIRP_{ϕ} are theta and phi polarizations of EIRP. The theta polarization is also called the “vertical polarization” and the phi one is called the “horizontal polarization.” The units of both are absolute power. In the real world, TRP is calculated from discrete measurement data. For a complete sphere measured with N theta intervals and M phi intervals, both with even angular spacing, the TRP can be written as follows:

$$\text{TRP} \approx \frac{\pi}{2NM} \sum_{i=1}^{N-1} \sum_{j=0}^{M-1} [\text{EIRP}_{\theta}(\theta_i, \phi_j) + \text{EIRP}_{\phi}(\theta_i, \phi_j)] \sin(\theta_i) \quad (5.6)$$

In the CTIA test plan, the angular step of a TRP test is 15° . No data need be recorded at positions corresponding to $\theta=0^\circ$ and 180° , nor at positions corresponding to $\phi=360^\circ$ ($\phi=0^\circ$ data

are recorded), because those points are not used. For transmit tests with $N=12$ and $M=24$, this means that only 11 theta cuts and 24 phi cuts, or 264 measurements in each polarization, need be taken.

Besides the integral formula, there is a more intuitive way to define TRP.

$$\text{TRP} = P_{\text{cond}} \text{ (dBm)} + \text{Efficiency (dB)} \quad (5.7)$$

The unit of TRP here is dBm. The difference between the conductive power and the TRP is the antenna efficiency. TRP is always lower than the conductive power. Now it is clear why people like to use dB as the unit when talking about an antenna's efficiency. When a project manager breaks down a phone's specification, the starting point is the TRP, which is provided by various service providers as a mandatory requirement. The P_{cond} is the sub-specification for RF engineers who are responsible for transceivers. The efficiency is for antenna engineers. Of course, project managers always reserve some "mark up" for themselves, in case something goes awry. The actual target is higher than the mathematically calculated one.

There are two other parameters [12], NFPRP ± 45 and NFPRP ± 30 , which are included in CTIA format reports but receive little attention from antenna engineers. HRP stands for near-horizon partial radiated power. One integrates EIRP between $\pm 45^\circ$, and the other one between $\pm 30^\circ$. They can be calculated from the same set of test data used by the TRP formula.

$$\begin{aligned} \text{NFPRP}_{\pm 45} &\approx \frac{\pi}{2NM} \left(\frac{\text{cut}_3 + \text{cut}_9}{2} + \sum_{i=4}^8 \text{cut}_i \right) \\ \text{NFPRP}_{\pm 30} &\approx \frac{\pi}{2NM} \left(\frac{\text{cut}_4 + \text{cut}_8}{2} + \sum_{i=5}^7 \text{cut}_i \right) \\ \text{cut}_i &\approx \sum_{j=0}^{M-1} \left[\text{EIRP}_\theta(\theta_i, \phi_j) + \text{EIRP}_\phi(\theta_i, \phi_j) \right] \sin(\theta_i) \end{aligned} \quad (5.8)$$

Shown in Figure 5.28 is a TRP test setup. In a TRP test, a phone is installed on the 3D rotation table without any cable attached. A base station simulator, which can establish a live connection with the phone and set the phone to the maximum transmitting power, is connected to a link antenna near the phone. The equipment which carries on the actual power measurement is a spectrum analyzer. The power of both the vertical and horizontal polarization fields is measured by switching an SPDT switch. In this setup, the position of the link antenna and the transmitting power of base station simulator are not important, as long as they can keep the wireless link connected during tests. As most phones are linear polarized, a circular polarized log spiral antenna is used as the link antenna in the test configuration shown in Figure 5.28. If a linear polarized antenna is used as the link antenna, the connection might break when the phone under test and the link antenna are cross-polarized.

The actual TRP specification depends on many factors; the following is a partial list:

- Service providers: Vodafone, Verizon, AT&T, China Mobile, and so on
- Technologies: AMPS, GSM, CDMA, WCDMA, CDMA2000, and so on
- Bands: 850 MHz, 900 MHz, DCS, PCS, UMTS, and so on
- Contents: voice, data, and so on

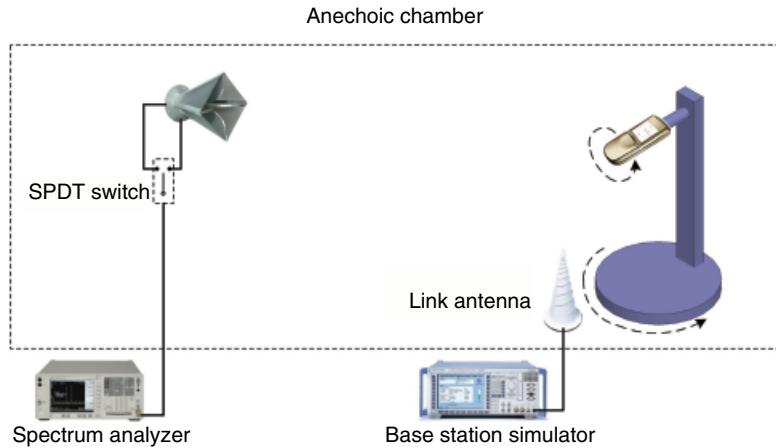


Figure 5.28 Simplified block diagram of a TRP test setup.

- Configurations: flip open/close, slide up/down, and antenna extended/retracted
- Positions: free space, with phantom head, with both phantom hand and head

Normally, only low, middle, and high channels are measured at each band. Depending on which market the antennas are being designed for, the TRP specification will be different. On the other hand, the TRP specifications are being continuously revised. For example, only free space TRP was required initially; then the phantom head was added; now people are discussing adding a new specification which requires a measurement with both phantom head and hand attached. Some typical TRP specifications can be found in the CTIA test plan [12].

From Equations 5.4 and 5.7, it is clear that the relation between a peak EIRP and a TRP is analogous to the one between gain and efficiency. Similarly, when both a phone's form factor and its antenna's shape have been determined, the peak EIRP and the TRP are highly correlated. However, the TRP requires a full 3D measurement and an EIRP can be obtained by a single-point measurement. The peak gain of most cellular antenna is quite low, which means the pattern's beamwidth is wide, which in turn means the EIRP measurement is not sensitive to slight positioning variation. To assess 20 antenna samples, the TRP measurement needs at least half a day, while the EIRP measurement can be done in several minutes. Before using the EIRP method to replace TRP measurements, remember that normalized patterns of all antenna samples must be consistent. If a phone has an intermittent connection between metal objects, the radiation patterns will change between tests. Under such circumstances, the TRP measurement is the only reliable method.

5.2.2 EIS and TIS

Effective isotropic sensitivity (EIS) and total isotropic sensitivity (TIS) are two parameters used to evaluate a cellular phone's performance as a receiver. EIS and TIS are analogous to EIRP and TRP. The EIS test method is similar to the EIRP method; in that, the range reference measurement is used to correct the unknown performance of a device back to values relative

to that of a theoretical isotropic receiver. In this case, the correction value offsets each sensitivity level measurement back to the equivalent sensitivity level of a theoretical isotropic receiver exposed to an incoming isotropic wave with the same magnitude. In the cellular industry, the unit normally used for EIS and TIS is dBm (decibels above 1 mW). In other wireless industries, such as broadcasting, dB μ V (decibels above 1 μ V) is more frequently used. For a 50Ω system, the relation between dBm and dB μ V is as follows:

$$\text{dB}\mu\text{V} = \text{dBm} + 107\text{dB} \quad (5.9)$$

Assuming that no interference was introduced into the receiver link, the relation between a device's peak EIS, conductive sensitivity and gain can be written as follows:

$$\text{EIS}_{\text{peak}}(\text{dBm}) = \text{Sensitivity}_{\text{cond}}(\text{dBm}) - \text{Gain}(\text{dB}) \quad (5.10)$$

For sensitivity, a lower value is better. It means that a receiver can detect a weaker signal. For example, a receiver of -105 dBm sensitivity is 10 dB more sensitive than a -95 dBm one. When a receiver of -95 dBm conductive sensitivity is connected to an antenna of 5 dB gain, the overall system's sensitivity is improved by 5 dB to -100 dBm. As a consequence, the sign before the gain is a minus in the EIS formula instead of a positive in the EIRP formula.

By integrating EIS across the total spherical surface, the TIS can be determined as follows:

$$\text{TIS} = \frac{4\pi}{\int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \left(\frac{1}{\text{EIS}_{\theta}(\theta, \phi)} + \frac{1}{\text{EIS}_{\phi}(\theta, \phi)} \right) \sin(\theta) d\phi d\theta} \quad (5.11)$$

Here, EIS_{θ} and EIS_{ϕ} are theta and phi components of EIS. As a smaller value is more significant from the sensitivity point of view, it should contribute more to the final TIS result. As shown in Equation 5.11, the reciprocals of both EIS components are integrated first. In this way, the smaller value contributes more to the integral. The final TIS value is the reciprocal of the integral.

In reality, TIS is calculated from discrete measurement data. For a complete sphere measured with N theta intervals and M phi intervals, both with even angular spacing, the TIS can be written as follows:

$$\text{TIS} \approx \frac{2NM}{\pi \sum_{i=1}^{N-1} \sum_{j=0}^{M-1} \left[\frac{1}{\text{EIS}_{\theta}(\theta_i, \phi_j)} + \frac{1}{\text{EIS}_{\phi}(\theta_i, \phi_j)} \right] \sin(\theta_i)} \quad (5.12)$$

In the CTIA test plan, the angular step of a TIS test is 30° . No data need be recorded at positions corresponding to $\theta=0^\circ$ and 180° , nor at positions corresponding to $\phi=360^\circ$. For receiver tests with $N=6$ and $M=12$, this means that only 5 theta cuts and 12 phi cuts, or 60 measurements in each polarization, need be taken.

The reason for using 30° as the angular step in TIS tests while using 15° in TRP tests is that TIS tests is very time-consuming. Even with the 30° step, a TIS test, which includes low, middle, and high channels of a few bands, can take hours.

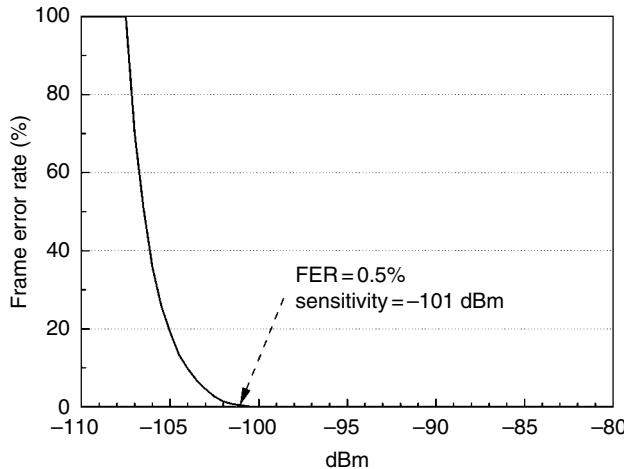


Figure 5.29 FER vs. signal power.

Shown in Figure 5.29 is the frame error rate (FER) curve of a sample device. According to the CTIA test plan, sensitivity measurements should be equivalent to the minimum RF power level that results in an FER of 0.5% or less with 95% confidence at each measured location on the sphere. For the phone shown in Figure 5.29, its sensitivity at the measured direction is -101 dBm. When the signal decreases to -107 dBm, the FER quickly degrades to 100%. It is very easy to drop the connection when the FER is higher than 50%. As in the existence of deep nulls in a radiation pattern and high isolation between cross-polarizations, the dynamic range of the EIS of a phone can easily exceed 20 dB. When doing an EIS test, it is desirable to maintain the connection during the whole test. The base station simulator must gradually dial down its power level from a power level 30 dB higher than the estimated TIS value, and repeat FER tests along the way. The power level cannot be decreased too rapidly, otherwise the connection might break. Telling whether a chamber is doing a TRP test or a TIS test is a foolproof task. Most of the time during a TIS test, the rotation table seems to be stuck in a fixed position.

There are two other parameters [12], NHPIS ± 45 and NHPIS ± 30 , which are included in CTIA format reports but receive little attention from antenna engineers. NHPIS stands for near-horizon partial isotropic sensitivity. One integrates EIS between $\pm 45^\circ$, the other one between $\pm 30^\circ$. They can be calculated from the same set of test data used by the TIS formula.

$$\begin{aligned}
 \text{NHPIS}_{\pm 45} &\approx \frac{2NM}{\pi \left(\frac{1}{8}(\text{cut}_1 + \text{cut}_5) + \frac{9}{8}(\text{cut}_2 + \text{cut}_4) + \text{cut}_3 \right)} \\
 \text{NHPIS}_{\pm 30} &\approx \frac{2NM}{\pi \left(\frac{1}{2}(\text{cut}_2 + \text{cut}_4) + \text{cut}_3 \right)} \\
 \text{cut}_i &\approx \sum_{j=0}^{M-1} \left[\frac{1}{\text{EIS}_\theta(\theta_i, \phi_j)} + \frac{1}{\text{EIS}_\phi(\theta_i, \phi_j)} \right] \sin(\theta_i)
 \end{aligned} \tag{5.13}$$

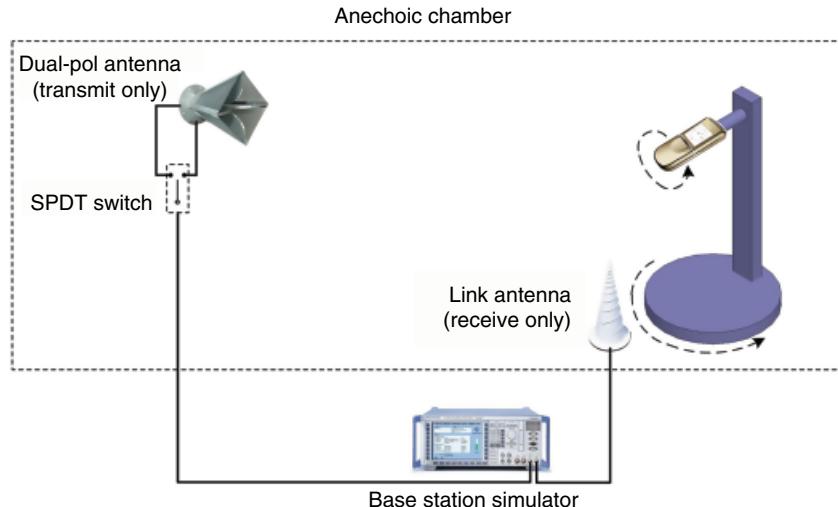


Figure 5.30 Simplified block diagram of a TIS test setup.

Shown in Figure 5.30 is a TIS test setup. A phone is installed on the 3D rotation table without any cable attached. A base station simulator must operate in the two-port mode. Its output port is connected to the dual-polarized horn antenna, while its input port is connected to a link antenna near the phone. The EIS of both vertical and horizontal polarizations is measured by switching an SPDT switch. In this setup, the position of the link antenna is not important as long as it can keep the wireless link connected during tests.

5.2.3 Sensitivity Degradation Due to Interference

If a phone's antenna and its transceiver are well matched, the phone's TRP is mostly decided by its conductive power and the antenna's efficiency. However, normally, a phone's sensitivity is not solely decided by its conductive sensitivity and the antenna's efficiency. In a real phone, the peak EIS formula shown in Equation 5.10 needs to be rewritten as follows:

$$\text{EIS}_{\text{peak}} (\text{dBm}) = \text{Sensitivity}_{\text{radiated}} (\text{dBm}) - \text{Gain} (\text{dB}) \quad (5.14)$$

Here, the $\text{Sensitivity}_{\text{cond}}$ of Equation 5.10 is replaced by $\text{Sensitivity}_{\text{radiated}}$. The radiated sensitivity of a phone is always worse or at best equal to the conductive sensitivity. The term for the degradation of sensitivity is “desense.” Desense means that something is desensitizing a receiver. Receivers often have filters on the front of their circuits to help prevent unwanted signals from getting to the receiver and damaging its operation. In the case of desense, however, the signal is often internal or within the allowable frequency band of the filter. Desense is mostly a problem for the RF engineer. For more information, refer to books on the design of RF circuits or systems [15]. Only antenna-related desense issues are discussed in this book.

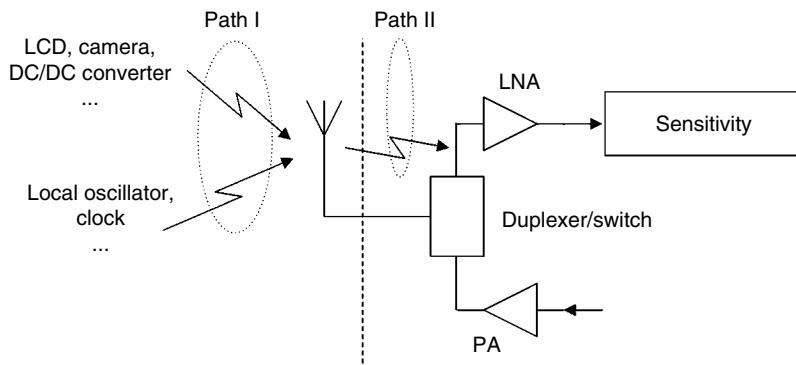


Figure 5.31 Some causes of desense.

Assuming the RF engineers have solved all the desense issues which can be measured in the conductive mode and the radiated sensitivity is still much worse than the specification, this means we have antenna-related desense issues. There are two dominant coupling paths which can cause desense. As shown in Figure 5.31, path I is directly antenna-related. Some components, such as LCDs, camera modules, and DC/DC converters, are known to have high noise across a wide-frequency spectrum and they introduce noise currents on the ground plane. In the absence of an antenna, those noise currents cannot affect a well-shielded receiver. When an antenna is installed, it is a different story. For most cellular antennas, the ground is part of antennas, so the noise will be picked up by the antenna. If those noises happen to fall in a phone's working band, the sensitivity of all channels will be affected. When that happens, a smooth inflation of EIS value across the whole band can be observed. For LCDs and cameras, the shield is the primary weapon in mitigating their effect. As LCDs and cameras are not required to work all the time, at least for 2G phones, so some level of desense can be tolerated. For some 3G phones, which support video calls, the desense issue must be handled more carefully. The primary noise source of a DC/DC converter is the μ H coil inductor. Shielding, solid ground, and high-frequency shunt capacitors can effectively mitigate desense caused by DC/DC converters.

The signals from local oscillators and various clock synthesizers are also a potential source for desense. As both of them are periodical signals, their noises' spectrum composes a series of discrete peaks. Their impact on the receiver is mostly constrained in selective channels. Besides shielding, using stripline, which sandwiches signal traces between two ground layers in a multilayer PCB, and placing two columns of grounded vias along the stripline are all useful techniques to eliminate unwanted noise.

Path II shown in Figure 5.31 affects systems where receiver and transmitter operate simultaneously. Such systems include CDMA, WCDMA, and CDMA2000. For a time division multiple-access (TDMA) system, such as GSM and Wi-Fi, as the transmitter and the receiver never work simultaneously, such a path is not an issue. Along path II, the high power signal from the antenna goes around the duplexer and goes back into the receiver link. This kind of desense is also called "self-jamming."

Shown in Figure 5.32 is an example of incorrect routing. The gray area is the ground on the backside of PCB. All traces and components are on the frontside. When measuring

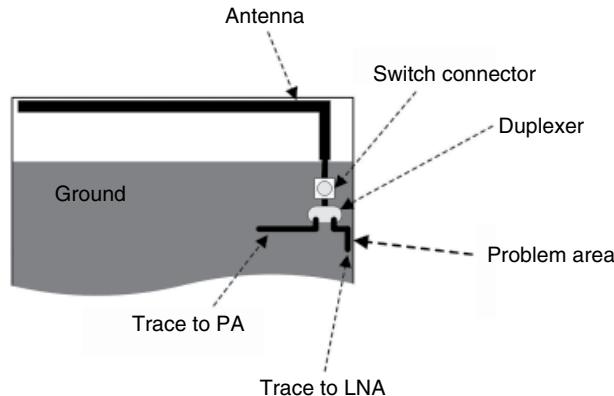


Figure 5.32 An example of incorrect routing.

this board conductively, everything may just be fine, because there is no direct path between the trace to PA (power amplifier) and the trace to LNA (low-noise amplifier). When measuring it in the radiative mode, it is most likely that self-jamming will be an issue. As we have previously discussed, the ground of a monopole antenna is part of the radiator and a strong current exists on the edge of the ground plane. In the example shown in Figure 5.32, some part of the trace to LNA is routed along the edge of the ground, which can pick up the high-power signal from the PA and cause self-jamming problems. If self-jamming is the root cause of desense, the level of sensitivity's degradation should correlate to the transmitting power. When the transmitting power is low enough, the desense seems to be fixed. To mitigate the self-jamming, the signal path of LNA must be well shielded.

In the CTIA test plan, only the TIS at low, middle, and high channels are fully measured. To catch the server desense issue, a relative sensitivity measurement must be done on all intermediate channels. Since some digital technologies have a very large number of intermediate channels, some channels may be omitted from actual testing provided there is no more than 500 kHz between any two successive intermediate channels. Shown in Figure 5.33 is the intermediate channel test according to the CTIA test plan. Based on the peak EIS of three fully tested channels, three segments pass/fail limit lines can be calculated. Each limit line is set to be 5 dB higher than the peak EIS of the corresponding reference channel. When measuring the intermediate channels, the transmitting power from the base station simulator is set according to the limit and the chamber positioner is moved to the location and polarization resulting in the best-radiated sensitivity according to the fully measured reference channel. For each channel, only one FER measurement is needed, the result is either pass or fail.

The curve shown in Figure 5.33 is an EIS result of a phone across all channels, which is not the intermediate channel test result. Two sparks marked by dashed circles are caused by desense. When doing an intermediate channel test, only the left spark can be detected and the right spark can pass the test without being noticed. If we want to fix these two sparks, the most likely suspects are local oscillators and clock traces.

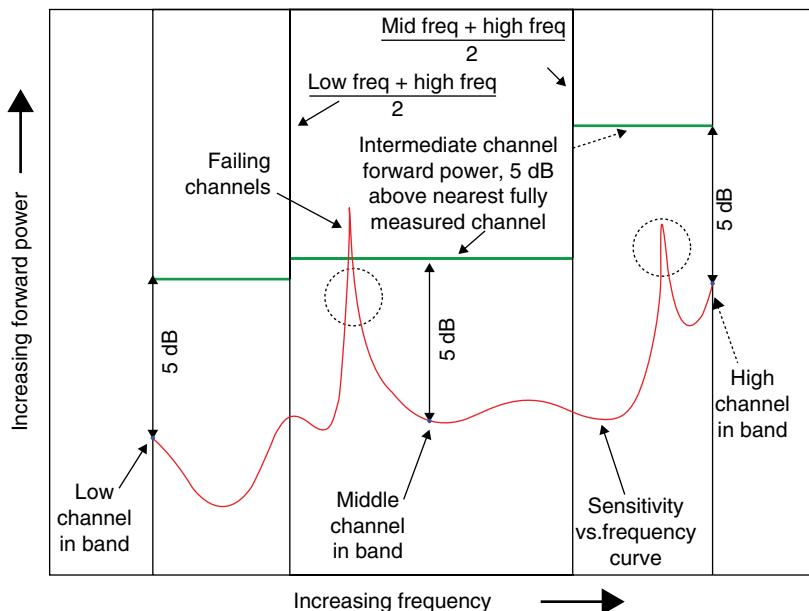


Figure 5.33 Illustration of intermediate channel test. (Source: CTIA.)

5.3 Antenna Measurements in the Production Line

As antenna engineers, our job responsibilities include not only designing antennas but also accompanying antennas through the whole production process. If you work for a phone company, you are on the customer side of the business and you may not need to do all the hands-on works related to the production. However, it is still good to know what is going on and be able to give clear instructions to your counterparts. If you work for an antenna or an original equipment manufacturer (OEM) company, you are on the vendor side and production measurements are certainly part of your job's responsibility.

In the production stage, a single model of an antenna can be manufactured in hundreds of thousands or even tens of millions. Some means of production control must be involved to ensure consistent quality. Some controls are mechanically related, such as dimension, visual quality, and antenna strength controls. Even if an antenna passes all mechanical inspections, that does not mean it will work properly when it is installed in a phone. Therefore, an electrical inspection is necessary on the production line.

Whether 100% of antennas need to be tested or a sampling test is enough is the call of the customer. Normally, for primary antennas, a 100% test rate is appropriate. A sampling test is enough for most single-band auxiliary antennas. The test equipments used in the production line are network analyzers. The SNAs, which is only capable of measuring amplitude, is good enough for line tests. SNAs are about half the price of entry level VNAs. Of course, if money is not an issue, VNAs also work fine. When using a network analyzer to design an antenna, we use a fixture made of a real phone. Each time we tweak the antenna, the fixture must be disassembled. On a production line, time is everything and the time spent in each test station is counted by seconds instead of minutes. Therefore, special antenna fixtures for antenna testing have been designed.

There are two ways of testing antenna modules. One is measuring the return loss (S_{11}) and the other is measuring the transmission loss (S_{21}). Designing a return loss fixture is quite straightforward. Most of the principles of designing R&D phone fixtures can be used here. For a production fixture, the most important consideration is consistency. Apart from the PCB and the necessary metal contacts, all the other components, such as the battery and the shielding box, should be excluded from the fixture. The life of a production fixture also needs to be considered. For example, some planar IFA (PIFAs), which are made of stamping metal, have built-in spring fingers serving as the contact mechanism. In a real phone, PCB pads are good enough to make a secure connection between a PCB and an antenna. If the same kind of pads are used in a production fixture, it is most likely that after several thousands of test cycles the fixture will be worn out. When making such a fixture, thin beryllium–copper pieces, which are cut according to the shapes of PCB pads, should be soldered on top of the pads. With the extra layer of metal, the life cycles of the fixture can be extended by several orders of magnitude. Pogo pins used in all phones have limited life cycles, so they shouldn't be used in production fixtures. For some phones, which have pogo pins on the PCB as the contact mechanism, spring fingers are a better choice when designing their fixtures.

If the relative position between an antenna and fixture varies, the antenna's response will change. Some alignment features, which keep an antenna in position, and some stopper posts, which maintain a constant distance between an antenna and a fixture, can all be used in fixtures' design. A good fixture should allow a well-trained operator to achieve a test cycle in less than 10 seconds.

It is quite difficult to guarantee that the fixture's connector is always at the sweat spot which can isolate the testing cable's influence, so some ferrite chokes should be used to minimize the possible impact from the cable. Touch the testing cable near the fixture with your hand; the antenna's response should be consistent with or without your hand's proximity.

For a return loss type fixture, shield boxes are not necessary. However, as antennas are radiating modules, any moving object should be kept out of the near-field region of the antenna under test. Shown in Figure 5.34 is a simplified illustration of a production test setup. The equipment of the setup is a VNA. The testing cable, which connects the VNA and the fixture, is omitted in the

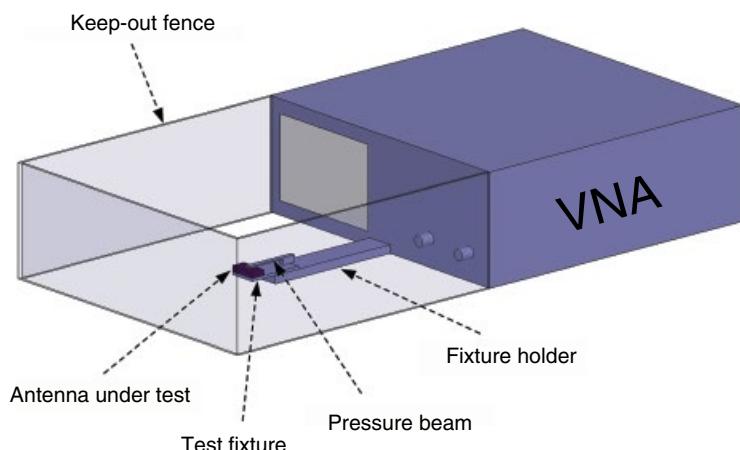


Figure 5.34 Simplified illustration of production test setup (S_{11} type).

illustration. The long board in front of the VNA is the fixture holder, which is secured on the test bench. The bench is also omitted in the illustration. One end of the holder extends out of the bench to provide a clear test environment for antennas. The test fixture is attached to the holder. The black module attached to the fixture is an antenna under test. A pressure beam provides the required contact force between an antenna and the fixture. Around the test fixture, there are three large semitransparent boards; they form the keep-out fence. An operator should keep both hands outside the fence when the VNA is collecting testing data. Apart from the test fixture, everything else should be made of plastic, such as polycarbonate or nylon.

In any test, there are always some inherent fluctuations. The S_{11} measures the mismatch between the impedances of a source and an antenna. The point which has the minimum return loss is normally defined as the resonant frequency. A slight antenna's impedance variation may affect both the measured resonant frequency and the amplitude of the return loss. To improve the test consistency, the S_{21} type test was introduced. Shown in Figure 5.35 is an S_{21} fixture. The fixture is composed of a shield box and two probes. Two probes are connected to the VNA's ports through two coaxial cables. In an S_{21} fixture, an antenna under test has no galvanic contact with either probe. They are electromagnetically coupled.

The principle behind S_{21} fixtures is the self-resonant modes. As is taught in electromagnetic courses [16–19], there exists a series of discrete eigenmodes for any resonators, and strong resonances can only be excited at those discrete modes. An antenna is a resonator, so it also has discrete modes. If a resonator is lossless, the 3 dB bandwidth of any its resonant mode will be 0 Hz. The more lossy a resonator, the wider the 3 dB bandwidth. To improve the measurement's accuracy, we need a narrower bandwidth, thus we must decrease the overall loss in a test setup. An antenna is a good radiator, which means it loses a lot of energy to the free space, at its resonant frequencies. By surrounding an antenna with a shielding box, all radiating paths are cut off and the antenna is converted to a resonator with little loss.

Shown in Figure 5.36 is data measured in an S_{21} fixture. The antenna is a single-band antenna. When there is no antenna in the fixture, the transmission loss is around -50 dB. After putting in an antenna, the S_{21} reaches -10 dB at 2.17 GHz. The fixture shown in this example is a constructive coupling one. In such fixtures, an antenna functions as a relay. Without the

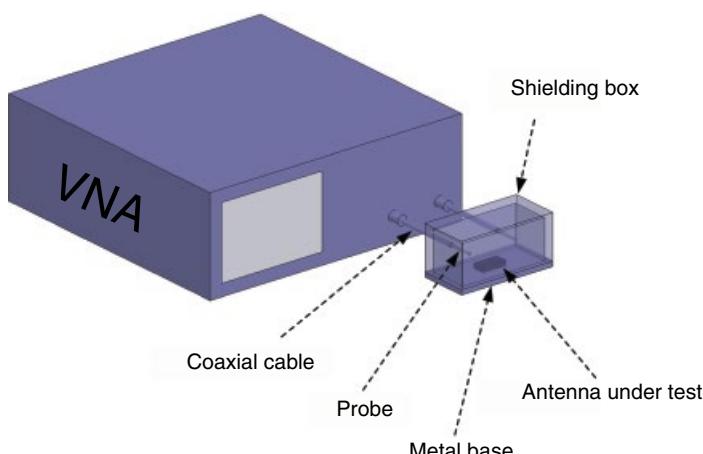


Figure 5.35 Simplified illustration of production test setup (S_{21} type).

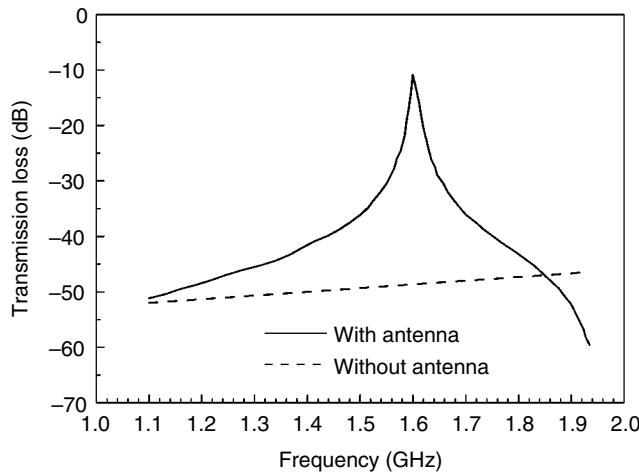


Figure 5.36 Measured result of a S_{21} setup, constructive coupling.

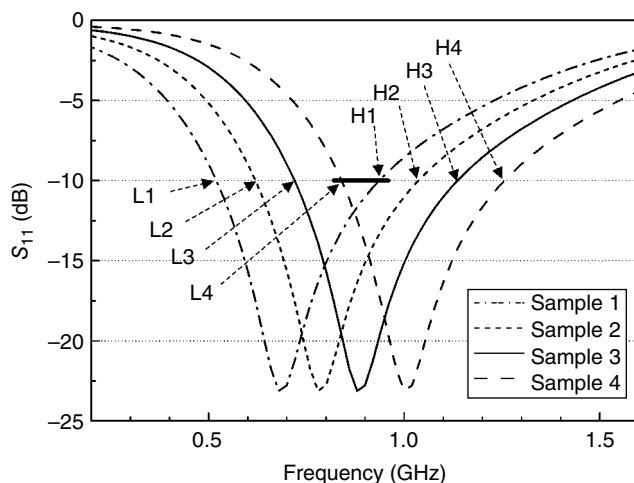


Figure 5.37 Antennas' response measured on a phone.

antenna, the energy from one port can barely reach the other port. At the antenna's resonant frequency, strong currents are excited on the antenna element by the input port; the radiation from the antenna then reaches the output port.

It is also possible to design destructive coupling fixtures. On such fixtures, the resonant current on an antenna blocks the transmission between two ports. The S_{21} curve of a destructive fixture looks like an upside-down version of Figure 5.36.

No matter whether it is an S_{11} fixture or an S_{21} one, as a fixture is different from a real phone, it is almost impossible to get identical responses. A correlation must be established between them. Let's use a single-band antenna to explain the process. Before we can set the pass-fail limit for a fixture, we must have a limit for the phone first. As shown in Figure 5.37, the thick

horizontal line at -10dB between 0.824 and 0.96GHz is the assumed limit. Sample 3 is the optimized antenna design. In this example, the antenna has huge bandwidth; this is for the convenience of demonstration. In reality, managers love to shrink antennas' size until it marginally meets the specification. A group of samples, at least ten, must be prepared to carry out the correlation. Those samples are not randomly picked from the production line. They are purposely made to represent wide variations of antenna responses, and some samples must be failed samples. For convenience, only four samples are used in the example. Samples 1, 2, and 4 are all specially made ones. Samples 1 and 4 are failed parts. Sample 1 has the lowest resonant frequency and it fails at the high band edge. Sample 4 resonates at the highest frequency and it fails at the low band edge. Markers L1–L4 are the four samples' low band edges, where the return loss is -10dB . To pass the limit, the low band edge must be smaller than 0.824GHz. Similarly, markers H1–H4 are high band edges and they must be larger than 0.96GHz to pass the limit.

Shown in Figure 5.38 are the measured results of above four samples on an S_{11} fixture. An S_{11} fixture should give a similar response to a real phone. In this example, the response curves move higher. Marks F1–F4 are the resonant frequencies, which are the lowest return loss points on respective curves, of the four samples.

After collecting all the data, it is the time to carry out the correlation. In this example, the goal of the correlation is to convert the specification of the band limit on a phone, as shown in Figure 5.37, into a frequency range limit on a fixture. There are two lines in Figure 5.39. The left and right lines correspond to the low and high band edges, respectively. Both X and Y axes are frequency ones. Let's look at the data line on the left. The X coordinates, L1–L4, are band edge frequencies obtained from Figure 5.37 and the Y coordinates, F1–F4, are resonant frequencies obtained from Figure 5.38. Plotting them in Figure 5.39 accordingly, we have four rectangular markers. A line can be drawn across these four markers. As has been shown in Figure 5.37, the high end of a fixture's limit is decided by a phone's low band edge. By interpolating from the line, it can be calculated that the low band edge of 0.824GHz corresponds to 1.28GHz on the fixture. Similarly, the high band edge of 0.96GHz corresponds to the low end of the fixture's limit, which is 0.91GHz in this example. The final fixture's limit is from

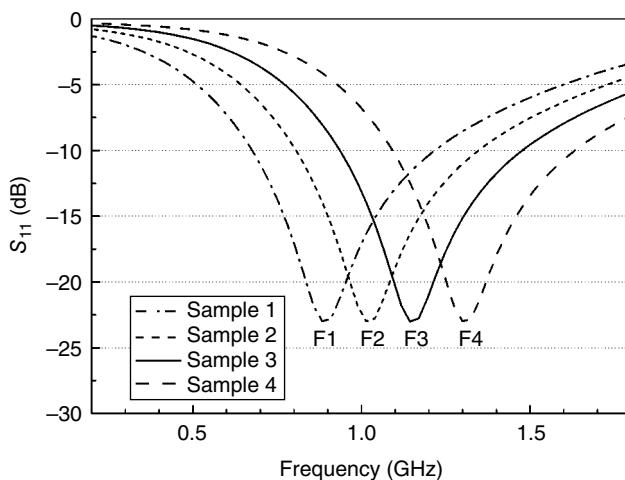


Figure 5.38 Antennas' response measured on an S_{11} fixture.

0.91 to 1.28 GHz, which is the gray area shown in Figure 5.39. If an antenna's resonant frequency on a fixture falls in the range between 0.91 and 1.28 GHz, it passes the limit. Otherwise, it is a failed part.

In reality, four antenna samples are not enough. The actual measured data will never fall nicely in a straight line. Ten to twenty samples are more reasonable. Of course, more is always better. The 12 solid rectangles shown in Figure 5.40 are measured data of the low band edge. The solid line is obtained through linear regression, which can be done quite easily though Microsoft Excel. There is a feature called trend line in Excel, which can generate a straight line, its formula, and the R^2 value. If the fixture functions well, all data points should evenly distribute around the trend line. The R^2 value, which ranges from 0 to 1, is an indication of how good the regression line represents all discrete data points. If R^2 equals 1, all points fall on the line. A R^2 value higher than 0.90 means there is good correlation between the fixture and the phone. If the R^2 value is too low, the fixture should not be used in the production line. In our case, the R^2 value is 0.9822. By substituting the X value of 0.824 GHz into the formula, we get the high limit of the fixture, which is 1.129 GHz.

The samples shown in Figure 5.40 are better distributed than those shown in Figure 5.39, which are exaggerated to fit two lines into the same figure. A good set of correlation antennas should be evenly distributed along a relative wide frequency range. It should include a representative number of failed and passed samples, so the interpolating frequency is away from the end of the trend line.

Designing an S_{11} fixture is relatively easy. An antenna's response on an S_{11} fixture resembles the one on a phone quite well. It is a totally different story when designing S_{21} fixtures. Most antennas used in current phones are basically monopoles, so they need a ground to work properly at the frequency they are designed for. In an S_{21} fixture, an antenna resonates all by itself. In theory, an antenna designed as a quarter-wavelength monopole will resonate at the frequencies twice the one it is supposed to work at, in the half-wavelength mode. In practice, it is even more complex, as the resonant frequencies of an antenna's different modes depend on various aspects, such as the antenna's orientation, the distance from the antenna to the metal shielding box, the size of the shielding box, and the position of probes. There is no clear

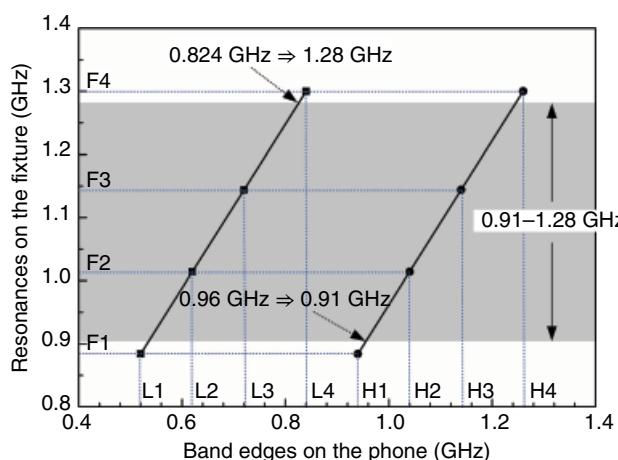


Figure 5.39 Correlations between a phone and a fixture.

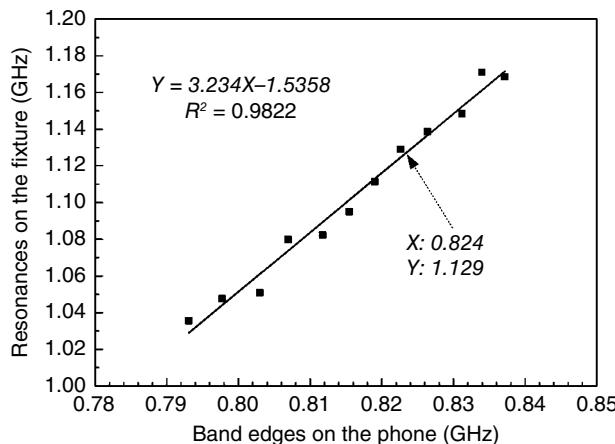


Figure 5.40 Correlation data of a real phone.

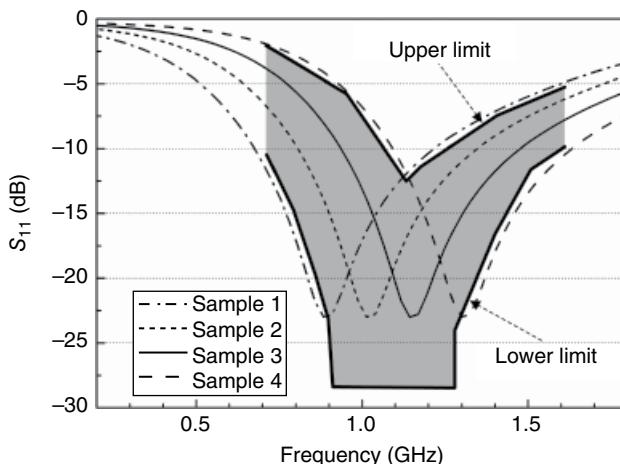


Figure 5.41 Set pass/fail region with upper and lower limits.

step-by-step game book on this issue. If you really want to master designing S_{21} fixtures, a lot of trial and error and good patience are required. From the correlation point of view, the calculation processes of both S_{11} and S_{21} fixtures are pretty much the same.

Besides the correlation method, there is a rougher way to set the limit. By superimposing the responses of large quantities of passed and failed samples, a boundary can be drawn to separate them. As shown in Figure 5.41, the gray area defined by the upper and lower limit lines is the pass region. Only four antennas are used here to demonstrate the concept. If an antenna's response, such as sample 2 or 3, can pass the gray area without crossing the upper or the lower limit, the antenna is a passed unit. Obviously, samples 1 and 4 are failed ones. Most network analyzers have the limit line as a built-in feature, so it is not too difficult to implement it on a production line.

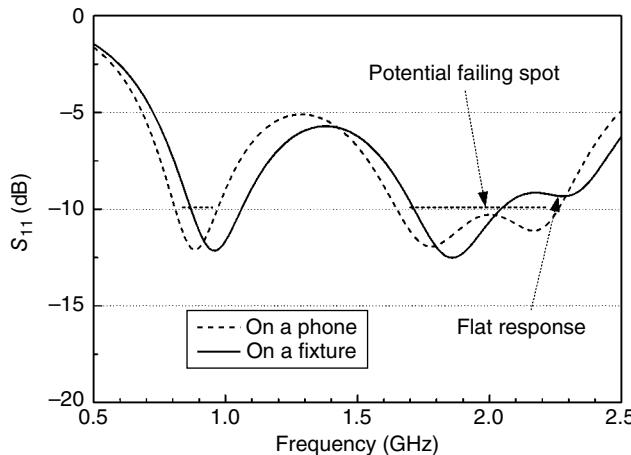


Figure 5.42 Some issues with the correlation method.

The limit line method is especially useful when an antenna has wideband response. The dashed line shown in Figure 5.42 is an antenna's response on a phone. The return loss specification for this phone is better than -10 dB across both $0.824\text{--}0.96\text{ GHz}$ and $1.71\text{--}2.17\text{ GHz}$ bands. At the high band, two resonances are used to provide the required bandwidth. For these kinds of antennas, using the correlation method is quite cumbersome. Two band edges of each band are not enough to represent an antenna's response. As marked in the illustration, the bump in the middle of the high band is a potential failure spot. When the return loss at this spot is higher than -10 dB , we get four -10 dB band edges at the high band instead of two.

There is another possible issue with the correlation method. When measuring an antenna, which has several resonances in one band, some resonances may become illegible on a fixture. As shown in Figure 5.42, the highest resonance on the fixture is quite flat, which makes accurate frequency measurement a challenge. In such circumstances, the limit line method is a better choice. In practice, there are more ways to set up the pass-fail criteria. As long as they can separate good parts from bad ones and customers buy the idea, they are usable ones.

After setting the pass-fail limit and calibrating the network analyzer, can we be confident about the measured results? In fact, we still can't. For instance, how do we know a failed antenna is really a defective one, instead of a false alarm caused by a worn-out fixture? This is where gold units fit in. Whenever we set up a test station, some gold units need to be prepared with it. Some of the gold units are passed ones and some of them are failed ones. Whenever we have doubts about the fixture, we can retest these gold units. If the test results turn out to be different from the original ones, it is time to check the test station. On a production line, it is unnecessary to calibrate the network analyzer too frequently. The consistency of most state-of-the-art network analyzers is extremely good. Frequent calibration might cause more problems than it solves, as there is always some possibility of human error in the calibration process. One dedicated gold unit can be used to quick check the calibration of a test station. By saving the response curve of the dedicated gold unit when setting up the station, then comparing its measurement with the saved one, we can have a pretty good idea of how well the station is performing.

The final production test related to antenna is the phone radiate test. This test station normally is located at the end of a phone's assembly line. To reach this stage, the transceiver on PCB has been conductively measured and calibrated; the antenna has also passed its stand-alone test. The possibility of failure is quite low. However, failures still happen. For instance, the contact mechanism on either PCBs or antennas can be damaged during the assembly process; the matching components are missing or installed with the wrong value, and the glue used in the underfill process spills over and covers the contact pad. These aren't pure hypothetical situations; they happen from time to time. This kind of failure is much more serious than the moderate degradation of an antenna's performance. If an antenna is outside the specification, a phone's performance gets a hit of a few decibels. Normally, people can't tell the difference. It is perceivable only at the fringe of a base station's coverage area. On the other hand, a failure that happens in the assembly process can easily introduce a degradation of more than 10 dB, which is notable to any customer. If such phones were shipped out, the damage to the brand and the loss associated with the logistic channel could be severe.

The aim of a phone's radiate test is to catch the abnormal in the assembly process. It should not be used to detect an antenna's performance degradation, which is the task of stand-alone antenna tests and is too time-consuming to be done in the final stage. In most circumstances, measuring the radiated power on a single channel is adequate. The station should be able to pick up some phones specifically prepared with different antenna-related defects. If the station does not do so, measurements at a few more bands might be necessary. Shown in Figure 5.43 is a setup of a radiating test. A base station simulator is used to establish the phone call and measure the transmit power of the phone under test. The coupler can be either an antenna or a transverse electromagnetic cell.

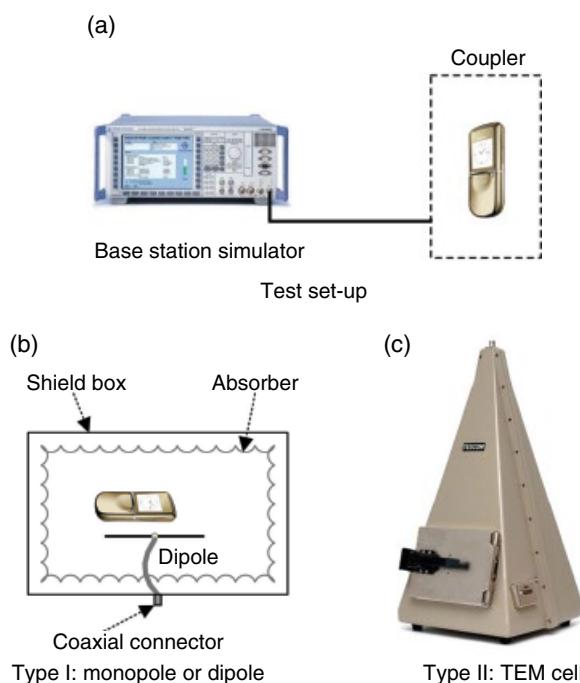


Figure 5.43 Radiating test on a production line. (Source: Reproduced with permission of TESCOM.)

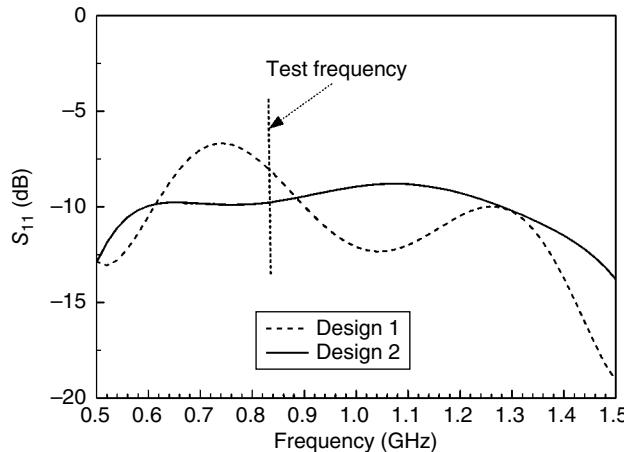


Figure 5.44 Coupling coefficient.

When designing a coupler, there are two considerations. The primary consideration is the coupling coefficient's flatness at the test frequency. The secondary consideration is how strong the coupling is, of course, higher is better. The coupling coefficient equals the transmission loss, which can be measured through the equivalent two-port network composed of a phone fixture and a coupler. Shown in Figure 5.44 is the coupling of two different designs. Although the coupling of design 1 is stronger, design 2 is actually better, because it has a flatter response on the frequency domain. The response curve of a coupler is sensitive to the gap between a phone and its holder. Due to the loading effect of the fixture, the coupling curve of the same antenna might drift around. As a consequence, design 1 generates more measurement variations than design 2.

By offsetting the transmission loss in the base station simulator, the radiated power measured through a fixture should be equal to the conductively measured power. The final pass-fail limit of a fixture depends on how good the coupler is and how consistent the transmitting power is. As the functionality test in a production line, a ± 3 dB limit should be fine. Of course, more time can always be taken to tighten the limit.

5.4 Multiple Input and Multiple Output Antenna Test

Multiple input and multiple output (MIMO) antenna requires both passive and active chamber tests. The passive chamber test has matured and the active chamber test is still a work in progress. Because MIMO antenna test is more than a stand-alone chamber test, this section is dedicated to this topic.

The passive MIMO test is actually measuring envelope correlations between different antennas. When measuring an N port MIMO antenna, first terminate all except one antenna ports with 50Ω loads, then measure the 3D radiation pattern of the unterminated port. Repeat this procedure until all ports are measured. Envelope correlation between any two ports can be obtained by substituting their corresponding 3D radiation patterns into Equations 4.12–4.15 of Section 4.9.

Antenna chamber used for passive MIMO test can be any 3D chamber as discussed in Section 5.1. However, before using any measured 3D radiation patterns, make sure the phase information has been included. By default, phase information is not included in most 3D chamber's exported data. You should be able to find the setting in the User Manual or ask their support engineers directly.

Comparing with passive MIMO test, the active MIMO test, which is also referred as the OTA MIMO test, is a totally different animal. Because 4G systems, both LTE-FDD and TD-LTE, support MIMO technologies, people are working toward a MIMO OTA standard since 2010. Several competing testing methods, such as multiprobe chamber [20], reverberation chamber [21], and two-stage method [22], have been proposed. In August 2015, CTIA released its Test Plan for 2×2 Downlink MIMO and Transmit Diversity Over-the-Air Performance Version 1.0 [23]. As the test plan is a working document and the section is only a brief discussion of the current version, one must refer to the latest version to keep up with the progress.

In Version 1.0, although the test plan mentioned both, only MIMO OTA has been finished and there is no content about Transmit Diversity. Other existing OTA test plans include both downlink (TIS) and uplink (TRP); there is only downlink for MIMO OTA. The 4G standards support 2×2 , 4×2 , and 4×4 MIMO; only 2×2 MIMO are included in the Version 1.0.

As has been demonstrated in Section 4.9, a MIMO antenna's performance is not only decided by the antenna itself but also its surrounding multipath environment. One important aspect of OTA test plan is how to reconstruct a multipath environment inside a chamber. To simplify the infinite possibilities of multipath environment in the real world, various propagation models or spatial channel models have been proposed.

One popular model for 4G system simulation is spatial channel model extended (SCME). The SCME is a comprehensive model. By adjusting model parameters, it can represent a wide range of application scenarios. The Version 1.0 adopted the urban macrocell (UMa) channel model. Parameters of the SCME UMa model are given in Table 5.3.

Table 5.3 SCME urban macrocell (UMa) channel model [23]

SCME urban macro-cell								
Cluster no.	Delay (ns)		Power (dB)			AoD (°)	AoA (°)	
1	0	5	10	-3	-5.2	-7	82.0	65.7
2	360	365	370	-5.2	-7.4	-9.2	80.5	45.6
3	255	260	265	-4.7	-6.9	-8.7	79.6	143.2
4	1040	1045	1050	-8.2	-10.4	-12.2	98.6	32.5
5	2730	2735	2740	-12.1	-14.3	-16.1	102.1	-91.1
6	4600	4605	4610	-15.5	-17.7	-19.5	107.1	-19.2
Delay spread (ns)							839.5	
Cluster AS AoD/AS AoA (°)							2/35	
Cluster PAS shape							Laplacian	
Total AS AoD/AS AoA (°)							7.9/62.4	
Mobile speed (km/h)/direction of travel (°)							30/120	
XPR NOTE: V & H components based on assumed BS antennas							9 dB	
Mid-paths share cluster parameter values for:							AoD, AoA, AS, and XPR	

Source: (CTIA.)

The SCME UMa model has six clusters, from 1 to 6. There are three taps in each cluster. Assume an antenna is sending signal at the base station, the signal passes through six paths (clusters) to reach the receiving antenna. In each cluster, one signal becomes three slightly delayed signals. Overall, there are total 18 multipath signals at the receiver side. Using cluster 1 as an example, three signals are delayed by 0, 5, and 10 ns, respectively. Their powers are -3, -5.2, and -7 dB lower than the transmitted signal. When leaving the base station, the angle of departure (AoD) of the cluster 1 is 82.0°. At the receiver end, the angle of arrival (AoA) of the cluster 1 is 65.7°. Inside each cluster, the actual AoD and AoA of each signal are randomly generated based on the angle spread (AS) value. The angle spread in each cluster are 2° at the base station and 35° at the receiver, respectively. The delay spread of all signals are 839.5 ns. The angle spread of all signals are 7.9° at the base station and 62.4° at the receiver, respectively.

The receiver is traveling toward 120° at speed of 30 km/h. At the receiver side, because multipath signals come in from different angles (AoA), their decomposed relative speed and thus Doppler frequency shift are all different.

The XPR is cross-polarization power ratio, which is a measure of how much the power in one polarization leaks to the orthogonal polarization. In SCME UMa model, when a 0 dBm signal is transmitted as vertical polarized, there is -9 dBm signal leaks to the horizontal polarization.

Shown in Figure 5.45 is the eight-element boundary array configurable recommended by Version 1.0. Eight dual-polarized antennas are uniformly distributed on a ring. The DUT, which can be a phone or a laptop, sits at the center of the ring. There are actually 16 independent antennas, 8 of them are vertically polarized and 8 are horizontally polarized.

The base station emulator shown in Figure 5.45 is similar to the one discussed in Section 5.2.1, except it has two output. As Version 1.0 is a test plan for 2×2 Downlink MIMO, which means there are two ports at the base station and two ports on the phone, the emulator only needs two ports. In the future, when 4×2 or 4×4 MIMO are also included in the test plan, an emulator should have four ports.

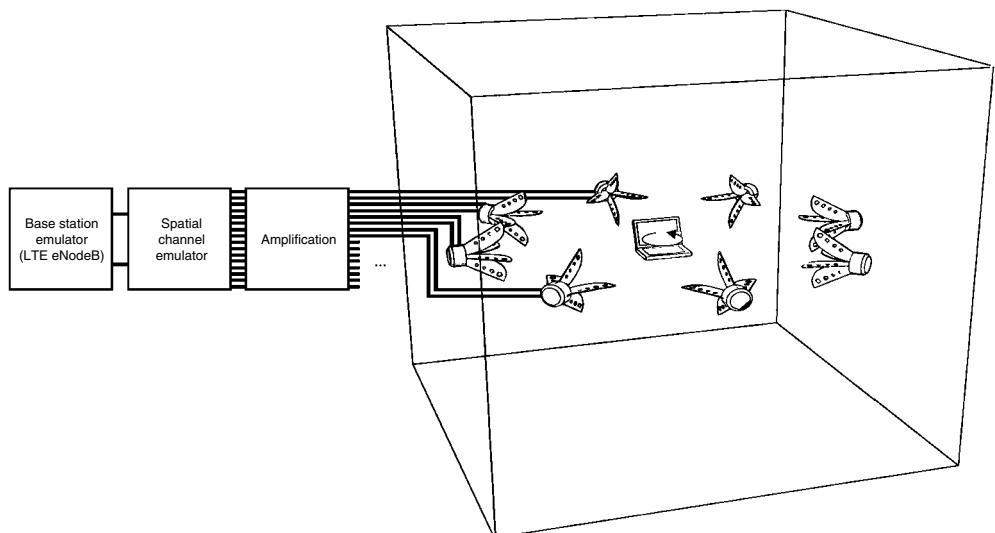


Figure 5.45 Block diagram of a typical boundary array RF environment simulation system [23]. (Source: CTIA.)

In real life, the two ports of base station emulator can be connected to two antennas directly. In the setup shown in Figure 5.45, it connects to a spatial channel emulator. When the spatial channel emulator gets a signal from the base station emulator, it regenerates 18 signals based on the SCME UMa model. Each signal has different delay, amplitude, and Doppler frequency shift. Because there are two outputs on the base station emulator, the spatial channel emulator needs to apply SCME UMa model to both data streams. In Version 1.0, two antennas at the base station have different polarizations and radiation patterns; the spatial channel emulator also needs to process signals accordingly.

Based on SCME UMa model, six multipath clusters are coming from 65.7° , 45.6° , 143.2° , 32.5° , -91.1° , and -19.2° , respectively. However, there are only eight dual-polarized antennas which are uniformly distributed with separation of 45° . Thus each multipath signal needs to be sent through some weighting methods. For example, if AoA of a multipath signal is in the middle of two adjacent antennas, both antennas need to send this signal and each antenna sends half of its total power.

EIS or TIS, which have been discussed in Section 5.2.2, are used to evaluate a pre-4G phone's sensitivity. To measure a MIMO phone, effective throughput SIR sensitivity (ETSS) and MIMO average radiated SIR sensitivity (MARSS) are introduced. ETSS is similar to EIS and is used to measure sensitivity of a single orientation. MARSS is similar to TIS and is an average value of many ETSS.

Unlike TIS test, which decreases the power of transmitted signal until the FER reaches a preset threshold, MARSS keeps signal's power constant and increases injected interference power. Interference should be uniformly distributed around the DUT. Normally, the interference are generated by the spatial channel emulator and transmitted together with multipath signals. Interference transmitted from each antenna should be independent and at the same power level.

The purpose of using MIMO technologies is to increase frequency efficiency and thus a phone's throughput. To make ETSS or MARSS measurements more meaningful, a phone's throughput is used to set various thresholds.

When measure ETSS of a phone, first place the phone in the center of the chamber as shown in Figure 5.45. Then dial down interference power and make sure the throughput of phone reaches 100% of the maximum throughput. Next, increase interference power until the throughput reaches 95% of the maximum throughput. The current signal-to-interference ratio or SIR (value in dB) is the measured ETSS value of 95% throughput, denoted as $P_{\text{ETSS},95}$.

MARSS is average value of multiple ETSS. In Version 1.0, a phone under test needs to rotate around its Z-axis as shown in Figure 5.45. One ETSS is measured at every 30° in azimuth. In total, there are 12 ETSS values, denoted as $P_{\text{ETSS},95,m}$ ($m = 1, 2, \dots, 12$).

Using Equation 5.15, a MARSS of 95% throughput, which is denoted as $P_{\text{MARSS},95}$, can be obtained.

$$P_{\text{MARSS},95} = 10 * \log_{10} \left[\frac{1}{M} \sum_{m=1}^M 10^{(P_{\text{ETSS},95,m}/10)} \right] \quad (5.15)$$

In Version 1.0, MARSS values of 90 and 70% are also required. They can be calculated by Equations 5.16 and 5.17, respectively.

$$P_{\text{MARSS},90} = 10 * \log_{10} \left[\frac{1}{M} \sum_{m=1}^M 10^{(P_{\text{ETSS},90,m}/10)} \right] \quad (5.16)$$

$$P_{\text{MARSS},70} = 10 * \log_{10} \left[\frac{1}{M} \sum_{m=1}^M 10^{(P_{\text{ETSS},70,m}/10)} \right] \quad (5.17)$$

This section has introduced the most important aspects of MIMO OTA test. However, many useful information, such as device orientation, chamber validation, and phase calibration are all omitted in the book. Before one starts any serious MIMO OTA test, especially when there is no support from a veteran engineer, it is recommended that one should read the latest CTIA test plan. The MIMO OTA test is indeed more complex than all previous OTA tests.

References

- [1] "Agilent Network Analyzer Basics," <http://cp.literature.agilent.com/litweb/pdf/5965-7917E.pdf>. Retrieved 25 October 2010.
- [2] Icheln, C., Krogerus, J., and Vainikainen, P. (2004) "Use of balun chokes in small-antenna radiation measurements," *IEEE Transactions on Instrumentation and Measurement*, **53**, 498–506.
- [3] Icheln, C. and Vainikainen, P. (2000) "Dual-frequency balun to decrease influence of RF feed cables in small antenna measurements," *Electronics Letters*, **36**, 1760–1761.
- [4] Phillips, J.P. and Krenz, E.L. (1998) Spherical-scan near-field chamber for cellular phones. Proc. AMTA, Montreal, QC, Canada, pp. 37–42.
- [5] "Applying Error Correction to Network Analyzer Measurements," <http://literature.agilent.com/litweb/pdf/5965-7709E.pdf>. Retrieved 25 October 2010.
- [6] "Specifying Calibration Standards and Kits for Agilent Vector Network Analyzers," <http://cp.literature.agilent.com/litweb/pdf/5989-4840EN.pdf>. Retrieved 25 October 2010.
- [7] Dash, G., "How RF Anechoic Chambers Work," http://glendash.com/Dash_of_EMCAnechoic_Chambers/Anechoic_Chambers.pdf. Retrieved 25 October 2010.
- [8] Hemming, L.H. (2002) *Electromagnetic Anechoic Chambers: A Fundamental Design and Specification Guide*, Wiley-IEEE Press.
- [9] King, H., Shimabukuro, F., and Wong, J. (1967) "Characteristics of a tapered anechoic chamber," *IEEE Transactions on Antennas and Propagation*, **15**, 488–490.
- [10] "Satimo Stargate 64," <http://www.satimo.com/content/products/sg-64>. Retrieved 25 October 2010.
- [11] "CTIA Certification," http://www.ctia.org/business_resources/certification/. Retrieved 25 October 2010.
- [12] "Test Plan for Mobile Station Over the Air Performance Rev 2.2," http://files.ctia.org/pdf/CTIA_TestPlanforMobileStationOTAPerformanceRevision_2_2_2_Final_121808.pdf. Retrieved 25 October 2010.
- [13] "3GPP TS 34.114: User Equipment (UE)/Mobile Station (MS) Over The Air (OTA) Antenna Performance; Conformance Testing," <http://www.3gpp.org/ftp/Specs/html-info/34114.htm>. Retrieved 25 October 2010.
- [14] "Test Plan for CDMA Mobile Stations Rev 5.4," http://files.ctia.org/pdf/CTIA_CDMA_Test_Plan_Rev_5.4.pdf. Retrieved 25 October 2010.
- [15] Gu, Q. (2005) *RF System Design of Transceivers for Wireless Communications*, Springer.
- [16] Iskander, M.F. (2000) *Electromagnetic Fields and Waves*, 1st edn, Waveland Press, Inc.
- [17] Sadiku, M.O. (2009) *Elements of Electromagnetics*, 5th edn, Oxford University Press, USA.
- [18] Ulaby, F.T., Michielssen, E., and Ravaioli, U. (2010) *Fundamentals of Applied Electromagnetics*, 6th edn, Prentice Hall.
- [19] Buck, J. and Hayt, W. (2005) *Engineering Electromagnetics*, 7th edn, McGraw-Hill Science/Engineering/Math.
- [20] Scannavini, A., Foged, L.J., Estrada, J. et al. (2015) 2×2 MIMO downlink OTA measurement based on CTIA guidelines. 2015 IEEE International Symposium on Antennas and Propagation, pp. 290–291.
- [21] Hussain, A., Por Einarrson, B., and Kildal, P.-S. (2015) "MIMO OTA testing of communication system using SDRs in reverberation chamber," *IEEE Antennas and Propagation Magazine*, **57**, 44–53.
- [22] Jing, Y., Wen, Z., and Kong, H. (2011) Two-stage over the air (OTA) test method for MIMO device performance evaluation. 2011 IEEE International Symposium on Antennas and Propagation, pp. 71–74.
- [23] "CTIA Test Plan for 2×2 Downlink MIMO and Transmit Diversity Over-the-Air Performance Ver. 1.0," http://www.ctia.org/docs/default-source/default-document-library/ctia_mimo_ota_v1_0NEW.pdf. Retrieved 25 October 2015.

6

Regulations Related to Antenna Engineers

There are three regulations which are of relevance to antenna engineers. The specific absorption rate (SAR) regulates the maximum power density inside human bodies when a phone is used. The hearing aid compatibility (HAC) regulates the maximum value of the electric (E) and magnetic (H) fields, which can interfere with the proper functionality of a hearing aid, around a phone's speaker. Both E and H fields regulated by the HAC are near field. The electromagnetic interference (EMI) or EM compatibility (EMC) regulates the maximum E field that can radiate into the surrounding environment. The E field regulated by the EMI is a far field which can interfere with the surrounding equipment. Most test procedures can be found on the Federal Communications Commission (FCC) website [1].

6.1 Specific Absorption Rate

There is a good reference book on the history of safety issues related to communication devices: *Handbook of Antennas in Wireless Communications*, edited by Lal Chand Godara [2]. Chapter 26, “Safety Aspects of Radio Frequency Effects in Humans from Communication Devices,” authored by Alan W. Preece, provides a comprehensive overview on the issue. More information can also be found in the “Telecommunication Health and Safety” column in the *IEEE Antenna and Propagation Magazine*.

As a brief summary, there is no consistent and solid evidence to support claims of non-thermal effects, such as brain tumors, DNA damage, and so on, when a biology object is exposed to normal dosage of radio frequency (RF) radiation. The exposure guidelines for RF radiation set by various governments cover an extreme wide frequency range, from a few KHz to several hundreds of GHz. The exposure limit not only depends on the working frequency but also depends on who is operating the device. At the commonly used cellular band, the limit for professional users can be five times higher than the one for the general population.

6.1.1 Definition and Measurement Method of SAR

Based on experimental studies on animals and humans, the external thermal load to create a 1°C temperature rise over about 15 minutes is about 4 W/kg or 4 mW/g. For partial body exposures, the presence of blood circulation can reduce the temperature increase. In particular, the limbs and head, which are the most likely to receive higher exposure, require around 10 W/kg to generate this rise. From a medical point of view, a 1°C temperature rise is not a severe event at all. This can easily be achieved by internal energy expenditure occasioned by gentle jogging.

Based on the definition of the biodosimetry, the SAR is the rate of absorption or dissipation of energy (W) in unit mass (M) as shown in Equation 6.1 [2]:

$$\text{SAR} = \frac{d}{dt} \left(\frac{dW}{dM} \right) = \frac{d}{dt} \left(\frac{dW}{\rho dV} \right) \quad (6.1)$$

Here, ρ is the specific density.

In the world of cellular phones, the SAR is a measure of the rate at which energy is absorbed by the body when exposed to an RF EM field. It can be calculated from the E field within the tissue as [3] follows:

$$\text{SAR} = \frac{\sigma E^2}{\rho} \quad (6.2)$$

Here, σ is the tissue's electrical conductivity, ρ is the tissue's density, and E is the root mean square (RMS) electric field. The commonly used unit of SAR is either W/kg or mW/g, which actually are identical. To smooth out the fluctuation in SAR measurement, SAR is usually averaged over a small sample volume (typically 1 or 10 g of tissue).

In practice, it would be too gross to use real tissues in SAR measurements, so various tissue-simulating liquids are introduced. The dielectric properties of different human tissues can vary over quite a wide range. To simplify the test procedure, only two human parts, head and body, are considered in SAR measurements. Shown in Tables 6.1 and 6.2 are target dielectric properties for head and body [4], respectively. The values listed here are target values according to IEEE 1528 standard. More than one standard is currently used, and the target values in different standards are also slightly different.

It is obvious that the dielectric properties of either head or body are variables of the working frequency. Ideally, people should find one liquid to simulate the head in all frequencies and

Table 6.1 Dielectric properties of head liquid (IEEE 1528 [3])

Freq (MHz)	Dielectric constant (ϵ_r)	Conductivity σ (S/m)
835	41.5	0.90
900	41.5	0.97
1800–2000	40.0	1.40
2450	39.2	1.80

Source: (Reproduced with permission of IEEE.)

Table 6.2 Dielectric properties of body/muscle liquid (IEEE 1528 [3])

Freq (MHz)	Dielectric constant (ϵ_r)	Conductivity σ (S/m)
835	55.2	0.97
900	55.0	1.05
1800–2000	53.3	1.52
2450	52.7	1.95

Source: (Reproduced with permission of IEEE.)

Table 6.3 Recipe for simulating liquid at different frequencies (IEEE 1528 [3])

Ingredients (% by weight)	Frequency (MHz)									
	450		835		915		1900		2450	
Tissue type	Head	Body	Head	Body	Head	Body	Head	Body	Head	Body
Water	38.56	51.16	41.45	52.4	41.05	56.0	54.9	40.4	62.7	73.2
Salt (NaCl)	3.95	1.49	1.45	1.4	1.35	0.76	0.18	0.5	0.5	0.04
Sugar	56.32	46.78	56.0	45.0	56.5	41.76	0.0	58.0	0.0	0.0
HEC	0.98	0.52	1.0	1.0	1.0	1.21	0.0	1.0	0.0	0.0
Bactericide	0.19	0.05	0.1	0.1	0.1	0.27	0.0	0.1	0.0	0.0
Triton X-100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	36.8	0.0
DGBE	0.0	0.0	0.0	0.0	0.0	0.0	44.92	0.0	0.0	26.7
Dielectric constant	43.42	58.0	42.54	56.1	42.0	56.8	39.9	54.0	39.8	52.5
Conductivity (S/m)	0.85	0.83	0.91	0.95	1.0	1.07	1.42	1.45	1.88	1.78

Source: (Reproduced with permission of IEEE.)

another one for the body. However, at present, there is no single head- or body-simulating fluid which can cover all frequencies with an acceptable accuracy and consistency. Different fluids must be used when carrying out SAR measurements in different frequency bands. Shown in Table 6.3 is a recipe for simulating liquid at different frequencies.

As the dielectric properties of head or body are close enough at adjacent frequency bands and a $\pm 5\%$ tolerance from the target value is allowed according to most standards, a single simulating liquid might be used when measuring adjacent bands, such as 850 and 900 MHz bands or 1800 and 1900 MHz bands.

As the dielectric properties are variables of temperature, the ambient temperature and also the tissue-simulating liquid temperature need to be well controlled within a certain range.

Besides tissue-simulating liquids, a container is required to hold the liquids. Shown in Figure 6.1 is a phantom produced by Schmid and Partner Engineering AG (SPEAG). The shell corresponds to the specifications of the specific anthropomorphic mannequin (SAM) phantom defined in IEEE 1528-2003, CENELEC 50361, and International Electrotechnical Commission (IEC) 62209. It enables the dosimetric evaluation of left and right hand phone usage as well as body-mounted usage at the flat phantom region. The whole container is installed on a wooden frame. The phone holder is made of low loss plastic.

A stand-alone phantom head, such as the one shown in Figure 6.2, can also be used as the container. The head has the same size and curvature as the one shown in Figure 6.1.

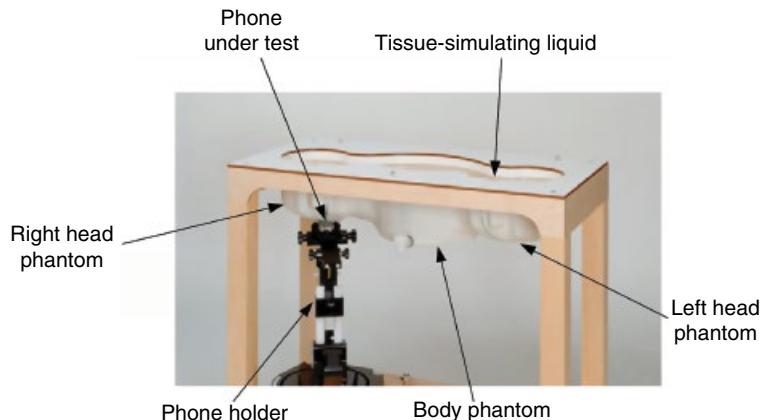


Figure 6.1 Phantom used in SAR measurements (Twin SAM). (*Source:* Reproduced with permission of Schmid & Partner Engineering AG.)



Figure 6.2 SAM head. (*Source:* Reproduced with permission of Schmid & Partner Engineering AG)

A head phantom, which meets the specifications defined by the standards, is often referred to as the SAM's head. The SAR of both left and right hand phone usage can be measured on a SAM's head. The mechanical parts on both sides of the SAM's head are phone holders.

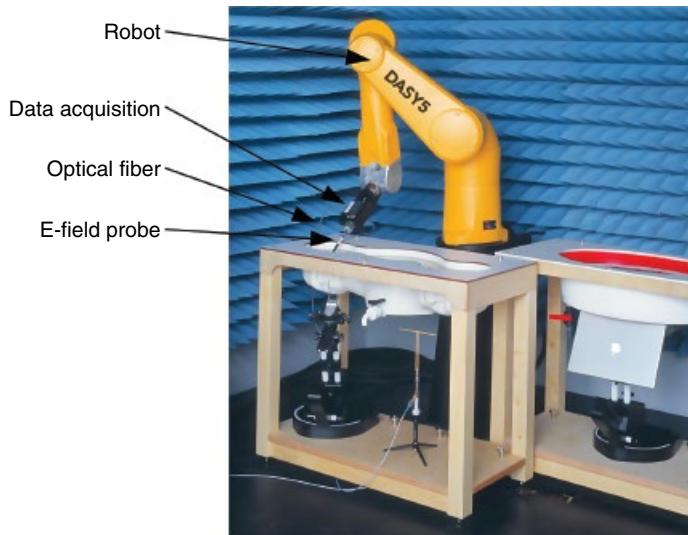


Figure 6.3 Robot and *E*-field probe. (Source: Reproduced with permission of Schmid & Partner Engineering AG)

According to the SAR's definition shown in Equation 6.2, only the *E* field is required when measuring a device. An *E*-field probe moved by a robot, as shown in Figure 6.3, is the heart of any SAR measurement setup. The robot is normally a five-axis one, which means it can move a probe to different locations in a three-dimensional (3D) space and also point the probe toward different angles. Comparing with the weak signal a probe detects, a robot is a device that drains enormous power. To mitigate the potential EMI problem, testing signals are amplified and converted into a digital signal in a shielded box, then transmitted to the processing computer through an optical fiber. The data acquisition circuit is powered by batteries to eliminate any galvanic connection toward the external environment.

As *E* field is a vector field, it is composed of three orthogonal components as shown in Equation 6.3. Inside an *E*-field probe, there are actually three individual *E*-field sensors. Each sensor measures one orthogonal *E*-field component.

$$E = \sqrt{E_x^2 + E_y^2 + E_z^2} \quad (6.3)$$

Shown in Figure 6.4 are two commonly used configurations: Δ and I [5]. Three short dipoles are used to detect those orthogonal components. In both configurations, the dipoles are arranged in a manner which ensures that they are perpendicular to each other.

It might be thought that the short dipoles are connected to the amplifier circuit through some normal transmission lines, such as a coaxial line or a microstrip line. In fact, they could not be arranged so. Any transmission line which can transmit an RF signal at 900 MHz or 1.8 GHz is composed of long metal traces. Those metal traces will disturb the field distribution around the device under test, thus invalidating the result. It seems that we are facing a dilemma. How can we measure something without putting a probe next to it?

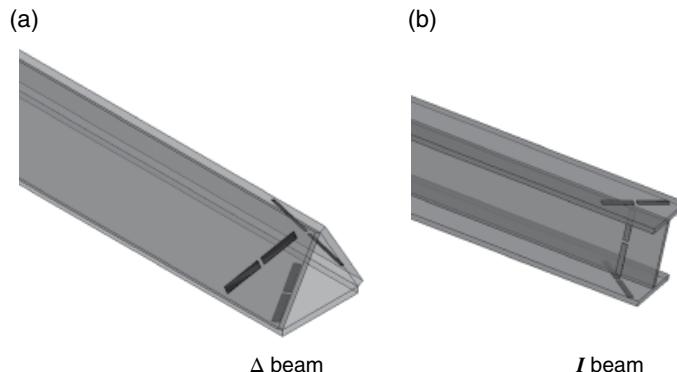


Figure 6.4 Configurations of orthogonal short dipoles.

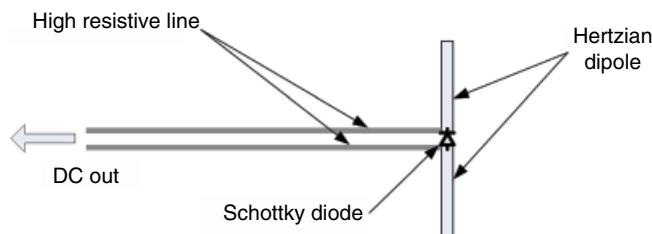


Figure 6.5 Sensor made of Schottky diode.

Shown in Figure 6.5 is the solution which is adopted by most commercial SAR systems. Of all the parts of a sensor, only the short dipole is made of metal. As the dimension of the short dipole is much less than one wavelength of even the highest working frequency, the disturbance from the dipole can be ignored. The parallel transmission line connected to the dipole is made of high resistive material and is not able to sustain a current. The transmission line is transparent from the RF point of view. The trick is done by the Schottky diode. A diode can function as a rectifier which converts an RF signal into a baseband envelope signal. The voltage value of the baseband envelope signal is related to the power level of the RF signal. However, the relation between the voltage value and the power level is not linear; the relation can be approximately expressed by the square law. Some means of calibration is required to establish the correlation between the two. The high resistive transmission line cannot carry a current, but it can transmit a baseband voltage signal as long as the input impedance of the signal amplifier is much higher than the total resistance of the transmission line. At the frequencies of cellular communication bands, the shunt capacitance required in a diode rectifier's circuit can be implemented either through the parasitic capacitance of the diode itself or the distributed capacitance of the layout.

Now, it is obvious that the probe converts RF power into baseband envelope voltage on the spot and transmits the voltage through an RF transparent transmission line. By doing this, we are successfully putting a probe next to a phone without allowing the phone to “notice” it. The conversion method is sufficient for SAR measurements; however, it might

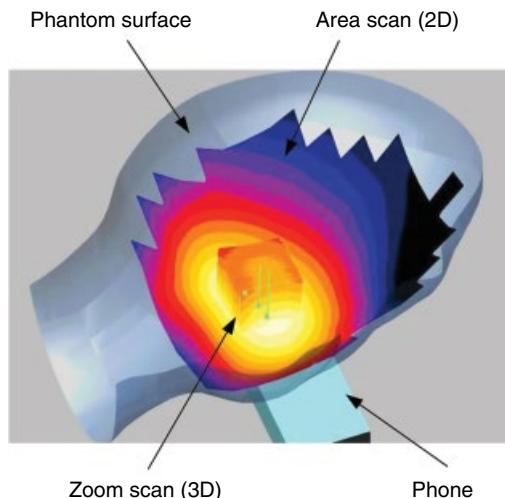


Figure 6.6 Area and zoom scanning on a left head phantom.

not suit other kinds of near-field scanning, because the phase information is discarded during the conversion.

When doing a SAR test, a phone should be in the normal working condition. There should not be any cable attached to the phone. The phone must only have all the features which will appear on the mass production phones. Similar to an over-the-air (OTA) test, a base station simulator is used to initiate and maintain a call with the phone under test. The phone is configured to transmit at its maximum power.

The goal of SAR measurement is to find the maximum value, which is averaged over either 1 or 10 g volume depending on different standards, inside the phantom head. The tissue, either a real one or a simulated one, is a lossy material. The deeper an RF signal penetrates, the more it is attenuated. As a matter of fact, the highest SAR always appears on the inner surface of the phantom. To expedite the test procedure, the whole test is divided into two steps, an area scan, and a zoom scan, as illustrated in Figure 6.6. The first step is the area scan. A two-dimensional (2D) area is scanned by following the inner surface of a phantom. The result is a SAR distribution on a curvature surface. Using the maximum point obtained from the first step as the center location, the second step scans a localized 3D cubic.

For the purpose of protection and sealing sensors from a tissue-simulating liquid, a plastic cover is used as a probe's outer shell. Thus, there is always a gap between the sensor and the inner surface of a phantom. The SAR value on the surface must be extrapolated from the measured 3D data by postprocessing software.

Shown in Figure 6.7 is an illustration of 1 and 10 g average. A body phantom is used in the measurement. The zoom scanning volume is marked by dashed lines. There are two cubes inside the scanning volume. The larger one is the 10 g average cube. The location of the 10 g cube represents where the maximum averaged SAR over 10 g is. The smaller one is the 1 g average cube. Similarly, its location corresponds to where the peak averaged SAR over 1 g is. The density of tissue-simulating liquid is approximately 1000 kg/m^3 , thus the volume of 1 g liquid is approximately 1 cm^3 .

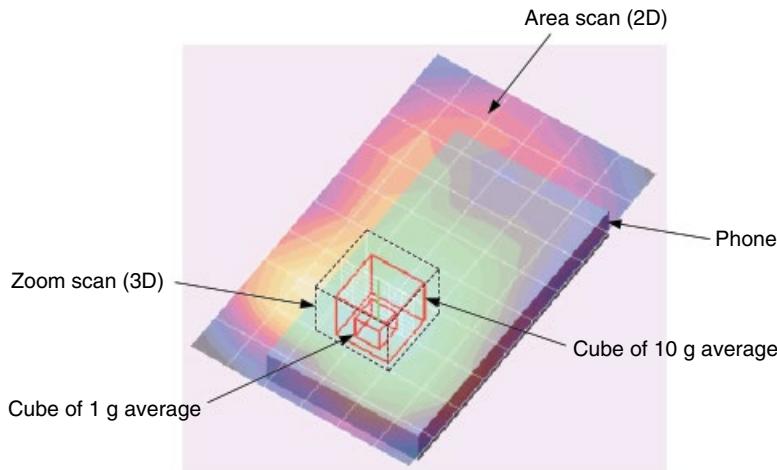


Figure 6.7 Cube of 1 and 10 g average on a body phantom.

Sometimes, there might be two regional maximums emerging from the area scan. If the difference between two peaks is less than 2 dB, two zoom scans which are centered on the two maximums, respectively, are required.

A thorough SAR test includes measurements at both sides of the head and the body. On each side of the head, there are two positions, the touch position and the 15° tilt position. There are also two positions, liquid crystal display (LCD) up and LCD down, for the body. At each position, the SAR measurement must be carried out at multiple bands. At each band, there are low, middle, and high channels. For some phones, there are multiple configurations by the phone itself, such as whip up and down or slide open and close.

It is obvious that an enormous number of measurements are required if all the combinations must be measured. Based on the current IEC 62209-1 standard [6], only the middle channel of all combinations is mandatory. For the configuration which generates the highest SAR, a full set of measurements at low, middle, and high channels are required. It is recommended that low, middle, and high channels are also measured for all combinations whose SAR reading is within 3 dB of the device's highest SAR value.

6.1.2 SAR Limits in the United States and Europe

There are quite a few regulations which are related to SAR limits. For example, in the United States, the FCC Rules Part 22H regulates the 824–849 MHz bands, Rule 24E regulates the 1850–1910 bands, and Rule 15.247 regulates the 2400–2483.5 MHz bands. As a design engineer, it is not necessary to remember all those regulations. All the regulations in the United States and Europe can be boiled down to the following two values:

1. In the United States, less than 1.6 mW/g or 1.6 W/kg, averaged over 1 g
2. In Europe, less than 2.0 mW/g or 2.0 W/kg, averaged over 10 g

If we recall that 10 mW/g is required to generate 1°C temperature increase inside a brain or limbs, both limits are well within the safety zone. Judging by the limit value itself, the US

Table 6.4 Comparison between US and European limits

Phone model	SAR value (mW/g)		$\frac{\text{SAR 1g}}{1.6 \text{ mW/g}} / \frac{\text{SAR 10g}}{2.0 \text{ mW/g}}$
	Averaged over 1 g	Averaged over 10 g	
Apple iPhone 3Gs	1.19	0.79	1.88
HTC diamond	1.40	0.78	2.24
LG CU920 Vu	1.26	0.76	2.07
Motorola Z9	0.94	0.56	2.10
Nokia N95	0.64	0.37	2.16
Nokia N97	1.01	0.74	1.71
Palm Pre	0.94	0.53	2.22
Samsung Omnia i910	1.25	0.89	1.76

limit is only 20% tighter than the European one. However, the European standard calculates the SAR value by averaging over a volume of 10 g instead the 1 g used in the US standard. The larger the average volume is, the lower the SAR value is. Thus, the US limit is significantly tighter than its European counterpart.

To give a more intuitive feeling about those two limits, SAR values, averaged over both 1 and 10 g, of eight phones are listed in Table 6.4. The data are collected from FCC reports [7]. The SAR values averaged over 10 g are collected from the same 3D scanning which generated the maximum SAR values averaged over 1 g. The fourth column is the comparison between the two limits. Using the Apple iPhone 3Gs as an example, the American limit is 1.88 times stricter than the European one. Overall, the ratio is around 2 between these two limits.

A comparison of some phones' SAR values between 1 and 10 g can be found on the Internet. Those values might be different from values listed in Table 6.4, because they are collected from certification reports filed in different countries. Let's use a 2G phone as an example. An international 2G phone might support 850, 900, 1800, and 1900 MHz bands. However, different countries support different portions of the overall bands. In the United States, when filing an FCC application, only the averaged SAR over 1 g at 850 and 1900 MHz are required. Similarly, in Europe, only 900 and 1900 MHz are measured for SAR over 10 g. As a phone's radiating power can be individually adjusted band by band, a comparison between SAR values over bands is not that meaningful.

For devices which have several coexisting transmitters, such as Global System for Mobile Communication (GSM) and Bluetooth, or GSM and wireless local area network (WLAN), the SAR value can be obtained through either a mathematical summation of individual SARs or a measurement with both modules transmitting simultaneously. This detailed information is omitted in the book. Interested readers can refer to related articles [8].

When working on SAR-related issues, one very useful resource is the FCC website. You can find the SAR reports of all FCC-approved phones under the "Equipment Authorization Search" page [7]. As required by the FCC filing process, detailed photos of a phone's internal circuit are also included in the report. From those photos, it might be possible to tell whether some special SAR reduction techniques have been used in certain phones.

6.1.3 Controlling SAR

The SAR might be the most dangerous time bomb for any antenna engineer. However, the SAR measurement itself needs to carry some of the blame. As the SAR test equipment is a quite complex system, the expanded uncertainty from the system itself is more than 20%. This means a phone that has a nice margin of 20% when it is measured in an internal R&D facility might fail when it is measured in an external certificated testing house. To mitigate that risk, most companies adopt an internal SAR limit which is lower than the official one. The internal limits are different from company to company. How much lower the internal limit is depends on a company's preference and judgment on risks. Of course, a lower internal benchmark only makes it more of a challenge to meet the limits as an antenna engineer.

Besides the measurement uncertainty, the SAR is also sensitive to a phone's mechanical parts, especially the metal ones. Some of the last-minute changes from other disciplines, such as adding extra grounding tabs for electrostatic discharge protection, changing the plating technique of the front bezel to improve the production yield, and so on, can all have an impact on the SAR. When those emergencies happen, it might be quite late in a phone's development cycle, which means more hard work is needed to solve the SAR problem in time.

Before discussing how to control SAR, we need to understand how a phone's near-field energy is distributed. As the whip antenna is disappearing from the phone market, the characteristic of a whip's SAR is not covered in the book. Only internal antennas and short external stubby antennas are discussed. As shown in Figure 6.8, the SAR distribution on a phone mostly depends on the working frequency bands. At the low bands, such as 850 and 900 MHz bands, the metal structure of a phone is the main radiator. The maximum E field on a phone normally appears in the middle of the phone. When a phone instead of the antenna on the phone is the main radiator, an antenna's form factor, whether it is a monopole or a planar inverted-F antenna (PIFA), has less impact on the SAR value.

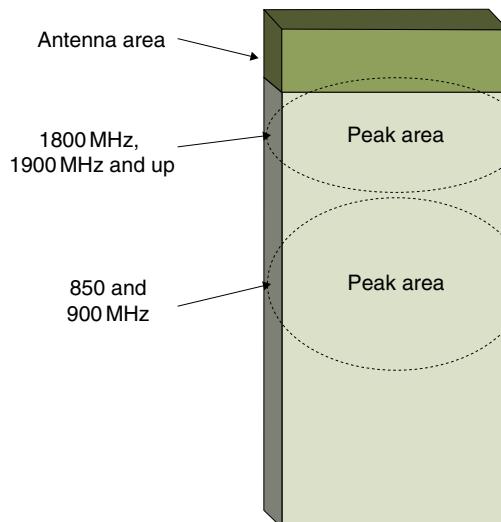


Figure 6.8 Peak energy location on a phone at different bands.

At the high bands, such as 1800 MHz, 1900 MHz and up, the peak near-field energy appears in locations which are closer to the antenna. At those bands, an antenna's form factor plays an important role. A normal PIFA has a directional radiating pattern, which means the ground underneath a PIFA can divert the energy away from a user's head, thus decreasing the field strength on the opposite side of the antenna.

When a phone is measured on a phantom body, the distance between its surface and the phantom's surface is a relatively consistent value, so the SAR distribution normally agrees pretty well with the near-field distribution of the phone. For antenna engineers, the body position is not a big concern, as there is no specification for how much the gap between a phone and a phantom should be. We can always decrease a phone's SAR value at the body position by increasing the gap. For phones with low SAR value, the gap normally chosen is 15 mm.

Unlike a phantom body, a phantom head is a totally different story. According to the standard [3], four positions must be measured on a phantom head. Shown in Figure 6.9 are illustrations of the left ear check position. It is also known as the touch position. Shown in Figure 6.10 are illustrations of the left ear tilt position. Repeating the two aforementioned measurements on the right ear, we have a total of four combinations.

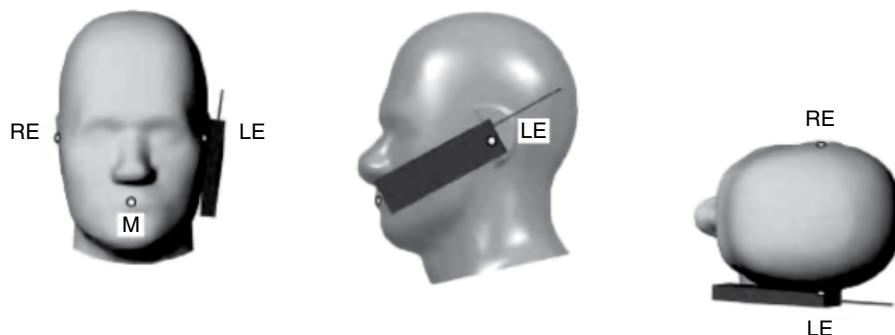


Figure 6.9 Front, side, and top views of left ear, check/touch position (IEEE [3]). M, mouth; RE, right ear; LE, left ear. (Source: Reproduced with permission of IEEE.)

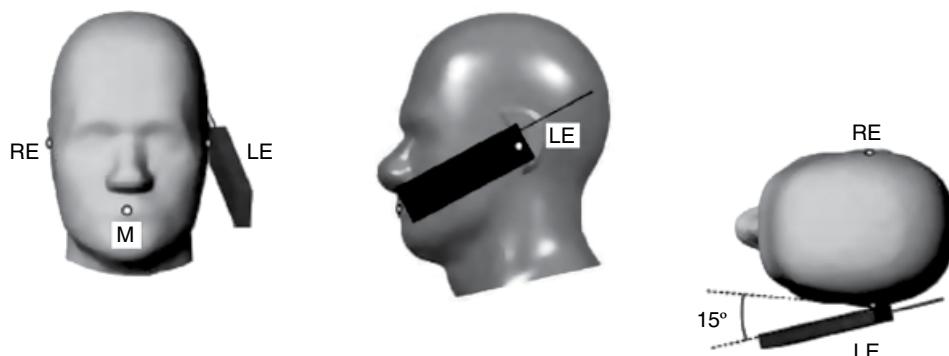


Figure 6.10 Front, side, and top views of left ear, 15° tilt position. (Source: Reproduced with permission of IEEE.)

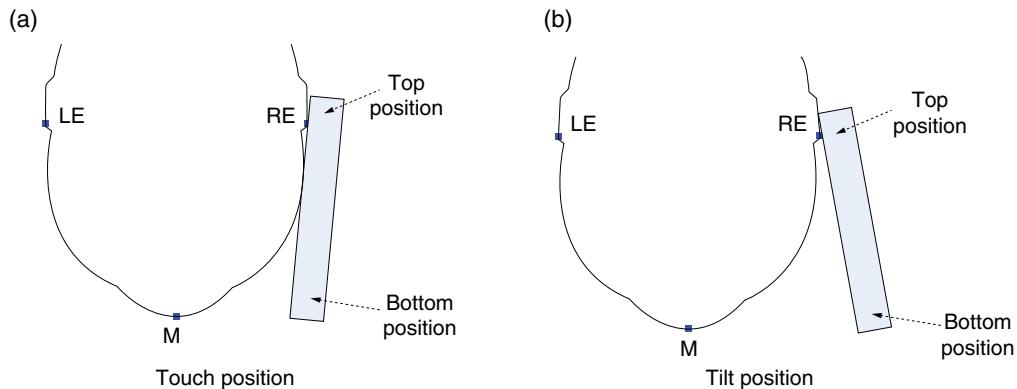


Figure 6.11 Putting a phone next to an LE–M–RE curve.

As shown in Figure 6.9, when a phone is correctly placed on a touch position, the phone's vertical center line should fall into the plane defined by RE (right ear), M (mouth), and LE (left ear) points. The center of the phone's earpiece should touch the LE point while another point on the front side is also in contact with the cheek of the phantom.

As shown in Figure 6.10, while maintaining the contact between the phone and the phantom, then pivoting against the ear, moving it outward away from the mouth by an angle of 15° , what we get is the tilt position. In both touch and tilt positions, the ear piece of a phone always touches the phantom ear.

It is obvious that the distance between a phone and the surface of a phantom head is not a consistent value. The distance is one of the two most important factors to control SAR. The further the distance, the lower the SAR value. The other most important factor is the near-field energy distribution. The actual SAR inside a phantom head is a combining effect of both factors.

To control the SAR is to play with those two factors. An antenna engineer should actively participate from the concept stage, which is the first stage when designing a phone. At the beginning stage of a project, most efforts should be focused on securing a correct location and a decent volume for antenna.

As per the requirement of SAR tests, the vertical center line of a phone is in the same plane defined by the LE, M, and RE points. We can extract the contours of a phantom and a phone in the LE–M–RE plane, as shown in Figure 6.11, to study their relative distance.

If an antenna is installed in the top position, there is little gap between the phone and the head in both touch and tilt positions. Whip antenna, stubby antenna, and PIFA can all be used in the top position. From the SAR point of view, the main difference between a PIFA and an internal monopole antenna is the existence of the ground beneath the radiating element. The ground in a PIFA can divert the energy away from the head; however, it can also shrink the antenna's bandwidth. If you want to put an internal antenna on the top, you have to use a PIFA. If you want to have a smaller antenna, the internal monopole antenna can be used and must be located at the bottom.

If you ever look at those phones with a stubby antenna, you might note that stubby antennas are normally off-center and are closer to the backside of phones, away from the head. This is a measure to increase the distance, thus decreasing the SAR. Another frequently used technique is tilting the stubby antenna, as shown in Figure 6.12. Tilting the stubby antenna has

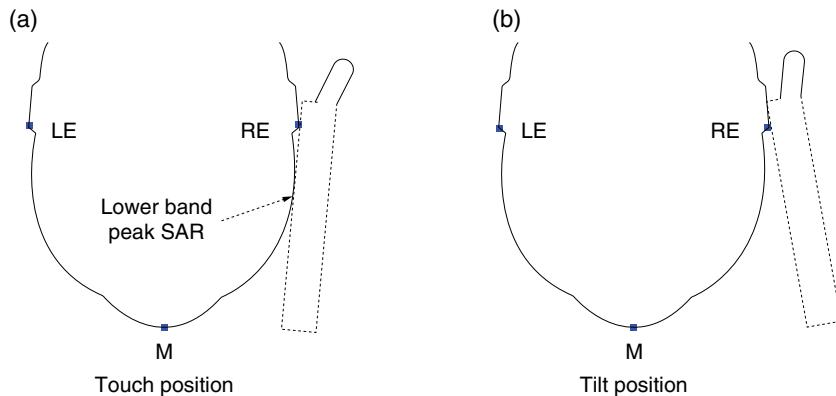


Figure 6.12 Tilting the stubby antenna.

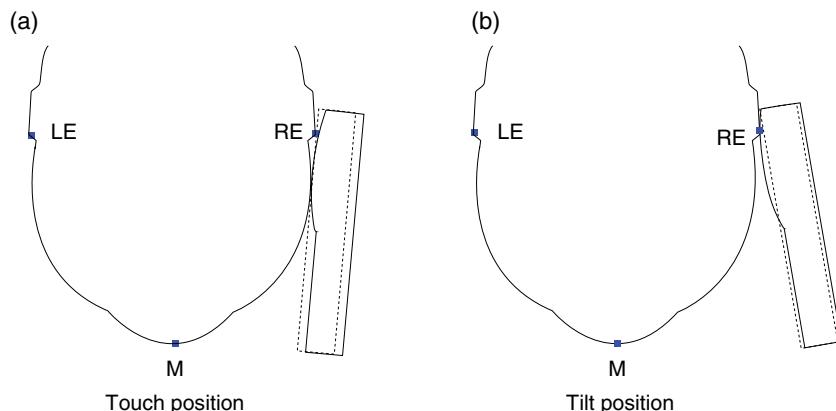


Figure 6.13 Localized thickness increasing.

more effect at the high band. At the lower band, the peak of near-field energy is located in the middle portion of the phone, which is also the contact point between a phone and a phantom head. Thus, the peak SAR is most likely to appear around the touch point on the phantom cheek. Tilting the antenna has little effect on the energy distribution on the board at the lower band, thus has little impact on the specific SAR.

Increasing a phone's thickness at the LCD display area is another commonly used technique. The solid line and the dashed line contours, as shown in Figure 6.13, illustrate phones with and without localized thickness increasing, respectively. It is obvious that the phone has effectively moved away from the phantom head.

As shown in Figure 6.14, by extruding the navigation key or adding a bumper at the center, a phone can also be effectively shifted away from the phantom head. For a bottom-installed antenna, the lever effect can magnify the small extrusion at the middle to a large separation at the bottom, thus decreasing the SAR more significantly. An extruded feature at the center is

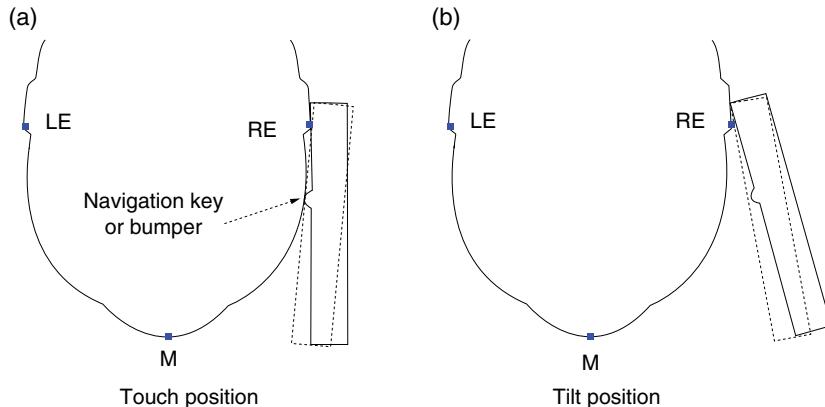


Figure 6.14 Extruded navigation key or bumper.

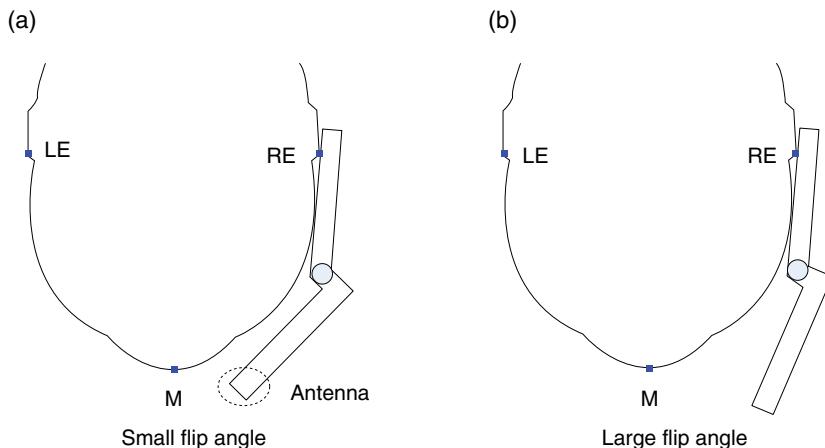


Figure 6.15 Flip angle on a clam shell phone, bottom antenna.

not only a useful SAR reduction measure but also an important feature from the mechanical point of view. Apart from glass screens used in some high-tier phones, most phones have LCD screens which are made of plastic and can easily be scratched. The extruded feature can keep the LCD screen away from the rough surface when the phone is placed faceside down.

For a clam shell phone, the flip angle is a critical parameter which decides the SAR. As shown in Figure 6.15, the larger the flip angle, the further the lower part of the phone away from the phantom head. If a phone has a bottom-installed antenna, the relation between the flip angle and the SAR is monotonous. The large angle is always better.

When a clam shell phone has a stubby antenna, the relation between the flip angle and the SAR value is a little more complex. At the low band, the circuit board inside the phone is the main radiator. So when the flip angle is small enough, as shown in Figure 6.16a, the bottom part of the phone is too close to the head and the SAR value starts to increase. On the other

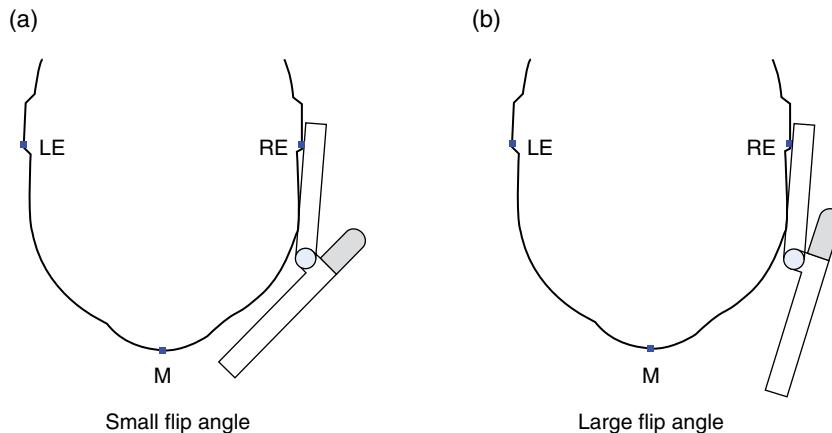


Figure 6.16 Flip angle on a clam shell phone, stubby antenna.

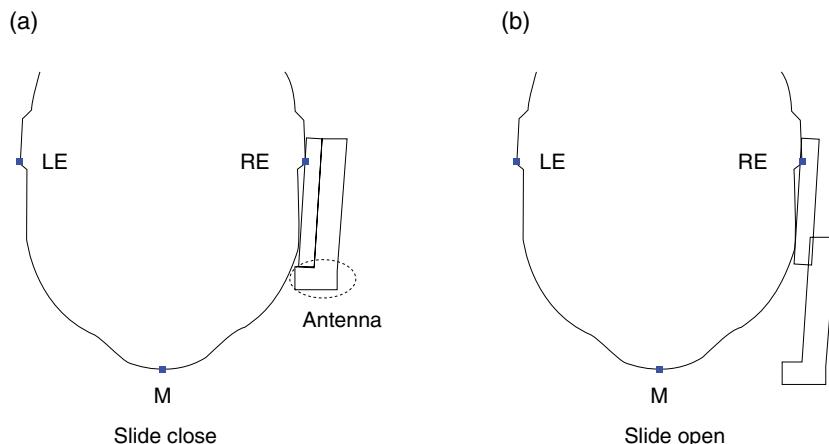


Figure 6.17 Slide phone.

hand, if the flip angle is too large, as shown in Figure 6.16b, the antenna itself becomes too close to the head, which also increases the SAR value. It is obvious that there should be an optimal flip angle in between which can balance the SAR contributions from both the phone body and the antenna radiator.

From some perspectives, slide phones are kind of similar to flip phones, as both of them have moving halves. Some slide phones have top-installed PIFAs, and the SAR design consideration of those slide phones is pretty much the same as a single-piece candy-bar phone. For some slide phones, the antenna is at the bottom. When an antenna is placed at the bottom, the design team expects the antenna solution to be a space-saving monopole antenna. The only issue here is that a slide phone can make a phone call in both close and open positions. When doing a SAR test, both slide open and close positions must be measured. As shown in Figure 6.17, in the slide close position, a bottom-installed antenna is too close to the head. Unlike a PIFA, the monopole

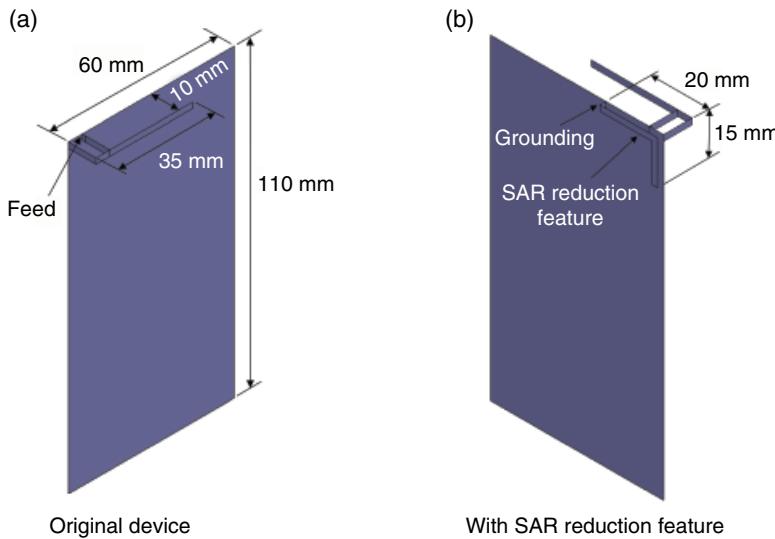


Figure 6.18 A demonstration of SAR reduction.

solution does not have a ground to divert the energy away from the head. A frequently used method to decrease SAR in such circumstances is mismatch. When a phone can be used in multiple configurations, most service providers only have a stringent requirement at the primary user configuration, which is the slide open position in this case, but a relative loose specification for other configurations. The antenna can be optimized for the slide open position and purposely mismatched in the slide close position, thus reducing the power radiated toward the head.

So far, all SAR reduction methods we have discussed focus on how to increase the distance between a phone and a phantom head. The other SAR impact factor, which is as important as the distance, is the near-field energy distribution. To some degree, by choosing an antenna solution, say, stubby, PIFA, or internal monopole, the near-field energy distribution is also selected. It is obvious that the near-field distribution of a PIFA is better than an internal monopole, and a bottom-installed monopole antenna is better than a top-installed one. However, there are other ways to influence the near-field distribution than merely selecting antenna type.

Figure 6.18 is a demonstration of a SAR reduction technique. Figure 6.18a shows the original device. The antenna used here is an IFA, and the length of the antenna's horizontal arm is 35 mm. In Figure 6.18b, an L-shaped SAR reduction feature is implemented, which is located on the opposite side of the ground plane to the antenna. The total length of the L strip is 35 mm, its width is 2 mm, and the distance between the strip and the ground is 2 mm.

The resonant frequency of the IFA is 1.8 GHz. As the dimensions of the IFA and the SAR reduction L strip are similar, it is safe to say the resonant frequency of the L strip is also around 1.8 GHz. Shown in Figure 6.19 are the reflection coefficients of the IFA with or without the SAR reduction feature. It is clear that the variation introduced by the SAR reduction feature is negligible.

Figure 6.20 shows the simulated SAR distribution plots. A flat body phantom is used in the simulation. The simulation liquid is 1800 MHz body fluid. The gap between the ground plane

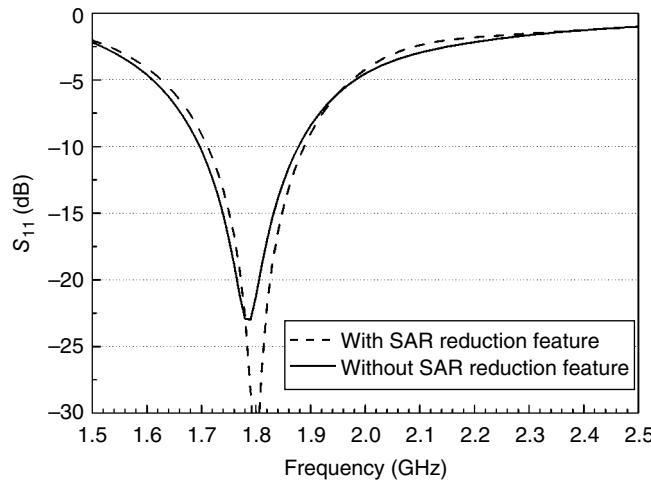


Figure 6.19 Impact of SAR reduction feature on reflection coefficient.

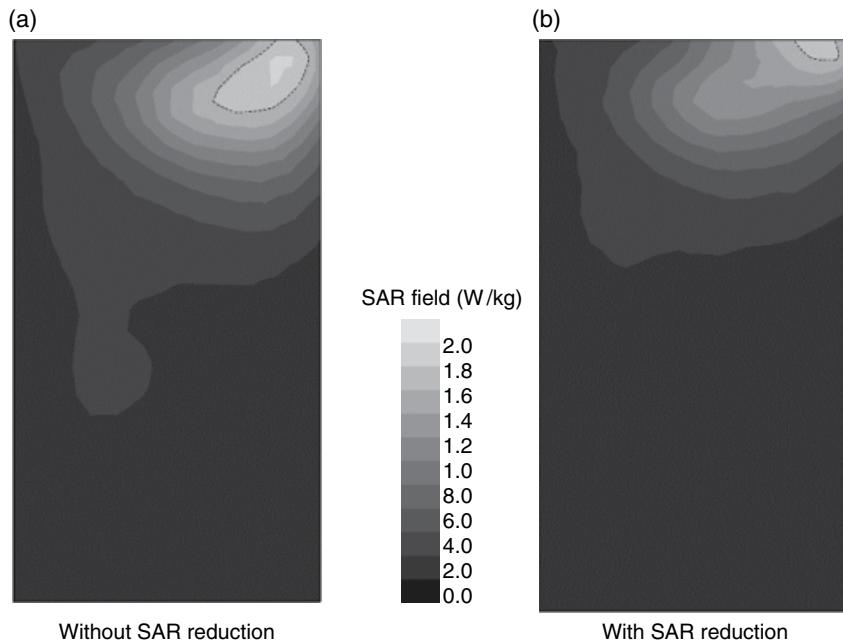


Figure 6.20 SAR distribution of an IFA.

and the phantom is 10 mm. The output power is 0.15 W. The dashed lines in Figure 6.20a and b are the contour lines in which SAR equals 1.6 W/kg. By introducing the SAR reduction feature, three effects are realized: (1) the peak SAR is decreased, (2) the area of SAR greater than 1.6 W/kg is shrunk, and (3) the location of the peak SAR is moved. The reason for SAR reduction in the first two is quite obvious. The third one, altering SAR distribution, is also an

important technique in decreasing SAR. As has been discussed in the earlier part of the section, when a phone is put next to a phantom head, the distance between the printed circuit board (PCB) and the head is not a constant. SAR reduction can be achieved even without suppressing the peak energy level on the PCB. By simply shifting the hot spot toward the less critical area on the PCB, where the distance between the PCB and the phantom head is greater, the SAR value can be decreased.

Some other guidelines related to the SAR reduction are listed as follows:

- As the SAR is a specification only with regard to the highest value which is averaged in a small cubic, the SAR can be decreased by spreading the energy more evenly over the PCB. By doing this, the total radiation toward the phantom head is pretty much the same, however, the SAR value will be lower.
- The location of a feeding point also plays a role on the SAR value. As a qualitative trend, the corner of a PCB is the best location from the point of view of antenna efficiency and bandwidth; however, it also generates a higher SAR value.
- Decreasing the conductive power of the RF transmitter is always the last resort. Nobody likes it, as it deteriorates a phone's performance and makes the R&D management team looks bad. Most of the time, their superior will only look at the numbers. For them, a 30 dBm power level is always much better than a 29.8 dBm one, although there is only 0.2 dB difference and nobody can tell the difference in the real world. As the SAR values at low, middle, and high channels are normally different, if it is possible, only adjust the power level at the channel which has SAR issues. By doing this, the perception of the phone specification value might look acceptable.

In some phones, there may not be enough space for any extra SAR reduction feature. The SAR reduction can be realized by utilizing existing metal structures. By selecting grounding locations of the metal frame of the LCD panel, the metal phone bezel, and other metal objects on the phone, a certain level of SAR reduction can still be achieved.

The SAR measurement is quite a time-consuming process. Even a single channel measurement, which includes an area scan and a fine scan, will take nearly 10 minutes. One way to save time is by using single-point measurement. You should be able to move the test probe to the peak SAR location and measure the instant SAR value at the location. Depending on the test equipment and the software version, the actual way to invoke the single-point measurement varies. If you have difficulty locating that function, refer to the equipment manufacturer. When you try out different SAR reduction configurations, you actually only need to check out the single-point instant SAR value. If that value does not decrease, it means the current configuration does not work. A single-point measurement takes at most 1 minute, so it can give you more time to try out different designs.

When you get some configurations which generate promising value at single-point measurements, you need to carry out a full SAR measurement. It is possible that only the location of the peak SAR has shifted and the SAR value has not been decreased. Then you must keep working. If a SAR reduction is really observed, you need to recheck the antenna's performance to make sure it still works well. Sometimes, the SAR reduction is due to the antenna's performance degradation and that is still not a working solution. After you are sure everything works correctly, you must duplicate the SAR solution on several phone samples and check the consistency of the solution. As an engineer, you should never broadcast the great news until you have done all the tests.

Always bear in mind that SAR reduction techniques can only help to some extent. If the SAR value is 20% above the specification, it is appropriate to seek help from SAR reduction techniques. If the SAR value is 100% higher than the specification, the design is totally flawed and it might have to be redesigned from scratch.

6.1.4 Updates on SAR Requirement

Since the first edition of the book, cellular technology has progressed a lot. The 3G network reached its prime and then handed the torch to the 4G network. Now 4G has also matured and even the most sub-100 USD phones are 4G ready. People are working toward 5G system. As most 4G phones are back-compatible, the SAR qualification procedure has become a very long laundry list. In the United States, the current standards applied to SAR measurement are FCC 47 CFR § 2.1093, IEEE STD 1528-2003, TCB workshop notes GPRS testing considerations, and FCC-published RF exposure KDB procedures listed later. One thing for certain is, with the progress of cellular technology, this list will be longer and longer.

- 447498 D01 General RF Exposure Guidance
- 648474 D04 Handset SAR
- 941225 D01 SAR test for 3G devices
- 941225 D02 HSPA and 1x Advanced
- 941225 D03 SAR Test Reduction GSM GPRS EDGE
- 941225 D04 SAR for GSM E GPRS Dual Xfer Mode
- 941225 D05 SAR for LTE Devices
- 941225 D05A LTE Rel.10 KDB Inquiry Sheet
- 941225 D06 Hotspot Mode SAR
- 248227 D01 SAR Meas for 802 11abg
- 865664 D01 SAR Measurement 100 MHz–6 GHz
- 865664 D02 SAR Reporting

When doing an official FCC qualification test, there are many detailed considerations. The best way to grab those details is from published SAR reports of FCC-approved phones [7]. The contents in the book are only a brief introduction. For each phone model, required SAR measurements of cellular bands are combinations of the following configurations.

- Multiple 2G, 3G, and 4G standards: For 2G, there are GSM and CDMA. For 3G, there are code division multiple access (CDMA) 2000, WCDMA, and TD-SCDMA. For 4G, there are TD-LTE and LTE FDD.
- For each standard, there are also multiple frequency bands.
- For each frequency band, there might be different substandards, such as voice and data.
- For each communication setting, there are several holding positions. Beside the head positions and body-worn positions discussed in Section 6.1.3, hotspot positions are added and will be discussed later.

Beside cellular bands, SAR of Wi-Fi bands (2.4 and/or 5 GHz) are also required. For those phones with more than one Wi-Fi antennas, SAR of each Wi-Fi antenna needs to be measured separately.

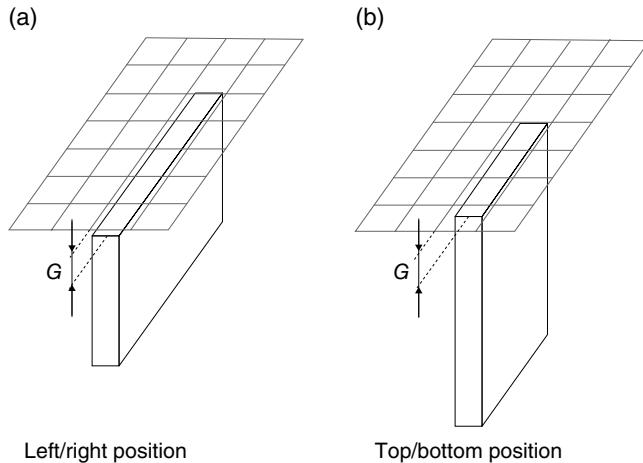


Figure 6.21 SAR distribution of an IFA.

The newly added hotspot positions include six positions and are all measured by using a body phantom. There are six surfaces, which are front, back, top, bottom, left, and right surfaces, on a phone. Each hotspot position is formed by placing one phone surface next to a body phantom.

Shown in Figure 6.21 are four out of six positions. Figure 6.21a are left or right hotspot positions and Figure 6.21b are hotspot top or bottom positions. The separation distance between a phone's surfaces and scanning grids, marked as G in Figure 6.21, is set by the FCC procedure “941225 D06 Hotspot Mode SAR.” For big devices ($>9\text{ cm} \times 5\text{ cm}$), the gap is 10cm. The gap is 5cm for smaller devices.

The front and back hotspot positions, which are omitted in Figure 6.21, are similar to body-worn positions. The only difference is the fixed separation distance G , which can be freely chosen when measuring body SAR. In most cases, the G of hotspot positions is considerably smaller than body positions.

When measuring body positions, G can be adjusted to meet SAR requirement. In the hotspot positions, the freedom of choosing G has gone and in return we get the freedom of adjusting the output RF power. Because a phone knows when it is in the hotspot mode, it indeed can dial down its output power based on its preset power table.

Once upon a time, a phone can only do one thing at a time and never turns on more than one transmitter. Nowadays, phones are just like human and capable multitasking. Today’s phone might simultaneously turn on up to four transmitters. For example, someone is sharing 4G data through Wi-Fi in the hotspot mode, a 2G voice call comes in, and he/she picks the call through a Bluetooth headset.

For simultaneously transmitters, one SAR evaluation method is adding SAR values of all simultaneously working transmitters. If the sum value can meet the SAR requirement, no more SAR evaluation is necessary. In fact, the simultaneous transmission of SAR is not really a challenge. Because a phone always knows how many transmitters are simultaneously working, it can legally dial down its output power accordingly. The hotspot mode is one kind of simultaneous transmission cases. It needs to turn on at least a Wi-Fi transmitter for routing and a cellular band transmitter for data connection. For some phones, which have the Wi-Fi multiple input and multiple output (MIMO) capability, Wi-Fi mode itself can also be a simultaneous transmission scenario.

6.2 Hearing Aid Compatibility

Just as its name implies, hearing aid compatibility (HAC) [9] is a standard related to hearing aids. Most users might have experienced some interference caused by cellular phones. When a cellular phone is put next to a computer's speaker, there is always some buzz noise out of the speaker before an incoming call rings the phone. Similar phenomena happen when an SMS has arrived. You can also hear this kind of noise when you are having a conversation over a landline phone if in the meantime someone is calling you on your cellular phone.

There is a misconception that what we hear is RF noise. The frequency range of RF signal used in cellular communication starts from several hundreds of megahertz. Nothing can generate an audio wave at that frequency. Even if someone could generate a sound at that frequency, we could not possibly hear it. The frequency of an RF signal is thousand times higher than the upper limit of what our ears can hear. However, the noise we hear is indirectly caused by the RF signal. Before a call is established or an SMS is received, a base station is communicating with a cellular phone through a burst of signaling. A speaker functions as an envelope detector, which can pick up the amplitude variation of the RF signal. The frequency spectrum of the RF signal's envelope falls within the audible range of our ears, that is the reason why we can hear it.

A hearing aid is much more sensitive than a speaker, and the distance between a phone and a hearing aid is much closer than between a phone and a speaker. So it is no surprise that a phone can cause a hearing aid to malfunction. The purpose of the HAC standard is to set a specification which can guarantee that a phone works well with a hearing aid.

If a phone cannot pass SAR, a customer is correct to wonder whether it is wise to buy the phone. For customers without hearing aids, whether a phone passes HAC or not is nothing to do with them. The HAC specification is not a safety-related issue, at least for most customers. However, it is mandatory in the United States that both phone manufacturers and wireless service providers must have a certain percentage of their phone models pass the HAC specification.

6.2.1 HAC Measurement

When measuring SAR, it is actually gauging the heating effect due to the conductive loss. The conductive losses are directly related to the E field, so only the electric field needs to be measured. When talking about HAC, as both the E field and the H field can interfere with a hearing aid, both fields must be measured.

The HAC measurement capability is generally provided as an extension of a SAR system. For antenna engineers, an HAC measurement relates to two probes: the E -field probe and the H -field probe. It should be no surprise that the E -field probe used in HAC is identical to the one used in SAR. The H -field probe is used for HAC measurement only. Similar to an E field, an H field is also a vector field and it is composed of three orthogonal components as shown in Equation 6.4:

$$H = \sqrt{H_x^2 + H_y^2 + H_z^2} \quad (6.4)$$

Inside an H -field probe, there are actually three individual H -field sensors. Each sensor measures one orthogonal H -field component. As shown in Figure 6.22, there are three metal wire rings inside an H -field probe. The axes of rings are perpendicular to one another. Each ring functions as a magnetic dipole and can detect the H -field along its axis. Of course, Schottky diodes are needed as rectifiers to convert the RF signal to the voltage signal; high

impedance lines are also needed to transmit the voltage signal back. Both of them are omitted in Figure 6.22. The *E*-field probe and the *H*-field probe provided by the same company normally use the same type of connectors. Thus, an *H*-field probe can be connected to the data acquisition unit used in SAR measurements.

Shown in Figure 6.23 are *E*-field and *H*-field probes. Although an *H*-field probe will never be inserted into tissue liquid during an HAC measurement, it still uses the same sealing package used by the *E*-field probe. It is quite easy to distinguish an *H*-field probe from an *E*-field probe if they are made by the same manufacturer. The dimensions of an *H*-field probe are normally larger than an *E*-field probe. One might think it is because a structure made of metal rings is more difficult to miniaturize. If it were the real reason, why couldn't we make the *E*-field probe bigger? This way the packages of both probes can be standardized and the logistical burden can be eased a little bit.

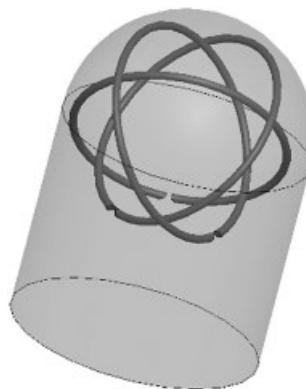


Figure 6.22 *H*-field probe.

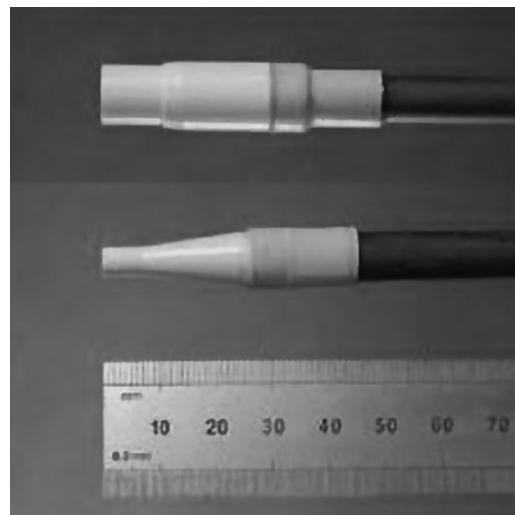


Figure 6.23 *H*-field and *E*-field probes. (Source: IndexSAR Ltd.)

The correct answer can be found in most EM textbooks [10]. The radiation resistance R of a Hertzian (electric) dipole is given by Equation 6.5. The radiation resistance is a quantitative property which tells how well a device is capable of radiating or receiving signals. The higher the value, the better the device.

$$R_{\text{Electric-Dipole}} = 80\pi^2 \left[\frac{dl}{\lambda} \right]^2 \quad (6.5)$$

Here, dl is the length of the electric dipole and λ is the wavelength. The radiation resistance R of a magnetic dipole is given by Equation 6.6:

$$R_{\text{Electric-Dipole}} = \frac{320\pi^4 S^2}{\lambda^4} = \frac{20\pi^6 S^4}{D^4} \quad (6.6)$$

Here, S is the area of the magnetic dipole and D is the diameter of the ring.

Both Equations 6.5 and 6.6 are valid approximations only when dipoles are electrically small, which means dl/λ or $D/\lambda \ll 1$. An electric dipole's radiation resistance is proportional to $(D/\lambda)^2$ and a magnetic dipole's is proportional to $(D/\lambda)^4$. The radiation resistance of an electric dipole is always higher than a magnetic dipole with the same external dimensions. For example, if the total length of an electric dipole is $\lambda/20$, its radiation resistance is 1.97Ω . The radiation resistance of a magnetic dipole with a $\lambda/20$ diameter is 0.12Ω . As both the E -field probe and the H -field probe share the same data acquired unit, it is preferable to minimize the difference between their output voltages. Therefore, the H -field probe is always bigger.

Shown in Figure 6.24 is the required HAC test grid [9, 11]. The grid is 50.0mm by 50.0mm area that is divided into nine evenly sized blocks or subgrids. The grid is centered on the audio frequency

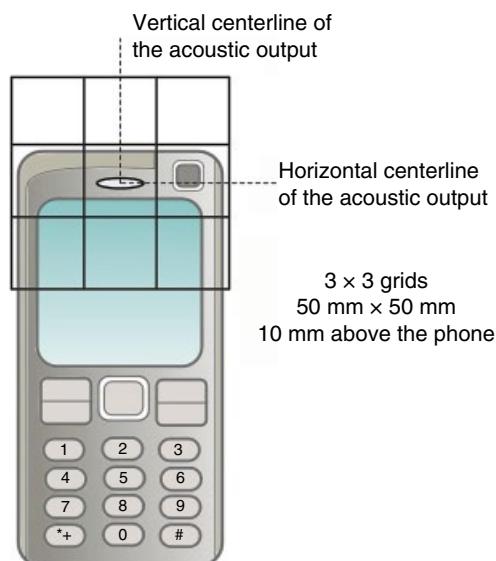


Figure 6.24 HAC measurement plane.

output speaker of the phone. The measurement plane is 10.0mm above the phone's top surface. Unlike a SAR test, which requires a 3D surface scanning, an HAC test only needs a simple 2D scanning. In a real HAC system, measurements are first taken on much finer grids, say, at 2 or 5 mm increments. The final 3×3 grid is then obtained through the average of the raw measurement data.

6.2.2 HAC Specification in the United States

The current HAC specification in the United States is according to ANSI C63.19-2007. The limit value of ANSI 63.19-2007 is the same as ANSI 63.19-2006. The version used before the 2006 one was ANSI C63.19-2001. Shown in Table 6.5 are the specifications of ANSI C63.19-2007. The top half of the table is the specification for devices operating in bands below 960 MHz, and the bottom half is for devices operating above 960 MHz. In each frequency range, there are four categories. An M4 grade device has the lowest RF emission, so it is the best from the RF emission point of view. The articulation weighting factors (AWFs) used for the standard transmission protocols is shown in Table 6.6. The AWF is set as 0 dB for

Table 6.5 Telephone near-field categories in linear units

Category	Telephone RF parameters <960 MHz		
	Near field	AWF	E-field emissions (V/m)
Category M1/T1	0	631.0–1122.0	1.91–3.39
	-5	473.2–841.4	1.43–2.54
Category M2/T2	0	354.8–631.0	1.07–1.91
	-5	266.1–473.2	0.80–1.43
Category M3/T3	0	199.5–354.8	0.60–1.07
	-5	149.6–266.1	0.45–0.80
Category M4/T4	0	<199.5	<0.60
	-5	<149.6	<0.45
Category M1/T1	0	199.5–354.8	0.60–1.07
	-5	149.6–266.1	0.45–0.80
Category M2/T2	0	112.2–199.5	0.34–0.60
	-5	84.1–149.6	0.25–0.45
Category M3/T3	0	63.1–112.2	0.19–0.34
	-5	47.3–84.1	0.14–0.25
Category M4/T4	0	<63.1	<0.19
	-5	<47.3	<0.14

Table 6.6 Articulation weighting factor

Standard	Technology	AWF (dB)
TIA/EIA/IS-2000	CDMA	0
TIA/EIA-136	TDMA (50Hz)	0
J-STD-007	GSM (217Hz)	-5
T1/T1P1/3GPP	UMTS (WCDMA)	0
iDEN	TDMA (22 and 11Hz)	0

almost all protocols except the GSM, whose AWF is 5 dB. Actually, the GSM protocol is also a TDMA system, just like EIA-136 and iDEN; however, its repeating frequency is 217 Hz and its fundamental and harmonical components fall within the audible range of human beings (20Hz–20kHz). To take this into consideration, a 5 dB penalty was enforced on GSM. The 5 dB is calculated by $20 \cdot \log_{10} (V/m \text{ or } A/m)$, so when calculated in V/m or A/m, the GSM's limit equals 56% of other standards.

For antenna engineers, the M3 category is what they are aiming for. When designing a GSM phone, the limits for *E*-field emission are 266.1 and 84.1 V/m for bands below and above 960 MHz, respectively. The limits for *H*-field emission are 0.8 and 0.25 A/m. A phone must pass all four limits to get an M3 rating, then it can be labeled as HAC approved. In reality, nobody really works toward an M4 grade. If a phone happens to pass the M4 limit, engineers will be happy with the result.

For CDMA or UMTS phones, the *E*-field limit is 354.8 and 112.2 V/m, respectively. The *H*-field limit is 1.07 and 0.34 A/m, respectively. The discrimination between bands below and above 960 MHz is a relatively new thing. Before the implementation of ANSI 63.19-2006, the limits for all frequency bands were the same, which is identical to what the high band value is today. During that period, hardly any dual-band candy-bar phone, except a few bulky models, could get M3 approval. Phone manufacturers depended on flip phones, slide phones, and so on, to comply with the FCC regulations. For most phones which could not get the HAC approval at that time, the low band is the problem. After many antenna engineers had struggled for thousands of hours, one fine day the HAC Committee made the famous decision: as the technology of the hearing aid has evolved significantly and almost nobody can pass the low band limit, let's increase the limit by 10 dB. The 10 dB increase equals to a 316% increase in V/m or A/m. This is the unofficial birth story of ANSI 63.19-2006. The official version of the story is that some studies showed that most contemporary hearing aids have more immunity in bands below 960 MHz than above. Then combined with the consideration of different transmitting power levels at different bands, the standard was revised to have the band-dependent limits. I still wonder why, if the lower band limit can be raised by 10 dB, why don't they lift the high band limit also by 10 dB? Then all antenna engineers can stop worrying about HAC altogether and live happily ever after.

As shown in Figure 6.25, the measurement grid defined in C63.19 consists of nine evenly sized blocks, which are used to define permissible exclusion areas. For both

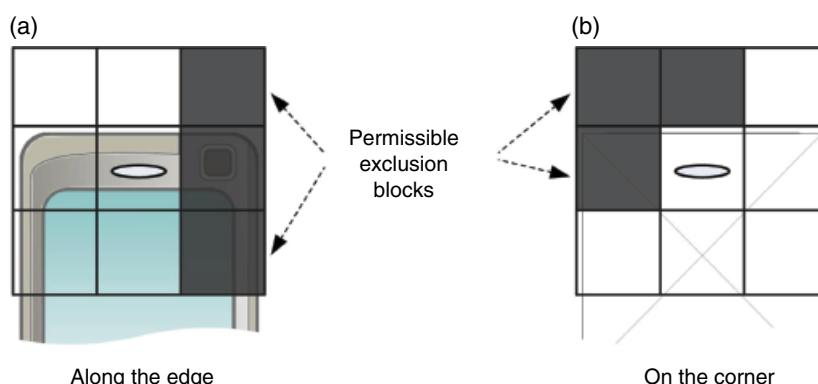


Figure 6.25 Exclusion block placement.

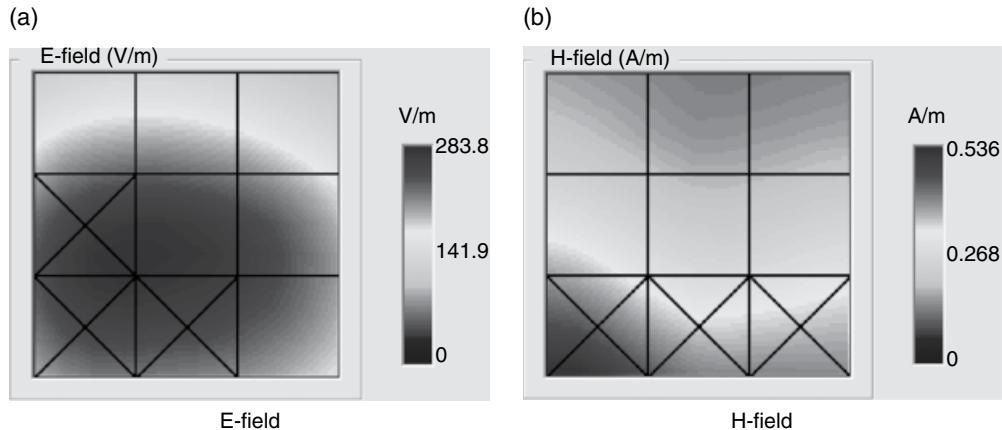


Figure 6.26 Measurement result of a sample phone (at CDMA 850, Channel 384).

E-field and *H*-field measurements, three contiguous blocks may be excluded from the measurements except for the center block that may never be excluded. Both exclusion examples shown in Figure 6.25 are permissible. There must be four blocks left that are common to both *E*-field and *H*-field measurements, so a maximum of five different blocks can be excluded (e.g., three blocks excluded from the *E*-field and two blocks from the *H*-field).

Shown in Figure 6.26 is the measurement result of a sample phone. In the *E*-field plot, three highest continuous blocks at the bottom-left corner are excluded. The exclusion does not help too much in this case, as the center block, which can never be excluded, has the strongest field distribution among all blocks. In the *H*-field plot, three blocks along the bottom edge are excluded. As the bottom-left block has the highest value, the exclusion of that block is actually helping the overall rating.

Most techniques used in SAR reduction can also be used in HAC reduction. However, the HAC is even trickier to work with. As the distribution of the *E* field and the *H* field are not correlated, if one field is suppressed, that might lead the other field to rise. Based on the rule of exclusion, shifting the peak away from the center block always helps.

So far, we have focused on the RF emission part of ANSI C63.19-2007. This is the part which is directly related to antenna engineers. However, the scope of ANSI C63.19-2007 is much wider than that. As an engineer, it is always good to know the whole picture. The standard was developed in cooperation with three main parties: representatives of organizations representing people with hearing loss, hearing aid manufacturers, and the digital wireless telephone industry. It regulates both hearing aids and cell phones. The overall frame of the standard can be simplified as follows:

- Wireless device
 - RF emission/category test (directly antenna related)
 - T-coil mode/category test.
- Hearing aid
 - RF immunity/category test
 - T-coil immunity/category test.

Table 6.7 Hearing aid near-field categories in linear units

Category	Hearing aid RF parameters	
	Near field	E-field emissions (CW) (V/m)
Category M1/T1	31.6–56.2	0.071–0.126
Category M2/T2	56.2–100.0	0.126–0.224
Category M3/T3	100.0–177.8	0.224–0.398
Category M4/T4	>177.8	>0.398

Note: Hearing aid must maintain lesser than 55 dB IRIL interference level and lesser than 6 dB gain compression.

Corresponding to the RF emission test for cellular phones, there is an RF immunity test for hearing aids. The RF immunity test evaluates how well a hearing aid can sustain RF emission. The ANSI C63.19-2007 limits for hearing aid are shown in Table 6.7. The hearing aid immunity is measured by using continuous wave (CW), and it is not frequency dependent. Similar to RF emission categories, the M4 grade is also the best one, which means a hearing aid can sustain the strongest RF emission.

To determine the compatibility of a phone and a particular hearing aid, simply add the numerical part of the hearing aid category with the numerical part of the phone emission rating to arrive at the system classification for this particular combination of phone and hearing aid. A total of four would indicate that the combination is usable; a total of five would indicate that the combination would provide normal use; and a total of six or greater would indicate that the combination would provide excellent performance. A category total of less than four would likely result in a performance that is judged unacceptable by the hearing aid user. In theory, the user experience of a combination of an M2 phone and an M3 hearing aid is similar to an M3 phone and an M2 hearing aid.

In ANSI C63.19-2007, there are contents about the T-coil in both the wireless device and hearing aid parts. For a phone, which must obtain HAC approval, the T-coil portion of the standard is not mandatory. The microphone inside any hearing aid can pick up the audio wave transmitted by a phone's speaker and then amplify it to a level which the user can hear it clearly. This coupling path is called "acoustic passage." For phones and hearing aids equipped with T-coils, there exists another passage. T-coil, also known as "Tele-Coil," was originally developed to support hearing aid use with landline telephone handsets that employ a magnetic earpiece transducer (such as those made in the mid-1980s). Later landline phones began to use piezoelectric transducers that did not generate an *H* field. Consequently, some landline phones and some cellular phones include a special inductor, which is also called a T-coil, specifically intended to generate a strong audio-band *H* field for T-coils in hearing aids. Hearing aids equipped with a T-coil can be configured by the user to disable the microphone and instead reproduce audio that is magnetically coupled to the T-coil.

If a phone is equipped with a T-coil, then the RF emission test might have to be repeated. For the T-Coil mode M-rating assessment, determine if the T-coil is contained in an included subgrid of the first scan, for both *E* fields and *H* fields. If so, then a second scan is not necessary. The first scan and resultant category rating may be used for the T-coil mode M rating. Otherwise, the test must be repeated with the 50 mm × 50 mm grid shifted so that it is centered

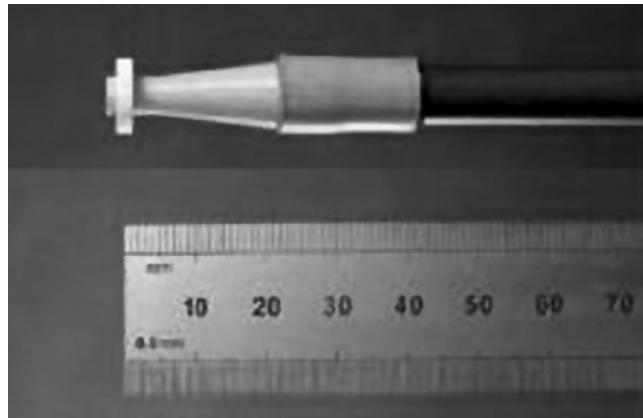


Figure 6.27 Probe for T-coil measurements. (Source: IndexSAR Ltd.)

on the axial measurement point of the T-coil. The lowest category obtained in the first or the repeated tests for either E -field or H -field determines the M category assessment.

Shown in Figure 6.27 is the probe for T-coil measurements. The probe is used to measure the audio-band magnetic field. Unlike the RF E -field or H -field probes, a T-coil probe can only measure one component of the vector H field at a time. Therefore the test must be repeated three times along three orthogonal axes. Detailed procedures of T-coil measurements are omitted in the book; more information can be found in ANSI C63.19-2006 [9].

Unlike SAR standards, a phone can still be sold in the United States even without an HAC certificate. The HAC regulation requests a certain percentage of all phone models sold in the United States by any device manufacturer or any wireless service provider are HAC certified [12]. The keyword here is “phone models.” In theory, a company can comply with the HAC regulation if only 1% of their phones have been HAC certified, as long as their model counts are more than the requirement.

6.2.3 Updates on HAC Requirement

The current HAC standard has been updated to ANSI C63.19-2011 [13]. The book only addresses some significant revisions. To get more information, please refer to the ANSI C63.19-2011, which can be purchased from either IEEE Xplore or webstore of ANSI.

The separation distance between test probes and a phone’s front surface used to be 10 mm for both E probe and H probe. It is now 15 mm for E probe and 10 mm for H probe. The distance is measured from the center of an E -probe’s dipole or an H -probe’s coil to the front surface of a phone.

AWFs have been replaced by modulation interference factor (MIF). The physics behind the MIF is the same as the AWF. If the MIF value of one standard is 3 dB, its actual E -field or H -field limit is 3 dB more stringent than the limit shown in Table 6.7. In the preceding standard, the AWF values are preset in the standard. However, the MIF values have to be measured for each phone models.

Shown in Table 6.8 are sample MIF values given in the ANSI C63.19-2011. The real values given in each HAC report are a little different, because they are measured on the spot.

Table 6.8 Sample MIF values

Transmission protocol	MIF (dB)
GSM; full-rate version 2; speech codec/handset low	+3.5
WCDMA; speech; speech codec low; AMR 12.2 kb/s	-20.0
CDMA; speech; SO3; RC3; full frame rate: 8kEVRC	-10.0
CDMA; speech; SO3; RC1; 1/8th frame rate; 8kEVRC	+3.3

Before starting any *E*-probe or *H*-probe measurement, MIF must be measured first. If one protocol's conducted power plus its MIF is less than +17 dBm for all its operating modes, the protocol can be exempt from further testing.

The required protocol lists are quite long, which can include GSM850, GSM1900, CDMA-Full, CDMA 1/8th, UMTS-RMC, UMTS-AMR, LTE-FDD, 2.4GHz WLAN, 5GHz WLAN, and so on. For each protocol, MIF needs to be measured on different channels, different data rates, different modulations, and any other combinations if they are applicable. However for most protocols, because either their transmit power is quite low or its MIF is significant small, their sum is less than +17 dBm and can be exempted. In most cases, the protocols, which have to go through *E*-probe or *H*-probe measurements, are still those old friends, GMS850, GSM1900, and CDMA 1/8th.

6.3 Electromagnetic Compatibility

Any phone model sold in the United States must pass the FCC EMC test. The EMC measurements are specified by requirements listed in FCC rules Parts 2, 22, and 24 of Title 47 of the Code of Federal Regulations [14–17]. For mobile phones, the EMC test is not a challenge at all. As most phones can pass them without a hitch, this topic is only touched upon in this book. However, for laptops and other devices with high-speed CPUs, passing the EMC requirement can be a challenge.

- RF power output
 - In an FCC report, both the conducted and radiated RF power are measured. However, only the radiated power needs to comply with the FCC limit. Based on FCC 24.232 (b) (c), mobile/portable stations are limited to 2 W effective isotropic radiated power (EIRP). For mobile phones, this limit is not a problem at all. If a phone has a 2 W EIRP, it will most likely have failed the SAR test already. The EIRP value given in an FCC report is measured in an EMC chamber, which is not as accurate as an antenna anechoic chamber. There might be a difference of several dB between EIRP values measured by an EMC chamber and an antenna chamber. Don't panic! This is normal and the value measured by the antenna chamber is more credible.
- Occupied bandwidth/emission bandwidth
 - Based on FCC 2.1049, for transmitters employing digital modulation techniques, the occupied bandwidth, that is the frequency bandwidth such as that below its lower and above its upper frequency limits, and the mean powers radiated are each equal to 0.5% of the total mean power radiated by a given emission will be measured when modulated by

- an input signal, such that its amplitude and symbol rate represent the maximum rated conditions under which the equipment will be operated.
- This test is a conductive test. The purpose of this test is to ensure that a working device will not interfere with other devices operating in adjacent channels.
 - Frequency stability
 - This test is also a conductive test. It measures a device's frequency stability under different temperatures and battery voltages. The purpose of this test is to ensure that the fundamental emission stays within the authorized frequency block.
 - Conducted spurious emissions
 - Based on FCC 2.1051, the RF voltage or power generated within the equipment and appearing on a spurious frequency will be checked at the equipment output terminals when properly loaded with a suitable artificial antenna. Curves or equivalent data will show the magnitude of each harmonic and other spurious emissions that can be detected when the equipment is operated under the conditions specified in FCC 2.1049 as appropriate. The magnitude of spurious emissions which are attenuated more than 20 dB below the permissible value need not be specified.
 - Depending on frequency bands, FCC 22.917 and FCC 24.238 set the limitations, respectively. The power of any emission outside of the authorized operating frequency ranges must be attenuated below the transmitting power (P) by a factor of at least $43 + 10 \log(P)$ dB. The purpose of this test is to ensure a device will not interfere with other devices working in the frequency spectrum between 30 MHz and 18 GHz.
 - Radiated spurious emissions
 - Based on FCC 2.1053, measurements will be made to detect spurious emissions that may be radiated directly from the device under normal conditions of operation. Curves or equivalent data will be supplied showing the magnitude of each harmonic and other spurious emissions.
 - The test is carried out by a spectrum analyzer and/or an EMC/EMI receiver. The measured frequency range is from 30 MHz to 18 GHz. A high-speed clock bus, leakage from the local oscillator, and so on, are all possible root causes of EMC issues. If the signal path of a device is well shielded, it can pass the conducted spurious emissions test but fail the radiated one. The solution to radiated spurious emissions is shielding. By confining noise sources in equivalent "Faraday cages," the problem can be solved.
 - AC power line conducted emissions
 - For equipment that is designed to be connected to the public utility (AC) power line, the RF voltage that is conducted back onto the AC power line on any frequency or frequencies within the band 150 kHz to 30 MHz will not exceed the limits. If a device fails the AC conducted emissions test, the easiest solution is to add a ferrite choke on the connection cable. If you look around your office, you can find "bumps," which is the nickname for choke, all over the place. They are on power cords, USB cables, and monitor VGA/DVI cables.

References

- [1] "FCC Measurement Procedures," <http://www.fcc.gov/oet/ea/eameasurements.html>. Retrieved 25 October 2010.
- [2] Godara, L.C. (2001) *Handbook of Antennas in Wireless Communications*, 1st edn, CRC Press.
- [3] "IEEE Standard. 1528–2013, Recommended Practice for Determining the Peak Spatial-Average Specific Absorption Rate (SAR) in the Human Body Due To Wireless Communications Devices," (2013). <http://standards.ieee.org/findstds/standard/1528-2013.html>. Retrieved 25 October 2010.

- [4] "Tissue Simulating Liquids," <http://www.speag.com/products/dasy6/tissue-simulating-liquids/>. Retrieved 25 October 2010.
- [5] Faraone, A., Mccoy, D.O., Chou, C.K., and Balzano, Q. (2000) Characterization of miniaturized E-field probes for SAR measurements. IEEE International Symposium on Electromagnetic Compatibility, Washington, DC, pp. 749–754.
- [6] "IEC 62209-1, Human Exposure to Radio Frequency Fields from Hand-Held and Bodymounted Wireless Communication Devices," (2005) http://webstore.iec.ch/webstore/webstore.nsf/Artnum_PK/33746. Retrieved 11 July 2016.
- [7] "Federal Communications Commission, Office of Engineering and Technology, Equipment Authorization Search," <https://apps.fcc.gov/oetcf/eas/reports/GenericSearch.cfm>. Retrieved 25 October 2015.
- [8] "FCC KDB Publication 648474: SAR Evaluation Considerations for Handsets with Multiple Transmitters and Antenna," <https://apps.fcc.gov/eas/comments/GetPublishedDocument.html?id=254&tn=555816>. Retrieved 25 October 2010.
- [9] "ANSI C63.19-2011, American National Standard for Methods of Measurement of Compatibility between Wireless Communication Devices and Hearing Aids," <https://standards.ieee.org/findstds/standard/C63.19-2011.html>. Retrieved 25 October 2010.
- [10] Sadiku, M.O. (2009) *Elements of Electromagnetics*, 5th edn, Oxford University Press, USA.
- [11] "Test Plan for Hearing Aid Compatibility," http://files.ctia.org/pdf/CTIA_HearingAidComp_TestPlan_Rev1.0.pdf. Retrieved 25 October 2010.
- [12] "FCC: Amendment of the Commission's Rules Governing Hearing Aid-Compatible Mobile Handsets," http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-08-68A1.pdf. Retrieved 25 October 2010.
- [13] "ANSI C63.19-2011, American National Standard for Methods of Measurement of Compatibility between Wireless Communication Devices and Hearing Aids," <http://webstore.ansi.org/RecordDetail.aspx?sku=IEEE%2fANSI+Std+C63.19-2011>. Retrieved 25 October 2015.
- [14] "FCC Part 2: Frequency Allocations and Radio Treaty Matters; General Rules and Regulations," http://www.access.gpo.gov/nara/cfr/waisidx_02/47cfr2_02.html. Retrieved 25 October 2010.
- [15] "FCC Part 22: Public Mobile Services," http://www.access.gpo.gov/nara/cfr/waisidx_00/47cfr22_00.html. Retrieved 25 October 2010.
- [16] "FCC Part 24: Personal Communications Services, PCS (Narrowband PCS 901–902, 930–931, 940–941 MHz. Broadband PCS 1850–1990 MHz)," <http://www.ecfr.gov/cgi-bin/text-idx?SID=53ecaf58804a520c2b16c9c6ea7d3941&mc=true&node=pt47.2.24&rgn=div5>. Retrieved 25 October 2010.
- [17] "FCC Parts (Code of Federal Regulations, Title 47-Telecommunication)," <http://www.scc-ares-races.org/FCCparttitles.html>. Retrieved 25 October 2010.

Appendix: User Manual for ZJ_Antenna_Matching Software

ZJ_Matching is an antenna matching network optimization tool. It is an installation-free software. Just download and unzip the file and run the ZJ_Antenna_Matching.exe. There is another strip-down version, ZJ_Matching_Mini, which fits into a 640×480 pixels monitor and can be used on a Windows®-based vector network analyzer (VNA).

The software can be used to design a matching network for an antenna which can have up to two different scenarios, for example

1. To design a matching network for a whip–stubby antenna which has two positions: extended and retracted
2. To design an antenna matching network that gives a balanced performance between the free space and the talking position

To design a matching network for a cell phone antenna, many commercial software packages, such as Agilent ADS®, Microwave Office®, and so on, can be used. When I was working for the industry as an antenna engineer, I did feel, although those commercial codes were very powerful, they were not very efficient and were a kind of overkill when used in an antenna matching network designing, which normally uses less than four components. Thus, I decided to write this software in my spare time. The first version was written in 2003, and I have been modifying it since then. After I made my career U-turn in 2007, the software can now be released to public.

This is a freeware, so you can use and distribute it as you wish. But I am not responsible for any consequences of using this software.

Have fun!

- **File lists:**

There are seven files in the package:

	COMDLG32.OCX	
	help.pdf	Help file
	ZJ_Antenna_Matching.exe	Main program
	extended.s1p	Example antenna data file #1
	retracted.s1p	Example antenna data file #2
	C_value.txt	
	L_value.txt	

Figure A.1 File list.

The values of C_value.txt and L_value.txt files are capacitor and inductor values used in the software. These values are good enough for most applications. But both files can be edited according to capacitors and inductors available to you.

- **Current supported data format:**

1. TOUCHSTN.S1P
2. TOUCHSTN.S2P (only the first port can be seen)
3. TOUCHSTN.SNP (only the first port can be seen)
4. CITIFILE (only the basic single segment format)

- **Version History:**

Table A.1 Version history

1.4.1	1 September 2007	First public release
1.0	20 February 2003	Initial version

1. You need to load at least one antenna data file to use the software. Two sample files, extended.s1p and retracted.s1p, are included in the software package. You can use them for practice.
 - a. When the software starts, only the top “load file” button is accessible.
 - b. After you load the first file, the second “load file” button will be accessible. Loading the second data file is optional. If you have two scenarios, you can load the second one. When you are optimizing an antenna, both antenna scenarios share the same matching network, so you can check out the antenna response of both scenarios simultaneously. When you load two data files, both of them MUST have identical frequency span and data points. Otherwise, the software will reject the second file.

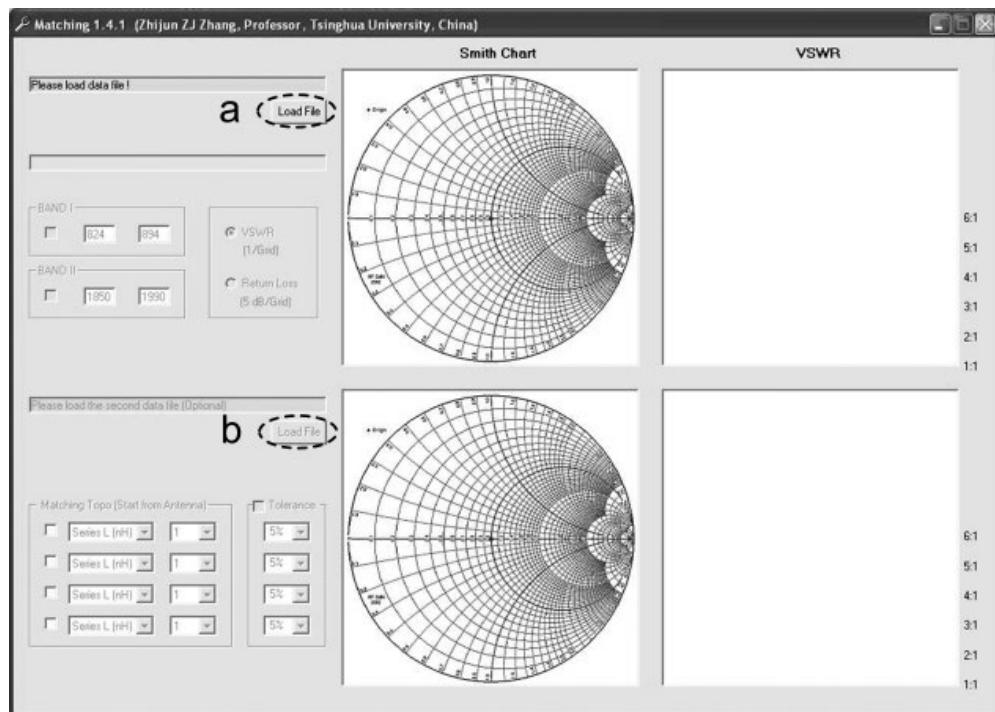


Figure A.2 Screenshot 2.

2. After loading a file, the software reports some information. If you cannot see the following information, it means the data format is not compatible.
 - a. The software indicates what data format the loaded file is. The software currently only supports TOUCHSTN and CITIFILE formats.
 - b. The software also indicates the minimum frequency and maximum frequency of the loaded file.
 - c. The plot on the Smith chart should be same as what you saw on your network analyzer; otherwise, it is most likely because you have saved the file as unformatted data which does not include the port extension you have made. Read the user manual shipped with your VNA and save the data file as formatted data.
 - d. Voltage standing wave ratio (VSWR) or return loss (RL) data should also be same as what you saw on your network analyzer.

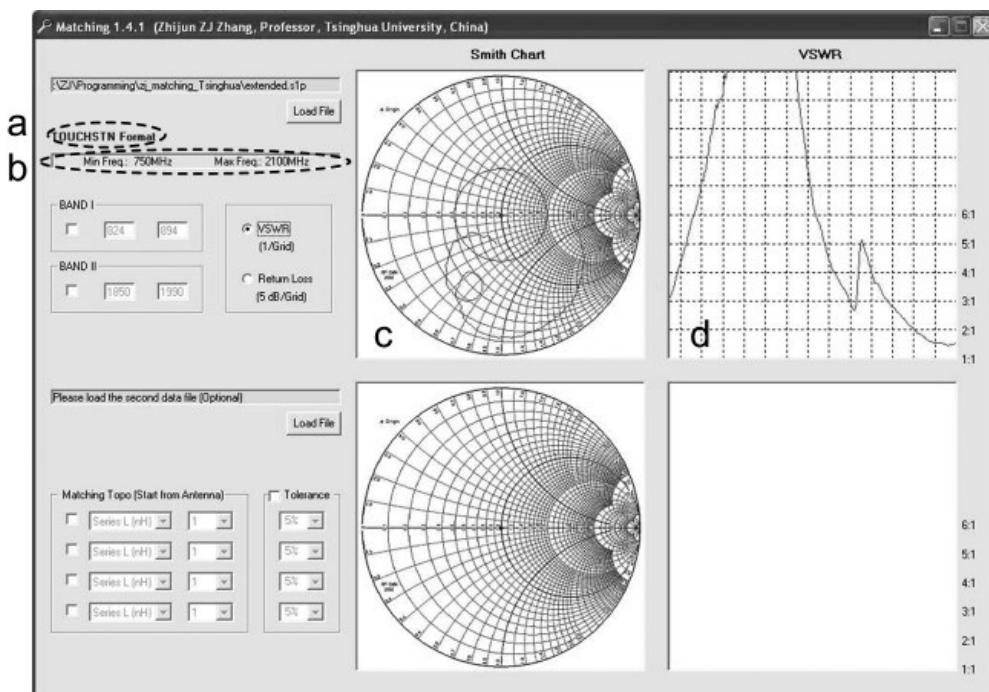


Figure A.3 Screenshot 3.

3. There are some settings you can change.

- a. You can highlight up to two interested bands. Band-I will be marked by blue color on both the Smith chart and the VSWR/RL plots. Band-II will be marked by green color.
- b. You can switch the display format between the VSWR and the RL.
- c. The red circle in the Smith chart is the VSWR 2:1 circle.
- d. In the VSWR/RL plot, the grid is 100MHz each grid along the x-axis. Along the y-axis, the unit of VSWR plot is 1/grid and that of the RL plot is 5dB/grid.

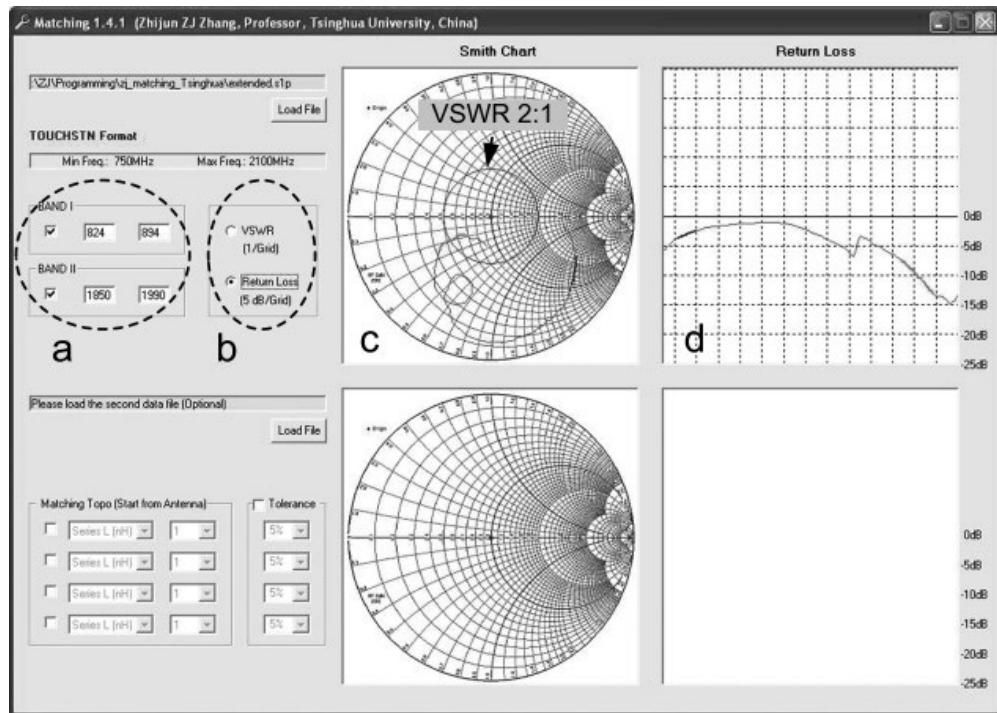


Figure A.4 Screenshot 4.

4. You can modify the matching network to simulate the response in the Smith chart, the VSWR, or the RL format.
- If looking from antenna side, the top one is the first component. You can add up to four matching components. Each component can be series inductor (Serial L), shunt inductor (Shunt L), series capacitor (Serial C), or shunt capacitor (Shunt C). The unit for capacitor is pF and for inductor is nH.
 - You can use pull-down menus to adjust the matching component value. If you cannot find the value you need, modify C_value.txt and L_value.txt files according to capacitors and inductors available to you. Restart the software with a valid modification.
 - The black line inside the graphic area is the original antenna response without the matching network. The blue line is the response of the matched antenna.

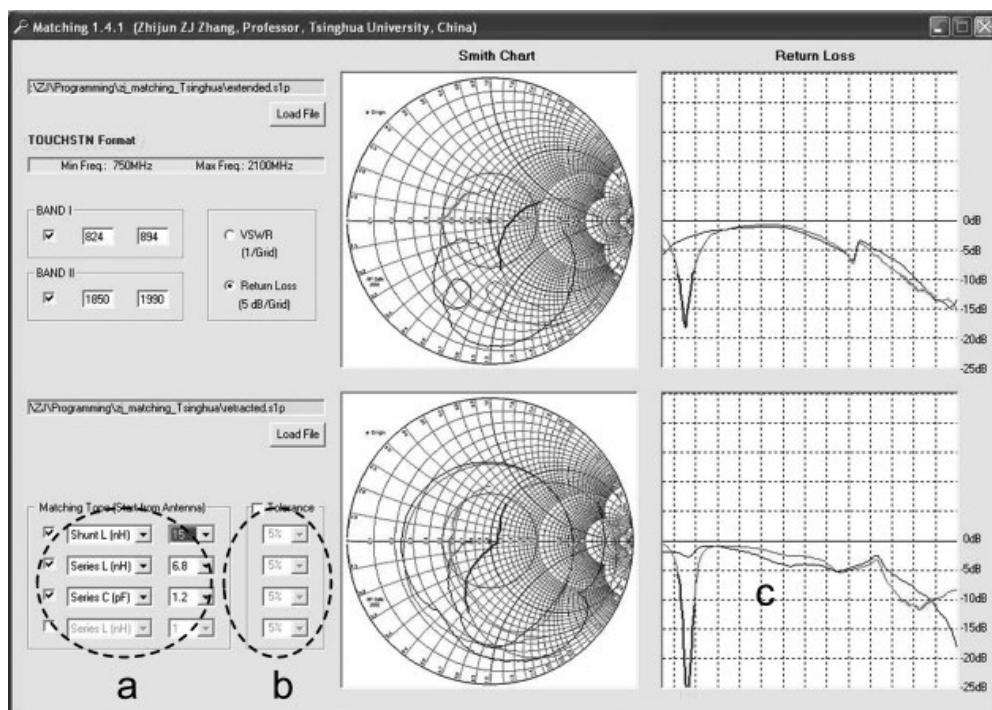


Figure A.5 Screenshot 5.

5. The software also supports tolerance analysis. Either Band I or II has to be selected before tolerance analysis can be done. The tolerance analysis shows the worst case that might happen.

- You can use pull-down menus to adjust the component tolerance from 0 to 20%.
- The tolerance result is only shown in VSWR or RL plots.

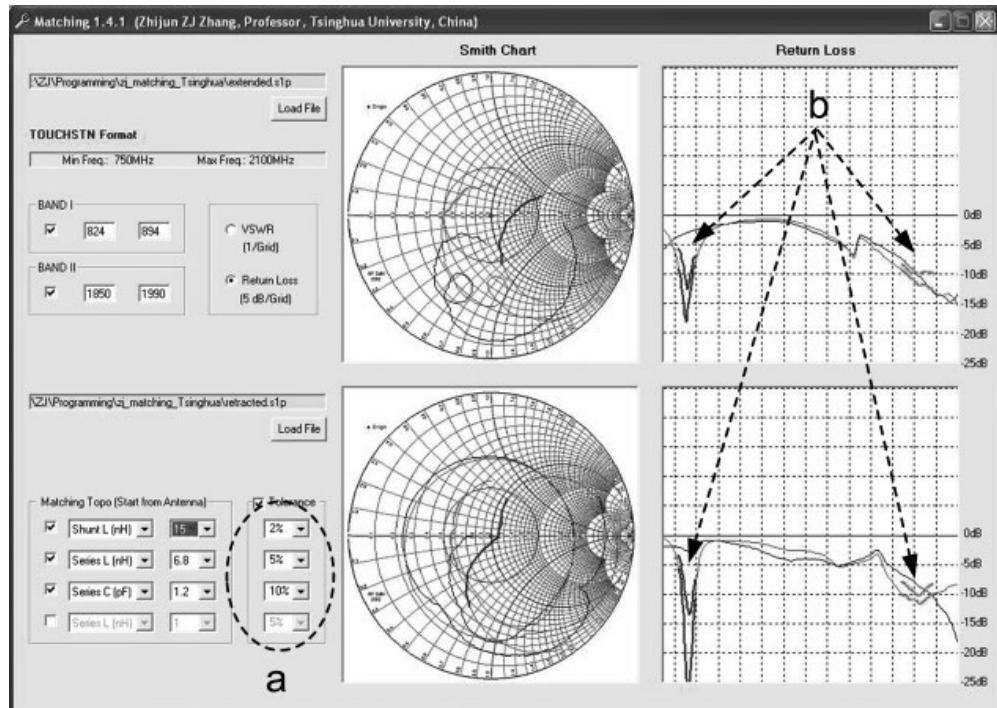


Figure A.6 Screenshot 6.

Index

- absorbing material, 246–8
active measurement
 EIRP, 253–6
 EIS, 256–9
 ERP, 253–6
 sensitivity degradation, interference, 259–62
 TIS, 256–9
 TRP, 253–6
anechoic chamber
 2D chamber, 248, 249
 3D chamber, 249
efficiency, 251–2
gain, 252
measurement 3D sphere, 249, 250
near-field chamber, 250
pyramid absorber, 247
RAM, 246, 247
Satimo chamber, 251
 wedge absorber, 248
antenna chamber, *see* anechoic chamber
antenna correlation, 207–8
antenna isolation, 207–11
antenna matching. *see* matching
antenna measurement. *see* measurement
antenna type
 dipole (*see* dipole antenna)
 internal (*see* internal antenna)
 retractable (*see* retractable antenna)
 stubby (*see* stubby antenna)
Apple, 7–9, 140, 211, 221, 284
band allocation, 16–18
bandwidth, 11–14, 42–50
candy-bar phones, 65, 66, 138, 139, 290, 300
ceramic antenna. *see also* internal antenna
 advantages, 172–3
 ceramic patch antennas, 172
 dielectric resonator antenna, 172
IFA, 176–7
loop, 177–9
monopole, 173–6
ceramic patch antennas, 172
choke, 2, 3, 21, 210, 237, 239–43, 245, 263
Chu limit, 13–14
coaxial cables, 235
coil antenna. *see* stubby antenna
correlation of fixture, 266–9
coupler, 270, 271
decoupled whip-stubby antenna, 119–21
desense, 259, 260
dielectric resonator antenna, 172
dipole antenna, 3, 4, 70–4, 82, 84, 95, 134, 167–70, 179, 180, 182, 254
direction-of-arrival (DOA) angle, 203, 204
direction-of-departure (DOD) angle, 203, 204
dual band matching, 50–4
dual-port MIMO antenna, 208
effective isotropic radiated power (EIRP), 253–6
effective isotropic sensitivity (EIS), 256–9

- effective radiated power (ERP), 253–6
efficiency, 12, 251–2
electromagnetic compatibility (EMC), 304–5
entry-level phone, 211–21. *see also* Hongmi 2A
external antenna
 ground effect, 131–6
 meander line antenna, 24, 26, 126–31
 retractable (*see* retractable antenna)
 stubby (*see* stubby antenna)
- ferrite beads, 239–43
field failure rate (FFR), 139
fixture. *see also* measurement
 calibration process, 243
 choke, 239–43
 coaxial cables, 235
 ferrite beads, 239–43
 port extension, 243–5
 SMA connectors, 236
 switch connector, 238
- flagship phone, 221–6
- flex antenna. *see* meander line antenna
- flexible print circuit (FPC), 213
- flip phones, 3, 138, 139, 290, 300
- folded monopole antenna, 163–7
- global positioning system (GPS), 16, 17, 107, 115, 145, 172, 173, 176, 177, 193, 195, 199, 213, 219, 220, 226, 238
- Google, 1, 140
- ground effect, 131–6
- hearing aid compatibility (HAC)
 acoustic passage, 302
 articulation weighting factor, 299
 E-field and H-field measurements, 300–1
 E-probe/H-probe measurements, 304
 exclusion block placement, 300
 measurement, 296–9
 near-field categories, 302
 T-coil axial measurement probe, 303
- helical antenna. *see* stubby antenna
- hepta-band antenna
 with multiple radiators and multiple modes, 185–91
 reconfigurable, 191–9
- Hongmi 2A
 back cover detached, 213
 bottom view of back cover, 214
 front, side, and back views, 211, 212
 matching circuit primary antenna, 216, 217
 matching circuit, secondary antenna, 218–19
 matching circuit, WLAN and GPS, 219, 220
 production primary antenna, 214, 215
 secondary antenna, 216, 217
 semifinished primary antenna, 215, 216
 semifinished secondary antenna, 216, 218
 semifinished WLAN and GPS antenna, 219, 220
 switch connectors, 220, 221
 WLAN and GPS antenna, 219
- Huawei, 8, 140, 211
- IFA ceramic antenna, 176–7
impedance matching. *see* matching
- internal antenna
 ceramic (*see* ceramic antenna)
 entry-level phone, 211–21
 flagship phone, 221–6
 folded monopole antenna, 163–7
 hepta-band (*see* hepta-band antenna)
 IFA, 141–6
 loop antenna, 167–72
 MIMO (*see* multiple input and multiple output (MIMO) antenna)
 on Motorola Razr V3, 7
 on Nokia 3210, 6
 PIFA (*see* planar inverted-F antenna (PIFA))
 slot, 179–85
- inverted-F antenna (IFA), 141–6
- iPhone, 7–9, 140, 213, 221, 284
- Kirchhoff's current law, 70
- Lenovo, 140
- limit line method, 261, 268, 269
- limits, 261, 265, 269, 271
- line-of-sight (LOS) path, 203–5
- loop antenna, 167–72. *see also* internal antenna
- loop ceramic antenna, 177–9
- low-temperature co-fired ceramic (LTCC) technology, 173, 174
- L-shaped antenna, 141, 142, 144, 185, 291
- lumped-element-only matching techniques, 41
- manufacturing
 flex, 129, 130
 folded monopole antenna, 166–7
 measurement (*see* production measurement)
 PIFA antenna, 159–63
 retractable, 126, 130
 stubby, 105, 118, 129
- matching
 bandwidth consideration, 42–50
 dual band, 50–4
 reconfigurable (*see* reconfigurable matching)
 single band (*see* single band matching)
 tolerance consideration, 42–3
- meander line antenna, 24, 26, 126–31
- measurement
 chamber active, 253–62
 chamber passive, 246–52

- measurement (*cont'd*)
 fixture, 234–45
 MIMO test, 271–5
 OTA, 253–62
 port extension, 243–5
 production line (*see* production measurement)
 Microsoft, 1, 140
 MIMO (*see* multiple input and multiple output (MIMO) antenna)
 monopole antenna
 bottom-installed, 170, 171
 vs. dipole, 73–4
 dual-branch multiband, 126, 128
 folded (*see* folded monopole antenna)
 ground effect, 131, 133–5
 helical, 95
 wideband cylindrical, 109
 monopole ceramic antenna, 173–6
 Motorola, 2–4, 7, 17, 65, 138–40, 163, 164, 284
 multiband helix stubby antenna
 multi-branch, 86–95
 single-branch, 95–109
 multiband PIFA antenna
 with parasitic element, 158–9
 with separate branches, 157, 158
 with slits, 149–57
 multi-branch multiband helix stubby antenna, 86–95
 multilayer ceramic antenna, 173
 multiple input and multiple output (MIMO) antenna
 active chamber test, 271–2
 antenna correlation and isolation, 207–8
 capacity boost, 200–6
 improving antenna isolation, 209–11
 passive chamber test, 271–2
 RF environment simulation system, 273
 SCME urban macrocell (UMa) channel model, 272, 273
 multiport network, 10
 near-horizon partial isotropic sensitivity (NHPIS), 258
 near-horizon partial radiated power (NHPRP), 255
 network analyzer, 18–19
 neutral line, 210–11
 Nextel, 66
 Nokia, 5, 6, 138–40, 163, 211, 284
 one-port network, 10
 over the air (OTA) measurement, 282. *see also* active measurement
 Panasonic, 173
 parasitic element, 158, 159
 pass/fail limits, 261, 265, 269, 271
 passive measurement
 anechoic chamber, 246–52
 fixture, 234–45
 VNAs, 229–34
 path loss, 65, 66
 phantom, 278, 279
 phantom head, 278, 279, 282
 phased arrays, 205, 206
 PIFA. *see* planar inverted-F antenna (PIFA)
 pigtail. *see* measurement
 PIN diode, 55, 60, 63, 192–4, 196, 198
 planar inverted-F antenna (PIFA), 285
 manufacturing, 159–63
 multiband with parasitic element, 158–9
 multiband with separate branches, 157, 158
 multiband with slits, 149–57
 single-band, 146–9
 port extension, 243–5
 probe
 E-field, 280, 297
 H-field, 297
 production measurement
 constructive coupling, 265
 correlation, 266–9
 coupling coefficient, 271
 phone's radiate test, 270
 S₁₁ fixture, 263, 266, 267
 S₂₁ fixture, 264, 265
 prototype. *see* fixture; measurement
 push-to-talk (PTT), 66
 reconfigurable matching
 switch-based, 60–3
 varactor-based, 55–9
 reflection coefficient, 10, 231
 retractable antenna, 5, 117–18
 decoupled whip-stubby antenna, 119–21
 semi-decoupled whip-stubby antenna, 121–6
 return loss, 11
 SAM head, 279
 Samsung, 140, 163, 284
 SAR. *see* specific absorption rate (SAR)
 SAR reduction, 291–294
 Satimo chamber, 249–51
 self-jamming, 260, 261
 semi-decoupled whip-stubby antenna, 121–6
 sensitivity. *see* effective isotropic sensitivity (EIS);
 total isotropic sensitivity (TIS)
 sensitivity degradation, 259–62
 Shannon limit, 200, 202
 single-band helix stubby antenna, 67–85
 single-band IFA, 22–3, 22–3
 single band matching
 bandwidth consideration, 42–3
 load impedance, 37
 with lumped elements, 33–6

- π shaped matching network, 46, 48
shunt *LC* resonator, 44
Smith chart, 36
tolerance analysis, 50
transmission line and lumped elements, 39–42
T-shaped matching circuit, 49
two-element matching circuit, 44, 45
single-band PIFA, 146–9
single-branch multiband helix stubby antenna, 95–109
singular value decomposition (SVD), 200–6
slot antenna, 179–85
SMA connectors, 236
Smith chart, 29–33
specific absorption rate (SAR)
area and zoom scanning, 282
definition, 277
dielectric properties, head/body, 277, 278
distribution, IFA, 291, 292, 295
FCC qualification test, 294
flip angle, clam shell phone, 289, 290
3G and 4G network, 294
left ear check position, 286
limits, 283–4
measurement method, 277–83
near-field energy, 288
phantom, 278, 279
reduction techniques, 291–294
slide phone, 290
stubby antenna, tilting, 287, 288
test equipment, 285
of Wi-Fi bands, 294
spectrum, 16
stubby antenna
meander line, 126–31
multiband helix, 86–109
single-band helix, 67–85
ultra-wideband, 109–17
whip-stubby antenna, 117–26
surface-mount technology (SMT), 36, 173
surface scan, 282
switch-based reconfigurable matching, 60–3
switch connector, 238
3D surface scanning, 299
tolerance analysis, 50
total isotropic sensitivity (TIS), 256–9
total radiated power, (TRP), 253–6
transverse electromagnetic (TEM) cell, 270
tuning resonant frequency, 24
ultra-wideband stubby antenna, 109–17
varactor-based reconfigurable matching, 55–9
vector network analyzers (VNAs)
configuration, 230
entry-level VNA, E5071C, 229, 230
hand usage, 233, 234
reflection coefficient, 231
VSWR, 231–3
voltage standing wave ratio (VSWR), 11
and reflection coefficient, 231
vs. return loss, 232
and Smith chart, 233
volume scan, 282
VSWR. *see* voltage standing wave ratio (VSWR)
walkie-talkie from coast to coast, 66
whip antenna, 3–5, 66, 78–84, 102, 104, 105, 117, 120, 121, 141, 144, 285, 287. *see also* retractable antenna
whip monopole antenna, 5, 74, 76–9
whip-stubby antenna. *see* retractable antenna
wideband dual antenna, 212
wireless local area network (WLAN) antennas, 16, 18, 126, 127, 173, 175, 177, 178, 193, 195, 196, 199, 213, 214, 219–24, 226, 238, 284, 304
Xiaomi 4
antennas, 221, 222
front, side, and back views, 221, 222
primary antenna, 224
secondary antenna, 225
WLAN, 226
XM satellite radio, 172