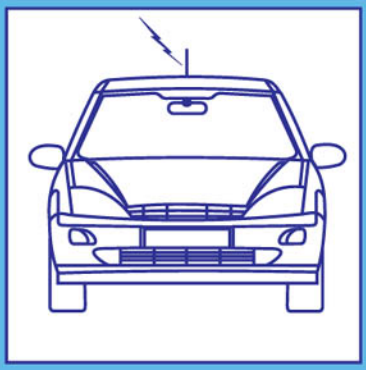
Antennas for Information Super Skyways:
An Exposition on Outdoor and Indoor Wireless Antennas



Perambur S. Neelakanta

Rajeswari Chatterjee

Antennas for Information Super Skyways:

An Exposition on Outdoor and Indoor Wireless Antennas

ELECTRONIC & ELECTRICAL ENGINEERING RESEARCH STUDIES

ANTENNAS SERIES

Series Editor: **Professor J. R. James**
The Royal Military College of Science
(Cranfield University), Shrivenham, Wiltshire, UK

10. Frequency Selective Surfaces: Analysis and Design
John C. Vardaxoglou

11. Dielectric Resonator Antennas *
Edited by **K. M. Luk** *and* **K. W. Leung**

12. Antennas for Information Super Skyways:
An Exposition on Outdoor and Indoor Wireless Antennas
P. S. Neelakanta *and* **R. Chatterjee**

*** Forthcoming**

Antennas for Information Super Skyways:

An Exposition on Outdoor and Indoor Wireless Antennas

Perambur S. Neelakanta
and
Rajeswari Chatterjee

RESEARCH STUDIES PRESS LTD.
Baldock, Hertfordshire, England

RESEARCH STUDIES PRESS LTD.

16 Coach House Cloisters, 10 Hitchin Street, Baldock, Hertfordshire, SG7 6AE, England
www.research-studies-press.co.uk

and

Institute *of* **Physics** PUBLISHING, Suite 929, The Public Ledger Building,
150 South Independence Mall West, Philadelphia, PA 19106, USA

Research Studies Press Ltd. is a partner imprint with the Institute *of* **Physics** PUBLISHING

Marketing:

Institute *of* **Physics** PUBLISHING, Dirac House, Temple Back, Bristol, BS1 6BE, England
www.bookmarkphysics.iop.org

Distribution:

NORTH AMERICA

AIDC, 50 Winter Sport Lane, PO Box 20, Williston, VT 05495-0020, USA
Tel: 1-800 632 0880 or outside USA 1-802 862 0095, Fax: 802 864 7626, E-mail: orders@aidcvt.com

UK AND THE REST OF WORLD

Marston Book Services Ltd, P.O. Box 269, Abingdon, Oxfordshire, OX14 4YN, England
Tel: + 44 (0)1235 465500 Fax: + 44 (0)1235 465555 E-mail: *direct.order@marston.co.uk*

Library of Congress Cataloguing-in-Publication Data

Neelakanta, Perambur S.
Antennas for information super skyways : an exposition on outdoor and indoor wireless antennas / Perambur S. Neelakanta and Rajeswari Chatterjee.
p. cm. -- (Electronic & electrical engineering research studies. Antennas series ; 12)
Includes bibliographical references and index.
ISBN 0-86380-267-2 (alk. paper)
1. Antennas (Electronics) I. Chatterjee, Rajeswari, 1922- II. Title.
III. Series.
TK7871.6 .N44 2002
621.384'135--dc21
2002073937

British Library Cataloguing in Publication Data

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

ISBN 0 86380 267 2

Printed in Great Britain by SRP Ltd., Exeter

Cover artwork by A3 grafix ltd.

Editorial Foreword

It is not uncommon for scientific and engineering topics to reflect the trends and upsurge of practices in society and this has been vividly witnessed with antenna research and technology during the past three decades. In the late 1960s I recall the words of a senior colleague "...that antennas had been exhaustively treated and there was little scope for further innovation...". Soon after we saw the profound developments in printed antenna technology, instigated by the onset of integrated electronics and the need for more compatible compact antennas that are easier to manufacture. Printed antenna research and development continues to this day unabated but is now overlaid with exciting developments in mobile base station and handset antennas, arising from the global expansion of new communications during the past decade. Even more recently, the old-fashioned word "Wireless" has been adopted to describe the latest communication trend of seamless connectivity using direct radiation between terminals, as opposed to guided waves in wires. Bluetooth, Wireless Local Area Networks (WLAN) systems, etc. are now almost household names, such is the penetration of these new communication concepts into society at large. For the antenna community, wireless presents both new design problems and outstanding opportunities. Every wireless link must have transmit and receive antennas designed precisely to satisfy both the link system requirements and the physical demands of a particular terminal; be it a personal computer, a refrigerator, an ear piece, a petrol pump and so on. Wireless antennas thus require bespoke individual design and this is an entirely new situation for equipment planners and suppliers, sales staff, antenna designers and students alike: antennas cannot be purchased off-the-shelf and each design needs to start afresh. The outcome is that a wider, and often disparate spectrum of people, now need to have a better understanding of these antennas at different levels and the present library stock of general elementary antenna texts on one hand and advanced research books on the other, needs supplementing with a new type of book: That is, a text specifically giving a comprehensive insight to a wide community of readers who are engaged in some way with the new wireless systems. This is the objective of their new book and Professors Neelakanta and Chatterjee are to be congratulated on creating such an expansive tutorial text without sacrificing technical depth and practical engineering content. Professor Neelakanta is well known for his books and research. Professor Chatterjee has made distinguished contributions to the profession and has already published in this Series with her notable 1985 book "Dielectric and Dielectric-loaded Antennas". Their present new book is an adventurous and enthusiastic response to the present needs of the antenna community in general and it is indeed a unique and valuable perspective. It is a sincere pleasure for me to welcome the authors of this new venture to RSP's "Antenna Series".

Professor Jim R James
December 2002

This page intentionally left blank

Preface

As the world marches along the information superhighway ahead into the first decade of the new century, the progress and pace of evolving telecommunications are seen incomprehensibly fast both in the wireline and wireless sectors. Specific to the panoramic realm of wireless telecommunications and networking (as conceived now and projected as future interests or "next-generations") exists a gamut of standards. These include multiple applications of electrical communication technology spanning across voice, video, and data transmissions in an integrated fashion. The associated radiation and propagation of electromagnetic (EM) energy pose unique considerations as a result of user mobility and dynamically changing ambient. As a result, the electromagnetics of wireless transmissions encounter complex interference situations. In order to combat such signal-impairing constraints and maintain robust wireless links, specific designs at the interface of the propagating medium (namely, the free-space) and the RF electronics of the wireless units are warranted. The vital part of this interface refers to the antenna structures.

The first step towards learning and comprehending antennas — whether in educating technicians or creating a new breed of wireless communication engineers — is to become proficient and conversant of the past profile of antenna systems as well as gain a comprehensive knowledge and hands-on perspective about the emerging EM radiating structures. Therefore, the broad scope of this book is to offer a comprehensive insight into such antennas. It includes the feasibility and implementation issues as well as design considerations used in modern wireless/mobile communication systems.

About this book …

The primary goal of this book, as stated above, is to cast the salient aspects of wireless communication antenna technology in the real-world perspectives. Generally speaking, there are three types of books on the subject matter of antennas: The first category refers to classical books on EM radiating systems written over the last several decades (but their contents revised and updated to include the changing trends and evolving new concepts). Such books address the global prospects of antenna systems and comprehensively project various antenna types against their suitable applications to radio/TV, radar, navigational aids and

traditional point-to-point communication systems. Detailed analytical considerations and/or numerical computations form the core theme of these books. However, they, to a large extent, are not written *per se*, with a mission to include exclusively (and exhaustively) modern wireless/mobile communication systems.

The next class of books on antennas contains details focused on wireless/mobile communication systems with the contents formatted in the style of a handbook. They are again written to include exhaustively the plethora of wireless/mobile communication antennas. Multiple authors have contributed chapterwise contents, each chapter devoted to describe a specific topic. These books, however, hardly include any theoretical formulations and/or analytical framework on the various structures of antennas elaborated upon, but they stand as excellent reference manuals on the subjects of interest.

The third version of books focused exclusively on applications of antennas for radar, satellite systems, wireless/mobile communications etc. They are written for an audience who are totally unacquainted with analytical considerations and have no flair for electromagnetic theory. Hence, intentionally avoided in such books are the mathematics and plug-in equations pertinent to elecromagnetics, radiation principles and EM propagation concepts.

The present book is written in a unique perspective — to present as its contents a mix of topics — covering both the analytical aspects of antennas (with their relevance to wireless/mobile communications) as well as descriptions on underlying principles and design considerations. In essence, this book includes chapters that supplement the descriptive portrayals of wireless/mobile communication systems with necessary analytical considerations along with necessary details on the associated antenna designs.

In writing this book, the authors have duly considered the wide audience profile that will reach out for this book — students and researchers, who pursue studies on antennas in the orthodox realm of EM field equations; and the technical staff of the wireless-communication industrial sector who would like to gain a working knowledge on the state-of-the-art antenna concepts (*sans* field equations!) so as to use them in their developmental efforts.

Hence, those sections and/or chapters that are significantly oriented in projecting analytical perspectives are duly identified (and indicated as footnotes) for those who may wish to skip them. Care was, however, exercised in formatting the chapters such that, such selective reading will not dislodge the sequence and/or understanding the contents in the rest of the book. The layout of the book is as follows: It is organised in eight chapters. The contents of each chapter are preceded by a chapter-opener portraying a preview on the real-world aspects and application considerations pertinent to the chapter. Further, some example-problems are presented on *ad hoc* basis within various chapters (with hints and solutions, as necessary). Lastly, the relevant bibliography is appended at the end of each chapter. Also included at the end of the book are acronyms and abbreviations commonly used in wireless communication parlance.

A brief outline of each chapter is indicated below:

Chapter 1
This is an introduction to wireless communications, and it describes the historical perspectives and evolution of modern wireless/mobile communication systems. It highlights the state-of-the-art systems, their general specifications and their functional attributes plus application profiles. This chapter is recommended to all readers to get a perception on wireless communication systems. It has hardly any exposé to mathematical principles.

Chapter 2
Here, a summary on the basic concepts of electromagnetic (EM) fields and waves is presented and the analytical aspects of electromagnetic wave theory are elaborated. This chapter requires background as well as in depth knowledge of electromagnetism. Readers who are either already familiar with such topics (or those who do not relish such mathematics!) may omit reading this chapter.

Chapter 3
This is an extension of Chapter 2 in presenting the underlying concepts of EM field theory applied to EM radiation, antennas and wave propagation. Again, considering its heavy mathematical outlay, those who desire so could skip this chapter. It is however stressed that, both in Chapters 2 and 3, the analytical presentations are tailored to appreciate the general scope and framework of this book.

Chapter 4
This chapter is written in two parts: The first part is again a compendium of analytical results on the basic antenna elements and, the second part is devoted to describe the selective elements that are widely used in wireless communication systems. It is recommended that the readers who do not need the first part may avoid it and proceed to the second part without any loss of generality of their reading sequel.

Chapter 5
Presented in this chapter are details on antenna arrays, constituted by the antenna elements described in Chapter 4. The specific uses of such arrays in wireless/mobile communications are identified. This chapter is a prologue to the smart antennas described in Chapter 6.

Chapter 6
Addressed here are exclusive considerations on intelligent antenna systems that combine the basic arrays (described in Chapter 5) and signal-processing techniques. Designated as “smart antennas”, these structures are viable radiators of the state-of-the-art wireless communication systems as presented in this chapter.

Chapter 7
The indoor RF communication links are part of modern wireless communications. Typically, the WLAN and the Bluetooth™ are examples of such systems. The antenna requirements of such indoor wireless communication systems are unique and warrant a distinct study. As such, Chapter 7 is devoted to present all the relevant topics.

Chapter 8
The trend in modern wireless communications is to support the so-called broadband transmissions of next-generation (3G) systems. Hence, focused in Chapter 8, is a study that summarises the antenna aspects of broadband wireless communication systems. The broadband considerations pertinent to wireless access networks as well as indoor applications are identified and relevant antenna requirements are described.

General layout of this book versus audience profile ...
This book is written to suit classroom presentations (for adoption as a textbook) as well as to guide design and development engineers. That is, it outlines the necessary underlying principles and implementation considerations of wireless communication antennas in a lucid manner for students, designers, as well as technical staff involved in development activities. This book is organised to help students as a companion text and guide design engineers and developmental staff of the industry on the changing trends in antenna concepts. In a nut-shell, the goal of the book is to make its audience appreciate the query "Why so many types of antennas at all?"

For student audience …
With reference to the class room environment, a modern approach to teaching engineering subjects (as encouraged by various accreditation bodies) refers to blending design considerations along with the theoretical contents. Hence design examples of practical interest and implementation are presented explicitly in this book. That is, the pedagogy of this book is conceived to meet the relevant objectives consistent with the student audience profile and its requirements. As stated before, across various chapters, presented are example-problems as well as some typical problems that are left as exercises to the readers. Should this book be adopted as a textbook, such exercises will be very useful. Where needed, some hints are indicated underneath the problem-statements.

The authors' vast experience as instructors of antennas-related courses provides insight into the students' needs on this subject leading to making this book student-friendly. This book will be ideal for a postgraduate and/or undergraduate elective course such as "Antennas and Propagation Aspects of

Modern RF/Wireless Communication Systems", which will attract a large enrollment in many universities.

For antenna researchers, designers and development staff …

In the industrial perspective, as indicated above, this book is intended to serve as a companion reference to researchers on "the cutting edge" aspects of wireless antenna systems and as a guide to the technical staff involved in the design, fabrication and testing of wireless communication antennas.

Wireless communication antennas are an exclusive subset of the traditional electromagnetic radiators. These antennas should be viewed for their applications tailored to meet specific performance aspects of wireless/mobile communication links that are beset by low-power transmissions through hostile, inference-prone settings. Their physical size becomes of utmost importance in view of the compactness and low visibility considerations attached to their applications in portable units. Further, the physical orientation of these antennas in portable units should be duly recognised in the design as well as in evaluating the overall performance. In addition, the polarisation considerations are also issues of concern.

As such, even for those who are familiar with antenna theory, the design perspectives of wireless/mobile communication antennas are challenging. Further, the staff involved in fabricating and testing these antennas should appreciate the exclusive attributes of their structures as a part of the RF-air interface. Taking these facts into consideration, in this book we present the descriptive and *ad hoc* requirements of the antennas under discussion, that can be appreciated by the technical staff of the industry involved in the design, development and testing efforts.

In short, this book will aptly educate students (at upper undergraduate and/or postgraduate levels), as well as those professionals belonging to the wireless communication industry, equipment/system manufacturers and wireless service providers.

Effort will be made to include any necessary updated information when the book will be revised in future editions. Comments, corrections and opinions are most welcome from the readers as a feedback and may be communicated to the publishers.

Perambur S. Neelakanta

Boca Raton, Florida, USA
2003

Rajeswari Chatterjee

Bangalore, India
2003

This page intentionally left blank

Acknowledgements

The authors' foremost thanks are due to Professor J. R. James, the Antenna Series Editor for inviting them to write this book. His technical opinion and constructive criticisms during the development of the manuscript are gratefully conceded. His editorial foreword on this book is also deeply appreciated. The authors also extend their thanks to the publisher and the staff of Research Studies Press Limited for providing them with an opportunity to write and publish this book. Their support and help are sincerely acknowledged.

Further acknowledged with thanks are the efforts of Dr. Jesada Sivaraks who immensely helped in preparing the manuscript of this book. Without his help, a timely release of this work would have been impossible. His untiring efforts in searching and locating the reference materials and his ardent exercise towards word-processing (especially of those equations!) are deeply appreciated. Further, his study on Bluetooth™ systems (presented in his doctoral dissertation of Department of Electrical Engineering at Florida Atlantic University) has been profusely adopted (in Chapter 7) and the authors duly acknowledge this contribution.

Pleasant support of friends and family members of the authors are facts of cognisance and worth of stating "thanks". Also, the authors like to thank the readers for their interest in this book.

One of the authors (P. S. Neelakanta) places on record his thanks to Florida Atlantic University for extending him Sabbatical leave (2001-2002), during which time this book was mostly completed.

Lastly (but heartily), the authors dedicate this book,

To all their students!

This page intentionally left blank

Table of contents

CHAPTER 6 SMART ANTENNAS FOR WIRELESS NETWORKS

CHAPTER 7 ANTENNAS FOR INDOOR WIRELESS COMMUNICATIONS

CHAPTER 1

An Introduction to Wireless Communication

1.1 INTRODUCTION

A unique age has dawned! It is an exotic era, which has set a pace in search of excellence against odds that diminish the physical separation of communications from infinite reality to virtual zero. This is the age of information in which the posentropy is assimilated and every niche in technology is astoundingly sought to bring the art of "communicating at a distance" into a nutshell of telecommunication with no separation in space or time being virtually observed.

The physical media of this information passage pose a two-prong pursuit – *wireline* and *wireless*. These pursuits are not new. They are known for a century. However, over the decades, the framework of *via media* between communicating entities have been wisely conceived, generously modified, smartly reengineered, profusely adopted and cyclically refined so as to match the needs of evolving trends in the emerging aspects of telecommunications.

Classically, wireline telecommunications (which was ushered into the society through telegraphy in the1800s) adopted the copper wire, well-known then and still known as a "good conductor" of electricity for transporting message in the form of dots and dashes of electrical pulses. This basic copper conductor in most part remains, even after a century and a half, as the vital physical link of telecommunications networks. Changes, of course, have been made on the geometry and structural aspects of copper wirelines; and, a variety of copper-wire pairs twisted, untwisted and concentrically placed has been introduced.

The information transfer in the realm of electrical communications has also presumed a new physical medium in the last few decades that allows the flow of light (in lieu of the conventional electricity) carrying the message being communicated. The optical fibre is a contemporary and competing physical transport medium for information transfer that runs parallel to the legacy of copper wire transmissions.

The potentials of copper lines and optical fibres as wireline media of transport are dominantly present and conceivably foreseen in explicit use over the years to come along the information highway of modern telecommunications.

Yet another "physical" medium (though not "physically" perceived!) has emerged as a major infrastructure for the telecommunications applications – it is the so-called "wireless" communication medium in which the link between the communicating entities is established through "radio" means. Again, the concept of wireless or radio transmission of information is not new. Its history dates back to the

1800s with associated scientific postulations posted in the 1700s. It is an engineered art based on the science of electricity and magnetism. It was realised as a technology by Marconi through the elegant equations of Maxwell, the grand proposals of Hertz and practical demonstrations by J. C. Bose.

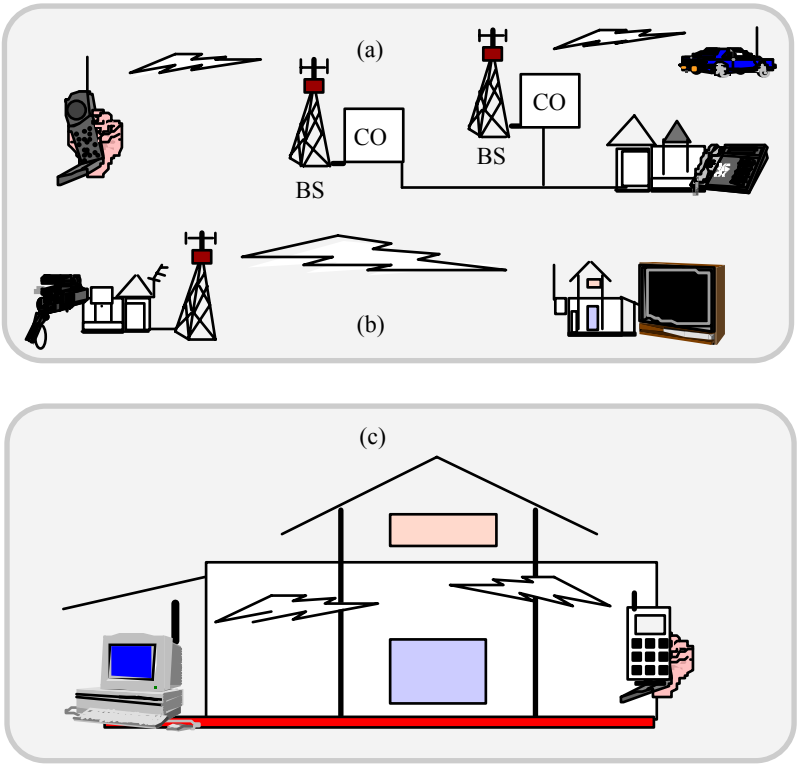


Fig. 1.1 Manifestations of wireless communication systems: (a) Outdoor system of cellular communication across voice sources; (b) broadband wireless video delivery at home *via* radio-port; and (c) indoor wireless data transmission (BS: Base station; CO: Central office/telephone exchange)

First conceived in its classical form as *radio telegraphy*, later as *radio telephony* and subsequently proliferated as broadcasting systems in the form of the radio, the television etc., the technique of radio communication has been adopted in the present times widely as a viable telecommunications alternative to the wireline technology.

The long standing concept of exercising communications anywhere, any time and of any type has now become a reality through wireless communications. Though essentially developed as wireless telephony, the underlying prospects of cellular systems catering for telecommunications involving mobile units were broadened in

their scope to include data and/or video transmissions. In other words, the so-called broadband wireless system is an agenda item of modern telecommunications branded as 3G/4G systems [1.1, 1.2].

Whether be in its primitive form of radiotelegraphy or in its present structure as a broadband wireless technique, the art of wireless communication systems, in essence, consists of the following considerations and constituents:

- Information source at the sending end
- Radio frequency transmitters
- Transmitting antenna
- Electromagnetic (EM) wave propagation
- Receiving antenna
- Radio frequency receivers
- Information sink at the receiver end.

The aforesaid forms of wireless communication systems are illustrated in Figure 1.1 and follow the generic architecture of electrical communication systems. Each system is comprised of an information source and sink pair. The transreceive wireless path across the antennas is established by electromagnetic wave propagation either indoor or outdoor.

The intervening wireless medium across which the information is transported by means of electromagnetic waves is known as the *channel*. (In the wireline systems, the physical media such as copper wires or optical fibres constitute this channel, as mentioned before.)

The general architecture of a communication system (wireline or wireless) is shown in Figure 1.2. The noise sources indicated are inevitable and pose impairments to the communication traffic, as will be discussed in detail later.

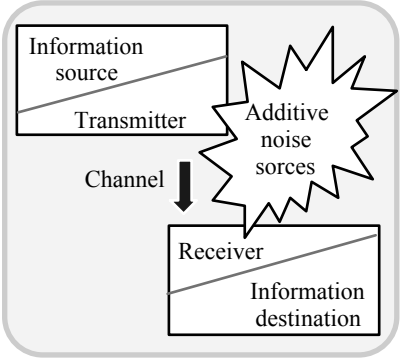


Fig. 1.2 Architecture of a wireline/wireless telecommunication system: Additive noise may prevail at the information source/sink and at the transreceive units, including the channel

1.2 A TECHNOLOGY OF SERVICES "UNTIED BY WIRES"

1.2.1 A historical perspective and state-of-the-art wireless systems

Wireless telecommunication is a "top-notch" technology of modern times. *Cellular telephones, paging, mobile radios*, and *personal communication systems* (PCS) are, for example, constituents of the state-of-the-art wireless system, which are growing into "easy-to-use" communication networks and have been proliferating extensively across the user community.

Wireless telecommunication, in a free sense, refers to a global, ubiquitous wireless network that permits its users to communicate with anyone, anywhere and at any time. Wireless *access points* (APs) can connect wandering users to wireline networks as well as to other wireless users. The access to wired infrastructure, in general, is provided for wireless/mobile network users *via* centralised access points.

The evolution of wireless telecommunication can be traced on the basis of its stratified generations. The classical era (termed as the *pioneer phase*) from 1921 to 1927 set the gears in motion to facilitate land mobile communication. The first experimental study refers to using mobile radios in police cars in Detroit (in the 1920s). That system used 2 MHz RF band. In 1934 several municipal police radios in the United States were placed in use serving more than 5,000 police cars. The FCC assigned 29 channels in the electromagnetic spectrum exclusively for police mobile radios. Until the early 1930s these mobile radios operated on the amplitude modulation (AM) principle. Later, frequency modulation (FM) mobile radios were found to be more resistant to electromagnetic propagation problems, and by the 1940s all police mobile radio systems in the United States became FM-based.

The scope of mobile radio applications was dramatically enhanced thanks to the implementation of such systems on a large-scale basis across the world during World War II for military purposes. Strides in performance, achieving reliability, and realising cost-effectiveness that were attempted in those war-time developments led ultimately to a very successful mobile communication system market in the post-war period.

Subsequent to World War II, the technological pursuits of mobile communication (during 1946 through 1968) refer to the first *commercial phase*. During this period, the demand for a mobile wireless system shifted from being solely restricted to the police and the military. Many civilian applications came into existence; as a result, there was inevitable congestion posed in utilising the available EM spectrum. Therefore, efforts were concentrated in multiplexing the channels and adopting a network-based centralised routing of messages, akin to the wireline systems, which were then in vogue. By 1949, mobile radio became a new class of telecommunication service, and in the two decades that followed, the mobile telecommunication user population in the United States alone exploded at least by an order of magnitude.

As a result, the mobile telephone service (which became a part of the *public switched telephone networks* (or PSTN) even in the 1940s) became commercial enterprise with AT&T as the service-provider in specified locales (in the United

States) under a license from the FCC. These services were operated in the very-high frequency (VHF) band (around 150 MHz). However, these were eventually shifted to the ultra-high frequency (UHF) band (around 890 MHz) using the FM technology in the middle of the1950s accommodating a FM bandwidth of 30 kHz for each voice transmission. Again, in order to serve multiple users, *multiplexing* (trunking) of a group of radio systems was the strategy that was used.

1.2.2 Cellular wireless technology

The original systems of wireless telecommunication were based on the radio-broadcasting model (with a high-power transmitter placed at an elevated location so as to serve the mobile units over a large area with an extended light-of-sight horizon). Eventually, the concept of having several stations (each of smaller RF power transmission capability and designated to serve only a small area called a *cell*) became popular. In this "cellular" system, the same frequencies used for a set of channels are "reused" in other cells as well. This *frequency reuse* strategy is done with minimal channel interference across the cells. For example, in Figure 1.3, the cells A and B, which are geographically well separated, may use the same frequency band without suffering mutual interference.

Further, the cells can be split judiciously into smaller cells in the event of increases in the user population per cell. When a mobile user goes from one cell region into another, the service responsibility is shifted from the first cell to the next one by means of a central *base station* control. This is called *hand-off*.

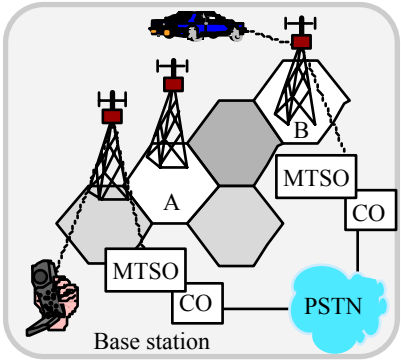


Fig. 1.3 Cellular telephone system and the associated networking (MTSO: Mobile telephone switching office: CO: Central office/exchange)

Thus what emerged was an organised cellular (analog as well as digital) telephone technology. In the early 1980s, Bell started operating a *High Capacity Mobile Telephone System* (HCMTS) in the FCC allotted band of 40 MHz in the 850 MHz spectral region. This became the forerunner to the so-called *Advanced Mobile Phone Service* (AMPS) of the 1980s through the 1990s constituting the first generation of commercial analog cellular telephone systems. Parallel developments in Japan and Europe emerged in the span of 1978 through 1986.

The UHF bands adopted are 870-960 MHz and 453-468 MHz and the number of channels serviced range from a few hundreds to a couple of thousands.

The European Total Access Communication System (ETAS) was developed in the mid-1980s, and is almost similar to AMPS, except that it is scaled to accommodate 25 kHz channels (unlike the 30 kHz channels of AMPS).

In the wireless cellular telephone system, the cellular users communicate with other subscribers on the PSTN through the access points (APs), which represent the *gateways* for such remote connectivity. A *central office* (CO) that supports the switching units is a *mobile telephone switching office* (MTSO) and it communicates with the APs through wireline PSTN (and/or *via* dedicated wireless channels) to perform necessary *signalling*; that is, it controls the call set-up, call-processing and call-release phases. Cell sizes of AMPS can range from less than a square kilometre in urban areas, up to 100 km^2 in the countryside. When a mobile telephone moves between neighbouring cells, communication responsibilities dictated by the associated protocols are automatically transferred between the corresponding APs.

For digital data transmission on analog wireless systems, the subscriber could use a conventional modem attached to their cell phones (*via* an adapter) and wireless service providers normally enable error correction feasibility in the packets supporting such data transmissions in order to achieve a desirable *bit error rate* (BER) performance. Relevant services also include pooling of modems that detects the protocol of incoming calls so as to set up a fast and robust connection to the wireline modem. However, data sending over current cellular analog networks is still highly error-prone and constitutes a delayed process: The user has to first establish the circuit connection, which often takes about 20 to 30 seconds. Once the connection is established, the propagation conditions and the transreceive separation would require hand-offs between cell sites, which are not handled by many modems. The data rates supported on these systems are in the order of 4.8 kbps.

Data transmission on conventional analog cellular voice telephony (such as AMPS) is also done by *cellular digital packet data* (CDPD) system [1.3]. It uses the idle time in the voice network to carry data at a rate up to 19.2 kbps. The relevant strategy is as follows: Wireless voice transmission statistics show that about 30% or more of channel's airtime remains idle (even during heavy traffic). Hence, an ample idle time-slot is available to establish an air-link to support short, bursty data transmissions. That is, CDPD awaits for open voice channels that appear between voice calls; and, an AP will pick up (or "seize") one of these unused (idle) time-slots and will transmit the data call awaiting in its locale over this time-slot. If a subsequent voice-call needs that time-slot channel, it gets the priority and the base station will give it up. The prevailing data transmission will then be hopped into another idle time-slot, when it becomes available. (This is known as *channel hopping*). An analog voice signal takes about 40 ms to set up before voice information is sent out, giving sufficient time for the CDPD data-link to get disconnected and hop into another channel. CDPD is targeted towards economical applications of transaction services such as credit card verification. It also uses authentication (through encryption) for security purposes. Further,

CDPD is built on the top of the existing cellular infrastructure (AMPS). It uses the *Internet Protocol* (IP) and packet-by-packet routing rather than circuit-routing.

The *first generation* AMPS indicated above was confined to a narrow-band standard and what followed next (in the mid-1990s) is the emergence of the so-called *second generation* (2G) cellular phones, which conform to three major standards, namely:

- *Group Special Mobile* (GSM): (also known as *global mobile system*). This is an European and international standard and the mobile unit to the base-station link operates at 890-915 MHz band and at 935-960 MHz in the forward link
- *IS-54*: North American Digital Cellular (NADC) standard operating at 824-849 MHz (mobile-to-base) and 869-894 MHz (base-to-mobile)
- *Japanese Digital Cellular* (JDC) standard operating at 810-915 MHz (mobile-to-base) and 940-960 MHz (base-to-mobile).

Unlike the first generation analog standards, the 2G-systems correspond to *digital cellular* deployment [1.4, 1.5]. This new adoption came into being to meet the popularity of cellular telephones in the 1990s. Going for digital cellular systems has definite advantages. For example:

- State-of-the-art advances in the digital modulation techniques facilitate high-performance (in terms of spectrum utilisation) of cellular telephones
- More voice channels on a single carrier can be accommodated thanks to developments in lower bit-rate digital voice encoders
- The digital technique allows reduction in overheads required for *signalling* (call set-up etc.)
- To meet the challenges of harsh, EM wave propagation environments faced by mobile systems, robust schemes have been developed for digital source and channel encoding strategies
- Digital techniques have also been developed to reduce the co- and adjacent-channel interference encountered in cellular telephone systems
- Digital schemes can be devised to accommodate flexible bandwidths
- Access and hand-off techniques can also be handled efficiently through digital methods.

Digital cellular deployment began in 1995 (in the United States) and two competing standards were prescribed toward providing high-capacity, namely, the *time-division multiple access* (TDMA) and the *code-division multiple access* (CDMA). The Telecommunication Industries Association (TIA) adopted both standards in 1993. The TDMA allows multiple user information to be placed in distinct time-slots supported by a carrier frequency. At any given receiver, the desired information

residing in the appropriate time-slots (ear-marked to that receiver) are identified and selected. The other, impertinent time-slot information are rejected.

The CDMA refers to a digital wideband, *spread-spectrum* (SS) technology that transmits multiple information in distinct codes ("chip-codes"). Hence, each receiver can selectively identify the information sent on the designated set of chip-codes. In the United States, both TDMA and CDMA wireless services in use are operated in FCC-auctioned geographical zones. The *US Digital Cellular* (USDC) is the IS-54 system that operates on TDMA scheme in the 824-849 MHz (reverse) and 869-894 MHz (forward) bands with each channel being 30 kHz in bandwidth. Subsequently, IS-95 was introduced with higher data rate for better speech quality using a 14.4 kbps speech encoder.

A *digital frequency-division multiple access* (FDMA) wireless system refers to CT-2 cordless telephone supported on 864.15-868.05 band (in the United States) and 864.10-868.10 (in Europe and Hong Kong). It assigns FDMA service to forty *time-division duplexed* (TDD) channels, each with 100 kHz bandwidth.

The *Digital European Cordless Telephone* (DECT) is a universal cordless telephone standard developed by the European Telecommunications Standards Institute (ETSI). It provides a cordless communication framework for high traffic density, short-range telecommunications, and covers a broad range of applications and environments. It enables a good quality of data and voice transmissions. Its main function is specified for portable users of the *private branch exchange* (PBX) in a building. It is based on the *open system interface* (OSI) standard and therefore, is compatible for interface with *integrated system of digital networks* (ISDN) as well as for GSM. However, it takes a long time for call set-up and/or call tear-down to complete due to its connection-oriented protocol.

In Japan, the digital cordless telephone service is known as *Personal Handyphone System* (PHS). Its telephone set is more compact and small-sized than other cordless telephones.

The 2G technology used in Japan for digital cellular telephony is known as *Personal Digital Cellular* (PDC). It uses a variation of TDMA. It operates in the 800 MHz and 1500 MHz with data rates of 9.6 kbps (full) and 5.6 kbps (half). Another 2G system is called TDMA IS-136 of TIA and is an evolution of IS-54.

The *third* and *fourth generation*s (3G/4G) of cellular telephone system refer to a contemplated service that stretches into the new century in their implementations and operation. Relevant technology has been conceived to include the state-of-the-art advances in FDMA, TDMA, CDMA, and *collision sense multiple access* (CSMA). Further, spread-spectrum considerations would play a significant role in the technology envisaged.

The European digital cellular system operates as the GSM indicated earlier. Introduced in 1992, the GSM is currently widespread in Europe, and is finding its way into Asia and Australia. Prior to the introduction of GSM, the European platform originally had incompatible national cellular standards and GSM was ushered in as an unified standard across the nations of Europe. Its services support data traffic at 9.6 kbps and *short message service* (SMS). It is also usable for computer communications.

Another service developed to provide packet data service to GSM and TDMA users is the *General Packet Radio System* (GPRS). It reserves radio resources only when there is data to send and it reduces reliance on traditional circuit-switched network elements. The enhanced functionality of GPRS will decrease the incremental cost of providing data services. It is an important step towards migration to 3G networks and allows service providers to implement an IP-based core architecture for data applications. GPRS can be regarded as an overlay network onto a 2G-GSM network facilitating packet data transport at rates from 9.6 to 171 kbps. It uses, as far as possible, the resources, interfaces, and protocols of existing GSM networks, but needs certain infrastructural modifications.

An upgrade of GPRS is known as the *Enhanced Data-rates for GSM Evolution* (EDGE) system. It has an enhanced capability to support 384 kbps. The building blocks of EDGE for radio access and packet-switching/IP are being streamlined toward multimedia services on mobile communication links.

IMT-2000 is another 3G mobile communication service evolved to support data transmissions with improved transmission speeds and applications of wideband technology. It is intended to work in the 2-GHz band and supports 1.25 MHz PCS and 5 MHz multimedia information.

The global GSM is being standardised for IMT-2000 use with *wideband CDMA* (WCDMA) implementation. From a radio-access point of view, adding 3G capabilities to existing systems means wideband operation. Therefore, the service providers have to work with a sufficiently wide spectrum by reframing the existing spectrum and/or acquiring new bands.

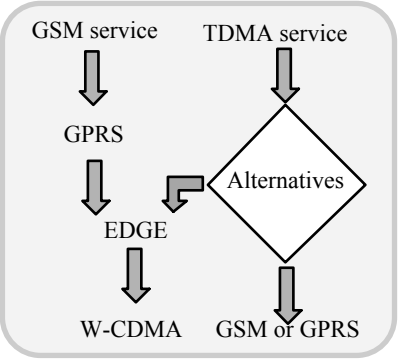


Fig. 1.4 Migration options across the present and next generations of wireless systems

The *Universal Mobile Telecommunication System* (UMTS) represents a radio-access network based on 5 MHz WCDMA. It is a member of the ITU's IMT-2000 global family. It is optimised for efficient support of 3G services. It can be used in both the existing and new spectra. It is expected to facilitate multimedia wireless communication services worldwide by 2010 to at least two billion subscribers.

Initially, it will deliver low-cost, high-capacity mobile communications offering data rates up to 2 Mbps with global roaming and other advanced capabilities.

Thus, wireless industries started with basic FDMA, TDMA, CDMA and GSM systems and gracefully migrate into 3G/4G technology. Some are rolling out of GPRS and enter into WCDMA, EDGE, UMTS etc. with their bearing set to find possible niches in the wireless market. The migration options are illustrated in Figure 1.4.

Whether they be simple voice telephony or high-speed Internet or multimedia transmissions, the wireless product lines are being directed to combine relevant protocols and content compression/optimisations to deliver high-speed access over the 2G through 4G networks. The technology evolved thereof involves conceiving efficient and compatible EM radiation and antenna considerations for robust wireless links. The task of identifying suitable radiating structures for these growing kaleidoscopic services forms in part the scope of this book.

1.3 WIRELESS NETWORKS

Wireless services need compatible networks. These networks bring the end-users within the realm of established wireless links so as to offer the service they need [1.2]. Wireless networking refers to interconnecting the end-users *via* radio means. Typically wireless communication involves modulating a high-frequency carrier by the baseband signal (or by a group of subcarriers, each modulated by a baseband voice or data signal) and radiating the modulated carrier as an electromagnetic wave using a suitable antenna. At the receiving end, an antenna system tuned to the central frequency, such as that of the carrier frequency, receives the EM wave. The baseband signal(s) are recovered from the passband of the tuned-in radio frequency spectrum using appropriate demodulation techniques. Further, as needed, regenerative repeaters are interposed between the transmitting and receiving end.

The range of frequencies in the electromagnetic spectrum compatible for radio transmissions with the available technology, in general, spans widely, stretching from almost 100 kHz (termed as *long-wave transmissions*) up to about 60 GHz (called *millimetre (mm) waves*). This wide range of the EM spectrum is divided into specific bands and these bands are designated for specific applications. Exclusively for modern wireless communication purposes, the UHF and/or *microwave* bands are used. (Classical radio-telegraphy and telephony adopted the short-wave band for long-distance communications.)

There are unique signal impairment situations (as will be discussed in detail later in this chapter) associated with wireless telecommunications. Wireless communication, in essence, is a point-to-point communication system, but there could be multiple transmission paths resulting from reflections and scattering of EM waves by physical structures such as buildings etc. or due to refractory effects caused by the atmosphere. The received signal is a vector sum of these multipath-traversed constituents, namely, the *primary ray* and the *delayed secondary rays*.

The extents of attenuation suffered by these rays would be different and may change with time. Such changes are significant in mobile and cellular phone applications. The attenuation is controlled by atmospheric conditions as well as by shadowing and other scattering of the EM waves involved. In effect, wireless telecommunication signals face what is known as time-dependent "signal fading".

To counter the effects of fading, a *fade-margin* is facilitated. Increasing the transmitted power and/or incorporating *frequency*- and *space-diversity* receptions are envisaged in practice to facilitate the fade-margin. (In frequency diversity systems, the same intelligence is transmitted over more than one carrier. It is expected that, even if one channel fades, the other channels are unlikely to fade. Hence, the information can be extracted from the unfaded channels. The space diversity system uses a single carrier but the reception is done at multiple, spatial-dispersed receiving antenna/receiver systems. Again, if a faded reception is perceived at one receiving locale, the other locales may probably receive unfaded signals. Therefore, the information can be recovered from these unfaded receptions.)

Facilitating reliable communication, despite the inevitable fading conditions, poses, however, a host of challenges in the operational scenario of modern wireless telecommunication services. Nevertheless, diversity-based system technology, different coding techniques, and spread-spectrum based strategies are adopted to minimise the effects of impairments in such wireless transmissions; and the networks implemented use such strategies consistent with the standards and application profiles.

Apropos the variety in the existing wireless standards, a number of associated networks have emerged and have been adopted. These are briefly reviewed below and are comprehensively addressed in the list of books referenced as [1.1] through [1.20].

1.3.1 Cellular voice networks

These essentially represent wireless networks of voice telephony, analog or digital. The analog system like AMPS uses frequency modulation with a peak deviation of 12 kHz. A duplex phone conversation requires one channel for transmitting and another channel for receiving.

The cellular system (as indicated before in Figure 1.3) includes a MTSO (sometimes situated in a CO of PSTN), the cell sites, and the mobile units. The central processor at the MTSO controls the switching equipment needed to interconnect the mobile users with the wireline telephone network (PSTN). It also controls cell-site actions and many of the mobile unit actions through commands relayed to them by the cell sites.

The MTSO is connected to each cell site *via* wireline PSTN, over which they exchange information required to process the calls. Each cell site has one transreceiver for each of its assigned voice channels and the corresponding transmitting/receiving antennas for those channels. The equipment at the mobile station contains a control unit, a transreceiver and an antenna, with a duplexer, to

segregate the transmitted and received signals. (In space diversity systems, if used to combat fading and interference problems, there will be more than one antenna in operation.) The control unit has user interfaces, namely, the handset, the dialler and a light-emitting diode (LED) indicator. The transreceiver includes a synthesizer to tune to all the allocated channels. The logic section interprets the user's actions and various system commands. Subsequently, it also controls the transreceiver and the control units.

Example 3.1
Determine the spectrum allocation for a metropolitan area service supporting 666 duplex AMPS channels.

Solution
Number of channels = 666
Type of service: AMPS duplex voice
Modulation: Frequency modulation with 12 kHz peak deviation ($[\Delta f_{peak}]$
Bandwidth (BW) of each channel: $2[\Delta f_{peak} + f_{m(max)}]$ $\Rightarrow$ *Carson's rule*
(where $f_{m(max)}$ is the highest modulating signal frequency)
$f_{m(max)} = 4$ kHz (for voice signal)

$$\therefore \quad BW = 2 \times (12 + 4) = 32 \text{ kHz}$$

Hence, total spectral allocation needed for the service $= 2 \times 32 \times 666$ kHz
$= 42.624$ MHz

A few radio channels are used for *call set-up* and *tear-down* procedures. First, they facilitate the necessary exchange of signalling information for call set-up. Whenever the mobile unit is "on", but not in use, the unit continuously monitors these set-up channels. And selects the strongest one. Thus, the base station/cell and the current location of this mobile unit (linked *via* that selected channel) are mutually identified. That is, the mobile unit and the base station are now synchronised through the set-up channel with an identification number (ID) prescribed for the mobile unit. Now, the base station can transfer any call coming for the mobile unit.

When the mobile unit is alerted of an incoming call, it again samples the signal-strength of all received signalling/set-up channels and responds through the cell-site offering the strongest signal. It then indicates its preference for communication through that cell-site/base station, gets an ID and the base station, in turn, responds by supplying the ringing tone so as to alert the user to pick up the phone. Similar sequence of events takes place at the calling subscriber side except in the reverse order.

Due to the mobility involved, the system examines the call being received every few seconds at the cell-site. If necessary, the system "looks" for another site

to serve the call. When such a need occurs, the mobile unit retunes itself to the newly found cell-site by receiving a command from the base station. This process of changing the base station or handing over the responsibility of servicing a mobile unit, as mentioned earlier, is known as "hand-off". It occurs in a brief interrupted period of about 50 ms.

The power output of the mobile unit is about 0.7 to 3W and is controlled by a power up/down "power control signal" from the base station in seven, 1-dB steps. This power control is necessary to reduce the interference with other phones and minimise the overloading of the base-station receiver.

After their inception, analog cellular networks soon exhausted the FDMA channels (with 30 kHz FM bandwidth for each channel) offered to the growing population of subscribers (within the allocated spectrum). Hence, the second generation of digital wireless telephony came into the picture with a channel allocation based on TDMA and/or CDMA schemes, as mentioned earlier.

1.3.2 Personal communication systems and networks

The personal communication systems and networks (PCS/PCN) are *location-independent* communication systems as illustrated in Figure 1.5. They allow freedom of communication for any type of information between any two points. The locales of end-entities can be indoor/outdoor, in a mobile unit, rural areas with sparse population, crowded metropolitan areas, in an aeroplane or at sea. The end-entities can be at a stand-still or be moving at jet-speed. The separation between the end-entities could be arbitrary.

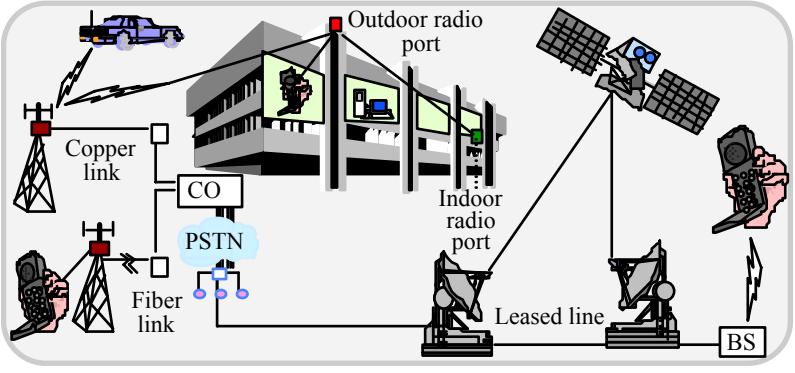


Fig. 1.5 Generic PCS framework and its global connectivity considerations involving both wireline and wireless transmissions

PCS/PCN is an emerging global system with international connectivity. A standing committee on ITU, namely the World Administrative Radio Conference (WARC) co-ordinates the *Future Public Land Mobile Telecommunication* (FPLMTS) systems so as to develop a global system of PCS/PCN for aggressive

deployment in the coming years. Some of the challenging aspects of the associated efforts are as follows:

- Judicious use of available EM spectrum
- Choosing appropriate technology by taking into account of spectrum scarcity
- Poising the usage demands and services to be offered.

TDMA techniques, together with SS methods and CDMA strategies, are targeted for use in PCS implementations. For example, in order to enhance the traffic capacity of PCS, digital multiple access methods such as time-division duplexed TDMA (TDD-TDMA), frequency-division duplexed TDMA (FDD-TDMA), slow frequency-hopped TDMA/CDMA etc. are being considered.

The general framework criterion towards PCS implementation is to render the system interconnected and internetworked so that the users are free of wirleline tether (tetherless) and/or cord (cordless). PCS should enable voice as well as fax and computer data transfers on a personal basis across an exhaustive internetwork of terrestrial (wireline and/or wireless) and satellite links with global coverage. Further, multiuser outdoor radio-ports would permit the users in a building to get access to remote end-entities. Likewise, an indoor radio-port would interconnect the users within a building.

Thus, regardless of the locale of a person and the type of end-entity, communication is to be facilitated in the envisioned global PCS. Eventually, adjunct to telephone service, PCS will include end-entities comprised of mobile personal computers — the laptop/palmtop version, handhelds, subnotebooks, and personal assistants (personal organiser plus pager and cellular telephone) [1.6 – 1.8]

Personal communication network (PCN) is a concept that includes a single identifying number with an individual subscriber. That is, each person will have a personal ID number (PIN), eliminating the need for separate home or work numbers. Any traffic intended for an individual reaches that person regardless of the location (of that person), rather reaching that user's (home or work) number. That is, the user can receive telecommunication services over a wide geographical area and the service is almost mobility-independent. The user is just expected to carry an operational wireless communication device throughout roaming. Unlike cellular systems, the PCN is designed to operate independently as well as interface with the standard wireline network. It is designated to operate in the 1.9 GHz band. Relevant radios are intended to be very small and the transmitted power is kept low. This warrants more cell-site deployment. Hence, PCNs are known as *microcellular systems*.

1.3.3 Wireless data networks

The wireless networks specified exclusively for data transports are termed as *wireless data networks*. These are classified in accordance with their coverage range. The most extensive network covering a wide area is the *wireless wide area network* (WWAN) that may span an entire country. The network that connects the

residents and visitors in a metropolitan area wirelessly is termed the *wireless metropolitan area network* (WMAN). A wireless community area network (*wireless CAN)* includes the systems intended for a small coverage area such as a university or a hospital campus. The wireless CAN, in a restricted sense when deployed within a localised area such as a building, office etc., becomes a *wireless local area network* (WLAN). It can be wired to the legacy LAN in the premises such as the Ethernet. Lastly, a wireless personal area network (WPAN) refers to the implementation of a wireless network in the smallest area possible, for example, a sub-region of a building or home, where the cluster of furniture or other contents in the surroundings could have an influence on the nature of the wireless propagation [1.9].

1.3.4 Wireless LAN

For wireless LAN operation, the regulatory bodies have permitted to share the so-called *ISM bands* (Industrial, Scientific and Medical bands: See Table 1.1 below) with existing systems. WLAN is classified as an "intentional radiator" in its scope of implementation in the ISM bands.

The Part 15 Rules of the FCC allows unrestricted (unlicensed) radio communications at these ISM bands, however, with constricted maximum power in the transmissions. In applications up to 1 W use of *spread-spectrum* (SS) technique is mandatory. In such cases, the radio implementations should ensure that the *spreading ratio* (ratio of signal bandwidth after spreading the spectrum to raw or "unspread" signal bandwidth) exceed a factor of ten in *direct sequence spread-spectrum* (DSSS) and requires a minimum of 50 and 75 hopping frequencies at 910 MHz and 2.4 GHz in the *frequency hopped spread-spectrum* (FHSS).

The typical WLAN implementations in the United States refer to AT & T's 2-Mbps WaveLAN™. Motorola's proprietary WLAN, namely Altair™, operates at 18 GHz and provides a 10 Mbps Ethernet interface. In Europe, the ETSI-defined WLAN standard, known as HIPERLAN, operates in the 5.1-5.3 GHz range and supports 23.5 Mbps. Its standard specifies 50 m coverage indoor, with options for relaying *via* HIPERLAN nodes and/or wired infrastructure. In the United States, the FCC has also allocated 300 MHz of bandwidth in the same 5 GHz regime as per the petition from the National Information Infrastructure (NII) and Supernet. Other formal workgroups (such as MM wave Working Group) have obtained a spectrum for millimetre wave medium access for WLAN deployment.

The WLAN standards are specified by the IEEE 802.11 working group, which define a number of services that need to be provided by the wireless LAN with the functionality equivalent to that of legacy wired LANs. The legacy LANs, interconnect the computers through copper and/or fibre lines with clients and servers placed at fixed locales. With roaming/mobile end-entities, the wireless connectivity is facilitated *via* WLAN systems. The traditional and wireless LAN configurations are illustrated in Figure 1.6.

The physical media prescribed for WLAN transmissions are the *infrared* (IR), the *radio frequency* (RF/UHF) and the *microwave* (including millimetre wave) ranges of the electromagnetic spectrum. Excluding the IR-band, the

RF/microwave ranges conform to ISM bands indicated earlier. Relevant United States prescriptions are: 902-928 MHz, 2.4000-2.4853 GHz, and 5.725-5.850 GHz. The operation is based on the spread-spectrum technique and the emitted spectrum is confined to the prescribed band. Further, these are low-power systems with a maximum permissible power level of 500 mW.

Table 1.1 The ISM bands: The frequency bands adopted in different countries

Geographical Area	ISM band (in GHz)	Guard bands (in MHz)		Number of channels
		Lower	Upper	
France	2.4465-2.4835	7.5	7.5	23
Other countries	2.4000-2.4835	2.0	3.5	79
USA	2.4000-2.4835	2.0	3.5	-
	0.9020-0.9280	-	-	-
	5.7250-5.8500	-	-	-

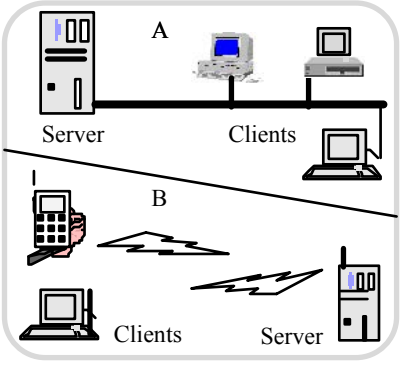


Fig. 1.6 Local area networking: (A) Legacy (copper line and/or fibre-based) wireline LAN and (B) wireless LAN

The IR system uses the same wavelengths used on fibre links. (Typically, the wavelengths are in the range 1300 nm and 1500 nm.) These systems are not bandwidth constrained and operate on *line-of-sight* (LoS) principle. The modulation scheme prescribed for WLAN, as stated before, is the spread-spectrum technology of either DSSS or FHSS type. A major consideration as regard to the

electromagnetics of WLAN transmissions deployed in the indoor environment is the prevalence of intense multipath transmissions and the resultant fading effects. The statistics of such fading is largely refers to the Rayleigh fading in which the signal intensity fluctuations follow the Rayleigh distribution. The antenna designs, therefore, are made to cope with such RF signal fading environments.

1.3.5 The Bluetooth™

The Bluetooth™ is an interesting up-coming technology. It was originally conceived by Ericsson and has been enthusiastically ushered in and promoted to enable devices of all kinds — from laptops and cell phones, to personal digital assistants and household appliances — to have the capability of communicating and interoperating with one another in the wireless medium [1.11 -1.13]. The underlying concept of the Bluetooth technology remains the transparent wireless RF communication between electronic devices. This mode of communication can offer a set of advantages encompassing a viable voice/data access point, no cumbersome cabling and creation of an "as-you-please" *ad hoc* wireless network. Bluetooth technology is characterised by personal connectivity of enabled devices, spontaneous creation of disposable networks and low-cost, low-power indoor wireless transmissions of short ranges supporting 721 kbps or 432.6 kbps over asynchronous and synchronous links respectively.

Essentially, the transmitter and receiver parts of the Bluetooth system use the hopping frequency scheme, thus facilitating modulation and demodulation of the baseband signal with a frequency-modulation scheme. That is, the Bluetooth conceptually connects everything wirelessly with a frequency-hopping radio link in short-range areas.

For example, a Bluetooth solution can replace the interconnecting cable across any electronic devices to be linked together for communication. Further, the Bluetooth carries a cost-attractive advantage and is expected to be the new hot-comer in the wireless technology, which may eventually replace existing wireless networking concepts, such as the RF-based WLAN (of IEEE 802.11) and/or infrared-based WLAN of IrDA (Infrared Data Association). Though an IrDA system is not wired, it is, however, limited for use only between two devices to communicate with each other at the same time along unobstructed LoS because of the optical (IR) link adopted. The Bluetooth on the other hand, can work with more than two devices and the line-of-sight requirement is not emphasised. Moreover, the Bluetooth technical specification is completely open, that is, no licensing fee is required for using the associated specifications by any developer and FCC regulations are just limited to ISM band constraints.

As indicated before, the Bluetooth operates around 2.4 GHz radio spectrum that belongs to the Industrial, Scientific, and Medical (ISM) band. This 2.4 GHz band is available world-wide. Specifically, a bandwidth between 2400 MHz and 2483.5 MHz is specified under FCC Part 15 regulations; and, ETS-300328 also regulates the same band. Further, Japan has also allowed the use of this band in commercial applications. The RF characteristics of the Bluetooth are summarised in Table 1.2.

(The details presented in Table 1.2 represent a summary of raw data compiled from various sources. For more details and updated specifications, the readers are encouraged to refer to publications and web presentations such as: Bluetooth Special Interest Group (SIG), "Specification on the Bluetooth System". Core Version 1.1, February 2001 and *http://www.bluetooth.com.*)

Bluetooth products are qualified through the Bluetooth Qualification Program. The specification supporting this programme lists rules for compliance, including interoperability with other Bluetooth devices from other manufacturers. Certified "Bluetooth-enable" devices must comply with this specification.

A Bluetooth system contains, in essence, a radio receiver, a link controller to connect the transreceiver to the link manager. The link manager interfaces with the application software, which talks to the host hardware layer. The antenna part performs RF-to-air interfacing in the transmit- and receive-operational phases.

Original equipment manufacturers (OEMs) can choose between a Bluetooth solution with the antenna incorporated in the package or with an antenna tailor-made and located externally to the unit. In the integrated module the embedded antenna eliminates requirements for tuning and the need for an additional connection on the printed circuit board (PCB). It is observed that a Bluetooth-enabled device (such as PCMCIA devices) may be prone to ineffective radiation from the external antennas as a result of improper implementation with the antenna too proximal to the grounded metal case. Hence, the antenna considerations in Bluetooth applications need a careful design and implementation strategies.

Table 1.2 Bluetooth RF characteristics: Its various parameters and their specifications

Parameters	Specifications
Carrier frequency	2400 to 2483.5 MHz: (ISM radio band) *Exceptions:* 2445 to 2475 MHz (Spain); 2446.5 to 2483.5 MHz (France) 2471 to 2497 MHz (Japan) $f = 2402 + N \times 1$ MHz, $N = 0,\ldots, 78$ $f = 2449 + N \times 1$ MHz, $N = 0,\ldots, 22$ $f = 2454 + N \times 1$ MHz, $N = 0,\ldots, 22$ $f = 2473 + N \times 1$ MHz, $N = 0,\ldots, 22$
Modulation	0.5 BT Gaussian-filtered 2 FSK at 1 Msymbol per second Digital FM scheme Modulation index: 0.28 to 0.35 (0.32 nominal) The peak frequency deviation allowed is 174 kHz

	Hopping rate: 1600 hops/s (in normal operation). This may vary along the packet length Each channel spacing : 1 MHz Channel hopping sequence: A pseudorandom sequence designed to visit each frequency regularly and with roughly equal probability. It has a periodicity of 23 hours and 18 minutes The system has five different hopping sequences:
Hopping	1. Page hopping sequence 2. Page response sequence 3. Inquiry sequence 4. Inquiry response sequence 5. Channel hopping sequence The first four hopping sequences are restricted hopping sequences used during connection set-up. The normal channel hopping-sequence is pseudorandom based on the master clock value and the device address
Transmit power	Power class 1: 1 mW (0 dBm) to 100 mW (+ 20 dBm) Power class 2: 0.25 mW (– 6 dBm) to 2.5 mW (+ 4 dBm) Power class 3: 1 mW (0 dBm) Class 1 power control: + 4 to + 20 dBm (required) – 30 to 0 dBm (optional) Class 2 power control: – 30 to 0 dBm (optional) Class 3 power control: – 30 to 0 dBm (optional)
Operating range	10 cm to 10 m (100 m with Power class 1)
Maximum data throughput	The asynchronous channel can support an asymmetric link of maximally 721 kbps in either direction while permitting 57.7 kbps in the return direction, or a 432.6 kbps symmetric link Data throughput is lower than the 1 Msymbol/s rate as a result of the overhead, which is inherent in the protocol.

1.3.6 Wireless ATM (WATM) networks

Thanks to the success of *asynchronous transfer mode* (ATM) on wireline networks [1.14], *wireless ATM* (WATM) has become an imposed strategy towards the "ATM everywhere" movement. The WATM can be viewed as a solution for

next generation PCNs, or a wireless extension of the broadband ISDN (B-ISDN) networks, which will support integrated transmissions of voice, data and video with guaranteed *quality of service* (QoS). The ATM Forum has been working on defining the baseline of WATM systems. The specification for both mobility control in ATM infrastructure networks and seamless radio extension of ATM to mobile devices are some embryonic considerations being mooted.

Table 1.3 WATM physical layers (PHY) specifications: Low-speed and high-speed PHY details

Parameters	Specifications	
	Low-speed ATM PHY	High-speed ATM PHY
Frequency band	5.150-5.350 GHz 5.725-5.875 GHz	59 GHz-64 GHz
Cell radius	80 m	10-15 m
Transmit power	100 mW	10-20 mW
Frequency reuse factor	Up to 12	7
Channel bandwidth	30 MHz	150/700 MHz
Data rate	25 Mbps	155/622 Mbps
Modulation	16 DQPSK	32 DQPSK
Medium access control (MAC) interface	Parallel, transfer speed: 3.127 Mbyte/s	Parallel, transfer speed: 87.5 Mbyte/s
ATM cell length	[PHY header + MAC header + (4 × ATM cell)]	

The overall WATM system consists of a fixed ATM network infrastructure and a radio access segment. In the fixed ATM networks, the switches that communicate directly with wireless station or wireless end-user entities are mobility-enhanced ATM switches. They set up connections on behalf of the wireless devices. They serve as the “entrance” to the infrastructure wireline ATM networks. The other ATM

switching elements in the ATM networks, however, remain unchanged.

While a fixed station may comfortably serve a 25-Mbps to 155-Mbps data rate ATM link, even an 25-Mbps data link in WATM environment is rather difficult to realise. In a wireless environment a several gigahertz spectrum is required to support a high-speed transmission. A recommended band for WATM is 5 GHz, which can support 51 Mbps channels only with advanced modulation and special coding techniques.

Although 155 Mbps is not reachable (at this time), it is anticipated that, when 60 GHz spectral operation is comprehended, even 622-Mbps rates could be achieved over WATM! Consistent with this utopian dream, two physical (PHY) layer specifications have been recommended, one for the 5-GHz band and the other for the 60-GHz band for the Herculean effort involved. These are summarised in Table 1.3.

To understand why wireless ATM has become a technology of interest, it is pertinent to understand the market drivers behind the scene. These market drivers indicate the existing co-operation between standards organisations, which has led to the agreements on wireless broadband ATM architectures on their performance.

Foreseeing the comforts of wireless ATM-based teletraffic, several vendors and education institutions have co-operated in realising a number of prototype networks. The Magic WAND, RACE II, and RDRN are a few to mention.

The quest for large bandwidth coupled with the ever-increasing mobility of today's society is the precursor for wireless ATM. It has been estimated that the need and market for wireless ATM can jump by about 100 % by 2005.

The conceivable applications of wireless ATM refer to the following locales: The workplace, home, and "wireless to the curb." The workplace application will allow multimedia, bandwidth-intensive applications to run on mobile devices. Video conferencing, access to multimedia databases, and Internet access will not be limited by the number and location of LAN ports and wiring. Mobile devices might also include ATM-ready PDAs. Workplace applications are likely to be the first applications implemented. Home-based applications might also include the following:

- Portable enhanced television services, including video-on-demand, home-shopping, banking, games etc.
- Wireless in-home networking and Internet access for traditional PCs and mobile network computers
- High-quality audio distribution throughout the home, allowing more flexible positioning of audio components, cordless telephony, and replacement of miscellaneous legacy wiring systems.

The "wireless to the curb" application of wireless ATM differs from the first two applications in that both the base station and subscriber devices would be fixed.

On the consideration of possible uses for a wireless ATM network, there are several constraints that must be looked into when the standards are developed. The primary factor is the QoS. Other hurdles to be faced are:

- *Access schemes* in reference to protocols that can overcome high error rates and noise inherent in radio systems

- *Reliability* and *availability* specified by coverage areas, fading temporary outages, error detection, and correction
- *Service ubiquity* posed by wireless reaching difficult downtown areas, carriers' responsibility to acquire licenses, acquiring rooftop space, and deploying access stations
- *QoS mobility* to ensure consistent QoS and hand-off as a user wanders
- *Applications* made to overcome the limitations inherent in wireless transmissions.

Standards on wireless ATM are still emerging. Several organisations have been working on standards from many different perspectives. Both European and North American standardisation efforts are on trial. The standards cover a wide range of topics pertaining to both additions to the ATM protocol standards for communication on air (both mobile and fixed) and enhancements to wireless standards for usage and globalisation.

The ITU-T has launched its International Mobile Telecommunications - 2000 (IMT-2000) initiative "to provide wireless access to the global telecommunication infrastructure through both satellite and terrestrial systems, serving fixed and mobile users in public and private networks" (ITU-IMT, 1999). Study Group 11 of the IMT met in July 1999 to address signalling for broadband and multimedia networks and services.

Working within the framework of IMT-2000 is the UMTS Forum (which was founded in 1996). It is a non-profit organisation with representation from over 180 member organisations. This group is working with other standards organisations including the ITU, ETSI, GSM Association, and ANSI to develop the Universal Mobil Telecommunication System (UMTS). The UMTS Forum met August 9 –12, 1999 for a workshop on 3G Mobile Broadband.

The ETSI has also established a standardisation project for *Broadband Radio Access Networks* (BRAN). Prior to the establishment of the BRAN project ETSI had released functional specification EN 300 652 HIPERLAN described earlier. The BRAN project will enhance the HIPERLAN functionality to HIPERLAN Type 2 and begin working on standards for wireless access and broadband interconnects.

At present there are several documents in various stages of approval under the BRAN project. The BRAN project has also been working closely with the ATM Forum, IEEE, and the ITU-T to avoid duplication of efforts and ensure a cohesive set of standards.

The ATM Forum had also begun looking into standards for wireless ATM. In 1996 the wireless ATM working group was formed. In part, the WATM working group will address two items, as follows:

- Radio access layer protocols including radio physical layer, medium access control for wireless channel errors, data link control for wireless channel errors, and wireless control protocol for radio resource management

- Mobile ATM protocol extensions including hand-off control (signalling, network-to-network interface (NNI) extensions, etc.), location management for mobile terminals, routing considerations for mobile connections, traffic/QoS control for mobile connections, and wireless network management.

In addition, there are smaller groups such as the Wireless Broadband Association and the Delson Telcom Group's Task Force — Wireless Mobile ATM (TF-WMATM). These are small independent organisations working on generic solutions to wireless ATM access. The TF-WMATM focuses on implementation issues and has held several conferences.

Pertinent to WATM architecture/networking, the wireless base stations act as an interface into the switched wireline ATM network. This network is based on the wireless ATM reference architecture. This reference architecture defines new interfaces for the ATM protocol stack. The *wireless user network interface* (or "W"-UNI) deals with handover signalling, location management, wireless link and QoS control. The *radio access layer*, (or "R"-AL) governs the signalling exchange (that is, channel access, data link control, etc.) between the WATM terminal adapter and the mobile base station. The *mobile network-network interface*, (or "M"-NNI) governs signalling exchange between the WATM base station and a mobile capable ATM switch, as well as mobility-related signalling.

Considering the WATM protocol stack, the bottom three layers make up the *radio access layer* (RAL). The RAL is required to support the new wireless specific protocols that will define the physical and access layers. At the physical layer there are multiple competing solutions. Currently the two front runners are CDMA and TDMA but new technologies are on the way. It is likely that there will be more than one answer. The *medium access control* (MAC) relates to the user's ability to access the wireless media in order to guarantee the QoS. The base station is the co-ordinator for channel access, and is aware of the wireless resource utilisation. The MAC layer will likely have to support multiple physical layer protocols.

The *data link control* (DLC) protocol will have to support new error handling to account for wireless channel errors. The DLC will be responsible for transmission and acknowledgement of frames, frame synchronisation and retransmission, and flow control.

There have been in the recent past, a few proof-of-concept projects on WATM. A consortium of vendors and educational institutions operates these projects. The *Magic Wireless ATM Network Demonstrator* (WAND) and the *Rapidly Deployable Radio Networks* (RDRN) are examples of two such projects to mention.

The Magic WAND is a project performed under the auspices of the *advanced communications technologies and services* (ACTS) research program of the European community through a consortium of vendors and educational institutions. The Magic WAND project chose the 5 GHz frequency band for the demonstration. The project covered design, implementation, and testing with actual subscribers.

The RDRN project is funded by the Information Technology Office (ITO) of the DARPA and is carried out at the University of Kansas. The goal of the project is to create architectures, protocols, and prototype hardware and software for a high-

speed network that can be deployed rapidly in areas of military conflicts or civilian disasters where communication infrastructures are lacking and or destroyed.

Although wireless ATM is still in its infancy, expectations are high. The world wide standards organisations, international educational institutions, vendors, and service providers are all working towards a standard WATM offering in the foreseeable future. When brought to fruition WATM will be able to provide today's mobile society with high-speed, high-quality mobile accesses from virtually anywhere.

1.3.7 Wireless application protocol (WAP) technology

This technology is intended to link the Internet to wireless portable devices such as mobile phones. This convergence has spearheaded the feasibility of mobile commerce (*m-commerce*). This technology, however, warrants new types of handsets with WAP capability.

Using a WAP phone involves more than "cutting through the cord". The content providers should know where their users are geographically located, and so they can direct them to specific destinations like restaurants, theatres etc. Further, they have to carry out a secured transaction. Also, the small screen requires simple but adequate data display covering rather a rich menu of goods and services meeting the mission of the m-commerce. Thus, WAP will underpin the convergence of mobile communications and the Internet.

1.3.8 Wireless local loop (WLL)

The WLL is designed to provide a wireless local loop access to telephone subscribers. It is installed (*in lieu* of copper- or fibre-based local loops) in "green-field start" projects. The DECT and a cordless system known as CT2 (that enables privacy through encryption of the coded voice signal) can be used for WLL services. In the relevant implementation, the subscriber has a cordless terminal adapter connected to the telephone equipment. A point-to-point wireless connectivity (CT2 or DECT) is enabled from the public network *via* a radio fixed part (up to 1 km distance). With more radio fixed parts, large-area coverage can also be achieved with the WLL implementation. Further, a modern trend is to improvise broadband transmissions on WLL. Relevant issues are elaborated in Chapter 8.

1.4 WIRELESS SYSTEMS OTHER THAN CELLULAR TELEPHONY

These refer to: (i) *Paging*; (ii) *private mobile radio* (PMR); (iii) *satellite mobile systems* and (iv) *personal communication systems*. The PCS system was discussed earlier and the satellite mobile systems will be detailed in the next section. The other two systems, namely, the paging and PMR are outlined here.

Paging

A simple pager is a wireless system that notifies a called-party *via* an alarm (beep) or a defined voice plus an alphanumeric display that someone (the calling-party) is interested in talking. It is essentially a one-way traffic. It could be a private (that is, local premises-oriented, as in hospital paging systems), or it could be a public

(wide area) paging system. The first generation of a paging system was introduced in the late 1950s and early 1960s. The windows of frequency band that are adopted for paging lies in the range 80-1000 MHz. Paging service is offered to several million users.

A modern version of paging refers to *smart pagers* (for example Motorola Pagewriter^{TM} 2000). This allows two-way communication to send and receive word messages. It is claimed to be smart enough to let one communicate with almost anyone else having a pager, Internet, e-mail address, or fax. It is also claimed that it is easy enough to create messages on the go with a full keyboard. Another claim is that, it is direct enough to respond from almost anywhere making it the most personal form of two-way communication.

Problem 1.1
The paging systems are designed for indoor and outdoor operational capabilities with utmost reliability and robustness for 2 to 5 km range operations with the paging transmitter antennas placed at tall towers/buildings.

In reference to this context, justify the following considerations:

- Simultaneous broadcasting a page from several locations (*simulcasting*)
- Facilitation of several transmitters on the perimeter of the city
- Use of small bandwidths
- Data rate being 6.4 kbps or lower
- Private mobile radio (PMR)

Private mobile radio (PMR)
The PMRs refer to a fixed base station serving a number of mobile units for private user applications (for example, to direct taxicabs to a customer). These operate in the VHF/UHF bands. The transmitter power could be in the range of 5 to 25 watts. Both FM and AM strategies are used.

Problem 1.2
Access the details on ETSI-specified, trunked-PMR that has all-digital signalling and voice/data handling and known as the *Trans-European Trunked Radio* (TETRA). Compare the TETRA service *versus* the GSM system.

1.5 SATELLITE-DEPENDENT MOBILE SYSTEMS

Satellite mobile systems are useful means of communication for long-distance travellers. Also, they are adopted in aircraft and ship navigational purposes. National Aeronautics and Space Administration (NASA) through the ATS-6

satellite undertook the premier efforts on mobile satellite communication. In 1979, the INMARSAT (International Maritime Satellite) organisation was set up to establish world-wide aeronautical satellite communication standards governing the telephone and telex services (Standard A), ISDN (Standard B), and low data rate services (Standard C). Since 1983, the International Civil Aviation Organisation (ICAO) is also involved in the related activities [1.15]. The other satellite-based mobile systems include the following:

- *Radio Determination Satellite System* (RDSS) of Geostar: This is a radio navigation/radio location system managed *via* a single satellite for US based domestic applications
- *OmniTracs* system of Qualcom Inc.: This is a two-way mobile satellite communication system intended for vehicle positioning and is adopted in the United States and Europe
- *MobileSat*: This is an Australian system which supports services such as circuit-switched voice, and data/packet-switched data transfers for land, aeronautical, and maritime users
- Telesat Mobile Inc. (TMI)/American Mobile Satellite Corporation (AMSC) systems: These systems provide mobile satellite services in the United States and in Canada.

Global communications are often based on satellite systems because of the relative ease with which they establish global coverage. Although fibre-optic networks yield world-wide coverage, it is only a parallel strategy to the satellite-based relaying. For the first time, the public satellite communication was deployed in 1962 with the efforts due to NASA through the Application Technology Satellite (ATS) program.

Satellite communication offers an extensive coverage or *footprint* on the earth and allows a seamless mobile service. The size of the footprint depends on the orbit of the *geostationary* (GEO) satellite deployed. For example, with an orbit at 36,000 km above the earth, the *field-of-view* (FoV) coverage is about 13,000 km diameter. There are *low orbit* (LEO) satellites, typically at an altitude of about 800 km and an FoV of about 1,500 km diameter. The communications *via* LEO satellites face small path-losses, which allow for small-sized earth antennas.

Inmarsat was developed in 1973 to provide maritime communications worldwide through GEOs and it offers services to all modes of transportation by air, land, or sea.

Iridium is a global satellite phone system, with voice and data traffic. It offers services like, digital voice, paging, faxing, *global positioning service* (GPS) etc. Iridium satellites are interconnected via microwave cross-links and interface gateways with the PSTN.

Problem 1.3

Although a geosynchronous satellite is stable, the earth station on a ship or a truck is not. Discuss the engineering of relevant antenna deployments in such earth stations.

(*Hint*: Consider the relevancy of stability/autostabilisation versus the weight of the antenna)

Example 1.2
For road and maritime traffic applications, the communication facility is usually integrated with a system that provides accurate (about 20-metre resolution) information as regard to the position of the vehicles on the earth *via* GPS.

Compare the antenna-specified performance limitations of GPS systems used in aircrafts as against the surface (road/marine) vehicles.

Solution
Drag-free mounting considerations in aircrafts/missiles warrant special antenna shapes and mounting arrangements (as will be detailed later in this book). Such tailored shapes may not be optimal for error-free (and resolution-constrained) GPS operation.

Further, Doppler effect is more pronounced in airborne systems, which should be duly accounted for in the signal intercepted at the antenna using signal-processing.

Lastly, the ground-reflected multipath effects are more severe in airborne systems at their normal cruising speeds.

1.6 IMPAIRMENTS TO WIRELESS COMMUNICATION

In the wireless system, the physical medium corresponds to free-space in which the electromagnetic wave propagates. As in any electrical/electronic system, the wireless signals are subjected to corruption by the inevitable presence of noise at the transmission and reception ends as well as along the transmission medium. Apart from the device/system based electronic noise (such as thermal noise) corrupting the signal, certain characteristics of EM propagation in the mobile environment would also impair the signals under transmission [1.16]. Such impairments arise from signal fading due to *scattering, reflection*, and *refraction* effects that the EM wave may face during propagation and attenuation of EM energy as a result of absorption by rain, snow etc. With the result, for robust implementation of wireless communication, the receiver technology is being trimmed continuously through available techniques and devices.

Thus, all along the passage of information transfer, from the transmitter through the channel to the receiver, there exists a host of noise sources (Figure 1.2), which may introduce undesirable effects to the signal being transported. Generally classified as noise sources, these are categorised into two versions in reference to the wireless communication systems. They are:

- Additive noise sources resulting from electronic parts of the transmitter and receiver and due to interference from extraneous electromagnetic effects

- Multiplicative noise due to nonlinear processes encountered by the electromagnetic wave propagation across the channel.

As indicated above, exclusive to wireless communication systems and the associated electromagnetic wave propagation, the signal along the channel may also suffer a loss in its intensity along the path. It could also be subjected to fading and shadowing arising from the reflection, refraction and diffraction effects encountered by the electromagnetic wave. These phenomena could significantly influence the signal intercepted at a mobile unit. As the mobile unit traverses its path, the propagation characteristics and interference effects could change dynamically due to the mobility involved. Corresponding considerations find due place in the studies devoted to modern wireless communication systems.

In short, the "genesis" of noisy wireless communication is as follows: "In the beginning of wireless, there was just 'noise', and life was simple. Engineers soon called this noise additive white Gaussian noise (AWGN), but even this title was not enough to capture all the idiosyncrasies of a real world RF channel. Then, engineers, scientists, and mathematicians got involved, and they soon gave names to more subtle effects, such as Rayleigh fading, Ricean fading, impulse noise, cyclostationary noise, and inter-symbol interference, to name just a few of the newer noise-family members. All these noises were in addition to basic channel difficulties of signal attenuation and loss".

Technological advances and consumer demands are changing the landscape of wireless communications. At present, commercially available wireless mobile communications transmit at rates up to 2 Mbps. At high data rates, radio channels are impaired and need to be reversed at the receiver to realise an acceptable BER. Such impairments, as indicated before, are mainly due to multipath fading. The trend in the technology is to enhance the bit rate, reduce the multipath fading induced impairments and facilitate mobility enhanced applications including broadband and multimedia services.

Problem 1.4

Perform a survey of major mobile radio standards adopted in North America, Europe and Japan. Hence, identify and tabulate all the applicable standards against the services indicated below:

Type of service	Compatible standards
High-power, wireless WAN/Cellular voice Low-power, wireless CAN/Data service Wireless LAN/Low data rate mobile Wireless LAN/High data rate mobile/Indoor Indoor *ad hoc* mobile	{?}

(*Hint*: See [1.1]/pages 7 and 8 and [1.12]/page 12)

1.7 WHITHER ANTENNAS FOR WIRELESS COMMUNICATIONS?

The overall perspective of wireless communication systems plus the pertinent details on the state-of-the-art aspects of associated technology as indicated above show that, in the modern deployment profile, wireless communications and networks pose multidisciplinary considerations. Relevant engineering involves robust telecommunication networking, skilled signal-processing techniques (covering voice, data, video and multimedia information), appropriate modulation schemes (for efficient and effective use of the limited spectral windows available), intricate RF technology (at the transmitter and receiver sides), and intelligent signal-recovery processes at the baseband level. Though all these endeavours are within the broad scope of electrical communication engineering concepts, each of them, however, belongs to a distinct discipline and specialisation.

Considering the RF technology part, there are three challenges posed: The first refers to conceiving the RF circuit topology (both at transmitter and receiver sides). This includes the use of appropriate active devices commensurate with the circuit configurations involved. The second consideration is concerned with realising effective ways of receiving the RF signal and delivering it robustly to the baseband circuits so as to recover the useful information in all its fidelity. Likewise, on the transmitter side, the baseband data from the source should be processed and delivered to the RF section for efficient wireless transmission. In either case, it should be noted that a *mixed–signal* environment prevails wherein both RF and low-frequency (baseband) signals exist in juxtaposition and have to be handled on a peer-to-peer basis.

The third consideration is the vital part of designing a way to link the RF circuits of the transmitter and the receiver by wireless means through the free-space. It is here that the “antenna” comes into the picture. The antenna represents a structure of specific type and geometry designed to mediate the transfer of EM energy from the RF circuit of the transmitter into the free-space or, reciprocally access the EM energy from the free-space and deliver it to the RF circuits of the receiver.

The concepts of “antenna” and “EM radiation” depicting the coupling of EM energy to the free-space for propagation have been studied and extensively reported in the literature pertinent to the classical systems of radio/TV broadcasting, radar and point-to-point and broadcast-mode wireless transmissions. Such studies through the past decades portray vividly the art of antenna engineering that took place concurrent to the developments in science and technology of radio/television, radar and wireless communications intended for police, military, ambulance services etc.

Thus, the underlying concepts of antennas and EM radiation as known today carry a certain level of maturity. Nevertheless, exclusive to modern wireless communications, there is a niche for a focused study on antennas. The reason can be listed as follows: Availability of limited EM spectrum, existence of variety in the information to be supported (namely, voice, data, video and multimedia), exponent growth in mobile user population, relative mobility of user end-entities, time-dependent EM-propagation ambient, outdoor and indoor service requirements, compact-sizing and portability of the transreceive units and low-

power considerations are unique attributes of modern wireless communications. These have pushed the wireless communication antenna engineering to conceive means of designing appropriate structures by identifying them compatible for various application profiles [1.17].

Hence, this book is written to provide an overview of such antenna requirements commensurate with the state-of-the-art wireless communications supported by the EM propagation ambient normally encountered. Typically, the modern wireless system, as discussed before, depicts a global, ubiquitous service network that can allow its subscribers to communicate with anyone, anywhere and anytime. Relevant systems are designed to support various applications — voice telephony, data transmissions of low- to high-bit rates, picture/video transmissions and/or multimedia traffic. In brief, the above comments, together with the particular mobile environment under consideration, will decide the appropriate antenna systems to be employed.

Further, such communication links may cover long distances (as in cellular telephony) or very short distances (as in indoor WLAN or Bluetooth systems). The information supported may belong to delay-sensitive, isochronous traffic (such as voice and video) or delay-insensitive transmissions of packet data, which can tolerate latency. The radio channels of wireless communications, in essence, are required to carry signals at bit rates from kilobits to hundreds of megabits per second. In this process, however, the wireless channels do not represent a low-bit error environment. Especially, mobility poses a time-dependent condition between the transmitter-receiver path. This changing environment, as discussed before, causes the signal to fade or change in an undesirable manner making the signal to pose random variations in its amplitude and/or phase. Robust recovery of the associated useful information from the EM wave subjected to this hostile environment depends on an efficient RF front-end plus an appropriate baseband signal-processing strategy.

The front-line of the RF technology involved at the transmitter and the receiver is the antenna structure. Its design, therefore, would dictate to a large extent a reliable wireless communication.

The motivation behind the emergence of wireless communication systems refers to the restriction placed on the mobility of end-users of wireline telephony. The cordless telephones slightly eased that restriction to allow the user to move around in a limited space, but still the base unit of the cordless phone is required to be attached to a wall-socket terminated to wireline telephony of a PSTN. Wireless communications break this barrier by allowing communication to and from a mobile node that does not require a wireline-constrained access point (such as the base unit of a cordless phone). The access point in the wireless communications is a base station, which can facilitate (establish/disconnect) wireless connectivity between roaming users.

The intervening wireline network between base stations could be the PSTN, the Internet, the ATM networks or others. Each of these follows its own private or standardised network protocols, independent of the wireless communication under discussion.

In an indoor environment, the access point may function as a gateway to translate the radio networking protocols into appropriate wireline networking protocols.

In all the aforesaid wireless communication efforts, an antenna is uniquely specified for optimal overall performance. Each application and system require distinct geometry and performance specifications imposed on the antenna structures. Large or small, compact or extensive, planar- or circular-polarised, single- or multi-band operational, outdoor or indoor specified, narrow- or broad-band oriented, concealed or prominently placed, low- or medium-power energised considerations are a few contras in the design that one may face in the design of an antenna for a specific wireless communication system.

To comprehend the associated principles of operation and pursuits of design of such antenna systems, presented in the ensuing chapters include a compendium of details, host of analyses, and a critical review of the state-of-the-art considerations.

Problem 1.5
Surveying the literature, catalogue the profiles of various antennas required for the wireless equipment of the following services:

- Land mobile systems (Automobiles and trains)
- Portable (hand-held and/or body-worn) units
- Base stations
- Maritime systems
- Airborne and spaceborne systems (Aircraft, missiles, rockets and satellites)

(*Hint*: See [1.17], pages 567-591)

1.8 CLOSURE

The purpose of the foregoing has been to introduce, partly in historical perspective, some of the underlying aspects of modern wireless communications. The contents of this chapter offer a prelude into the technological basis for the type of antenna one may choose for a specified wireless communication application (as will be elaborated in the later chapters). The rest of this book is devoted to discussions on the antenna theory and practice (plus the relevant electromagnetic principles) trimmed and tailored for wireless communication applications.

This page intentionally left blank

REFERENCES

[1.1] Rappaport, T. S., *Wireless Communications – Principles and Practice*, Upper Saddle River, NJ: Prentice Hall PTR, 1996

[1.2] van Duuren, J., Kastelein, P., and F. C. Shoute, *Fixed and Mobile Telecommunication Networks, Systems and Services*, Harlow, UK: Addison-Wesley, 1996

[1.3] Sreetharan, M., and R. Kumar, *Cellular Digital Packet Data*, Norwood, MA: Artech House, 1996

[1.4] Feher, K., *Wireless Digital Communications – Modulation and Spread Spectrum Applications*, Upper Saddle River, NJ: Prentice Hall PTR, 1995

[1.5] Sampei S., *Applications of Digital Wireless Technologies to Global Wireless Communications*, Upper Saddle River, NJ: Prentice Hall PTR, 1997

[1.6] Garg, V. K., and J. E. Wilkes, *Wireless and Personal Communications Systems*, Upper Saddle River, NJ: Prentice Hall PTR, 1996

[1.7] Garg, V. K., Smolik, K., and J. E. Wilkes, *Application of CDMA in Wireless/Personal Communications,* Upper Saddle River, NJ: Prentice Hall PTR, 1997

[1.8] Goodman, D. J., *Wireless Personal Communications Systems*, Reading, MA: Addison-Wesley, 1997

[1.9] Muller, N. J., *Wireless Data Networking*, Boston, MA: Artech House Publishers, 1995

[1.10] Davis, P. T., and C. R. McGuffin, *Wireless Local Area Networks*, Boston, New York, NY: McGraw-Hill Book Co., 1994

[1.11] Bray J., and C.F. Sturman, *Bluetooth - Connect Without Cables*, Upper Saddle River, NJ: Prentice Hall PTR, 2001

[1.12] Muller, N. J., *Bluetooth Dymystified*, New York, NY: McGraw-Hill, 2001

[1.13] Miller B. A., *Bluetooth Revealed*, Upper Saddle River, NJ: Prentice Hall PTR, 2001

[1.14] Neelakanta, P. S., *A Textbook on ATM Telecommunications*, Boca Raton, FL: CRC Press, 2000

[1.15] Logton, T., *Mobile Communication Satellites – Theory and Applications*, New York, NY: McGraw-Hill, 1995

[1.16] Siwiak K., *Radiowave Propagation and Antennas for Personal Communications*, Boston, MA: Artech House, 1995

[1.17] Fujimoto, K., and J. R. James, eds., *Mobile Antenna Systems Handbook*, Bodton, MA: Artech House, 2001

[1.18] Chang, K., *RF and Microwave Wireless Systems*, New York, NY: John Wiley & Sons, Inc., 2000

[1.19] Tranter, W. H., and B. D. Woerner, J. H. Reed, T. S. Rappaport, and M. Robert, eds., *Wireless Personal Communications Bluetooth and Other Technologies*, Boston, MA: Kluwer Academic Publishers, 2001

[1.20] Pahlavan, K., and P. Krishnamoorthy, *Principles of Wireless Networks*, Upper Saddle River, NJ: Prentice Hall PTR, 2002

CHAPTER 2

Electromagnetic Fields and Waves

2.1 INTRODUCTION

The technology of wireless communications is built on the principles of electromagnetic fields and waves. It is facilitated by the underlying concepts of electromagnetic (EM) radiation, which is mediated by antennas and associated wave propagation. As such, in order to understand the principles of antennas and EM radiation pertinent to wireless communication services, it is useful to review the fundamentals of electromagnetic theory and the associated radiation mechanism. This would enable grasping a working knowledge on the principles of antenna structures adopted in wireless technology. Therefore, presented in the section ahead is a treatise on the basic concepts of electromagnetic fields and waves. In the subsequent sections, relevant concepts of electromagnetic theory are elaborated [2.1 - 2.11].

2.2 CONCEPTS OF ELECTROMAGNETISM

Experimental studies and observations in the science of electricity and magnetism have led to postulate a number of laws pertinent to the associated phenomenological aspects perceived as electric and/or magnetic field forces. Relevant definitions, concepts and analytical considerations are summarised in this section and elaborated in the later sections[*]. The SI units are adopted throughout unless specified otherwise.

2.2.1 Electric field

An *electric field force* is perceived in the presence of *electric charges* in a physical domain, normally designated as a *medium*. The electric charge is a physical attribution to electrons and protons depicting their electrical property. Denoted by e, the electronic charge is negative (– e) and the protonic charge is positive (+ e). The magnitude of electronic charge, |e| is equal to 1.6×10^{-19} coulomb.

A set of electric charges $\{Q_n\}$ is always quatitised and prevails as a collection of positive or negative integral multiples of the electronic charge. That is, $Q_n = \pm$ ne, where, n = 1, 2, 3..., etc. and this is known as *law of charge quantitisation*. Another important property of electric charge is that like-charges repel each other and unlike-charges attract each other. Further, electric charges cannot be created nor destroyed in conformance with the *law of conservation of charge*.

---

[*]*Note*: This chapter can be skipped if the reader is familiar (or unfamiliar!) with electromagnetic theory.

An experimental postulation quantifying the force of attraction or repulsion between charges refers to the *Coulomb's force law*. Given two point-charges separated by a distance r in a medium, the force (of attraction or repulsion) between them is given by, $\mathbf{F} = (1/4\pi\varepsilon) \times Q_1 \times Q_2 \times \mathbf{a}_r/r^2$ newton. Here, $\mathbf{a}_r$ is the unit vector along r and ε depicts the property of the medium and is known as the *permittivity*. Quantitatively, $\varepsilon = \varepsilon_o\varepsilon_r$ with $\varepsilon_o = (1/36\pi) \times 10^{-19}$ farad/m representing the *absolute permittivity of free-space* and ε_r refers to the *relative permittivity* of the medium, also known as the *dielectric constant*.

Electric field intensity

Pertinent to an electric force field perceived in a medium due to the presence of charges, a measure of the *field strength* depicting the intensity of this force field (at a point) can be quantitatively specified by *electric field intensity* (**E**). It denotes the force exerted on a unit charge at a point due to the presence of another charge, Q (or a collection of charges).

Using Coulomb's law, $\mathbf{E} = Q\mathbf{a}_r/4\pi\varepsilon r^2$ newton/coulomb (or volt/m), where $\mathbf{a}_r$ is, again the unit vector along r. Due to a system of N point charges (Q_1, Q_2, ..., Q_n, …, Q_N) located at vector distances $|\mathbf{r} - \mathbf{r}_n|$, n = 1, 2, ..., N, the electric field intensity, $\mathbf{E}_P(\mathbf{r})$ at a location P situated at a distance r from the origin, is equal to $(1/4\pi\varepsilon)\sum_n Q_n\mathbf{a}_{r_n} / |\mathbf{r} - \mathbf{r}_n|^2$ with $\mathbf{a}_{r_n}$ being the unit vector along r_n (and r_n is the distance of the discrete charge Q_n from the origin).

Electric flux and electric displacement

The force field due to electric charges can be represented by a set of *flux lines* (per unit area) depicting an *electric flux density* given by $\mathbf{D} = \varepsilon\mathbf{E}$ C/m^2. The vector **D** is also known as *electric displacement*.

Considering an elemental electric flux (dψ) streaming through an (elemental) area dS in a given medium, it can be obtained in terms of the dot product, **D.**d**S**. Hence, $\psi = \iint_S \mathbf{D}.d\mathbf{S}$ denotes the total flux across the entire area S. Further, the total flux out of any closed surface is equal to the total charge enclosed in the volume (v) within the closed surface. That is, $\oiint_S \mathbf{D}.d\mathbf{S} = Q_{enclosed}$ with the double integration performed over the closed surface area. This is known as *Gauss' law* and the total charge enclosed within the volume is specified by $Q_{enclosed} = \iiint_v \rho_v dv$ where ρ_v (also alternatively denoted by q_v) is called *volume charge density* that has the unit C/m^3. The triple integration governs the volume enclosed by the closed surface.

Electrical potential

In an electric force field, work is exerted in moving a charge against the force (of repulsion). Quantitatively, the work done in moving a unit positive charge (against the force of repulsion), from an initial position at infinity to a point of observation

at r located in the electric field is equal to $-\int_{\infty}^{r} \mathbf{E}.d\mathbf{r}$ with the negative sign implying that the work done is against the force of repulsion. This work performed or energy expended on a unit charge is known as the *absolute electric potential* at that point. Its unit is joule/coulomb or volt. In moving a unit charge from r_1 to r_2, the corresponding potential difference is given by $-\int_{r_1}^{r_2} \mathbf{E}.d\mathbf{r}$. The work field $V = -\int \mathbf{E}.d\mathbf{r}$ can also be written in terms of a differential operation, namely, $\mathbf{E} = -\nabla V$ where ∇ is the *differential gradient operator*, which specifies a maximum spatial derivative of a *scalar point-function* V representing the *electric potential* at that point. That is, $\nabla V \Rightarrow (dV/dr)_{max}$. Furthermore, considering the integral $-\oint(\nabla V).d\mathbf{r}$ over a closed path, it is always equal to zero as a result of the conservation property of energy. In circuit theory, this condition is stipulated by the *Kirchhoff's voltage law*, namely, the algebraic sum of electric potentials around a closed loop is equal to zero.

Charge distribution

Electric charges can exist in the following forms: (i) Discrete, point charges; (ii) a cloud of bulk (volume) charges; (iii) charges distributed along a line (linear charges); and, (iv) charges distributed over a surface area (surface charges). The volumetric charge density is denoted by ρ_v coulomb/m^3, linear charge density is ρ_ℓ coulomb/m and surface charge density is ρ_s coulomb/m^2. (Note: In this text, q and ρ may be interchangeably used.)

Boundary conditions for electric field

At a dielectric-dielectric interface the tangential components of an electric field across the interface of two dielectric media are continuous regardless of the presence or absence of surface charges at the boundary.

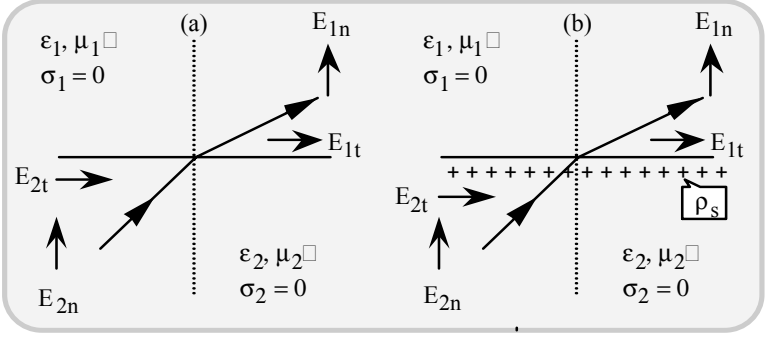


Fig. 2.1 Boundary conditions of electric field components at the interface between two dielectric media: (a) Charge-free interface and (b) an interface that has a surface charge, ρ_s

That is, $E_{t1} = E_{t2}$ in Figure 2.1. Further, the normal components of electric flux density across a charge-free interface of two dielectric media are also continuous. That is, $(D_{n1} = \varepsilon_1 E_{n1}) = (D_{n2} = \varepsilon_2 E_{n2})$ in Figure 2.1(a). Should surface charges of density ρ_s coulomb/m^2 exist at the interface, then, $D_{n1} - D_{n2} = \rho_s$ (Figure 2.1(b); however, the tangential components are unaffected or remain continuous as indicated above.

Divergence of electric flux

The net electric flux through a charge-free, closed volume is zero. That is, in view of Gauss' law, the total flux entering a closed volume of a charge-free medium will always be the same as that emerges out of the volume. Or, $\oiint_S \mathbf{D}.d\mathbf{S} = 0$. Alternatively, this is specified mathematically by a *divergence operation* leading to a *solenoidal condition* at a point, namely, $\nabla\cdot\mathbf{D} = 0$. In the limiting condition of the volume shrinking to zero (depicting a point) where the charge density is ρ_v, then $\nabla\cdot\mathbf{D}$ represents the net flux through this volume and is equal to ρ_v. This relation, namely $\nabla\cdot\mathbf{D} = \rho_v$ is the differential form of Gauss' law and is known as the *microscopic Gauss' law*.

Poisson's and Laplace's equations

In reference to a medium where charges are present, $\nabla\cdot\mathbf{D} = \rho_v$; and, since $\mathbf{D} = \varepsilon\mathbf{E}$ with $\mathbf{E} = -\nabla V$, it follows that $\nabla\cdot(\nabla V) = -\rho_v/\varepsilon$. The operation $\nabla\cdot\nabla(\cdot)$ refers to taking the divergence of a gradient of a scalar and it is known as *the Laplacian operator* denoted by ∇^2 $(= \nabla\cdot\nabla)$. Hence, $\nabla^2 V = -\rho_v/\varepsilon$, which is known as *Poisson's equation*. If $\rho_v = 0$, $\nabla^2 V \equiv 0$, which is called *Laplace's equation*. Solutions of Poisson's or Laplace's equations with appropriate boundary conditions of the region under consideration lead to explicit determination of the electric potential function V at any point in the medium.

2.2.2 Electric current

In the context of electricity, there are two electric currents defined: The first one refers to the *conduction current* defined as the time rate of change of charges. That is, the conduction current I_c is equal to dQ/dt coulomb/s and it is designated by the unit *ampere*. Conduction current could be constituted by the flow of free electrons in a conductor, by free carriers (electrons and holes) in a semiconductor or by a flow of ions in a medium like electrolyte or plasma. The corresponding *conduction current density* $\mathbf{J}_c$, is defined as the vector depicting the flow of (conduction) current normal to the cross-sectioned area and it has the unit ampere/m^2. Further, $\mathbf{J}_c$ depicts a vector point-function and is directly proportional to the electric force field (at that point), which causes the charges to flow due to the coulombic force exerted on them. Or, $\mathbf{J}_c = \sigma\mathbf{E}$ where σ is a material property known as the

electrical conductivity with the unit siemen/m. The reciprocal of σ, namely, $1/\sigma = \rho_R$ (ohm-metre) is defined as the *resistivity* of the medium.

The second version of electric current is known as the *displacement current*, which will be described later in this section.

Ohm's law

The relation depicting the conduction current density and the electric field intensity, namely $\mathbf{J}_c = \sigma\mathbf{E}$ indicated above is known as *microscopic Ohm's law*. For a bulk material of length ℓ and an area of cross-section A, it follows from the microscopic Ohm's law that $I_c = J_cA = V/R$, where V represents the potential difference across the length and, R is called the *resistance* of the bulk material. It has the unit volt/ampere or ohm. Further, $R = \ell/\sigma A = \rho_R\ell/A = V/I_c$. Thus, the conduction current is directly proportional to the applied voltage (V) with 1/R representing the constant of proportionality. This Ohm's law is valid under the condition that the ambient temperature is invariant. The reason is explained below.

Dependence of electrical conductivity on temperature

The material property of electrical conductivity σ is decided by the statistics of charge movement within the material as a result of externally applied voltage. Quantitatively, it is given by $ne^2\tau/m_e$ where n is the number of electrons under motion, e is the electronic charge, m_e is the electronic mass and τ represents the *mean collision time* of moving electrons. (The term τ is also known as the *relaxation time*.) Should temperature change, the mean collision time will be affected due to thermodynamic energy acquired. Hence, the values of σ, ρ_R and R would depend on temperature. Therefore, Ohm's law is stated validly for a specified temperature only.

Perfect conductors, metals, semiconductors and insulators

When $\sigma \to \infty$, it refers to the material being a *perfect conductor*. However, the perfect conductor is only conceptual. In practice, the value of electric conductivity depends on the material. For metals like, copper, σ is very high (5.7×10^7 siemen/m). For pure (intrinsic) semiconductors such as germanium, σ is low (2.3 siemen/m). For ideal insulators (or *perfect dielectrics*), $\sigma = 0$. However, for practical insulators like Bakelite™, $\sigma \cong 10^{-9}$ siemen/m. The flow of ions can also contribute electric conductivity. For example, the conductivity of salt water is approximately 4 siemen/m at room temperature.

Boundary conditions at a conductor-dielectric interface

Electric field (**E**) within a perfect conductor is equal to zero. That is, $\mathbf{E} = 0$ in a material with $\sigma = \infty$ since there is no potential gradient possible (inasmuch as $\nabla V = 0$ as a result of a perfect conductor posing zero resistivity).

The absence of a potential gradient (in a perfect conductor) implies that V is a constant throughout, making the perfect conductor an *equipotential region*. As a

result, the electric charges tend to reside on the surface of a conductor. This is illustrated in Figure 2.2.

The electric flux lines per unit area (**D**) emanating from the surface of a conductor are identically equal to the charge density ρ_s, on the surface. That is, $|\mathbf{D}|_{\text{normal}} = \rho_s$. Since a perfect conductor has an equipotential surface, there is no voltage drop tangential to the surface. As such, $|\mathbf{E}|_{\text{tangential}} = 0$ and any electric flux lines at the surface of a perfect conductor would terminate normal to the surface upon the surface charges (Figure 2.2). If a tangential component of **E** exists at the surface of a conductor, it should be supported by a tangential current density. This is feasible only when the conductor is of finite conductivity. (Otherwise for a perfect conductor, $\sigma \to \infty$, $E_t \to 0$ since there cannot be a potential gradient on the equipotential surface of the perfect conductor, as mentioned above.)

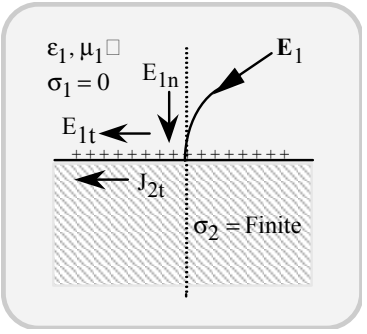


Fig. 2.2 Boundary conditions of electric field components at the interface between a dielectric and a conductor

Since $\mathbf{E} = 0$ in the interior of a conductor, $\nabla \cdot \mathbf{D} = 0$ and, hence $\nabla \cdot \mathbf{J}_c = 0$. This condition, namely, the divergence of **D** equal to zero specifies that the interior of a metallic body is point-wise neutral. Further, $\nabla \cdot \mathbf{J}_c = 0$ depicts the *solenoidal property* of the conduction current. That is, I_c has no source or sink and the current entering a point should be equal the current leaving that point depicting the *continuity of current* concept. In circuit theory point of view, it represents the *Kirchhoff's current law*, namely, the algebraic sum of currents at a node is equal to zero.

Refraction of current at the boundary of two imperfect conductors

In reference to Figure 2.3, since $\mathbf{J}_c = \sigma\mathbf{E}$, the normal and tangential currents at the boundary between two imperfect conductors are as follows: $(E_{t1} = E_{t2}) \equiv (J_{ct1}/\sigma_1 = J_{ct2}/\sigma_2)$, and $(J_{cn1} = J_{cn2}) \equiv (\sigma_1 E_{n1} = \sigma_2 E_{n2})$. Therefore, $\tan(\theta_1)/\tan(\theta_2) = \sigma_1/\sigma_2$, which depicts the *current refraction condition* at the interface between two imperfect conductors.

Further, considering the relations, ($\varepsilon_1 E_{n1} = \varepsilon_2 E_{n2}$), ($E_{n1} = J_{cn1}/\sigma_1$), ($E_{n2} = J_{cn2}/\sigma_2$) and the current continuity condition, namely, ($J_{cn1} = J_{cn2} \equiv J_n$), it follows that $J_n[(\varepsilon_1/\sigma_1) - (\varepsilon_2/\sigma_2)]$ is finite and equal to a charge density ρ_s C/m^2 that would appear at the interface. In other words, there is an inevitable induction of a surface charge, ρ_s whenever there is a refraction of current at the boundary of two imperfect conductors.

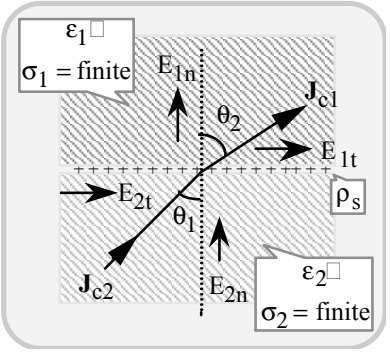


Fig. 2.3 Boundary conditions of electric field components across the interface between two imperfect conductors

Laplace's equation for conduction media

Considering a conducting medium, it can be observed that, as a result of solenoidal condition, namely, $\nabla \cdot \mathbf{J}_c = 0$ and the equipotential state specified by $\nabla V = 0$, it follows that $\nabla(\sigma \mathbf{E}) = \nabla \cdot (\nabla V) = \nabla^2 V = 0$. That is, the second-order partial differential equation (the Laplace's equation), namely ($\nabla^2 V = 0$), defines V at any point in the conductor (specified in a given co-ordinate system). To solve ($\nabla^2 V = 0$), the general steps involved are as follows:

- Specify the geometry of the region/medium in which the scalar potential is to be ascertained (*via* the solution of the Laplace's equation), and the associated boundaries in an appropriate co-ordinate system
- Stipulate the forced boundary conditions pertinent to the electric field at the boundaries involved. There are three possible forced boundary conditions: (i) Voltage specified all over the boundaries (*Dirichlet's* condition); or (ii) gradient of the voltage prescribed all over the boundaries (*Neumann* condition); or, (iii) the *mixed* (Dirichlet and Neumann) conditions specified in sections over the boundaries

- Also, stipulate the necessary natural boundary conditions, namely, $V \rightarrow 0$ at infinity in the geometry conceived and the continuity (or discontinuity) of **E**-field components at the boundaries, as appropriate. Hence, $(\nabla^2 V = 0)$ can be solved as a boundary-value problem in the co-ordinate system accommodating the geometry of the region under consideration (for single, two- or three-dimensional configurations).

2.2.3 Dielectric media

In microscopic characterisation of dielectric materials, the neutral molecules form a set of equal positive and negative charges separated by a distance constituting electric dipoles; these dipoles align or orient themselves along the electric force field applied on the material. This state of alignment along the **E**-field is known as *dielectric polarization*. The dipole pairs of (+ Q) and (– Q) charges separated by a small distance d, in the medium, would experience a *dielectric moment*, $\mathbf{p} = Q\mathbf{d}$ coulomb-metre in the process of aligning along the **E**-field. The corresponding torque exerted is $\mathbf{T} = \mathbf{p} \times \mathbf{E}$.

Dielectrics may consist of two types of molecules. The first version refers to *nonpolar molecules* in which the (+ Q) and (– Q) are located at the same locale as (± Q) and the second type has (+ Q) and (– Q) separated by a distance d within a molecule forming a *polar dipole*: (+ Q) ⇔ (– Q). The dielectrics made of such polar molecules are known as *polar dielectrics*. Nonpolar molecules, which under an **E**-field get stretched, acquire a dipole moment and experience a torque. On the other hand, polar molecules have a permanent dipole moment (in contrast with nonpolar molecules, which have no dipole moment in the absence of an electric field).

Polarisation vector of a dielectric media

The dielectric polarisation is quantified in terms of *dielectric moment per unit volume* and denoted by a *polarisation vector* **P** having the unit C/m^2. Suppose a dielectric material is placed in the free-space. Subjecting this dielectric medium to an external field **E** of flux density $\varepsilon_o\mathbf{E}$, the rotating force realised as a result of the torque exerted on the dipoles of the material is given by $\mathbf{F} = \nabla(\mathbf{p}\cdot\mathbf{E})$. Hence, the work done when a dipole orients or flips itself (as a result of rotation), corresponds to a potential energy $W = -\mathbf{p}\cdot\mathbf{E}$ joules. With the polarisation set up inside a dielectric as above, the resulting total electric flux density ($\mathbf{D}_d$) within the medium is given by: $\mathbf{D}_d = \varepsilon_o\mathbf{E} + \mathbf{P}$, where **P** implicitly depicts the additional flux density "permitted" in the medium as a result of dielectric (molecular) polarisation.

If $\mathbf{D}_d$ is identically set equal to $\varepsilon_d\mathbf{E}$, then ε_d represents the permittivity of the dielectric (which is distinct from that of free-space). Then, it follows that $\varepsilon_d = \varepsilon_o + |\mathbf{P}/\mathbf{E}|$. Further, in terms of the ratio $|\mathbf{P}/\mathbf{E}| = \chi$ (known as *dielectric susceptibility*), $\varepsilon_d = \varepsilon_o(1 + \chi)$; or, $(\varepsilon_{rd} - 1) = \chi$ with $\varepsilon_d = \varepsilon_o\varepsilon_{rd}$ and ε_{rd} denotes the *relative permittivity* or dielectric constant of the dielectric material under

discussion. It can be shown that the polarised dipoles in a dielectric equivalently represent a collection of volume charges of density equal to $-\nabla.\mathbf{P}$ (presumed to be within the dielectric volume) plus a system of surface charges of density $\mathbf{P}\cdot\mathbf{n}$ on the surface enclosing the dielectric volume. (Here, **n** is the normal vector to the surface.) Therefore, the electric field exterior to the dielectric (under polarised state) at a vector distance of R from an elemental volume dv of the dielectric, can be written as,

$$\mathbf{E}(\mathbf{r}) = -\nabla V(r) = (\mathbf{a}_R/4\pi\varepsilon_o) \iint_S (\mathbf{P}.\mathbf{n})dS/R^2 + (\mathbf{a}_R/4\pi\varepsilon_o) \iiint_v (\nabla.\mathbf{P})dv/R^2 \quad (2.1)$$

where $\mathbf{a}_R$ denotes the unit vector along R and **r** is the vector distance of the point of observation from the origin.

Electric displacement current

Whenever the polarisation in a dielectric changes with time, it would amount to a time-varying set of bound charges constituting a *polarisation current*, which is specified by a current density $\mathbf{J_p} = d\mathbf{P}/dt$ amperes/m^2. Corresponding to the time-varying electric displacement vector **D**(t), perceived in the dielectric medium as a result of polarisation, the vector $d\mathbf{D}/dt$ is defined as the *displacement current density* $\mathbf{J}_d$ expressed in amperes/m^2.

Capacitance

Capacitance refers to the state of acquiring electric charges by a particular geometry of conducting objects separated by a distance and maintained at a potential difference. Considering a set of charges (of opposite nature) on a pair of conductors, (+ Q) and (– Q), if V is the potential difference observed across the conductors, then the *capacitance* C refers to a proportionality relation between Q and V specified by Q = CV. The value of C depends on the geometrical surface (area) of the conductor (A), the distance (d) between the surfaces holding the charges and the permittivity (ε) of the medium between the conductors. Explicitly, it is given by C = εA/d farads.

Capacitor is an electric passive component formed by a pair of conductors with an interposed dielectric to yield a specified value of capacitance. The electrostatic energy stored in a capacitance system is given by $W = (1/2)\,CV^2 = Q^2/2C$ joules. Expressed in terms of the electric field (**E**) associated with the capacitor system $W = (\varepsilon/2)\iiint \mathbf{E}^2 dv$, with the triple integral specified over the volume (v) constituting the capacitor geometry.

2.2.4 Magnetic field

Magnetic field is a force field perceived in a region where electric charges are time varying. Analogous to electric field, the magnetic force field can be depicted by a set of flux lines. The density of such flux lines per unit area is known as *magnetic flux density* (**B**) and its SI unit is tesla. (1 tesla = 1 weber/m^2 = 10^4 gauss in CGS

units.) Also, similar to electric field intensity, the driving force causing a magnetic flux density is known as *magnetic field intensity* (**H**) with a unit ampere-turn per metre. The vector entities **B** and **H** are related by the *constitutive relation* $\mathbf{B} = \mu\mathbf{H}$, where μ is called the *permeability* of the medium; Further, $\mu = \mu_o\mu_r$ where μ_o is the *absolute permeability of free-space* and is equal to $4\pi \times 10^{-7}$ henry per metre. The absolute permeability μ_o represents the magnetic property of free-space or the extent to which magnetic flux is set-up in free-space due to a magnetising force (caused by time-varying charges) and μ_r is known as the *relative permeability* of a medium. It denotes the magnetic characteristics of the medium relative to free-space. It is a measure of the extent of *magnetic induction* in that medium (for a given magnetising field strength).

Magnetic field postulations

The experimental observations made on the physics of magnetism have enabled the postulations of a set of laws that characterise the associated phenomenology and quantitatively describe the underlying parameters. Relevant information is summarised below.

Ampere's force law

Considering an elemental length $d\ell$ carrying a current **I**, then the product $\mathbf{I}d\ell$ is defined as a *current element* (representing otherwise equivalently a flow of an elemental charge, dQ with a velocity, **v**). Assuming that this current element is immersed in a magnetic field **B**, it will experience a force d**F** given by: $d\mathbf{F} = (\mathbf{I} \times \mathbf{B}d\ell$ or $(dQ)(\mathbf{v} \times \mathbf{B})$ newtons. The vector cross-product indicates that **F**, **I** and **B** are mutually perpendicular to each other as per the left-hand thumb rule.

Biot-Savart's law

A current element $\mathbf{I}\, d\ell$ produces a magnetic field **B,** which at a distance R from the element is driven by $d\mathbf{B} = (\mu_o/4\pi)[(\mathbf{I}) \times (\mathbf{a}_R d\ell/R^2)]$, where $\mathbf{a}_R$ is the unit vector along R and the magnitude $|d\mathbf{B}| = (\mu_o/4\pi)\,(Id\ell)\,[\sin(\theta)/R^2]$.

Lorentz force

It refers to the force on a moving charge Q with a velocity **v** in a magnetic field **B** and an electric field **E**. It is given by: $\mathbf{F} = Q\,[\mathbf{E} + (\mathbf{v} \times \mathbf{B})]$ newtons.

Magnetic moment

A current element $\mathbf{I}d\ell$ in a magnetic field **B** would experience a torque $\mathbf{T} = \mathbf{m} \times \mathbf{B}$ newton-metre (as a result of Ampere's force), where **m** is the *magnetic moment* of the current-carrying loop encircling an area A and $|\mathbf{m}| = IA$.

Suppose IA is identically equal to $(q_m d)$. Then $(+ q_m)$ can be assumed as a fictitious magnetic (isolated) charge (analogous to electric charge), separated from an opposite magnetic charge $(- q_m)$ by a distance d. The entities $(+ q_m)$ and $(- q_m)$ are said to constitute a *magnetic dipole* pair. A bar magnet of length L across its

north and south poles can be equated to a corresponding magnetic dipole having *pole strengths* of (+ q_m) and (– q_m) at the two ends of the bar. This relevant concept helps quantifying the "strength" of a permanent magnet.

Magnetic properties of matter

It is known that the electrons associated with the atoms of a material are in a state of motion. Relevant orbital motion, clockwise and anticlockwise spins and vibrational states of electrons denote the prevalence of time-varying charges in the medium of the material. As indicated earlier, a state of time-varying charges represents implicitly a current flow that can induce a flux loop of magnetic flux. If the resultant of the vectors depicting all the electron motions indicated above cancels to zero, the equivalent current flow would not prevail. Therefore, the magnetic field induced will be zero. The materials exhibiting this property are designated as *diamagnetic* materials.

In some materials known as *ferromagnetics*, the atomic or molecular arrangement permits a finite set of resultant current loops resulting from finite (non-cancelling) vectors of electron movements. They depict implicitly a set of magnetic dipoles (induced as a result of uncancelled time-varying electron states). When the random arrangement of such magnetic dipoles mutually let the associated magnetic field components cancel each other, then the result will be a *null magnetism* exhibited by (diamagnetic) materials. However, if these dipole magnets are exposed to an external source of magnetisation, then, they in turn will become directionally aligned. In the case of some materials (like steel), even after the external magnetisation is removed, some dipoles remain polarised; that is, there will be some remnant dipole orientation still sustained. Therefore, the material would pose the property of magnetism. In other words, the material is said to have acquired a "permanent magnetism".

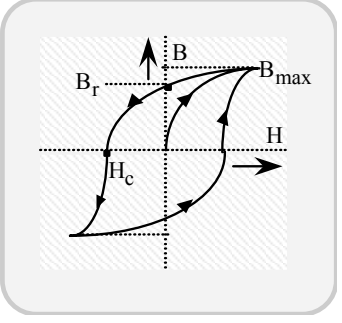


Fig. 2.4 Magnetic hysteresis curve (B_r: Remnant magnetic field and H_c: Coercive magnetic field intensity)

In reality, however, the posture of permanent magnetism is only "temporary" in the sense that, in the event of a demagnetisation influence imparted to the "permanent

magnet", the associated remnant dipole orientation will be lost leaving the material unmagnetised. (This demagnetising influence could either be a counter-magnetic force or extraneous influences such as heat or mechanical shock.)

Thus, "permanent magnetism" is just the result of remnant magnetism in a material. This remnant field can, however, be coerced to become null by applying a counter-magnetic field intensity. Further, when a ferromagnetic material is subjected to a cyclic magnetisation, the relation between B and H forms a hysteresis loop as shown in Figure 2.4. This curve also depicts the resulting *remnant magnetism* (B_r) and its removal through demagnetising or *coercive magnetic field intensity* (H_c). The ratio of B/H at any point on the hysteresis curve (Figure 2.4) is the *permeability* of the material at that state of magnetisation.

Relative permeability

For *diamagnetic materials* (such as free-space, water, copper etc.) the magnetic property is characterised with $\mu_r = 1$, or slightly less than 1; for *ferromagnetics* like iron, cobalt, or nickel as well as for *ferrites*, $\mu_r >> 1$. In some other materials (for example, titanium, aluminium), μ_r is slightly larger than one, characterising them as *paramagnetics*.

Ampere's current law

The line integral of a magnetic field **B** around any closed path (CLP) is equal to $\mu\mathbf{I}$. That is $\oint_{CLP} \mathbf{B}.\,d\,\ell = \mu\mathbf{I}$, if the closed path of integration encloses I; otherwise, the integration is zero. In terms of current density $\oint_{CLP} \mathbf{H}.d\ell = \iint_A \mathbf{J}.d\mathbf{A}$, with dA being the area bounded by the closed path.

By Stokes theorem, it follows that $\oint_{CLP} \mathbf{B}.d\ell = \iint_A (\nabla \times \mathbf{B}).d\mathbf{A}$. Hence, it follows that $\nabla \times \mathbf{B} \equiv \mu\mathbf{J}$, which represents the differential form of Ampere's current law. Considering $\oint_{CLP} \mathbf{H}.d\ell = NI$ with N being the number of turns of the current-carrying conductor, NI refers to the magnetomotive force (mmf). It is the driving influence in the induction of B (or μH) as a result of a current flowing through a loop of N turns.

Inductance

Self-inductance L is a measure of magnetic flux, which links the circuit when a current I flows in the circuit. It (like a capacitor) depends on the physical arrangement of conductors. Denoting the flux linkage as Φ, L is defined as equal to $\Phi/I = (N/I) \iint_A \mathbf{B}.d\mathbf{A}$ henries; and, the energy stored in an inductor is given by $W = (1/2)LI^2 \equiv 1/2\mu H^2$ with H representing the magnetic field intensity caused by the current in the circuit. The inductance is an electrical dual of capacitance.

Mutual inductance

It relates to the flux linkage produced in circuit # 2 as caused by the current in circuit # 1 or *vice versa*. That is, $M_{21} = d\Phi_{21}/dI_1$; or, $M_{12} = d\Phi_{12}/dI_2$.

Faraday-Lenz's law

An *electromotive force* (emf) is induced in a conductor and refers to the voltage across the terminals (of the conductor), whenever the magnetic flux (Φ) linking the conductor varies with time (*Faraday's law*). The magnitude of the emf is proportional to the time rate of change of flux linkage with the conductor; the polarity of the induced emf sets-up a current flow (in the conductor) in a direction such that the resulting magnetic field (formed as per Biot-Savart's law) will tend to oppose the cause, namely, the original magnetic flux linking the conductor. This is known as *Lenz's law*. Hence, the combined Faraday-Lenz's law specifies that emf $= - d\Phi/dt$.

2.2.5 Maxwell's equations

The Maxwell's equations are cohesive statements of the postulations of electricity and magnetism discussed in the above sections. The four equations of Maxwell's can be described as follows:

Magnetomotive force around a closed path is equal to the conduction current plus the displacement current through any surface bounded by the path,

$$\oint \mathbf{H}.d\ell = \iint(\mathbf{J}_c + \mathbf{J}_d).dA \Leftrightarrow \nabla \times \mathbf{H} = (\mathbf{J}_c + \mathbf{J}_d) \qquad (2.2)$$

where $\mathbf{J}_c = \sigma\mathbf{E}$ represents the conduction current density and $\mathbf{J}_d$ is the displacement current density equal to $\varepsilon(d\mathbf{E}/dt)$.

The electromotive force around a closed path is equal to the time derivative of the magnetic flux density through any surface bounded by the path:

$$\oint \mathbf{E}.d\ell = -\iint (d\mathbf{B}/dt).d\mathbf{A} \Leftrightarrow \nabla \times \mathbf{E} = - d\mathbf{B}/dt \qquad (2.3)$$

where $\mathbf{B} = \mu\mathbf{H}$.

The total electric flux density through the surface enclosing a volume is equal to the total charge within the volume:

$$\iint \mathbf{D}.d\mathbf{A} = \iiint q_v dv \Leftrightarrow \nabla.\mathbf{D} = q_v \qquad (2.4)$$

where $\mathbf{D} = \varepsilon\mathbf{E}$.

The net magnetic flux emanating from a closed surface is zero; that is,

$$\iint \mathbf{B}.d\mathbf{A} = 0 \Leftrightarrow \nabla.\mathbf{B} = 0 \qquad (2.5)$$

2.2.6 Wave equations

Consistent with the Maxwell's equations, which inter-relate the **E**- and **H**-field components perceived in a medium, the time-space profile (that is, the variations in time and space) of these force functions **E**(r, t) and **H**(r, t) can be written in terms of what are known as Helmoltz wave functions. These wave functions also include explicitly the constitutive relations of the medium, namely $\mathbf{D} = \varepsilon\mathbf{E}$, $\mathbf{B} = \mu\mathbf{H}$, $\mathbf{J}_c =$

$\sigma\mathbf{E}$ and $\mathbf{J}_d = \varepsilon\partial\mathbf{E}/\partial t$.

For example, in reference to an isotropic, homogeneous, linear medium (with ε, μ, σ),

$$\nabla^2\mathbf{E} - (\mu\varepsilon)[d^2\mathbf{E}/dt^2] - \mu\sigma[d\mathbf{E}/dt] - \nabla q_V/\varepsilon = 0 \quad (2.6a)$$

$$\nabla^2\mathbf{H} - (\mu\varepsilon)[d^2\mathbf{H}/dt^2] + \nabla \times \mathbf{J} = 0 \quad (2.6b)$$

where $1/(\mu\varepsilon)^{1/2}$ m/s is the velocity of propagation of the electromagnetic wave in the medium and, q_V coulomb/m^3 is the volume charge-density in the medium. Considering a charge-free medium (that is, with $q_V = 0$) and the medium being lossless ($\sigma = 0$), the wave equations simplify to:

$$\nabla^2\mathbf{E} = \mu\varepsilon\, d^2E/dt^2 \quad (2.7a)$$

$$\nabla^2\mathbf{H} = \mu\varepsilon\, d^2H/dt^2 \quad (2.7b)$$

In case of harmonic excitation, the general solutions of these equations are of the form (for example, in Cartesian co-ordinates) as follows:

$$E(x,y,z,t) = E_o\,(\gamma_x x, \gamma_y y, \gamma_z z; \omega t) \quad (2.8a)$$

$$H(x,y,z,t) = H_o\,(\gamma_x x, \gamma_y y, \gamma_z z; \omega t) \quad (2.8b)$$

where γ_x, γ_y, and γ_z are the propagation constants along x-, y-, and z-directions respectively; $\omega = 2\pi f$ and the wavelength (λ) in metres is equal to (velocity in m/s)/(frequency in Hz).

Further, in a charge-free medium of finite conductivity, the propagation constant γ specified in equations (2.8a) and (2.8b) can be identically set equal to $(\alpha + j\beta)$ where, explicitly (α, β, Z_o, and v) denote respectively, the attenuation constant, the phase constant, the characteristic impedance and the velocity of EM wave propagation in the medium. Explicitly, they are given by:

$$\left.\begin{aligned} \alpha &= \omega[(\mu\varepsilon/2)\{(1+\sigma^2/\omega^2\varepsilon^2)^{1/2} - 1\}]^{1/2} \\ \beta &= \omega[(\mu\varepsilon/2)\{(1+\sigma^2/\omega^2\varepsilon^2)^{1/2} + 1\}]^{1/2} \\ Z_o &= [j(\omega\mu/\sigma) + j\omega\varepsilon)^{1/2}] \\ v &= \omega/\beta \end{aligned}\right\} \quad (2.8c)$$

Table 2.1 EM wave parameters of different versions of dielectric and conducting media

Parameters	Media		
	Perfect dielectric ($\sigma = 0$)	Lossy dielectric ($\sigma/\omega\varepsilon << 1$)	Good (imperfect) conductor ($\sigma/\omega\varepsilon >> 1$)
α	0	$(\sigma/2)\,(\mu/\varepsilon)^{1/2}$	$(\omega\mu\sigma/2)^{1/2}$
β	$\omega(\mu\varepsilon)^{1/2}$	$\omega(\mu\varepsilon)^{1/2}\,(1 + \sigma^2/8\omega^2\varepsilon^2)$	$(\omega\mu\sigma/2)^{1/2}$
Z_o	$(\mu/\varepsilon)^{1/2}$	$(\mu/\varepsilon)^{1/2}(1 + j\sigma/2\omega\varepsilon)$	$(\omega\mu/\sigma)^{1/2}\,\angle 45^o$
v	$1/(\mu\varepsilon)^{1/2} = v_d$	$v_d[(1 + \sigma^2/8\omega^2\varepsilon^2)]^{-1}$	$(2\omega/\mu\sigma)^{1/2}$

The parameters of EM wave depend on the ratio $\sigma/\omega\varepsilon$ that can be attributed to different types of media. The extreme situations refer to: (i) a perfect dielectric, which is characterised by the limiting case of $\sigma/\omega\varepsilon \to 0$; and, (ii) a perfect conductor that is presumed to have $\sigma \to 0$, so that $\sigma/\omega\varepsilon \to \infty$.

In practice, the material media often exhibit intermediate characteristics. That is, when $\sigma/\omega\varepsilon << 1$, the material represents a lossy dielectric; and, with $\sigma/\omega\varepsilon >> 1$, the material tends to be a good (but imperfect) conductor. Taking the considerations as above, the relevant EM parameters are tabulated in Table 2.1.

Poynting vector

This is defined as electromagnetic power flow per unit area along a given direction and is specified by the vector: $\mathbf{P} = \mathbf{E} \times \mathbf{H}$ watts/m.

Plane wave

It is a transverse EM wave (*TEM wave*) with a uniform planar wavefront having the electric and magnetic field components being entirely transversal to the direction of power flow.

Reflection and refraction of plane waves

At a dielectric-dielectric interface, for normal incidence of an EM wave, the (complex) reflection coefficient is given by $R = (\eta_2 - \eta_1)/\eta_2 + \eta_1)$; the transmission coefficient $T = \eta_2/(\eta_2 + \eta_1)$, where η_1 is the intrinsic (complex) impedance of medium # 1 and, η_2 denotes the intrinsic (complex) impedance of medium # 2.

The resulting *standing wave* in the incident region has a *voltage standing wave ratio* (VSWR) $S = (1 + |R|)/(1 - |R|)$. It should be noted that R and S are bounded by $0 \leq R \leq 1$ and $1 \leq S \leq \infty$.

For oblique incidence, there are two cases of polarisations (as illustrated in Figures 2.5(a) and 2.5(b) and the corresponding reflection coefficients ($R_{||}$) and ($R_{\perp}$) are specified as follows:

Case (i): **E** vector lies in the plane of incidence (parallel polarisation) as in Figure 2.5(a):

$$R_{||} = [\eta_2 \cos(\theta_t) - \eta_1 \cos(\theta_i)]/[\eta_2 \cos(\theta_t) + \eta_1 \cos(\theta_i)] \qquad (2.9a)$$

Case (ii): **E** vector is normal to the plane of incidence (perpendicular polarisation) as in Figure 2.5(b):

$$R_{\perp} = [\eta_2 \cos(\theta_i) - \eta_1 \cos(\theta_t)]/[\eta_2 \cos(\theta_i) + \eta_2 \cos(\theta_t)] \qquad (2.9b)$$

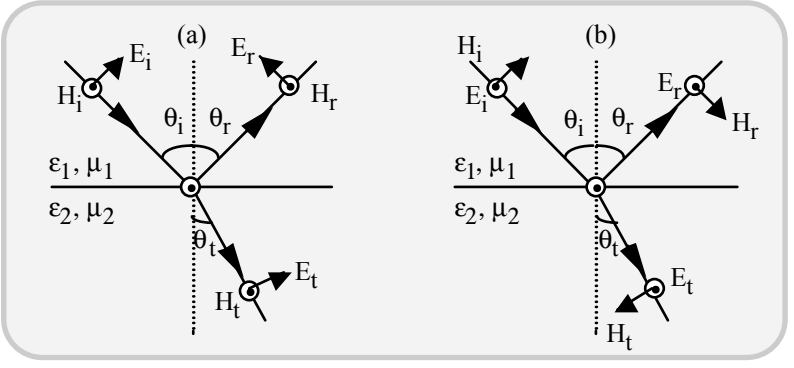


Fig. 2.5 Oblique incidence of EM wave at a dielectric-dielectric interface: Reflection and refraction as per *Snell's law*, $\theta_i = \theta_r$; $\sin(\theta_i)/\sin(\theta_t) = (\varepsilon_2/\varepsilon_1)^{1/2}$: (a) Parallel polarisation and (b) perpendicular polarisation

Brewster phenomenon

This refers to the condition of no reflection prevailing in the case of parallel polarisation, that is, $R_{||} = 0$. The corresponding angle of incidence, known as the *Brewster angle (polarising angle)* is given by,

$$\theta_B = \arcsin[\varepsilon_2/(\varepsilon_1 + \varepsilon_2)]^{1/2} = \arctan[\varepsilon_2/\varepsilon_1]^{1/2} \qquad (2.10)$$

Total internal reflection

With reference to both parallel and perpendicular polarisations, there is a condition for which the incident wave could be totally reflected back into the incident medium (with no transmission into the other medium). This *total internal reflection* happens when the permittivity (ε_1) of the medium of the incident wave is larger than the permittivity (ε_2) of the other medium. That is, when $\varepsilon_1 > \varepsilon_2$ in Figure 2.6, the total internal reflection would occur at a *critical angle* of incidence given by: $\theta_c = \arcsin[(\varepsilon_2/\varepsilon_1)]^{1/2}$.

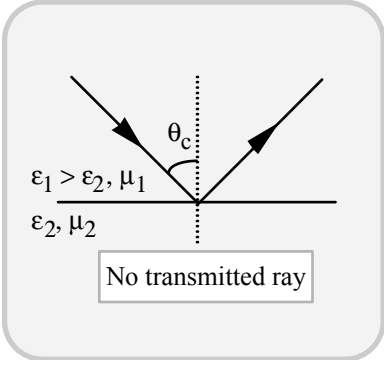


Fig. 2.6 Total internal reflection phenomenon. (θ_c is the critical angle of incidence). (*Note*: The total internal reflection can, under the condition, prevail for both parallel and perpendicular polarisations)

Wave polarisations

The orientations of the field vectors in the individual plane waves and their resultant are described by the wave polarisation. Suppose the E-field has two components, E_x and E_y, along x- and y-directions respectively. Let ψ be the phase difference between them. If $\psi = 0$, these components add at every plane along the direction of propagation, namely the z-direction in Figure 2.7.

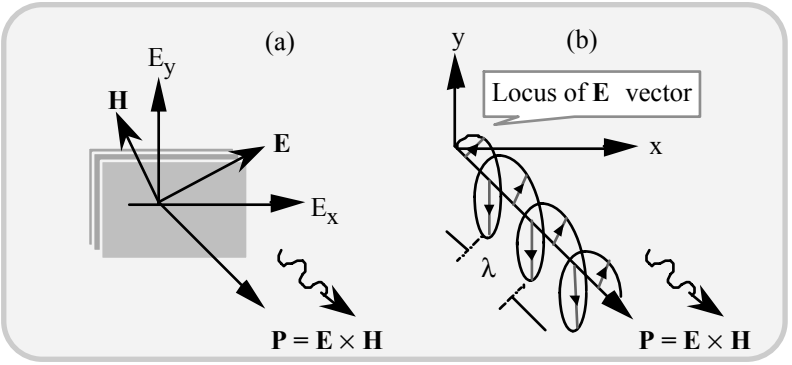


Fig. 2.7 EM wave polarisations. (a) Linear polarisation and (b) circular polarisation

The resultant of E_x and E_y is E, which maintains its constant direction in space at any specified plane as the wave propagates (along the z-direction). If **E** is vertical, the wave is said to be vertically (plane) polarised. That is, the wave is polarised with the electric field lying in the vertical plane. Likewise, a wave with **E** being horizontal is termed as horizontally (plane) polarised. Thus, the linear or plane polarisation may refer to either vertical or horizontal (azimuthal). In antenna practice, a vertically polarised wave has its **E**-field normal (vertical) to the ground-plane and a horizontally polarised wave has its **E**-field parallel (tangential) to the ground-plane.

The other type of wave polarisation is known as circular/elliptical polarisation as illustrated in Figure 2.7(b). If the two components of the electric field, namely E_x and E_y, have a phase difference $\psi = \pm \pi/2$, the instantaneous angle of the resultant vector of E_x and E_y rotates with a constant angular velocity, $(\mp\omega t)$. That is, the resultant vector **E** traces out a spiral or a cork-screw movement along the z-direction. If $E_x = E_y$, then this polarisation refers to circular polarisation; and, if $E_x \neq E_y$, the wave is said to be elliptically polarised. Further, when $\psi = +\pi/2$, the sense of polarisation is left-handed. Otherwise it is right-handed.

Thus, as regard to circular/elliptical polarisation shown in Figure 2.7(b), the tip of the **E** vector traces a circle (or an ellipse) as the wave progresses with time along the z-direction. Considering the sense of rotation being either right-handed or left-handed, the polarisation is classified as right-handed circular-polarisation (RHCP) or left-handed circular polarisation (LHCP) respectively.

Example 2.1
Determine the loss per kilometre for a plane EM wave propagating in a wet soil at f = 1 MHz.

Solution

For wet soil, approximately $\sigma = 10^{-4}$ S/m and $\varepsilon_r = 7$. Therefore, $\sigma/\omega\varepsilon = 0.257$ (< 1). Hence, from Table 2.1, $\alpha \approx (\sigma/2)(\mu/\varepsilon)^{1/2} = 1.37 \times 10^{-2}$ neper/m. Therefore it follows that, $\exp(-\alpha\ell) = 1.2 \times 10^{-6}$ for $\ell = 1$ km. Or, loss per km is given by:
$-20 \log_{10}(1.2 \times 10^{-6})$ dB/km = 118.64 dB/km (This attenuation faced by the EM wave at this low-end radio frequency (RF) is considered as a result of the power loss caused by the finite conductivity (and hence finite resistivity) of the medium. Thus, the attenuation in a material depends on the relative extent of the ratio $\sigma/\omega\varepsilon$, as specified in the formulas of Table 2.1. The reader may recalculate the attenuation in the wet soil of a plane EM wave in terms of $(\sigma/\omega\varepsilon)$ factor at a different frequency (say, f = 100 MHz) and compare it with the answer obtained for 1 MHz).

Problem 2.1

Calculate the power loss for a plane EM wave at 1.8 GHz through a buttressed concrete-filled tunnel wall of average thickness equal to 2 metres. Assume a dissipation factor of the concrete medium as 0.01 and $\varepsilon_r = 3$.

2.2.7 Transmission-line theory

A transmission line is used for the transport of an electrical/optical signal using a physical structure, which "guides" the flow of EM energy (or the electrical signal) from the generator to a load.

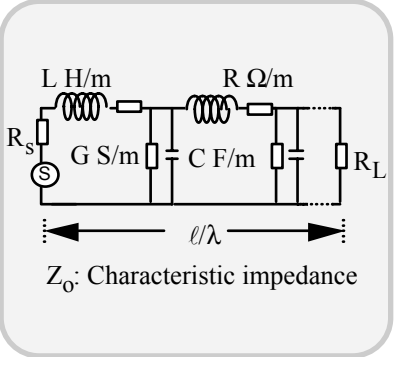


Fig. 2.8 Iterated network equivalence of a transmission line having distributed line parameters R, G, L and C

A uniform line that supports the transmission of electrical signals is characterised by the invariancy in its dimensions and/or electrical parameters along its length.

Such a line can be modelled as an iterated network of distributed (per unit length) parameters, namely R, G, L and C as shown in Figure 2.8.

A transmission line, in general, can be characterised by the attributes listed below:

* A lossless line refers to the conditions that R = 0 and G = ∞. It denotes, therefore, a pure, nondissipative iterated network of distributed LC parameters.

* A transmission line supports an EM wave in transporting signal energy. That is, the alternating (time-varying) signal on the line represents a time and distance dependency of voltage V(z, t) and current I(z, t) along a lossless line as described by the following *transmission line equations*:

$$d^2V/dz^2 = LC\, d^2V/dt^2 \quad (2.11a)$$

$$d^2I/dz^2 = LC\, d^2I/dt^2 \quad (2.11b)$$

* The solution of the differential equations as above are given by:

$$V = V^i(t - z/v) + V^r(t + z/v) \quad (2.12a)$$

$$I = I^i(t - z/v) + I^r(t + z/v) \quad (2.12b)$$

where the associated (unknown) coefficients can be determined by applying appropriate boundary and initial conditions. The boundary represents the line terminals at the source and load ends. Further, v is the velocity of propagation of the guided wave along the line and is equal to $1/(LC)^{1/2}$ m/s.

* The *characteristic impedance* (Z_o) of the line is defined as the ratio of voltage and current of a wave on the line at any point and $Z_o = (L/C)^{1/2} = R_o$ ohm (for a lossless line).

* Reflection at the termination of a line with a load $Z_L \neq Z_o$ is specified by a coefficient, $\Gamma = (Z_L - R_o)/(Z_L + R_o)$, which in general, could be complex. Hence, the complex load Z_L can be determined in terms of the complex reflection coefficient as $Z_o(1 + \Gamma)/(1 - \Gamma)$.

* When $R_L = R_o$, it refers to a flat-line and there is no reflection. It specifies the condition for impedance matching. In reference to a flat line, the input impedance Z_{in}, is measured at any distance ℓ from the load-end given by: $Z_{in} = Z_o$.

* A short-circuit line with $Z_L = 0$ renders the input impedance Z_{in} at any distance □ from the load-end and is given by: $Z_{in} = jZ_o\tan(2\pi\ell/\lambda)$, where λ depicts the wavelength of EM energy exciting the line. It is given by: λ (in

metres) = c/f with $c = 3 \times 10^6$ m/s and f is the frequency of the exciting signal in Hz. Further, $2\pi/\lambda = \beta$ represents the *phase constant* of the wave.

* An open-circuit line with $Z_L = \infty$ renders the input impedance Z_{in} at any distance ℓ from the load-end and is given by: $Z_{in} = -jZ_o \cot(2\pi\ell/\lambda)$.

* The *voltage standing wave ratio* (VSWR) resulting from the mismatch conditions is given by $S = (1 + |\Gamma|)/(1 - |\Gamma|)$ and $|\Gamma| = (S - 1)/(S + 1)$. It can be noted that $|\Gamma|$ and S are bounded by $0 \le |\Gamma| \le 1$ and $1 \le S \le \infty$.

* *Impedance matching*: In the event of a load being complex and not matched to a line on which it is terminated, the impedance-matching to be pursued refers to the following. It should facilitate a *conjugate impedance-matching*, meaning that a matching network should be interposed between the load and the line; and, the output impedance of this matching network should correspond to the conjugate of the load impedance as shown in Figure 2.9(a). Further, the input impedance of the network should be equal to the characteristic impedance of the line. At radio frequencies, there are two main options in obtaining a matching network, namely, by means of *stubs* or *via transformers.*

* *Single-stub matching* is a technique of placing a short-circuited stub on the line ahead of the terminal (where the complex load to be matched is connected). Proper choice of the location (at d_1) and the length (d_2) of this stub facilitate the required matching.

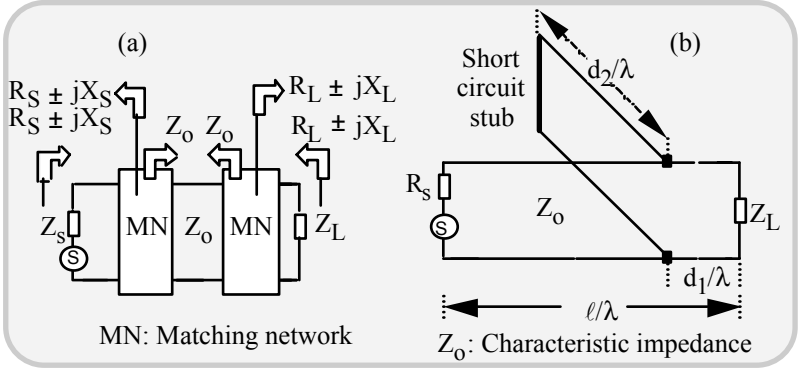


Fig. 2.9(a) Conjugate matching principle for maximum power transfer and (b) single-stub matching technique

Relevant design formulations in reference to Figure 2.8(b) are as follows: $\tan(\beta d_1) = (S)^{1/2}$ and $\tan(\beta d_2) = [S/(S - 1)]^{1/2}$.Two stubs separated by a fixed distance (normally of $\lambda/4$) and of designed lengths can also be used (instead of

a single stub). In this design, there is no restriction on the exact location where the stub structure should be placed. This technique is known as *double-stub matching*.

It should be noted that the matching *via* stubs indicated above is essentially prescribed for a single frequency. However, a broadband matching can also be done with additional changes on the line. Further, the line, in general, can be lossy. In such cases, again necessary changes in the design are warranted to achieve a desired matching performance. For details one may refer to [2.12, 2.13].

* *Half-wave* ($\lambda/2$) and *quarter-wave* ($\lambda/4$) *transformers*: A transformer in transmission line systems denotes a section of a line whose length can be of $\lambda/2$ or $\lambda/4$ (or integral multiples of them) and it can be used for impedance transformation along the line.

$\lambda/2$ transformer: This is a transmission line of $\lambda/2$ length and can be used as an 1-to-1 impedance transformer.

$\lambda/4$ transformer: A transmission line of $\lambda/4$ length with a characteristic impedance $Z_{To} = (Z_o Z_R)^{1/2}$ can function as a matching section between Z_R and Z_o, with Z_R being real. Again, it should be noted that the aforesaid transformer functions are specified at a single frequency.

Smith-chart: This is a graphical representation in the complex plane of the function, $|\Gamma| \angle(\phi \angle 2\beta\lambda = [-2/z_{in} + 1] + 1$, where $z_{in} = Z_{in}/Z_o$. It is useful in the design and calculations pertinent to transmission lines.

* *Guided wave propagation*: Transmission lines support guided electromagnetic wave propagations. Single-wire/ground-return system (as in telegraphic transmissions), two-wire lines of twisted/untwisted and shielded or unshielded versions (used in telephony), coaxial lines of different types, waveguides, microstrip lines, image-lines and fibres represent the gamut of transmission-line media adopted in telecommunications.

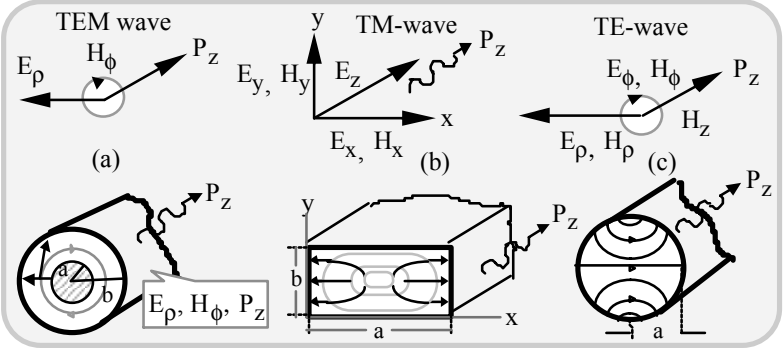


Fig. 2.10 TEM, TE-and TM-waves and examples of guided wave structures supporting these waves

Typically, a single-wire, two-wire parallel lines and coaxial lines support TEM mode where the electric and magnetic fields are mutually perpendicular to the direction of power flow. Waveguides (rectangular or circular type) support TE or TM (or mixed TE + TM) modes as illustrated in Figures 2.10. (TEM mode, however, cannot be supported on such lines because of incompatible boundary conditions.)

* *Parameters of transmission lines:* Presented in Table 2.2 are relevant parameters of typical transmission lines used in practice. In RF practice, coaxial lines, waveguides (rectangular and circular) and microstrip lines are popular transmission lines. The antennas are connected into (or supplied from) the RF section *via* these transmission lines using appropriate connectors.

Table 2.2 Parameters of various transmission lines: Coaxial lines and waveguides

Transmission line type	Parameters
Coaxial lines	C F/m = $2\pi\varepsilon/[\ell n\ (b/a)]$ L H/m = $(\mu/2\pi)[\ell n\ (b/a)]$ Z_o ohm = $(\mu/\varepsilon)^{1/2}$ v m/s = $1/(LC)^{1/2} = 1/(\mu\varepsilon)^{1/2}$
Waveguides	
Rectangular	$\beta = \omega\ (\mu\varepsilon)^{1/2}\ (1 - \omega_c/\omega^2)^{1/2}$ $\omega_c^2\mu\varepsilon = (m\pi/a)^2 + (n\pi/b)^2$, m, n = 0, 1, 2, 3, ... $\omega_c = 2\pi \times f_c$, f_c: Cut-off frequency v_p: phase velocity = ω/β λ_g : guide-wavelength = $\lambda_o/(1 - \omega_c^2/\omega^2)^{1/2}$ $\lambda_o = c/f$ metres TE_{mn}- wave (H -wave with $E_z = 0$) Dominant mode $TE_{10} \Rightarrow m = 1,\ n = 0$ (m and n are mode numbers denoting half-cycle variations of the fields in x- and y-directions respectively)
Circular	TM_{mn}- wave (E -wave with $H_z = 0$) Dominant mode $TM_{11} \Rightarrow m = 1,\ n = 1$ (m and n are mode numbers denoting half-cycle variations of the fields in ρ- and ϕ-directions respectively)

	TE-mode $\Rightarrow E_Z = 0$ TM-mode $\Rightarrow H_Z = 0$ For the dominant mode (TM_{11}): $\beta_{11} = (\omega^2\mu\varepsilon - 1.84)^{1/2}$ $f_{c(11)} = 1.84/2\pi a(\mu\varepsilon)^{1/2}$ $\lambda_g = \lambda_o/[1 - (f_c/f)^2]^{1/2}$ $v = v_o/[1 - (f_c/f)^2]^{1/2}$
Microstrip line	Transmission mode: TEM mode (under quasi-static approximation)

* Microstrip lines are planar transmission lines constructed by a copper trace on a dielectric substrate with a conducting ground-plane backing as shown in Figure 2.11. They support the lowest order TEM-mode under quasistatic approximations. (In quasistatic representation, the higher order modes resulting from fringing of the fields at the edges are ignored.) The microstrip characteristics are specified (approximately) by a set of *Wheeler's formulas*: $Z_o = Z_{oo}(\varepsilon_{eff})^{1/2}$ with $t \to 0$ and $W/d > 0.06$; $Z_{oo} = 377 \times [(W/d) + 1.98 \times (W/d)^{0.172}]^{-1}$; and $(\varepsilon_{eff})^{1/2} = 1 + [(\varepsilon_r - 1)/2] \times [1 + 1/(1 + 10d/W)^{1/2}]$.

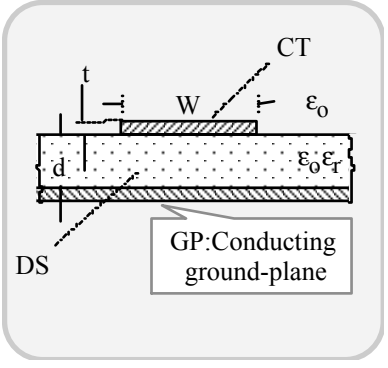


Fig. 2.11 A typical microstrip structure. (The thickness of the copper trace (CT) on the dielectric substrate (DS) is assumed to be very small as compared to other dimensions)

Example 2.2
The distributed constants of an open-wire transmission line made of copper are as follows: R = 6.50 ohm/km; L = 2.3 H/km; C = 0.0052 μF/km and G = 0.5 μS/km. Determine the line attenuation, phase constant and phase velocity at a low frequency of 796 Hz.

Solution
$\omega = 2\pi f = 5000$ $(f = 796$ Hz)
$Z = R + j\omega L = 13.1\angle 60.45^{o}$ ohm/km; $Y = G + j\omega C = 26.1\angle 41.76^{o}$ S/km
$\gamma = (Z/Y)^{1/2} = (0.005 + j0.018) \equiv (\alpha + j\beta)$
$\therefore$ Attenuation constant, $\alpha = 0.005$ neper/km
Phase constant, $\beta = 0.018$ radian/km
Phase velocity, $v = 1/(LC)^{1/2} = \omega/\beta = 278$ km/s.

Problem 2.2
Repeat the exercise of Example 2.2, for a frequency of 900 MHz.

Problem 2.3
Calculate the characteristic impedance of a microstrip line made with a copper trace on an alumina substrate (with $\varepsilon_r = 8.8$). Assume W/d = 7.5 and d = 0.015λ. Determine the fractional change in Z_o over the frequency range of 2 to 3 GHz.

Problem 2.4
Illustrated in Figure 2.12 is a system of a transposed half-wave line section used to feed two antenna elements in phase.

The two centre-fed antenna elements are spaced λ/4 apart in the vertical plane. They are fed in phase from a common transmission line *via* a pair of λ/2 section. The lines are transposed to satisfy the in-phase condition.

Suppose each antenna element has a driving point impedance of (72 + j0) ohm and the feed-line has $Z_o = 300$ ohm. Design (1) a single-stub matching system; and, (2) a double-stub matching system.

Problem 2.5
An antenna typically exhibits symmetric or asymmetric conjugate driving-point impedance characteristics about the resonant frequency. Develop a design procedure for a λ/4 transformer matching with a VSWR constrain of $S \leq 2$ over a band of interest about the resonant frequency. Assume a perfect matching at the resonant matching to the line that has a characteristic impedance of Z_o.

(*Hint*: See [2.12], page 498-503])

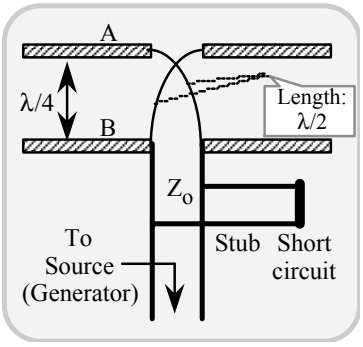


Fig. 2.12 An in-phase feeding of an antenna system from a common transmission line

Problem 2.6(a)

A tapped λ/4 line is illustrated in Figure 2.13. It acts as an impedance transformer in matching a line to an antenna.

Show that the resistive input impedance R_{in} (as seen by the exciting source) is given by: $R_{in} = (4Z_o/\alpha\lambda)(1/[1+4Z_o/\alpha\lambda R_L)\sin^2(2\pi d/\lambda)$ where R_L is the load presented by the antenna.

(*Hint*: See [2.13], page 347).

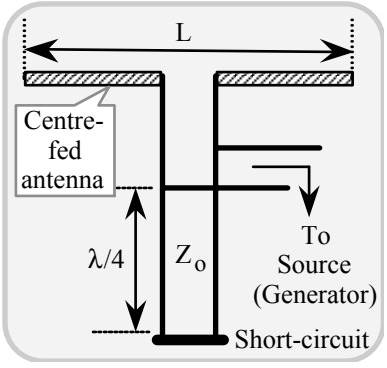


Fig. 2.13 An antenna feed arrangement based on tapped quarter-wave line principle

Problem 2.6(b)
Using the concept of Problem 2.5(a), design a tapped $\lambda/4$ line to match a source having an internal impedance of 50 ohm connecting an antenna via a transmission line of $Z_o = 50$ ohm. Assume that the antenna is a $\lambda/2$ dipole, centre-fed operating at 900 MHz. The line parameters are as follows: R = 0.008 ohm/m; L = 2.0 μH/m and C = 5 pF/m.

(*Hint*: $\alpha = R/2Z_o$).

2.3 ELECTROMAGNETIC THEORY

Principles of EM radiation and antenna theory are built on the classical concepts of electromagnetic postulations described by Maxwell's equations. Consistent with the review of electromagnetic phenomena presented above, the underlying analytical considerations are elaborated in this section; and, relevant perspectives of analyses are extended to explain the principle of radiation and wave propagation in the next chapter.

As detailed in the previous sections, the electric and magnetic fields are force fields perceived in a medium and are, in general, specified in spatial co-ordinates (like x, y, z in Cartesian system) and at an instant of time, t. The analytical descriptions of these force fields in the light of the associated physics form the scope of electromagnetic theory.

The important vector and scalar quantities pertinent to the EM field theory considerations are indicated below and are used consistently throughout the text:

Vector field entities

E Electric field intensity in volts/m
H Magnetic field intensity in amperes/m
D Electric displacement density in coulombs/m^2
B Magnetic flux density in webers/^{2}
A Magnetic vector potential in amperes/m^2
F Electric vector potential in volts/m^2
P Poynting vector in watts/m^2

Scalar field entities

V Electric scalar potential in volts
U Magnetic scalar potential in amperes

Medium parameters

ε Permittivity of the medium in farads/m
- ε_r Relative permittivity or dielectric constant
- ε_o Absolute permittivity of free-space
  $= (1/36\pi) \times 10^{-9}$ farads/m

$(\varepsilon_r' - j\varepsilon_r'')$ Complex (relative) permittivity
$\tan(\delta)$ Loss-tangent of the medium [$= \tan(\varepsilon_\rho'/\varepsilon_\rho'')$]

μ Permeability of the medium in henrys/m
- μ_r Relative permeability
- μ_o Absolute permeability of free-space $= 4\pi \times 10^{-7}$ henries/m
- $(\mu_r' - j\mu_r'')$ Complex (relative) permeability

σ Conductivity of the medium in siemens/m
- ρ_R Resistivity of the medium ($= 1/\sigma$) in ohm-m

Vector source parameters
M Magnetic current density vector in volts/m^2
J Electric current density vector in amperes/m^2

Scalar source parameters
(ρ or q) Electric charge density
- (ρ_v or q_v) Volumetric electric charge density in coulombs/m^3
- (ρ_s or q_s) Surface electric charge density in coulombs/m^2
- (ρ_ℓ or q_ℓ) Line electric charge density in coulombs/m

EM wave parameters
- v Velocity of EM wave in a medium (in m/s)
- c Velocity of light or EM wave in free-space ($= 3 \times 10^8$ m/s)
- f Frequency (in Hz); $\omega = 2\pi f$
- λ Wavelength (in m); $f = v/\lambda$ in Hz; $k = \beta = 2\pi/\lambda$.

2.3.1 Electromagnetic waves

The analytical description of electromagnetic wave motion can be enunciated from the laws of electromagnetic induction specified by Maxwell's equations as indicated in the previous sections. Consider the following two equations of Maxwell:

$$\nabla \times \mathbf{E} - \nabla \times (\mathbf{v} \times \mathbf{B}) = \partial \mathbf{B}/\partial t \qquad (2.13a)$$

$$\nabla \times \mathbf{H} = \sigma \mathbf{E} + \partial \mathbf{D}/\partial t \qquad (2.13b)$$

In equation (2.13a), the term curl ($\mathbf{v} \times \mathbf{B}$) depicts the excitation of a Lorentz force field. It results from free charges under motion (with a velocity, v) in an ambient of a magnetic field, **B**.

In most problems (relevant to wireless communications taking place in a free-space regime), it is not necessary to include the Lorentz field term since the

prevalence of free charges in the medium of propagation is not significant so as to influence the EM field. Hence, equations (2.13a) and (2.13b) may be rewritten as follows:

$$\nabla \times \mathbf{E} = \partial \mathbf{B}/\partial t \tag{2.14a}$$

$$\nabla \times \mathbf{H} = \sigma \mathbf{E} + \partial \mathbf{D}/\partial t \tag{2.14b}$$

In equations (2.14a) and (2.14b), the sources of the electromagnetic field, namely, the electric and magnetic currents, have been omitted. However, if these sources are included, the two fundamental electromagnetic equations cited above may be written as,

$$\nabla \times \mathbf{E} - \mathbf{M}^i = \partial \mathbf{B}/\partial t \tag{2.15a}$$

$$\nabla \times \mathbf{H} = \sigma \mathbf{E} + \partial \mathbf{D}/\partial t + \mathbf{J}^i \tag{2.15b}$$

where $\mathbf{M}^i$ and $\mathbf{J}^i$ may be called respectively, as the *magnetic-* and *electric-source current densities*. They are also known, alternatively as *impressed magnetic and electric current densities* respectively.

In integral form, the equations (2.15a) and (2.15b) can be written as,

$$\oint_c \mathbf{E}.d\ell = -\iint_S (\partial \mathbf{B}/\partial t + \mathbf{M}_n^i).d\mathbf{S} \tag{2.16a}$$

$$\oint_c \mathbf{H}.d\ell = -\iint_S (\sigma \mathbf{E}_n + \partial \mathbf{D}_n/\partial t + \mathbf{J}_n^i).d\mathbf{S} \tag{2.16b}$$

In practical situations, only the sources and fields that vary harmonically with time are of concern. Hence, equations (2.16a) and (2.16b) become

$$\oint_c \mathbf{E}.d\ell = -\iint_S (j\omega\mu \mathbf{H}_n + \mathbf{M}_n^i).d\mathbf{S} \tag{2.17a}$$

$$\oint_c \mathbf{H}.d\ell = -\iint_S [(\sigma + j\omega\varepsilon)\mathbf{E}_n + \mathbf{J}_n^i].d\mathbf{S} \tag{2.17b}$$

using the constitutive relations $\mathbf{B} = \mu\mathbf{H}$ and $\mathbf{D} = \varepsilon\mathbf{E}$ and assuming that all quantities vary as $\exp(+ j\omega t)$.

When Maxwell's equations are expressed in the form of equations (2.17a) and (2.17b), they exhibit the characteristics of *symmetry*. That is, **E** and **H** correspond to each other, being expressed in volts per metre and amperes per metre, respectively. Likewise, **D** and **B** correspond to each other, being expressed

in coulombs (or ampere-seconds) per square metre and webers (or volt-seconds) per square metre, respectively; further, electric and magnetic currents correspond to each other, being measured in amperes and volts, respectively. Hence, it can be stated that **E** (force per unit positive charge) and **B** (force per unit electric current element) form a pair and **D** and **H** another. These symmetry considerations implicitly depict the associated *principle of duality*. Further, it can be noted that **D** and **B** are respectively determined from the charges and currents in the field, and they represent lines of force (or *flux line densities*) stemming from these charges and currents. In other words, these charges and currents depict sources for the force fields represented by the corresponding flux lines.

The total electric and magnetic currents indicated above when considered across a closed-surface would vanish. Relevant considerations lead to the following relations by taking the surface integrals (over a closed surface) of $\mathbf{M}_n^i$ and $\mathbf{J}_n^i$.

$$\iint_S (\partial \mathbf{B}/\partial t).d\mathbf{S} = -\iint_S \mathbf{M}_n^i.d\mathbf{S} = -\mathbf{K} \tag{2.18a}$$

$$\iint_S (\sigma \mathbf{E}_n + \partial \mathbf{D}_n/\partial t).d\mathbf{S} = -\iint_S \mathbf{J}_n^i.d\mathbf{S} = \mathbf{I} \tag{2.18b}$$

where **K** and **I** are respectively, the magnetic- and electric-source currents flowing out of a closed surface S.

Considering a perfect dielectric medium with $\sigma = 0$, equation (2.18b) upon integration with respect to time becomes,

$$\iint_S \mathbf{D}_n.d\mathbf{S} = -\int_{-\infty}^{t} \mathbf{I}dt = Q_e \tag{2.19a}$$

where Q_e is the total electric charge enclosed by the closed surface S. Equation (2.19a) specifies the *Gauss' theorem*. Similarly, integrating equation (2.18a) with respect to time, it follows that

$$\iint_S \mathbf{B}_n.d\mathbf{S} = -\int_{-\infty}^{t} \mathbf{K}dt = Q_m \tag{2.19b}$$

where Q_m is the total magnetic pole enclosed by S. However, inasmuch as no isolated positive or negative magnetic poles ("charges") exist, Q_m is always equal to zero. Therefore,

$$\iint_S \mathbf{B}_n.d\mathbf{S} = Q_m \equiv 0 \tag{2.19c}$$

In view of the above, the concept of existence of $\mathbf{M}^i$ or source magnetic current, appears to be questionable. However, this is not true of source electric current $\mathbf{J}^i$, which can be catered through a circuit loop by various types of physically-known generators of electric energy such as batteries, electric dynamos, and electron devices. Despite the uncertainty about the existence of source magnetic current prevailing, the concept of a fictitious source of magnetic current is very useful in practice. Besides, this concept renders the fundamental electromagnetic (or Maxwell's) equations mathematically symmetric and helps to solve such practical problems as radiation from antenna structures.

2.4 BOUNDARY CONDITIONS IN THE EM FIELD

Applying the fundamental relations of equations (16a) and (16b) to a rectangle ABB'A' across a boundary S' specified between two media, as shown in Figure 2.14, it follows that

$$E_t.AB + E'_n.BB' - E'_t.A'B' - E_n.AA' = \iint_S (\mu \partial \mathbf{H}_n / \partial t + \mathbf{M}^i_n).d\mathbf{S} \tag{2.20}$$

where E_n and E'_n are the normal components of **E** along AA' and BB', respectively. In the limit of AA' and BB' tending to zero, then

$$E_t - E'_t = 0 \tag{2.21a}$$

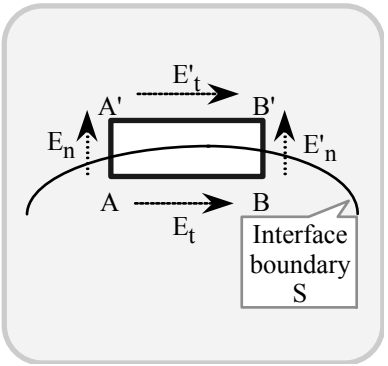


Fig. 2.14 Boundary between two media: An elemental closed path across the boundary with relevant normal and tangential electric field components

Applying a similar strategy to a magnetic field across the boundary, it can be shown that

$$H_t - H'_t = 0 \tag{2.21b}$$

Hence, the tangential components of **E** and **H** at the interface across the two dielectric media are *continuous* as indicated in an earlier section.

Since the circulation of the tangential component of **H** per unit area is the normal component of the electric current density **J**, it can be stated that the normal component of **J** would also be continuous across the boundary implying that,

$$J_n = J'_n \tag{2.21c}$$

Likewise,

$$M_n = M'_n \tag{2.21d}$$

For fields varying harmonically with time, specified by $\exp(j\omega t)$, the boundary conditions indicated *via* equations (2.21c) and (2.21d) become

$$(\sigma + j\omega\varepsilon)E_n = (\sigma' + j\omega\varepsilon')E'_n \tag{2.21e}$$

and,

$$\mu H_n = \mu' H'_n \tag{2.21f}$$

For static fields, the above relations reduce to $\sigma E_n = \sigma' E'_n$ and $\mu H_n = \mu' H'_n$. Further, if the two media are perfect dielectrics with $\sigma = \sigma' = 0$, equations (2.21e) and (2.21f) become $\varepsilon E_n = \varepsilon' E'_n$ (or) $D_n = D'_n$ and $\mu H_n = \mu' H'_n$ (or) $B_n = B'_n$. These relations lead to the boundary conditions stated in the earlier sections, namely, the normal components of **D** and **B** are continuous across the boundary of two dielectrics provided that the interface has no charge distribution. Should there exist surface charges of density q_s C/m^2 on this interface, then the normal component of **D** will be discontinuous to an extent of q_s. But, the continuity relation of the normal component of **B** will, however, remain unaffected (since there could be no isolated magnetic poles present at the interface to render the magnetic field nonsolenoidal).

In perfect conductors (with $\sigma = \infty$), the electric field intensity is zero for finite currents, and hence, at the interface between a perfect conductor and a perfect dielectric, $E_t = 0$ and $H_n = 0$. The concept of perfect conductors is valuable because it helps in simplifying mathematical calculations and in providing approximation to solutions of problems involving good conductors.

2.4.1 Boundary conditions in the vicinity of current sheet

A *current sheet* is defined as an infinitely thin sheet carrying finite current (electric

or magnetic) per unit length normal to the lines of flow. Figure 2.15 shows the cross-section of an electric current sheet whose linear current density vector $\mathbf{J}_S$ (in amperes per metre) is normal to the plane of the paper and is directed towards the reader.

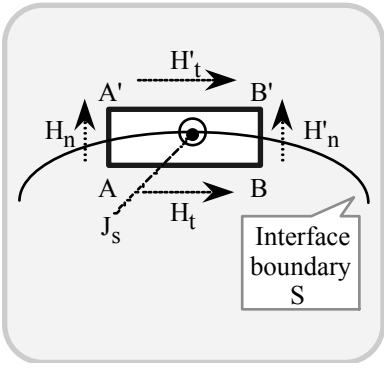


Fig. 2.15 Boundary between two media: An elemental closed path across the boundary with relevant normal and tangential magnetic field components

Applying the equation $\int_c \mathbf{H}.\mathrm{d}\ell = \iint_S (\sigma \mathbf{E}_n + \varepsilon \partial \mathbf{E}_n / \partial t).\mathrm{d}\mathbf{S} + \iint_S \mathbf{J}'_n.\mathrm{d}\mathbf{S}$ to a rectangle indicated as ABB'A' in Figure 2.12, it can be seen that,

$$H'_t.AB + H'_n.BB' - H''_t.A'B' - H''_n.AA' = \iint_S (\sigma \mathbf{E}_n + \varepsilon \partial \mathbf{E}_n / \partial t).\mathrm{d}\mathbf{S} + \iint_S \mathbf{J}^i_n.\mathrm{d}\mathbf{S} \quad (2.22)$$

and, as AA' and BB' tend to zero, the resulting limiting conditions lead to $H'_t - H''_t = J_S$, where the positive direction of current density, the tangential component of **H** and the normal to the sheet constitute a right-handed triad. Further, the relation $\oint_{CLP} \mathbf{E}.\mathrm{d}\ell = \iint_S (\mu \partial \mathbf{H}_n / \partial t + \mathbf{M}^i_n).\mathrm{d}\mathbf{S}$ applied to the closed path (CLP) ABB'A' (enclosing an area S), leads to the condition $E'_t = E''_t$. in the limit AA' $\rightarrow$ 0.Therefore, it follows that the tangential components of the electric field intensity in the vicinity of an electric current sheet of density $\mathbf{J}_s$ are continuous. However, the tangential components of the magnetic field intensity are discontinuous by an amount equal to the linear current density $\mathbf{J}_\ell$.

Similarly, for magnetic current-sheet density $\mathbf{M}_s$ volts per metre, it can be shown that $E'_t - E''_t = -M_S$ and $H'_t = H''_t$. In other words, there is a discontinuity in the tangential components of electric field intensity (equal to the linear magnetic current density) $\mathbf{M}_s$, and the tangential components of the magnetic field intensity

are continuous. The discontinuities in the tangential components of the field intensities imply discontinuities in the normal components of the current densities. Imagine a pill-box with its broad faces infinitesimally close and parallel to the electric current sheet on its opposite sides, as illustrated in Figure 2.16. Let A be the cross-section and $d\ell$ be the height of this pill-box. Suppose J'_n and J''_n are the normal components of the electric current density **J** on either side of the electric current-sheet and $\mathbf{J}_s$ is the surface current density on the electric current-sheet; then, using the *current continuity theorem* stipulating that the total electric current across the closed-surface of the pill-box should be zero, one obtains in the limiting cases of $d\ell \to 0$ and $dA \to 0$, the surface divergence condition on $\mathbf{J}_S$, namely $J'_n - J''_n = \text{div}(\mathbf{J}_S)$.

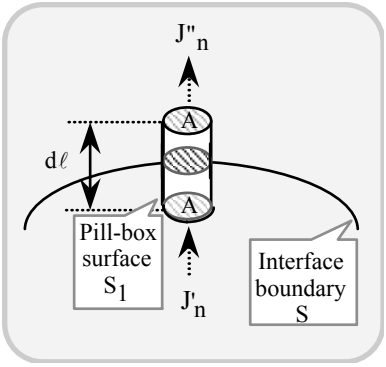


Fig. 2.16 Pill-box volume across a current sheet that could pose a discontinuity in the flow of a normal current component

Thus, the discontinuity in the normal component of the electric current density is equal to the surface divergence of the surface electric current density of the electric current sheet. Similarly, the discontinuity in the normal component of the magnetic current density is equal to the surface divergence of the magnetic surface current density of the magnetic current sheet.

2.4.2 Boundary conditions in the vicinity of infinitely thin linear current filaments

For an infinitely-thin electric current filament carrying I amperes per metre in the z-direction, the magnetic field intensity at a distance ρ metre is specified by $H_\rho = I/2\pi\rho$. Likewise, for an infinitely-thin magnetic current filament carrying K amperes per metre in the z-direction, the corresponding electric field intensity at a distance ρ metre is given by $E_\rho = K/2\pi\rho$.

2.5 THE POYNTING VECTOR

Consider the fundamental equations of electromagnetic induction given by equations (15a) and (15b). By taking the scalar product of equation (15a) and **H** and subtracting the result from the scalar product of the equation (15b) and **E**, the expression obtained can be subjected to integration over a closed volume. The result, when integrated with respect to time, would lead to the following relation (using the necessary vector identity and Gauss' theorem):

$$-\int_{-\infty}^{t} dt \iiint_v (\mathbf{E}.\mathbf{J}^i + \mathbf{M}^i.\mathbf{H}) dv = \int_{-\infty}^{t} dt \iiint_v \sigma E^2 dv + \iiint_v (1/2)(\varepsilon E^2 + \mu H^2) dv + \int_{-\infty}^{t} dt \iint_S (\mathbf{E} \times \mathbf{H})_n . d\mathbf{S} \qquad (2.23)$$

The left-hand side of equation (2.23) represents the total work performed by the impressed forces (up to the instant t), against the forces of the field so as to sustain the impressed or source currents, $\mathbf{J}_i$ and $\mathbf{M}_i$. In accordance with the principle of conservation of energy, one can say that this work appears as electromagnetic energy as specified by the right-hand side of the equation. The various terms on the right-hand side of equation (2.23) explicitly denote the following:

The first term is the part of EM energy converted into heat as a result of finite conductivity (σ) of the medium. The second term depicts the total electric and magnetic energies stored within S; and, the last term can be interpreted as a flow of energy across S (up to time t). That is, $\iint_S (\mathbf{E} \times \mathbf{H})_n . d\mathbf{S}$ can be treated as the time rate of flow of energy (or power flow) across S. Hence, as mentioned in an earlier section, the vector $\mathbf{E} \times \mathbf{H} = \mathbf{P}$, namely the Poynting vector, represents the time rate of energy flow or power flow per unit area. Its unit is watt per square metre.

The surface integral of the vector $\mathbf{P} = \mathbf{E} \times \mathbf{H}$ over a closed surface depicts the difference between the energy contributed to the field inside S and the energy accounted for within S. Further, the vector $\mathbf{P}^* = \mathbf{E} \times \mathbf{H}^*$ can be regarded as the *complex Poynting vector* with its real part depicting the average power flow per unit area.

If S corresponds to a perfect conductor, the tangential component of **E** on S vanishes; and hence, there is no flow of energy across S and the tangential component of **H** is parallel to the surface. In the physical world, metals represent good approximations to perfect conductors, but there are no good approximation to sheets on which **H** is zero except in certain substances with extremely high permeability at zero (or near-zero) frequency.

2.6 NORMAL AND SURFACE IMPEDANCE CONCEPTS

Only the tangential components of **E** and **H** contribute to the complex power flow ψ across a surface S specified by a pair of orthogonal co-ordinates u and v (on S). If u, v and n form a right-handed triplet of directions, of directions, then,

$$\psi = (1/2)\iint_S (\mathbf{E}_u \mathbf{H}_v^* - \mathbf{E}_v \mathbf{H}_u^*)dS = (1/2)\iint_S Z_n (\mathbf{H}_u \mathbf{H}_u^* - \mathbf{H}_v \mathbf{H}_v^*)dS \tag{2.24}$$

where Z_n is called the *impedance normal to the surface* S.

Now, consider a conducting surface of thickness t_o. The surface current density J_s in amperes per metre is equal to Jt_o, where J is the volume current density in amperes per square metre. If σ is the conductivity of the material, then by microscopic Ohm's law, $\mathbf{J} = \sigma \mathbf{E}$; and hence, $J_s = \sigma t_o E$. Suppose t_o approaches zero and σ increases to a very large value (so that the product $G = \sigma t_o$ remains constant). Then $J_s = G_s E$ and $E = R_s J$ where G_s and R_s are known respectively as the *surface conductance* (in siemens) and *surface resistance* (in ohms) of the sheet.

Pertinent to a medium where the displacement current density is considered in addition to the conduction current density, then for time-harmonic fields $J_s = (\sigma + j\omega\varepsilon)t_o E$ and as $t_o \to 0$ and $\sigma \to \infty$, then the following results are valid: $J_s = (G_s + jB_s)E = Y_s E$ and $E = (R_s + jX_s)J_s = Z_s J_s$. Here, Y_s denotes the *surface admittance*, G_s is the *surface conductance*, B_s is the *surface susceptance*, Z_s is the *surface impedance*, R_s is the *surface resistance*, and X_s represents the *surface reactance*.

A perfect conducting surface has zero normal impedance because the tangential component of **E** at this surface is zero; and hence, this surface is called a *surface of zero impedance*. Conversely, a surface of infinite impedance may be defined as one with zero tangential component of **H**, and it can be approximated by a surface of high permeability only at zero frequency.

In reference to characterising a material in terms of its bulk and/or surface impedance (or admittance) as above, the conductivity (σ), permittivity (ε) and permeability (μ) parameters adopted denote the *primary electromagnetic constants* of a medium. This definition is true in the sense that the presence of these parameters is explicit in the formulation of relevant electromagnetic equations pertinent to the medium in terms of the constitutive relations.

2.7 TRANSMISSION LINE AND MAXWELL'S EQUATIONS

Consider the transmission line equations namely,

$$dV/dx = -(R + j\omega L)I + E(x) \tag{2.25a}$$

$$dI/dx = -(G + j\omega C)V \tag{2.25b}$$

where E(x) is the voltage per unit length impressed along the line (that is, in series with the line). Further, V and I represent the instantaneous voltage across the line and the electric current flowing through the line respectively. Considering the (per unit length) iterated circuit representation of a transmission line, R and L respectively are the per unit length values of series elements, namely, resistance and inductance of the line; G and C respectively denote per unit length values of (shunt) elements, namely conductance and capacitance of the line as discussed earlier.

It can be observed that the transmission line equations indicated above form a special one-dimensional case of Maxwell's equations. In other words, the concepts and terminology of Maxwell's equations correspond to those of transmission-line equations. Thus, by designating $j\omega\mu$, $j\omega\varepsilon$, and σ as the distributed series impedance, shunt susceptance and shunt conductance entities, then $(\sigma + j\omega\varepsilon)$ would denote the distributed shunt admittance per metre of the bulk medium in which an EM wave propagates. Hence, the constants, μ, σ, and ε may be regarded respectively, as the "distributed parameters" depicting the series inductance, shunt conductance and shunt capacitance per metre of the bulk medium.

Further, in transmission-line theory, there are two secondary constants introduced, namely, the propagation constant (γ) and the characteristic impedance (Z_o). Correspondingly, considering the three-dimensional Maxwell's equations specified for a bulk medium, one may introduce a pair of secondary electromagnetic constants, namely, the wave propagation constant γ and the intrinsic impedance of the medium, η. These secondary constants are summarised in Table 2.3:

Table 2.3 Secondary parameters of transmission lines and EM wave propagating media

Secondary parameters	Transmission lines	Medium supporting an EM wave
Propagation constant (γ)	$[(R + j\omega L) \times (G + j\omega C)]^{1/2}$	$[j\omega\mu \times (\sigma + j\omega\varepsilon)]^{/2}$
Characteristic/intrinsic impedance (Z_o or η)	$[(R + j\omega L)/(G + j\omega C)]^{1/2}$	$[j\omega\mu/(\sigma + j\omega\varepsilon)]^{1/2}$

As a matter of fact, the transmission lines support EM energy in wave format subject to the postulations of Maxwell. For convenience, however, the characteristics of the line are studied in terms of voltage (instead of **E**) and current (instead of **H**) functions in view of the familiar circuit-theory principles normally applied to networks and lines at low frequencies. Therefore, the functional attributes and parameters specified to transmission lines are just another manifestations of EM wave propagation considerations; a transmission line simply depicts a *bounded* medium supporting relevant wave propagation mode. (On the other hand, in a bulk medium of large size (relative to wavelength), the EM wave propagates as an *unbounded* wave.) In an unbounded medium, the constants γ and

η are independent of the geometry of the wave but are "intrinsic" or "characteristic" of the medium.

In the RF applications, the primary constants of the medium are positive and hence $\gamma = \alpha + j\beta$ and $\eta = R + j X$ are either in the first quadrant or in the third quadrant. The definitions of γ and η are unambiguous if they always lie in the first quadrant or on its boundaries. For perfect conductors, $\sigma = \infty$, and hence both γ and η_o lie on the bisector of the first quadrant. For perfect dielectrics, $\sigma = 0$, and hence γ is on the positive imaginary and η on the positive real axis. In general, both γ and η are complex quantities specified by $\gamma = \alpha + j\beta$ and $\eta = R + j X$ where α, β, R and X are known respectively, as the *intrinsic attenuation constant*, *intrinsic phase constant*, *intrinsic resistance*, and *intrinsic reactance* of the medium. For different media of practical interest, the simplified formulations presented in Table 2.4 will be useful in the designs.

In Table 2.4, the entities v and λ are called the *intrinsic* or *characteristic velocity* and λ is the *wavelength* of the EM wave in the medium.

The quality factor (Q) of a medium is defined as the ratio of the displacement current density to the conduction current density, and hence, it is equal to $\omega\varepsilon/\sigma = 1/D$ where D is called the *dissipation factor*. For good dielectrics $Q > 1$, and for good conductors, $Q \sim 1$. The Q and D of a medium are implicit parameters representing the lossy nature of a medium. More explicitly, the lossy characteristics of a medium can be denoted by a *complex (relative) permittivity* given by: $(\varepsilon_r' - j\varepsilon_r'')$ where ε_r' represents the dielectric constant of the medium and ε_r'' denotes the loss component introduced in the *dielectric relaxation process* of the dipole molecules (in the medium subjected to an external EM excitation). This loss is equivalently specified by the conductivity parameter σ. Also, more commonly, another term known as the *loss-tangent* is used in practice to depict the dielectric loss. It is equal to $\tan(\delta) = \tan(\varepsilon_r''/\varepsilon_r')$ where $\varepsilon_r'' = \sigma/\omega\varepsilon_o$ with ω being $2\pi \times$ (frequency of the EM excitation).

Table 2.5 gives the values of ε_r, for various materials including those of interest in wireless communication media. An exhaustive treatment on electromagnetic materials has been presented by one of the authors in [2.13].

2.7.1 EM wave equation in dielectrics and conductors

From the considerations deliberated earlier, it follows that in homogeneous source-free regions, the wave equations are,

$$\nabla^2\mathbf{E} = \mu\varepsilon\partial^2\mathbf{E}/\partial^2 t + \mu\sigma\partial\mathbf{E}/\partial t \qquad (2.26a)$$

$$\nabla^2\mathbf{H} = \mu\varepsilon\partial^2\mathbf{H}/\partial^2 t + \mu\sigma\partial\mathbf{H}/\partial t \qquad (2.26b)$$

In perfect dielectrics with $\sigma = 0$, equations (2.26a) and (2.26b) become

Table 2.4 EM parameters for different media: Free-space, perfect dielectric and good conductor

Parameters	Free-space	Perfect dielectric	Good conductor
ε, μ and σ	ε_o (F/m) $= (1/36\pi) \times 10^{-9}$ μ_o (H/m) $= (4\pi) \times 10^{-7}$ $\sigma = 0$	$\varepsilon_o\varepsilon_\rho$ (F/m) $\mu_o\,\mu_\rho$ (H/m) $\sigma = 0$	σ (S/m) >> $\omega\varepsilon$
γ	$j\beta_o$ $\beta_o = \omega(\mu_o\varepsilon_o)^{1/2}$ $= \omega/v_o = 2\pi/\lambda_o$ $v_o = c = (\mu_o\varepsilon_o)^{-1/2}$ $= 3\times 10^8$ m/s $= f\lambda_o$ $\lambda_o = 2\pi/\beta_o$	$j\beta$ $\beta = \omega(\mu\varepsilon)^{1/2}$ $= \omega/v = 2\pi/\lambda$ $v = c \times (\varepsilon_r)^{-1/2}$ $= f\lambda$ $\lambda = 2\pi/\beta$	$\alpha + j\beta$ $\alpha = \beta$ $= (\omega\mu\sigma/2)^{1/2}$
η	$\eta_o = (\mu_o/\varepsilon_o)^{1/2}$ $= 120\pi$ ohms	$(\mu/\varepsilon)^{1/2}$ $= 120\pi \times(\varepsilon_r)^{-1/2}$ ohms	$(j\omega\mu/\sigma)^{1/2}$ ohms
R and X	—	—	R = X $= (\omega\mu/2\sigma)^{1/2}$ (in ohms)

$$\nabla^2\mathbf{E} = \mu\varepsilon\partial^2\mathbf{E}/\partial^2 t \tag{2.27a}$$

$$\nabla^2\mathbf{H} = \mu\varepsilon\partial^2\mathbf{H}/\partial^2 t \tag{2.27b}$$

And, for time-harmonic fields, the above wave equations can be written as

$$\nabla^2\mathbf{E} = \gamma^2\mathbf{E} \tag{2.28a}$$

$$\nabla^2\mathbf{H} = \gamma^2\mathbf{H} \tag{2.28b}$$

Table 2.5 Complex relative permittivity and conductivity of some materials at UHF and/or microwave frequencies

Material/medium	Relative complex permittivity $\varepsilon_r' - j\varepsilon_r''$		Conductivity σ (siemens/m)
	ε_p'	ε_p''	
Air (0°C) and vacuum	1.00	0	0
Aerated concrete	2.00 – 3.00	—	—
Alumina	8.80	0.014	—
Asphalt	2.68	—	—
Bakelite™	5.00	0.035	—
Barium titanate ($BaTiO_2$)	1200	—	—
Carbon dioxide (0°C)	1.000985	0	0
Chip board	2.90	—	—
Dry/wet soil	3.40/7.00	—	10^{-5} /10^{-3}
Dry brick	40	—	—
Dry concrete	4.00 – 6.00	—	—
Flintglass	6.60 - 9.90	—	—
Formica™	6.00	—	—
Gypsum board	2.80	—	—
Ice	3.20	—	—
Lead glass	5.40- 8.00	—	10^{-12}
Lime stone	7.50	—	—
Marble/calcarious stones	11.00	—	—
Mica	5.60-6.00	—	—
Nylon (hard)	4.00-6.00	—	—
Paper	2.00-2.50	—	—
Paraffin	2.10-2.20	0.003	—
Plexiglas™	3.40	—	—
Plywood	2.10	—	—
Polyethylene	2.60	0.002	—
Polystyrene	2.60-2.70	0.002	—
Porcelain	5.70	—	—
Quartz	5.00	—	—
Rubber (Chloroprene)	2.30-4.00	0.004	10^{-15}
Rutile (TiO_2)	89-173	—	—
Snow	1.20 – 1.50	—	—
Styrofoam	1.03	0	0
Teflon™	2.1	0.001	—

Water (distilled)/Fresh	81.1	—	10^{-3}
Water (sea/salt)	75.00-78.00	—	4.00
Water vapour (145°C)	1.00705	0	0
Wood	2.50-7.70	—	—

2.7.2 Solution of EM wave equation in Cartesian co-ordinates

Considering a homogeneous, linear, isotropic and charge-free medium, the vector wave equations of equation (2.28) representing a uniform plane wave in Cartesian co-ordinates reduces to a set of scalar wave equations given by,

$$\nabla^2 E_x = \gamma^2 E_x \tag{2.29a}$$

$$\nabla^2 E_y = \gamma^2 E_y \tag{2.29b}$$

$$\nabla^2 E_z = \gamma^2 E_z \tag{2.29c}$$

A solution of any of the equation (2.29) is of the form

$$E_x = \exp(\pm\Gamma_x \pm\Gamma_y \pm \Gamma z) \tag{2.30}$$

where the propagation constants Γ_x, Γ_y, and Γ_z in the x, y, and z co-ordinate axes respectively satisfy the condition

$$\Gamma_x^2 + \Gamma_y^2 + \Gamma_z^2 = j\omega\mu(\sigma + j\omega\varepsilon) \tag{2.31a}$$

Further, in nondissipative media $\sigma = 0$; hence,

$$\Gamma_x^2 + \Gamma_y^2 + \Gamma_z^2 = -\omega^2\mu\varepsilon = -\beta^2 = 4\pi^2/\lambda^2 \tag{2.31b}$$

With reference to equation (2.31), only two of the propagation constants of the set (Γ_x, Γ_y, Γ_z) are independent and they depend on the distribution of sources producing the field. If the distribution of sources is uniform in planes parallel to the xy-plane, the field would be uniform in these planes. Further, if $\Gamma_x = \Gamma_y = 0$, then the propagation constant (Γ_z) in the z-direction is equal to the intrinsic propagation constant γ. Such a wave depicts a *uniform plane wave* whose equiphase and equiamplitude planes are parallel to the xy-plane. Relevant expression for any of the components of the field is of the form:

$$E_x = \exp(\pm\Gamma_z)\exp(j\omega t) \equiv \exp[\pm(\alpha + j\beta)z]\exp(j\omega t) \tag{2.32a}$$

The positive signs in equation (2.32a) would let a wave increase amplitude indefinitely as z increases. This is not possible in a physical situation unless there is a (regenerative) conversion of energy of some other form into electromagnetic energy (as occurs in an electron device where the kinetic or potential energy of the charge carriers is converted into electromagnetic energy). Hence, it is appropriate to use only the negative sign. Therefore, it follows that,

$$E_x = \exp(-\alpha z)\exp[j(\omega t - \beta z)] \tag{2.32b}$$

which represents an attenuated uniform plane wave travelling in the positive z-direction with a phase velocity v equal to ω/β. In a nondissipative or perfect dielectric medium, $\alpha = 0$; as such, there is no attenuation of the wave amplitude.
For a uniform plane wave, whose equiamplitude and equiphase planes are parallel to a plane, it can be observed that,

$$x\cos(A) + y\cos(B) + z\cos(C) = \text{constant} \tag{2.33}$$

where cos(A), cos(B) and cos(C) are the *direction-cosines* of a normal to the plane. As a result, the expression for any of the field components is of the form,

$$\begin{aligned} E_x &= \exp\{-\gamma[x\cos(A) + y\cos(B) + z\cos(C)]\} \\ &= \exp(-\Gamma_x x - \Gamma_y y - \Gamma_z z) \end{aligned} \tag{2.34}$$

so that the propagation constants along the three axes are given by:

$$\Gamma_x = \gamma\cos(A),\ \Gamma_y = \gamma\cos(B) \text{ and } \Gamma_z = \gamma\cos(C) \tag{2.35}$$

If the medium is nondissipative or lossless, $\sigma = 0$ yielding $\gamma = j\beta$. Therefore, $\Gamma_x = j\beta_x$ $\Gamma_y = j\beta_y$ and $\Gamma_z = j\beta_z$. Or, $\beta_x = \beta\cos(A)$, $\beta_y = \beta\cos(B)$ and $\beta_z = \beta\cos(C)$. Hence, the following relations can be explicitly stipulated:

$$\left.\begin{aligned} \lambda_x &= 2\pi/\beta_x = \lambda/\cos(A) \\ \lambda_y &= 2\pi/\beta_y = \lambda/\cos(B) \\ \lambda_z &= 2\pi/\beta_z = \lambda/\cos(C) \end{aligned}\right\} \tag{2.36}$$

$$\left.\begin{aligned} v_x &= \omega/\beta_x = v/\cos(A) \\ v_y &= \omega/\beta_y = v/\cos(B) \\ v_z &= \omega/\beta_z = v/\cos(C) \end{aligned}\right\} \tag{2.37}$$

$$\beta_x^2 + \beta_y^2 + \beta_z^2 = \beta^2 \tag{2.38}$$

$$1/\lambda_x^2 + 1/\lambda_y^2 + 1/\lambda_z^2 = 1/\lambda^2 \tag{2.39}$$

$$1/v_x^2 + 1/v_y^2 + 1/v_z^2 = 1/v^2 \tag{2.40}$$

Thus, for uniform plane waves in a nondissipative medium, the phase velocities in various directions are always greater than the characteristic phase velocity, and the phase constants β_x, β_y, and β_z are always less than the characteristic phase constant β of the medium. This property can be observed explicitly in guided wave propagations such as those supported in a rectangular waveguide.

There are some types of physical situations, for example, in the case *of surface waves* supported by surface-wave structures (or slow-wave structures) in non-dissipative media, where the phase constant in some direction, for example, the z-direction, could be greater than β; or, in other words, $\beta_z > \beta$.

In general, concerning a nondissipative medium, if the propagation constants Γ_x, Γ_y, and Γ_z are complex, then $\Gamma_x^2 + \Gamma_y^2 + \Gamma_z^2 = -\beta^2$ and it leads to the condition $\alpha_x\beta_x + \alpha_y\beta y + \alpha_z\beta_z = 0$, which shows that the equiamplitude planes $\alpha_x x + \alpha_y y + \alpha_z z$ = constant are perpendicular to the equiphase planes, namely, $\beta_x x + \beta_y y + \beta_z z$ = constant.

Thus, in nondissipative media, equiamplitude and equiphase planes will either coincide with each other (as for uniform plane waves) or could be orthogonal to each other. In the first instance, the waves are uniform on equiphase planes in the sense that **E** and **H** fields have constant values at all points of a given equiphase plane at any given instant. In the second case, the amplitude varies exponentially, the fastest variation being in the direction given by the direction components α_x, α_y, and α_z.

The foregoing general conclusions on waves of exponential type of the form $E_x = [\exp(-\Gamma_x - \Gamma_y - \Gamma_z)][\exp(j\omega t)]$, have a broader significance than it appears at first sight. The constant Γ_x represents the relative rate of change of E_x in the x-direction specified explicitly by $\Gamma_x = -(1/E_x)\partial E_x/\partial x$ and hence,

$$\left.\begin{aligned} \Gamma_x^2 &= (1/E_x)\partial^2 E_x/\partial^2 x^2 \\ \Gamma_x^2 &= (1/E_x)\partial^2 E_x/\partial^2 y^2 \\ \Gamma_x^2 &= (1/E_x)\partial^2 E_x/\partial^2 z^2 \end{aligned}\right\} \tag{2.41}$$

If the quantities Γ_x, Γ_y, and Γ_z vary slowly from one point to another, the solution of the wave equation will be approximately exponential, and the properties of exponential waves will be applicable even in sufficiently small regions.

2.7.3 EM waves at the interface between conductors and dielectrics

Consider a plane interface (the xy-plane) between air (substantially representing the free-space) above the plane and a conductor below the plane, as shown in Figure 2.17.

For an exponential wave of the functional form $F = \exp(-\Gamma_x - \Gamma_y - \Gamma_z)$, the propagation constants Γ_x and Γ_y in directions parallel to the boundary must be the same in both media so that the boundary conditions are satisfied at all points of the air-to-conductor interface. This is obvious because F represents a component of either **E** or **H** field parallel to the boundary. In the present case, F is continuous across the boundary and the continuity cannot be satisfied at all points unless Γ_x and Γ_y are the same on both sides of the boundary. The same condition applies to the normal components of the current densities, namely, $(\sigma + j\omega\varepsilon)E_z$ and $j\omega\mu H_z$. Thus, it follows that,

$$\Gamma_x^2 + \Gamma_y^2 + \Gamma_{z,o}^2 = -\beta_o^2 \qquad \text{(in air)} \qquad (2.42a)$$

$$\Gamma_x^2 + \Gamma_y^2 + \Gamma_z^2 = -\gamma^2 \qquad \text{(in conductor)} \qquad (2.42b)$$

where β_o is the characteristic phase constant of free-space equal to $\omega(\mu_o\varepsilon_o)^{1/2}$; and γ is the characteristic propagation constant of the conductor. Subtracting equation (2.42a) from equation (2.42b), it follows that $\Gamma_z^2 = \gamma^2 + \beta_o^2 + \Gamma_{z,o}^2$.

The propagation constant γ in the conductor is much larger than that in free-space (or air). That is, $\gamma^2 >> \beta_o^2$. Also $\Gamma_{z,o}$, representing the propagation constant in free-space in the direction normal to the interface, is comparable to β_o, if the wave direction is nearly normal to the interface; or, it is much smaller than β_o if the wave direction is nearly parallel to the interface. In other words, $\Gamma_{z,o}$ is almost equal to or much smaller than β_o. Hence $\Gamma_{z,o}$ can be assumed to be much smaller than γ. Therefore, it follows from the relation $\Gamma_z^2 = \gamma^2 + \beta_o^2 + \Gamma_{z,o}^2$ that $\Gamma_z = \gamma$. In other words, in a conducting medium, the propagation constant normal to the interface is substantially equal to the intrinsic propagation constant of the conductor.

Further, inasmuch as the current density normal to the interface is continuous, the following relations can be observed:

$$\sigma E_n'' = j\omega\varepsilon_o E_n' \qquad (2.43)$$

where E_n'' and E_n' are the normal components of **E** field in air and the conductor, respectively. Since σ for the conductor is very large, $E_n'' << E_n'$; or, the normal component of **E** in the conductor is much smaller than the normal component of **E** in air (free-space).

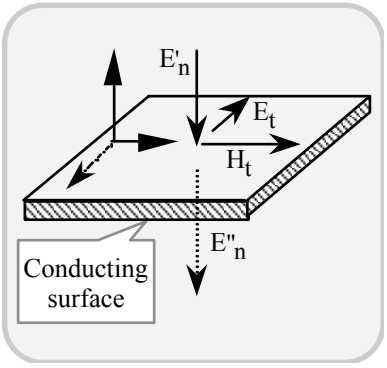


Fig. 2.17 Plane boundary between two semi-infinite media (separated by a conductor)

Even at moderately high frequencies, the attenuation constant in a conductor is large, and the field becomes quite small even at short distances from the interface. That is, except at very low frequencies, the fields are confined largely to thin skins of conductors, and this effect is known as the *skin effect*.

The current density at the surface of the conductor is σE_t, where E_t is the tangential component of **E** at the interface. Elsewhere inside the conductor it is $\sigma E_t \exp(-\gamma z)$, where z is the perpendicular distance from the surface. The total current per unit length normal to the lines of flow in a conductor is given by:

$$J_S = \int_0^\infty \sigma E_t \exp(-\gamma z) dz = \sigma E_t / \gamma = E_t / \eta \qquad (2.44)$$

where $\eta = (j\omega\mu/\sigma)^{1/2}$ and $\gamma = (j\omega\mu\sigma)^{1/2}$. Also, on the other hand, at the surface of the conductor the tangential component of **H**, namely H_t, is equal to J_s, and is given by $H_t = J_S = E_t / \eta$. Or $E_t/H_t = Z_n$, which depicts the normal impedance of the surface equal to the intrinsic impedance of the conductor (η). By definition, E_t/J_s is the *surface impedance* Z_s, of the interface, as mentioned earlier. Therefore, Z_s of an interface between a good conductor and a good dielectric is equal to the *normal impedance* Z_n. The conductor can also be replaced by a sheet (with a surface impedance equal to η) adjacent to a sheet of infinite impedance (that would effectively exclude the space previously occupied by the conductor).

The parameters Z_s and γ are explicitly given by: $Z_s = R + jX$ with $R = \alpha/\sigma$ and $\gamma = (\alpha + j\beta) = (j\omega\mu\varepsilon)^{1/2}$. The reciprocal of α denotes the skin-depth (δ) of wave penetration into the conducting medium. Hence, the surface resistance R is equal to the dc resistance of a plate of thickness equal to $1/\alpha$, namely, the skin

depth (δ). The field reduces to (1/e) of its value at the surface at a distance equal to the skin depth, and the attenuation through the skin depth corresponds to one neper.

2.8 MAXWELL'S EQUATIONS IN SOURCE-FREE REGIONS

The fundamental electromagnetic (or Maxwell's) equations in source-free regions for time-harmonic fields are of usual interest in describing the EM fields associated with transmission lines, antenna structures and in wave propagations. Such field components have to be specified in an appropriate co-ordinate system consistent with the geometry of the EM structure and/or media of interest. Hence, presented in Appendix 2.1 (Subsection 2A.4) are the sets of Maxwell's relations in the rectangular, cylindrical and spherical co-ordinates for convenient reference. They refer to the curl relations, namely, curl($\mathbf{E}$) = $-j\omega\mu\mathbf{H}$ and curl($\mathbf{H}$) = $j\omega\varepsilon\mathbf{E} + \sigma\mathbf{E}$ in linear, homogeneous, isotropic and charge-free media subjected to harmonic excitation at an angular frequency, ω.

2.9 CONCLUDING REMARKS

The subject of electromagnetic fields and waves stemmed from experimental postulations and deliberated *via* Maxwell's equations. An exhaustive and excellent selection of books exists in the literature written on EM theory and some are cited here [2.1 – 2.7]. These books are written to cover both fundamental considerations as well as advanced perspectives of the subject matter. The classical books by Jordan and Balmain [2.1], Ramo, Whinnery and Van Duzer [2.2], Harrington [2.5], Balanis [2.15] and Stratton [2.6] are a few mentioned for their diversity in presentations. A host of other books, such as [2.10] and [2.11], have emerged in the recent past, which provide adequate insight into the academic aspects of electromagnetic engineering.

The concepts of transmission line theory have been treated in the books on EM theory as well as in books on radio communications. The grand, old books by Terman [2.16], Everitt and Anner [2.12] and Ryder [2.13] still bear their comprehensiveness and practical considerations as warranted in the design and understanding of transmission line theory applied to modern radio/wireless systems.

What is attempted in this chapter is an outline presentation on EM fields and waves tailored to understand the underlying considerations as would be required in developing EM radiation and antenna principles presented in the chapters that follow. Readers are encouraged to supplement the contents of this chapter with the details available in the archival of books on this subject.

REFERENCE

[2.1] Jordan, E. C., and K. G. Balamain, *Electromagnetic Waves and Radiating Systems*, Englewood Cliff, NJ: Prentice-Hall, 1968

[2.2] Ramo, S., Whinnery, J. R., and T. Van Duzer, *Fields and Waves in Communication Electronics*, New York, NY: John Wiley & Sons, Inc., 1994

[2.3] Plonus, M. A., *Applied Electromagnetics*, New York, NY: McGraw-Hill Book Co., 1978

[2.4] Cheng, D. K., *Field and Wave Electromagnetics*, Reading, MA: Addison-Wesley Pub. Co., 1989

[2.5] Harrington, R. F., *Time-Harmonic Electromagnetic Fields*, New York, NY: McGraw-Hill Book Co., 1961

[2.6] Stratton, J. A., *Electromagnetic Theory*, New York, NY: McGraw-Hill Book Co., 1941

[2.7] Plonsey, R., and R. E. Collin, *Principles and Application of Electromagnetic Fields*, New York, NY: McGraw-Hill Book Co., 1961

[2.8] Schelkunoff, S. A., *Electromagnetic Waves*, New York, NY: Van Nostrand, 1943

[2.9] Wait, J. R., *Electromagnetic Wave Theory*, New York, NY: Harper & Row Publishers, 1985

[2.10] Chatterjee, R., *Antenna Theory and Practice*, New Delhi, India: New Age International (P) Ltd. Publishers, 1996

[2.11] Sadiku, M. N. O., *Elements of Electromagnetics*, New York, NY: Oxford University Press, 2001

[2.12] Everitt, W. L., and G. E. Anner, *Communication Engineering*, New York, NY: McGraw-Hill Book Co., 1956

[2.13] Ryder, J. D., *Networks, Lines and Fields*, Bombay, India: Asia Publishing House, 1962

[2.14] Neelakanta, P. S., *Handbook of Electromagnetic Materials*, Boca Raton, FL: CRC Press., 1995

[2.15] Balanis, C. A., *Advanced Engineering Electromagnetics*, New York, NY: John Wiley and Sons., 1989

[2.16] Terman, F. E., *Electronic and Radio Engineering*, New York, NY: McGraw-Hill Book Co., 1955

This page intentionally left blank

APPENDIX 2.1

Vector Calculus and Generalised Co-ordinate System

2A.1 Vector calculus: Notations

Del Operator:	∇
Gradient of a scalar, Φ:	$\nabla.\Phi$
Divergence of a vector, **B**:	$\nabla.\mathbf{B}$
Curl of a vector, **E**:	$\nabla \times \mathbf{E}$
Laplacian of a scalar, V:	$\nabla.(\nabla V) = \nabla^2 V$

2A.2 Vector identities

$$\nabla.\ (\nabla \times \mathbf{F}) = 0$$
$$\nabla \times (\nabla f) = 0$$
$$\nabla \times \nabla \times \mathbf{F} = \nabla\ (\nabla.\mathbf{F}) - \nabla^2\mathbf{F}$$
$$\nabla\ (fg) = f\,\nabla g + g\,\nabla f$$
$$\nabla.(g\mathbf{F}) = g\nabla.\mathbf{F} + \mathbf{F}.\nabla g$$
$$\nabla \times (g\mathbf{F}) = g\,\nabla \times \mathbf{F} + \nabla g \times \mathbf{F}$$

2A.3 Generalised rectangular, cylindrical and spherical co-ordinate transformations

Co-ordinate system

Rectangular	Cylindrical	Spherical
$u_1 = x, h_1 = 1, a_1 = a_x$ $u_2 = y, h_2 = 1, a_2 = a_y$ $u_3 = z, h_3 = 1, a_3 = a_z$	$u_1 = r, u_2 = \phi, u_3 = z$ $h_1 = 1, h_2 = r, h_3 = 1$	$h_1 = 1, h_2 = r, h_3 = r \sin(\theta)$

$$\nabla\,\phi = (1/h_1)\,(d\mathbf{a}_1/du_1) + (1/h_2)\,(d\mathbf{a}_2/du_2) + (1/h_3)\,(d\mathbf{a}_3/du_3)$$

$$\nabla.\mathbf{D} = (1/h_1h_2h_3)\,[d(h_2h_3\,D_1)/du_1 + d(h_1h_3\,D_2)/du_2 + d(h_2h_1\,D_3)/du_3]$$

$$\nabla \times \mathbf{H} = (1/h_1h_2h_3)\begin{vmatrix} \mathbf{a_1}h_1 & \mathbf{a_2}h_2 & \mathbf{a_3}h3 \\ \partial/\partial u_1 & \partial/\partial u_2 & \partial/\partial u_3 \\ h_1H_1 & h_2H_2 & h_3H3 \end{vmatrix}$$

$$\nabla^2\Phi = (1/h_1h_2h_3)[d\{(h_2h_3/h_1)d\Phi/du_1\}/du_1 + d\{(h_3h_1/h_2)d\Phi/du_2\}/du_2 + d\{(h_1h_2/h_3)\, d\Phi/du_3\}/du_3]$$

2A.4 Maxwell's Equations in Source-Free Regions

Maxwell's equations in Cartesian (x, y, z) co-ordinate system

$$\partial E_z/\partial y - \partial E_y/\partial z = -j\omega\mu H_x$$
$$\partial E_x/\partial z - \partial E_z/\partial x = -j\omega\mu H_y$$
$$\partial E_y/\partial x - \partial E_x/\partial y = -j\omega\mu H_z$$
$$\partial H_z/\partial y - \partial H_y/\partial z = j(\sigma + \omega\varepsilon)E_x$$
$$\partial H_x/\partial z - \partial H_z/\partial x = j(\sigma + \omega\varepsilon)E_y$$
$$\partial H_y/\partial x - \partial H_x/\partial y = j(\sigma + \omega\varepsilon)E_z$$

Maxwell's equations in cylindrical (ρ, ϕ, z) co-ordinate system

$$\partial E_z/\partial\phi - \partial E_\phi/\partial z = -j\omega\mu\rho H_\rho$$
$$\partial E_\rho/\partial z - \partial E_z/\partial\rho = -j\omega\mu\rho H_\phi$$
$$\partial(\rho E_\phi)/\partial\rho - \partial E_\rho/\partial\phi = -j\omega\mu\rho H_z$$
$$\partial H_z/\partial\phi - \partial H_\phi/\partial z = j(\sigma + j\omega\varepsilon)\rho E_\rho$$
$$\partial H_\rho/\partial z - \partial H_z/\partial\rho = j(\sigma + j\omega\varepsilon)\rho E_\phi$$
$$\partial(\rho H_\phi)/\partial\rho - \partial H_\rho/\partial\phi = j(\sigma + j\omega\varepsilon)\rho E_z$$

Maxwell's equations in spherical (r, θ, ϕ) co-ordinate system

$$\partial[\sin(\theta)E_\phi]/\partial\theta - \partial E_\theta/\partial\phi = -j\omega\mu r\sin(\theta)H_r$$
$$\partial(rE_\theta)/\partial r - \partial E_r/\partial\theta = -j\omega\mu rH_\phi$$
$$\partial E_r/\partial\phi - \sin(\theta)\partial(rE_\phi)/\partial r = -j\omega\mu r\sin(\theta)H_\theta$$
$$\partial[\sin(\theta)H_\phi]/\partial\theta - \partial H_\theta/\partial\phi = j(\sigma + j\omega\varepsilon)r\sin(\theta)E_r$$
$$\partial(rH_\theta)/\partial r - \partial H_r/\partial\theta = j(\sigma + j\omega\varepsilon)rE_\phi$$
$$\partial H_r/\partial\phi - \sin(\theta)\partial(rH_\phi)/\partial r = j(\sigma + j\omega\varepsilon)r\sin(\theta)E_\theta$$

CHAPTER 3

Radiation, Antennas and EM Wave Propagation

3.1 INTRODUCTION

Radiation of electromagnetic (EM) energy is a "leakage phenomenon" that takes place at a physical structure coupled to a surrounding space and excited by a signal source. The space around the physical structure is known as the *propagation medium* and it supports the radiated electromagnetic energy in the form of a propagating wave. Further, the physical structure in question could manifest in different forms. For example, a circuit or a transmission line of appropriate size and configuration can be designed to transfer electromagnetic energy from a source and excite the surrounding medium (such as free-space) so that it results in a leakage of that energy in the form of an electromagnetic wave.

Thus, the physical structure, in essence, is a "radiator" and mediates the EM energy transfer between the source and the surrounding space. When it represents a transition unit bridging the signal-source and the medium that supports the EM wave is technically known as the " transmitting antenna". Reciprocally, a similar (physical) structure residing in a medium could intercept or "receive" the EM wave incident on it and transfer the energy into a receiving unit. In this case, it is known as the "receiving antenna". In this receiving mode, the antenna gets excited by the EM wave intercepted from a desired direction (or a set of directions) and efficiently transfers the energy to the receiver system. Likewise, in the transmitting mode also, the antenna can be designed to radiate in a preferred direction or directions.

In short, the underlying considerations behind realising an effective electromagnetic radiator (or the antenna) refer to achieving a controlled "leakage" of the electromagnetic energy, which would let a propagating wave to proliferate across the surrounding medium. Controlling the radiation (or reception) of electromagnetic energy at an antenna structure governs the desired excitation of an EM wave in a preferred direction (or directions) as efficiently as possible.

Hence, the subject of antenna theory* is concerned with a cohesive analysis of the methods by which EM energy is radiated (or received) by a physical structure. Such an analysis also portrays the associated *mechanism of radiation*. Further, antenna theory covers the endeavour of understanding the structural (geometry, size, material and spatial orientation) aspects of the physical radiator (in facilitating the desired

---

* *Note*: The analytical presentations in this chapter on antenna theory are intensely tied to relevant electromagnetic equations. It is left to reader's choice whether to skip those sections.

radiation characteristics) in relation to the wave characteristics, such as wavelength and orientation of **E**- or **H**-vector in space. (This orientation specifies the polarisation of the wave as discussed in the earlier chapters.) In other words, the theory of antenna elaborates the mechanism of EM radiation and leads to the evolution of engineering plus technological considerations pertinent to the design and evaluation of various antenna structures that can be conceived for specified applications.

Parallel to the theory and techniques of antennas, exits a related consideration of interest in radio engineering: It refers to the *propagation of the EM wave*, which is radiated by an antenna (or eventually intercepted by an antenna). This EM wave propagation is characteristically decided by the medium that supports it. Ideally, the "free-space" is presumed as the vehicle for RF channels. But, in reality, wireless communication links face more than the simple "vacuum" description of the "free-space"! The associated RF channels often see a distinct ambient, both in outdoor and indoor deployments. It is an environment infested with obstacles, which may reflect, scatter and/or diffract the incident EM waves. That is, the EM wave radiated or received proliferates in a medium of distinct characteristics in any given wireless communication application. In an outdoor wireless communication scenario, the radio transmissions of concern may prevail in urban locales and/or in the countryside, with the features of the terrain and surroundings changing significantly. Considering the wireless communications deployed exclusively in the indoor environments, the indoor building architecture and the contents (such as furniture etc.) in the house, dictate the associated radio channel profiles. Hence, in the art of relevant antenna designs, due considerations are needed to cope with the EM radiation *vis-à-vis* the changes in propagation characteristics across the area of service deployment.

In view of the above, this chapter is written with a three-fold objective. They refer to the following: (1) To illustrate the underlying principle of EM radiation (and/or reception) by an antenna; (2) to derive the radiated field relations from the basic Maxwell's equations and (3) to outline the EM wave propagation characteristics in wireless communication applications.

3.2 EM RADIATION AND ANTENNA PRINCIPLES

In ascertaining the general perspectives of EM radiation from an antenna structure, the salient features of radiation phenomenon can be summarised as follows:

- The functional attribute of an antenna refers to realising desired field strength (specified in terms of **E-** and/or **H**-field intensities) in specified directions. The map of this spatial field strength distribution is known as the *antenna pattern*
- In terms of circuit attributes (of current and voltage), a corresponding power delivery at the antenna can be identified as the "radiated power" from the antenna as excited by the source into the medium or *vice versa*
- The source exciting an antenna (represented either in Thevenin's or Norton's form) has an internal impedance. Correspondingly, an *input impedance* (or *driving-point impedance*) can be specified to characterise the extent that a source excites an antenna in terms of power transfer considerations. The maximum transfer of power

from the source to the antenna (or from the medium through the antenna into a receiver) corresponds to *conjugate impedance matching* (as governed by the *maximum power transfer theorem*) between the source and the load

- In general, the radiating structure, namely the antenna, has a specified bandwidth characteristics
- The ratio of power radiated to the total power supplied by the source refers to the *radiation efficiency*. This efficiency results from the fact that, the total power supplied by a source to the antenna structure is divided into a radiated power component and a nonradiated part constituting dissipation
- The phenomenon of radiation or the excitation of electromagnetic energy as waves into (or from) the surrounding medium is subject to supporting the **E-** and **H**-fields at the antenna structure and solving the associated Maxwell's equations with the relevant boundary conditions (including those that apply to infinity).

3.2.1 Condition for radiation

Given a source coupled to a radiating structure, the question is, what condition should be stipulated for the electromagnetic radiation to occur. Relevant considerations are pertinent to analysing the source and the associated structure (or circuit that radiates) under the constraint that the physical dimension(s) of the circuit (or the structure) are comparable to the wavelength of the associated electromagnetic energy. The underlying consideration behind the size of the radiator versus the wavelength relation relies on two significant features, namely, the *distributed* and the *retardation effects* that characterise the radiated EM wave.

The distributed effect refers to the following: The associated inductive, capacitive and ohmic effects in the circuitry coupled to an antenna from the source (or from the antenna to the receiver) are regarded as *spatially distributed*. That is, the electric and/or magnetic field components, which provide the reactive attributions to the associated circuitry representing the antenna structure can be recognized as being dispersed in space so that, the corresponding inductive and/or capacitive elements do not represent lumped elements. Instead, they denote the dispersed elements of a transmission line. This representation is feasible when the spatial region over which the circuitry (or the associated antenna geometry) is identified as having a dimension comparable to the wavelength of electromagnetic energy prevailing in that region.

The *retardation* indicated amounts to a phase-shift perceived in phase of the observed electromagnetic phenomenon (usually in the associated potential functions) proportional to the distance from the source to the point of observation at which the EM wave is assessed. That is, in reference to the causative mechanism of EM wave radiation, namely a time-varying EM source (of charges and/or currents), the resulting wave at some point of observation cannot be instantaneously sensed. There must be some finite delay or retardation equal to the time it takes a signal generated at the source to travel through the intervening medium with a finite velocity and reach the point of observation in the medium.

3.2.2 Mechanism of EM radiation

EM radiation, as indicated earlier, is the proliferation of an unbounded wave that occurs as a result of discontinuity in a guided-wave transmission system. It takes place in the medium surrounding the locale of discontinuity, subject to the conditions for radiation mentioned above. The Maxwell's equations and the underlying concepts govern the propagation or transmission of EM energy in both regions (namely, the guiding structure and the medium in which the radiation is supported). The associated EM wave motion (in either case) is specified and described by the solution of Maxwell's equations (*via* the Helmoltz wave function). Appropriate boundary conditions prevail in the domain of interest that supports the EM energy. That is, wave transmission through a distributed framework of a guiding structure is decided by the status of electric and magnetic fields at the bounding geometry; and, the radiated wave propagation exhibiting retardation faces degeneration (to zero at infinity) in the propagating medium. Thus, guided-wave transmission (through a transmission line, for example,) is constrained by identifiable boundaries prescribed by the geometrical dimensions (relative to wavelength of EM excitation in the guiding structure). Further, the nature of boundary surfaces (specified in terms of their electric and magnetic properties) stipulate the boundary conditions and hence the type of the mode of guided-wave transmissions (such as TE, TM, TE + TM or TEM) supported. Examples of guided-wave transmissions are those observed in parallel wires, coaxial lines, waveguides, microstrip lines etc.

Should there be a discontinuity on such guided transmission structures, EM leakage would occur at this discontinuity (manifesting as radiation) so as to form a propagating mode of EM energy into the surrounding, which depicts an extended or unbounded medium (in terms of its size relative to wavelength of operation). The guiding surfaces or the boundary of this medium are specified as of infinite extent and the associated wave motion is dictated by the bulk electromagnetic primary constants, namely, σ, μ and ε of the medium. Air (approximately depicting the free-space) in which radio transmissions of wireless communications take place depicts, for example, an unbounded medium.

Typical transition from a guided EM transmission into an unbounded EM wave propagation is illustrated in Figure 3.1. An oscillator (representing a transmitter) excites a transmission line, which supports a guided wave. A discontinuity (abrupt or gradual) at the termination of the line, for example, marks a transition and lets the EM energy from the line to "leak" (or "radiate) into the unbounded free-space. The discontinuity regime or the transition part denotes the "antenna" (and in the configuration depicted in Figure 3.1, it is more appropriately a "transmitting antenna" structure).

The wavelets that emanate from the radiating structure eventually spread into the propagating medium as spherical wavefronts. At a sufficiently large distance (compared to wavelength) from the source, the spherical wavefront can be approximately regarded as a planar wavefront as illustrated in Figure 3.1.

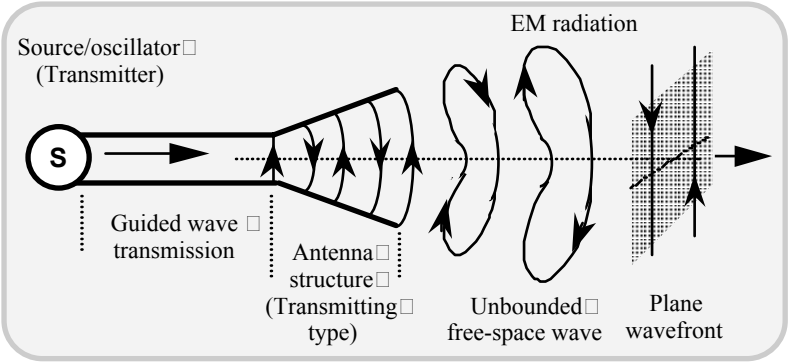


Fig. 3.1 Concept of EM radiation mechanism: Release of bounded EM energy from a guided-wave structure (transmission line) into the free-space by a radiator unit (antenna) as an unbounded wave

The *principle of reciprocity* specifies that the system indicated in Figure 3.1 can also function in reverse towards EM wave *reception*. That is, an EM wave incident at the transition device (namely at the "antenna") could be intercepted and coupled into the guided wave structure. In other words, a "receiving antenna" intercepts an EM wave incident on it and allows it to flow through a guided structure into a receiver.

The modes of EM wave transmission and propagation as indicated in the last chapter are essentially in three classes, namely, *transverse electric* (TE), *transverse magnetic* (TM) and *transverse electromagnetic* (TEM) waves. These are illustrated in Figure 3.2.

The TE wave has the **E**-field always transverse to the direction of propagation and its magnetic (**H**) field has components both transverse and parallel to the direction of propagation. Likewise, the TM wave has the **H**-field always transverse to the direction of propagation and its electric (**E**) field components are both transverse and parallel to the direction of propagation. In TEM mode of propagation, both **E**- and **H**-field components are entirely transverse to the direction of propagation.

In reference to guided-wave propagation restricted by the boundaries of the media supporting the transmission of EM energy (as in transmission lines), it can be of either TE, or TM or TEM mode. Also, it is possible to support a hybrid mode (of TE plus TM combination) on certain structures. In general, the method of excitation, dimensions (relative to wavelength), geometry of the structure and the electrical and magnetic properties of the supporting medium (as well as those of the bounding surfaces) decide the type of the mode supported. For example, a coaxial line can support only a TEM wave, whereas a metallic waveguide can support either a TE- or TM- or a mixed- (TE + TM) mode.

In an unbounded medium in which a radiated field prevails, the mode refers to the TEM wave, often known as the *plane wave*. It is illustrated in Figure 3.1. The EM radiation (emanating from a discontinuity on a guided-wave structure (as shown in Figure 3.1) into the propagating medium assumes a plane-wave form made of spherical wavelets set out from the source. The plane wave is characterised by the

TEM mode in the sense that, at a relatively large distance from the source, the **E**- and **H**-field components of the wave can be regarded to lie in planar surfaces transverse to the direction of propagation.

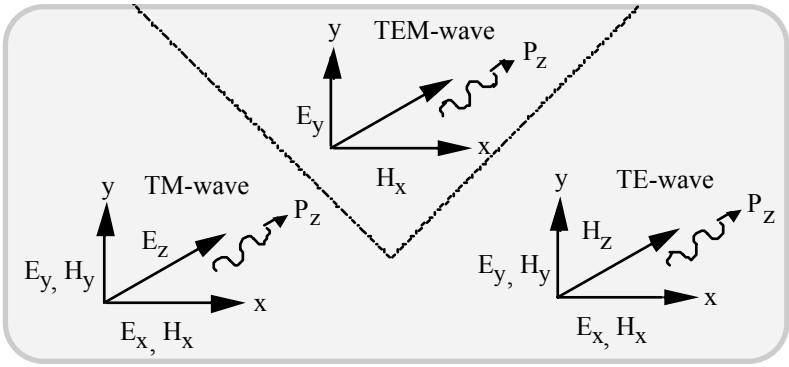


Fig. 3.2 Modes of EM wave transmission illustrating the transverse magnetic (TM), transverse electromagnetic (TEM) and transverse electric (TE) waves specified in the Cartesian co-ordinate system

The characteristics of EM wave transmission or propagation (such as the associated velocity, attenuation, phase constant etc.) are governed by the primary parameters (ε, μ, σ) of the medium as discussed in the last chapter. This is true for both guided and unbounded waves.

Consistent with the radiation and wave propagation considerations indicated above, the scope of this chapter is to outline the concepts of EM radiation and elaborate the principle behind conceiving an antenna/radiating structure of desired functional characteristics. Ideally, an antenna should create sufficiently strong electromagnetic fields at large distances; and, reciprocally, it should be quite sensitive to the electromagnetic fields impressed on it externally. However, in practice, the coupling between a transmitting antenna and a receiving antenna pertinent to a radio link is so small that the signals have to be amplified both at the transmitting and receiving stations.

As indicated before, antennas are usually conceived by modifying transmission lines to work effectively as EM radiators; relevant physical as well as mathematical considerations inculcated to describe the underlying characteristics are aimed at meaningful design formulations pertinent to various antenna geometry.

In practice, the antennas required for various radio communications, such as radio broadcasting and receiving, radar, satellite communication, cellular wireless systems etc. depend on the band of operating wavelengths (from tens of kilometres to a few millimetres), the amount of power radiated or received and the transmission range. For example, large horizontal grids or bimetallic (often steel-aluminium) wires of diameter about 20 mm or more, are used in *long-wave* (LW) transmissions.

Medium-wave (MW) antennas are usually mast or tower antennas. For *short-wave* (SW) transmission, arrays of wire dipoles, single and double rhomboid antennas suspended from towers 60 m high, are deployed. There are also other classes of antennas such as travelling-wave antennas and loop antennas, which are mostly used for receiving purposes. In the microwave and millimetre-wave (mm) range of wavelengths (varying from a few centimetres to millimetres), a plethora of antenna structures have been successfully designed and deployed in radar, satellite and communication systems. Dipole antennas, slot antennas, parabolic and other reflectors, horn antennas, periscopic antennas, helical antennas, spiral antennas, surface-wave and leaky-wave antennas including dielectric and dielectric-loaded antennas, microstrip antennas, etc. are typical examples.

However, for wireless communication purposes, only a select group of antennas can be identified to match the working outdoor and/or indoor ambient. The chapters that follow are written to elaborate the relevant structures. Prior to analysing such radiators, indicated here as a prelude in the following sections are certain preliminary considerations on the quantitative descriptions and qualitative attributes of antennas in general, as well as descriptions on the EM environments of wireless communications. What is presented in this chapter is only a summary on the salient aspects of radiation, antennas and EM wave propagation with the scope limited to modern wireless communication systems. Readers can refer to the vast literature available on this subject-matter. For example, the classical considerations are amply discussed in [3.1 - 3.18] and specific topics related to wireless communications can be found in [3.19 - 3.25].

3.3 ANTENNA PARAMETERS

As indicated earlier, an antenna used as a radiator or receiver of electromagnetic energy may be defined as the transition region between free-space and a guiding structure such as a transmission line. The guiding structure transports electromagnetic energy to or from the antenna. As mentioned earlier, in the former case it refers to a "transmitting antenna" and in the latter case, it is a "receiving antenna".

In either case, the *radiation pattern* is an important property of the antenna. It depicts the pattern of electromagnetic field perceptible at an antenna structure. The power received at a point is a function of the position of the receiving antenna with respect to the transmitting antenna. The graphical plot of the received power at a constant radius from the transmitting antenna is called the *power patt*ern of the antenna and, in essence, it represents a *spatial pattern*. This spatial pattern specified in terms of the electric (or magnetic field) is called a *field pattern*. A cross-section of the field pattern in any particular plane is known as the *radiation pattern* in that plane.

The *power density* is defined as the power per unit area in the field of the antenna. The power density multiplied by the square of the radial distance from the antenna gives the *radiation intensity*, which denotes the power per unit solid angle.

The *directivity* of an antenna is defined as the ratio of the radiation intensity of the antenna in the direction of maximum radiation to the average radiation intensity of the antenna. If $U(\theta, \phi)$ is the radiation intensity in the direction (θ, ϕ), then the directivity D is given by

$$D = 4\pi U_{max}(\theta, \phi) / \int U(\theta, \phi) d\Omega \qquad (3.1)$$

where Ω denotes the solid angle and $d\Omega = \sin(\theta) d\theta d\phi$.

The *gain* of an antenna is defined as the ratio of the maximum radiation intensity of the antenna to the maximum radiation intensity of a reference antenna with the same power input. If the reference antenna is an isotropic source then this ratio is the *absolute* gain of the antenna. If any other reference antenna such as a dipole or a horn is used, then it should be explicitly stated. The antenna gain is the product of the directivity and the radiation efficiency. The directivity can be calculated whereas the gain is usually measured.

As described in Chapter 2, the *polarisation* of an antenna specifies the orientation of the electric field of an antenna. With reference to the earth's surface, the two principal-plane polarisations are *horizontal* and *vertical*. That is, antennas, in general, are designed to radiate or receive either in vertical or in horizontal polarization or, sometimes in other orientation of the electric field as well. These antennas are called *linearly polarized* antennas. Further, it can be recalled from Chapter 2 that an antenna is *elliptically polarised* if it responds to two orthogonal field components with some phase difference. Specifically when the magnitudes of the orthogonal field components are equal with a phase difference of ± 90°, then the antenna is said to be *circularly polarized.* Further, in the case of elliptical or circular polarisation of the wave, if the electric field rotates in the clockwise direction, it is called *right-handed circular polarisation* (RHCP*)* and if it rotates in the counter-clockwise direction it is called *left-handed circular polarisation* (LHCP).

The *impedance* (or more specifically the input or driving-point impedance) of an antenna mentioned earlier, explicitly refers to the following: It is the impedance measured at the antenna terminals with no load attached. It may be defined as the ratio of voltage to current at the terminals or as the ratio of appropriate components of electric and magnetic fields at a point. In general, the antenna impedance is complex.

Relevant to antenna impedance, a parameter of interest refers to the *voltage standing wave ratio* (VSWR) posed by the antenna when it is connected to a feed line leading to the RF electronics that excites the antenna with (or receives from the antenna) an electrical signal. This VSWR consideration arises due to the mismatch condition at the antenna terminals. Its value would change as a function of frequency. For a desirable performance, the VSWR is expected to be close to unity (guaranteeing matching). In practice, the VSWR should be less than 2 for the range of frequency over the bandwidth of interest.

The practical considerations also specify the type of connector through which the antenna is interfaced to the RF electronics. Again, such a connection could be either direct or *via* a pigtail feed. In either case, matching is assured in terms of the VSWR being less than 2.

A structure, when used as a receiving antenna, collects a certain amount of energy from the incident electromagnetic wave. Suppose W is the power in watts delivered to the load impedance at the antenna terminals and P_i is the power density in watts per square meter incident on the antenna. Then, $A_e = W/P_i$ is defined as the *effective aperture* of the antenna. The effective aperture depends on the load

impedance and the polarisation as well as the direction of the incident wave. In the special case where polarisation of the antenna is identical to the incident EM wave and the incoming wave is in the direction of maximum directivity of the antenna, then A_e is given by the following expression (under matched impedance considerations):

$$A_e = \lambda^2 D / 4\pi \qquad (3.2)$$

where D is the directivity and λ is the wavelength. (The wavelength as discussed in Chapter 2, is expressed in metres. It is equal to c/f, where $c = 3 \times 10^8$ m/s denotes the velocity of propagation of EM energy in free-space and f is the frequency in Hz.) For large aperture antennas, the effective aperture is of the same order as the physical aperture. For antennas of finite aperture size (comparable to λ^2) such as pyramidal horns, the effective aperture is about 50 to 80 per cent of the physical aperture, and for parabolic reflector antennas it is about 50 to 65 per cent of the physical aperture.

When the antenna is designed, due consideration should be given to the RF power that it handles. Normally, both *average and peak powers* are furnished in the specifications.

The gain of an antenna, defined above, specifies the ratio of RF power-in to the power-out, usually measured in *decibels* (dBi) with the index i depicting explicitly the relative gain with respect to an isotropic antenna. (The other reference used in practice is the dipole antenna, in which case the decibel is denoted as dBd with the index d depicting the dipole radiator.)

The difference in the input and output RF powers stems from finite losses encountered in the antenna structure. Such losses are governed mostly by the reflective losses caused by mismatched conditions at the antenna terminal plus ohmic losses in the structure itself.

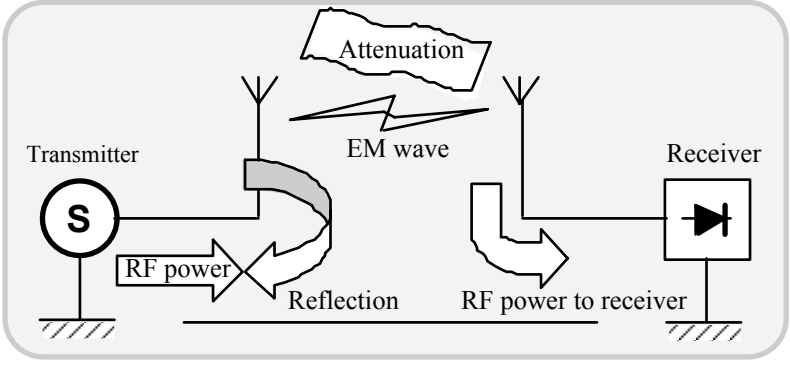


Fig. 3.3 Loss-budget of electric signal in a typical wireless system: The reflection loss corresponds to impedance mismatch conditions and the attenuation shown is medium-dependent

During propagation, the signal in the radio channel is invariably attenuated as a result

of the EM energy spreading out as well as due to the medium having lossy material attributes. Shown in Figure 3.3 is the loss-budget encountered in typical wireless communication scenarios.

In loss-budget calculations, many times, the antenna gain is either misunderstood or misinterpreted in reference to mobile measurement contexts. Understanding the true meaning of "antenna gain" can help improve the accuracy of the so-called "handset measurements" and eventual use of the measured gain in loss-budget evaluations [3.26].

Basically, the antenna gain as indicated above, is the term used to relate the *isotropic radiative power* (IRP) of a particular antenna to the peak radiated power of other "standard" antennas such as dipole or monopole structures. At times, the industry may lump losses and compensation factors together with antenna gain in order to relate the response of the physical RF connector to that of the antenna. This (mis)interpretation often poses confusion while trying to accurately determine the response of a mobile telephone through the antenna. The precaution consideration thereof is as follows.

In all simplicity, the antenna gain refers to the radiated power of the maximum radiated power density at a particular point from the antenna to the power density radiated by a reference antenna (typically the isotropic radiator) at the same point as illustrated in Figure 3.4.

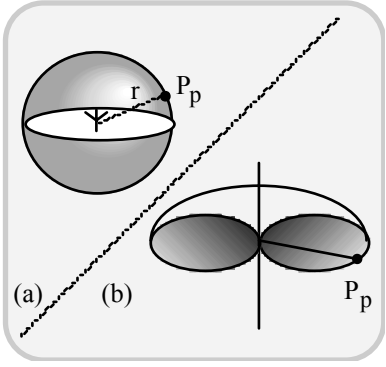


Fig. 3.4 Antenna gain specification with respect to an isotropic radiator: (a) Isotropic radiation pattern and (b) directed radiation pattern

Denoting the isotropic power at a particular point at a distance r as P_p, it is equal to $P_r/4\pi r^2$ with P_r being the radiated power. For example, if P_r = 100 W and r = 100 m, then P_p = 0.795 mW/m^2, which in dB measure is equal to – 0.996 dBm/m^2. As will be indicated in the next chapter (Chapter 4), the antenna gain for a short dipole antenna is $P_{dipole}/P_{isotropic}$ = G = k D, where D is the directivity equal to 1.64 and k = $R_r/(R_r + R_\ell)$. Here, k is the antenna efficiency, R_r denotes the radiation resistance and R_ℓ is the ohmic losses in the antenna. For k = 1, G = 1.64 or 2.15 dB for a dipole antenna. It means that an equivalent power transmitted by an isotropic antenna presumed to be at

the same point as the test dipole is $P_p = 1.64 \times 0.795 = 1.3$ mW/m^2 = 1.14 dBm. As mentioned before, the antenna gain is normally assessed by measurements. Suppose such a measurement is applied to a mobile telephone unit. Often, in such mobile telephones, the dipole structure used is constituted by a monopole (as one part of the dipole) and the body of the mobile phone, as the other part.

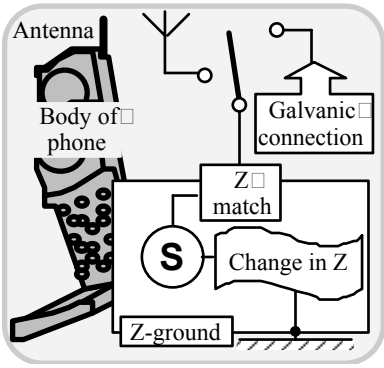


Fig. 3.5 Mobile station under test (MSUT) with the RF path at the receiver switched between the antenna and a galvanic connection

The impedance of this antenna (specified by the radiation resistance) is matched to the RF front-end of the mobile phone circuit for maximum power transfer. In this configuration, suppose the *mobile station under test* (MSUT) is evaluated for its performance, including its antenna gain. In such measurements, the RF path is switched between the antenna and the galvanic connection as shown in Figure 3.5.

The galvanic connection is improvised with characteristic impedance of 50 ohms and interfaced with standard test equipment through 50-ohm RF cables. The RF characterisation of this galvanic connection, in real-world situations, do not replicate that of the RF connection of mobile transmission through an antenna coupled to the free-space (with η_o = 377 ohms). For optimum coupling of the dipole between the free-space and the RF front-end, the necessary matching corresponds to the matching network interposed having an input impedance equal to the radiation resistance of the dipole, namely R_r = 76 ohms (as will be derived in the next chapter). However, in practice, the equivalent impedance of a mobile unit may not exactly correspond to that of R_r. This is because the mobile unit has always other bodies (such as a human hand, head or automobile-specific mountings etc.) in its proximity, which would reactively couple it to the radiation section altering the value of R_r. Such coupling is also subject to change dynamically. As such, the controlled environment measurement of the gain may not be the true representation of the actual "gain" that a mobile unit presents in the real world condition. Suppose the measurement is performed with the galvanic connection indicated in Figure 3.5, and performance calibration on the gain of the unit is specified thereof the results obtained may not reflect the actual implications and

therefore, a correction factor is needed on the calibrated data. Such a correction factor should explicitly account for any antenna *versus* galvanic connection inconsistencies. In [3.26], the suggested term for this correction coefficient is "galvanic-to-antenna port compensation factor". Thus, the classical definition and measurement of antenna gain should be carefully interpreted and adopted in the case of mobile antenna systems.

3.4 EM FIELDS IN AN UNBOUNDED MEDIUM

Relevant to the discussions above, the field components in an unbounded medium, as described in Chapter 2, can be described by the parameters (ε, μ, σ). Further, concerning practical wireless communication links, the medium can be assumed to be lossless and charge-free. Also, the medium is invariably linear, homogeneous, and isotropic. Hence, the field produced by a distribution of charges or currents in an infinite homogeneous medium depicting the ideal "free-space" attributes of the radio channels *vis-à-vis* wireless applications can be evaluated using the concepts of vector and scalar potential functions as described below.

3.4.1 Vector and scalar wave potentials

When the source (or impressed) currents are known throughout an infinite homogeneous medium, the field calculation is fairly straightforward. That is, when the field components of a current element are known, then by using the *principle of superposition*, the radiated field due to any such current distribution on the element can be ascertained.

Considering the electromagnetic equations for harmonic fields, namely, curl ($\mathbf{E}$) $= - j\omega\mu\mathbf{H} - \mathbf{M}^i$ and curl ($\mathbf{H}$) $= (\sigma + j\omega\varepsilon)\mathbf{E} + \mathbf{J}^i$, upon solving these equations, the solution of the most general case can be expressed in the form of a contour integral in the oscillation constant plane. If $\mathbf{M}^i$ is a continuous and differentiable function, then it is possible to eliminate $\mathbf{E}$ from the curl relations indicated above. Hence, it is possible to obtain a second-order differential equation for $\mathbf{H}$. Similarly, if $\mathbf{J}^i$ is a continuous and differentiable function, one can eliminate $\mathbf{H}$ from those equations. Hence, a second-order differential equation for $\mathbf{E}$ can be derived thereof [3.1]. However, in practical problems, $\mathbf{J}^i$ and $\mathbf{M}^i$ are localised and the regions occupied by them have sharp boundaries. When this happens, the differential equations have to be solved by introducing a set of auxiliary functions, generally known as the *potential functions*. Hence, designating $\mathbf{E} = \mathbf{E}' + \mathbf{E}''$ and $\mathbf{H} = \mathbf{H}' + \mathbf{H}''$, the sets ($\mathbf{E}'$, $\mathbf{H}'$) and ($\mathbf{E}''$, $\mathbf{H}''$) respectively could form the solutions of,

$$\nabla \times \mathbf{E}' = -j\omega\mu\mathbf{H}' \tag{3.3a}$$

$$\nabla \times \mathbf{H}' = (\sigma + j\omega\varepsilon)\mathbf{E}' + \mathbf{J}^i \tag{3.3b}$$

and

$$\nabla \times \mathbf{E}'' = - j\omega\mu\mathbf{H}'' - \mathbf{M}^i \tag{3.4a}$$

$$\nabla \times \mathbf{H}'' = (\sigma + j\omega\varepsilon)\mathbf{E}'' \tag{3.4b}$$

The field (**E**', **H**') results from the electric current $\mathbf{J}^i$ and the field (**E**", **H**") are due to the magnetic current $\mathbf{M}^i$. The sum of these fields also satisfies the curl relations and hence it can be presumed to be caused by $\mathbf{J}^i$ and $\mathbf{M}^i$.

Taking the divergence of each of the equations (3.3) and (3.4), one obtains the following set of relations:

$$\left.\begin{aligned} \nabla.\mathbf{H}' &= 0 \\ \nabla.\mathbf{E}' &= -\nabla.\mathbf{J}'/(\sigma + j\omega\varepsilon) \\ \nabla.\mathbf{H}'' &= -\nabla.\mathbf{M}'/j\omega\mu \\ \nabla.\mathbf{E}'' &= 0 \end{aligned}\right\} \qquad (3.5)$$

For valid stipulations of equations (3.5), the entities $\mathbf{J}^i$ and $\mathbf{M}^i$ are required to be continuous and differentiable. But, it is possible to obtain one form of the solution to the problem in hand by circumventing the conditional relations of equations (3.5). (In the other form of the solution that depends on these equations, one may assume to begin with that $\mathbf{J}^i$ and $\mathbf{M}^i$ are differentiable and then extend the results to include any discontinuous distributions.)

The set of equations (3.5) shows that **H**' and **E**" can be represented as the curls of some *vector point functions*, **A** and **F**. That is,

$$\nabla \times \mathbf{A} = \mathbf{H}' \qquad (3.6a)$$

$$-\nabla \times \mathbf{F} = \mathbf{E}'' \qquad (3.6b)$$

Substituting equations (3.6) in equations (3.3a and 3.4b), it follows that,

$$\left.\begin{aligned} \mathbf{E}' &= -j\omega\mu\mathbf{A} - \nabla V \\ \mathbf{H}'' &= -(\sigma + j\omega\varepsilon)\mathbf{F} - \nabla U \end{aligned}\right\} \qquad (3.7)$$

where V and U are two new *scalar point functions*, which are introduced inasmuch as the equality of the curls of the two vectors does not imply that the vectors are identical. From equations (3.7, 3.3b and 3.4b), the following results are obtained:

$$\left.\begin{aligned} \nabla \times \nabla \times \mathbf{A} &= \mathbf{J}^i - \gamma^2\mathbf{A} - (\sigma + j\omega\varepsilon)(\nabla V) \\ \nabla \times \nabla \times \mathbf{F} &= \mathbf{M}^i - \gamma^2\mathbf{F} - (j\omega\mu)(\nabla U) \end{aligned}\right\} \qquad (3.8)$$

Using the vector identity, namely, $\nabla \times \nabla \times \mathbf{A} = -\nabla^2\mathbf{A} + \nabla(\nabla.\mathbf{A})$, one can get the following relations:

$$\left.\begin{aligned}\nabla^2\mathbf{A} &= \nabla(\nabla.\mathbf{A}) = -\mathbf{J}^i - \gamma^2\mathbf{A} - (\sigma + j\omega\varepsilon)(\nabla V)\\ \nabla^2\mathbf{F} &= \nabla(\nabla.\mathbf{F}) = \mathbf{M}^i - \gamma^2\mathbf{F} - (j\omega\mu)(\nabla U)\end{aligned}\right\} \qquad (3.9)$$

Thus, it is possible to express **E** and **H** in terms of the vectors **A** and **F** and the scalars V and U *via* the new functions (called *auxiliary functions*) interrelated to each other as above. Moreover, if one adds the gradients of arbitrary functions to **A** and **F**, then equations (3.6a) and (3.6b) remain unchanged. Hence, the choice of the pairs (**A** and **F)** and (V and U) is arbitrary. Therefore, as a matter of convenience, it is possible to impose further conditions on these functions. For instance, by substituting $V = -\nabla.\mathbf{A}/(\sigma + j\omega\varepsilon)$ and $U = -\nabla.\mathbf{F}/j\omega\mu$, equations (3.9) reduce to,

$$\left.\begin{aligned}\nabla^2\mathbf{A} &= -\mathbf{J}^i - \gamma^2\mathbf{A}\\ \nabla^2\mathbf{F} &= \mathbf{M}^i - \gamma^2\mathbf{F}\end{aligned}\right\} \qquad (3.10)$$

When the auxiliary functions, namely, (**A**, **F**, V and U) are specified by the relations $V = -\nabla.\mathbf{A}/(\sigma + j\omega\varepsilon)$ and $U = -\nabla.\mathbf{F}/j\omega\mu$, they are called *wave potentials*. The entity **A** is known as the *magnetic vector potential* and **F** denotes the *electric vector potential*. Further, V is the *electric scalar potential*, and U is known as the *magnetic scalar potential*. H. A. Lorentz, while considering electromagnetic wave propagation in nondissipative (lossless) media, introduced these wave potentials for the first time and called them *retarded potentials* [3.1]. These wave potentials are not only retarded but also attenuated, and, therefore, the term wave potentials is appropriate in the context of EM radiation and propagation problems.

In terms of these wave potentials **A**, **F**, V and U, the field vectors **E** and **H** can be expressed as follows:

$$\mathbf{E} = \mathbf{E}' + \mathbf{E}'' = -j\omega\mu\,\mathbf{A} - \nabla V - \nabla\times\mathbf{F} \qquad (3.11a)$$

$$\mathbf{H} = \mathbf{H}' + \mathbf{H}'' = \nabla\times\mathbf{A} - \nabla U - (\sigma + j\omega\varepsilon)\mathbf{F} \qquad (3.11b)$$

If $\mathbf{J}^i$ and $\mathbf{M}^i$ are differentiable functions, V and U satisfy equations similar to those satisfied by **A** and **F**. Taking the divergence of equations (3.10) and using the inter-relations, namely $V = -\nabla.\mathbf{A}/(\sigma + j\omega\varepsilon)$ and $U = -\nabla.\mathbf{F}/j\omega\mu$, the following equations are obtained for nondissipative media;

$$\nabla^2 V = -\beta^2 V - q_v/\varepsilon \qquad (3.12a)$$

$$\nabla^2 U = -\beta^2 U - q_m/\mu = -\beta^2 U \qquad (\text{since } q_m = 0) \qquad (3.12b)$$

where q_v and q_m depict the volumetric electric and magnetic charge densities respectively.

3.5 CURRENT ELEMENT AS A RADIATOR

The radiation phenomenon associated with an antenna structure can be viewed as the result of the causative mechanism of time-varying current distribution in the geometry of the radiator. Suppose an infinitesimally small, electric current element, specified by the product of an elemental length ($d\ell$) and the current (I) that flows through, namely, ($Id\ell$) represents a radiating structure (as illustrated in Figure 3.6). In this section we attempt to determine the radiation field components due to this current-element. In the following chapter, a relevant strategy to deduce the radiation characteristics of a host of practical structures is presented.

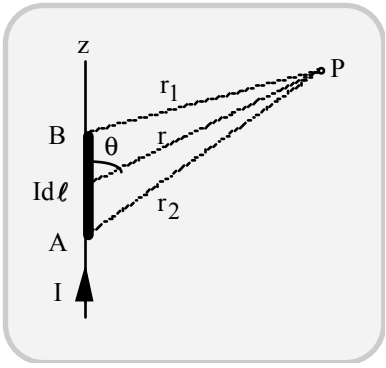


Fig. 3.6 Electric current element. (P is a remote point of observation specified by the co-ordinates (r, θ) in space)

For the purpose of analysis, the current I between the terminals A and B in Figure 3.6 is forced to flow out of B into the external medium and returned to A. As such, if the external medium were a perfect dielectric, then there would be an aggregation of positive electric charges at B and negative electric charges at A at any given instant of time. Under time-varying conditions, this conglomeration of charges at A and B would occur at a rate of I coulombs per second; this would eventually result in a growing electric field around the filament of the current element under consideration.

The *moment* of the electric current element is defined as the product of the current and length of the filament. If the flow of current is along the z-axis and presumed to be centred at the origin of the co-ordinates (as shown in Figure 3.6), then the current density $\mathbf{J}_S$ at a point of observation (P) on a surface S, situated at a radial distance r from the origin, is given by

$$\mathbf{J}_S = \mathbf{a}_r J_r = \mathbf{a}_r I / 4\pi r^2 = \mathbf{a}_r \partial(-I / 4\pi r)\partial r \qquad (3.13)$$

where $\mathbf{a}_r$ is the unit vector along r.

Assuming that the current element consists of two point sources namely, + I at A and – I at B, the current density at a distant point P ($r >> \ell$) is the gradient of the following function:

$$\begin{aligned} f(I, r, \theta) &= (-I/4\pi r_1) + (+I/4\pi r_2) \\ &\quad + (I/4\pi)\{-1/[r - (\ell/2)\cos(\theta)] + /[r + (\ell/2)\cos(\theta)]\} \\ &\approx -I\ell\cos(\theta)/4\pi r^2 \end{aligned} \tag{3.14}$$

(since $r >> \ell$). Hence, the following current density components of $\mathbf{J}_S$ can be deduced:

$$J_r = \partial f/\partial r = I\ell\cos(\theta)/2\pi r^3 \tag{3.15a}$$

$$J_\theta = \partial f / r\partial\theta = I\ell\sin(\theta)/2\pi r^3 \tag{3.15b}$$

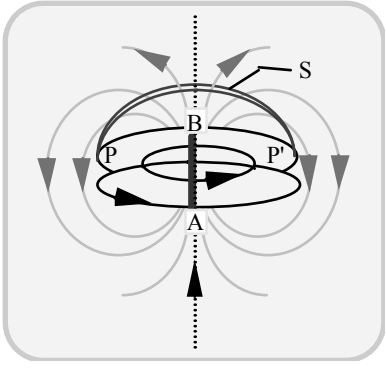


Fig. 3.7 Flux lines representing the current flow (grey lines) and magnetic field (dark lines) in the vicinity of a current element

In Figure 3.7, shown are current flux lines along with the magnetic flux lines. The grey lines represent the current-flow lines and the dark lines show the magnetic field lines, which are coaxial with the element.

The magnetomotive force around the circumference of a circle PP' coaxial with the element is equal to the total electric current I(θ) flowing through any surface S bounded by the circumference of this circle. If S is chosen as part of a sphere concentric with the origin, as shown in Figure 3.7, then it follows that,

$$I(\theta) = \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\theta} J_r r^2 \sin(\theta) d\theta d\phi$$

$$= I\ell \sin^2(\theta)/2r \qquad (3.16)$$

Further, since $I(\theta) = H_\phi 2\pi r \sin(\theta)$, $H_\phi = (I\ell)\sin(\theta)/4\pi r^2$ and inasmuch as $\mathbf{H} = \nabla \times \mathbf{A}$ and $\nabla^2 \mathbf{A} = \gamma^2 \mathbf{A} - \mathbf{J}^i$, the component A_z of $\mathbf{A}$ depends only on the z-component of the source current $\mathbf{J}^i$; and, the current element specified in the z-direction $\mathbf{A}$ has only a z-component. Hence, the following relations can be deduced:

$$H_\phi = -\partial A_z/\partial\rho = I\ell \sin(\theta)/4\pi r^2 \qquad (3.17)$$

Further, since $r^2 = \rho^2 + z^2$ and $\sin(\theta) = \rho/r = \rho/(\rho^2 + z^2)^{1/2}$, the relation given by $\partial A_z/\partial\rho = -I\ell \sin(\theta)/4\pi r^2$ leads to: $A_z = I\ell/4\pi r$. Suppose the current is an harmonic function of time. That is, $I = I_o \exp(j\omega t)$. As ω approaches zero, the field of the current element must still be consistent with the relations for H_ϕ and A_z indicated above corresponding to a steady current element. Hence, the magnetic vector potential $\mathbf{A}$ should satisfy $\nabla^2 \mathbf{A} = \gamma^2 \mathbf{A}$ at all points external to the current element; since $\mathbf{A} = \mathbf{a}_z A_z$ (being independent of θ and ϕ), $\mathbf{A}$ satisfying the Poisson equation yields a solution for A_z as:

$$A_z = G_1 \exp(-\gamma r)/r + G_2 \exp(+\gamma r)/r \qquad (3.18)$$

The term $[G_2 \exp(+\gamma r)]/r]$ is, however, inadmissible since it would increase exponentially with the distance r from the element and is inconsistent of being a physically realisable field. Hence, only the first term, namely $[G_1 \exp(-\gamma r)]/r]$ is retained to represent A_z. In a nondissipative medium $\gamma = j\beta$ and hence $\mathbf{A} = \mathbf{a}_z A_z \equiv \mathbf{a}_z [I\ell \exp(-j\beta r)]/4\pi r]$, since $G_1 = I\ell/4\pi$, as deduced from the steady-state relation, namely $[G_1 \exp(-\gamma r)]/r]_{\text{steady-state}} \equiv A_z = I\ell/4\pi r$.

From the expression derived for A_z, it is possible to determine the associated **E**- and **H**-field components. Inasmuch as $V = -\nabla.\mathbf{A}/(\sigma + j\omega\varepsilon)$, the potential at P is given explicitly by,

$$V = -[1/(\sigma + j\omega\varepsilon)]\partial A_z/\partial z$$

$$= (\eta I\ell/4\pi r)(1 + 1/\gamma r)\exp(-\gamma r)\cos(\theta) \qquad (3.19)$$

where $\eta = [j\omega\mu/(\sigma + j\omega\varepsilon)]^{1/2}$. Now, the following relations can be used to determine **E** and **H**:

$$\mathbf{E} = -j\omega\mu\mathbf{A} - \nabla V \tag{3.20a}$$

$$\mathbf{H} = \nabla \times \mathbf{A} \tag{3.20b}$$

whose relevant components are given by:

$$\left.\begin{aligned} E_r &= (\eta I\ell / 2\pi r^2)(1 + 1/\gamma r)\exp(-\gamma r)\cos(\theta) \\ E_\theta &= (j\omega\mu I\ell / 4\pi r)[1 + (1/\gamma r) + (1/\gamma^2 r^2)]\times \\ &\quad \exp(-\gamma r)\sin(\theta) \\ H_\phi &= (\gamma I\ell / 4\pi r)(1 + 1/\gamma r)\exp(-\gamma r)\sin(\theta) \end{aligned}\right\} \tag{3.21}$$

where the (1/r) term is known as the *radiation field* term. Further, the ($1/r^2$) term depicts the *induction field* and the ($1/r^3$) term represents the so-called *electrostatic field*. At large distances, only the (1/r) term predominates and specifies the following radiated field components:

$$\left.\begin{aligned} E_r &= 0 \\ E_\theta &= \eta H_\phi \\ H_\phi &= (jI\ell / 2\pi r)\exp(-j\beta r)\sin(\theta) \end{aligned}\right\} \tag{3.22}$$

Similarly, the field components of a magnetic current element $K\ell$ can be obtained from the electric vector potential $\mathbf{F} = \mathbf{a}_Z F_Z = \mathbf{a}_Z K\ell\exp(-\gamma r)/4\pi r$.

3.5.1 Radiation from electric current element

The total power W radiated from an electric current element is calculated as the total flow of power across a large sphere concentric with the element. That is,

$$\begin{aligned} W &= \iint_{\text{sphere}} P_r\, dS = (1/2)\int_{\phi=o}^{2\pi} d\phi \int_{\theta=0}^{\pi} E_\theta H_\phi^* r^2 \sin(\theta)\, d\theta \\ &= (1/3)\pi\eta(I\ell/\lambda)^2 \end{aligned} \tag{3.23}$$

where P_r is the radial component of the Poynting vector. The radiated power W, in free-space, becomes equal to $40\pi^2(I\ell/\lambda_o)^2$. In the radiated field regime only the (1/r) term contributes the power. The ($1/r^2$) and ($1/r^3$) terms contribute to the reactive power that is exchanged back and forth between the source and the medium.

The radiated power can also be calculated from the work done by the electromotive force impressed on the current element. Further, the electric intensity on the axis of

the element in a nondissipative medium is given by,

$$E_z = E_r = (\eta I \ell / 2\pi)[(1/j\beta r^3) - (j\beta / r) - (\beta^2 / 3) - ...] \qquad (3.24)$$

The first two terms in equation (3.24) are in quadrature with I and on average do not do any work; but the third term is 180° out of phase with I, and work is done against the field by the impressed electromotive force. Then, the in-phase component of this force field is given by:

$$\mathrm{Re}(V^i) = -\ell\, \mathrm{Re}(E_z) = 2\pi\eta \ell^2 I / 3\lambda^2 \qquad (3.25)$$

The work resulting from this force field per second is $2\pi\eta \ell^2 I^2 / (2 \times 3\lambda^2)$, which is equal to the total power radiated, W. The ratio $R = \mathrm{Re}(V^i)/I = 80\pi^2(\ell/\lambda_o)^2$ (when specified for the free-space) denotes the *radiation resistance* of the current element in free-space.

3.5.2 EM field produced by a given distribution of applied electric and magnetic currents

Any given distribution of applied electric or magnetic currents can be subdivided into elements, and the resulting field is obtained by the superposition of the fields of individual elements. If an infinitesimal volume bounded by the lines of flow and two surfaces normal to them are considered, then the current I in this element is $\mathbf{J}^i dS$, where dS is the cross-section of the tube of current flow. If $d\ell$ is the length of this tube, then the moment $\mathbf{I}d\ell$ of the current element is $\mathbf{J}^i dS d\ell = \mathbf{J}^i dv$, where $dv = (dS d\ell)$ is the volume of the element illustrated in Figure 3.8.

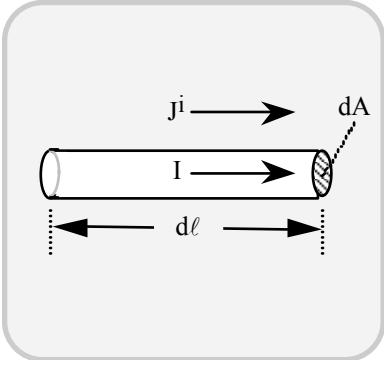


Fig. 3.8 Infinitesimally small current element: It is a representative of an elemental (filamentary) current-carrying conductor

The magnetic vector potential due to the electric current distribution is given by

$$\mathbf{A} = \iiint_v (1/4\pi r)\mathbf{J}^i \exp(-\gamma r)dv \qquad (3.26)$$

where r is the distance between a typical current element and a point in space. It is taken over the volume over which the current is distributed. Similarly, for a magnetic current distribution, the electric vector potential is

$$\mathbf{F} = \iiint_v (1/4\pi r)\mathbf{M}^i \exp(-\gamma r)dv \qquad (3.27)$$

From **A** and **F** as above, the scalar potentials V and U can be evaluated and hence **E** and **H** can be ascertained.

3.5.3 EM field due to impressed currents varying arbitrarily with time

In nondissipative media, the magnetic vector potential of a given electric current distribution varying harmonically with time, that is, varying as exp(jωt), is given by,

$$\mathbf{A} = \iiint_v (1/4\pi r)\mathbf{J}^i \exp(-j\beta r)\exp(j\omega t)dv \qquad (3.28)$$

The phase of $\mathbf{J}^i$ is ωt, and the phase of the corresponding component of the retarded vector potential is $\omega t - \beta r = \omega(t - r/v_m)$, where $v_m = \omega/\beta$ depicting the characteristic velocity of the medium. The time-delay r/v_m is independent of the frequency; hence, all frequency components of a general function $\mathbf{J}^i(x, y, z; t)$ are shifted equally on the time scale, and **A** will depend on $\mathbf{J}^i(x, y, z; t - r/v_m)$. Thus, one has,

$$\mathbf{A}(x, y, z; t) = \iiint_v (1/4\pi r)\mathbf{J}^i(x, y, z; t - r/v_m)dv \qquad (3.29)$$

3.5.4 Field of electric current element whose current varies arbitrarily with time

Consider an electric current element of length ℓ along the z-axis located at the origin, and suppose that current I(t) is given by,

$$\begin{aligned} I(t) &= 0 \quad \text{for} \quad t \le 0 \\ &= I(t) \quad \text{for } t > 0 \end{aligned} \qquad (3.30)$$

Let dI/dt be finite. The charge q(t) at the upper end is zero when $t \le 0$, and is given by

$$q(t) = \int_0^t I(t)dt \quad \text{for } t > 0 \qquad (3.31)$$

and at the lower end the charge is equal to – q(t).

The corresponding field of the current element can now be evaluated. Both **E** and **H** consist of three parts, namely, (**E**', **H**') that depends only on the time-derivative of the current, (**E**", **H**") which depends on the current alone, and (**E**"', **H**"') components relying on the charge present. Thus, **E** = (**E**' + **E**" + **E**"') and **H** = (**H**' + **H**" + **H**"') with the following being their explicit components:

$$\left.\begin{aligned} E'_\theta &= \eta H'_\phi, H'_\phi = [\ell I(t - r/v)\sin(\theta)]/4\pi vr \\ E'_r &= 0 \\ E''_\theta &= \eta H''_\phi, H''_\phi = [\ell I(t - r/v)\sin(\theta)]/4\pi r^2 \\ E''_r &= 2E''_\theta \cos(\theta) \\ E'''_\theta &= [\ell q(t - r/v)\sin(\theta)]/4\pi\varepsilon r^3 \\ E'''_r &= [\ell q(t - r/v)\cos(\theta)]/2\pi\varepsilon r^3 \end{aligned}\right\} \qquad (3.32)$$

At sufficiently large distances, as stated earlier, only the radiation field (**E**', **H**') is significant. At this distance, the entire field outside the spherical surface of radius vt, with its centre at the element, is zero. This sphere is the *wavefront* of the wave radiated by the element; and on this wavefront, both (**E**", **H**") and (**E**"', **H**"') would vanish. Further, at the wavefront, **E** and **H** are perpendicular to the radius (or are tangential to the wavefront).

The above EM field considerations with relevance to the radiation mediated by a current element, can be extended to conceive various antenna structures as will be deliberated in the next chapter (Chapter 4).

3.6 EM WAVE PROPAGATION MODELS OF WIRELESS COMMUNICATION CHANNEL

The physical medium of wireless communication, namely the "free-space" radio channels, could place adverse influences on the performance of mobile communication systems. This is because the paths of EM waves between the transmitter and the receiver, though theoretically an unobstructed "line-of-sight" (LoS) propagation in "free-space", are invariably obstructed and shadowed by the buildings and natural foliage of the outdoor environment. Likewise, in an indoor there exists a variety of obstacles and reflectors (such as walls, floor, ceiling and furniture) to EM propagation that would impair the signal transmissions [3.27 - 3.29].

In a mobile environment (where the receiver may be moving at an unspecified speed), the ambient of associated obstructions could strongly influence the variation (or *fading*) of the signal intensity at the receiver's end. This is in contrast to wireline systems where the channel characteristics are invariant,

predictable and mostly amenable for analysis; whereas, the wireless communication channels are stochastical in nature as regard to their influence on EM waves propagating between the transmitter and receiver(s). The underlying, nondeterministic considerations pose difficulties in predicting the signal-strength variations at the receiving end. Therefore, it has become a common practice to "model" the propagation of EM wave in the wireless communication scenario. These models are useful in designing the link budget considerations as well as help the designers to choose appropriate hardware and software support for the system. On the hardware side, the signal-processing at baseband level and RF signal conditioning at the front-end largely benefit from the heuristics of these models. The antenna being a part of the front-end system, it has a significant disposition in such channel models.

Channel models are done *via* analytical tools and numerical methods with empirical data collected from field measurements. Nevertheless, such models should be consistent with the basic mechanism of EM wave propagation in the ambient in question, pertinent to:

EM wave parameters

- Mode of propagation (invariably, the TEM mode)
- Polarization (linear or circular)
    - Directional attributions/spatial scanning, if deployed
    - Frequency spectrum of operation/frequency-scanning, if adopted

EM propagation considerations

- LoS transmissions under micro- and/or macro-cell environments
- Single and multiple reflection of the wave (by horizontal ground and vertical structures)
- Diffraction of the wave and shadowing effects at the obstacles
- Scattering of the wave by the illuminated objects
    - Multipaths (of varying length) of the reflected waves
    - Distance between the transmitter and receiver

In view of the above considerations, the wireless communication signals (especially in the cellular urban environment and in indoor applications) could suffer severe diffraction losses as a result of destructive interference of multipath components at the receiver leading to time-varying signal fading and fall in signal intensity as the transreceiver separation increases.

Relevant propagation models perceived thereof aim at predicting the average signal strength at a given distance from the transmitter and the time-dependent profile of such signal-strength variations in spatial proximity to a particular location. Such average (stochastical) description of the signal for an arbitrary transreceiver separation is meaningful in deciding the radio coverage area as well as in the design of wireless systems *vis-à-vis* antenna diversity and coding requirements.

The propagation models, in general, can be classified as follows:

- *Large-scale models* depicting the signal strength characteristics over extensive transreceive separations (of the order of thousands of wavelengths) observed over a long period of time
- *Small-scale models* representing short-time, rapid fluctuation of the signal strength over short regimes of transreceive separations.

The modeling strategies and the underlying considerations are briefly addressed in the following sections.

3.7 OUTDOOR EM WAVE PROPAGATION MODELS

3.7.1 EM propagation in a simple LoS link

Assuming the transreceive path is ideally devoid of any obstacles and the transreceive transmissions occurs over a line-of-sight (specular) ray, the EM wave propagation is then dictated purely by "free-space" transmission principles. (It should be noted that this ideal condition could apply only in satellite communications and/or in a very short line-of-sight radio links.)

For a transmitted power P_t by an antenna of gain G_t, the receiving antenna (with a gain G_r) kept at a line-of-sight distance d from the transmitter (in free-space) will intercept a power P_r given by:

$$P_r = \left(\frac{P_t \lambda^2}{4\pi d^2}\right)\left(\frac{G_t G_r}{4\pi}\right)\left(\frac{1}{L}\right) \tag{3.33}$$

where L denotes the system loss and λ is the wavelength.

The corresponding *free-space path loss* (FSPL) in dB is given by:

$$PL = -\log_{10}\left(\frac{P_r}{P_t}\right) = -10\log_{10}\left[\frac{G_t G_r}{(4\pi)^2}\cdot\left(\frac{\lambda^2}{d^2}\right)\cdot\frac{1}{L}\right] \tag{3.34}$$

The free-space model as above accounts for the path loss due to the transreceiver separation for a set of antennas at the transmitter and receiver of specified gains and the system impaired by an associated loss of $L > 1$. The corresponding received power decays (falls off) at a rate of 20 dB per decade with distance.

The above model, however, depicts an "ideal" non-obstruction situation and/or multipath-free spatial considerations. In practice, the relevant modeling should be modified to include one or more of the adjoining EM-related mechanisms of the wave propagation, namely, *reflection*, *diffraction* and *scattering*.

3.7.2 Reflection-specified propagation model

Reflection of an EM wave is governed by the interface boundary conditions, whenever the wave faces an obstruction in its free-space path. The dimensions of the obstruction/interface are assumed to be very much larger than the wavelength of incident radiation. The reflection phenomenon, in general, can be characterized by the following considerations:

- The incidence of EM wave on the surface of the obstacle could be normal (perpendicular) or oblique
- The polarisation of such a wave could be *horizontal* implying that associated **E** vector is parallel (tangential) to the surface of incidence; or, it could be *vertical* meaning that the **H**-vector of the wave is parallel (or tangential) to the surface of incidence
- The electrical characteristics of the obstacle could be such that the obstacle is a pure dielectric medium characterised by its (real) dielectric constant ε_r. If this medium is lossy, then its permittivity is complex (as indicated in Chapter 2) with a complex dielectric constant specified by $(\varepsilon'_r - j\varepsilon''_r)$; and, the loss-tangent $\tan(\delta) = (\varepsilon''_r/\varepsilon'_r)$ depicts the lossy nature of the medium whose conductivity σ is and the frequency f implicitly decide the term ε''_r. That is, $\varepsilon_o\varepsilon''_r = (\sigma/\omega\varepsilon'_r\varepsilon_o)$ where ε_o is the free-space permittivity equal to $(1/36\pi)\times10^{-9}$ F/m
- In case of a lossless or a lossy dielectric obstacle, there is a reflected as well a transmitted component of the EM wave (that enters into the medium of the obstacle). The lossy medium would eventually absorb/dissipate this transmitted EM energy through it
- When the medium is a good conductor (with $\sigma \to \infty$), most of the incident energy will be reflected back
- Thus the electromagnetic parameters (complex or real) namely, the permittivity $(\varepsilon_r\varepsilon_o)$, the permeability $(\mu_o\mu_r)$ and the conductivity (σ) eventually dictate the boundary conditions at the interface of the free-space to obstacle medium
- Snell's law and relevant boundary conditions decide the resulting reflection coefficients (that are of interest in the wireless communication transmissions).

When the reflection coefficient is finite, the resulting ground-reflected EM wave would interfere at the receiver with the direct-ray as illustrated in Figure 3.9.

The vector sum of direct/LoS wave and the reflected wave at the receiver is given by:

$$P_r = [P_tG_tG_r\lambda^2/(4\pi d)^2][1+\Gamma\exp(j\Delta\phi)]^2 \qquad (3.35)$$

where Γ denotes the reflection coefficient $\Delta\phi = (2\pi\Delta d/\lambda)$, with Δd being the path difference between direct and reflected rays.

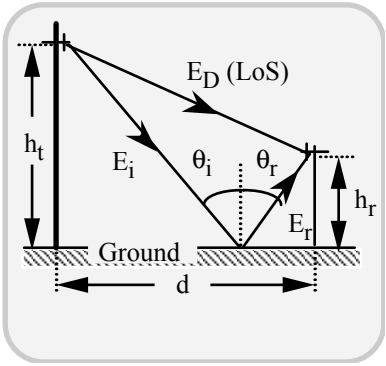


Fig. 3.9 Direct and reflected wave model: (*Two-wave model* pertinent to a flat ground)

For large values of d, $(2\pi\Delta d/\lambda) \approx (2\pi h_t h_r/\lambda d)$ and P_r simplified to $P_t G_t G_r (h_t h_r)^2/d^4$. The attenuation of the wave can then be expressed in terms of the *path loss* (PL) in dB as follows:

$$PL = -10\log_{10}(G_t) - 10\log_{10}(G_r) - 20\log_{10}(h_t) - 20\log_{10}(h_r) + 40\log_{10}(d) \tag{3.36}$$

Hence, at large distances, the received power falls off at a rate of 40dB/decade, (which is twice the extent of fall-off in free-space (LoS) situation with no ground reflection).

The aforesaid path loss prediction again refers to an ideal flat-ground with reflection coefficient approaching a value equal to –1. However, in practice, the terrain where the mobile receiver exists could be of such a nature (with lots of undulations, buildings, trees, rocks etc.) the prediction error with respect to measured data could pose a standard deviation in the range 6 to 8 dB. Therefore, there are empirical models that have been developed using measured data. For example, the so-called *Okumura model* [3.30] based on average measured data (collected in Japan) is specified by Hata [3.31] in terms of *median path loss* given by,

$$\begin{aligned} PL_{50} &= 69.55 + 26.16\log_{10}(f_o) - 13.82\log_{10}(h_t) \\ &\quad + [44.9 - 65.5\log_{10}(h_t)]\log_{10}(d) - [1.1\log_{10}(f_o) - 0.7]h_r \\ &\quad - [1.56\log_{10}(f_o) - 0.8] \end{aligned} \tag{3.37}$$

where f_o is the carrier frequency in MHz; h_t and h_r are base station and mobile unit antenna heights respectively, in metres; and d is the base station to mobile unit separation in kilometres.

With reference to the irregular terrain where the radio transmissions of a mobile communications system take place, the terrain profile significantly influences the path loss as could be evinced from the Okumura's model (which is applicable for the frequency range of 150-2000 MHz, base station to mobile distances in the range of 1 to 100 km and effective height of the base station in the range 30 to 1000 m). The general expression for Okumura's model can be written as:

$$PL_{50} = PL_{FS} + A_{mu}(f, d) + G_t(h_t) + G_r(h_r) \quad (3.38)$$

where the subscript 50 depicts the median propagation loss and the subscript FS denotes the free-space. Further, A_{mu} is the median attenuation relative to free-space.

The above model is quite popular in modeling the path loss of urban areas and it offers a standard deviation within 12 dB of measured data. It is used often in planning land mobile systems in Japan. It is not, however, good for rural terrain.

For suburban areas, the *Okumuara-Hata model* has been modified as follows:

$$PL_{50} = L_{50}(\text{Urban}) - 2[\log_{10}(f_o/28)]^2 - 5.4 \quad (3.39a)$$

and for an open area,

$$PL_{50} = L_{50}(\text{Urban}) - 4.78[\log_{10}(f_o)]^2 - 18.33 \log_{10}(f_o) - 40.98 \quad (3.39b)$$

The aforesaid models become even more complex considering other EM phenomena, namely diffraction and scattering, which are briefly outlined in the following section.

3.8 REFLECTION OF EM WAVE AT A LOSSY SURFACE

A reflecting medium faced by an EM wave in a wireless communication could also be lossy. For example, a wet brick or concrete wall or wet soil could typically pose a lossy medium with the result that the transmitted wave through such media will dampen significantly.

In literature [3.32], other practical civil engineering structures such as gypsum board walls built on metal studs etc. have been considered to study the EM wave reflection properties at frequencies of interest in wireless communications.

Typically, multilayered structures, that is, structures made of composite layers of different materials could also be of interest in knowing their EM reflection properties. For example, brick walls with cement mortar plastering are common in the buildings of Asia and tropical countries.

Bricks and mortar have distinct dielectric properties and mortar could be of significant thickness (1 cm to 2.5 cm). For example, when considering indoor environments, in countries like India, the ceiling and side walls are made of reinforced concrete while the partition walls are a layer of single-bricks. In these particular conditions, unique indoor EM reflection conditions may prevail.

By working out the problem indicated below, one can appreciate the extent of such EM wave attenuation through some practical civil engineering structures.

Problem 3.1
Assuming a wet wall with $\varepsilon_r = (4 - j0.1)$ at 4 GHz, show that the standing EM waves inside the wall (of thickness 30 cm) will strongly be damped.

(*Hint*: See [3.27], Page 74)

3.9 EM WAVE BOUNCING AT ROOF-TOPS OF BUILDINGS

Typically in urban areas, the high-rise buildings are disposed in a row roughly of uniform height. In small towns and villages, the buildings could be scattered and are of uneven heights. These ambient would pose distinct EM propagation scenarios. A comprehensive analysis of relevant consideration is presented in [3.27].

3.10 REFLECTIONS OF TM AND TE WAVES

As discussed in Chapter 2, the reflection of EM waves at a surface (such as ground, brick wall etc.) depends on the parallel or perpendicular polarisation of the incident wave.

The parallel polarisation refers to the incident **E**-field in parallel to the reflecting surface. That is, it is perpendicular to the plane of incidence. Hence, it is also known as TE wave.

In the perpendicular polarisation, the incident **H**-field in parallel to the reflecting surface; and, therefore this **H**-field is perpendicular to the plane of incidence classifying it as TM wave.

It was shown in Chapter 2 that these two versions of plane-polarised EM waves pose distinct reflection coefficients $|\Gamma|$ for oblique incidence of EM waves.

Problem 3.2
Plot $|\Gamma|$ as a function of the angle of incidence θ, for both TE- and TM-polarized waves reflected at a brick wall of thickness 20 cm, assuming the dielectric constant of the brick wall composite as 4.44, and frequencies being 900 and 1800 MHz.

Determine the Brewster and the critical angles.

With the finite thickness of the wall (equal to 20 cm), show that for both polarisations the reflection coefficient would vanish at $\theta = 18°$ and f = 1800 MHz.

Plot Γ_{TM} over $\theta = 18°$ to 65° at 1800 MHz and discuss its significance.

3.11 HEIGHT-GAIN FOR ANTENNAS

This refers to an increased signal strength observed when the antenna is raised in radio systems located on the earth's surface. Assuming that the transreceive links correspond to vertical polarization, the spatial dependence of the vertical **E**-field component, namely E_y, can be written as follows:

$$E_y(x, y) = H^i\{\exp[jky\cos(\theta)] + \Gamma_H \exp[-jky\cos(\theta)]\} \times \{\exp[-jky\sin(\theta)] \times [\eta\sin(\theta)]\} \quad (3.40)$$

where xz-plane depicts the ground-plane, η is the free-space intrinsic impedance, H_i is the incident magnetic field for the TM polarisation (perpendicular polarisation) with **H**-field parallel to the ground surface (xz-plane) and transverse to the plane of incidence and Γ_H is the reflection coefficient for the incident wave of perpendicular polarisation.

In typical radio links with the transreceive separation being very large (compared to the height at which the transmitter is located), the angle of incidence (θ) tends to 90° and hence $\Gamma_H \rightarrow -1$. Hence,

$$|E_y(x, y)| \rightarrow 2\eta H_i |\sin[ky\cos(\theta)]| \quad (3.41)$$

which shows that $|E_y|$ increases approximately linearly with respect to y (for small values of y depicting the receiving antenna height above the ground). Hence for $ky[\cos(\theta)] << \pi$, a linear gain of 6 dB is achieved by doubling the height.

Example 3.1

Suppose the transreceive separation (R) between a base station and the mobile unit is 10 km (Figure 3.10).

Assume the height of the base-station antenna (H_{BS}) location as 100 m. Determine the limit on the receiver (mobile) antenna located y metres above the ground, so as to achieve a linear height-gain assuming 1 GHz operation.

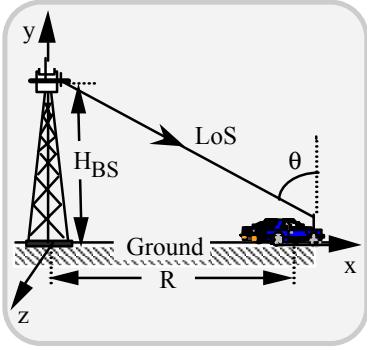


Fig. 3.10 Antenna height-gain scenario for a specified transreceive separation in a mobile phone communication. (Mobile antenna is y m above the ground)

Solution
From the condition that $kycos(\theta) << \pi$ with $k = 2\pi/\lambda$ and θ specified as shown in Figure 3.10, it follows that y should be far less than 50λ so that the linear height-gain conditions can be met. That is, the limit on y is about 15 m for 1 GHz signal.

3.12 REFLECTION OF CIRCULARLY POLARISED EM WAVES

Suppose the radio link supports a circularly-polarised wave propagation. Assuming the propagation along a positive x-direction, the EM wave can be specified by:

$$E(x,t) = (\mathbf{a}_y A_y + \mathbf{a}_z A_z)\exp(j\phi)\exp(-jkx)\exp(j\omega t) \quad (3.42a)$$

$$H(x,t) = (\mathbf{a}_z A_y + \mathbf{a}_y A_z)\exp(j\phi)\exp(-jkx)\exp(j\omega t) \quad (3.42b)$$

where A_y and A_z are complex amplitudes and ($\mathbf{a}_y$, $\mathbf{a}_z$) depicts the unit vector set along y and z respectively. Assuming that $A_y = A_z = A$ as real values, the instantaneous **E**-field reduces to:

$$\mathbf{E}(x,t) = A[\mathbf{a}_y \cos(\omega t - kx) \mp \mathbf{a}_z \sin(\omega t - kx)] \quad (3.43)$$

where ($\mp$) depict LHCP and RHCP respectively and the tip of the **E** vector along x will trace out a circle. (If $A_y \neq A_z$, then this trace will be an ellipse.)

The reflection of a circularly polarised wave at a good conducting surface will lead to reflected components with the sense of polarisation reversed with respect to the incident wave; the circularity of polarisation will approximately be retained. However, considering the reflection of a circularly-polarized wave at a dielectric surface, the circularity will be affected. But, when the angle of incidence θ is greater than the Brewster angle (θ_B), the sense of rotation will be preserved. The reversal of sense of polarisation could be of interest in wireless communications towards achieving reduced interference-based signal fading effects. The relevant application is discussed in Chapter 7.

3.13 DIFFRACTION OF EM WAVES IN WIRELESS TRANSMISSIONS

Diffraction refers to the "bending" of the EM at curvature and edges. For example, an EM wave can go beyond the rectilinear path-specified horizon over the curved earth's surface. Likewise, EM wave prevails in the shadow zone of an obstacle, when the edge of the obstacle is illuminated by an EM wave. The observed EM wave in the shadow zone refers to Huyghen's secondary wavelets produced at the point of illumination on the objects. Typically, considering the boulders/rocks of the mountains, large buildings etc., they can cast a shadow in reference to the EM wave incident on them. But the diffraction at their corners and edges would give rise to secondary waves in the shadow zone [3.27] as indicated in Figure 3.11. For

an incident ray from the source, S, the diffracted field in the shadow or diffraction zone where the receiver (R) is located and specified by the Fresnel integral, which is a function of Fresnel-Kirchhoff diffraction parameter.

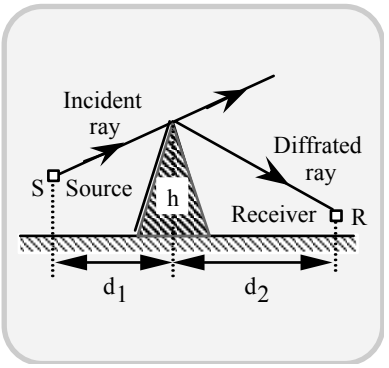


Fig. 3.11 Edge-diffraction phenomenon: Diffraction of the EM wave into the shadow-zone

Explicitly, the diffraction-zone field $F(\vartheta)$ is given by,

$$F(\vartheta) = \int_{\vartheta}^{\infty} \exp(-j\pi t^2/2)dt \qquad (3.44)$$

where $\vartheta = h[2(d_1+d_2)/\lambda d_1 d_2]^{1/2}$; and, the diffraction gain in dB resulting from the presence of a diffraction edge is equal to $20\log_{10}|F(\vartheta)|$.

Most often, the measured path-loss in a mobile communication radio environment is less than the prediction values specified by ground reflection and/or diffraction-specified calculations. This is because the incident radiation on the ground of an irregular terrain will be diffused as a result of scattering. Relevant considerations are described below.

3.14 SCATTERING OF EM WAVES IN MOBILE COMMUNICATION SCENARIO

Scattering of EM waves refers to a diffused reflection of incident energy impinging on a surface having an irregular profile. That is, when the surface undulations are significant (in dimensions relative to wavelength of incident energy), the laws of regular reflection from the surface are not strictly applicable. For example, shown in Figure 3.12 is an irregular surface, which diffuses/scatters the incident EM wave.

The random undulations or roughness of the surface is often specified by the *Rayleigh criterion*, which defines the critical height (h_c) of surface protuberances for a given angle of incidence θ_i. This condition is given by,

$$h_c/\lambda \geq [(1/8)\cos(\theta_i)] \tag{3.45}$$

When the height/wavelength ratio is less than the critical value h_c/λ, the surface is "smooth" and the incident wave will be specularly reflected (as per Snell's laws) and no diffusion/scattering prevails.

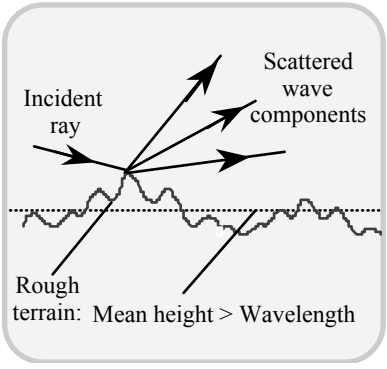


Fig. 3.12 EM wave scattering by a terrain of irregular surface: The extent of surface-irregularity or undulations dictate the Rayleigh criterion

For a rough surface, the Snell's law-based reflection coefficient is modified by a scattering loss coefficient ρ_s to account for the diminished extent of specular reflection caused by the diffused state of EM scattered energy. It is given by,

$$\rho_S = \exp[-(\pi\sigma_h/h_c)^2] \tag{3.46a}$$

where σ_h is the standard deviation of surface height about the mean height. A better expression for ρ_s is as follows:

$$\rho_S = \{\exp[-(\pi\sigma_h/h_c)^2]\}\{I_o(\pi\sigma_h/h_c)\} \tag{3.46b}$$

where $I_o(.)$ is the modified Bessel function of the first kind and zeroth order.

Concepts of geometric theory of diffraction and physical optics are, in practice, applied to deduce the scattered field strength. Further, considering the urban mobile radio systems, the scattering surface can be characterized by its *radar*

cross-section (RCS)*. A bistatic radar equation can be written to evaluate the power reradiated in the direction of the receiver (Figure 3.13):

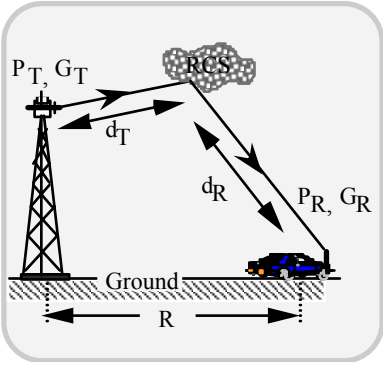


Fig. 3.13 Bistatic radar cross-section model of a wireless communication radio channel. (The term "bistatic" implies that the transmitter and receiver are in different locales.)

$$P_R = \left(\frac{P_T G_T}{4\pi d_T^2/\lambda^2}\right)(RCS)\left(\frac{G_R}{4\pi d_R^2/\lambda^2}\right) \qquad (3.47a)$$

which can expressed in dB as follows

$$P_R(dB_m) = P_T(dB_m) + G_T(dB_i) - 20\log_{10}(d_T) + RCS(dB_m^2) + G_R(dB_i) - 20\log_{10}(d_R) + 20\log_{10}(\lambda) - 20\log_{10}(4\pi)$$

$$= P_T(dB_m) + G_T(dB_i) - 20\log_{10}(d_T) + RCS(dB_m^2) - 30\log_{10}(4\pi) - 20\log_{10}(d_R) + 20\log_{10}(\lambda) \qquad (3.47b)$$

The simple propagation models indicated above, when extended to practical terrain, should include not only the effects of various terrain conditions, but also the man-

*RCS of a scattering object is defined as the ratio of the power density of the signal scattered in the direction of the receiver to the power density of EM wave incident upon the object. The unit for RCS is m^2.

made structural influences. In such models, relevant considerations are specified to include identifying the following:

- Nonflat terrain conditions with specular reflection and the diffused reflection point under non-obstructive conditions
- Nonflat terrain with single or multiple knife-edge diffraction
- Terrain made of land patches plus water beds
- Macro-cell (> 1km diameter) and micro-cell regimes

Comprehensive details, including the aforesaid considerations, in the mobile communication transmission modeling refer to the so-called "Lee's model" [3.33].

As mentioned earlier, practical propagation models are largely empirical. They have been proposed taking into account both the analytical considerations and measured data. They may also include the associated statistics of signal fluctuations with respect to time as well as varying mobile unit to base station separation.

There are two considerations in modeling the path loss in practical mobile system propagation pertinent to specified terrain conditions. The first one refers to specifying the average path-loss for an arbitrary transreceiver separation (d) in terms of a path-loss exponent η. Explicitly,

$$[PL(d)]_{mean} = k\left(\frac{d}{d_o}\right)^{\eta} \qquad (3.48a)$$

where k is a constant and d_o is a reference distance. In decibels

$$(PL)_{dB} = 10\log_{10}(k) + \eta \times 10\log_{10}(d/d_o) \qquad (3.48b)$$

where $10\log_{10}(k)$ can be set equal to the free-space path-loss with appropriate reference distance factor, d_o. Typical path-loss exponents are: $\eta = 2$ for free-space, $\eta = 2.7$ to 4 for urban areas, and $\eta = 5$ to 6 for shadowed urban conditions. Corresponding details as above, in reference to indoor radio channels are indicated in Chapter 7.

The other modeling consideration refers to what is known as the *local-mean prediction*. That is, the predicted local mean is used as a comparison with the measured local mean for a given test site. This local-mean prediction is different from the path-loss prediction in view of the distinct natural terrain configuration and the coexisting man-made structures. The so-called *Lee's model* mentioned before separates these two effects.

The path-loss can be empirically modeled at a flat test site of an urban area by an exponent model as follows:

$$P_r = P_{ro}(d/d_o)^{-\gamma}\alpha_o \text{ mw} \qquad (3.49a)$$

or

$$(P_r)_{dBm} = (P_{ro})_{dBm} - 10\gamma\log_{10}(d/d_o) + (\alpha_o)_{dBm} \qquad (3.49b)$$

where P_{ro} is the received power (in mw) at the (reference) distance d_o (=1 mile or 1 km), γ (≈ 4) is the slope of the path-loss deduced *via* measured data and α_o is a correction factor when actual conditions are different from the reference conditions. The above model applies to nonobstructive situations and for macro-cells. With obstructive paths, an additional diffraction loss (in dB) can be added to equation (3.49b).

There are models to specify the path-loss and received power under micro-cell conditions and also for obstructive path environments. When an obstruction prevails, diffraction effects are judiciously introduced; for an irregular terrain the model is explicitly modified to include the scattering.

3.15 SIGNAL FADING

In a mobile radio ambient, the multipath propagation involved leads to short-term fluctuations. Such fluctuation depicts small-scale signal variations distinct from the large-scale swings (from the mean level) in the received signal intensity. The received signal strength depends on transreceiver separation. Small-scale fading in received signal strength occurs as a result of the interference between two or more multipath components arriving at the receiver when the mobile traverses a short distance (in a short time). The amplitude and phase conditions of the multipath-specified vector components decide the received signal statistics. The associated factors are as follows:

- *Speed of the mobile station*: The relative motion of the mobile unit with respect to the base station would cause *Doppler shifts* in multipath components
- *Transmission bandwidth*: This refers to the signal bandwidth supported by the wireless transmission. If this bandwidth is larger than the so-called "flat-fading" bandwidth of multipath channel, the received signal will be distorted. (Relevant considerations are elaborated in the next section)
- *Time-delay and its spread associated with the received signal*: Since the spatial disposition of participating reflectors, diffractors and/or scatterers (which dictate the extent of multipaths) are a function of time, the relative phase and time of arrival of multipath signals at the receiving antenna are time-variant, posing varying extents of delay and arrival time in their composure
- *Random phase and amplitude*: The signal received will have random phase and amplitude as a result of the random characteristics of multipath components constituting the signal
- *Channel impulse response*: This characteristic is also time-variant as decided by the rate of change of channel properties *vis-á-vis* multipath geometry constituted by relative transreceive disposition and the objects in the vicinity.

3.15.1 Flat-fading

The small-scale fading indicated above could be either "flat" or "frequency-selective". The "flat-fading" refers to the condition when the mobile radio channel has a constant gain and linear phase response over a bandwidth that is larger than the transmitted signal bandwidth. In such situations, the signal integrity of spectral response is preserved, though the signal amplitude may fluctuate as a function of time as caused by multipath effects. The channels characterised by a "flat-fading" response are also known as *amplitude-fading channels* or *narrow band channels*.

The statistics of flat-fading is governed by the cluttered surroundings of reflectors/scatterers with none of them posing a dominant component. That is, the received multipath constituents of the signal at the receiving end are assumed to be almost identical with the dominant component being present. The signal at the receiver can be characterised by a probability density function (pdf) of a resultant vector **r** constituted by two independent components x and y, each of which is a Gaussian random variable. The pdf of r (depicting the envelope of received signal), namely p(r), is then *Rayleigh distributed* and is given by,

$$p(r) = [r \exp(-r^2/2\sigma^2)]/\sigma^2 \qquad 0 \le r \le \infty \qquad (3.50)$$

where σ is the variance of the received signal r.

In the Rayleigh fading environment two other statistical aspects of importance are the *level crossing rate* (LCR) and the *average fade duration*. The LCR is defined as the expected rate at which the received signal envelope (normalized to the local mean signal) crossed a specified threshold R in a positive going direction. The number of level crossings per second (N_R) is given by:

$$N_R = \sqrt{2\pi} f_m \rho \exp(-\rho^2) \qquad (3.51)$$

where f_m is equal to $(v/\lambda)\cos(\theta)$ and it depicts the maximum Doppler frequency; $\rho = R/R_{rms}$ where R_{rms} is the local root mean-squared (rms) amplitude of the fading envelope and v is the speed of the mobile unit.

The *average fade duration* (AFD) is defined as the average period of time (τ) for which the signal is below the threshold, R. It is also dictated by the Doppler effect due to the speed of the mobile. AFD is given by:

$$AFD = \langle\tau\rangle = [\exp(\rho^2) - 1]/\sqrt{2\pi}\,\rho f_m \qquad (3.52)$$

The most likely member of signaling bits that may be lost during a fade outage event is decided by the duration of signal fading occurring below a specified value of ρ.

Another class of fading that may occur in mobile communication transmission is due to the situation where there is a dominant signal component (among the multipath components) as a result of the direct line-of-sight reception (or due to a strong reflection from a nearby object). The joint distribution of this dominant part and other random components arriving at different angles represents

a statistics of small fluctuations superimposed on a dominant stationary signal. The relevant small-scale fading distribution is Rician. The pdf of the envelope of the signal r is then given by:

$$p(r) = \frac{r}{\sigma^2}\exp\left[-\frac{\left(r^2 + A^2\right)}{2\sigma^2}\right] I_o\left(\frac{Ar}{\sigma^2}\right) \qquad \text{for } \{A \geq 0,\ r \geq 0\}$$

$$= 0 \qquad \text{for } r < 0 \qquad (3.53)$$

where A denotes the dominant component and $I_o(.)$ is the modified Bessel function (of zeroth order and first kind) as mentioned before. The ratio of the deterministic signal power (A^2) and the variance of multipath components (σ^2) is expressed in dB is known as the *K parameter*. It is given by,

$$K = 10\log_{10}(A^2/2\sigma^2) \qquad (3.54)$$

As $A^2/\sigma^2 \rightarrow 1$ and $K \rightarrow 0$, the dominant part vanishes leading the statistics to represent the Rayleigh distribution.

3.15.2 Frequency selective fading

Suppose the channel supporting the wireless transmissions experiencing fading has a transfer function of constant gain and a linear phase over a bandwidth (which is much smaller than the bandwidth of the transmitted signal). Then the channel is said to induce a *frequency selective fading*. In such cases, the impulse response of the channel has a multipath delay spread greater than the time duration of the transmitted waveform. As such, the received signal would include several versions of transmitted waveform each with a delay. Hence, the resultant signal waveform would be distorted. In the frequency domain, this amounts to certain selective frequency components having larger gain than others. Modeling of frequency selective fading channels is usually difficult since it would warrant modeling and characterizing individual multipath involved. However, there are models like SURP [3.34a] and SIRCIM*[3.34b], which are generally used to study the effects of frequency selective small-scale fading in outdoor channels.

3.15.3 Fast-fading for indoor links

Indoor wireless ambient refers to a small-area coverage. Here, the average signal

* SIRCIM: This is acronym of the computer code simulation of indoor radio channel impulse-response model due to Rappaport et al. [3.34b]. It is based on measured data at 1300 MHz in five factory buildings and subsequently in more buildings. It refers to multipath delays in the range 40 to 800 ns. The delay is found to be uncorrelated to the transreceiver separation but is dependent on the building interior architecture and contents.

refers to averaging the signal measured as the portable end of radio-link is moved over a path whose length is about 20λ. In practice, this mobility is specified around a circle of radius about 1 m or on a raster path [3.35a, 3.35b].
Typically, a fast-fading is observed on an indoor link on the received power as the portable unit is moved over a circle (of about 3 ft) when the portable unit and the base unit are separately located in two different rooms (in a large office building). For the receiver locations within a building, the transreceive path of radio links can be fairly symmetrical; With the prevailing scattering environment, any fast-fading observed is justifiable when either of the end-entities moves. This situation is in contrast with the base stations used in outdoor cellular mobile radio where base station antennas are located well above the surrounding scatterers and any multipath effect perceived at the tower antennas refers to localised regions close to the subscriber. As such, the corresponding spatial variation at the base station antenna is not intense.

The unique fast-fading in indoor radio links suggests the use of antenna diversity schemes at both ends of the transreceive system. The small-area average will vary with location within a building as a result of the signal passing through walls and/or floors. The exact nature of such variations will depend on the architecture of the building interior geometry.

3.15.4 Electromagnetics of fast-fading

The fast-fading effects can be modeled as due to the superposition of several plane waves obliquely received at a specified location. Suppose the plane waves propagating in air parallel to the xz-plane with wave-vector k making an angle θ_i with the positive z-axis, as indicated in Figure 3.14.

The decomposed components of the wave-vector are $k_1 = k \sin(\theta_i)$, $k_2 = 0$ and $k_3 = k\cos(\theta_i)$. Further, for the plane-waves, polarised vertically, the total field along the x axis (at z = 0) of the E_y component can be given by:

$$E_y(x,0) = \sum_i A_i \exp(j\phi_i)\exp[-jkx\sin(\theta_i)] \qquad (3.55)$$

where A_i and ϕ_i are the amplitude and phase of the plane-waves, which are random parameters as decided by the statistics of the ambience and $k = 2\pi/\lambda$. Simulated studies on E_y indicate that the statistical distribution of the signal variation corresponds to a Rayleigh distribution [3.27].

The expression for E_y can be used to find the *autocorrelation function* (ACF) leading to the determination of the correlation length of the fading and Doppler spread when the subscriber is moving. For an observation distance much larger than the wavelength, the ACF reduces to $\sum_i A_i^2 \exp[-j(ks)\sin(\theta_i)]$ where s is the sample distance over which the ACF is assessed. In the limiting case, the summation assumes an integration in the ACF formulation over θ, leading to the result ACF(s) = $J_o(ks)$. Here, the Bessel function $J_o(.)$ describes the spatial correlation function of arriving plane waves (assumed to be uniformly distributed in the azimuthal plane) and the corresponding fast-fading remains decorrelated for $s > 0.4\lambda$.

The above spatial correlation consideration can be used to determine the Doppler spread resulting from the user mobility. Suppose the user moves at a velocity v_s, then the corresponding s can be replaced by $v_s\tau$ (with τ depicting the time of traverse over the distance s). Then, by taking the Fourier transform of ACF($v_s\tau$), the *power spectral density* (PSD) of the received signal can be specified as a function of $\Delta\omega$ depicting the deviation of the radian frequency from ω. Hence,

$$PSD(\Delta\omega) = (2/kv_s)U(kv_s - |\Delta\omega|) / [1 - (\Delta\omega/kv_s)^2]^{1/2} \qquad (3.56)$$

where $U(\cdot)$ is the unit-step function and kv_s denotes the maximum Doppler shift.

3.15.5 Macro- and micro-diversity considerations

Macro-diversity

In a cellular system, the base station is usually located in the heart of the service area, typically in the center of the cell. The base antenna is omnidirectional in azimuth but with about 6-10 dBi gain in elevation. It is intended to serve most of the cell area (≈ 95%). Stray pockets within the cell may experience a lower quality of service because the direct-path signal may be attenuated due to obstruction losses caused by buildings and terrain conditions.

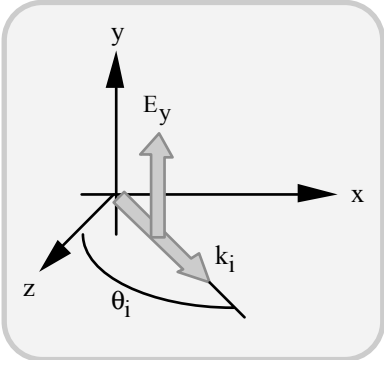


Fig. 3.14 Oblique incidence of a plane-wave at a specified location: Co-ordinate system and the associated vector components

The closest neighbouring (the first tier) base stations serve corresponding neighbouring area cells by using different sets of frequencies, eventually causing *adjacent channel interference*. The second closest neighbouring (the second tier) base stations might use the same frequencies (*frequency reuse*) causing *co-channel interference*. If the same real estate (base stations) is used in conjunction with 120° directional (in azimuth) antennas, the designated area may be served by three base stations. In this scheme one base station serves three cells by using three 120°

directional antennas. As a result, the same number of existing base stations equipped with new directional antennas and additional combining circuitry may be required to serve the same number of cells, though in a different fashion. The mode of operation in which two or more base stations serve the same area is known as *macro-diversity*. On a statistical basis, three base stations can provide a better coverage of an area similar in size to the system with a centrally-located base station. The directivity of a base station antenna (120° or even 60°) provides additional discrimination against signals for neighboring cells, therefore reducing adjacent and co-channel interference (that is, improving re-use efficiency and capacity). Actual improvement depends on the terrain configuration and the combining strategy and efficiency. However, it would warrant more complex antenna systems and combining devices.

Micro-diversity

Micro-diversity refers to the state of receiving two or more signals at one site (base or mobile). Typical micro-diversity schemes are as follows:

Space diversity systems employ two or more antennas spaced a certain distance apart from one another. For example, a separation of only $\lambda/2 = 15$ cm at f = 1 GHz, which is suitable for implementation on the mobile side, can provide a notable improvement in some mobile radio channel environments. Microspace diversity is routinely used on cellular base sites. (Macro-diverstiy is also a form of *space diversity*.)

Field component diversity system employs different types of antennas receiving either the electric or the magnetic component of an electromagnetic signal.

Frequency diversity systems use two or more different carrier frequencies to transmit the same information. In such cases, the same information signal may or may not fade at the same time at different carrier frequencies. Frequency-hopping and very wideband signaling can be viewed as frequency diversity techniques.

Time diversity systems are primarily used for the transmission of data. The same data is sent through the channel, as many times as necessary, until the required quality of transmission is achieved. (This function is done by *automatic repeat request* (ARQ) protocols.)

The improvement of any diversity scheme is highly dependent on the combining techniques employed, namely, the *selective (switched) combining*, the *maximal ratio combining*, the *equal gain combining*, the *feedforward combining*, the *feedback combining*, *majority vote scheme*, etc.

3.16 ANTENNA SELECTION AND SPECIFICATIONS

An antenna constitutes a critical juncture in any wireless communication system. It viably determines the quality of transreceive characteristics of RF communications. It plays a pivotal role in the overall system design and system integration under performance versus cost constraints. In selecting an antenna for a specific application consistent with the wireless ambient in question, one needs to follow the basic underlying parameters and related questions:

Operating frequency of the system:	*What is the range?*
Gain:	*How much?*
Radiation pattern:	*What is the type desired?*
VSWR:	*What is the maximum allowable value?*
Polarisation:	*What is the type required?*
Connector interface:	*What is the type compatible with the RF front-end?*
Power:	*How much power to be handled?*
Placement:	*Where will be the antenna placed (mounted) on the system?*
Radome requirement:	*Is a radome required?*
Life time:	*What is the expected life time?*

The frequency of operation of an antenna and the specified band are designated values for a given wireless application. Hence, the antenna deployed should be of such dimensions commensurate with the wavelength of operation over a specified range.

The gain of the antenna is the next primary factor of interest pertinent to antenna design and performance issues. Common wireless systems currently operate at VHF/UHF, microwave and millimetre wave ranges and the gains of typical antenna configurations (relative with respect to a dipole antenna gain) are listed in Table 3.1. (Details on these antennas and more related versions are presented in the ensuing chapters.)

Table 3.1 Gains (relative to a dipole) of typical antenna structures used in wireless communications

Antenna type	Gain (dBd)
Dipole antenna	0
Mobile whip antenna	– 0.6 to + 5.5
Corner reflector	4 to 10
Log-periodic antenna	3 to 8
Horn antenna	5 to 12
Helical antenna	5 to 15
Microstrip patch antennas	3 to 15
Yagi-Uda arrays	3 to 20
Panel antennas	5 to 20
Parabolic dish antennas	10 to 30

The next major delineator of antenna selection is the desired antenna pattern. It implicitly decides the gain *versus* concentrating the radiated power in a specified beamwidth. The narrower the beamwidth, the greater the gain of the antenna. The antenna design is aimed at enabling radiation in the azimuthal and/or vertical planes over a selective beamwidth where it is most beneficial for the RF link. For example, for a line-of-sight communication with mobile stations, the pattern is set mostly

omnidirectional in the horizontal plane with decreased vertical plane coverage (where the illumination is unwanted). Likewise, panel antennas are used to optimize a sector (typically 90° to 120° in azimuth) with narrow elevation beamwidth. The so-called parabolic reflectors, Yagi-Uda and patch arrays are designed to focus the EM energy into pencil-beams (posing high gain) with narrow beamwidths both in azimuthal and elevational planes.

Now, considering the VSWR parameter, what is its maximum allowable value? As indicated before, the VSWR represents the extent to which an antenna is matched to the system impedance. For maximum power transfer across the antenna and the system, the matching relation implies that the antenna impedance should be a complex conjugate of the system impedance.

In most of modern wireless transreceivers, ideally, the antennas are required to operate into a (50 + j0) ohm system impedance so that VSWR $\rightarrow$ 1. However, in practice, the preferable standard for maximum allowable VSWR across the entire bandwidth of a system is 1.5:1. It means that the antenna impedance should lie around 37.5 to 75 ohms. A VSWR in excess of 2:1 may not be acceptable since it amounts to undue reflection losses caused by the associated mismatching.

The next in the list of antenna parameters of interest is the polarisation. In general, the polarisation of EM radiation from an antenna in wireless communication systems could be vertical, horizontal or circular/elliptical (with right-handed or left-handed rotations). Choice of polarisation is specified to achieve isolation from possible interferers. That is, for example, a vertically polarised transreceive link is well isolated from the possible interference from a horizontally polarised transreceive link.

Many wireless communication systems, however, have to face transreceive conditions with antennas often pointing many degrees off-axis. This is especially true with the handheld units. To accommodate such situations, the antenna is configured for circular/elliptical polarisation with a hemispherically shaped pattern trading off high gain for reasonable gain in all directions.

Another important system design consideration is the selection of connector interface. Typically at UHF band applications type N-, TNC- and BNC-connectors are used because they maintain a good VSWR. Also, they can handle power up to 250 watts. In 1 to 10 GHz operations, low-profile antennas are interfaced with SMA-connectors. Regardless of the type of connector used, it is essential that the ground rules on noise, signal loss, reliability and matching are closely guarded in the designs.

An antenna with an integral jumper or "pigtail," which extends from the antenna to the feedline on the mounting structure (mast), may be preferred. Though an antenna with its connector mounted right on the antenna feedpoint is economical, it would however add-in the cost of a "jumper" cable with the associated signal losses and maintenance requirements.

The other crucial consideration is the power handling capability of the antenna. Generally, the passive antenna structures are capable of handling much lower powers than those operational in wireless communication systems. The power handling capability of the antenna is linked to that of the connector that may form the integral part of the antenna. When considering the power-handling capability of the antenna, it is normally the peak-power that is of importance.

Mounting an antenna refers to the way the antenna is "mounted" or installed in a transmitter or receiving system and it also specifies where the antenna is mounted. The location, that is the physical presence of an antenna structure in wireless communication system, refers to the following:

- Outdoor antennas
- Indoor antennas
- Antennas on portable systems

The placement of outdoor structures (such as those of base station antennas) is subject to zoning ordinances and regulations on installation *vis-à-vis* towers, buildings etc.

The so-called indoor antennas (like cordless phone base antennas, and access-point antennas used in wireless LANs (WLANs) and Bluetooth-enabled systems) are strategically placed for robust operations matching the clustered indoor ambient involving reflections, scattering and diffraction of EM waves at the operating wavelengths.

For the third category of antennas, the location is not normally definable and their in situ placement depends on the handheld or body-mounted conditions of the transreceive units. Their disposition in close proximity to the user's body has two design implications: The first one refers to the extent to which the human body would affect the performance of the antenna by virtue of the proximity involved; the second one is the *specific absorption rate* (SAR) depicting the EM power absorbed by the human body being close to the antenna.

The antenna design is tailored as far as possible not to let the power unnecessarily wasted via absorption by the body. This would allow more efficient radiation towards the RF link as well as prudently avoid excessive SAR (inasmuch as bioelectromagnetic health concerns have been expressed concerning EM radiation influences on living systems).

In the exterior placements of antennas on towers or masts, the integral strength of the design should also address the structural compatibility, namely, the ability to withstand wind, rain, ice and other climatic extremities. Provisions of feeds, radomes, connectors and cabling should be consistent with the exterior ambient. Material selection, construction of the tower etc. and hardware used play vital roles in the associated design. Use of durable materials is mandatory. The performance of the antenna under adverse weather conditions (such as encrusting of ice) should also be carefully considered.

3.17 OUTDOOR ANTENNAS: SITING CRITERIA

In installing antennas in outdoor environment there are certain general siting criteria to be followed. Briefly they refer to the following:

- Antenna sites should be designed for outdoor installations such that the public may not access areas proximal to them consistent with the public exposure standards of ANSI and FCC. As a general rule, the uncontrolled (public) exposure standard cannot be exceeded more than 20 feet from an antenna

- If there are areas accessible to workers that exceed the 1992 ANSI and/or FCC standards for controlled (occupational) exposure, the workers should know such areas and they should shut down the transmitters while accessing these antennas (normally specified within 10 feet from the antennas).

The outdoor antennas are generally either of low-gain or of high-gain type. The low-gain omnidirectional antennas are of whip types and pole-mounted, one for transmission and two for reception.

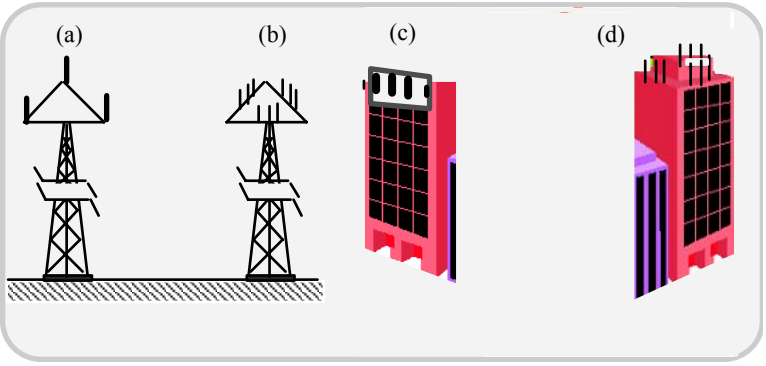


Fig.3.15 Mounting of outdoor antennas: (a) Low-gain pole/tower mounted; (b) high-gain pole/tower mounted; (c) high-gain penthouse installed; and (d) high-gain roof-mounted

The high-gain antennas are sector types with radiation pattern specified over 90°-120° azimuth. The so-called tripodal or 3-sector version consists of one transmitting and two receiving antennas per sector. They are either pole tower mounted or on penthouse or roof-mounted as illustrated in Figure 3.15.

3.17.1 Antenna installation guidelines

The following are general guidelines towards the installation of outdoor wireless communication antenna systems.

- For roof-mounted antennas, the transmitting antennas should be elevated above the height of people who may access the roof
- Further for such antennas, the transmitting radiators should be kept away from the areas where people are most likely to be (for example, roof access points, telephone service points, high-voltage AC equipment etc.)
- Also in roof-mounted directional radiating structures, the antennas should be near the periphery (of the roof) and point away from the building

- A trade off between large aperture size (supporting the lower end of radio frequency) and small aperture configuration (presenting low visual impact) should be duly considered
- RF standards are stricter for lower-frequency antennas (at 900 MHz) than for higher-frequency antennas (such as 1800 MHz)
- Special precautions should be taken to keep higher-power antennas away from accessible areas
- Antennas should be kept at a site as far apart as possible (although this may run contrary to local zoning requirements)
- Special precautions should be taken when designing "co-location" sites, where multiple antennas owned by different companies are on the same structure. This applies particularly to sites that include high-power broadcast (FM/TV) antennas. Local zoning often favors co-location, but co-location can provide "challenging" RF safety problems
- Land-based radio facilities often operate in hostile lightning environment. Any relevant installation should conform to standards on protection, not only for antennas and feeder-lines, but also for personnel and the associated radio equipment as well as the power-supply facilities. Necessary provisions of shielding, bonding and grounding are mandated in the specifications.

The steel antenna towers and supporting structures are governed by TIA/EIA 222 standards. This standard is reviewed every five years to maintain that it always complies with the state-of-the-art tower designs. Currently the existing version is as per Revision F; Revision G is being prescribed. The anticipated revisions are concerned with wind loading (specified by Three-Second Gust Wind Speed/1-in-50-year speed), ice loading (as per a prescribed ice-map of the regions concerned), appurtenance, distinguished features of rural and urban environments, seismic loading solutions and accounting for the site topography.

3.17.2 Work practices to reduce RF radiation exposure

- Individuals working at antenna sites should be informed about the presence of RF radiation, the potential for exposure and the steps they can take to reduce their exposure
- If RF radiation at a site can exceed the FCC standard for general public/controlled exposure, then the site should be posted with appropriate signs
- Radio-frequency radiation levels at a site should be modeled before the site is built
- Radio-frequency radiation levels at a site should be measured
- It should be assumed that all antennas are active at all times
- All attached transmitters before working on an antenna should be locked out
- Personal monitors to ensure that all transmitters have actually been shut down should be used

- A safe distance from antennas should be kept. As a practical guide for keeping [radio-frequency radiation] exposure low, maintain a 1-1.2m distance from any (telecommunications) antenna
- Posting of sign-boards like "Keep on moving" and "Avoid unnecessary and prolonged exposure in close proximity to antennas" are encouraged
- At some sites (for example, where multiple antennas are used in a restricted space and where some antennas cannot be shut down), it may be necessary to use protective clothing
- It should be remembered that there are many non-RF hazards at most sites (such as operating machinery, electric shock hazard, falling hazard etc.). Hence, only authorized and trained personnel should be allowed at a site.

Example 3.2
An RF transmitter operating at f Hz emits P watts. What is the safe distance to the antenna of this transmitter?

Solution
The International Radiation Protection Agency (IRPA) advises a specific absorption rate (SAR) level under 0.08 W per kg. Relevant maximum field strengths (E_{max}) are as follows:

0.1-1 MHz	75 v/m
10-40 MHz	28 v/m
1,200-200,000 MHz	60 v/m

Corresponding recommended IRPA safe distance (r) is specified by: $E_{max}^2 = 30P/r^2$ where E_{max} is in v/m, P in watts and r in metres.

For the handheld devices, another antenna placement consideration refers to its "low profile". In many systems, compactness and aesthetic considerations lead to placing the antenna interior to the unit or "embedded" as a part of the electronics or the printed circuit board (PCB). Unique design considerations are warranted in such built-in systems. If exterior mounting is inevitable, low-profile and "small-antenna" configurations are design options. Whether the antenna is interior or exterior to the handheld units, all the relevant radiation performance considerations should be met with including the SAR parameter specifications.

Concerning external antennas mounted on towers etc., provision of a radome has a dual propose: The radome protects the enclosed antenna system from the beatings of the harsh weather and it also "hides" the antenna to reduce visual impacts as well as offer some aesthetic attributes. The presence of a radome may

alter the radiation characteristics of the antenna (such as off-setting the bore sight axis of the main beam). Proper design counter measures should be duly specified.

The life time of an antennas is rather an *ad hoc* issue. Minimal replacement and maintenance considerations specify the adjacent features of antenna longevity and the cost effectiveness.

3.18 ANTENNA REQUIREMENTS QUESTIONNAIRE

In the design and implementation strategies, the designer and/or system engineer should consider answers for the queries indicated in Table 3.2 so that the designed/procured antenna meets the application profile warranted. Further, indicated in Tables 3.3. through 3.7, are specifications of typical antenna structures deployed in wireless communication applications. These specifications match the questionnaire. The details of various antennas mentioned in Tables 3.3-3.7 can be found in the ensuing chapters.

3.19 CONCLUDING REMARKS

Like electromagnetic theory, the subject of EM wave radiation and propagation considerations is vast and has been comprehensively studied. Ample literature covering the analyses of EM wave radiation and propagation viewed in various perspectives has piled up over the past decades concurrent to the emergence of radio engineering. Such studies also span almost the entire frequencies of EM waves of interest — from the LF band through millimetre waves and stretching into the infrared and light spectrum.

Classically, the RF transmissions propagated as ground and/or space/sky waves were subjected to intense studies during the evolution of long wave through VHF band radio operations. Subsequently, the UHF and microwave/millimetre wave propagation were also deeply scrutinised thanks to their deployments in radar, satellite and other communication systems.

The advent of cellular wireless communication systems (and subsequent wireless data communication efforts) has shifted the focus of EM radiation and propagation studies towards the analysis of unique mobile ambient encountered in both indoor and outdoor situations. The combination of analytical principles and measured data pertinent to practical wireless environments has enabled the postulation of many propagation models. Such models also include considerations relevant to terrain- and/or surrounding-specified fading conditions. Where necessary, the Doppler effects caused by the mobility of the end-units are appropriately adopted in model specifications.

The antenna and desired EM radiation considerations also warrant certain re-engineering of the classical strategies adopted in antenna designs. However, the basic radiation mechanism that can be stipulated in terms of associated electromagnetics of a simple current element should be well understood in order to conceive antennas for the exclusive applications in the wireless communications scenario. Hence, summarised in this chapter, are essentially the basics of EM radiation mechanism and analytical perspectives of the radiation from a current element.

Both antenna principles and propagation aspects of wireless communication, as addressed in this chapter, are used in the chapters that follow.

Table 3.2 Questionnaire on antenna design requirements

1. Frequency (MHz)
   Center frequency _____
   Bandwidth _____

2. Minimum gain:
   Reference (dBd or dBi) _____

3. Radiation pattern
   (Omni or directional) _____

4. Maximum allowable
   VSWR _____

5. Polarisation
   □ Vertical □ Horizontal
   □ Elliptical (□ RHCP or □ LHCP)
   _____

6. Connector type
   □ Male □ Female
   Is pigtail feed beneficial? _____
   If so, what is the length? _____

7. Power of antennas (in watts)
   Average or peak _____

8. Where will the antenna be mounted?
   _____

9. Maximum size of antenna:
   _____

10. Environment
   □ Indoors □ Outdoors
   Wind-speed _____
   Ice _____
Hardware required _____
Stainless steel-hardware _____
Rear- or center-mount _____
Mast diameter (if required) _____

11. Is radome required? _____
   Size _____
   Color _____

12. What is the expected lifetime of the antenna?
   _____

Table 3.3 Specifications of typical base-station transmitter and receiver antennas used in PCS applications

Parameters	Transmitter antenna (PCS 1900 MHz)	Receiver antenna (PCS 1900 MHz)
Elements	One air-loaded patch in the centre	Two air-loaded patches, + 45° slant and – 45° slant
Frequency range	1930-1990 MHz	1850-1910 MHz
Peak gain	3.0 dBi	3.0 dBi
Polarisation (vertical or horizontal)	Linear	Linear
Pattern	Omnidirectional	Omnidirectional
3 dB azimuthal beamwidth	130° sector	130° sector
3 dB elevational beamwidth	130°	130°
Impedance	50 ohms	50 ohms
Maximum input power	50 W	50 W
VSWR	2.0:1.0	2.0:1.0
Applications	Base-station, micro-cell or pico-cell	
Connector Cable Radome Operational ambient (Temperature, humidity, etc.)	Custom-specified	

Table 3.4 Typical Cellular/GSM and GPS integrated antenna specifications

Parameters	Cellular/GSM antenna	GPS integrated antenna
Elements	Microstrip (internal to automobile) Air-loaded patch (external to automobile)	Dielectric-loaded patch
Frequency range	824-896 MHz (Cellular) 880-960 MHz (GSM)	1575.42 ± 1.023 MHz
Peak gain	3.0 dBi	26 dBi (with LNA)
Axial ratio	—	3 dB (maximum)
Polarisation (vertical or horizontal)	Linear (along the axis of the antenna)	RHCP
3 dB azimuthal beamwidth	245°	100°
3 dB elevational beamwidth	80°	100°
Impedance	50 ohms	50 ohms
Maximum input power	50 Watts	50 Watts
VSWR	1.8:1.0	1.8:1.0
Low noise amplifier (LNA) provision	—	Gain: 26 dB Maximum NF: 2.5 (NF: Noise figure) DC operation: 4.5-5.0 volt with ~ 30 mA drain
Applications	Automobile installation	
Connector Cable length Radome Operating conditions (Humidity, temperature etc.)	Custom-specified	

Table 3.5 Specifications on typical wireless communication antennas designed for almost omnidirectional patterns

Parameters	Single-band versions	Dual-band versions
Elements	Microstrip	Microstrip
Frequency range	420-480 MHz 806-896 MHz 880-960 MHz 1710-1880 MHz 1850-1990 MHz 2400-2500 MHz	806-896 MHz and 1850-1990 MHz 880-960 MHz and 1710-1880 MHz
Gain	3.0 dBi	3 dBi
Polarisation (vertical or horizontal)	Linear	Linear
3 dB azimuthal beamwidth	240°	240°
3 dB elevational beamwidth	80°	80°
Impedance	50 ohms	50 ohms
Maximum input power	50 watts	50 watts
VSWR	1.5:1	1.5:1
Applications	In-vehicle or imbedded	
Connector Cable length Radome Operating conditions (Humidity, temperature etc.)	Custom-specified	

Table 3.6 Typical specifications of diversity antennas used in wireless communication antennas

Parameters	Omnidirectional Type	Directional versions	
		Bi-directional	Uni-directional
Elements	Air-loaded patch	Air-loaded patch	
Frequency range	1850-1990 MHz 2400-2500 MHz	2400-2500 MHz	
Gain	3.0 dBi	5 dBi	9.0dBi
Polarisation	Linear	Linear	
Power pattern	Omnidirectional	Bi-directional	Uni-directional
Impedance	50 ohms	50 ohms	
Maximum input power	50 Watts	50 Watts	
VSWR	2.0:1	1.8:1	1.5:1
Applications	In-door: Hallways, corridor or ceiling-mounting		
Connector	Custom-specified		
Cable length			
Radome			
Operating conditions (Humidity, temperature etc.)			

Table 3.7 Typical specifications of omnidirectional spherical coverage antennas used in wireless communication

Parameters	UHF version	Microwave versions	
		1900 MHz	2400 MHz
Elements	Air-loaded patch	Air-loaded patch	
Frequency range	824-896 MHz	1850-1990 MHz	2400-2500 MHz
Peak gain	3.0 dBi	3 dBi	
Polarisation	Linear	Linear	
Power pattern	Omnidirectional	Omnidirectional	
Impedance	50 ohms	50 ohms	
Maximum input power	50 Watts	50 Watts	
VSWR	2.0:1	2.0:1	1.5:1
Applications	Indoor: Ceiling-mounting		
Connector	Custom-specified		
Cable length			
Radome			
Operating conditions (Humidity, temperature etc.)			

In addition to EM principles and their applications to wireless communication antennas, there are a few important intutive aspects that we need to address before designing and implementing antennas. This approach will alleviate possible "design pitfalls", which may otherwise ensue if we omit these obvious practicalities. These can be identified and summarised by posing the following queries and answers [3.36]:

- *At what stage should the antenna part of the radio unit be designed?*
  It should be (one of) the first items designed so as to account for the influence of the shape and size of the platform in the gain and pattern performance issues.

- *Should the antenna be an internal unit or an external part?*
  While external antennas were popular in the legacy systems (because of their simplicity and sparse mutual-coupling effects on

the RF circuit), more and more concerns have been expressed taking into account their explicit presence, assembly considerations and fragility in the modern compact design requirements.

- *Where should the antenna be placed in the cellular unit?*
  The antenna placement in the radio unit should be consistent with its viable proximity to metal parts (RF shields) and user anatomy.

- *How should one optimise the antenna size?*
  If compact design is warranted, the size of the antenna (relative to the wavelength) should be optimised with respect to the bandwidth performance; an early design would permit a compromise between the antenna size *versus* available space constraint.

- *How should the antenna be oriented in its placement?*
  This again should be viewed tactfully taking into consideration of the user's body effects while in use.

- *How about the hazardous effects on the user due to possible RF energy interactions with the biology (of the user)?*
  In view of the much-debated biological influences of RF energy on human beings, a prudent approach will be to meet the antenna designs for minimal specific absorption rate (SAR) mandated dosage of RF energy on the users.

- *How far should the antenna be placed from the RF circuits?* The separation should be such that mutual detuning effects are avoided.

- *What is the concern about radio-unit shielding requirements versus antenna disposition?*
  The criterion should be such that any RF shielding used on the RF circuit side should minimally cause shielding effects on the antenna. This applies also to the encasing adopted.

- *What about the testing and evaluation of the designed antenna?*
  The antenna designed should be tested all along the evolution of the radio unit for its performance attributes.

This page intentionally left blank

REFERENCES

[3.1] Chatterjee, R., *Antenna Theory and Practice*, New Delhi, India: New Age International (P) Ltd. Publishers, 1996

[3.2] Kraus, J. D., *Antennas*, New York, NY: McGraw-Hill Book Co., 1988

[3.3] Balanis, C. A., *Antenna Theory Analysis and Design,* New York, NY: Harper & Row Publishers, 1982

[3.4] Johnson, R. C., and H. Jasik, eds., *Antenna Engineering Handbook*, New York, NY: McGraw-Hill Book Co., 1984

[3.5] Blake, L. V., *Antennas,* Boston, MA: Artech House Publishers, 1984

[3.6] Collin, R. E., and F. J. Zucker, *Antenna Theory,* New York, NY: McGraw-Hill Book Co., 1969

[3.7] Collin, R. E., *Antenna Radiowave Propagation,* New York, NY: McGraw-Hill Book Co., 1985

[3.8] Elliot, R. S., *Antenna Theory and Design,* Englewood Cliff, NJ: Prentice-Hall, Inc., 1981

[3.9] Jordan, E. C., and K. G. Balamain, *Electromagnetic Waves and Radiating Systems,* Englewood Cliff, NJ: Prentice-Hall, Inc., 1968

[3.10] Pozar, D. C., *Antenna Design Using Personal Computers,* Boston, MA: Artech House Publishers, 1985

[3.11] King, R. W. P., *The Theory of Linear Antennas,* Cambridge, MA: Harvard University Press, 1956

[3.12] Lo, Y. T., ed. *Handbook of Antenna Theory and Design,* New York, NY: Van Nostrand Rheinhold, 1987

[3.13] Rudge, A. W., K. Milne, A. D. Oliver and P. Knight, eds., *Handbook of Antenna Design,* Stevenage, UK: Peter Peregrinus Ltd, 1983

[3.14] Schekunoff, S. A., and H. T. Friss, *Antenna Theory and Practice,* New York, NY: John Wiley & Sons, 1952

[3.15] Silver, S., Microwave *Antenna Theory and Design,* New York, NY: McGraw-Hill Book Co., 1949

[3.16] Stutzman, W. L., and G. A. Thiele, *Antenna Theory and Design,* New York, NY: John Wiley & Sons, 1981

[3.17] Wait, J. R., *Antennas and Propagation,* Stevenage, UK: Peter Peregrinus Ltd, 1986

[3.18] Weeks, W. l., *Antenna Engineering,* New York, NY: McGraw-Hill Book Co., 1968

[3.19] Fujimoto, K., and J. R. James, eds., *Mobile Antenna Systems Handbook,* Boston, MA: Artech House Inc., 1994

[3.20] Setian, L., *Antennas with Wireless Applications,* Upper Saddle River, NJ: Prentice Hall PTR, 1998

[3.21] Saunders, S. R., *Antennas and Propagation for Wireless Communication Systems,* New York, NY: John Wiley & Sons, 1999

[3.22] Siwiak, K., *Radiowave Propagation and Antennas for Personal Communications,* Boston, NJ: Artech House Inc., 1995

[3.23] Pattan, B., *Robust Modulation Methods and Smart Antennas in Wireless Communications,* Upper Saddle River, NJ: Prentice Hall PTR, 2000

[3.24] Liberti Jr., B., and T. S. Rappaport, *Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications,* Upper Saddle River, NJ: Prentice Hall PTR, 1999

[3.25] Hirasawa K., and M. Haneishi, *Analysis, Design and Measurement of Small and Low-Profile Antenna,* Boston, MA: Artech House Inc., 1992

[3.26] Hendricksen, R., Clarify antenna gain for accurate mobile measurements, *Microwaves & RF*, Feb. 2001, 103-106

[3.27] Bertoni, H. L., *Radio Propagation for Wireless Communications,* Upper Saddle River, NJ: Prentice Hall PTR, 2000

[3.28] Rappaport, T. S., *Wireless Communications: Principles and Practice,* Upper Saddle River, NJ: Prentice Hall PTR, 1996

[3.29] Lee, W. C. Y., *Mobile Cellular Telecommunications,* New York, NY: McGraw-Hill, Inc., 1995

[3.30] Okumura, T., E. Ohmori, and K. Fukuda, Field strength and its variability in VHF and UHF land mobile service, *Rev. Elec. Comm. Lab,*, Sept./Oct. 1968, vol. 16 (9-10), 825-873

[3.31] Hata, M., Empirical formula for propagation loss in land mobile radio services, *IEEE Trans. Veh. Tech.,* Aug. 1968, vol. VT-29(3), 317-325

[3.32] Kim, S., B. Bougerolles and H.L. Bertoni, Transmission and reflection properties of interior walls, *Proc. IEEE ICUPC'94*, 1994, 1124-1128

[3.33] Lee, W. C. Y., Lee's model, *IEEE Trans. Veh. Tech.*, Aug. 1998, vol. VT- 47 (3), 68-70

[3.34a] Turin, G. L., F. D. Clapp, T. L. Johnston, S. B. Fine, and D. Lavry, A statistical model of urban multipath propagation. *IEEE Trans. Veh. Tech*., Feb.1972, vol. VT-21(1), 1-9

[3.34b] Rappaport, T.S., Performance decision feedback equalizers via simulated urban and indoor channels, *Special Issue on Land/mobile/Portable Propagation, IEICE Trans. Comm. (Japan),* Feb 93, vol. E76-B(2)

[3.35a] Honcharenko, W., H.L. Bertoni, and J. Dailing, Bilateral averaging over receiving and transmitting areas for accurate measurements of sector average signal strength inside buildings, *IEEE Trans. Antennas Propagat*., May 1995, AP-43 (5) 508-512

[3.35b] Lafortune, J. F., and M. Lecours, Measurement and modeling of propagation losses in a Building at 900 MHz, *IEEE Trans Veh. Tech.,* 1990, vol. VT-39, 101-108

[3.36] Blyer, J. Guidelines geared toward antenna design, *Wireless Syst. Design*, 2001, vol. 6(6), 52-53

This page intentionally left blank

CHAPTER 4

Antenna Elements in Wireless System Applications

4.1 INTRODUCTION

Antennas refer to gateways of wireless communications interfacing the free-space medium and the RF electronics of transreceive systems. In wireless communication applications, the operational push and design pull considerations direct the choice of antennas *vis-à-vis* achieving optimised range/coverage without adding a high visibility to the system. The relevant designs applied are characterised by a number of key parameters discussed in Chapter 3, which include mainly, the size of the antenna (relative to wavelength), shape (geometry), bandwidth/Q-factor, beamwidth, diversity, radiation efficiency, directivity/gain, polarisation, radiation pattern, relative levels of the lobes, and VSWR-specified matching considerations.

A variety of antenna structures – small and large – have been conceived and adopted in modern wireless systems. Such radiators range from giant parabolic dish-antennas for satellite communications systems to tiny dipole radiators used in cellular communication handsets, in addition to other basic radiating elements such as monopole or dipole antennas that are combined to make large arrays and smart systems deployed in base stations.

The large structure family of reflector antennas are formed by surrounding one or more radiating elements (primary feeds) with a reflective secondary surface that focuses the radiated EM energy in the receive or transmission phase. The size of the reflector (relative to the wavelength of operation) decides primarily the directivity of the radiated pattern. As the size of the reflector increases (relative to wavelength), the beamwidth of the reflector becomes narrower yielding a pencil-beam radiation compatible for point-to-point links.

The smaller versions of antennas used in wireless communications refer to portable radio designs wherein the trend has been to opt for smaller, less expensive and low-profile structures. Such compact antennas, for example, also replace traditional, whip-style antennas in ultra-high-frequency (UHF) mobile communications for vehicle-installed radios as well as in handheld equipment. The low-profile antennas, depending upon their applications, can handle up to 100 W power with 2-dBi gain and VSWR of less than 2.0:1.

Still another class of miniaturised designs is based on the microstrip technology outlined in Chapter 3. They are formed by metalised circuit traces formed on thin dielectric substrates supported on a ground-plane. That is, a

conducting patch on a dielectric substrate (supported on a ground-plane) forms a microstrip antenna. Relevant considerations rely on microstrip transmission-line concepts with the contour of the patch appropriately chosen and excited, so that the resulting current distribution on the patch leads to a desired EM radiation. As the thickness of the dielectric substrate increases, the bandwidth of operation also increases. However, in practice, due to dimensional constraint, the bandwidth of a simple microstrip antenna is somewhat limited; but such an antenna can be made extremely compact, with significant gain and low in cost. For example, the microstrip antennas used in GSM applications are low-profile structures but have a gain in the order of 30 dB with 360° azimuthal and 0 to 90° elevational coverage.

Another recent trend in antenna development refers to the design of so-called "smart antennas" formed by combining antenna hardware and *digital signal-processing* (DSP) techniques using spatial-processing software. Such systems intelligently adjust the antenna patterns according to changing operating conditions. These antennas, which borrow the underlying concepts from the traditional phased-array technology (which is widely adopted in radar and military applications), use the switched-beam and adaptive array techniques as well as adaptive beamforming methods. The switched-beam antennas employ a fixed-phase feed network (known as the *beamforming network*) and yield different (switched) outputs corresponding to beams in specified (fixed) directions. The beams are selected appropriately so that the outputs, when processed, give optimum performance while matching the operating conditions. In an adaptive array, the patterns are controlled dynamically, by means of a suitable feedback mechanism that controls the phase and amplitude of the beams. Adaptive beamformers use DSP techniques and iterative algorithms to smartly adjust the contributions of radiating elements. The smart antennas are exclusively addressed in Chapter 5.

Beam-switching technology is applied to improve the capacity and performance of GSM networks. Significant capacity gains are achieved in such systems by replacing conventional antennas with a narrow-beam antenna array plus necessary DSP hardware.

Apart from the antenna types indicated above, there are also a variety of specific radiators (of different sizes and shapes) that have been in use for wireless communications —horns, corner-reflectors, lens structures, helical antennas etc., are a few to mention. Further dual/multiple polarisation and multiband operation have become a part of wireless communication antenna designs.

The basic concepts of antenna elements have been thoroughly analysed and evaluated in all the classical books on antennas such as [4.1-4.20]. In this chapter, selective details from this vast literature are identified and summarised in two parts. The first part (Part I) largely covers the analytical aspects of antenna elements; Part II is written to identify and describe those elements which widely play a role in modern wireless communication systems.

It is suggested that the readers may choose to omit the first part if they are already familiar with the concepts of antenna theory. Also, for those who are not interested in the electromagnetics of antennas as such, it is recommended that they

can directly proceed to Part II without the loss of continuity. Sections in Part I can be read independently, as needed.

Part I

4.2 ELECTROMAGNETICS OF ANTENNA STRUCTURES

The basic radiating structures described in the previous chapter (Chapter 3), fall in the main streams of the following antennas in the context of wireless communication purposes:

- Linear antennas
- Aperture and slot antennas
- Planar and microstrip patch antennas

Linear antennas use a transmission line with an appropriate termination that facilitates desired radiation condition. (These antenna structures are "linear" in the sense that they are based on the linear, passive transmission-line considerations.) In the simplest form, a set of short conductors in series, excited by a transmission line can represent a linear antenna and each short element that is a part of this linear antenna is termed as a *short (elementary) dipole*.

An *aperture* on the other hand, refers to a designated region on a surface illuminated by an electromagnetic wave (EM) so that the aperture exhibits radiation characteristics consistent with the electromagnetic field impressed on it. For example, an open-ended waveguide excited by an appropriate EM field can be a radiating aperture yielding a radiation pattern.

A *slot* is a specific class of aperture representing an EM radiator. It refers to an opening on a conducting plane with dimensions much smaller than the dimensions of the conducting plane and the wavelength of EM excitation placed on it. A slot is essentially of one-dimensional geometry.

Unlike slot radiators, aperture antennas carry neither dimensional constraints nor size restrictions. They belong to two categories, namely, *primary* and *secondary* radiating apertures. The primary radiating aperture is normally a part of a transmission line representing a discontinuous end. For example, an open-ended coaxial-line or a waveguide can radiate EM energy and depict a primary aperture radiator. Such open-ended transmission line apertures can also be modified to obtain desired radiation characteristics and/or match the radiating-end to the transmission-line that excites it. Essentially, horn antennas represent such modified open-ended waveguide radiators and are commonly used as primary radiators.

A secondary aperture antenna is a reflector surface illuminated by an antenna element (which could either be a linear antenna or an aperture radiator). For example, a parabolic reflector is a secondary radiator and a primary antenna, (which is also known as a *feed antenna*) is used (at the focal plane of the parabolic surface) to illuminate the reflector. As a result, the parabolic reflector dish yields a (secondary) radiation pattern as a result of EM illumination on its surface.

Normally, this secondary reflector is much larger in physical size as compared to the primary feed.
Falling in the generic classes of antennas as above, there exists a plethora of radiators in the realm of wireless communications of modern times which are used either independently (as discrete units) or in groups (constituting an array). As stated before, this part of the chapter is dedicated to describe and analyse such antenna elements in terms of the electromagnetic concepts associated with them. (The next part (Part II) of this chapter is devoted to identify and elaborate the elements selectively adopted in wireless communication systems. The antenna arrays formed by a collection of discrete antenna elements and used in mobile communication systems are addressed in Chapters 5 and 6.)

4.2.1 Discrete antenna elements

There are various considerations specified in selecting a discrete (single) version of an antenna for use in mobile communication systems. The salient aspects, which can be linked to achieve the desired performance in such applications, are as follows:

- Desirable quality in transmission and reception modes
- Facilitation of low-power budgets
- Compactness and aesthetic features (for marketability of the cell phones)
- Light-weight and low profile attributes
- Omnidirectional pattern in the azimuthal plane
- Compatibility with EM environment
- Robustness against mechanical and environmental hazards
- Mountability and portability issues
- Compatibility for array configurations (when needed)
- Compatibility for use in diversity-specified installations
- Amicable interface feasibility with the front-end circuitry
- Cost-effectiveness.

In addition to application-specific features as above, the antenna elements in question should be compatible with the following general design parameters of the system:

- Frequency of operation
- Bandwidth
- Channel capacity
- Modulation and demodulation considerations
- Service area coverage and radiated power
- Desired-to-undesired signal-strength ratio (D/U ratio)
- Radiation pattern and gain
- Sidelobe level(s)
- Wave polarisation (linear, cross and circular)
- Matching (VSWR) performance

- Materials used: Electromagnetic, mechanical, structural and thermal properties
- Design, manufacture and installation considerations.

Consistent with the above design considerations and system requirements, the antenna elements as adopted in their discrete forms (as single units) are described in the forthcoming sections. Prior to considering the unique structures of these elements adopted in wireless communications, a systematic analysis of the radiation characteristics pertinent to the generic versions of these elements is needed. Hence, presented in the section below, are relevant analytical modelling methods and exposition of radiation parameters for linear and aperture antennas.

4.2.2 Linear antenna theory

Short (elementary) dipole/doublet

As indicated earlier, a linear antenna may be considered as made of a large number of very short conductors cascaded in series. The elementary section of a linear antenna, constituted by a current-carrying (short) conductor, is designated as a *current element*. This element under consideration is "short" in the sense that its physical length (L) is much smaller than the wavelength (λ) of the signal exciting it. That is, $L/\lambda << 1$. For analysis purposes, this elementary unit of a current element can be considered as a simple EM radiator as detailed below.

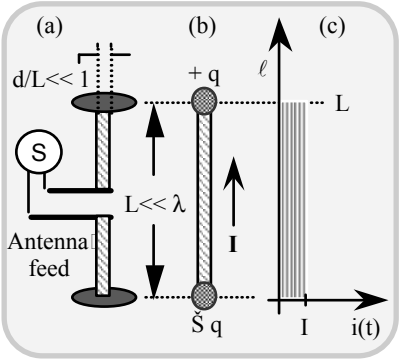


Fig. 4.1 Short (elementary) dipole: (a) Geometry (with disk-loading at the ends); (b) equivalent dipole representation; and, (c) the associated (uniform) current distribution along the length of the dipole

Figure 4.1(a) depicts a current-carrying, short element driven by a time-varying electrical signal. This radiating structure is also known as an *elementary dipole radiator*. Conducting disks shown at the ends of this radiator provide a capacitive loading. These disks together with the fact that $L << \lambda$, ensure an uniform current I along the entire length of the antenna. Further, it is assumed that this short dipole

is energised by a balanced transmission line and the radiation from the end-plates is negligible. Also, the short dipole under consideration is thin and slender with its diameter d << L. Hence, a simple equivalent representation of this short dipole can be specified as in Figure 4.1(b), which consists of a thin conductor of length L carrying a uniform current I, as shown in Figure 4.1(c), with point charges (+ q) and (– q) localised at, the ends. The current I, through the conductor and the associated charge q, are related by the conduction current relation, namely dq/dt = I amperes with q expressed in coulombs and the time (t) unit is in seconds.

Assuming that the medium surrounding the elementary dipole is free-space, one can proceed to find the EM field produced by this radiator. The co-ordinate system used for this purpose is presented in Figure 4.2, where the centre of the short-dipole is located at the origin of the co-ordinates.

In reference to Figure 4.2, inasmuch as the current I is specified only in the z-direction, the *retarded magnetic vector potential* has a single z-component, namely $\mathbf{A}_z$, which is given by,

$$A_z = \frac{\mu_o}{4\pi} \int_{-L/2}^{+L/2} [\mathbf{I}]/s \, dz \qquad (4.1a)$$

where [**I**] is the retarded current vector, namely,

$$|[\mathbf{I}]| = I_o \exp[j\omega\{t - (s/c)\}] \qquad (4.1b)$$

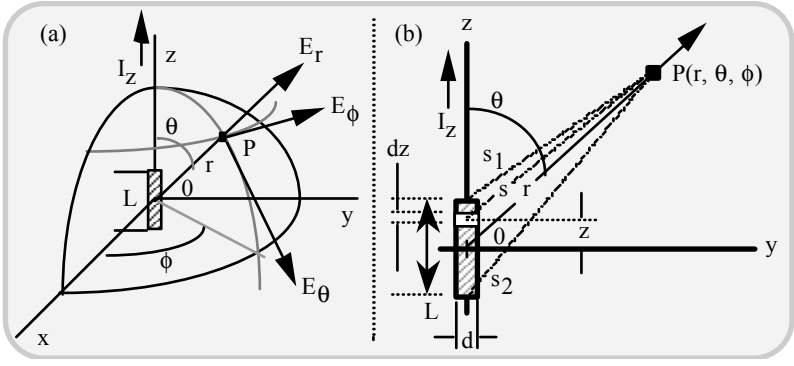


Fig. 4.2 A short dipole: (a) Vector components of the field in spherical polar co-ordinates and (b) geometry used in the analysis. (P is a remote point of observation)

Further, z is the distance to a point on the conductor from the origin; I_o is the peak value of current observed in time; μ_o is the permeability of free-space; c is the

velocity of propagation of EM wave in free-space and other entities are as indicated in Figure 4.2.

If r >> L and λ >> L, then s ≅ r; hence, the phase difference of the field contributions from different parts of the wire can be neglected. Then the magnetic vector potential can be rewritten as follows:

$$A_z = \frac{\mu_o L I_o \exp[j\omega\{t - (r/c)\}]}{4\pi r} \tag{4.1c}$$

The retarded scalar electric potential V at P due to the two charges + q and – q presumed at the ends of the short dipole is given by,

$$V = \frac{1}{4\pi\varepsilon_o}\{[q]/s_1 - [q]/s_2\} \tag{4.2}$$

Further, since $I = dq/dt \equiv I_o \exp[j\omega(t - s/c)]$ is consistent with an assumed harmonic excitation, it follows that

$$[q] = \int [I]dt = I_o \int \exp[j\omega(t - s/c)]\, dt = [I]/j\omega \tag{4.3}$$

From equations (4.2) and (4.3), the scalar electric potential at P can be written explicitly as,

$$V = \frac{1}{4\pi\varepsilon_o (j\omega)}\left\{\frac{[\exp[j\omega\{t - (s_1/c)\}]]}{s_1} - \frac{[\exp[j\omega\{t - (s_2/c)\}]]}{s_2}\right\} \tag{4.4}$$

When r >> L, the lines connecting the ends of the dipole and the point of observation, namely P, can be considered parallel so that,

$$s_1 \approx r - (L/2)\cos(\theta) \tag{4.5a}$$

and

$$s_2 \approx r + (L/2)\cos(\theta) \tag{4.5b}$$

Hence, it follows that

$$V = \frac{I_o \exp[j\omega\{t - (r/c)\}]}{4\pi\varepsilon_o (j\omega)} \times \left[\frac{\exp(\varphi_1) - \exp(\varphi_2)}{r^2}\right] \tag{4.6a}$$

where

$$\varphi_1 = \left(+ j[\omega L \cos(\theta)/2c]\right)\left\{r + [L/2]\cos(\theta)\right\} \tag{4.6b}$$

and

$$\varphi_2 = \left(- j[\omega L \cos(\theta)/2c]\right)\left\{r - [L/2]\cos(\theta)\right\} \tag{4.6c}$$

Further, by neglecting the term $L^2/4\cos^2(\theta)$ in the denominator of equation (4.6a) in comparison with r^2 (since $r >> L$), and inasmuch as $\lambda >> L$, it leads to the conditions $\cos[\omega L \times \cos(\theta)/2c] \approx 1$ and $\sin[\omega L \times \cos(\theta)/2c] \approx [\omega L \times \cos(\theta)/2c]$; hence, the expression for V simplifies to:

$$V = \left[\frac{I_o L \cos(\theta)\exp[j\omega\{t - (r/c)\}]}{4\pi\varepsilon_o c}\right] \times [(1/r) + (c/j\omega) \times (1/r^2)] \tag{4.7}$$

Thus, with the knowledge of magnetic vector potential **A** and electric scalar potential V, the electric and magnetic field intensities at P can be deduced explicitly using the following relations:

$$\mathbf{E} = -j\omega\mathbf{A} - \nabla V \tag{4.8a}$$

$$\mathbf{H} = \frac{1}{\mu_o}\nabla \times \mathbf{A} \tag{4.8b}$$

In the present analysis, since **A** has only a *z*-component, A_z is specified by equation (4.1c) and V is given by equation (4.7). Further, in spherical polar co-ordinates, the components A_r and A_θ of **A** are specified by,

$$\left.\begin{aligned} A_r &= A_z\cos(\theta) \\ A_\theta &= -A_z\sin(\theta) \end{aligned}\right\} \tag{4.9}$$

and

$$\nabla V = \mathbf{u}_r\frac{\partial V}{\partial r} + \mathbf{u}_\theta\frac{1}{r}\frac{\partial V}{\partial \theta} + \mathbf{u}_\phi + \frac{1}{r\sin\theta}\frac{\partial V}{\partial \phi} \tag{4.10}$$

Using equations (4.1c) and (4.7) in equations.(4.8a) and (4.8b), together with the relation specified by equations (4.9) and (4.10), the components of **E** are obtained as follows:

$$E_r = \frac{I_o L \cos(\theta) \exp[j\omega\{t-(r/c)\}]}{2\pi\varepsilon_o}\left(\frac{1}{cr^2}+\frac{1}{j\omega r^3}\right), \qquad (4.11a)$$

$$E_\theta = \frac{I_o L \cos(\theta) \exp[j\omega\{t-(r/c)\}]}{4\pi\varepsilon_o}\left(\frac{j\omega}{c^2 r}+\frac{1}{cr^2}+\frac{1}{j\omega r^3}\right) \qquad (4.11b)$$

$$E_\phi = 0 \qquad (4.11c)$$

with $1/c^2 = \mu_0\varepsilon_0$. Similarly, the (r, θ and ϕ) components of **H** are given by,

$$H_r = H_\theta = 0 \qquad (4.12a)$$

and

$$H_\phi = \frac{I_o L \sin(\theta) \exp[j\omega\{t-(r/c)\}]}{4\pi}\left(\frac{j\omega}{cr}+\frac{1}{r^2}\right) \qquad (4.12b)$$

Therefore, the EM field of a short dipole has only three components, namely, E_r, E_θ and H_ϕ as indicated above. Examining the expressions of these three components, it can be noticed that E_r has $1/r^2$ and $1/r^3$ terms; E_θ has $1/r$, $1/r^2$ and $1/r^3$ terms; and H_ϕ has $1/r$ and $1/r^2$ terms. The $1/r^2$ term is called the *induction field* and the $1/r^3$ term represents the *electrostatic field*. These two terms are significant only in the region very close to the dipole and therefore, are important only when considering the *near-field* region of the short dipole. For very large values of r/λ (relative to the size of the radiator expressed in wavelengths), the $1/r^2$ and $1/r^3$ terms can be neglected leaving only the $1/r$ term being significant. This $1/r$ term represents the *far-field*. In this far-field region, E_r is negligible and only E_θ and H_ϕ components are important. They are given by,

$$E_\theta = \frac{j\omega I_o L \sin(\theta) \exp[j\omega\{t-(r/c)\}]}{4\pi\varepsilon_o c^2 r} \qquad (4.13a)$$

$$H_\phi = \frac{j\omega I_o L \sin(\theta) \exp[j\omega\{t-(r/c)\}]}{4\pi cr} \qquad (4.13b)$$

Hence, the following ratio can be specified:

$$\frac{E_\theta}{H_\phi} = \frac{1}{\varepsilon_o c} = \sqrt{\frac{\mu_o}{\varepsilon_o}} = (120\pi \approx 377) \quad \text{ohms}, \qquad (4.14)$$

which denotes the *intrinsic* or *characteristic impedance* of the free-space. Examining the E_θ and H_ϕ components, it can be seen that E_θ and H_ϕ are in time-phase (with respect to each other) in the far-field, and that the field patterns of both are proportional to $\sin(\theta)$ but independent of ϕ. Therefore, the space pattern is a *figure of revolution* and is doughnut-shaped as illustrated in Figure 4.3.

Further, referring to equations (4.11) and (4.12), it can be noted that for small values of *r*, the electric field has two components, E_r and E_θ, which are both in time-phase quadrature with the magnetic field, as in a resonator. At intermediate distances E_θ and E_ϕ can approach time quadrature, so that the total electric field vector rotates in a plane parallel to the direction of propagation, thus exhibiting the phenomenon of *cross-field*. The near-field pattern for E is proportional to $\cos(\theta)$, as shown in Figure 4.3.

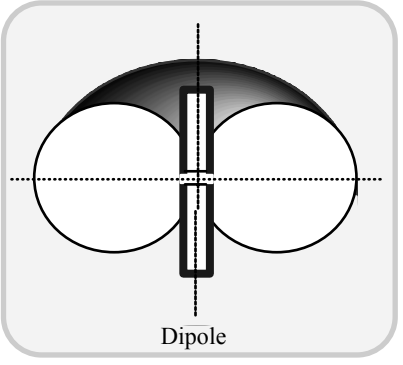


Fig. 4.3 Near- and far-field patterns of E_θ and H_ϕ components of a short-dipole: Doughnut-shaped 3D-pattern

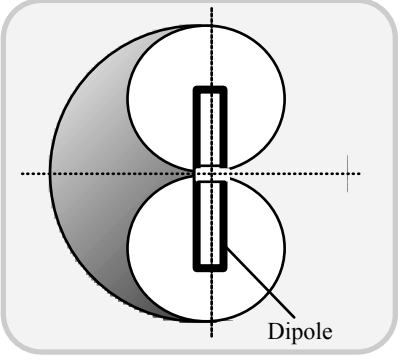


Fig. 4.4 Near-field patterns of E_r component of a short-dipole: Doughnut-shaped 3D-pattern

At very low frequencies (known as the *quasi-stationary* condition), the radiation characteristics can be deduced as follows: Since [I] = jω[q] for harmonic excitation, equations (4.11) and (4.12) can be written simplified as follows:

$$E_r = \frac{[q]L\cos(\theta)}{2\pi\varepsilon_o}\left(\frac{j\omega}{cr^2}+\frac{1}{r^3}\right) \tag{4.15a}$$

$$E_\theta = \frac{[q]L\sin(\theta)}{2\pi\varepsilon_o}\left(-\frac{\omega^2}{c^2 r}+\frac{j}{cr^2}+\frac{1}{r^3}\right) \tag{4.15b}$$

$$E_\phi = \frac{[I]L\sin(\theta)}{4\pi}\left(\frac{j\omega}{cr}+\frac{1}{r^2}\right) \tag{4.15c}$$

As $\omega \to 0$ (namely, at low frequencies), the terms with ω in the numerator of the above relations can be neglected. Further, under this quasistatic situation,

$$[q] = q_o \exp[\,j\omega(t - r/c)] = q_o = \text{a constant} \tag{4.16a}$$

Also,

$$[I] = I_o = \text{a constant} \tag{4.16b}$$

Hence, for a quasi-stationary case (that is, when approaching the DC condition of excitation), the field components of a short dipole reduce to,

$$\left.\begin{aligned} E_r &= \frac{q_o L\cos(\theta)}{2\pi\varepsilon_o r^3}, \\ E_\theta &= \frac{q_o L\sin(\theta)}{4\pi\varepsilon_o r^3} \\ H_\phi &= \frac{I_o L\sin(\theta)}{4\pi r^2} \end{aligned}\right\} \tag{4.17}$$

The expressions for E_r and E_θ, as above, are identical to the electrostatic field of two point charges $+q_o$ and $-q_o$ separated by a distance L_o. The expression for H_ϕ (as above) is the well-known Biot-Savart relation for the magnetic field induced by a short element carrying a steady or slowly varying electric current.

Considering the far-filed components E_θ and H_ϕ, the Poynting vector has a radial component P_r given by,

$$P_r = \frac{1}{2}\mathrm{Re}(E_\theta H_\phi^*) = \frac{1}{2}\mathrm{Re}(Z_o H_\phi H_\phi^*)$$
$$= \frac{1}{2}|H_\phi|^2\,\mathrm{Re}(Z_o) = \frac{1}{2}|H_\phi|^2\sqrt{\frac{\mu_o}{\varepsilon_o}} \qquad (4.18)$$

where Z_o is the characteristics impedance of free-space, namely, $(\mu_0/\varepsilon_0)^{1/2} = 120\pi$ ohms. The total power radiated is given by,

$$W = \iint_{LS} P_r\,da = \frac{1}{2}(\mu_o/\varepsilon_o)^{1/2}\int_0^{2\pi}\int_0^{\pi}|H_\phi|^2\,r^2\sin(\theta)d\theta d\phi$$
$$= \frac{1}{32}(\mu_o/\varepsilon_o)^{1/2}(\beta^2 I_o^2 L^2/\pi^2)\int_0^{2\pi}\int_0^{\pi}\sin^3(\theta)d\theta d\phi$$

$$= (\mu_o/\varepsilon_o)^{1/2}(\beta^2 I_o^2 L^2/12\pi) \qquad (4.19)$$

where β is the phase constant (that is $\beta = k = 2\pi/\lambda$) and LS denotes a large, spherical surface over which the integration is performed. Further, the derivation of equation (4.19) includes the substitution of $|H_\phi| = \omega I_0 L\sin(\theta)/4\pi cr$ from equation (4.13). The power, obtained as above, can be identically equated to I^2R where I is the root mean-square (rms) current impressed on the short dipole and R is a resistance, termed as the *radiation resistance* of the radiator. Hence, from the resulting identity, namely,

$$(\mu_o/\varepsilon_o)^{1/2}(\beta^2 I_o^2 L^2/12\pi) \equiv I_o^2 R/2 \qquad \text{watts} \qquad (4.20a)$$

it follows that

$$R = (\mu_o/\varepsilon_o)^{1/2}(\beta^2 L^2/6\pi) = 80\pi^2(L_\lambda^2) \qquad \text{ohms} \qquad (4.20b)$$

where $L_\lambda = L/\lambda$.

Centre-fed dipole with sinusoidal current distribution

In a linear antenna symmetrically fed at the centre (by a balanced transmission line) the current distribution can be assumed to be sinusoidal. (Whenever the diameter of the conductor is less than $\lambda/100$, this assumption of sinusoidal distribution of current along the length of the dipole is approximately valid.) Such

sinusoidal current distribution profiles along different lengths of dipole are illustrated in Figure 4.5.
In order to study the properties of a dipole structure as an EM radiator, its far-field should be first derived. Relevant approach is as follows: The retarded value of the current at any point z on the antenna in reference to the observation point P(r, θ, ϕ) in Figure 4.6, which is at radial distance s (from the point z), is given by,

$$I = I_o \sin\{\frac{2\pi}{\lambda}[(L/2) \pm z]\}\exp\left[j\omega(t - s/c)\right] \qquad (4.21)$$

where (L/2 ± z) correspond to z < 0 and z > 0, respectively.

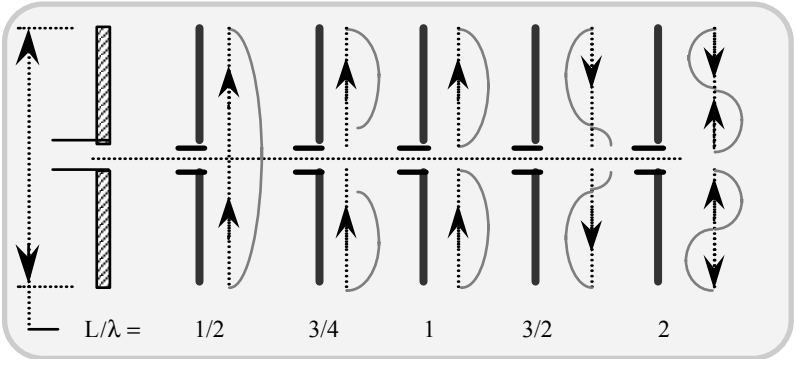


Fig. 4.5 Current distribution profiles on thin centre-fed dipoles (along the lengths)

The radiator under discussion can be modelled as a series of infinitesimally short dipoles, each of length dz. The far-field components of one of these short dipoles at P (in Figure 4.6) can be written using equation (4.13) as follows:

$$dE_\theta = \frac{j60\pi[I]\sin(\theta)dz}{s\lambda} \qquad (4.22a)$$

$$dH_\phi = \frac{j(I)\sin(\theta)dz}{2s\lambda} \qquad (4.22b)$$

Since $E_\theta/H_\phi = (\mu_o\varepsilon_o)^{1/2} = Z_0$, it is enough if H_ϕ is evaluated by integrating dH_ϕ over the length of the antenna and E_θ is obtained from the relation $E_\theta = Z_o H_\phi$. Thus, H_ϕ is given by $H_\phi = \int_{-L/2}^{+L/2} dH_\phi$, which by using the current-distribution relation given by equation (4.21) reduces to:

$$H_\phi = \left\{\frac{jI_o \sin(\theta)\exp(j\omega t)}{2\lambda}\right\} \times \left[\int_{-L/2}^{0} \frac{\sin(\xi_1)\exp(-j\omega s/c)dz}{s} + \int_{0}^{+L/2} \frac{\sin(\xi_2)\exp(-j\omega s/c)dz}{s}\right] \quad (4.23)$$

where $\xi_1 = \frac{2\pi}{\lambda}\left(\frac{L}{2} + z\right)$ and $\xi_2 = \frac{2\pi}{\lambda}\left(\frac{L}{2} - z\right)$.

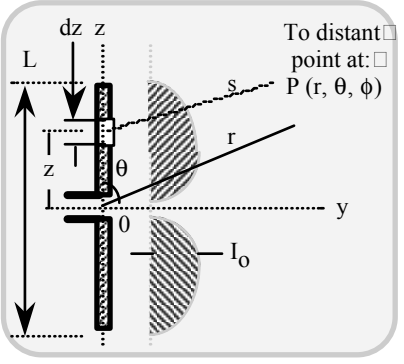


Fig. 4.6 Thin linear centre-fed dipole: The size of the antenna (L/λ) is assumed to be very much smaller than the distance of the point of observation (r/λ)

Further, from Figure 4.6, s = [r – z × cos(θ)], and at a large distance with L << r, s ≈ r. Hence, equation (4.23) becomes,

$$H_\phi = \frac{jI_o \sin(\theta)\exp[j\omega(t - r/c)]}{2\lambda r}\left\{\int_{-L/2}^{0} \sin(\xi_1)\exp[+j\omega\cos(\theta)/c]dz + \int_{0}^{+L/2} \sin(\xi_2)\exp[+j\omega\cos(\theta)/c]dz\right\}$$

$$= \frac{j\beta I_o \sin(\theta)\exp(j\omega\{t-(r/c)\})}{4\pi\lambda}\left\{\begin{array}{l}\int_{-L/2}^{0}\exp[\,j\beta z\cos(\theta)]\sin\left[\beta\left(\frac{L}{2}+z\right)\right]dz+\\ \int_{0}^{+L/2}\exp[\,j\beta z\cos(\theta)]\sin\left[\beta\left(\frac{L}{2}-z\right)\right]dz\end{array}\right\}$$

(4.24a)

where again, $\beta = k = \omega/c = 2\pi/\lambda$. Performing relevant integrations involved, equation (4.24a) reduces to $H_\phi = (j[I_o]/2\pi r) \times SF$ and $E_\theta = 120\pi H_\phi = (j[I_o] \times 60/r) \times SF$ with $[I_0] = I_0\exp[j\omega(t - r/c)]$ and,

$$SF = \left[\frac{\cos\{[\beta L/2]\cos(\theta)\} - \cos[\beta L/2]}{\sin(\theta)}\right] \qquad (4.24b)$$

Equations (4.24b) represents the *shape factor* of the radiation patterns of the far-field components of centre-fed, thin dipole-antennas under discussion. Figure 4.7, for example, shows the far-field pattern of the linear antennas of lengths $\lambda/2$ and $3\lambda/2$. A linear radiator of this type with $L = \lambda/2$ is known, popularly as, a *Hertzian dipole*.

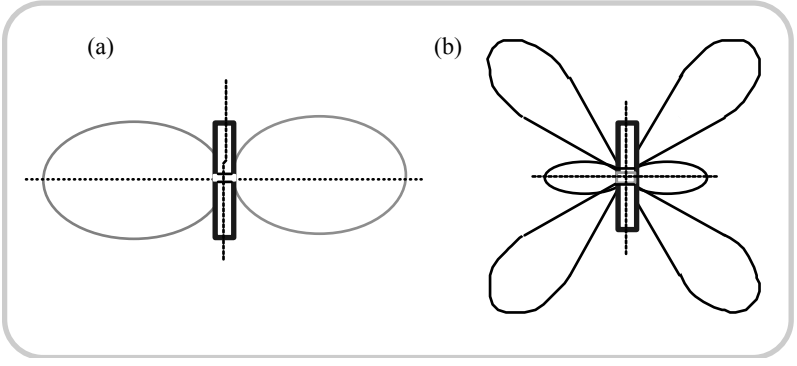


Fig. 4.7 Far-field patterns of thin centre-fed dipoles: (a) $L = 2\lambda$ and (b) $L = 3\lambda/2$

To find the radiation resistance of this thin, linear antenna, the Poynting vector relevant to the field components can be integrated over a large sphere so as to obtain the total power (W) radiated. Then this total power is identically equated to $I_o^2 R/2$, where R depicts the radiation resistance at a current maximum point. Explicitly, the total power W is given by,

$$W = \frac{15I_o^2}{\pi}\int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi}(SF)d\theta d\phi = 30I_0^2\int_{\theta=0}^{\pi}(SF)d\theta \quad \text{watts} \qquad (4.25)$$

Equating W to $I_o^2 R/2$, the radiation resistance R is obtained as,

$$R = 60\int_0^{\pi}\frac{[\cos\{\beta L\cos(\theta)/2\} - \cos\{\beta L/2\}]^2}{\sin(\theta)}d\theta \quad \text{ohms} \qquad (4.26)$$

The radiation resistance given by equation (4.26) refers to the current-maximum on the radiator. For a λ/2-dipole, the current-maximum is at the centre of the antenna as shown in Figure 4.5. Hence substituting L = λ/2, R is obtained from equation (4.26) as,

$$R = 30\int_0^{2\pi}\frac{[1-\cos(\nu)]}{\nu}d\nu = 30\times Cin(2\pi) \qquad \text{ohms} \qquad (4.27)$$

where

$$Cin(x) = \int_0^{x}\frac{[1-\cos(\nu)]}{\nu}d\nu = \ln(\gamma x) - Ci(x) \qquad (4.28a)$$

with ln(γ) representing the Euler's constant equal to 0.577 and Ci(x) is the cosine integral equal to $\ln(\gamma x) - \frac{x^2}{2!2} + \frac{x^4}{4!4} - \frac{x^6}{6!6} + \ldots$. When x is small (< 0.2),

$$Ci(x) \approx \ln(\gamma x) = 0.577 + \ln(x) \qquad (4.28b)$$

And, when x is large (x >> 1),

$$Ci(x) \approx \sin(x)/x \qquad (4.28c)$$

Using equation (4.27) the radiation resistance of an Hertzian dipole is obtained as: $R_{(\text{Hertzian dipole})} = 30 \times Cin(2\pi) = 30 \times 2.44 = 73$ ohms. While equation (4.27) determines the value of the radiation resistance of a half-wave dipole, the terminal impedance of the radiator, however, may also include some stray inductive reactance. In order to make the antenna to resonate, this reactance should be made equal to zero. This would require the antenna be made of length a little less than λ/2 so that the resulting capacitive reactance will cancel out the inductive effects. As such, in practice, the radiation resistance of a Hertzian dipole with a length (taken slightly less than λ/2) would correspond to R being somewhat less than 73 ohms.

The radiation resistance (R_1) of a Hertzian dipole at a point other than the locale of current-maximum can be determined in terms of the corresponding value (R) specified at the point of current-maximum. The relevant relation is decided by considering the total radiated power set equal to the corresponding power at the point where R_1 is ascertained. That is $I_o^2R \equiv I_1^2R_1$, where I_o and I_1 are values of current at the centre and the off-centre points respectively. Corresponding to an off-centre point at a distance z_1 from the antenna centre (where R_1 is elucidated), and assuming a sinusoidal current distribution along the length of the conductor given by $I_1 = I_0\cos(\beta z_1)$, R_1 becomes $R/\cos^2(\beta z_1)$.

Linear antenna on a ground-plane

In the discussions of previous sections, it was assumed that the linear antenna is placed in free-space and the ground has no effects on it. This is true only when the radiating element is far away from the ground. Otherwise, the presence of ground-plane in the vicinity of the antenna would affect the radiation pattern significantly.

To analyse the relevant scenario, two possible configurations can be considered. In the first case, the linear radiating element is vertically placed with respect to the ground-plane; and, in the second case, the antenna is placed parallel to the ground-plane.

Vertical antennas above a ground-plane

A typical configuration of a linear antenna placed vertically on a ground-plane is illustrated in Figure 4.8. It is assumed that the ground-plane is of infinite extent with a conductivity $\sigma \to \infty$; that is, the antenna is presumed to be on a perfectly conducting, infinite ground-plane and its image cast thereof is as illustrated in Figure 4.8.

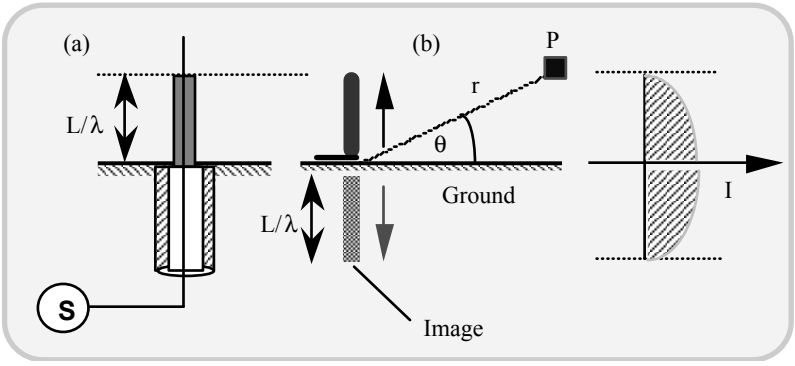


Fig. 4.8 A vertical linear antenna of length L/λ placed over a perfectly conducting, infinite ground-plane
(a) The geometry of the antenna, its disposition with respect to the ground-plane and the associated excitation method.
(b) Image-specified dipole configuration and the current distribution along the length, L.

In terms of the power input P, the electric field intensity **E**(r, θ) at a far-field point P(r, θ) can be deduced from equation (4.24a) and using the relation $E_\theta = 120\pi H_\phi$. The result is as follows:

$$E(r,\theta) = \frac{60}{r}\left(\frac{P}{R_{11}+R_{1L}}\right)^2 \times \left\{\frac{\cos[(2\pi L/\lambda)\sin(\theta)] - \cos[2\pi L/\lambda]}{\cos(\theta)}\right\} \qquad (4.29)$$

where R_{11} depicts the *self-resistance* of the antenna referred to the point of current maximum and R_{1L} is the *loss resistance* (if any) of the antenna referred to the same point. (It should be noted that, in deriving equation (4.15), the effect of loss resistance was ignored; that is, $R_{1L}$was taken equal to zero). The self-resistance indicated above is same as the radiation resistance defined earlier.

The radiation pattern in the vertical plane depicting **E**(r, θ) of the antenna geometry in question is invariably influenced by the following two considerations:

- The conduction-loss introduced by the ground-plane (due to the finite conductivity)
- The edge-diffracted field from the edges of the finite-sized ground-plane superimposing on the main field of the antenna.

In reference to the first, the field at a large distance is the resultant of the direct ray from the vertical dipole and the ray reflected by the ground of finite conductivity. In this case, for the geometry shown in Figure 4.8(a), the vertical plane pattern $E_{11}(\theta)$ can be written as,

$$E_{11}(\theta) = [\cos(\theta)](1+\rho_{11}\angle\psi) \qquad (4.30)$$

where ρ_{11} is the reflection coefficient at the interface of free-space and the ground-plane for vertical polarisation of the incident radiation; and ψ is given by,

$$\psi = \cos\left[\frac{2\pi h}{\lambda}\sin(\theta)\right] + j\sin\left[\frac{2\pi h}{\lambda}\sin(\theta)\right] \qquad (4.31)$$

where the argument $(2\pi h/\lambda)\sin(\theta)$ accounts for the difference in the path lengths between the direct and reflected rays.

Thus, the pattern distortion that may arise when an antenna is placed in the vicinity of a ground-plane stems from the relative amplitude and phase of the direct ray (from the antenna) and the reflected ray (from the ground). For an imperfectly conducting ground-plane, the reflected-ray amplitude would tend to be small and the associated phase relations may not cause a complete signal cancellation (implying a null in the pattern). Further, the gain in the direction

where the fields add would be less than the case where the ground-plane is a perfect conductor. A qualitative set of sketches comparing the field patterns of an antenna structure of Figure 4.8(a) is presented in Figure. 4.9.

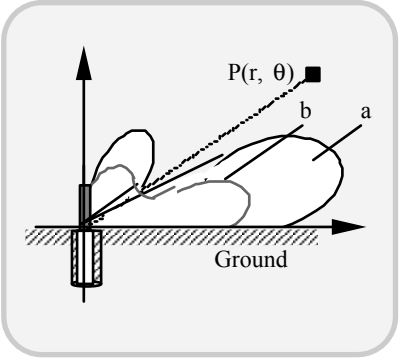


Fig. 4.9 Effects of finite conductivity of the ground-plane on the antenna pattern of a vertical antenna
Case (a) Ground conductivity tends to infinity
Case (b) Finite ground conductivity leading to the absence of a total null in the pattern and a reduced maximum directivity

Horizontal linear antennas over a ground-plane

Suppose a Hertzian dipole is placed at a height, h horizontal to the ground-plane, as illustrated in Figure 4.10. In this case, the image of the antenna can be modelled such that it supports a current distribution same as that of the antenna, but with a phase reversal of 180°. It can further be assumed that the tangential **E**-field is zero everywhere on the perfectly conducting ground-plane.

Relative to a λ/2-antenna placed in the free-space, an horizontal λ/2-antenna (kept at a height h above the ground-plane) has a gain that can be written as follows:

$$G(\alpha)\big|_{GP-FS} = \left[\frac{R_{11}+R_{1L}}{R_{11}+R_{1L}-R_m}\right]\left[2\sin\{\frac{2\pi h}{\lambda}\sin(\theta)\}\right] \qquad (4.32)$$

where the index (GP-FS) explicitly specifies that "the gain of the antenna over the ground-plane under discussion is relative to that of a λ/2-antenna placed in free-space". Further, R_{11} is the self-resistance of a λ/2-antenna, R_{1L} is any loss-resistance associated with the radiator and R_m is the *mutual resistance* of the λ/2-antenna with its image located at a distance of 2h.

When the ground-plane is not a perfect conductor, the reflected ray from the ground surface would correspond to a reflection coefficient pertinent to the

horizontal polarisation, namely $\rho_{\perp}$, which is given explicitly by the following relation:

$$\rho_{\perp} = \frac{\sin(\theta) - \sqrt{\varepsilon_r - \cos^2(\theta)}}{\sin(\theta) + \sqrt{\varepsilon_r - \cos^2(\theta)}} \tag{4.33}$$

And, the corresponding far-field for an incident field amplitude E_o is specified by,

$$E_{\perp}(\theta) = E_o[1 + |\rho_{\perp}|\angle\psi] \tag{4.34}$$

where ψ is the phase angle of the reflection coefficient, $\rho_{\perp}$. It is given explicitly by,

$$\psi = \arctan[I(\rho_{\perp})/R(\rho_{\perp})] \tag{4.35}$$

with $R(\rho_{\perp})$ and $I(\rho_{\perp})$ denoting respectively, the real and imaginary parts of $\rho_{\perp}$.

Thus, the influence of finite conductivity of the ground-plane on the radiation pattern of a horizontal antenna over a ground plane is again similar to that of a vertical antenna over a ground-plane. In the relevant antenna patterns, a general reduction in gain along the beam maximum and an ill-defined (nonzero) null are observed characteristically.

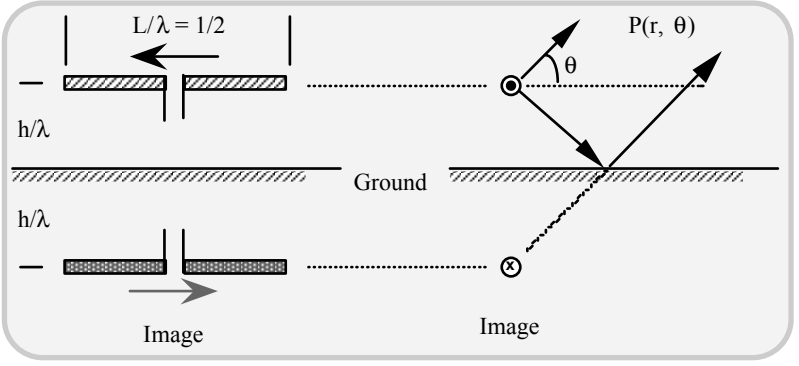


Fig. 4.10 A λ/2-dipole placed horizontally over a ground-plane: (a) Geometry and (b) image consideration

If a ground-plane of finite size is considered, the relevant analysis will be more involved since the resultant field is a superposition of direct-, reflected-, and edge-diffracted rays arising from the inevitable current discontinuity at the edges of the finite-sized ground-plane. A comprehensive analysis of such edge-diffraction

conditions can, however, be done *via geometrical theory of diffraction* (GTD) as discussed in [4.21]. For example, considering an antenna mounted on the body of an automobile, the ground-plane can be modelled as three finite-sized surfaces comprised of the trunk region, the roof and the bonnet section of the vehicle as depicted in Figure. 4.11.

The GTD analysis in the case of a radiator mounted on the body of a vehicle can be applied for the edges along z- and x-directions. Similar GTD-based analysis is also applicable to antennas mounted on a portable radio. The influence of a conducting body that supports the antenna can then be assessed by depicting the body by a suitable (volumetric) geometry with known locales of edge diffraction.

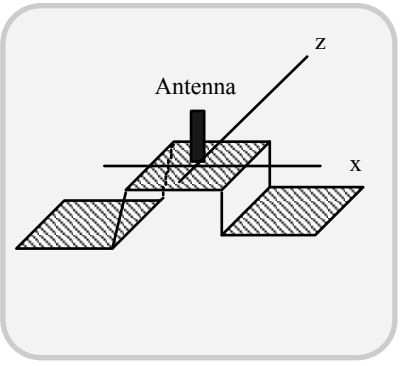


Fig. 4.11 A segmented, finite-sized ground-plane representation of an automobile with a roof-top mounted antenna

Sleeve antenna

In view of the details furnished in the previous section, it can be concluded that the radiation characteristics of a linear antenna will be modified whenever the antenna is mounted on, or placed in proximity to a ground-plane. Therefore, either by proper design of the ground-plane where feasible (in terms of its size, shape and conductivity) or by choosing an appropriate location on the ground-plane (again, subject to such a feasibility), the radiation patterns with certain desirable properties can be achieved to a certain extent.

For example, a $\lambda/4$-stub with an outer conductor shaped in the form of a conical skirt, as illustrated in Figure 4.12(a), constitutes a typical ground-plane modified antenna. When the cone or skirt part of this antenna is degenerated into a sleeve (cylinder), as shown in Figure 4.12(b), the resulting radiator is popularly known as a *sleeve antenna*.

Usually, the sleeve is made of a centre conductor of a coaxial cable connected to an element whose length is $\lambda/4$, and the outer conductor of the coaxial cable is connected to a cylindrical skirt (whose length is again equal to

λ/4). The coaxial (cylindrical) skirt behaves like a λ/4-choke and prevents the RF current leaking into the outer surface of the coaxial feeder line.

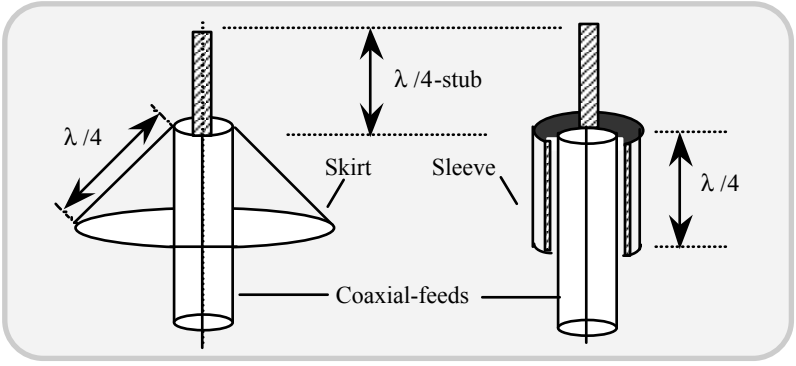


Fig. 4.12 Examples of ground-plane dependent antennas: (a) Conical-skirt antenna and (b) sleeve antenna

The concept of λ/4-choke is based on the following transmission line principle: Consider the sleeve structure shown in Figure 4.13(a). The input impedance indicated at the open end of the sleeve in Figure 4.13(b) corresponds to the input impedance (Z_{in}) of a short-circuited transmission line. Explicitly, $Z_{in} = jZ_0\tan(2\pi\ell/\lambda)$, assuming a lossless line of characteristic impedance Z_0. For $\ell = \lambda/4$, $Z_{in} \rightarrow \infty$ posing an open-circuit at the rim of the sleeve, as illustrated in Figure 4.13(b). Thus, at this point, the function of the RF-choke is to emulate a condition that impedes any current leakage beyond this point on to the coaxial outer conductor, thus preventing the edge-diffraction effects on the radiation patterns.

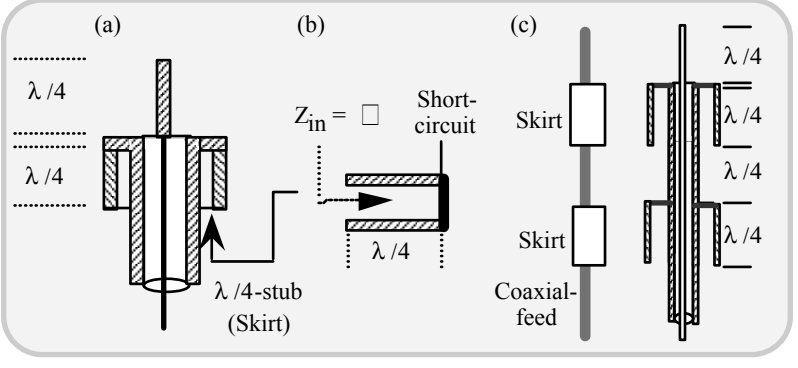


Fig. 4.13 Transmission line concept of a sleeve antenna: (a) Structure; (b) short-circuited λ/4-line and (c) half-wave dipole with a choke

In practice, the RF-choke is added as a "skirt" on the lower section of the coaxial feed-line indicated in Figure 4.13(c) in order to suppress the current leakage beyond the short-circuit. As a result, this structure exhibits almost the same radiation characteristics of a λ/2 Hertzian dipole. Further, since this antenna has suppressed "ground-plane effects" by means of the RF-choke deployed, any related gain degradation and/or pattern distortion are minimal. Also, the associated feed arrangement is simple and is compatible for coaxial lines and standard connectors. This geometry is popular in 800 MHz mobile radio systems in Japan and also been found compatible for portable telephone systems.

Problem 4.1

The sleeve antenna shown in Figure 4.13 emulates a RF-choke as explained above. It is also possible to improvise a set of multiple, concentric sleeves corresponding to a "corrugated" skirt as illustrated in Figure 4.14. Such modifications can yield a better control on RF-choke-specified, undesired spill-over of EM energy and an improved input-VSWR performance.

Elaborate a conceivable design with a 3-sleeve RF-choke arrangement for antenna applications at 900 MHz.

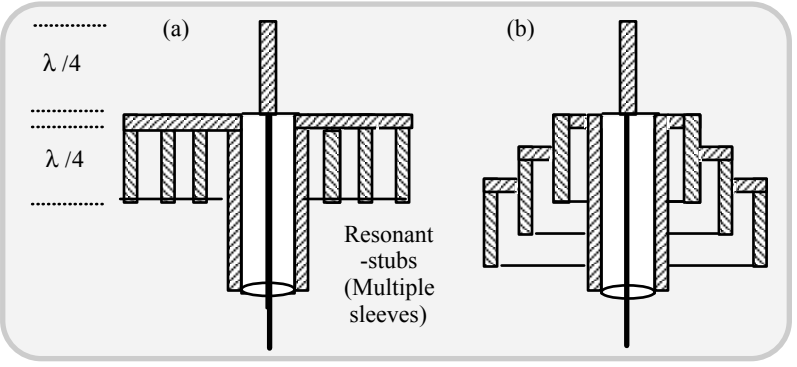


Fig. 4.14 Sleeve antenna with an RF-choke with (multiple) corrugations constituting a multiple-groove skirt: (a) All corrugations set in the same plane; and, (b) corrugations placed off-set from the aperture of the coaxial line

(*Hint*: See [4.22 - 4.24])

Example 4.1

Improvising a single or multiple, concentric sleeves as indicated in the previous problem can be further augmented with a filling of the grooves (single or multiple)

in the choke structure. Again, such dielectric filling could offer a controllable design parameter to realise a better RF-choke-specified reduction of (undesired) spill-over of EM energy and an improved input-VSWR performance. Justify the underlying concept.

Solution

The concept is illustrated in Figure 4.15. The grooves of the corrugations are filled with a dielectric material of relative permittivity, ε_r. This filling would effectively reduce the $\lambda/4$-depths to $\lambda/[4 \times (\varepsilon_r)^{1/2}]$. This offers an additional control feature on the associated antenna/RF-choke design. (See [4.25]).

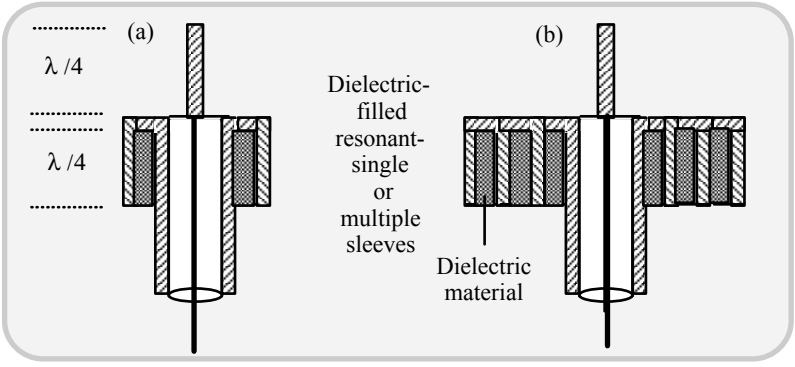


Fig. 4.15 Sleeve antenna with a RF-choke with single/multiple corrugations filled with a dielectric material

---

Problem 4.2

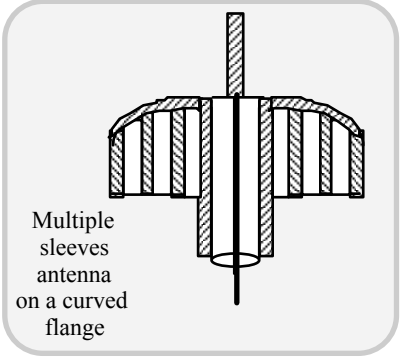


Fig. 4.16 Sleeve antenna with a RF-choke with single/multiple corrugations set on a curved flange (with or without dielectric-filling)

The single or multiple, concentric sleeves specified in Problem 4.1 can further be modified with a curved flange supporting the choke structure as illustrated in Figure 4.16. Discuss the possible merits and demerits of this structure.

(*Hint*: See the publications [4.26] and [4.27] of one of the authors of this book)

Short inverted L- and F-antennas

The geometry of these antennas is ground-plane based, as shown in Figure 4.17. These antennas are low-profile versions and they offer a narrow bandwidth. Largely, they are used on cars, portable units, missiles and so on. Necessary impedance matching is achieved by trimming the dimensions h, d and ℓ.

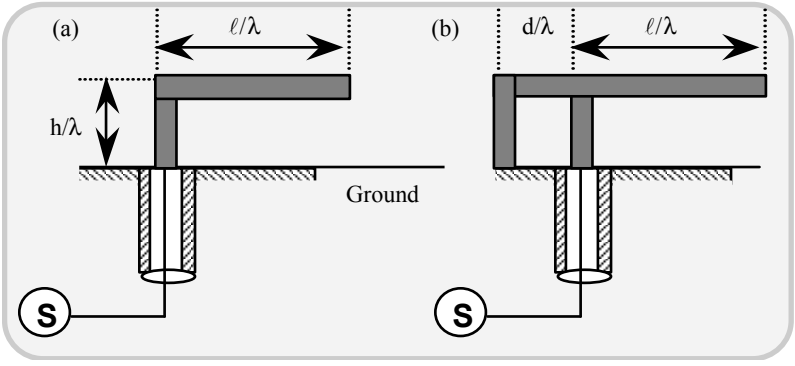


Fig. 4.17 Short inverted antennas: (a) L-type ($h + \ell << \lambda/4$); and, (b) F-type ($h + \ell \approx \lambda/4$ and $d << \lambda$)

King, Harrison and Denton [4.28] developed relevant theories on these antennas. Further, Prasad and King [4.29] have described an experimental study concerning the impedance characteristics of these structures. Originally, these radiators were developed for telemetry and missile applications. Their low silhouette is aptly compatible for their use on the exterior of missiles. These antennas are characteristically of narrow bandwidth and their impedance can be matched through proportioning the associated dimensions. Taga and Tsunekawa [4.30] have done a comprehensive analysis on the performance of built-in planar inverted F-antennas.

The low profile of inverted antennas and the proven use of such antennas in missile structures promoted the wide spread use of these radiators in modern portable and automobile-mounted wireless communication units. RF engineering and the successful integration of multiple wireless standards into low profile antennas are playing significant roles in the development and proliferation of in-

vehicle systems. Before the end of this decade, these systems will transition from an optional luxury to a standard features in automobiles.

Problem 4.3
In reference to the inverted antenna structures shown in Figure 4.17, analyse possible alternatives of improvising a dielectric-coated ground-plane (replacing the traditional ground-plane). Discuss the possible advantages and disadvantages.

(*Hint*: See [4.31-4.32])

Monopoles in wireless communication Systems
A monopole is a λ/4-whip placed over a ground-plane, as illustrated in Figure 4.18(a). The ground-plane of this antenna can be, for example, the body of an automobile or an aircraft. The pattern of this antenna corresponds to that of a λ/2-dipole as a result of the image element posed by the ground-plane. However, its input impedance is only half of the dipole (namely, 36 ohms).

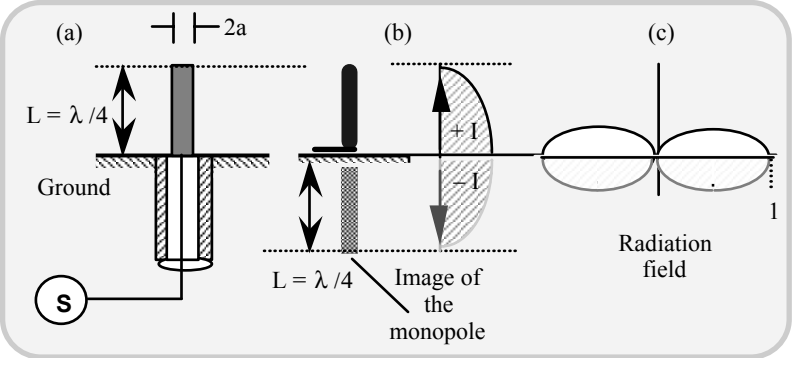


Fig. 4.18 (a) Monopole antenna structure; (b) image of the monopole and the associated current distribution and (c) the radiation pattern

Theoretically, the directivity of a monopole antenna is 3 dB larger (than that of a λ/2-dipole) since the power is confined only above the ground-plane. However, the ideal pattern depicted in Figure 4.18(c) may get distorted, should the ground-plane not be of infinite size. For a finite-sized ground-plane of radius r, the pattern will tilt upward (from the ground-plane) and this tilt would vary inversely with r/λ ratio. More changes in the radiation pattern can be anticipated, if the ground-plane poses a finite conductivity and/or finite complex (surface) impedance [4.31, 4.32].

A typical monopole structure of practical use and of length other than λ/4 is described below.

5/8 wavelength monopole

By increasing the height of the monopole from λ/4 to 5λ/8, the directivity of the antenna will increase facilitating a high-gain structure. A typical directive pattern is shown in Figure 4.19. Because of the non-resonant length involved, a series inductor is used for impedance-matching purposes.

The pattern of the antenna with increased length will be, in general, infested with minor lobes. That is, as the main beam becomes narrower (yielding a higher directive gain), the excess energy will spill out as sidelobes. As long as such sidelobes do not significantly influence the transreceive signal integrity (along the main beam), such minor lobes can be tolerated, in the context of higher directivity achieved along the major lobe.

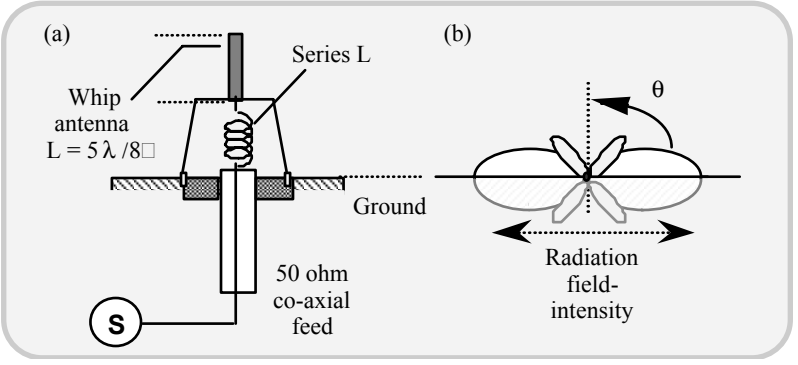


Fig. 4.19 A 5/8λ-whip antenna: (a) Structural details and (b) radiation pattern

Mounting schemes for monopole whips

The categories and general characteristics of monopole whips adopted for automobile applications are indicated in Table 4.1. Further, different ways of mounting a monopole whip antenna on an automobile are illustrated in Figure 4.20.

Auto antennas have classically been in the form of whip, helical or vertical monopoles because these antennas typically produce better RF performance for applications like AM/FM and cellular communication. However, each these antennas may present some drawbacks. For example, a whip antenna needs to radiate the power in a vertical polarisation. But, as the antenna industry drives towards shorter antenna designs pertinent to λ/4 standards, the RF performance of such whip antennas invariably suffers. Further, monopoles and whip antennas are not conducive to multiband configurations.

Table 4.1 Characteristics and application of automobile-mounted short whip antennas

Type of whip mounting	Description	Application
Front/rear-end side-mounted monopole: Figure 4.20(a)	• Mounted on front fender or on trunk-lid • Single-element (1/300λ for MF or 1/3λ for VHF) • Appropriate inductive matching is needed to cancel out short-antenna drive-in capacitive reactance	AM/FM broadcast reception
Slant-element mounted along the pillar in the front: Figure 4.20(b)	Telescopic element	AM/FM broadcast reception
Rear-side wind-shield mounted type: Figure 4.20 (c)	• Mounted on glass • EM coupling *via* window glass • Directivity: 3 dBd	• Car-radios • Mobile phones at 800 MHz
Roof-top mounted single or dual monopoles: Figure 4.20(d) (Two monopoles are used for spatial diversity)	• Single-type (detachable) • Dual-type for spatial diversity • Omnidirectional azimuthal pattern (in both cases)	▪ VHF/UHF band ▪ Mobile phone in UHF band
Bumper-mounted monopole: Figure 4.20 (e)	Inductive loading for matching purposes	HF/VHF band radios

The growing trend in providing multiple, in-vehicle wireless services – cellular and PCS communications, GPS and satellite radio – operating on different frequency bands (ranging from 800 MHz to 6 GHz) requires a separate antenna for

each function. However, automakers are reticent to disturb the aesthetic line of their vehicles by having multiple antennas protruding from a car's body. Hence, to attain the anticipated success of in-vehicle wireless systems, low profile and unobtrusive structures are preferred *in lieu* of whip or monopole structures.

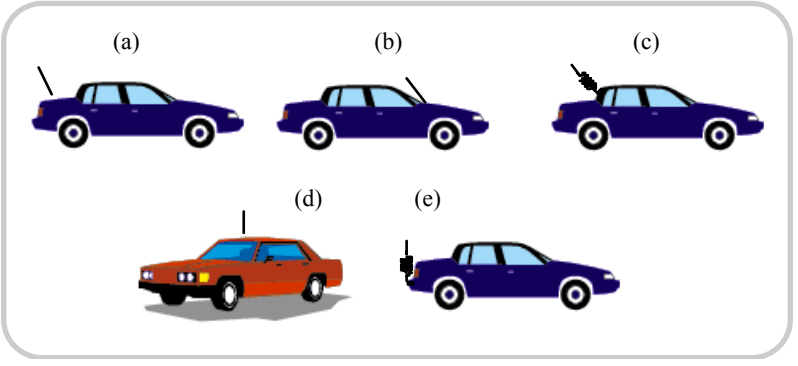


Fig. 4.20 Mounting alternatives for the whip antenna on the body of an automobile. (Details are presented in Table 4.1)

Monopole antennas for airborne applications

Considering the common airborne vehicle geometry, a monopole on a ground-plane refers to a short-antenna on a long, horizontal cylinder. The corresponding azimuth pattern is largely circular (provided that the antenna is at least 0.5λ away from the cylinder edge). This version of antenna is mounted, for example, on the fuselage of an aircraft such that the resulting pattern is compatible for the intended antenna application regardless of the pitch, yaw and roll manoeuvres of the vehicle. A typical, short monopole with a top-loaded disk is illustrated in Figure. 4.21.

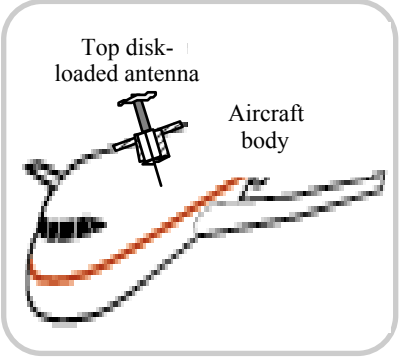


Fig. 4.21 Top-loaded short antenna used in aircraft for ADF applications: Typically mounted on the fuselage

The antenna shown in Figure 4.21 is a low-profile, low-drag structure. Its bandwidth is narrow and it is adopted for LF/MF bands in automatic direction finding (ADF) applications.

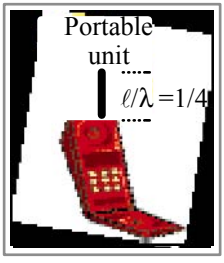
Monopoles in handheld and mobile-computing units

The short-antenna concept is well suited for systems where low-profile considerations are important. For example, in handheld/body-mounted pagers and/or in cell-phones, the monopoles find significant deployment. Also, in mobile computing devices, such as lap-top/palm-held computers with wireless connectivity, the use of low-profile monopole antennas is obvious.

Table 4.2 Monopoles in portable communication devices: Applications and illustrations

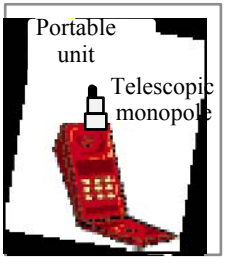
Device	Application
Monopole plus the body of a portable unit as a radiator	VHF/UHF portable devices

Illustration



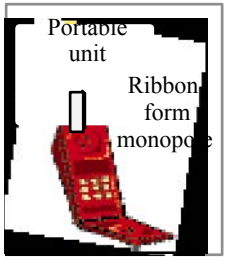
Device	Application
Telescopic version of a λ/4-monopole on a portable unit	VHF/UHF band cordless and portable phones

Illustration

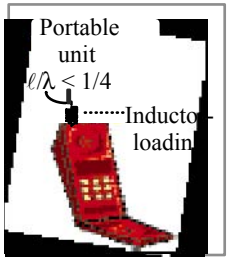


Device	Application
Ribbon-form monopole on a portable unit. (Ribbon-form is flexible and could be of push-in or push-out type)	VHF/UHF band portable phones

Illustration



Device	Application
Subquarter wavelength ($\lambda < \ell/4$) radiator on a portable unit (with inductive-loading)	VHF band, self-tuned devices



Illustration

There are a variety of short antenna configurations that has been conceived and successfully adopted in wireless communication applications. The associated mobile units refer to those body-worn and/or vehicle-mounted. They have been designed to operate at different frequency bands, depending on their applications. Further they are configured for single or multiple band functions. They may also facilitate transreception at linear and/or circular polarisations. In all, the growing trends posed in the recent years in wireless communications have set a pace for visualising the need-based antenna structures in their compact forms.

Shown in Table 4.2 is a set of monopole applications in portable equipment along with their characteristics.

4.3 DIPOLES IN WIRELESS COMMUNICATION SYSTEMS

Among the family of dipoles, sleeve antennas, V-shaped antennas, inverted-L and F radiators, printed dipoles, inductor-loaded simple or sleeve dipoles and crossed dipoles are widely used as stand-alone antenna elements in automobiles.

The sleeve antenna is used in VHF/UHF bands and so are the V-shaped structures. The V-structure dipole is conceived for the purpose of TV reception (VHF/UHF bands) in cars. It gives a directional pattern having horizontal polarisation. Its low-drag characteristics are compatible for automobile installation. This antenna is illustrated in Figure 4.22.

Low-profile radiators giving both vertical and horizontal polarisations are realised using inverted L and/of F structures at HF/VHF/UHF bands.

Printed form of dipole exists in various forms. Printed wire facilitated on the rear window of an automobile (for AM/FM radio) along with a similar element on the quarter window offers diversity reception in practice. Additional TV reception is facilitated with a monopole element, if needed.

Planar printed structure refers to another form of dipole for automobile applications. They are thin and light-weight elements adopted for VHF/UHF mobile telephones. They are made as dashboard mounted units and offer about 8 % bandwidth with a radiation pattern, which is semicircular in the horizontal plane.

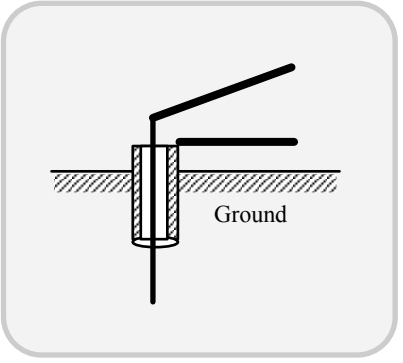


Fig. 4.22 V-shaped antenna placed over a ground-plane: The antenna is fed through a coaxial line

Folded printed dipoles are also used in practice for dual frequency operation. The bands prescribed are: 900MHz and 1.5 GHz. Again, this antenna structure is thin, light-weight and compact yielding a semicircular horizontal pattern with a diversity of about 4.5 dBi.

A simple vertical dipole with inductive loading is used in roof-mounted and/or bumper-mounted configurations for UHF band applications. More than one such antenna element may be used under space diversity considerations.

Collinear $\lambda/2$-dipoles formed by two sleeve dipoles have been conceived for VHF band, in vertical, spatial-diversity specified applications. This structure is mounted on the bumper of the automobile.

Cross-dipoles refer to λ/2-dipole pairs arranged perpendicular to each other so as to achieve polarisation diversity.

Considering portable unit applications, sleeve dipole and planar inverted F-antenna are adopted for UHF band telephony. The λ/2-structure poses less body effects than a λ/4-monopole.

The inverted version is normally built into a portable unit and it offers vertical and horizontal polarisations. They are used at UHF frequencies in portable telephones and relay transmitters.

Inverted-L element finds application in locomotive trains. It is mounted on the roof of the train and used for 150 MHz train-to-station communications. The bent/inverted geometry facilitates low drag.

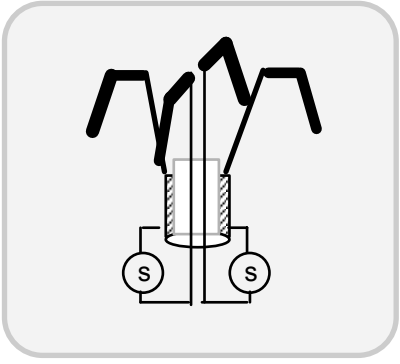


Fig. 4.23 Crossed drooping dipoles: The drooping angle is a design compromise

Dipoles in printed versions with an integrated feeding circuit are used for base station applications (UHF band). A wide beamwidth (180°) is utilised in systems with a corner reflector.

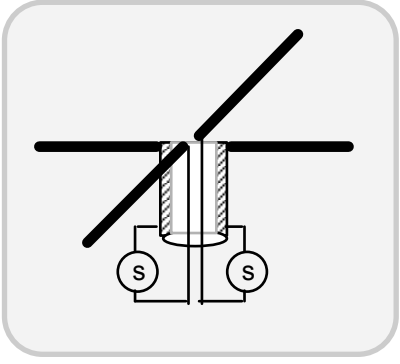


Fig. 4.24 Crossed dipoles: Each dipole is independently fed. (This a special case of drooping crossed dipoles with zero drooping angle)

In maritime applications, crossed (drooping or straight) dipoles are used to get omnidirectional pattern with a low gain (4 dBi) at L-band frequencies used in INMARSAT-C systems. These structures are depicted in Figures 4.23 and 4.24 Another form of cross-dipole structure that is based on inverted V-structure (with thick elements) is used in aeronautical systems at Ku-bands. It offers circular polarisation with good axial ratio over wide angle. It is used for satellite communications.

In aeronautical systems bent/inverted dipoles find wide applications because of their low-drag geometry.

4.4 LINEAR TRAVELLING WAVE ANTENNAS

The linear antennas considered so far have a sinusoidal current distribution along the length, which may be regarded as a result of standing wave produced by two uniform travelling waves of equal amplitude travelling in opposite directions along the length of the antenna.

However, there are other types of antennas such as beverage, rhombic, helical, thick conducting or dielectric cylindrical structures (Figure. 4.25), which support only a single travelling wave (at least to an approximate extent). On such antenna structures, the current amplitude distribution is uniform, but the phase may change linearly as shown in Figure. 4.26. Figure 4.27 depicts a linear travelling wave antenna of length b along the z-axis. A single travelling wave is assumed to be travelling along the antenna in the z-direction and the retarded current distribution is given by,

$$[I] = I_o \sin\left[\omega\left(t - \frac{r}{c} - \frac{z_1}{v}\right)\right] \tag{4.36}$$

where z_1 is a point on the conductor. The phase velocity of the travelling wave is v and a relative phase velocity p can be defined as: p = v/c. Further, since the current is in the z-direction, the Hertz vector (has only a z-component, which leads to a H(component at a distant point.

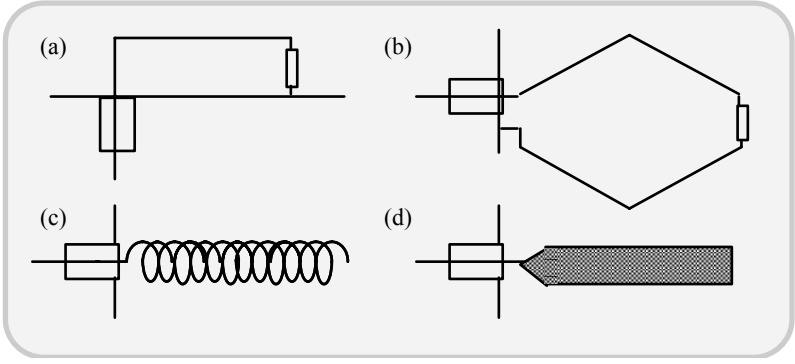


Fig. 4.25 Travelling wave antennas: (a) Beverage antenna; (b) rhombic antenna, (c) helical antenna; and (d) thick conducting or dielectric cylindrical antenna

In reference to a point of observation in the far-field at r >> b and z_1/z (<< 1) → 0, the corresponding H_ϕ field can be deduced with necessary simplifications as:

$$H_\phi = \frac{I_o p}{2\pi r_1}\left\{\frac{\sin(\theta)}{\left[1 - p\cos(\theta)\right]} - \left[\sin\left(\frac{\omega b\left[1 - p\cos(\theta)\right]}{2pc}\right)\right]\right\} \angle\xi \quad (4.37)$$

where $\xi = \omega\,(t - r_1/c) + \omega b[1 - p\cos(\theta)]/2pc$ and $E_\theta = Z_o H_\phi$ with $Z_o = (\mu_o/\varepsilon_o)^{1/2} = 120\pi$ ohms.

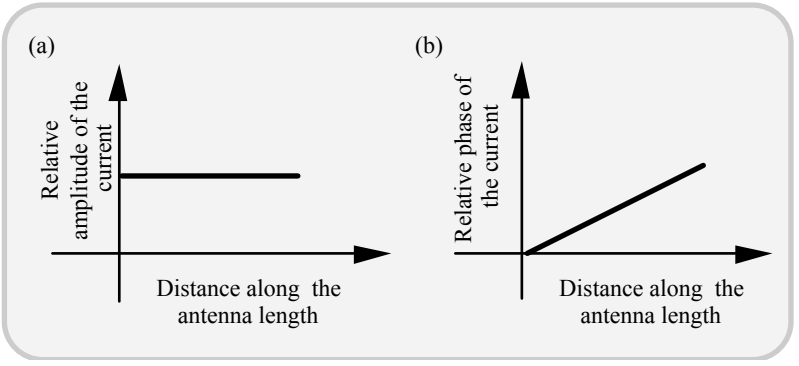


Fig. 4.26 Amplitude and phase distributions of current along linear travelling-wave antenna. (a) Amplitude and (b) phase

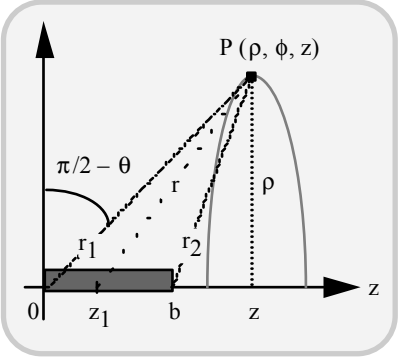


Fig. 4.27 Cylindrical co-ordinate system to analyse linear travelling wave antennas

The amplitude of H_ϕ decides the shape of far-field pattern, and the phase angle determines the phase of the far-field (referred to the origin of the co-ordinates taken as the *phase centre*).

Figure. 4.28 shows the far-field patterns of a λ/2 linear travelling wave antenna for two values of p; and, Figure. 4.29 illustrates the far-field pattern of a 5λ linear travelling wave antenna. It can be noticed that, as the phase velocity v of the travelling wave is reduced, the tilt angle is increased and the beamwidth is reduced. Even for v = c (that is, with p = 1), the lobes are sharper and tilted forward as compared to the standing wave antenna. As the length of the antenna is increased, the tilt angle increases further, reaching a value of 78° when the length is equal to 20λ (for p = 1).

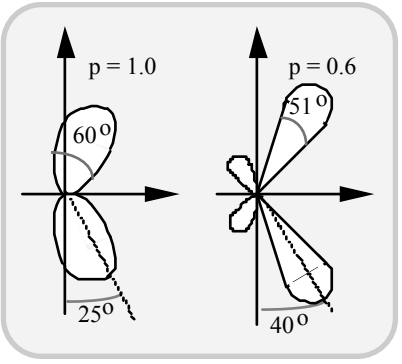


Fig. 4.28 Far-field patterns of λ/2 linear travelling wave antenna: (a) p = 1.0 and (b) p = 0.6

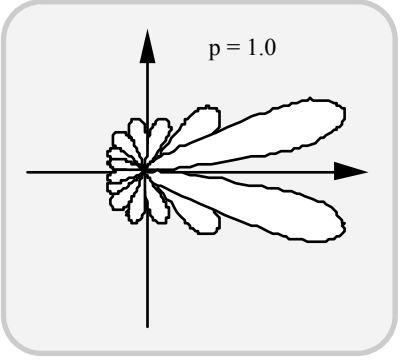


Fig. 4.29 Far-field patterns of a 5λ-long linear travelling wave antenna

4.5 LOOP ANTENNAS

The *loop antenna* refers to a radiating element made of a coil of one or more turns. Both ferrite- and air-core loops are commonly used in RF applications. Classically, these antennas were developed for deployment in direction finders, aircraft receivers and UHF transmitters.

In the current times, loop antennas have also been adopted for wireless communications. In this context, they are conceived as low profile, electrically-small magnetic loops compatible for small radio devices. They are often placed internal to the radio case (unlike the electric dipoles, which are usually appendages to the radio case or used with the radio housing itself as one of the dipole elements).

The radiation pattern of a small loop antenna is identical with that of a small dipole oriented normal to the plane of the loop with the **E-** and **H**-fields interchanged. This arises from the *duality* aspect of Maxwell's equations. For example, the Faraday law governing the **E**-field, namely, $\nabla \times \mathbf{E} = -j\omega\mathbf{B} - \mathbf{M}$ and the Ampere's current law governing the **H**-field, namely, $\nabla \times \mathbf{H} = j\omega\mathbf{D} + \sigma\mathbf{E}$ are mathematically dual to each other.

Hence the radiation characteristics of a magnetic loop and those of an electric dipole would represent a mathematical dual to each other. As such, applying this principle of duality, the radiation characteristics of an infinitesimal current loop made of a circulating current I enclosing a surface area S can be deduced by analogy to the infinitesimal (short) dipole radiation considerations discussed earlier.

The relevant results are summarised in Table 4.3 where the relation depicting the duality aspects of radiation fields due to a small electric dipole and a small magnetic loop are illustrated.

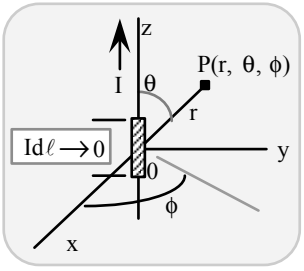
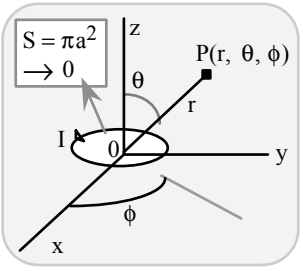
Considering the radiation (far-field) zone where the angular radiation field is essentially independent of distance, the two field components of the small loop antenna of significance are as follows:

$$\left.\begin{aligned} E_\phi &= \left[\frac{120\pi^2}{r}\right] \times \left[\frac{SI}{\lambda^2}\right] \times \sin(\theta) \\ H_\theta &= \left[\frac{\pi}{r}\right] \times \left[\frac{SI}{\lambda^2}\right] \times \sin(\theta) \end{aligned}\right\} \qquad (4.38)$$

(If there are N turns in the loops, then the current term I is replaced by NI in the field expressions.)

The relations of equation (4.38) are valid under the assumptions that the excitation current is uniform and the loop diameter (2a) is invariant, as well as that $2a < \lambda/3$. The relevant pattern corresponds to a figure-of-eight in the zy-plane as shown in Figure 4.30.

Table 4.3 Dual relation depicting the radiation fields characteristics of a small electric dipole and a small magnetic loop

Field components (Electric dipole)	Field components (Magnetic dipole)
	
$E_r = -j\frac{Id\ell}{2\pi} e^{-jkr} Z_o k^2 \times \left[\frac{j}{(kr)^2} + \frac{1}{(kr)^3}\right] \times \cos(\theta)$	$H_r = \frac{kIS}{2\pi} e^{-jkr} k^2 \times \left[\frac{j}{(kr)^2} + \frac{1}{(kr)^3}\right] \times \cos(\theta)$
$E_\theta = -j\frac{Id\ell}{4\pi} e^{-jkr} Z_o k^2 \times \left[-\frac{1}{kr} + \frac{j}{(kr)^2} + \frac{1}{(kr)^3}\right] \times \sin(\theta)$	$H_\theta = \frac{kIS}{4\pi} e^{-jkr} k^2 \times \left[-\frac{1}{kr} + \frac{j}{(kr)^2} + \frac{1}{(kr)^2}\right] \times \sin(\theta)$
$H_\phi = j\frac{Id\ell}{4\pi} e^{-jkr} k^2 \times \left[\frac{1}{kr} - \frac{j}{(kr)^2}\right] \times \sin(\theta)$	$E_\phi = Z_o \frac{kIS}{4\pi} e^{-jkr} k^2 \times \left[\frac{1}{kr} - \frac{j}{(kr)^2}\right] \times \sin(\theta)$

Now considering finite-sized loops of radius a (that are not "small" and carry a uniform current I contrary to the earlier assumption), the radiated field components are as follows:

$$\left.\begin{aligned} E_\phi &= \left(\frac{120\pi^2 aI}{\lambda r}\right) \times J_1\left[\frac{2\pi a}{\lambda}\sin(\theta)\right] \\ H_\theta &= \left(\frac{\pi aI}{\lambda r}\right) \times J_1\left[\frac{2\pi a}{\lambda}\sin(\theta)\right] \end{aligned}\right\} \tag{4.39}$$

where $J_1(\bullet)$ is the Bessel function of the first kind and first order. (The derivation of the relations given by equation (4.39) will be furnished later.) Shown in Figure 4.31 are far-field patterns of loops with different diameters.

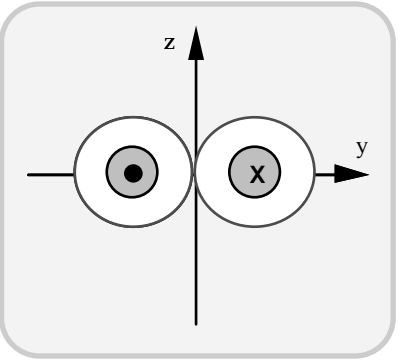


Fig. 4.30 The principal plane pattern of a small magnetic loop placed in the xy-plane

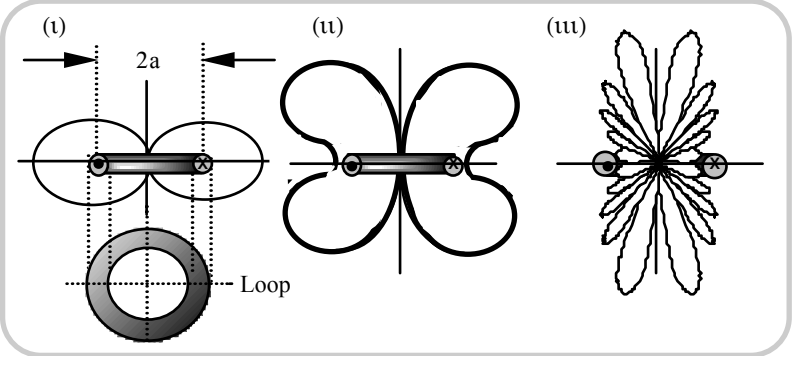


Fig. 4.31 Far field pattern of loops of different diameters: (i) $2a = \lambda/10$; (ii) $2a = \lambda$ and (iii) $2a = 5\lambda$

4.5.1 Radiation resistance of a small loop

By defining the radiation resistance (R_r) of a small loop (in the same manner as that of a linear dipole antenna) namely, $R_r = 2W/I_0^2$, where W is the total power radiated and I_o is the peak current in line on the loop, relevant derivation yields the following result:

$$R_r = 31{,}171(NS/\lambda^2)^2 \text{ ohms} \qquad (4.40)$$

where N is the number of turns in the loop.

For large-sized loops (of finite radius a), the corresponding result on the radiation resistance is given by:

$$R_r = 60\pi^2 (2\pi a/\lambda) \int_0^{4\pi a/\lambda} J_2(y)dy \quad \text{ohms} \qquad (4.41)$$

where $J_2(\bullet)$ is the Bessel function of the first kind and second order. When $2\pi a/\lambda \geq 5$, R_r is approximately equal to $3720 \times (a/\lambda)$ ohms.

4.5.2 Directivity of a circular loop antenna

Consistent with the definition of directivity (D) of an antenna being the ratio of maximum radiation intensity to the average radiation intensity, the parameter D specific to a loop antenna can be specified as follows:

$$D = (r^2 \mid \mathbf{P}_r \mid)/(W/4\pi) \qquad (4.42)$$

where $\mathbf{P}_r$ is the radial component of the Poynting vector and is equal to $(1/2)|\mathbf{H}|^2 Z_0$ (with $|\mathbf{H}|$ being the magnitude of H_θ component) and W is the total radiated power, namely, $\iint_S |\mathbf{P}_r| da$.

A rigorous analysis of a circular loop antenna (of finite radius) leading to its H_θ component is as follows: Consider a large-sized circular loop of radius a, and carrying uniform, in-phase current as shown in Figure 4.32. The loop is placed in the xy-plane. (While the assumption of uniform in-phase current can be easily realised in a small loop, for large loops with a perimeter greater than $\lambda/4$, phase shifters should be introduced around the periphery of the loop in order to obtain a uniform in-phase current carried by the loop.)

Now, consider a pair of infinitesimally short and diametrically opposite electric dipoles, each of length $d\ell = (ad\phi)$ as shown in Figure 4.32. Since the current flows along the loop, it has a component only in the ϕ-direction, and hence the magnetic vector potential has also only a ϕ-component.

The vector potential dA_ϕ at a distant point P from the two diametrically-opposite and infinitesimally-short electric dipoles under consideration is given by:

$$dA_\phi = \frac{\mu_o d\mathbf{M}}{4\pi r} \tag{4.43}$$

where d**M** is the current moment due to this pair of electric dipoles carrying current I and of length $\ell = (ad\phi)$.

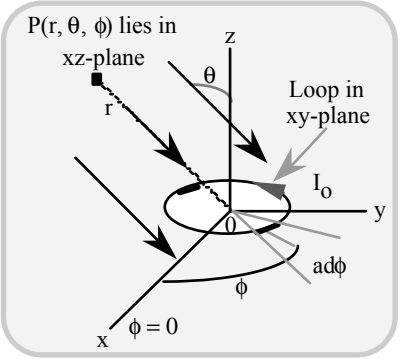


Fig. 4.32 A loop of radius 'a' placed in the xy-plane: Illustration of the co-ordinate system

The cross-section of the loop in the xz-plane is shown in Figure 4.33. In this plane $\phi = 0°$, and the ϕ-component of retarded current moment due to a single dipole is:

$$[I](a\, d\phi) \times \cos(\phi) \tag{4.44}$$

where $[I] = I_o \exp\{j\omega(t - r/c)\}$.

Now, considering both diametrically opposite dipoles, the resultant moment due to this pair, namely dM, is given by,

$$dM = 2j[I](a\, d\phi) \times \cos(\phi) \times \sin(\psi/2) \tag{4.45}$$

where $\psi = (2ka) \times \cos(\phi) \times \sin(\theta)$. Hence, considering the entire loop, the total magnetic vector potential can be written as:

$$A_\phi = \int_{\phi=0}^{\pi} dA_\phi$$

$$= \frac{j\mu_o[I]a}{2\pi r} \int_0^{\pi} \sin[\beta a \cos(\phi)\sin(\theta)]\cos(\phi)\, d\phi \tag{4.46}$$

where J_1 is the Bessel function of the first kind and first order.

Correspondingly, the far electric field will have only a ϕ-component, which can be explicitly written as:

$$E_\phi = -j\omega A_\phi$$

$$= \frac{\mu_o \omega [I] a}{2r} J_1[(\beta a) \times \sin(\theta)]$$

$$= \frac{60\pi\beta a[I]}{r} J_1[(\beta a) \times \sin(\theta)] \qquad (4.47)$$

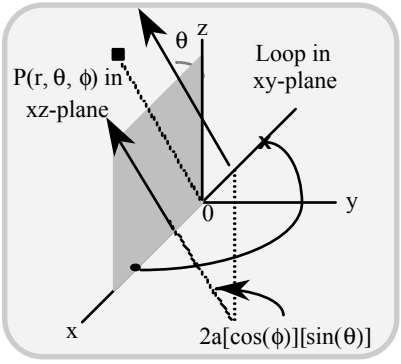


Fig. 4.33 Cross-section of a loop placed in the xy-plane: P is a remote point of observation

On the other hand, the magnetic field at a large distance will have only a θ-component given by,

$$H_\theta = E_\phi / Z_o$$

$$= \frac{\beta a[I]}{2r} J_1[(\beta a) \times \sin(\theta)] \qquad (4.48)$$

As such, it follows that,

$$W = \iint P_r da = \iint_S \frac{1}{2} |\mathbf{H}|^2 da$$

$$= 15\pi (kaI_o)^2 \times \left[\int_{\phi=0}^{2\pi} \int_{\theta=0}^{p} J_1^2 \left[\frac{2\pi a}{\lambda} \sin(\theta) \right] \times [\sin(\theta)] d\theta d\phi \right] \qquad (4.49)$$

Performing the integrations in equation (4.49) and simplifying, it leads to:

$$W = [30\pi^2 kaI_0^2] \int_{\phi=0}^{2\pi a/\lambda} J_2(y)dy \tag{4.50}$$

where $J_2(\bullet)$ is the Bessel function of the first kind and second order.

Using equation (4.50), the directivity as defined earlier *via* equation (4.42), can be obtained as:

$$D = \frac{2C_\lambda \left\{ J_1^2 \left[C_\lambda \times \sin(\theta) \right] \right\}_{max}}{\int_0^{2\pi a/\lambda} J_2(y)dy} \tag{4.51}$$

where C_λ is the circumference of the loop normalised with respect to the wavelength. For a small loop ($C_\lambda < 1/3$), D reduces to:

$$D = \left[\frac{3}{2} \sin^2(\theta) \right]_{max} = \frac{3}{2} \tag{4.52}$$

(It may be noted that this value of 3/2, also refers to the directivity of a short electric current element or dipole).

For a large loop, D can be determined as,

$$D = 2(C_\lambda) \times J_1^2 \left[(C_\lambda) \times \sin(\theta) \right] \tag{4.53a}$$

For loops with $C_\lambda \geq 1.84$, the maximum value of $J_1[(C_\lambda) \times \sin(\theta)]$ is equal to 0.582. Hence, for such loops,

$$D = 0.68 \times (2\pi a/\lambda) \tag{4.53b}$$

Figure 4.34 shows the directivity D as a function of $C_\lambda = 2\pi a/\lambda$.

4.5.3 Fresnel zone and induction zone fields of a radiating magnetic loop

The region where $C_\lambda = 2\pi r/\lambda \rightarrow 1$ refers to the so-called the *Fresnel*- or the *near-field zone*. The boundary of transition from the near-field to the far-field in antenna technology is not well defined. However, it is approximately specified by an inner boundary with $r^2 > 0.38D^3/\lambda$ and an outer boundary specified by $r < 2D^2/\lambda$. Here, D refers to the largest dimension of the antenna and r is the radius of the sphere (in the propagating space) at which the EM field is observed and the antenna in

question is assumed to reside at the centre of this sphere. The outer boundary is limited by the phase error not exceeding $\pi/8$.

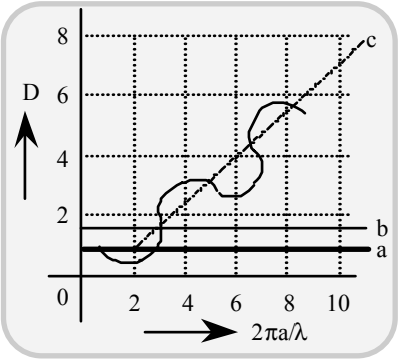


Fig. 4.34 Directivity of a circular loop carrying uniform, in-phase current as a function of $2\pi a/\lambda$: (a) Directivity of an isotropic source; (b) directivity of a loop antenna of the size $1.5 \times 2\pi a/\lambda$ and (c) linear fit on the directivity of a loop antenna of the size $0.68 \times 2\pi a/\lambda$

The near-field zone is characterised by the prevalence of a significant radial component of the field. Considering the far-zone, the field components (**E** and **H**) are entirely transversal to the radial direction of the EM wave propagation. That is, the wave is essentially a TEM wave. On the contrary, in the near-field there could be radial **E** or **H** components leading to TM and/or TE modes of propagation respectively. While the far-field zone field varies inversely as r, the field variation in the near field domain is proportional to $(1/r^2)$.

The so-called *induction zone* applies to the regime where the field components are still "tied" or coupled to the radiating element. In reference to wireless small radio modules, the radiator (such as a part of a pair of loops) is placed or geometrically oriented in such a fashion, that it poses minimum coupling in the induction zone. Should any such coupling still prevails, the resulting far-field pattern (due to the pair of loops in question) may change.

4.5.4 Q-factor of a small loop antenna

The small loop antenna can be characterised by its radiation resistance R_r, (which is proportional to the fourth power of the loop radius, a). Its reactance part, however, is proportional to the loop radius. Therefore, the Q-factor of this antenna will vary inversely proportional to the third power of the loop radius. Further, for a loop made of a wire with a diameter d, and inner and outer diameter of the loop, (respectively) being 2a and 2a′ (so that d = a′ – a), the corresponding feed-point impedance (Z_{in}) can be determined as,

$$(Z_{in})_{Loop} = Z_o[\frac{\pi}{6}(ka')^4] + jZ_o ka'[\ln(\frac{2a'}{d}) + 0.0966] \qquad (4.54)$$

Hence, the corresponding unloaded Q of the loop antenna is as follows:

$$Q_{unloaded|loop} = \frac{\frac{6}{\pi}[\ln(\frac{2a'}{d}) + 0.0966]}{(ka')^3} \qquad (4.55)$$

which for large values of 2a′/d (for example, 10) will reduce to $4.6/(ka')^3$. The above expression of equation (4.55) for Q does not include the dissipative losses. Further, the value obtained from this expression will be larger than $1/(kr)^3$, which denotes the free-space, *fundamental mode Q* of a radiator. (The fundamental mode Q represents the minimum possible Q-factor of a radiating system.)

In practice, due to inevitable presence of finite dissipation effects, the value of Q obtained *via* equation (4.55) will tend to depict the quality factor of free-space modes, namely $1/(kr)^3$.

A simple-loop antenna with its feed reactance being inductive, can be resonated with a series capacitor that nullifies the inductive influence of the loop. The value of this capacitance is given by $C = 1/(2\pi f)^2 L$, where $L = \mu_0 a'[\ln(2a'/d) + 0.0966]$.

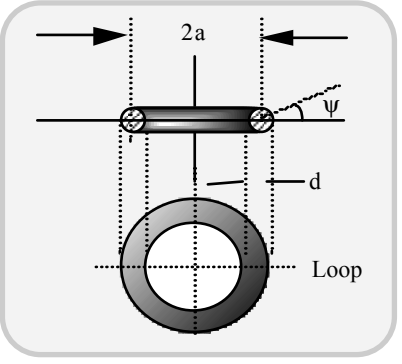


Fig. 4.35 Cross-section of loop made of a finite-diameter wire: Mean loop diameter is "2a"

A loop antenna that has a large wire diameter (d) will have a current density that varies along the circumference of the wire. In such cases, relevant to the cross-section of the wire shown in Figure 4.35, an approximate expression for the current density can be deduced as:

$$J_\phi = \left(\frac{I_\phi}{\pi d}\right)\frac{1}{\left[1+\frac{d}{2a'}\cos(\psi)\right]} \qquad (4.56)$$

where I_ϕ denotes the loop current varying along the loop perimeter.

4.5.5 Non-circular loops

The far-field patterns of a circular loop and that of a square loop (both having the same loop area, S) can be considered as identical as long as the loops are electrically "small" (with $S < \lambda^2/100$ or so). For large sizes however, the circular loop has a pattern, which is a function of θ only. On the other hand, a large square loop has a pattern that depends on both θ and ϕ; and, the variation along θ of the square loop would be different from that of circular loop. (In making, the above comparisons, uniform in-phase currents are assumed in both cases.)

4.5.6 Radiation efficiency of a loop antenna

Specifying the gain of an antenna (with respect to an isotropic radiator) as G, it will be identically equal to the directivity D, if the radiation process poses no losses. Should losses, however, be present, G is then equal to $\eta_R D$, where $0 \le \eta_R \le 1$ is an *antenna efficiency factor* that can be explicitly related to the radiation resistance R_r and the loss resistance R_{loss} by the following expression:

$$\eta_R = R_r/(R_r + R_L) \qquad (4.57)$$

For a small loop antenna, $R_r \approx 31{,}200(S/\lambda^2)^2$ and R_{Loss} can be deduced from the skin effect considerations associated with the conductor of the loop. Explicitly,

$$R_{Loss} = (\text{Perimeter of the loop})/(\pi d\sigma\delta)$$

$$= (\text{Perimeter of the loop}/d)[f\mu_o/\pi\sigma]^{1/2} \quad \text{ohms} \qquad (4.58)$$

where d is the diameter of the wire, σ is the conductivity of the wire material, δ is the skin-depth, f is the operating frequency and μ_o is the absolute permeability of free-space.

Hence for a loop of perimeter C and N turns (as decided by its radius $(a + a')/2$), $\eta_R = 1/[1+(R_{Loss}/R_r)]$ with,

$$\frac{R_{Loss}}{R_r} = \frac{3430}{[C^3 \times (f_{MHz})^{7/2} \times d \times N]} \qquad (4.59)$$

where C and d are expressed in metres.

The radiation efficiency (η_R) of a loop antenna can be enhanced by introducing a low-loss or a lossless ferromagnetic core of high permeability. A

ferrite rod can serve this purpose. Ferrite is essentially a low-loss ceramic constituted by a stoichiometric formulation, $x(MO).yFe_2O_3.zH_2O$, where M is a divalent material (such as Mn) and (x, y and z) are the stoichiometric proportions of the constituting compounds. Ferrite can be formed to have a high permeability [4.32].

With ferrite-loading of a loop of N turns, the radiation resistance (R_r) and the loss resistance (R_f) in ohms can be ascertained from the following relations:

$$\left.\begin{aligned} R_r &= 31,200 \times \mu_{er}^2 \times N^2 \times (S/\lambda^2) \\ &= 197 \times \mu_{er}^2 \times N^2 \times (C/\lambda)^4 \\ R_f &= (2\pi f) \times \mu_{er} \times (\mu_r'' / \mu_r') \times \mu_o \times N^2 \times (A_o / \ell) \end{aligned}\right\} \quad (4.60)$$

where f is in Hz; A_o is in m^2; ℓ is in meter; $\mu_o = 4\pi \times 10^{-7}$ H/m (and other terms are all dimensionless). The term $(\mu'_r - j\mu''_r)$ denotes the relative (complex) permeability of the ferrite material. Further, μ_{er} corresponds to the relative effective permeability of the ferrite rod, A_o is the cross-sectional area of the ferrite core and ℓ is the stretched length of the ferrite ring. (*Note*: The factor μ_{er} is introduced to account for the *demagnetisation effect* that may be encountered in the ring geometry of the ferrite core in contrast to its open geometry. This value of μ_{er} is dependent on length/diameter ratio of the ferrite rod.)

Hence with ferrite-loading, the radiation efficiency of an (N-turn) loop can be written as follows:

$$\eta_R = \frac{R_r}{R_r + R_f + R_L} = \left\{1 + \left(\frac{R_f + R_L}{R_r}\right)\right\}^{-1} \quad (4.61)$$

And, the Q-factor (depicting the ratio of energy stored to energy lost per cycle) of a ferrite-loaded loop-antenna can be specified as:

$$Q = \frac{(2\pi \times f_o \times L_c)}{(R_r + R_f + R_L)} \quad (4.62)$$

where L_c is the inductance of the ferrite-loaded coil and it can be explicitly determined from the expression $L_c = (\mu_{er} N^2 A_o \mu_o / \ell)$ H; and, f_o denotes the frequency of operation in Hz.

The high permeability of ferrite cores renders a ferrite-loaded coil an ideal magnetic dipole. In practical units, the actual loaded-Q of the ferrite-core antenna may be close to 150; the unloaded-Q would correspond to an ideal magnetic or electric dipole, which in practice, will lead to the unloaded-Q being in the order of 10^6. The resulting antenna efficiency namely $Q_{actual}/Q_{unloaded}$ will be about – 41 dB and the actual bandwidth will correspond nearly equal to 200 kHz.

4.5.7 Loop antennas in wireless communication systems

Loop antennas (with or without ferrite-loading) may be used in wireless communication applications, either independently as stand-alone structures, or in conjunction with another form of antenna element (such as a dipole).

Essentially this family of antennas facilitates small-size and low-profile considerations compatible for small radios and/or pager units. They are usually placed internal to the radio case and are gap-fed. They can be resonated with a series capacitor if needed.

Both air-cored as well as ferrite-cored loops are adopted in practice. The number of turns is another design parameter of these radiators as could be evinced from the analytical expressions indicated above.

On automobiles for example, a small loop built inside the mirror box can be designed for MF band radios, as well as for remote-control of doors facilitating a keyless entry using a radio-card transponder.

A combination of a monopole and a loop element is also adopted in UHF bands to offer polarisation diversity as described in [4.34].

A ferrite-loaded multiturn loop-antenna can also be used in series with a single turn, air-core loop, as illustrated in Figure 4.36. Such structures are compatible for wrist-mounted pager applications [4.35].

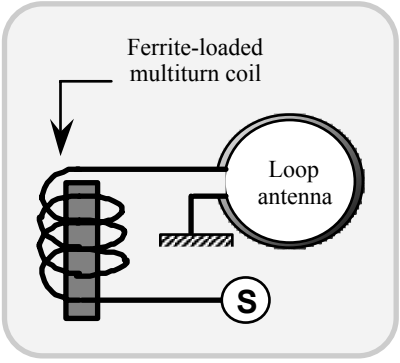


Fig. 4.36 Loop-plus-(ferrite-loaded) coil antenna: For inductive pick-up/near-field applications

A ferrite-loaded coil made as a small cylindrical geometry can be used in LF bands, for a narrow bandwidth transreceive applications pertinent to bus-to-roadside inductive communication systems. In such systems, an electromagnetic near-field coupling is used at the operating frequency (such as 100 MHz). A set of stations are established along the roadside (called roadside or signpost stations), which communicate with a mobile unit (a bus, for example) travelling on the road using this system of antennas. A practical transreceive system of this type, for example, can support 200 bit data transmissions at 4.8 kbps.

Notably, the ferrite-loading allows the realisation of low profile/small-size antennas, enables high (inductive) coupling efficiency and facilitates a desired zone pattern that can be synthesized via appropriate transmitter-receiver antennas.

4.5.8 Loop plus dipole antenna

A field polarisation diversity antenna for portable phones can be designed using a dipole terminated with a parallel loop capacitor. It has a single-point feed arrangement incorporated at a specified point at the opposite junction-points of the loop, as indicated in Figure 4.37. The capacitor in Figure 4.37 enables a phase-shift between the dipole and the loop currents so that the presence of any deep null can be negated in the pattern in the presence of reflecting objects [4.36].

The design of a loop-plus-dipole antenna can be optimised for portable transceivers (including cellular and PC handsets) operating at 900 to 2000 MHz. Specifically, considering the operations in urban environments (where near planar reflectors such as the walls of a building etc. may prevail in the vicinity of the phone), this antenna is conceivably useful. The probability of null in those circumstances is reduced as a result of the set of using the collocated antenna elements.

The superposition of individual field patterns (due to the dipole and the loop) enables such reductions. In these electrically-small antenna designs, the output voltage given by:

$$V = K[E_a + Z_o e^{j(\gamma+\pi/2)} H_a] \qquad (4.63)$$

where K is a constant (with units of length), E_a and H_a are field components respectively along the *a*-direction of the dipole orientation, picked up by the loop and the dipole respectively. Further, Z_o is the free-space characteristic impedance of the free-space (= 120π ohm), and γ (= $\pi/2$) depicts the phase-shift introduced by the capacitor. Further, as mentioned earlier, the antenna geometry is tailored so that the loop and dipole voltages induced are equal in magnitude with $H_a = E_a/Z_o$.

The motivation behind this two-element antenna design, as discussed in [4.37], is as follows: Suppose the antenna under discussion has no capacitor connected. Then $\gamma = 0$ and $V = KE_{oz}$ since the a-axis of the antenna (lying in the yz-plane) will be tilted at an angle θ_o with respect to the z-axis. Therefore, $E_a = E_{oz}\cos(\theta)$ and $H_a = (E_{oz}/Z_o)\sin(\theta_o)$. Thus, in the absence of the capacitor, the signal-voltage induced at the antenna output is independent of the tilt-angle θ_o as well as the position of x-axis. On the other hand, with the capacitor connected, ($\gamma = \pi/2$ leading to $V = KE_o|\cos(\theta_a) - \sin(\theta_a)|$, which would cause a null at $\theta_0 = \pi/4$.

When the antenna under consideration is placed in front of a reflecting object with its plane with normal along the x-axis, as shown in Figure 4.37, the field components are $E_a = -2jE_0\cos(\theta_a)\sin(k_o d_{ar})$ and $H_a = -2(E_o/Z_o)\sin(\theta_o)\cos(k_o d_{ar})$. This is under the assumption that the reflection corresponds to a perfect reflector. Further, d_{ar} denotes the spacing between the antenna and the reflector and $k_o = 2\pi/\lambda_o$, with λ_o being the free-space wavelength of operation.

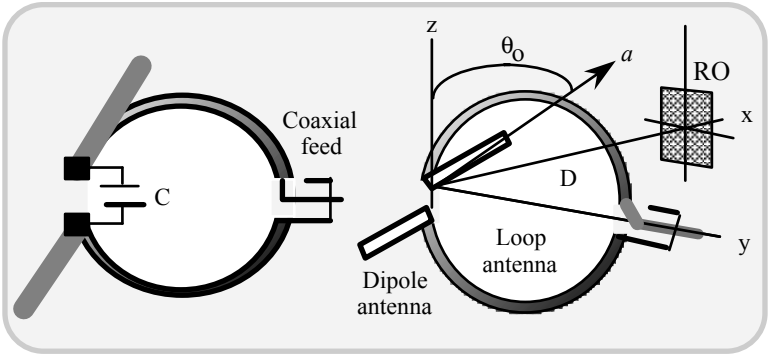


Fig. 4.37 Loop-plus-dipole antenna: C is a capacitor introduced for a necessary phase-shift between the loop and the dipole; RO: Reflecting object in the vicinity

Under the assumption, $\gamma = 0$, $V = 2KE_0|\cos(\theta_o)\sin(k_0 d_{ar}) + \sin(\theta_o)\cos(k_0 d_{ar})|$, which will become zero for certain discrete values of θ_o and at periodic distances from the reflector. If a capacitor is introduced, $\gamma = \pi/2$ leading to a quadrature addition of E_a and H_a. Hence, $V = 2KE_0|\cos^2(\theta_o)\sin^2(k_0 d_{ar}) + \sin^2(\theta_o)\cos^2(k_0 d_{ar})|^{1/2}$ and nulls would disappear. By choosing γ appropriately, the probability of a null occurrence can be reduced with and without the reflector being present. The respective conditions indicated in [4.37] are: $\pi/8 \le \gamma \le \pi/2$ and $\pi/8 \le \gamma \le \pi/4$.

4.6 HELICAL ANTENNAS

A *helical antenna* has a basic three-dimensional geometry formed by a helical winding of a wire on a cylinder as illustrated in Figure 4.38. The sense of winding could be right-handed or left-handed. Also, the base cylinder on which the helical geometry is conceived could be free-space.

Thanks to the extensive research due to Kraus [4.2], a wealth of knowledge on this class of antenna has been gathered. There are basically two versions of helical antennas. The first one refers to an *axial-mode* type, which yields a beam-mode of end-fire pattern as shown in Figure 4.38(b), the other version is known as an *omnidirectional mode* (or *perpendicular mode*), whose pattern is depicted in Figure 4.38(c).

In reference to the helical geometry, the associated dimensions in Figure 4.38(a) are explicitly as follows:

- D Diameter of the helix; (and, $C = \pi D$: Circumference of the helix)
- s Spacing between turns (centre-to-centre)
- L Length of a single turn
- A Axial length of the helix ($= N \times s$, with N = Number of turns)
- d Diameter of the helix conductor

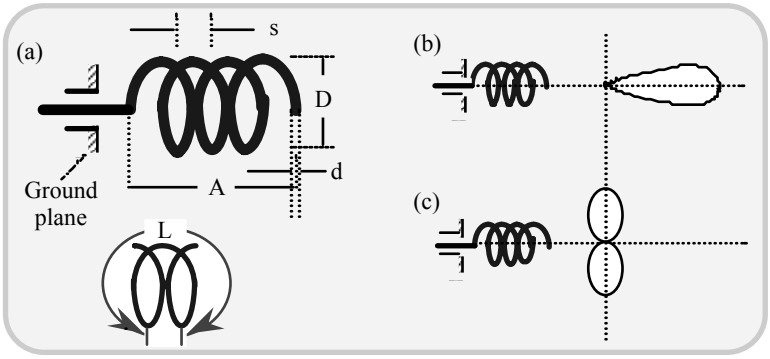


Fig. 4.38 Helical antenna: (a) Basic geometry; (b) axial-mode type yielding an end-fire pattern and (c) omnidirectional mode yielding a broad-side pattern

In addition, the other dimensions of importance are: the distance of the helix from the ground plane and the diameter of the (finite-sized) ground-plane. The pitch angle (α) of the helix is equal to [arctan(s/πD)].

An infinitely long helix can support an EM wave propagating along its length constituting a *transmission mode* (as in a transmission line). A variety of such transmission modes are possible.

A finite helical structure can yield a *radiation mode* with a specific far-field pattern. As indicated earlier such a radiation mode could be either of axial/end-fire type (R_1 *mode*) or normal/broad-side type (R_0 *mode*). In the axial mode of radiation, the field will be maximum in the direction of the helix axis and it will be circularly (or near-circularly) polarised. This is the most important mode of radiation, and it occurs whenever the helix circumference is of the order of a wavelength. Further, it persists over a fairly wide frequency range.

In the normal (perpendicular) mode of radiation, the field maximum occurs in a direction normal to the helix axis, and, under certain conditions, the radiation field will be circularly polarised. This normal mode occurs if the dimensions of the helix are small compared to a wavelength, and as such, it bears neither a wide-band nor a high efficiency capability.

4.6.1 Transmission modes of helices

The term "transmission mode" is used to describe how an electromagnetic wave propagates along an infinitely long helix. In the lowest transmission mode, a helix has adjacent regions of positive and negative charge separated by many turns as shown in Figure 4.39(a). This mode is designated as the T_o*-mode*, and it occurs when the length L of one turn of the helix is small as compared to the wavelength λ (that is, L << λ).

The T_o-mode is essentially the one that occurs in a low-frequency inductance. It is also the important mode on the helix used in the travelling-wave devices such as the travelling wave tube. (In a traveling-wave tube, this transmission mode field is deliberately excited to interact with an electron. Relevant design enables an appreciable axial distance between the adjacent regions of positive and negative charge for optimal wave-beam interactions.)

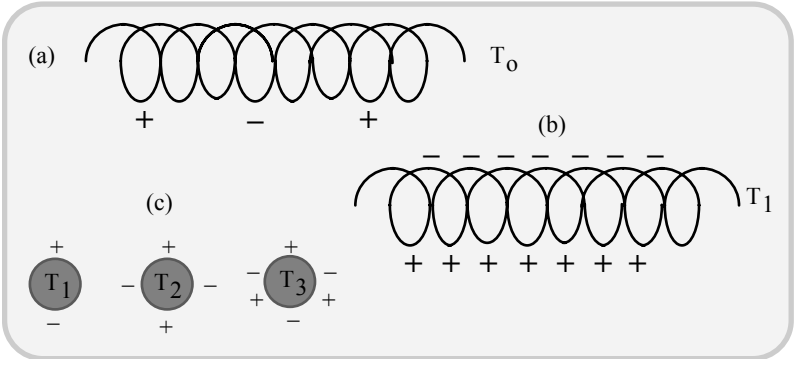


Fig. 4.39 Approximate instantaneous charge distributions on helices for different transmission modes: (a) T_0; (b) T_1 and (c) T_1, T_2 and T_3

When the helix circumference, expressed in free-space wavelength, is of the order of one wavelength (that is, as $C/\lambda \rightarrow 1$), the transmission mode induced on the helix corresponds to a first order known as the *T_1-mode*. For small pitch angles, it has regions of adjacent positive and negative charge separated approximately by half a turn as shown in Figure 4.39(b).

Higher-order transmission modes are designated T_2, T_3, etc. and for small pitch angles, the approximate charge distribution around the helix for such modes are shown in Figure. 4.39(c).

4.6.2 Radiation modes of helices

For a short helix ($NL << \lambda$), the current may be assumed to be of uniform magnitude and in phase along the helix, and it may further be assumed that a T_0-mode is propagating. Such a helix radiates with the maximum field in a direction normal to the helix axis as shown in Figure 4.40(a). The spatial pattern is a figure of revolution, and this is called a "normal radiation mode" (designated as R_0-mode).

With C/λ is of the order of a wavelength and as a T_1-transmission mode propagates along the helix, the corresponding radiation from the helix is referred to as the *R_1-mode*. It has the field maximum in the axial direction as shown in Figure 4.40(b). A helix radiating in this axial mode is usually called a *helical beam antenna*. Figure. 4.40(c) shows a corresponding multilobed pattern for a helix of large dimensions.

Considering the geometry of a helix, if the dimensions of the helix are small (NL << λ), then the maximum radiation is always in the normal direction. When the pitch angle is zero, the helix becomes a loop, and when the pitch angle is 90°, it straightens out into a linear antenna.

The helix may be regarded as a number of small loops and short dipoles connected in series as illustrated in Figure 4.41(a). The diameter D of each loop is same as that of the helix and the length of the dipoles corresponds to the spacing between turns. This assumption is fairly sufficient to characterise the radiation properties of a small helix. With the current assumed to be uniform (in magnitude and in-phase) over the entire length of this small helix, the far-field can be regarded as independent of the number of turns. As such, it is enough to derive the far-field pattern of a single loop and one short dipole connected in series as shown in Figure 4.41(b).

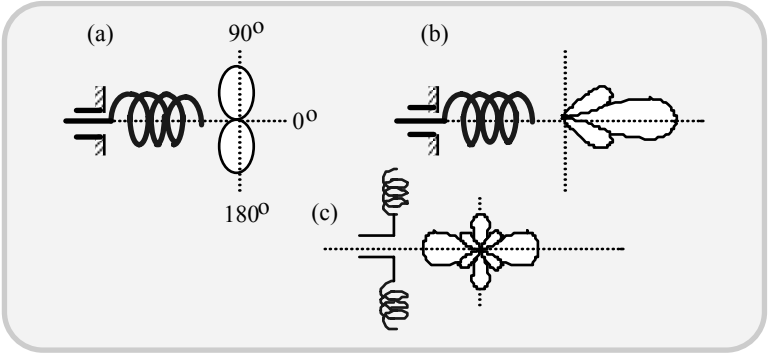


Fig. 4.40 Radiation patterns of helices: (a) Normal mode; (b) axial mode and (c) a multilobed pattern

The far-field of the small loop of Figure 4.41(b) has an E_ϕ-component given by,

$$E_\phi = \frac{120\pi^2 [I]\sin(\theta)}{r} \frac{A_L}{\lambda^2} \tag{4.64}$$

where A_L is the area of the loop equal to $\pi D^2/4$. The far-field of the short dipole of length s has an E_θ-component given by,

$$E_\theta = j\frac{60\pi[I]\sin(\theta)}{r} \frac{s}{\lambda} \tag{4.65}$$

It can be seen from equations (4.64) and (4.65) that E_θ and E_ϕ components are 90° out of phase. The ratio of the magnitudes of E_θ and E_ϕ is given by,

$$AR = \frac{|E_\theta|}{|E_\phi|} = \frac{s\lambda}{2\pi A_L} = \frac{2s\lambda}{\pi^2 D^2} \tag{4.66}$$

and AR is known as the *axial ratio* of the polarisation ellipse. The polarisation ellipse degenerates to a vertical line with $E_\phi = 0$, depicting a vertical polarisation and the helix becomes a vertical dipole. When $E_\theta = 0$, it becomes a horizontal line indicating a horizontal polarisation, and the helix becomes a horizontal loop. When $|E_\theta| = |E_\phi|$, AR is unity and the polarisation ellipse is a circle, indicating circular polarisation. Therefore, the resulting radiation is circularly polarised with the condition that $\pi D = \sqrt{2s\lambda}$. In most situations, the radiation will be elliptically polarised.

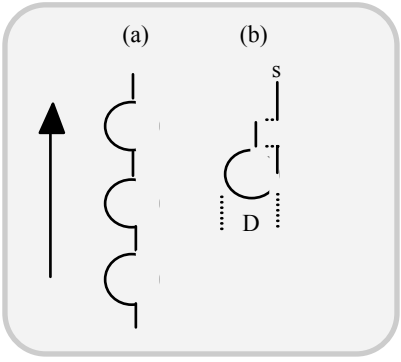


Fig. 4.41 Loop-plus dipole geometrical representation of a small helix under normal mode excitation: (a) Multiple turns structure and (b) a single turn unit

To derive the radiation pattern of a helix radiating in the axial mode, it can be assumed that, approximately a single travelling wave of constant amplitude is travelling along the length of the helix. Then, using the principle of pattern multiplication, the radiation pattern of such a helix can be obtained by the product of the pattern of a single turn and the pattern of an array of N isotropic point sources. Relevant consideration is indicated in Figure 4.42, where N is the number of turns. As mentioned before, the spacing s between the sources refers to the spacing between the turns of the long helix under consideration. The array pattern of this structure can be expected to be much sharper than the pattern of a single-turn radiator. Therefore, the array pattern essentially decides the shape of the total far-field pattern.

The array factor AF corresponding to an array of n isotropic point sources is given by,

$$AF = \frac{\sin(n\psi/2)}{\sin(\psi/2)} \tag{4.67}$$

where $\psi = s_r \cos(\phi) + \delta$, with $s_r = 2\pi(s/\lambda) \equiv 2\pi\, s_\lambda$ and $\delta = -2\pi(L_\lambda/p)$ with $L_\lambda = L/\lambda$. Further, p equal to v/c denotes the relative phase velocity of wave propagation along the helix with v being the phase velocity along the helix and c is the velocity of EM wave in free-space.

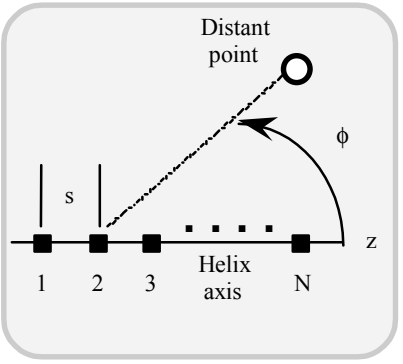


Fig. 4.42 Array of a collinear set of N isotropic point sources adopted to represent the helix

The radiation characteristic of the array depicted in Figure 4.42 will be in the axial mode if the fields from all the sources are in-phase at a point on the helix axis ($\phi = 0°$), or under ordinary end-fire condition, which requires that $\psi = -2\pi m$ where m = 0, 1, 2, 3, … The minus sign in the expression for ψ is due to the fact that the phase of source 2 is retarded by $2\pi L_\lambda/p$ with respect to source 1. Similarly, source 3 is retarded by the same amount with respect to source 2 and so on. Substituting $\phi = 0°$ in equation (4.67) and equating it to $-2\pi m$, it follows that,

$$\frac{L_\lambda}{p} = s_\lambda + m \tag{4.68}$$

Now, substituting p = 1 and m = 1, the resulting $(L - s) = \lambda$ is an approximate relation between L and s for a helix radiating in the axial mode. Further, since $L^2 = \pi^2 D + s^2$, $D = (2s_\lambda+1)/\pi$, $C_\lambda = \sqrt{2s_\lambda + 1}$. This relation $C_\lambda = \sqrt{2s_\lambda + 1}$ defines the approximate upper limit of the axial or beam-mode region. It describes a helix operating in the first order T_1-transmission mode. When m = 2, $L_\lambda = (s_\lambda + 2)$ and $C_\lambda = 2\sqrt{s_\lambda + 1}$, it corresponds to the T_2-transmission mode. Hence, m denotes the order of the transmission mode on the helix radiating a maximum field in the axial

direction. The important case, however, refers mostly to the dominant case, specified by m = 1.

Further, when m = 0, it does not represent a realisable condition, unless p > 1, since when m = 0 and p = 1 would lead to L = s, and the helix will become a straight line and the radiated field in the axial direction of a straight wire is zero. Therefore, there can be no axial field with a helix when m = 0.

Using equation (4.68), $p = L_\lambda/(s_\lambda+1)$ together with the *Hansen and Woodyards' condition of increased directivity* [4.38], the following can be specified:

$$\psi = -\left(2\pi m + \frac{\pi}{N}\right), \tag{4.69}$$

which in turn leads to the relation:

$$p = \frac{L_\lambda}{s_\lambda + m + (1/2N)} \tag{4.70}$$

4.6.3 Axial ratio and conditions for circular polarisation for helix

The axial ratio of the radiation in the direction of the axis of the helix can be determined using the helix geometry shown in Figure 4.43. Relevant condition for circular polarisation can be obtained by referring to electric field components E_θ and E_ϕ at a distant point on the axis (assuming a single uniform traveling wave is supported on the helix).

For harmonic time excitation, the axial ratio $|E_\theta/E_\phi| = AR$ is obtained as:

$$AR = \frac{1}{L_\lambda\left[\sin(\alpha) - 1/p\right]} \equiv \frac{1}{K} \tag{4.71}$$

When K = – 1, this condition refers to circular polarisation leading to the phase velocity $p = L_\lambda/(s_\lambda + 1)$, depicting an end-fire radiation. That is, the natural condition on a helix radiating in the axial mode corresponds to an increased directivity with AR = (2N + 1)/2N. If N is a large, AR approaches unity and the polarization is circular.

Thus, the end-fire condition as well as the increased directivity condition would render the field to be on the helical axis. This field would also be almost circularly-polarised. This circularly-polarised field of the helix is advantageous while using the helical antenna in the receiving mode of the signals of arbitrary polarisation.

4.6.4 Feed arrangements and physical forms of helical antennas

Shown in Figure 4.44 are different methods of feeding axial-mode helical antennas. It can be seen that the ground-plane can be plane or beveled. Also, the

helix can be air-cored or ferrite-cored. The dual helix structures can support right- or left-circular polarisations. Single- or multi-filar windings can be used in making the helices.

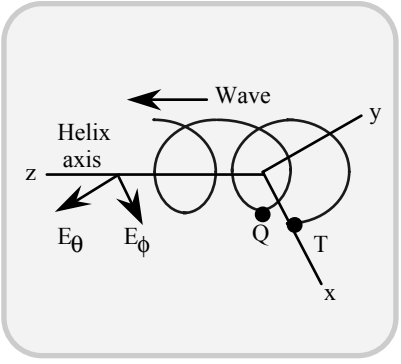


Fig. 4.43 The helix geometry and the field components expressed in the appropriate co-ordinate system

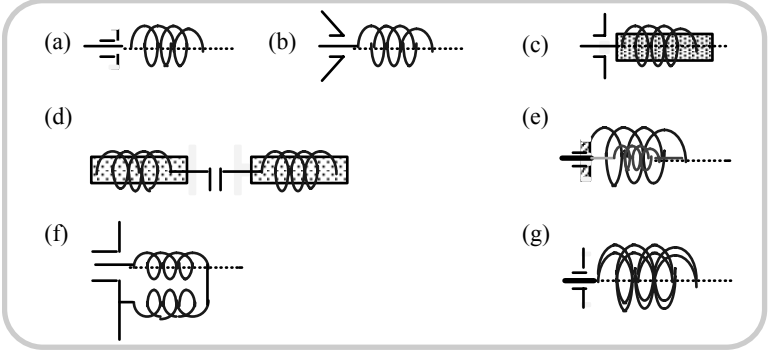


Fig. 4.44 Axial-mode helices showing various constructional and feed arrangements: (a) Simple helix with flat ground-plane; (b) simple helix with flared ground-plane; (c) ferrite-loaded helix; (d) dual helix structure for right- and left-handed polarisations; (e) coaxially placed two helices; (f) ground-returned helix and (g) multi-filar helix

The tapered helical antennas illustrated in Figure 4.45 exhibit more broadband characteristics than their untapered counterparts. Again, the basic tapered geometry can be modified in various forms, as illustrated.

Depicted in Figure 4.46 are slotted metallic cylinder and conical structures, both representing other feasible helical antenna configurations made with slots cut on their surfaces.

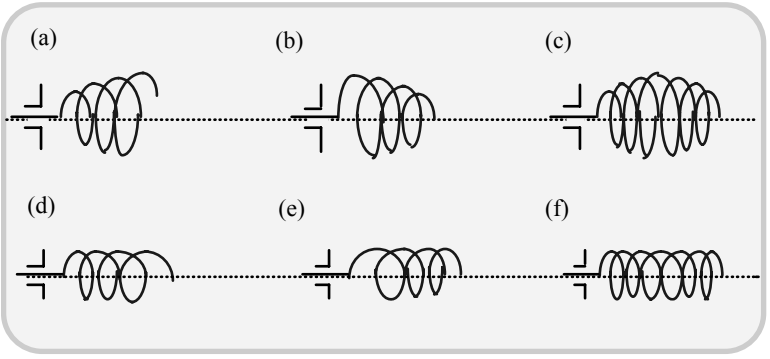


Fig. 4.45 Tapered axial-mode helical antennas: Tapering refers to helix diameter or pitch (a) Diverging helix; (b) converging helix; (c) helix with duocone geometry; (d) helix with pitch increasing outward; (e) helix with pitch contracting outward and (f) helix axially expanded and contracted

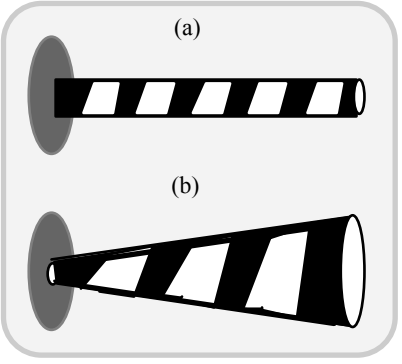


Fig. 4.46 Slotted helical antennas: (a) On a metallic cylinder, and (b) on a metallic cone

Arrangements that can facilitate linearly-polarised radiation with axial-mode helical antennas are illustrated in Figure 4.47. Properly excited right-handed and left-handed polarised waves combine to yield a linearly polarised radiation.These wire-wound helical antennas can also be made of monofilar, bifilar, quadrifilar or multifilar windings.

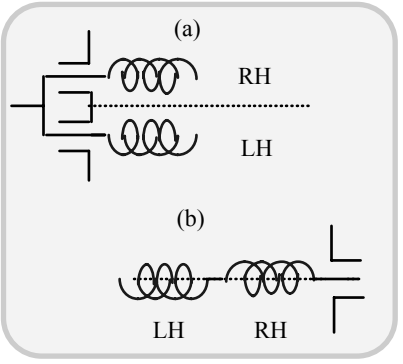


Fig. 4.47 Axial mode helical antennas for producing linear polarization by exciting LH and RH circular polarisations

4.6.5 Circumference-spacing chart of helical antennas

The circumference-spacing chart is a graphical representation of the circumference of the helix (in wavelengths) *versus* the spacing of the turns (in wavelengths) presented in the rectangular coordinate system. This chart is also superimposed with the graphs of corresponding turn-length (in wavelengths) and pitch angle variations, both expressed in polar coordinates. These plots are drawn for different orders of space harmonics (m) of the EM wave supported on the helix. A typical chart is presented in Figure 4.48.

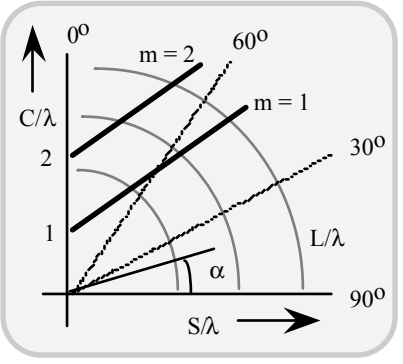


Fig. **4.48** Circumference-spacing chart of helical antennas: (c/λ) *versus* (s/λ) for a set of (α, L/λ) and m values

4.6.6 Helical antennas in wireless applications

In VHF portable transreceivers, the antennas used may correspond to a helix mounted as a part of the radio casing. For example, combining a helix element and a radio case (as shown in Figure 4.49) can constitute a helix-radio dipole. In such designs, the length of the helical element is decided by the frequency of operation. Typically, the helix is made with a length of 35 cm at 40 MHz, 15 cm at 150 MHz and 5 cm at 460 MHz. At frequencies above 800 MHz, the helix-radio pair is generally made longer than $\lambda/2$. As a result, there will not be any peak along the azimuth ($\theta = 0$), which is rather contrary to the desired performance.

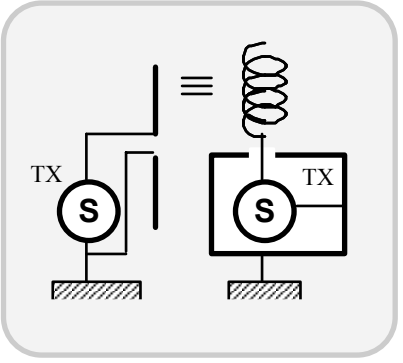


Fig. 4.49 Helix plus radio case constituting an effective dipole structure

The average gain characteristics of a helix-radio dipole (used at level of the head) are presented in Figure 4.50.

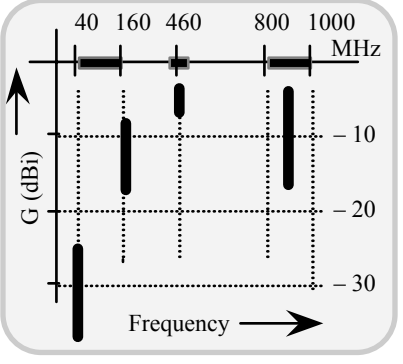


Fig. 4.50 Average gain performance of helix-radio dipole

4.6.7 Land mobile/satellite mobile compatible helix antenna

This refers to a composite structure comprised of a sleeve antenna and a two-wire helical radiator, each yielding a distinct radiation pattern. These antennas are compatible for mobile units for satellite and cellular communication systems [4.39].

4.6.8 Bifilar/quadrifilar helical antennas

Found compatible for satellite mobile telephones, this antenna offers a conical beam radiation pattern in UHF bands. This structure is usually supported on a balanced feed. A simple form of this radiator is illustrated in Figure 4.51.

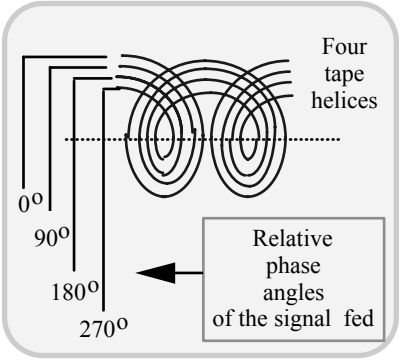


Fig. 4.51 Quadrifilar helix made of a tape material

A multifilar helix antenna can be made with four tape helices that are equally spaced and placed along the circumference of a cylinder. For example, in Figure 4.51, the quadrifilar structure has four elements that are fed with equal amplitude signals with relative phases of 0°, 90°, 180° and 270°. Typically, a fractional-turn resonant quadrifiler helix can produce a *cardiod-shaped* radiation pattern with circular polarisation over a wide range of frequency. If the number of turns is increased, the pattern will become conical with its maximum being at the edges.

It is desirable in many satellite and ground-station applications to concentrate energy into a cone with 120° to 180° cone-angle. In such applications, optimally-shaped conical patterns are often synthesised for maximum gain at the rim of the cone and less gain (by 10 dB) at the cone centre. This pattern shaping or control is done by appropriate choice of the number of turns (to about 1 to 5) of the helix.

The quadrifilar helix is commonly adopted in the designs pertinent to GPS applications. Relevant structure is made usually small in size (fractional turn volutes) and it has circular polarisation properties. The frequencies for which this kind of antenna is designed correspond to the microwave bands at 1.57542 GHz (L_1) and 1.2276 GHz (L_2).

In practice, the GPS operation warrants coverage (of approximate full hemisphere) visible to satellites. This coverage should almost be up to the horizon (with $\theta \approx 10°$) below which the satellite cannot be tracked. Over the visible coverage enabled, a fairly uniform response (in amplitude and phase) can be expected.

Studies [4.40] indicate that such antennas (of resonant fractional volutes) offer high accuracy due to their small size. The accuracy is also dependent on how the phase centre fluctuates with respect to observation angle.

4.6.9 Normal mode helical antennas for portable phones

A normal mode helical antenna (shorter than a $\lambda/4$-monopole) can be used as a mountable radiator on portable units. It is self-resonant, yields a narrow bandwidth and is of flexible structure. It is used in portable telephones in UHF/VHF bands [4.41].

4.6.10 Helix antennas used in maritime and aeronautical systems

It maritime systems, a simple helix is adopted along with a reflector so as to offer an axial-mode, wide-band radiation pattern with circular polarisation and gain in the range 8 to 15 dBi.

In aeronautical systems, the monofilar and/or multifilar cylindrical helix giving normal, axial or conical mode of patterns with circular polarisation at L-band frequency is used for satellite communication purposes.

Conical helix structure facilitating a half-power beamwidth of about 100° in the pattern with circular polarisation is used in aeronautical satellite communications (L-band) with 4-7 dBi gain.

Quadrifilar helix antenna, made of two bifilar helices, has been designed and used in aeronautical satellite communications (at L band). It gives a high front-to-back ratio without a ground-plane. It is good for narrow band operations (3% to 5 % for VSWR < 2).

4.7 SPIRAL ANTENNAS

These antennas have been evolved as a class of what is known as the *frequency-independent* or *wideband antennas*. They are essentially based on *Rumsey's concept* [4.42] of realising EM radiators classified as *self-complementary antennas*. These antennas are characterised by constant terminal impedance of $Z_o/2$ at all frequencies and self-complementary shapes specify relevant antenna configurations. A self-complementary planar antenna, for example, has a metal area congruent to the open area. That is, the metal and open areas are congruent in the sense that a rotation of either brings both into overlap.

Relevant to such self-complementary geometries, Rumsey [4.42] has indicated that the impedance and pattern properties of an antenna will be frequency-independent if the antenna shape is specified only in terms of angles. Typically, an *infinite logarithmic spiral* meets this requirement.

Rumsey's concept is based on the fact that the properties of an antenna (such as the radiation pattern, the gain and the impedance) should depend on the shape, size and on dimensions of the antenna relative to the wavelength.

If, by any arbitrary scaling, the antenna is transformed into a structure, which is equal to the original one, its properties will be independent of the frequency of operation. The antenna is then said to satisfy the *angle condition*, which makes its form to be specified only by angles and not by any dimensions. Infinitely-long *conical antennas* of any arbitrary cross-section with a common apex, and *equiangular antennas* having their surfaces generated by (equiangular) spirals of same defining parameter with a common axis, belong to this class of radiators.

Another relevant concept is due to Duhamel and Isbell [4.43] who indicated that a structure that becomes equal to itself as a result of certain scaling of dimensions with some ratio τ, will have the same properties at a frequency f as well as at the frequency $\tau \times f$. The characteristics (like impedance properties) of such antenna structures will then be a periodic function with the period equal to $\log(\tau)$ of the logarithm of frequency. Antennas based on this concept are known as *log-periodic* antennas The associated variations in the antenna properties over the frequency band (f, $\tau \times f$) are very small in such radiators.

To satisfy the Duhamel and Isbell conditions [4.43], equiangular as well as log-periodic antennas should extend from the origin at 0 (which corresponds to the centre of expansion as well as the feed point), up to infinity. However, in practice, the antenna forms only a section of the ideal (infinite) structure and is contained between two spheres of radii r_1 and r_2 and with their centres at 0. The feed region is inside the smaller sphere of radius r_1. The length r_1 determines the highest frequency f_1 of operation, and r_1 should be much smaller than the wavelength λ_1 (corresponding to the frequency f_1). Thus, the feed-coupling mechanism will have little influence on the impedance or the resulting current distribution over the antenna. The dimension r_1 is also determined by the size of the waveguide or transmission line connected to the antenna.

The radius r_2 of the larger sphere determines the outer dimension of the antenna, and it also decides the lowest frequency of operation. In practice, the current in such structures may decrease rapidly with the distance from the centre 0. Therefore, it is possible to terminate the geometry of the structure at a distance where the current becomes very small as compared to the feed current I_0. This happens usually at a distance proportional to the wavelength of operation.

For conical antennas excited at the apex 0, the field is transverse to the radial direction; and, it decreases as a function of (1/r). Correspondingly, the current also decreases as (1/r).

But in equiangular and log-periodic antennas, the decrease in current is faster than 1/r. Therefore, the end-effects for equiangular and log-periodic antennas will become negligible at a distance of about one wavelength. In comparison with conical antennas such negligible end-effects are not perceived.

4.7.1 Equiangular antennas

When an antenna satisfies the angle condition and if it is expanded in an arbitrary ratio τ about the feed point 0, the resulting new structure has the following properties. This structure will coincide either with the original structure, or with a geometry that can be deduced from the original structure by a rotation about some axis passing through the point 0.

Conical structures having a common apex satisfy the first condition, namely, the expanded structure coincides with the original structure. For example, a *discone antenna* shown in Figure 4.52 exhibits good wideband properties. However, any finite length antenna of this type will pose end-effects. It means that the bandwidth cannot be increased indefinitely, just by increasing the size.

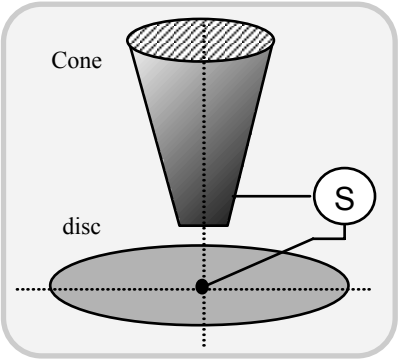


Fig. 4.52 A discone antenna: It is a combination of a disc at the base and a vertical cone

The equiangular structures characterised by the following conditions are normally much more frequency-independent:

- The axis of rotation is independent of τ.
- The angle of rotation is proportional to $\log(\tau)$, or $\tau = \exp(A\phi)$

where ϕ is the angle of rotation and A is a constant. Such equiangular structures also show fast decrease in current with distance.

If the axis of rotation is the z-axis, the equiangular property may be expressed by stating that the surfaces bounding the antenna have a parametric equation of the form,

$$F(\theta, re^{-A\phi}) = 0 \qquad (4.72)$$

where (r, θ, ϕ) depict the spherical polar coordinates. The rotation ϕ about the z-axis together with an expansion about 0 in the ratio $\tau = e^{A\phi}$ will carry the structure onto itself. Figure 4.53 shows the geometry of an equiangular spiral.

The point M in Figure 4.53 traces the locus of an equiangular spiral, either in the plane $z = 0$ or on a cone having the z-axis as the axis of revolution. Suppose the cone has a semiangle θ_o. Then, the equation of the equiangular spiral is given by,

$$r = r_o \exp(A\phi) \tag{4.73}$$

or in terms of ρ as,

$$\rho = \rho_\theta \exp(A\phi) \tag{4.74}$$

The spiral (in Figure 4.53) makes a constant angle ψ with the radius vector r and hence it is designated as equiangular. This angle ψ is given by,

$$\tan(\psi) = \sin(\theta_o / A) \tag{4.75}$$

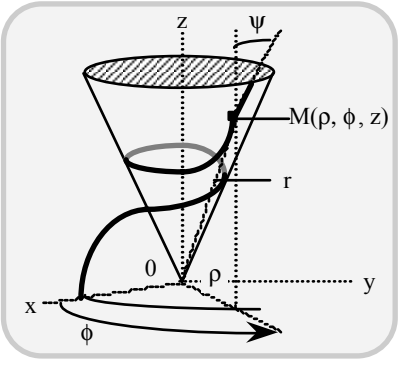


Fig. 4.53 An equiangular spiral set on a cone geometry

Further, the length s of the arc from the origin 0 is proportional to the radial distance **r**, and it is given by:

$$s = r/\cos(\psi) \tag{4.76}$$

The projection of the spiral on the z = 0 plane is also equiangular with the same parameter A. The equiangular spirals with the parameter A′ = – 1/A are orthogonal to those depicted by the parameter A.

If an infinitely long equiangular spiral is fed at the centre 0 by a constant current source of unit magnitude, the electric field at frequency f at point M, is E(M, f). If the frequency is changed to f′ = (f/τ), then all dimensions are scaled by the factor τ, and the field becomes,

$$E(M, \frac{f}{\tau}) = \frac{1}{\tau^2} T^{\phi}(T^{-\phi} M, f) \tag{4.77}$$

where T^{ϕ} corresponds to $\phi = (1/A)\ln(\tau)$. The radiation pattern is given by the field on a large sphere of radius R and is given by,

$$E(\theta, \phi, R) = \frac{e^{-jkR}}{R}[\mathbf{u}_{\theta} P_{\theta}(\theta, \phi) + \mathbf{u}_{\phi} P_{\phi}(\theta, \phi)] \tag{4.78}$$

Both P_{θ} and P_{ϕ} satisfy the relation,

$$P(\theta, \phi, \frac{f}{\tau}) = P[\theta, \phi - \frac{1}{A}\ln(\tau, f)] \tag{4.79}$$

When the frequency is changed from f to f/τ, the pattern will be rotated about the z-axis through an angle, $(1/A) \times \ln(\tau)$.

Practical equiangular spiral structures can be made of conducting wires or sheets (tapes). The wires or sheets should be tapered with increasing distance to satisfy the equiangular condition, but in practice, it has been observed that they can be used over a fairly wide frequency band even if they are not tapered.

The planar structures made with two symmetrical equiangular arms are shown in Figure 4.54 where the expansion ratio is $e^{A\delta}$, corresponding to the rotation ϕ. The angle ψ is given by,

$$\psi = \tan^{-1}(1/A) \tag{4.80}$$

and, it is the angle at which the radial lines cut the spirals.

The arm width of the planar structure is specified by the angle δ or by the ratio $K = e^{-A\delta}$. Further, when $\delta < 90°$, the structure may be considered as a slot of angular width $(180° - \delta)$ rather than as a dipole. The cut-off radii of the finite antenna are r_1 and r_2. The exact shape of the termination is not very important if the antenna is large enough to accommodate the given frequency band.

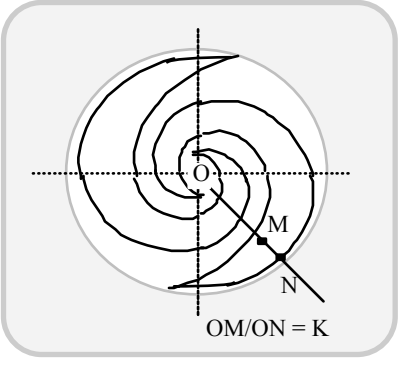


Fig. 4.54 Two-arm planar equiangular spiral antenna

Typical radiation patterns as functions of the cone angle of equiangular spiral are shown in Figure 4.55. The antenna is usually fed by a balanced line or by a coaxial cable embedded into one arm of the spiral.

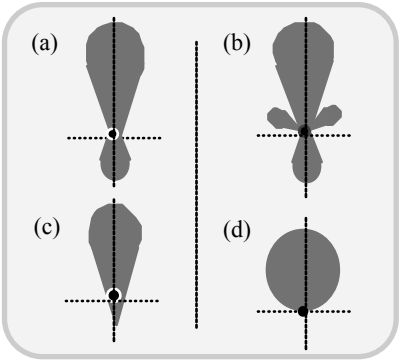


Fig. 4.55 Radiation patterns of equiangular spiral as functions of cone half angle. Approximate cone half-angles: (a) 180°; (b) 30°; (c) 15°; and (d) 10°

Though the impedance of the infinitely-long equiangular spiral is real and independent of frequency, the impedance of a finite length antenna may differ from this ideal characteristic; but the VSWR would remain nearly close to unity over the band of operation.

A typical equiangular spiral set on a cone is illustrated in Figure 4.56. This spiral structure on the cone can be made with tapes and/or single/multifilar windings. Necessary impedance matching can be done with matching circuits as well as by adjusting the dimensions of the cone and the spiral.

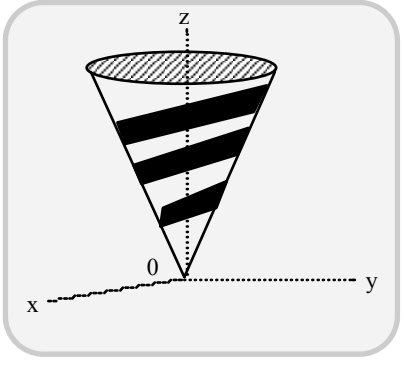


Fig. 4.56 Equiangular spiral antenna on cone

4.7.2 Log-periodic antennas

Log-periodic antennas are designed so that the electrical properties repeat periodically with the logarithm of the frequency. This frequency independence is obtained if the variation of the properties over a period is small. Figure 4.57 illustrates a log-periodic antenna with trapezoidal tooth where θ is a function of $\ln(r_n)$. The analysis of this structure indicates that the relations $\ln(r_n)$ *versus* θ, as well as the input impedance Z_{1N} *versus* ln(f), are periodic.

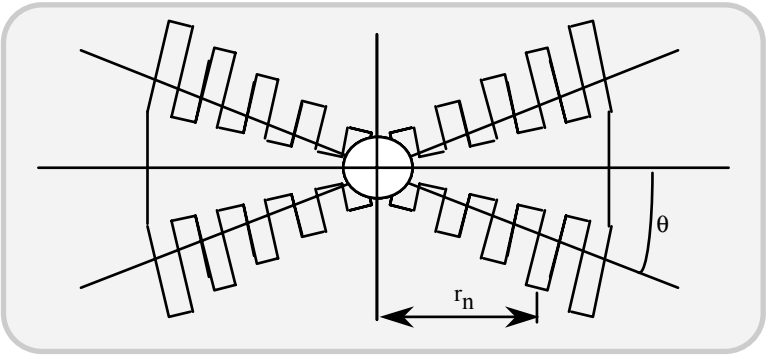


Fig. 4.57 Log periodic antenna: Trapezoidal-tooth log-periodic structure

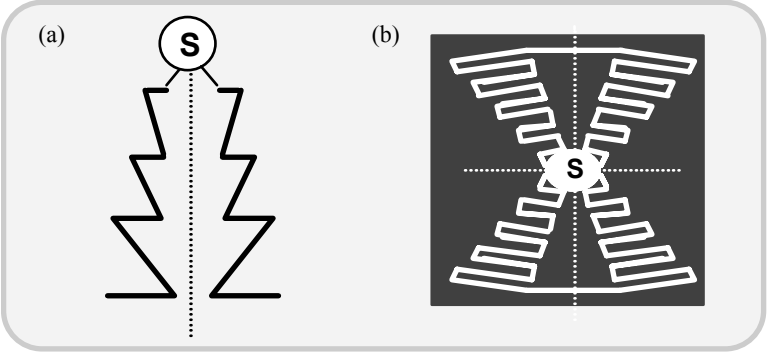


Fig. 4.58 Other types of log-periodic antennas: (a) Sheet circular tooth and (b) wire triangular tooth

Other shapes of teeth like triangular or curved teeth have also been used either in wire or sheet form as shown in Figure 4.58. Theoretically, the thickness of the sheets and diameter of the wires should increase linearly with distance from the vertex of the half structures, though this is not totally necessary for bandwidths

less than 5:1. The greater the bandwidth, the greater should be the tapering. In practice, bandwidth of the ratio in the order of 20:1 can be easily achieved.

Log-periodic dipole arrays as shown in Figure. 4.59 are also useful radiating structures. In such arrays, a uniform two-wire line with the line transposed between the dipoles excites the dipoles.

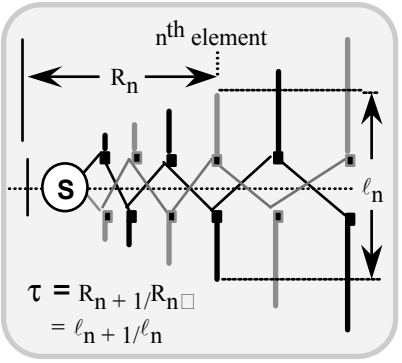


Fig. 4.59 A log-periodic dipole array

4.7.3 Self-complementary antennas

As described earlier, a *self-complementary* structure is a planar structure in which the metal area is congruent to the open area and has a constant input impedance. This follows from Babinet's principle, which specifies that the product of the impedance of two complementary planar structures is equal to $Z_0^2/4$, where Z_0 is the free-space intrinsic impedance of 377 ohms.

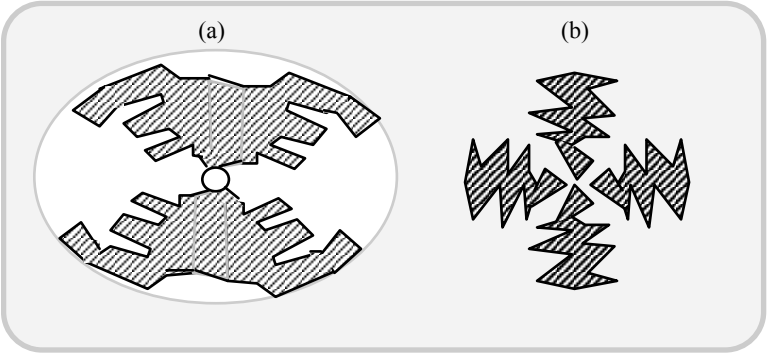


Fig. 4.60 Self-complementary log-periodic antennas: (a) Made of two elements; and (b) made of four elements

Therefore, a self-complementary structure has an impedance of $Z_0/2 = 189$ ohms, which is independent of frequency. This does not necessarily mean that the radiation pattern is also independent of frequency. However, log-periodic structures, which are also self-complementary have a radiation pattern that remains more or less constant with frequency. Two such structures are shown in Figures 4.60(a) and 4.60(b).

4.7.4 Application of log-periodic antennas

The arrays of log-periodic antennas have frequency-independent operation exhibiting a wideband operational characteristic. If the log-periodic elements are inclined with respect to the ground and with their feed points set at ground level, then the resultant radiation pattern will be frequency-independent. In such cases, the elements can be placed such that their images would form E-, H- or a combined E- and H-plane array.

A wire log-periodic with a trapezoidal tooth structure is a useful HF communication antenna. Two log-periodic structures placed at right angles to each other give a radiation pattern that is omnidirectional in the plane transverse to the axis of the antennas.

Two oppositely placed log-periodic half-structures when fed against each other provide a linearly polarised unidirectional beam; two similar structures when fed (symmetrically) against each other (each with a linearly-polarised unidirectional beam orthogonal to the other) would lead to a radiation of circular polarisation over wide bandwidths. Any asymmetry introduced in such configurations will make the resultant radiation elliptically polarised.

4.7.5 Spiral antennas used in wireless communication systems

Typical spiral antennas used in wireless communication systems are illustrated in Figure 4.61. These structures are based on discussions presented earlier. The characteristics of the individual radiator are briefly indicated below.

Conical spiral

This is a bifilar, self-complementary structure that yields circular polarisation without requiring a hybrid. It is used in frequency-independent L to Ku-band satellite communication applications.

Logarithmic planar spiral

This structure again refers to a bifilar radiator formed with a logarithmic period. It is cavity-backed and flush-mounted. This is also used for L to Ku-band satellite communication applications with circular polarisation needing no hybrids.

Archimedean spiral

This is a thin wire bifilar radiator cavity-backed and flush-mounted radiator. This again is used for broadband L to Ku-band satellite communication applications with circular polarisation without hybrid.

Equiangular planar spiral

This bifilar radiator with logarithmic period is cavity-backed and flush-mounted. It gives a constant impedance of 60π ohms. Circular polarisation without using a hybrid can be realised with this structure. This antenna is also compatible for L to Ku-band satellite applications.

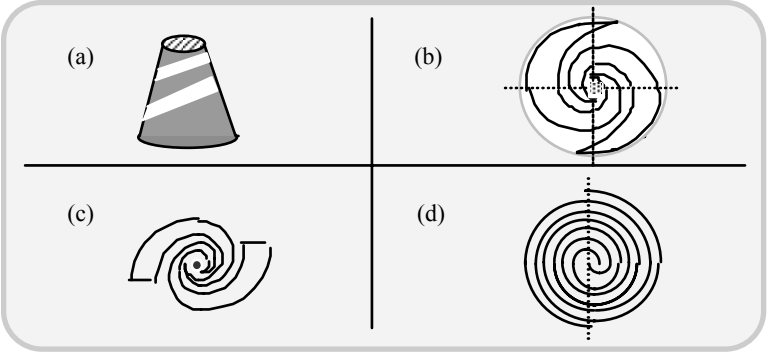


Fig. 4.61 Spiral antennas: (a) Conical spiral, (b) equiangular planar spiral, (c) logarithmic planar spiral and (d) Archimedian spiral

4.8 SLOT AND APERTURE ANTENNAS

4.8.1 Slot antenna

As defined earlier, a slot represents an opening or an aperture on a conducting (ground) plane with dimensions relatively much smaller than the size of the conducting plane as well as the wavelength of EM excitation impressed on the slot. A slot can be excited by means of an energised cavity placed behind it or through a waveguide or by an appropriate transmission-line system.

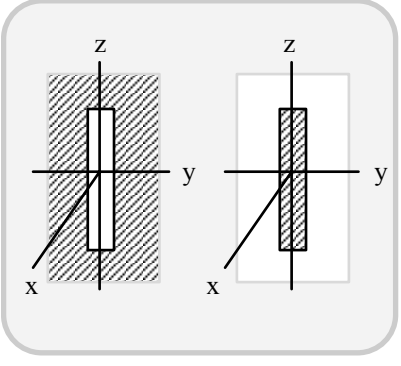


Fig. 4.62 Slot antenna in a conducting screen and the complementary flat dipole

The EM field distribution on a slot is usually described using *Babinet's principle*, which is based on the duality aspects of **E** and **H** fields. Essentially, a slot and a dipole can be regarded as duals. A slot in a conducting plane sheet and its complementary dual, namely a flat dipole antenna, is shown in Figure 4.62. Considering the analysis of these two antennas as a boundary-value problems the appropriate Maxwell's relations and wave equations should satisfy relevant boundary conditions. That is, either one of the wave equation, namely $\nabla^2\mathbf{E} = \mu\varepsilon\mathbf{E}$ or $\nabla^2\mathbf{H} = \mu\varepsilon\mathbf{H}$, has to be solved satisfying proper boundary conditions that are valid for the structure in question.

Considering the dipole $\nabla^2\mathbf{H} = \mu\varepsilon\mathbf{H}$, for example, has to be solved to satisfy the following boundary conditions:

(i) The tangential components of the magnetic field, namely, H_y and H_z are zero in the yz-plane and outside the perimeter of the dipole

(ii) The normal component H_x of the magnetic field is zero in the yz-plane and within the perimeter of the dipole.

In reference to the slot, the dual equation, namely $\nabla^2\mathbf{E} = \mu\varepsilon\mathbf{E}$, should be solved to satisfy the following boundary conditions:

(iii) The tangential components of the electric field, namely E_y and E_z, are zero in the y-z plane and outside the perimeter of the slot

(iv) The normal component E_x of the electricc field is zero in the yz-plane and with in the perimeter of the slot.

It can be seen that mathematically the two aforesaid problems are identical except that **E** and **H** are interchanged. Hence, the two problems follow the principle of duality. Therefore, except for a constant coefficient, the solution obtained for **E** for the slot will be identical to the solution for **H** solved for the dipole. Hence, one can formulate the following relations:

$$\mathbf{E}_s = k_1\mathbf{H}_d \tag{4.81}$$

$$\mathbf{H}_s = k_2\mathbf{E}_d \tag{4.82}$$

where $\mathbf{E}_s$ and $\mathbf{H}_s$ are respectively the electric and magnetic fields of the slot; and $\mathbf{E}_d$ and $\mathbf{H}_d$ respectively denote the electric and magnetic fields of the dipole. From equations (4.81) and (4.82), it can be seen that the electric field distribution of the slot is the same as the magnetic field distribution of the dipole and *vice versa*. Also the impedance of the dipole is proportional to the admittance of the slot and *vice versa.*

The impedance or admittance relations between these two dual structures can be explicitly specified in terms of Z_s and Z_d. Here, $Y_s = 1/Z_s$ denotes the admittance of the slot as a result of the current set up at the edge due to the voltage induced across the gap from the EM excitation. Further, Z_d is the impedance of the dipole depicting the ratio of voltage deriving the dipole and the current induced.

It can be shown that $Z_s Z_d = Z_o^2/4$, where Z_o is the free-space intrinsic impedance (equal to 120π ohms). For a theoretical $\lambda/2$-dipole $Z_d \approx (73 + j43)$ ohms and as such, $Z_s = 418 \angle -30.5°$ ohms refers to the impedance of a theoretical $\lambda/2$-slot.

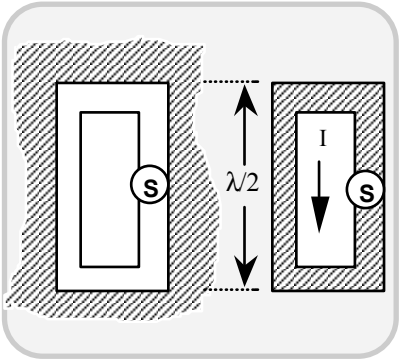


Fig.4.63 Folded slot and folded dipole

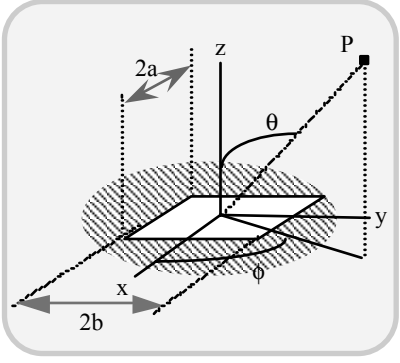


Fig. 4.64 Radiation from a half-wavelength slot in infinite ground-plane

For a resonant-length dipole, the input impedance is a pure resistance. It depends on the thickness of the dipole, and is usually of the order of 65 ohms for a practical half-wave dipole. Hence, the impedance of a practical resonant-length slot is close to 550 ohms. Considering a folded half-wave dipole shown in Figure 4.63, its input impedance is approximately four times that of a half-wave dipole. This is because the two half-wave dipoles of the folded dipole are in parallel and, hence, the total effective current is 2I. Therefore, the power radiated from a half-wave folded dipole is four times that radiated by a simple half-wave dipole. However, the current that is delivered by the generator to the half-wave dipole is only I. The

input resistance of the λ/2 folded structure is therefore four times that of a simple dipole. Thus, the corresponding input impedance of a folded half-wave slot (shown in Figure 4.63) is approximately 550/4 = 137.5 ohms.

Now, consider a half-wavelength slot cut in a perfectly conducting plane as shown in Figure 4.64. Such a slot may be fed either by a coaxial line connected to an RF source as shown in Figure 4. 65(a). Alternatively, the slot can be excited along the side-wall or across at the end of a waveguide as shown in Figure 4.65 (b). Slot excitation is also feasible by means of a cavity as shown in Figure 4.65(c). For any of these feeding arrangements, the tangential electric field in the aperture can be assumed as follows:

$$\mathbf{E}_o = \mathbf{u}_x E_{oo} \cos\left(\frac{\pi y}{2a}\right) \text{ for } |x| < a, |y| \le b \tag{4.83}$$

The tangential electric field in the xy-plane, outside the slot, can be assumed to be zero. The above assumptions are quite valid since they typically specify the distribution of TE_{10} mode at the waveguide aperture that feeds the slot. The assumptions, as above, are also quite reasonable for a cavity-fed slot excited by the fundamental mode of the cavity. The assumed field distribution at the slot will also closely correspond to a coaxial feed excitation.

The Fourier transform of the EM field distribution (as assumed across the slot), will lead to the evaluation of the radiated far-field from the slot. The relevant expressions are as follows:

$$\left.\begin{aligned} E_\phi &= (+jk_o E_o ab) \times \left[\frac{\exp(-jk_o r)}{r}\right] \times \sin(\phi) \times \cos(\theta) \times C_1(\theta,\phi) \times C_2(\theta,\phi) \\ E_\theta &= (-jk_o E_o ab) \times \left[\frac{\exp(-jk_o r)}{r}\right] \times \cos(\phi) \times C_1(\theta,\phi) \times C_2(\theta,\phi) \end{aligned}\right\} \tag{4.84a}$$

where

$$C_1(\theta,\phi) = \left[\frac{\sin[(k_o a) \times \cos(\phi) \times \sin(\theta)]}{(k_o a) \times \cos(\phi) \times \sin(\theta)}\right] \tag{4.84b}$$

and

$$C_2(\theta,\phi) = \left[\frac{\cos[(k_o b) \times \sin(\phi) \times \sin(\theta)]}{[(k_o b) \times \sin(\phi) \times \sin(\theta)]^2 - (\pi/2)^2}\right] \tag{4.84c}$$

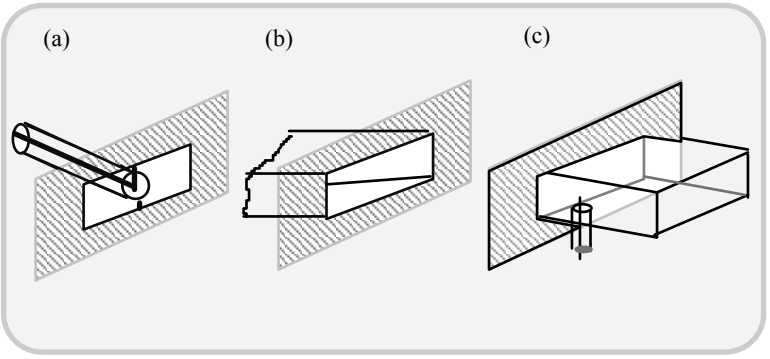


Fig. 4.65 Slots cut in a ground-conducting plane and methods of feeding such slots (a) *Via* a coaxial line; (b) by using an aperture-end of waveguide and (c) by means of a cavity-backed excitation

When the width 2a is small (so that $k_0 a << 1$), and if the length 2b of the slot is a half-wavelength (so that $k_0 b = \pi/2$), the field components E_ϕ and E_θ given by the above expressions simplify to:

$$\left.\begin{aligned} E_\phi &= -j\frac{V_o}{\pi}\left[\frac{\exp(-jk_o r)}{r}\right]\sin(\phi)\cos(\theta)\left[\frac{\cos[(0.5\pi)\sin(\phi)\sin(\theta)]}{[1-\sin^2(\phi)\sin^2(\theta)]}\right] \\ E_\theta &= +j\frac{V_o}{\pi}\left[\frac{\exp(-jk_o r)}{r}\right]\cos(\phi)\left[\frac{\cos[(0.5\pi)\sin(\phi)\sin(\theta)]}{[1-\sin^2(\phi)\sin^2(\theta)]}\right] \end{aligned}\right\} \quad (4.85)$$

where $V_o = 2aE_o$ depicts the voltage across the slot. In the ($\phi = 0°$)-plane, E_ϕ is zero and E_θ is constant, so that the pattern in this plane will be a circle. In the ($\phi = 2\pi$)- plane ($x = 0$), E_θ is zero and,

$$E_\phi = -\left(\frac{jV_o}{\pi}\right)\left[\frac{\exp(-jk_o r)}{r}\right]\left[\frac{\cos[(0.5\pi)\sin(\theta)]}{\cos(\theta)}\right] \quad (4.86)$$

The pattern for E_ϕ in this ($\phi = \pi/2$)-plane will be the same as that of a half wavelength electric dipole oriented along the y-axis.

The aperture of a narrow, half-wavelength slot would produce exactly the same fields as a magnetic dipole with a magnetic current given by:

$$\mathbf{I}_m = -4aE_o \cos(\pi y/2b)\mathbf{u}_y \qquad \text{for } |y| \le b \tag{4.87}$$

radiating in free-space (with no ground-plane) indicating that the slot is just a dual of an electric dipole.

The total power radiated by the slot is obtained by integrating (1/2) × Re(**E** × **H***) over a hemisphere of large radius centred at the aperture and it is given by,

$$P = \left[\frac{Z_o^{-1}}{2}\frac{V_o^2}{\pi^2}\right] \times \int_{\psi=0}^{\pi}\int_{\zeta=-\pi/2}^{+\pi/2} \frac{\cos^2[(\pi/2)\times\cos(\psi)]}{\sin^2(\psi)}\sin(\psi)\,d\psi\,d\zeta \tag{4.88}$$

where $\sin(\theta)\sin(\phi) = \cos(\psi)$ and ζ is the angle made by the projection of r in the xz-plane with the z-axis.

For a half-wavelength thin slot, $k_o a << 1$ and $k_o b = \pi/2$ so that,

$$P = \left[\frac{Z_o^{-1}V_o^2}{4\pi}\right][\ln(2\gamma\pi) - \mathrm{Ci}(2\pi)] = 2.436\left[\frac{Z_o^{-1}V_o^2}{4\pi}\right] \tag{4.89}$$

Since V_o represents the voltage across the slot, a radiation conductance G_r can be defined as the conductance, which when placed across the voltage V_0, would dissipate the same power as that radiated by the slot; that is,

$$\frac{1}{2}V_o^2 G_r \equiv P, \tag{4.90}$$

which leads to:

$$G_r = 2.436\left[\frac{Z_o^{-1}}{2\pi}\right] = 1.03\times10^{-3}\,\text{siemens} \tag{4.91}$$

4.8.2 Slots on conducting cylinders

A slot can be formed by a waveguide-opening through the side of a vehicle (such as rocket body or an aircraft wing). An approximate model of this slot antenna corresponds to a slot on a conducting cylinder. A slot so realised, can be either a circumferential or of axial type as illustrated in Figures 4.66(a) and 4.66(b). If necessary, slots (both circumferential and axial) can be skewed to trim the resulting radiation pattern to match the application in hand.

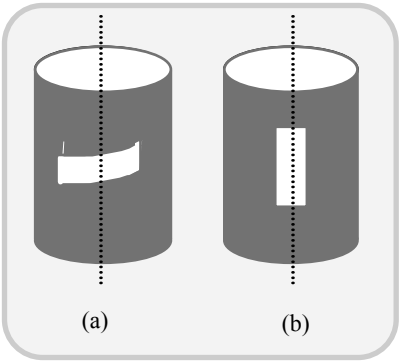


Fig. 4.66 Slots on conducting cylinders: (a) Peripheral slot and (b) longitudinal axial slot

If the electric field across the aperture of the slot is known, the field in the region outside the cylinder can be found in the following manner. First, the general expressions depicting the cylindrical mode functions for an arbitrary field in a cylindrical co-ordinate system are decided. Then, by identically equating these general expressions for the field components to those assumed on the surface of the cylinder (including the slot region), the unknown coefficients are determined [4.44].

The technique indicated above can be applied to any arbitrarily-shaped slot. Results on the radiation patterns pertinent to a narrow $\lambda/2$-slot on a large conducting cylinder are shown in Figures 4.67. These results correspond to a cosinusoidal distribution of E_ϕ-component on the aperture.

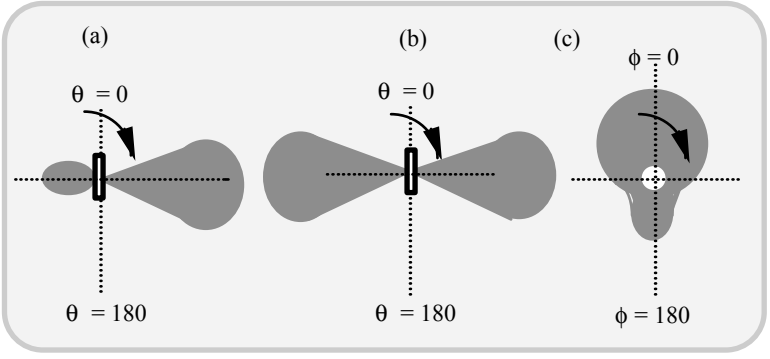


Fig. 4.67 Radiation patterns of an axial half-wave slot cut on a conducting cylinder: (a) ($|E_\phi|$ *versus* θ) in the ($\phi = 0$)-plane; (b) ($|E_\phi|$ *versus* θ) in the ($\theta = \pm \pi/2$)-plane and (c) ($|E_\phi|$ versus ϕ) in the ($\theta = \pi/2$)-plane

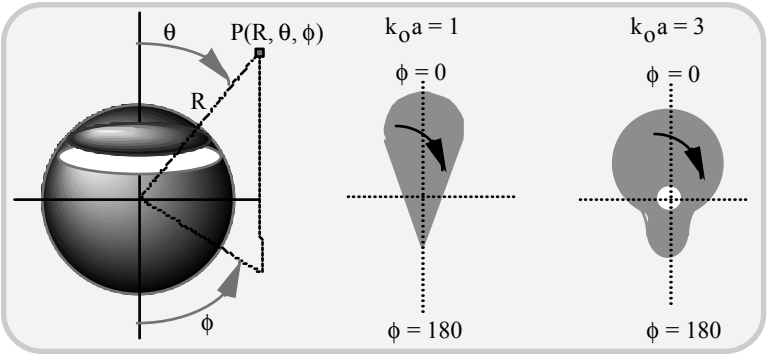


Fig. 4.68 Slot on a conducting sphere and its radiation patterns: ($|E_\theta|$ *versus* ϕ) in the ($\theta = \pi/2$)-plane for a half-wave slot cut on a conducting sphere

Similar to evaluating the radiation field components of a slot on a cylinder, the corresponding EM problem for a slot on a conducting sphere can also be solved. Here, the resulting field components (due to the slot) outside the sphere should be specified as spherical mode functions. Typical results obtained for half-wave slots on a conducting sphere are presented in Figure 4.68. A set of slots on a surface can be used to construct a slot-array antenna. Relevant considerations will be presented in Chapter 5.

4.8.3 Aperture antennas

The underlying concept of an aperture antenna is an extension of the principle of a slot radiator. Unlike in a slot, the aperture can have dimensions comparable, or even larger than the wavelength of the EM field exciting it. Typically, a radiating aperture can be regarded as an area with a specified EM field distribution.

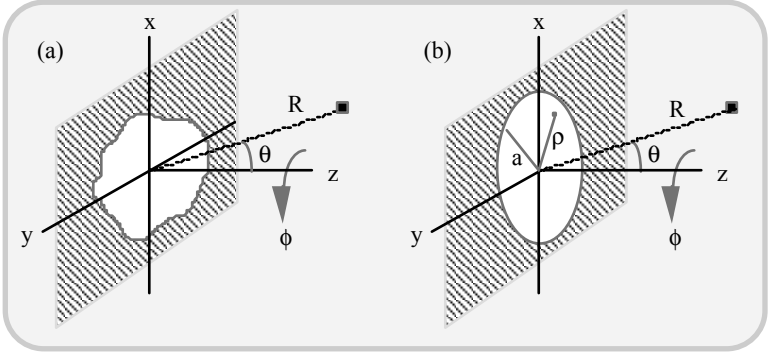


Fig. 4.69 Aperture radiators: (a) An arbitrary aperture and (b) a circular aperture

For example, in reference to the illustration of an arbitrary aperture shown in Figure 4.69(a), if the aperture distribution is given by f(x, y), where x and y are the two associated coordinates, then the directivity pattern of the radiation from the aperture can be written as follows:

$$E(\theta,\phi) = \iint_{x,y} f(x,y)\exp[jk\sin(\theta)[x\cos(\phi)+y\sin(\phi)]]dxdy \quad (4.92)$$

Hence, considering a circular aperture of Figure 4.69(b), the directivity pattern can be explicitly written as:

$$E(u,\phi') = a^2 \int_{\phi'=0}^{2\pi}\int_{r=0}^{1} f(r,\phi')\exp[j(ur)\times\cos(\theta)]\times \cos(\phi-\phi')rdrd\phi' \quad (4.93)$$

where a is the radius of the aperture, ρ is the radius at any point of aperture, $r = \rho/a$, $u = (2\pi a/\lambda)\sin(\theta)=(\pi D/\lambda)\sin(\theta)$ with D = 2a and $f(r, \phi')$ denotes the normalised aperture-field distribution function. Suppose $f(r, \phi') = f(r)$ (depicting explicitly that the aperture distribution is not dependent on ϕ and it is a function of r only), the directivity pattern can be simplified to:

$$E(u) = (2\pi a^2)\int_{r=0}^{1} f(r)J_o(ur)dr \quad (4.94)$$

where $J_o(\cdot)$ is the Bessel function of first kind and zeroth order. For a general distribution of $f(r) = (1 - r^2)^p$, E(u) simplifies to:

$$E(u) = (\pi a^2)\frac{2^{(p+1)}\,p!\,J_{p+1}(u)}{u^{p+1}} \quad (4.95)$$

where $J_{p+1}(\cdot)$ is the Bessel function of first-kind and $(p+1)^{th}$ order. For a constant aperture illumination $f(r) \equiv 1$, therefore $E(u) = (2\pi a)2J_1(u)/u$.

The circular aperture discussed above, for example, can be realised with an open-ended circular waveguide as shown in Figure 4.70(a). The type of field distribution at the open-ended aperture of a circular waveguide, namely $f(r, \phi)$ would depend on the TE_{mn} or TM_{mn} mode of excitation of the EM field in the waveguide.

By flaring the circular waveguide into a conical structure as shown in Figure 4.70(b), the open-end of a circular waveguide can be structurally modified. The result will be a *conical horn antenna*.

Similar to an open-ended circular waveguide radiator and its modified form of conical horn antenna, an open-ended rectangular waveguide radiator (or its

modified versions, such as pyramidal and sectoral horns) can be conceived as illustrated in Figures 4.70(d) – 4.70(f).

The horn structures of Figure 4.70, which are devised by flaring out a waveguide (circular or rectangular) yield patterns that are, in general, asymmetrical in the principal planes. The E- and H-plane horns are deliberately made with the flares confined to vertical and horizontal planes, so that these horns yield beams that fan-out in the E- or H-planes respectively. They are used for those applications in which the beam is confined to one plane.

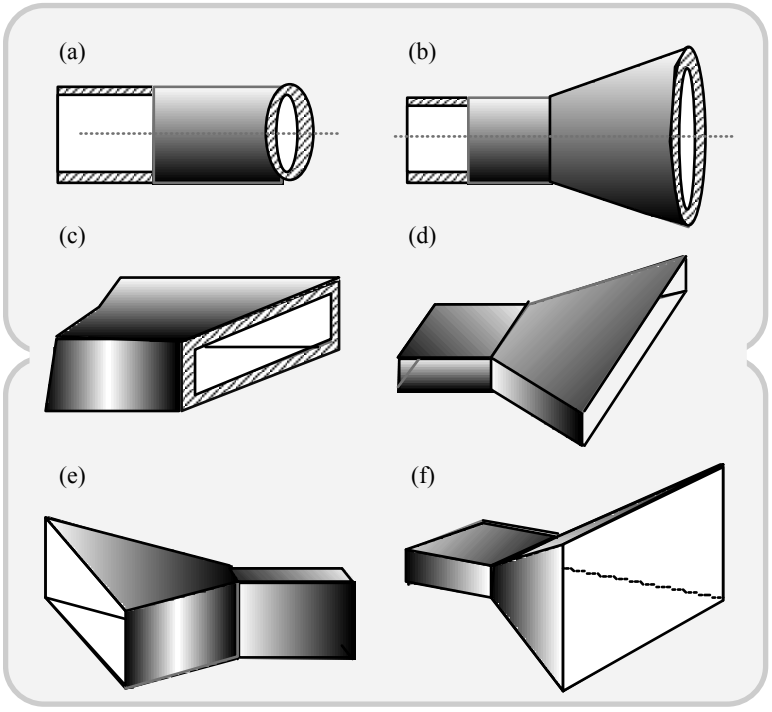


Fig. 4.70 Open-ended waveguide apertures and horn structures. Circular waveguide geometry: (a) Open-ended aperture and (b) conical horn; Rectangular waveguide geometry: (c) Open-ended aperture; (d) H-plane sectoral horn; (e) E-plane sectoral horn and (f) pyramidal horn

Conical and pyramidal horns excited in the dominant modes exhibit patterns asymmetric in the principal planes. In order to achieve symmetrical patterns (in the principal planes), these horns can be modified to support hybrid mode excitation, so that the beam pattern has a near-circular cross-section. Such horns are known as *corrugated or scalar horns* and will be described later.

Horn antennas, as indicated earlier, are mainly used as primary antennas to illuminate secondary reflectors. The E- and H-plane antennas are used as primary radiators in (two-dimensional) cylindrical reflector antenna systems to give fan-out beams.

The pyramidal and conical horns (or their modified corrugated versions) are used to excite parabolic reflectors to yield pencil-beams. They are used in satellite as well as in point-to-point radio links at microwave frequencies. In such applications, they are invariably kept fixed in a particular direction. (In radar applications, horn antennas can form a part of a mechanical (movable) scanning system.)

4.8.4 Rectangular aperture on an infinite ground-plane

Suppose an open-ended rectangular waveguide is set at an infinite ground-plane as illustrated in Figure 4.71. Again, the radiated far-field from this aperture would directly correspond to the EM field distribution at the aperture plane. Considering that the waveguide is excited in its dominant mode, namely TE_{10}, for the geometry presented in Figure 4.71, the far-field components are given by:

$$\left.\begin{aligned} E_r &= H_r = 0 \\ E_\theta &= -\frac{\pi C E_o e^{-jkr}}{2}\sin(\phi)\times[A(\theta,\phi)]\times[B(\theta,\phi)] \\ E_\phi &= -\frac{\pi C E_o}{cr}\cos(\theta)\cos(\phi)\times[A(\theta,\phi)]\times[B(\theta,\phi)] \end{aligned}\right\} \qquad (4.96a)$$

where

$$\left.\begin{aligned} A(\theta,\phi) &= \left[\frac{\cos[(ka/2)\sin(\theta)\cos(\phi)]}{[(ka/2)\sin(\theta)\cos(\phi)]^2 - [\pi/2]^2}\right] \\ B(\theta,\phi) &= \left[\frac{\sin[(kb/2)\sin(\theta)\sin(\phi)]}{(kb/2)\sin(\theta)\sin(\phi)}\right] \end{aligned}\right\} \qquad (4.96b)$$

and

$$C = \left[j\frac{abk}{2\pi r}\right] \qquad (4.96c)$$

Further, a and b are the dimensions of the waveguide in the x- and y-directions respectively and $k = 2\pi/\lambda$. The other magnetic field components are as follows: $H_\theta = -E_\phi/\eta_o$ and $H_\phi = E_\theta/\eta_o$, where η_o is the intrinsic impedance of the free-space.

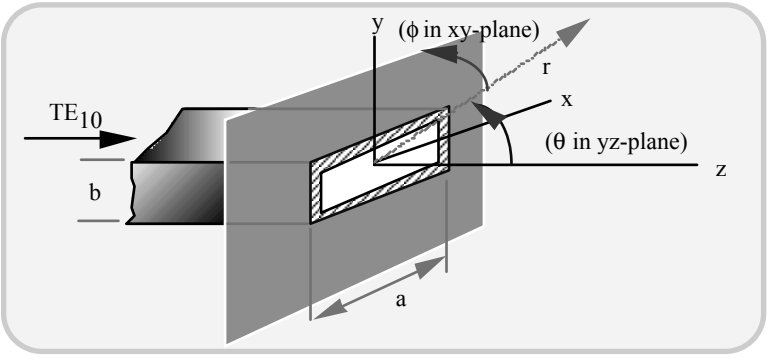


Fig. 4.71 Open-ended waveguide aperture on an infinite ground-plane

Similar to an open-ended aperture of a waveguide, the open end of a coaxial-line can also be realised as an aperture radiator [4.45]. Apart from horn antennas conceived from the open-ended geometry of waveguides, there are also other aperture radiators evolved by modifying the waveguide apertures. Typically, the family of so-called *scalar horns* (or *corrugated horns*) [4.46], and dielectric-loaded waveguide apertures [4.47, 4.48] are examples of such modified versions adopted in practice.

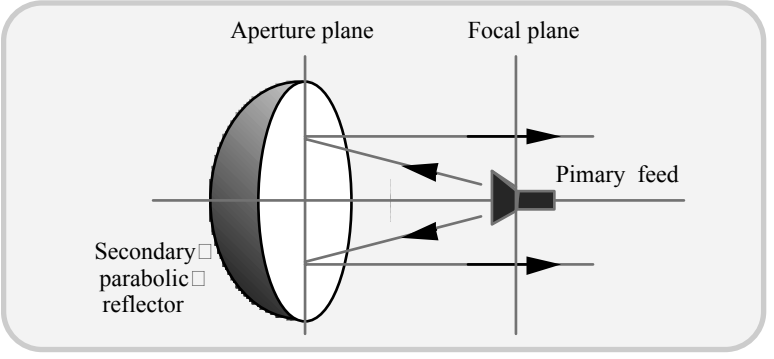


Fig. 4.72 A parabolic secondary reflector aperture excited by a primary feed at the focal plane

The open-ended waveguide structure (or its variations) indicated above are largely used as primary feeds in conjunction with a secondary reflector. For example, a primary feed can be placed and excited at the focal plane of a parabolic reflector. The resulting field on the parabolic reflector would correspond to a large-aperture

illumination with a specified field pattern, which in turn produces a far-field radiation as shown in Figure 4.72.

In the following sections, the radiation characteristics of primary and secondary aperture antennas are briefly addressed.

4.9 HORN ANTENNAS

4.9.1 Conical horn antenna

As indicated before, a conical horn antenna is a flared geometry evolved from an open-ended circular waveguide aperture. The relevant modification of the aperture-end geometry, as envisaged, has two motives: The first one refers to achieving a larger radiating aperture size (relative to wavelength) that would eventually lead to a better directivity. The second reason is that the smooth flaring (from the waveguide to a horn geometry) allows a continuous impedance transition. An abrupt open-ended waveguide aperture on the other hand, would represent a discontinuity. Furthermore it will reflect EM waves back into the transmission line (waveguide) causing excessive VSWR on the line; also, the discontinuity at the open-ended waveguide aperture would induce an edge current at the rim of the aperture, leading to excessive sidelobes and/or backlobes in the radiation pattern [4.49].

Ideally, a horn should be flared to an infinite extent so as to obviate totally the negative considerations characteristic of an open-ended waveguide aperture. However, for practical reasons, the size of the horn is made to a finite extent with a compromised VSWR in the exciting guide as well as giving due regard to the sidelobes and/or backlobes induced in the radiation pattern. Hence, shown in Figure 4.73 is a conical horn antenna (with appropriate co-ordinates specified) to deduce the radiation pattern.

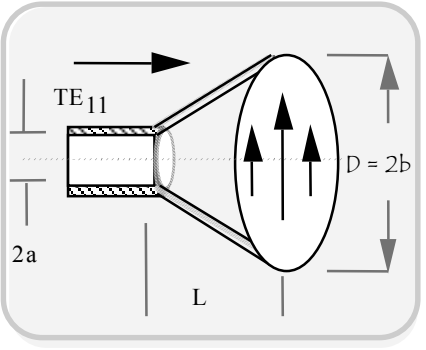


Fig. 4.73 A simple conical horn excited by a circular waveguide supporting the dominant mode

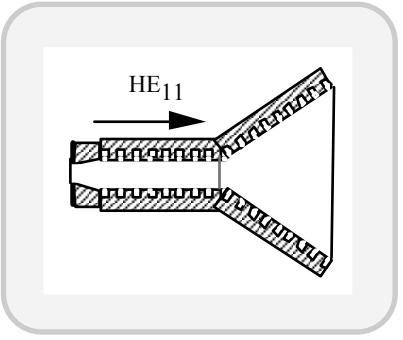


Fig. 4.74 A corrugated conical horn (scalar horn) excited by a hybrid mode supported on a circular corrugated waveguide

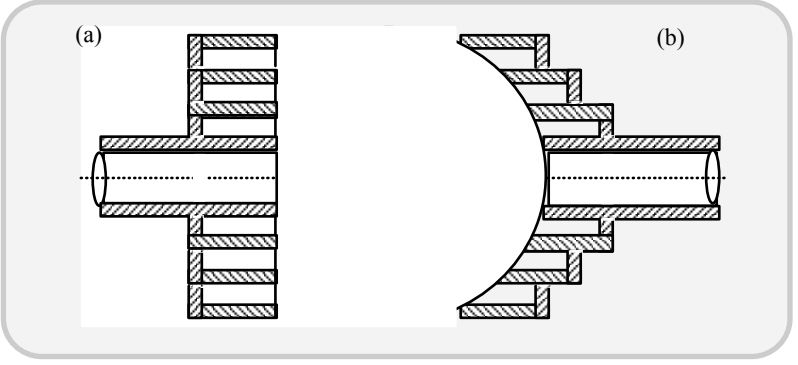


Fig. 4.75 Corrugated corner with 180° flare-angle: (a) Flat flange structure and (b) curved flange structure

When a conical horn is excited through a circular waveguide in the dominant TE_{11} mode, the resulting electric field distribution at the aperture of the cone are illustrated by a set of vertical arrows (Figure. 4.73). The major shortcoming of a conical horn with such an excitation is the resulting asymmetric field pattern in the principal planes (as indicated earlier). Further, the TE_{11} excitation would lead to an induction of edge currents and hence, edge-diffraction effects thereof, causing undesirable sidelobes in the far-field pattern. A modification proposed on the conical structure that allows reduced edge diffraction, improved pattern symmetry and low cross-polarisation refers to the so-called *corrugated conical horn* (also known as *scalar horns*), illustrated in Figure 4.74.

The scalar horn supports a hybrid mode HE_{11} (a balanced TE_{11} plus TM_{11} mode). The corrugation on the horn walls act as $\lambda/4$-chokes leading to reduced E-

field at the rim of the horn for all polarisations. This, in effect reduces the edge current-induced edge diffractions.

The corrugated horn can also be modified to have a flare-angle equal to 180°, as shown in Figure 4.75. Further, the resulting corrugated flange can also be made curved for better radiation characteristics, as studied in [4.26, 4.27].

4.9.2 Rectangular horn antennas

As illustrated earlier, there are three versions of rectangular horn antennas, namely, pyramidal, H- and E-plane sectoral horns.The pyramidal horn has finite flare angles in the E-and H-planes and the extents of corresponding aperture are a_1 and b_1, as illustrated in Figure 4.76.

The horn in Figure 4.76 is excited *via* a rectangular waveguide supporting a TE_{10} y-directed field. Under this condition, the far-fields of the pyramidal horn are given by:

$$\left.\begin{aligned} E_\theta &= \left[\frac{jkE_o e^{-jkr}}{4\pi r}\right] \times \sin(\phi) \times [1 + \cos(\theta)] \times A \\ E_\phi &= \left[\frac{jkE_o e^{-jkr}}{4\pi r}\right] \times \cos(\phi) \times [1 + \cos(\theta)] \times A \end{aligned}\right\} \quad (4.97a)$$

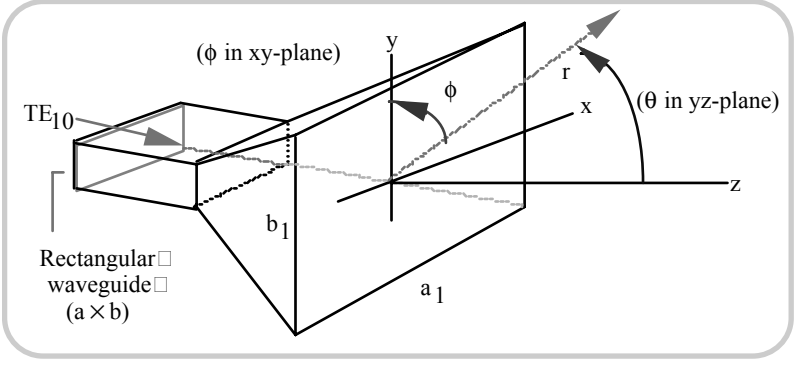


Fig. 4.76 A pyramidal horn antenna and the associated co-ordinate system at the aperture plane

In equation (4.97a) $A = A_1 \times A_2$; and A_1 and A_2 are explicitly specified by the following expressions:

$$A_1 = C_1 \times [e^{jk_x'^2 r_2/2k}\{[C(\tau_4) - C(\tau_3)] - j[S(\tau_4) - S(\tau_3)]\} + e^{jk_x''^2 r_2/2k}\{[C(\tau_6) - C(\tau_5)] - j[S(\tau_6) - S(\tau_5)]\}]$$

$$A_2 = C_2 \times [[e^{jk_y^2 r_1/2k}\{[C(\tau_2) - C(\tau_1)] - j[S(\tau_2) - S(\tau_1)]\}] \quad (4.97b)$$

with

$$C_1 = \left(\frac{1}{2}\sqrt{\frac{\pi\rho_2}{k}}\right)$$

$$C_2 = \left(\frac{1}{2}\sqrt{\frac{\pi\rho_1}{k}}\right)$$

$$\left.\begin{aligned} k &= 2\pi/\lambda \\ k'_x &= k\sin(\theta)\cos(\phi) + (\pi/a_1) \\ k''_x &= k\sin(\theta)\cos(\phi) - (\pi/a_1) \\ k_y &= k\sin(\theta)\sin(\phi) \end{aligned}\right\}$$

and

$$\tau_1 = \sqrt{\frac{1}{\pi k\rho_2}}\left(-\frac{ka_1}{2} - k'_x\rho_2\right); \quad \tau_2 = \sqrt{\frac{1}{\pi k\rho_2}}\left(\frac{ka_1}{2} - k'_x\rho_2\right),;$$

$$\tau_3 = \sqrt{\frac{1}{\pi k\rho_2}}\left(-\frac{ka_1}{2} - k''_x\rho_2\right); \quad \tau_4 = \sqrt{\frac{1}{\pi k\rho_2}}\left(\frac{ka_1}{2} - k''_x\rho_2\right);$$

$$\tau_5 = \sqrt{\frac{1}{\pi k\rho 1}}\left(-\frac{kb_1}{2} - k_y\rho_1\right); \quad \tau_6 = \sqrt{\frac{1}{\pi k\rho_1}}\left(\frac{kb_1}{2} - k_y\rho_1\right).$$

Further, C(·) and S(·) represent the following *cosine* and *sine (Fresnel) integrals* given by:

$$\left.\begin{aligned} C(x) &= \int_0^t \cos\left(\frac{1}{2}\pi t^2\right)dt \\ S(x) &= \int_0^t \sin\left(\frac{1}{2}\pi t^2\right)dt \end{aligned}\right\} \qquad (4.98)$$

Also, the entities ρ_1 and ρ_2 in the above expressions denote the following: Suppose the upper and lower surfaces of the horn are extended into the waveguide, they meet at a horizontal line.

The distant of this line from the aperture plane of the horn (specified at z = 0) is ρ_1. Likewise, if the two side surfaces of the horn are extended to meet in a vertical line in the waveguide, the distance of this line from the aperture plane of the horn is ρ_2.

The directivity of a pyramidal horn is given by:

$$D_p = (\pi\lambda^2/32ab) \times D_E \times D_H \qquad (4.99a)$$

where

$$D_E = (64a\rho_1/\pi\lambda b_1 \times [C^2(b_1/\sqrt{2\lambda\rho_1}) + S^2(b_1/\sqrt{2\lambda\rho_1})] \qquad (4.99b)$$

and

$$D_H = (4\pi b\rho_2/\lambda a_1) \times \{[C(u) \pm C(v)]^2 + [S(u) - S(v)]^2\} \qquad (4.99c)$$

with $u = (u_1 + 1/u_1)/\sqrt{2}$ and $v = (v_1 - 1/v_1)/\sqrt{2}$, where $u_1 = \sqrt{\lambda\rho_2}/a_1$ and $v_1 = \sqrt{\lambda\rho_2}/a_1$.

Further, the entities D_E and D_H refer to the directivities of E- and H-plane horns respectively, which are discussed in the following section.

Approximate E-and H-plane patterns of a typical pyramidal horn with $\rho_1 = \rho_2 = 6\lambda$, $a_1 = 12\lambda$, $b_1 = 6\lambda$, $a = 0.5\lambda$ and $b = 0.25\lambda$, are sketched in Figure 4.77.

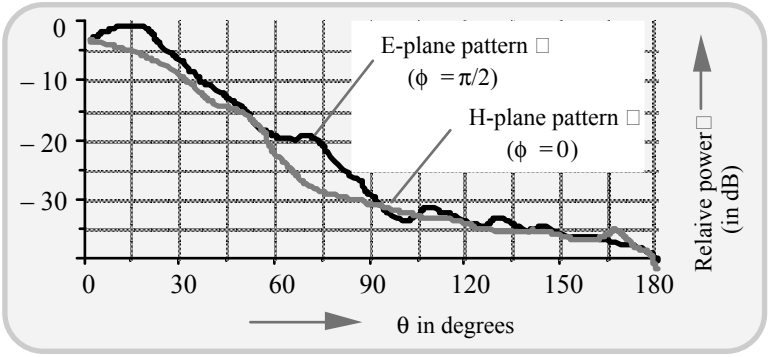


Fig. 4.77 Typical E- and H-plane patterns of a pyramidal horn excited by a rectangular waveguide supporting the dominant mode

4.9.3 Sectoral horns

The E- and H-plane sectoral horns illustrated earlier (in Figure 4.71), refer to sectoral structures conceived by flaring an open-ended rectangular waveguide in the E- and H-planes respectively.

The E-plane sectoral horn antenna

Illustrated in Figure 4.78 is again an E-plane horn with relevant details and considerations of the fields exciting the aperture and those in the radiated far-field.

Following the rigorous analysis due to Balanis [4.3] the electric field components radiated by the horn under discussion are given by a set of relations of equation (4.100). They are derived by considering appropriate aperture field distribution at the horn aperture.

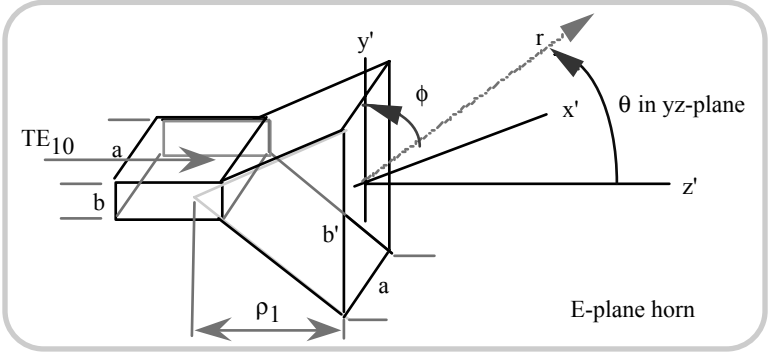


Fig. 4.78 E- plane horn and the associated co-ordinate system at its aperture plane

The far-field components of the electromagnetic radiation from an E-plane sectoral horn:

$$E_\theta = E(kr)\times\left\{e^{j(k_y^2\rho_1/2k)}\sin(\phi)\times[\cos(\theta)+1]\times G(k_x,a)\times F(\tau_1,\tau_2)\right\}$$

$$E_\phi = E(kr)\times\left\{e^{j(k_y^2\rho_1/2k)}\cos(\phi)\times[\cos(\theta)+1]\times G(k_x,a)\times F(\tau_1,\tau_2)\right\}$$

(4.100a)

where

$$E(kr) = -j\frac{a\sqrt{\pi k\rho_1}E_1 e^{-jkr}}{8r} \tag{4.100b}$$

and

$$G(k_x,a) = \cos\left(\frac{k_x a}{2}\right)\times\left[\left(\frac{k_x a}{2}\right)^2-\left(\frac{\pi}{2}\right)^2\right]^{-1} \tag{4.100c}$$

The parameter ρ_1 (as specified earlier for a pyramidal horn) represents the distance of the point/line of convergence of the upper and lower surfaces of the E-plane horn from the aperture plane, as illustrated in Figure 4.78.

Further, $k_x = k\sin(\theta)\cos(\phi)$, $k_y = k\sin(\theta)\sin(\phi)$ and $F(\tau_1, \tau_2) = [C(\tau_2) - C(\tau_1)] - j[S(\tau_2) - S(\tau_1)]$ with τ_1 and τ_2 given by:

$$\tau_1 = \sqrt{\frac{1}{\pi k\rho_1}}\left(-\frac{kb_1}{2}-k_y\rho_1\right) \tag{4.101a}$$

and

$$\tau_2 = \sqrt{\frac{1}{\pi\rho_2}}\left(+\frac{kb_1}{2}-k_y\rho_1\right) \tag{4.101b}$$

In the principal planes, namely E-plane (with $\phi = \pi/2$) and H-plane (with $\phi = 0$), the radiated field components refer to: $\{E_r = E_\phi = 0)$; $E_\theta\ (r, \theta)\}_{\phi=\pi/2}$ and $\{E_r = E_\theta = 0)$; $E_\phi\ (r, \theta)\}_{\phi=0}$ respectively. Typical E- and H-plane patterns of an E-plane

sectoral horn with $b_1 \approx 10b$, $\rho_1 \approx 25b$, $a \approx 2b$ and $b \approx \lambda/4$ are presented in Figure 4.79.

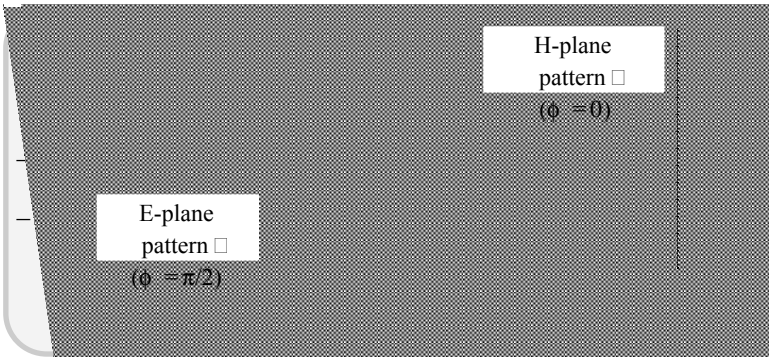


Fig. 4.79 E- and H-plane patterns of a typical E-plane sectoral horn with dominant mode excitation

The directivity of an E-plane sectoral horn is given by:

$$D_E = (64a\rho_1/\pi\lambda b_1)\{C^2[b_1/(2\lambda\rho_1)^{1/2}] + S^2[b_1/(2\lambda\rho_1)^{1/2}]\} \qquad (4.102)$$

and the optimum directivity (considering the associated beamwidth performance) occurs with $b_1 = (2\lambda\rho_1)^{1/2}$.

H-plane sectoral horn

As mentioned earlier, flaring the dimensions of a rectangular waveguide along the directions of the H-field (while keeping the other constant) results in an H-plane sectoral horn depicted in Figure 4.80.

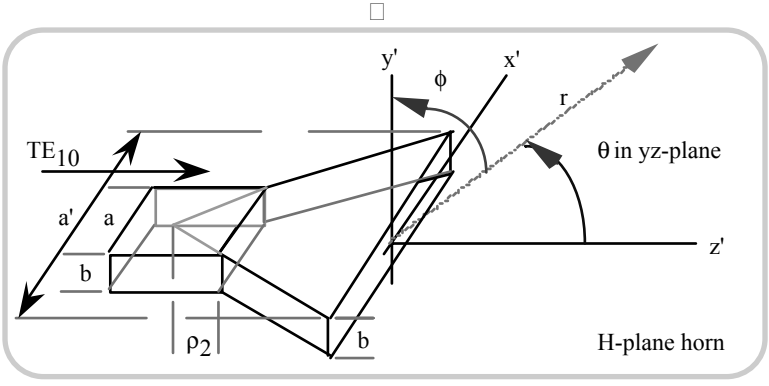


Fig. 4.80 H-plane sectoral horn and the associated co-ordinate system at its aperture plane

Again, following the analysis due to Balanis [4.3], the radiated far-field components of this antenna are given by:

$$E_\theta = jE_2(b/8)\left(\frac{k\rho_2}{\pi}\right)^{1/2}\left(\frac{e^{-jkr}}{r}\right)\times \left\{\sin(\phi)[\cos(\theta)+1]\times\frac{\sin(Y)}{Y}\left[e^{jf_1}F(\tau_3,\tau_4)+e^{jf_2}F(\tau_5,\tau_6)\right]\right\} \quad (4.103a)$$

$$E_\varphi = jE_2(b/8)\left(\frac{k\rho_2}{\pi}\right)\left(\frac{e^{-jkr}}{r}\right)\times \left\{\cos(\varphi)[\cos(\theta)+1]\times\frac{\sin(Y)}{Y}\left[e^{jf_1}F(\tau_3,\tau_4)+e^{jf_2}F(\tau_5,\tau_6)\right]\right\} \quad (4.103b)$$

where

$$\tau_3 = \sqrt{\frac{1}{\pi k\rho_2}}\left(-\frac{ka_1}{2}-k'_x\rho_2\right);\ \tau_4 = \sqrt{\frac{1}{\pi k\rho_2}}\left(+\frac{ka_1}{2}-k'_x\rho_2\right);$$

$$\tau_5 = \sqrt{\frac{1}{\pi k\rho_2}}\left(-\frac{ka_1}{2}-k''_x\rho_2\right);\ \tau_6 = \sqrt{\frac{1}{\pi k\rho_2}}\left(+\frac{ka_1}{2}-k''_x\rho_2\right); \quad (4.103c)$$

Further, $f_1 = [(k'_x)^2\rho_2]/2k$; $f_2 = [(k''_x)^2\rho_2]/2k$; $Y = (kb/2)\sin(\theta)\sin(\phi)$; $F(x_1, x_2) = [C(x_2) - C(x_1)] - j[S(x_2) - S(x_1)]$; $k'_x = k\sin(\theta)\cos(\phi) + (\pi/a_1)$; and $k''_x = k\sin(\theta)\cos(\phi) - (\pi/a_1)$.

The principal plane patterns, namely the E-plane pattern (with $\phi = \pi/2$) and the H-plane pattern (with $\phi = 0$) of a typical H-plane sectoral horn with $a_1 \approx 10a$, $\rho_2 \approx 12a$, $b = 0.5a$ and $a = \lambda/2$ are presented in Figure 4.81.

The directive gain of an H-plane sectoral horn is given by:

$$D_H = (4\pi b\rho_2/\lambda a_1)\{[C(u) - C(v)]^2 + [S(u) - S(v)]^2\} \quad (4.104a)$$

with

$$u = [(\sqrt{\lambda\rho_2} / a_1) + (a_1 / \sqrt{\lambda\rho_2})]/\sqrt{2} \qquad (4.104b)$$

and

$$v = [(\sqrt{\lambda\rho_2} / a_1) - (a_1 / \sqrt{\lambda\rho_2})]/\sqrt{2} \qquad (4.104c)$$

For a specified set of values of a_1/λ and the corresponding set of values of ρ_2 for which maximum directivities are realised correspond to an optimal relation given by: $a_1 \approx \sqrt{3\lambda\rho_2}$.

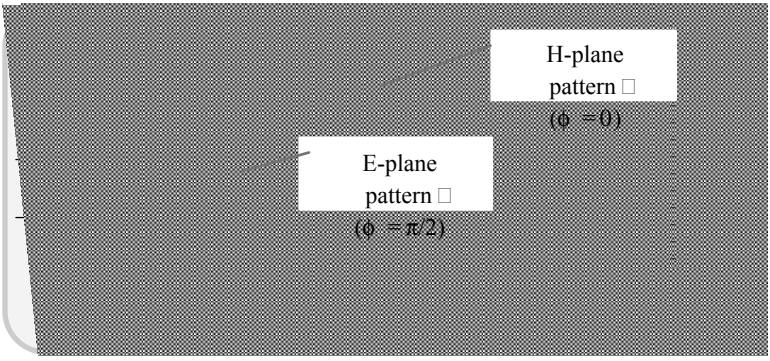


Fig. 4.81 E- and H-plane patterns of a typical H-plane sectoral horn with dominant mode excitation

The scalar horn version of a pyramidal horn is shown in Figure 4.82. It has corrugations in the E-plane wall. Since the H-plane edges do not contribute to edge diffraction to a significant extent, no E-plane wall corrugation is necessary. Again, this corrugated pyramidal horn essentially suppresses the E-wall edge currents and hence the undesirable radiation causing minor lobes/back lobes is minimised.

With a corrugated pyramidal horn having a square aperture, an almost rotationally symmetric radiation pattern with very low side-lobe levels can be obtained. Such antennas are useful as primary feeds for symmetric illumination (in the principal planes) of secondary reflectors so as to minimise the cross-polarisation losses as well as reduce the spill-over efficiency.

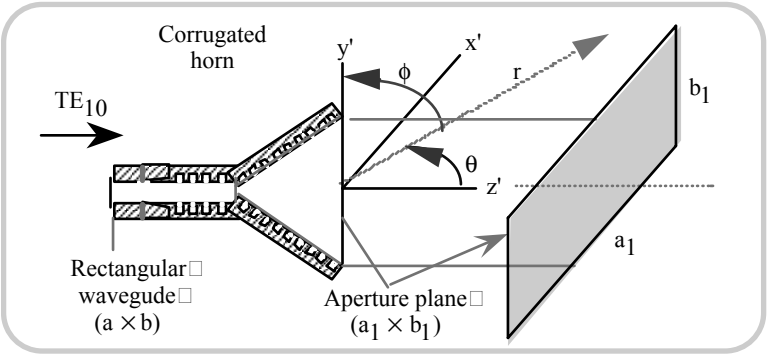


Fig. 4.82 Corrugated pyramidal horn geometry and the associated co-ordinate system

4.10 REFLECTOR ANTENNAS

As indicated before, the reflector antenna is a secondary radiator in the sense that a primary feed (such as a horn) illuminates its surface, which in turn reflects the EM energy to form the far-field radiation. Since the aperture size of the secondary reflector could be made very large (in relation to the operating wavelength), the associated gain will also be significantly high. Such gains are difficult to be realised just with a single (primary) antenna alone. Gain in excess of 30dB in the microwave region is not uncommon with secondary reflector antennas.

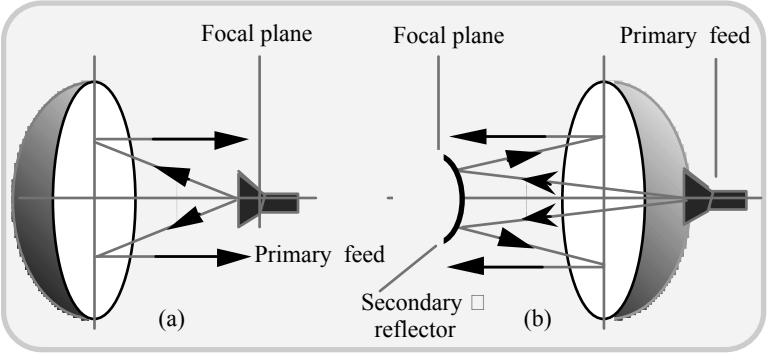


Fig. 4.83 (a) Parabolic reflector excited by a primary feed at the focal plane and (b) Cassegrainian system

A typical secondary antenna is a parabolic reflector (of relatively large size in terms of the wavelength) and it is illuminated by a relatively small primary feed placed in the focal region of the reflector. There are two ways of illuminating the

parabola: the first one as said above, refers to having a primary feed at the focal plane illustrated in Figure 4.83(a). Alternatively, a Cassegrainian version can be adopted in which, a feed illuminates a small secondary reflector, which in turn illuminates the primary reflector, as shown in Figure 4.83(b).

An aperture of a parabolic reflector of diameter D illuminated by a field $F(\rho, \phi')$, leads to a radiation pattern derived *via* Huyghen's principle (Figure 4.84). It is given by:

$$S(\theta, \phi) = \int_{\phi'=0}^{2\pi} \int_{\rho=0}^{D/2} F(\rho, \phi') \times \exp[jk\rho \sin(\theta) \cos(\phi - \phi')] \rho \, d\rho d\phi \qquad (4.105a)$$

which simplifies to the following relation when $F(\rho, \phi') = 1$ (corresponding to an uniform illumination):

$$S(\theta, \phi)\Big|_{F(\rho,\phi')} = \frac{\pi D^2}{2} \frac{J_1\left[\frac{\pi D}{\lambda} \sin(\theta)\right]}{\left[\frac{\pi D}{\lambda} \sin(\theta)\right]} \qquad (4.105b)$$

There are two versions of parabolic reflectors. The reflector illustrated in Figure 4.85 is a paraboloid of revolution, which is illuminated, as mentioned before, by a primary feed (or by a secondary reflector) placed at the focal plane of the paraboloid. Considering the relative size of the parabolic reflector and the feed, the source of EM energy illuminating the parabolic surface can be regarded as a point source.

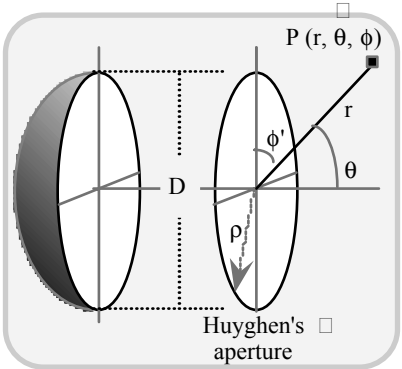


Fig. 4.84 Radiation from a parabolic reflector: Application of Huyghen's aperture concept

The other version of a parabolic reflector is formed by a part of a parabolic cylinder with a cheese-like shape as shown in Figure 4.85(b). The source that is placed at the focal plane of this (singly curved) parabolic cylinder should be a line source compatible for illuminating the reflecting surface geometry.

Considering a rectangular aperture depicting a line-source that illuminates the surface of a parabolic cylindrical reflector and assuming the field distribution on the reflector as F(y), the radiation pattern (obtained via Huyghen's aperture principle) as a function of ϕ in the xy-plane is given by:

$$E(\phi) = \int_{\phi'=0}^{2\pi} F(y) \exp[(jky) \times \sin(\phi)]dy \tag{4.106}$$

which to the following in the case of uniform aperture field distribution (denoted by F(y) = 1):

$$E(\phi)\Big|_{F(y)=1} = \frac{\sin\{(kL/2)\sin(\phi)\}}{(kL/2)\sin(\phi)} \tag{4.107}$$

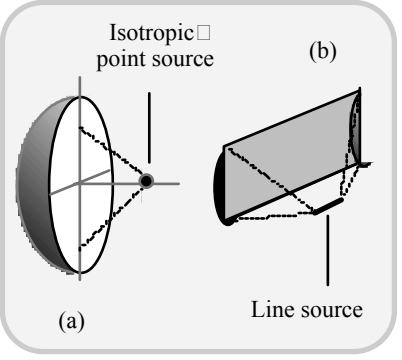


Fig. 4.85 (a) Paraboloid of revolutions and (b) parabolic cylinder: Concepts of point- and line-source excitations

For other aperture field distributions, such as F(y) = cos($\pi\gamma$/L), for example, the resulting expression is follows:

$$E(\phi)\Big|_{F(y)=\cos(\pi yL/\lambda)} = \frac{(\pi/2)^2 \cos\{(kL/2)\sin(\phi)\}}{(\pi/2)^2 - [(kL/2)\sin(\phi)]^2} \tag{4.108}$$

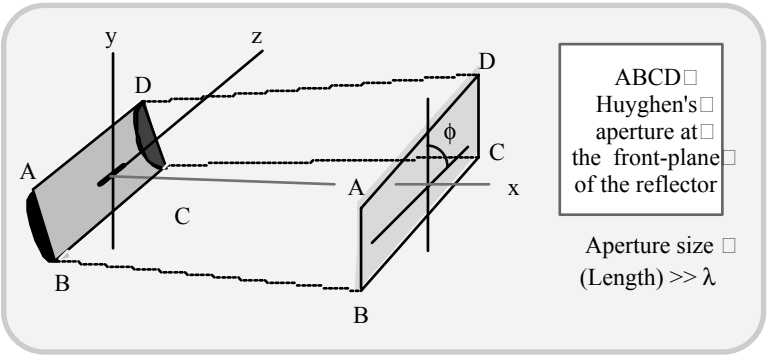


Fig. 4.86 The Huyghen's aperture of the parabolic cylindrical reflector and the associated co-ordinate system at its aperture

4.10.1 Paraboloidal reflector *versus* parabolic cylindrical reflector

The following are the contrasts that can be observed in the design and characteristics of a paraboloidal reflector and a parabolic cylindrical reflector:

Paraboloidal reflector	Parabolic cylindrical reflector
• Doubly-curved surface	• Singly-curved surface
• Aperture field amplitude fall-off from the feed to the reflecting surface is proportional to $1/r^2$ (due to point-source excitation)	• Aperture field amplitude fall-off from the feed to the reflecting surface is proportional to $1/r$ (due to line-source excitation)
• Cross-polarisation components prevail	• Cross-polarisation components do not exist
• Manufacture requires casting or fabrication of segmented pieces to form the paraboloid	• Mechanically simpler to construct
• Less aperture blockage	• More aperture blockage

Some additional characteristics of paraboloidal reflector antennas are as follows:

- *Gain and diversity*: With large D/λ ratio, the directivity with uniform aperture field distribution is equal to $(4\pi/\lambda^2) \times$ (area of the aperture). The power gain relative to a λ/2-dipole is $G = 6(D/\lambda)^2$. The directivity and gain will be less for tapered illumination of the aperture. Other reasons for such reductions are: Pattern asymmetry, nonaligned phase-center, the existence of cross-polarised fields, presence of primary feed blockage and the prevalence of random surface errors
- *Aperture efficiency*: This defines how efficiently the physical area of the reflecting surface is used, and it is decided by focal length-to-diameter ratio (f/D) and by the directive pattern of the primary feed
- The extent of utilization of the surface of a reflector is decided by a factor *u* given by:

$$u = 5.15 \times [0.423 \times J_1(3.5D/4f) + 0.26 \times (D/4f)]^2 \quad (4.109)$$

which identically depicts the aperture efficiency. Thus, u is essentially a function of f/D ratio.

4.10.2 Concept of phase centre

Considering the far-zone radiated field components of an antenna, they can be specified by an amplitude function $E(\theta, \phi)$, and an associated phase variation, $\psi(\theta, \phi)$.

Suppose a reference point at the location of the antenna (for a given operating frequency) is assigned such that $\psi(\theta, \phi)$ is constant. Hence, $\psi(\theta, \phi)$ is invariant of θ and ϕ. Such a reference point is known as the *phase-centre* of the antenna. In reference to this phase-centre, the far-fields of the antenna can be described in terms of spherical wavelets (emanating from the phase-centre location) with ideal spherical phase fronts depicting equiphase surfaces.

Having a reference point like the phase-centre is crucial in applications like global positioning, homing, landing and other navigational systems.

For antenna structures extended in space (such as a reflector or an array), specifying a unique phase-centre may not be feasible, though over the angular space covered by the main lobe, an approximate phase-centre can be specified.

The need for a specified phase-centre can be understood by considering a paraboloidal reflector. Suppose a plane wave is incident on the reflector. It can be collected at the focal point. Likewise, a point-source at the focal point will set a spherical wavefront, which when reflected by the reflector will form a plane wave. In practice, no antenna would facilitate an ideal point-source so as to cause

spherical wavefronts in the desired angular space. However, *via* proper design, a point can designated, where, if a source is placed would enable spherical fronts over most of the angular space.

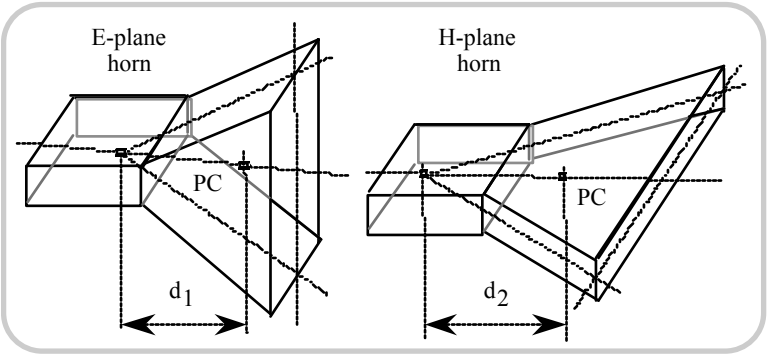


Fig. 4.87 Phase centres (PC) of E-and H-plane horns. (In a pyramidal horn, the location of the phase centre is given approximately by the average of d_1 and d_2)

Suppose such a point is identified for a primary feed and it coincides with the focal point of the secondary reflector illuminated by the feed. For horn antennas, locations of the phase centre are shown in Figure 4.87. For a given horn geometry, the phase centre is an imaginary apex point (from which the spherical wave fronts are assumed to emerge out). Its location is a function of the flare angle [4.50]. The larger the flare angles, the further will be the phase centre from the aperture of the horn.

For pyramidal horns, the phase centre is approximately the average of d_1 and d_2 indicated for the E-and H-plane horns. For conical horns, the location of the phase center has been computed in [4.51].

4.10.3 Plane sheet reflector

Among the family of reflector antennas, the simplest one corresponds to a large, flat conducting sheet placed behind a linear dipole antenna. Essentially this reflector reduces the backward radiation and offers substantial gain in the forward direction for a small spacing between the dipole and the sheet.

The size of the reflecting sheet in relation to the size of the dipole and/or the wavelength of operation together with the dipole-to-sheet spacing decides eventually the radiation pattern of the system. Planar reflector concepts are used in the developments of the so-called *corner reflectors* described in the following section.

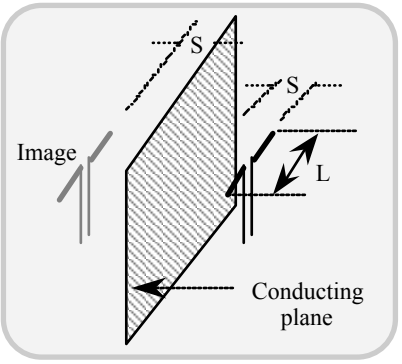


Fig. 4.88 (a) Flat reflector plus a dipole source and (b) image of the source decided by plane-mirror analogy

4.10.4 Corner reflector antennas

A *corner reflector* is a hollow structure whose internal planar surface(s) are used to reflect the electromagnetic energy falling on them from an excitation source (a driven radiator) placed within the corner reflector volume.

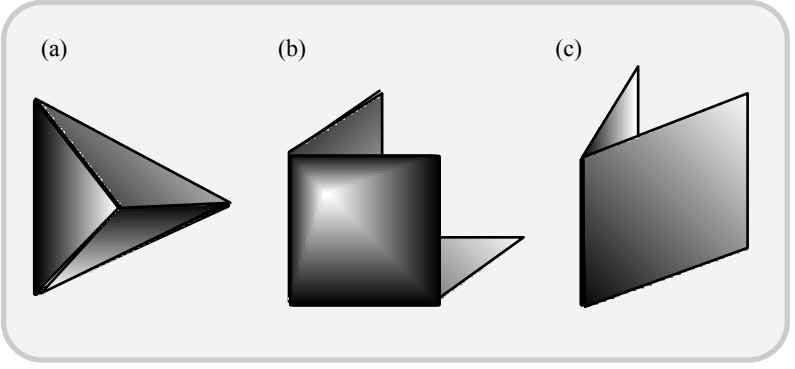


Fig. 4.89 Corner reflectors: (a) Rectangular type (square trihedral), (b) triangular type (triangular trihedral) and (c) dihedral type

Such systems are essentially secondary radiators designed to give a directional, sharp radiation pattern. A corner reflector can be made either from a part of a hollow rectangular pyramidal structure (square trihedral) or from a part of a hollow triangular pyramidal structure (triangular trihederal) or a dihedral corner as illustrated in Figure 4.89.

The reflecting surfaces of the corner reflector can be made of a conducting plate or with closely placed parallel grid wires. Normally, such a grid type corner reflector is adopted to realize a dihedral, V-type reflecting radiator as shown in Figure 4.90.

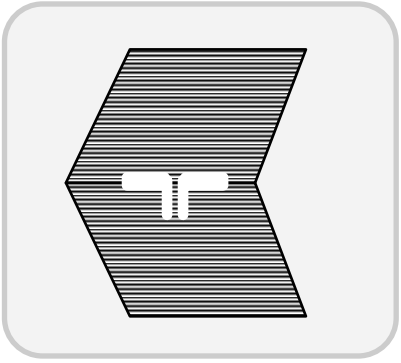


Fig. 4.90 V-type (dihedral) corner reflector made of parallel grid wires and excited by a dipole, as illustrated

The relation between the aperture angle and the half-power beamwidth of a corner reflector is deduced in [4.52].

When the aperture angle is 180° (corresponding to a flat plate), the beamwidth is nearly equal to 120°. If more than 120° is required (so as to realise a sector beam), the corner reflector to be designed should have an aperture angle of more than 180° (in which case it is known as *superior angle corner reflector*).

A dual frequency radiating system can also be obtained with a corner reflector sector-beam antenna. This will require a dual frequency primary radiator. A dipole plus a closely-spaced parasite can be used for this purpose. For a pair of dual frequencies f_1 and f_2, the aperture angle, the distance of the primary source from the reflector and the width of the corner reflector would decide the design in achieving sector beams of equal beamwidths, at f_1 and f_2. Mostly, the width of the corner reflector would dominate in such designs than the aperture angle.

The coordinate system specified for a corner reflector and how it is designated to various reflectors are shown in Figure 4.91.

The corner reflector antenna can be analysed via *method of images*. Typically, the primary feed that excites a corner reflector could be a simple dipole or an aperture (such as a horn antennas). By assuming symmetry (where applicable) of the set of (perfectly) conducting intersecting planes, a set of images of the primary source are assumed to form as a result of the reflecting surfaces. These images are then presumed to constitute a set of array elements and the corresponding radiation pattern thereof is determined.

For example, suppose a current element source (feed) is presumed on the bisector plane of a corner angle π/N, where N is an integer as depicted in Figure

4.92. Referring to Figure 4.92, application of the principle of images would lead to 2N equi-spaced current elements on a circle of radius R_o in the adjacent (image) elements at ($-\pi/2N \le \theta \le +\pi/2N$) in the plane $\phi = 0$. Hence, the corresponding E-field parallel to the z-axis can be written in the form,

$$E(\phi) = NI_o \int_{m=0}^{\infty} j^{(2m+1)N} J_{(2m+1)N}[kR_o \cos(\phi)]\cos[(2m+1)\theta]d\theta \tag{4.110}$$

where I_o is the current through the source element. This E-field will become zero at $\theta = \pi/N$, namely at the planes of the conducting (reflector) planes.

Should the element have a pattern $f(\theta, \phi)$, the resultant pattern of the corner reflector is obtained by multiplying the array factor by this element pattern of equation (4.110).

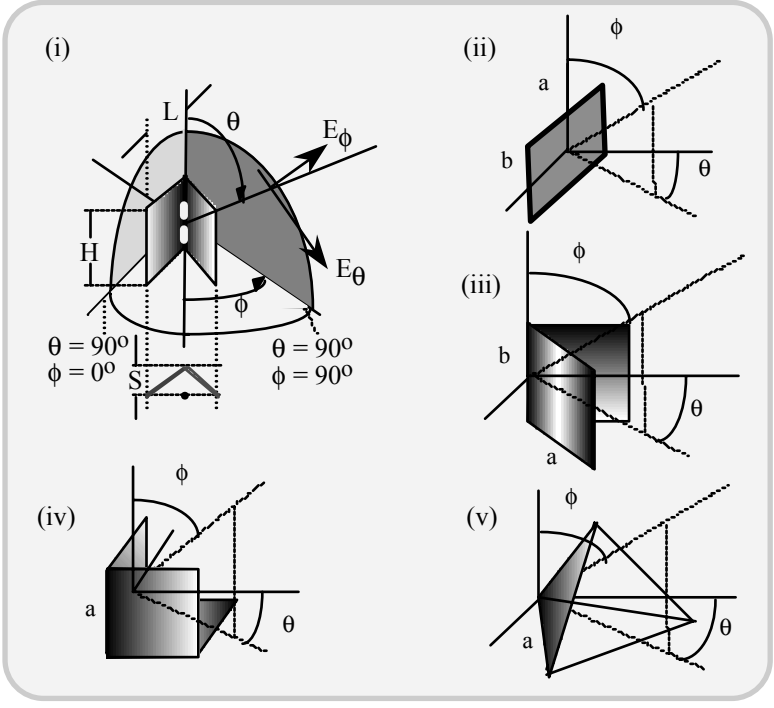


Fig. 4.91 Corner-reflector antenna with a feed source(s): (a) Co-ordinate system; (b) flat reflector; (c) dihedral reflector; (d) trihedral square reflector and (e) trihedral triangular reflector

A corner reflector in typical wireless communication applications are designed to provide a sector beam with controllable beamwidth by adjusting the aperture angle between the reflecting surfaces. For example, with a λ/2-dipole excitation, sector beams of widths ranging from 60° to 180° can be obtained by setting the corner aperture angle from 60° to 270°.

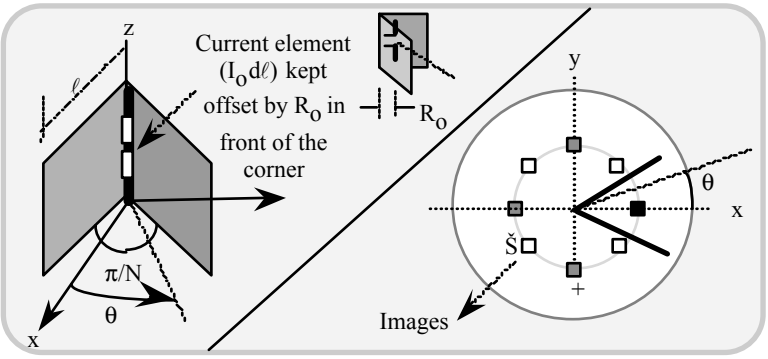


Fig. 4.92 (a) Corner reflector and (b) circular array elements due to source and multiple images

4.11 SURFACE-INSTALLED LOW PROFILE ANTENNAS

A surface-mounted antenna or a radiating structure refers to a system made flush to a surface of a body conformal to the shape. The surface that supports the antenna could be a simple planar or a curved surface.

Planar-surface installed flush and/or low profile antennas are designed to be a two-dimensional (planar) configuration. In general, a conformal antenna "conforms" to the shape of the surface on which it is installed. The shape (curvature) of the surface (of the body) could be two- or three-dimensional geometry. For example, a conformal antenna is of a two-dimensional type when the surface on which it is flush-mounted is a planar or a cylindrical surface; whereas, if it conforms to a paraboloidal or a spherical surface, it is a three-dimensional flush antenna.

The conformal antenna that is designed to reside flush over a surface (planar or curved) can be made of a simple element or by a set of (array of) elements. Typically, such elements could be of the following types:

- Linear wire antennas
- Loops, spirals, bifilar/quadrifilar helices etc.
- Patches (conducting or dielectric patches)
- Simple slots
- Apertures

Typical examples of these antennas are presented in Figure 4.93.

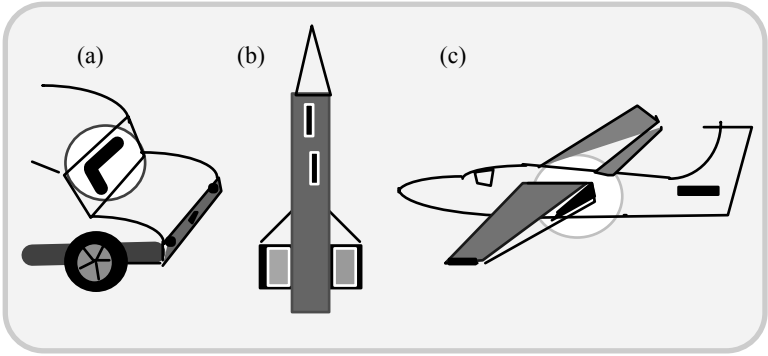


Fig. 4.93 Conformal, flush-mounted surface antennas: (a) Conducting patch/wire antenna printed on the rear wind-shield; (b) slot antennas on the surface of a missile etc. and (c) a cavity-backed aperture flush-type antenna installed on an aircraft surface

In the context of wireless systems, a surface-mounted antenna may not be strictly "flush" to the surface in the sense that, it could be slightly raised above the surface but conformed to the surface, so as to maintain a low-profile and/or low-drag considerations.

The generic versions of such antennas designed on planar surfaces can be categorized as follows:

- Wire-bonded or wire-printed linear antenna or loop structures on an insulating surface (such as glass). (The insulating surface mostly acts as a mechanical support and does not form a part of radiating process)
- Inverted L-shaped/F-shaped linear antennas, slightly raised above the planar conducting surface
- Vertical semi-loop projecting slightly on a planar conducting surface
- Conducting patch constituting a linear antenna or a loop on a dielectric surface
- Slots and apertures flush to a conducting surface with a cavity-backing
- Planar conducting bifilar spirals (equiangular, logarithmic and Archimedian type) on a dielectric base
- Planar dielectric radiating "patch" on a conducting surface
- Planar arrays formed with a set of elements of the types indicated above

Wire-bonded/wire-printed linear and loop antennas

These are basically the same as those of traditional linear or loop antennas, except that they are mounted flush conformal to a planar insulating surface (such as the windshield of an automobile). Typically, their geometry and form are meant to realise low-profile and low-drag characteristics. Since the insulating base simply acts as a mechanical support structure, the radiating characteristics of these structures are almost akin to those placed in free-space.

Typical examples of these types correspond to AM/FM radio, TV reception antennas mounted on the rear windshield and/or side-quarter windows of automobiles. Such structures can be used in multiple locations (such as rear and quarter windows) for space-diversity purposes. A small loop built inside a minor box can operate as a MF band antenna for door remote control applications.

Surface-mounted inverted L- and F-linear antennas

The inverted L- and F-linear antennas are described in Section 4.2. In flush-mounting form, they represent monopole wire structures having a low profile and projecting slightly above the planar surface.

Further, the surface-mounted radiators can be configured for both horizontal and vertical polarisations. Among the possible surface-mounted structures, a form of inverted antenna refers to a V-type dipole shown in Figure 4.94. This again is made as a low-profile structure (for low-drag considerations when mounted on automobiles). It gives a directional pattern with horizontal polarisation and used for VHF/UFH band communications and TV reception.

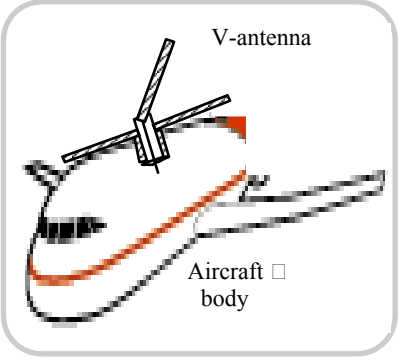


Fig. 4.94 V-shaped dipole on a planar surface such as the fuselage of an aircraft

Vertical semi-loop installed on a surface

This refers to a half-loop mounted vertically on a planar conducting surface (with a low profile). Its polarisation is parallel to the axis of the surface. The loop is also loaded with a reactance.

The structure is illustrated in Figure 4.95 and adopted on the fuselage of an aircraft for VHF band beacon marker applications.

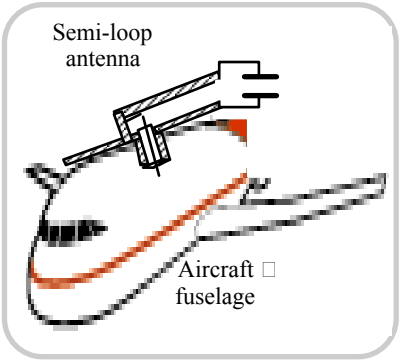


Fig. 4.95 Half-loop antenna mounted on a fuselage surface. (Both V- and loop-structure on the fuselage are designed as streamlined geometry)

Conducting patch planar antennas

These are more specifically known as *microstrip antennas*. A microstrip is an open guided-wave structure constituting a transmission line. The microstrip antenna is a radiating structure conceived out of a microstrip transmission line assembly, which is physically simple and flat. Figure 4.96 shows a simple microstrip patch antenna.

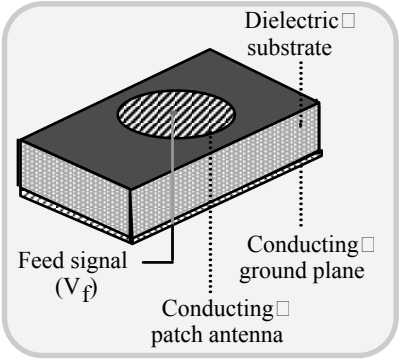


Fig. 4.96 Microstrip patch antenna driven against ground-plane: Patch geometry is a design choice

The upper surface of the dielectric substrate supports a printed conducting strip (usually a copper trace), which is suitably contoured (while the entire lower surface of the dielectric substrate is backed by a conducting ground plate). Such an antenna is sometimes referred to as a *printed antenna* because of the associated fabrication process. The microstrip antenna is similar to a slot antenna not only

because it is a flat antenna, but also due to its radiation characteristics, which can be derived by considering it as a slot antenna with magnetic walls.

Many types of microstrip antennas have been evolved, which are primarily variations of the basic sandwich structure shown in Figure 4.96. The common property of all these antennas is that they are electrically driven with respect to the ground-plane. A wide range of dielectric substrate thickness and permittivity is permissible and judiciously adopted in practice. Air (free-space) is also a feasible substrate alternative.

Microstrip antennas can be designed as very thin planar printed antennas and they are very useful elements in making different types of arrays (especially conformal arrays that can be designed on a surface of any type and shape).

4.11.1 The microstrip as a transmission line

Figure 4.97 illustrates a microstrip line. Usually, the substrate is a dielectric of relative permittivity $\varepsilon_r > 1$ and relative permeability $\mu_r = 1$. A simple analysis assumes that the line supports a TEM wave and the characteristic line impedance is deduced approximately as,

$$Z_d = 1/v_d C_d \quad \text{ohms} \tag{4.111a}$$

where v_d is the velocity of propagation along the line and C_d is the line capacity per unit length. If Z_0, v_0 and C_0 denote the corresponding parameters for an air-spaced line, then,

$$\left.\begin{aligned} v_d &= \frac{v_o}{\sqrt{\varepsilon_r}} \quad C_d = C_o \varepsilon_r \\ v_o &= \frac{1}{\sqrt{\mu_o \varepsilon_r}} \\ Z_d &= \frac{Z_o}{\sqrt{\varepsilon_r}} \end{aligned}\right\} \tag{4.111b}$$

The dielectric material fills only a partial space above the ground-plane of the microstrip. Therefore, it is reasonable to expect that the corresponding values of velocity, capacity and impedance (denoted by v_m, C_m and Z_m respectively), to lie somewhere between the values corresponding to two extreme cases, namely a completely dielectric-filled case and the other of empty space. Thus, $v_d < v_m < v_o$, $C_o < C_m < C_d$ and $Z_d < Z_m < Z_o$.

Hence, it is convenient to introduce the concept of a *filling-factor*, ($0 \le q \le 1$) and an *effective relative permittivity*, ($1 \le \varepsilon_e \le \varepsilon_r$), so that the following set of transmission line relations can be explicitly specified:

$$\varepsilon_e = 1 + q(\varepsilon_r - 1)$$

$$v_m = \frac{v_o}{\sqrt{\varepsilon_e}}$$

$$C_m = C_o \varepsilon_e$$

$$Z_m = \frac{Z_o}{\sqrt{\varepsilon_e}}$$

$$\lambda_m = \frac{\lambda_o}{\sqrt{\varepsilon_e}}$$

$$\beta = \frac{2\pi}{\lambda_m} \qquad (4.112)$$

where λ_m is the wavelength in the dielectric-filled line and λ_o is the wavelength in free-space. The parameters C_o and Z_o can be calculated by a conformal transformation method as shown by Wheeler [4.53, 4.54].

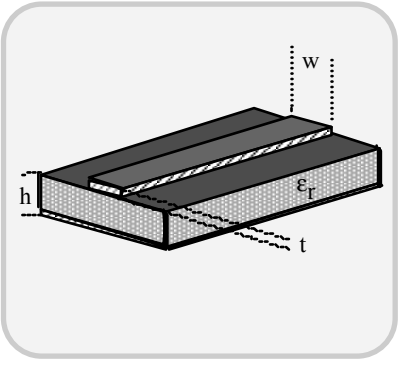


Fig. 4.97 A microstrip line: A planar transmission line set on a dielectric substrate backed by a conducting ground-plane

Using the relations given by equation (4.112), the properties of a microstrip can be approximately decided. In general, a microstrip structure can support not only the TEM mode but also a set of discrete hybrid modes having non-zero E_z and H_z components, where z denotes the direction of propagation. However, the lowest-order hybrid mode strongly resembles a TEM mode and it is called a *quasi-TEM mode* or *wave*.

Wheeler's basic expression for Z_m is given by:

$$Z_m = \frac{Z_o}{\pi\sqrt{2}\sqrt{\varepsilon_{r+1}}} \left\{ \begin{array}{l} \ln\left(\frac{8h}{w}\right) + \frac{1}{32}\left(\frac{w}{h}\right)^2 - \\ \frac{1}{2}\left(\frac{\varepsilon_r - 1}{\varepsilon_r + 1}\right)\left(\ln\left[\frac{\pi}{2}\right] + \frac{1}{\varepsilon_r}\ln\left[\frac{4}{\pi}\right]\right) \end{array} \right\} \quad \text{ohms} \tag{4.113a}$$

for w/h < 1; and,

$$Z_m = \frac{Z_o}{2\sqrt{\varepsilon_r}} \left\{ \begin{array}{l} \frac{w}{2h} + 0.441 + 0.082\left(\frac{\varepsilon_r - 1}{\varepsilon_r^2}\right) + \\ \left(\frac{\varepsilon_r + 1}{2\pi\varepsilon_r}\right)\left[1.451 + \ln\left(\frac{w}{2h} + 0.94\right)\right] \end{array} \right\}^{-1} \quad \text{ohms} \tag{4.113b}$$

for w/h > 1. In the above expressions Z_o denotes the free-space intrinsic impedance equal to 120π (= 377) ohms.

The radiation from a patch antenna can be derived by assuming that, either it is due to the current distribution on the patch, or due to the field that prevails on the patch-edge (which can be identically represented by equivalent magnetic currents existing at a slot). The method based on the field in the slot is easier to handle.

Suppose V is the voltage at the patch-edge with respect to ground. Then the far-field electric radiation vector is given by:

$$\mathbf{L} = -2 \int_{C_\ell} \mathbf{u}_Z \times \mathbf{u}_n V(c_\ell) \exp[(jk_o r)\cos(\psi)] dc_\ell \tag{4.114}$$

where the integration range C_ℓ refers to the closed loop around the effective edge of the patch as depicted in Figure 4.98. The corresponding radiation field components E_θ and E_ϕ are given by

$$\left. \begin{array}{l} E_\theta = -j\left[\frac{L_\phi}{2\lambda_o R}\right]\exp[-jk_o R] \\ E_\phi = +j\left[\frac{L_\theta}{2\lambda_o R}\right]\exp[-jk_o R] \end{array} \right\} \tag{4.115}$$

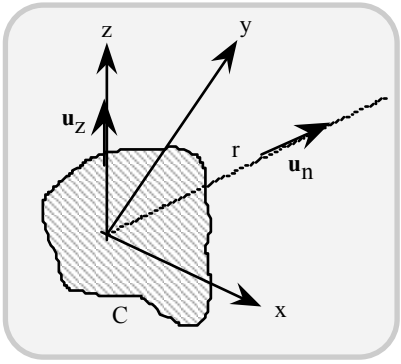


Fig. 4.98 Coordinate system for calculating the radiation from an arbitrary patch

The associated magnetic field components are: $H_\theta = -E_\phi/Z_o$ and $H_\phi = -E_\theta/Z_o$. Applying the above relations of equation (4.115) to the rectangular patch shown in Figure 4.99, and assuming the field components as duals of the well-known rectangular waveguide TE modes, the resulting field components are as follows:

$$\left.\begin{aligned} E_x &= \left(\frac{V_o}{h}\right)\cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right) \\ H_x &= -\left(\frac{j\omega\varepsilon_o}{k_o^2}\frac{n\pi}{b}\frac{V_o}{h}\right)\cos\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right) \\ H_y &= +\left(\frac{j\omega\varepsilon_o}{k_o^2}\frac{n\pi}{a}\frac{V_o}{h}\right)\sin\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right) \end{aligned}\right\} \qquad (4.116)$$

In equation (4.116), V_o depicts the peak voltage at the corner of the patch. The resonant frequency of the patch is the same as the cut-off frequency of the rectangular waveguide and is given by,

$$f_r = \frac{1}{2\sqrt{\mu_o\varepsilon_o\varepsilon_r}}\left\{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2\right\}^{1/2} \qquad (4.117)$$

Using equation (4.97), the voltage at the edges of the patch are deduced as follows:

$$\begin{aligned} V(y) &= V_o\cos(n\pi y/b) && \text{at } x = 0;\ 0 \le y \le b \\ V(y) &= V_o\cos(n\pi y/b)\cos(m\pi) && \text{at } x = a;\ 0 \le y \le b \end{aligned}$$

$$V(x) = V_o\cos(m\pi x/a) \qquad \text{at } x = 0;\ 0 \le y \le a$$
$$V(x) = V_o\cos(m\pi x/a)\cos(n\pi) \qquad \text{at } y = b;\ 0 \le x \le a$$

(4.118)

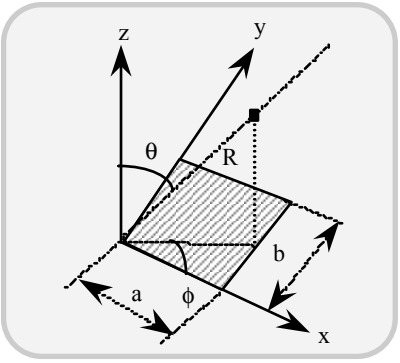


Fig. 4.99 A rectangular (a × b) patch antenna placed in the xy-plane and the associated co-ordinate system

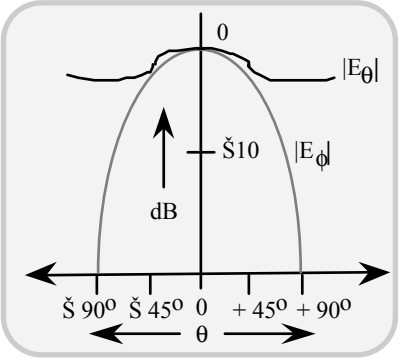


Fig. 4.100 Theoretical radiation patterns of a square patch antenna in ϕ = 45°/225° plane $\varepsilon_r = 2.0$

Further substituting equation (4.118) in equation (4.115), it follows that

$$\left.\begin{aligned} L = 2jV_o\Big[1 - \cos(m\pi)\exp\{jk_o a\cos(\phi)\sin(\theta)\}\Big]\times \\ \Big[1 - \cos(n\pi)\exp\{jk_o b\sin(\phi)\sin(\theta)\}\Big]\times \\ \Big[\mathbf{u}_x k_o\cos(\phi)\sin(\theta)/D_1 - \mathbf{u}_y k_o\sin(\phi)\sin(\theta)/D_2\Big] \end{aligned}\right\}$$

(4.119a)

where

$$\left.\begin{aligned} D_1 &= k_o^2 \cos^2(\phi)\sin^2(\theta) - \left[\frac{m^2\pi^2}{a^2}\right] \\ D_2 &= k_o^2 \sin^2(\phi)\sin^2(\theta) - \left[\frac{n^2\pi^2}{b^2}\right] \end{aligned}\right\} \qquad (4.119b)$$

Converting equation (4.119a) to spherical polar co-ordinates, and substituting in equations. (4.115) results in,

$$E_\theta = -\left[\frac{\exp(-jk_oR)}{2\lambda_oR}\right]\times 2V_o \times \left[1-\cos(m\pi)\exp\{jk_oa\cos(\phi)\sin(\theta)\}\right]\times$$
$$\left[1-\cos(n\pi)\exp\{jk_ob\sin(\phi)\sin(\theta)\}\right]\times\left[k_o\sin(\theta)\cos(\phi)\sin(\phi)\right]\times\left[\frac{1}{D_1}+\frac{1}{D_2}\right] \qquad (4.120a)$$

$$E_\phi = \left[\frac{\exp(-jk_oR)}{2\lambda_oR}\right]\times 2V_o \times \left[1-\cos(m\pi)\exp\{jk_oa\cos(\phi)\sin(\theta)\}\right]\times$$
$$\left[1-\cos(n\pi)\exp\{jk_ob\sin(\phi)\sin(\theta)\}\right]\times\left[k_o\sin(\theta)\cos(\theta)\right]\times$$
$$\left[\frac{\sin^2(\phi)}{D_2}-\frac{\cos^2(\phi)}{D_1}\right] \qquad (4.120b)$$

The lowest-order mode for which m = 0 and n = 1 is the most important (dominant) case. For this mode the radiation fields in the $\phi = 0°$ and $\phi = 90°$ planes are given by,

$$E_\theta(\phi = 0^\circ) = 0 \qquad (4.121a)$$

$$E_\phi(\phi = 0^\circ) = -j\left[\frac{\exp(-jk_oR)}{2\lambda_oR}\right]\times(4V_oa)\times\exp\{j\frac{k_oa}{2}\sin(\theta)\}\cos(\theta)\times$$
$$\left[\sin\left\{\left(\frac{k_oa}{2}\right)\sin(\theta)\right\}/\left\{\left(\frac{k_oa}{2}\right)\sin(\theta)\right\}\right] \qquad (4.121b)$$

$$E_\phi(\phi = 90^\circ) = -j\left[\frac{\exp(-jk_o R)}{2\lambda_o R}\right] \times (4V_o a) \times \exp\left\{j\left(\frac{k_o b}{2}\right)\sin(\theta)\right\} \times \cos\left\{\left(\frac{k_o b}{2}\right)\sin(\theta)\right\} \quad (4.121c)$$

$$E_\theta(\phi = 90^\circ) = 0 \quad (4.121d)$$

Figure 4.100 shows the theoretical radiation patterns of a square patch antenna in ϕ = 45°/225° plane. It can be seen that the radiation patterns are not very directional. However, by using arrays of patch antennas, highly directional patterns can be obtained.

The bandwidth of a square or a circular patch antenna for a voltage standing wave ratio S, is given by:

$$\text{Bandwidth} = \left[\frac{100(S-1)}{\sqrt{S}}\right]\left[\frac{8}{\varepsilon_r}\right]\left[\frac{h}{\lambda_o}\right]\%, \quad (4.122)$$

which shows that the bandwidth decreases with decreasing h. Thus, thinner antennas will have lesser bandwidth.

The feed structure of a microstrip antenna is often printed on the substrate surface as a part of the radiating element. The feeder lines may introduce additional losses and hence a reduced overall efficiency can be anticipated. The shape of the patch is chosen to obtain desirable radiation characteristics and impedance matching considerations. Typically, simple geometry such as rectangular, square, circular, triangular, bow-tie shape etc. is used in the designs.The mechanical tolerances of thin microstrip antennas usually place a limit on the precision with which the aperture phase and amplitude distributions can be controlled in manufacture. Further, choice of substrate dielectric is crucial in realising specified bandwidth performance.

Specific types of microstrip antennas adopted in wireless communication applications are described in the following sections (of Part II).

Part II

4.12 MICROSTRIP ANTENNAS FOR WIRELESS APPLICATIONS

There are number of considerations in the choice of an antenna element for wireless communication applications that meet specific requirements such as,

compactness (low-profile), bandwidth performance, beamwidth requirements, radiation patterns in specified planes, polarisation needs and multifunctional characteristics (like dual-polarisation feasibility and multifrequency operation etc.). Hence, different versions of antenna elements either in their basic configuration or in certain specified modified forms have been conceived and adopted in both fixed basestation applications as well as in portable units.

Specifically, a variety of microstrip patch antennas have been found compatible with wireless communication systems, mainly due to their low profile, light weight and compactness with the possibility of making such an antenna as a part of the PCB used in the device. For example, printed-dipoles (with a parasitic reflector) yielding a semicircular horizontal radiation pattern have been conceived for use in automobiles. This relevant concept has also been extended for dual frequency operations at 900 MHz and 1.5 GHz with a directivity of 4.5 dBi. Likewise, a simple circular patch for UHF band GPS applications is conceivable to get a low-profile narrow bandwidth pattern with semihemispherical coverage and circular polarisation.

Discussed in this part of the chapter are relevant heuristics, design perspectives and practical considerations addressed to microstrip patch antennas.

4.12.1 Circular and rectangular microstrip patch antennas

Microstrip patches of circular and rectangular geometry are useful as antenna elements for compact units of wireless communication systems Considering a circular patch antenna that supports a TM_{110}-mode excitation, the radiating (effective) surface radius (in m) is specified approximately by:

$$a = \left[1 + \frac{2\tau}{\pi a_o \varepsilon_r}\left\{\ln\left(\frac{\pi a_o}{2\tau}\right) + 1.7726\right\}\right]^{-1/2} \qquad (4.123)$$

where $a_0 = 1.8412\lambda_o/2\pi\sqrt{\varepsilon_r}$ and (τ, ε_r) depict the thickness (in m) and dielectric constant of the substrate respectively. Further, λ_o is the resonant free-space wavelength equal to c/f with $c = 3 \times 10^8$ m/s and f is the frequency in Hz.

For a rectangular patch of (length × width) equal to (L × W), the relevant dimensions for TM_{010}- or TM_{00}-mode excitation are:

$$L = \frac{\lambda_o}{2\sqrt{\varepsilon_e}} - 2\Delta\ell \qquad \text{(in metre)} \qquad (4.124a)$$

$$W = (\lambda_o/2)[(\varepsilon_r+1)/2]^{-1/2} \qquad \text{(in metre)} \qquad (4.124b)$$

where

$$\varepsilon_e = [(\varepsilon_r + 1)/2] + [(\varepsilon - 1)/2] + [1+12\tau/W]^{-1} \qquad (4.124c)$$

and

$$\Delta\ell = (0.412\tau)[(\varepsilon_e + 0.3)(0.264 + W/\tau)]/[(\varepsilon_e - 0.258)(0.8 + W/\tau)] \quad (4.124d)$$

The patches designed as above, will be highly resonant yielding a narrow bandwidth of only 2%. Though broadbanding can be attempted *via* proper matching networks, such network syntheses are invariably complex. Alternatively, the thickness of the substrate can be increased and its dielectric constant can be decreased so as to achieve broader bandwidths.

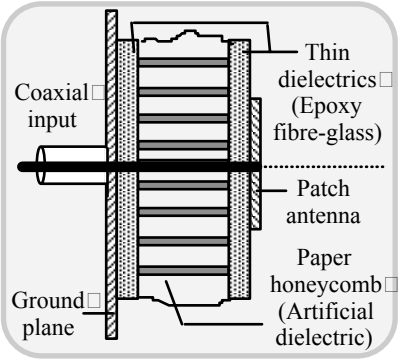


Fig. 4.101 Patch antenna backed by a thick composite substrate (artificial dielectric)

A method of accomplishing this strategy is as follows: The substrate used can be made of an artificial dielectric [4.33] facilitating a large thickness with reduced effective permittivity as shown in Figure. 4.101. The geometry of the patch antenna indicated in Figure 4.101 offers an improved bandwidth to an extent of about 8% with VSWR < 2. However, should this structure be deployed for circular polarisation in the pattern, some changes are warranted to suppress higher order modes, which otherwise may cause a degradation in the axial ratio.

Since microstrip antennas are inherently narrowband (typically of a few percent) structures, parasitic loading in addition to using thick substrates are indicated as other possible solutions for bandwidth enhancement [4.55-4.57]. While a single patch antenna gives a bandwidth of 3 %, a stacked configuration with cavity backed patch may offer a bandwidth of 15 %. The size of the stacked patches can be changed to dual-frequency (resonance) operation.

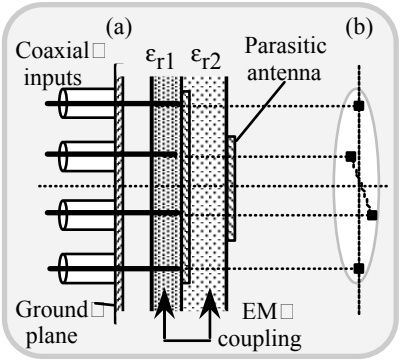


Fig. 4.102 (a) Stacked patch antenna and (b) four-point feeding on the periphery of a circle drawn on the patch

Another way of improving the bandwidth performance of a patch antenna is to make use of stacked patches that are coupled *via* EM fields. Relevant structure is illustrated in Figure 4.102 where the lower (excited) patch and the upper parasitic element are electromagnetically coupled. A stripline circuit attached to the ground-plane enables a four-point excitation of the lower patch in order to get circular polarisation.

4.12.2 Dual-frequency patch antenna

This is a compatible structure whenever receiving and transmitting frequencies are well separated. This technique assigns separate frequency bands to each patch antenna (and therefore, no external duplexer is needed). The relevant structure is illustrated in Figure 4.103.

The microstrip radiator shown in Figure 4.103 is a composite structure that consists of two distinct parts, each with a microstrip patch and a dielectric substrate. Both parts have a common ground-plane through which transmit and receive ports are facilitated. The microstrip structure of each part is designed for a distinct resonant frequency, one for transmit mode and the other for receive mode. Proper choice of patch dimensions, thickness of the substrates and values of dielectric constants (of the substrates used) enable the frequency separation between these modes of operation.

The design principle includes allowing for proper cutoff values of the frequency band supported by each microstrip radiator so as to allow selective bandpass of signals (being transmitted and received without mutual interference).

Careful trimming of the dimensions is required to achieve acceptable performance of the entire unit. Another factor of crucial issue is the VSWR performance of the individual patch units at the frequencies of operation.

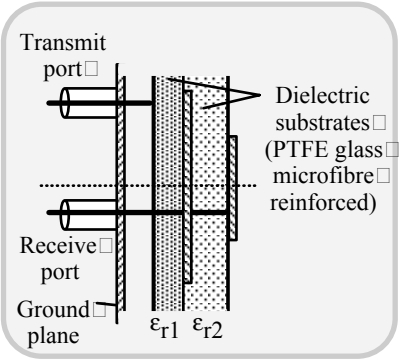


Fig. 4.103 Dual-frequency microstrip patch antenna: Patch dimensions and substrate properties (dimensions and permittivity) are design parameters to achieve the desired frequency separation

4.12.3 Circularly polarised microstrip antenna

Microstrip antennas are primarily linearly polarised radiators. However, circular polarisation can be obtained by exciting two orthogonal modes with a 90° time-phase difference. There are two ways of doing this as indicated below.

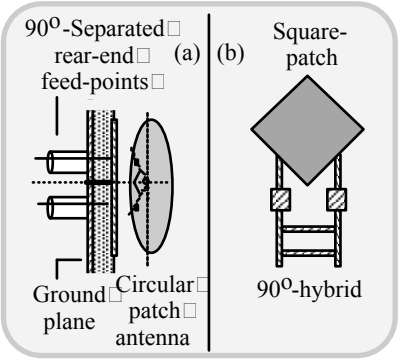


Fig. 4.104 Circularly polarised patch antennas: (a) Circular patch antenna and (b) square patch antenna

The first method, as shown in Figure 4.104(a), uses a centre-grounded circular disk driven from the back by two probes located 90° apart. In the second technique, a square radiator, as illustrated in Figure 4.104(b), can be used with its two adjacent sides being driven in *via* 90° hybrid.

Single-fed circularly polarised (SFCP) patch antennas
These are patch antennas having some perturbation deliberately introduced to the patch geometry and fed by a single source. Such an antenna can simultaneously radiate a pair of orthogonally polarised waves at a frequency that lies between the resonant frequencies.

Typical patch structures are shown in Figure 4.105 where S denotes the patch area, ΔS is the perturbed area and Q_o is the unloaded-Q of the unperturbed patch. A relevant concept has been adopted with a pentagonal patch with complementing feed locations to produce RHCP and LHCP waves [4.58].

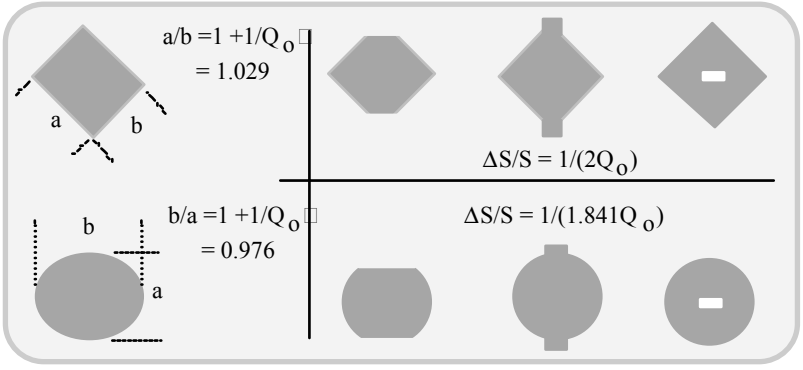


Fig. 4.105 Single-fed circularly-polarised patch antenna with deliberately introduced perturbations. (The perturbations could be notches or projections as shown)

4.12.4 Other versions of microstripline-based antennas of wireless units

There are also certain versions of microstrip-based planar structures that are widely adopted in wireless communication units. Identified below are some typical examples.

PIFA structures
The planar inverted F-antennas (PIFA) mounted on to or built inside a portable mobile telephone can be designed to work at UHF bands yielding vertical and/or horizontal polarisations.

Microstripline-fed slot antennas
A narrow rectangular slot cut in the ground-plane and excited by a microstrip feedline with either a short or an open termination. Such structures yield an impedance bandwidth performance close to 20 %. The shape of the feedline (such as T-shape) can be chosen as a design consideration in realising a required bandwidth.

Patch antenna plus a parasitic director

A typical circular patch antenna (with a parasitic director) compatible for mobile telephones, 800-MHz operation, is illustrated in Figure 4.106. It can offer broadband performance both in vertical and horizontal polarisations. It is a self-resonating structure and has a low profile.

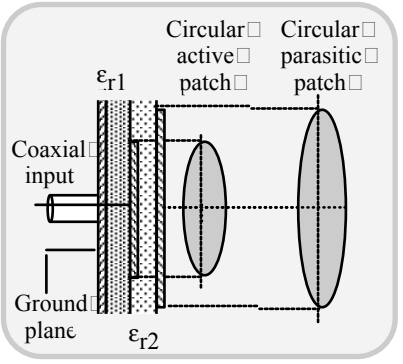


Fig. 4.106 Circular microstrip antenna for 800-MHz mobile phone with a parasitic director

Microstrip radiators used in satellite-based units

For satellite communications, a host of microstrip patches have been developed: A standard square or a circular patch of conformal structure, self-resonant, yielding circular polarisation provisioned by a hybrid, at L- to Ku-band operations, is a typical example.

A modified structure with a thick substrate (of low dielectric constant), again of conformal structure, and enabling higher order mode suppression via balanced feeding technique has been developed for L-band satellite communications applications allowing circular polarisation with the use of a hybrid.

Further, a patch stacked with a parasitic element can be designed for dual resonance and higher order mode suppression (*via* balanced feeding technique). It is useful in L-band satellite communications. It supports circular polarisation realised through the use of a hybrid.

Stacked patch configurations can also be modified for dual frequency self-duplexing operations.

Circularly-polarised patterns produced by orthogonal mode degeneration are compatible for narrowband, L- and Ku band satellite communication applications. Relevant approach uses a singly-fed circularly polarised (SFCP) geometry of patch antenna.

Planar radiators for base station applications
In base station applications, a printed dipole integrated with a feeding circuit and offering a beamwidth (*via* corner reflector geometry principle) of 180° at UHF bands is shown in Figure 4.107.

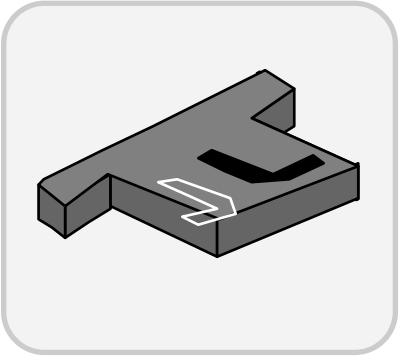


Fig. 4.107 Printed dipole for base station applications: The printed structures lie on the top and bottom of the substrate

4.13 CAVITY-BACKED PATCH ANTENNA

For UHF, microwave and millimeter wave applications, the use of microstrip patch antennas may pose the following inherent disadvantages:

- Poor scan performance in array environment
- Low gain
- Limited polarisation diversity
- Lack of frequency agility
- Narrow impedance bandwidth.

One or more of the limitations listed above can be overcome by using *multi-layer, multi-probe cavity-backed* (MMCB) patch antennas, which can be designed to offer significant flexibility pertinent to structural as well as electromagnetic considerations and improved performance (when compared to a simple patch antenna). A typical configuration of a MMCB patch antenna is illustrated in Figure 4.108. Cavity-backing indicated in Figure 4.108 is used to suppress guided-wave effects leading to enhanced impedance bandwidth and scan performance. The short-posts (or alternatively switching diodes) can be used to provide polarisation diversity, frequency agility and bandwidth enhancement. Additional dielectric layers used enable frequency agility and gain enhancement and facilitate beamforming.
The cavity-backed structures are desirable in phased array applications due to reduced guided-wave effects. Further, in practical situations the antenna structure

can be embedded in the body of a platform, effectively creating a cavity structure around it.

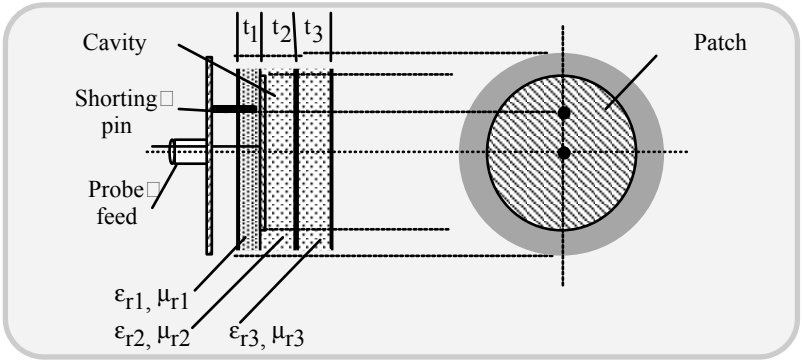


Fig. 4.108 Multi-layer, multi-probe cavity-backed patch antenna with a shorting pin

When used in arrays, the cavity-backed patches offer almost nil scan blindness when compared to conventional patch antenna arrays. With increase in substrate thickness, the performance of MMCB patch antenna array is reported to improve [4.59]. But, cavity-backing may warrant an increase in the cost due to additional waveguide requirements. The MMCB patch antenna, in general, gives more design options for performance optimization *via* shorting pins, cavity design alternatives and the incorporated dielectric layers.

As regard to gain, the traditional patch antennas offer low gain (≈ 3 to 5 dBi) and this gain may not be significantly controllable with substrate thickness and/or by the dielectric constant: However, with additional dielectric-layering and by including a cavity-backing (as in MMCB patch antenna), the gain can be increased. Also, the broadside and sidelobe levels can be decreased in addition to any undesired guided-wave effects that may be present. However, any gain enhancement effort will concurrently reduce the bandwidth performance (to an extent of 1-2%).

Considering beamforming, the desirability is to have a narrow beam with its axis laid along an arbitrary angle. This can be achieved *via* phased arrays, as will be explained in Chapter 5. An inexpensive alternative is to use substrate-loading of the patches. Relevant designs can offer maximum radiation at a specified angle by proper choice of superstrate parameters. Corresponding bandwidth will, however, be constrained.

As mentioned earlier, microstrip antennas in simple form would lead to only linear polarisations. However, if circular polarisation is needed (as in some mobile and indoor wireless communications), the simplest way is to use shorting posts (or switching diodes) as shown in Figure 4.108. By properly placing two or more

posts on the patch, right- and/or left-circular polarisation as well as horizontal and vertical polarisations can be realised. Further, polarisation optimisation can be done in conjunction with other design parameters, such as impedance bandwidth.

Another merit of cavity-backing of patch antennas is as follows: The large size of the patch antenna at UHF may cause concern to designers. Use of high permittivity substrates may reduce the size, but the high dielectric constant of the substrate would lead to surface effects, increased weight and reduction in efficiency.

Alternatively, a superstrate structure plus cavity-backing can be adopted with a nominal dielectric constant (≈ 10). Such a structure allows relatively small-sized configurations that can operate at lower resonance frequencies (corresponding to the lower operating bands). However, these structures may call for a resonance-shift versus impedance-matching design compromise. Also, since, the cavity can suppress the guided waves (excited in the superstrate), no reduction in efficiency occurs.

The shorting posts used accomplish the resonance shifts. The number and locations of these posts are the associated design parameters. In a practical system, a number of switching diodes can be used (*in lieu* of passive shorting posts) and appropriate diodes can be switched on or off so as to tune the antenna to the operating frequency, and conform to a desired frequency agile performance.

Bandwidth-specified impedance matching is normally done with shorting pins with appropriate optimisation of the distances between the pins. Such posts act as reactive elements and control the impedance of the structure. For a 2:1 VSWR performance, the bandwidth can be almost doubled with the posts, from the value obtained without the posts.

4.13.1 Loaded and cavity-backed small patch antennas

Shorted patch antennas, as indicated above, can overcome the limitations of antenna real estate in mobile communication systems posed at low-frequency microwave regimes. Further, at such frequencies, the conventionally-sized patches and even a quarter-wave patch would occupy considerable surface area.

In order to alleviate this size problem, use of different patch geometries plus probe- or aperture-coupled feed arrangements have been developed, based on the "shorted-patch" concept. That is, a shorting pin is located near the feed of a size-reduced patch posing an intense capacitive coupling or loading. This reactive loading counteracts any inductive-loading caused by patch size reduction [4.60-4.62].

With reduced size and through the use of shorted pins, the small patch radiators conceived, however, may not yield a bandwidth of 10%. A method of augmenting the bandwidth is to load the patch resistively. This would, however, cause a reduction in the gain. An alternate approach is to back the loaded patch with a cavity. Relevant configuration is illustrated in Figure 4.109.

In Figure 4.109(a), an impedance-loaded probe feeds a circular patch. In Figure 4.109(b), a loaded-cavity backed probe-fed microstrip patch is illustrated. The patch is etched on the substrate and resides within a cavity of larger radius. A

conventional shorted pin patch antenna refers to Figure 4.109(a) with $Z_L = 0$. A qualitative comparison of the performance characteristics of these structures is presented in Table 4.4.

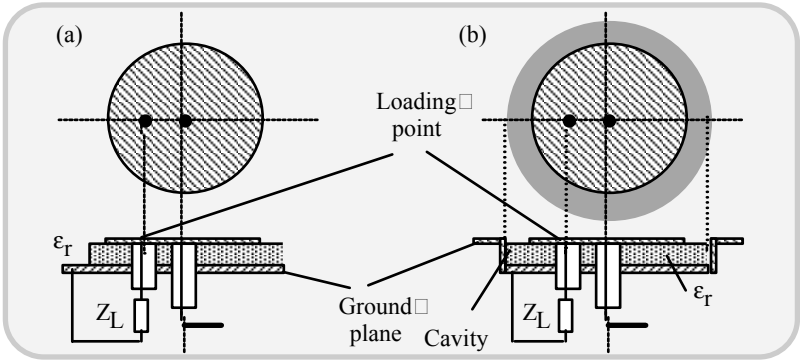


Fig. 4.109 Schematics of: (a) An impedance-loaded patch; and (b) a loaded cavity-backed patch

Table 4.4 Relative performance characteristics of shorted-, loaded-, and cavity-backed patch antennas

Characteristics and features	Shorted-patch antenna	Loaded-patch antenna	Cavity-backed patch antenna	
			Loaded	Unloaded
Gain	Highest	Lower (reduced by 2 dB)	Lower	Higher
Bandwidth for a given return loss	Lower	Lower	Highest	Lower
Spacing between the probe feed and shorting or loading post	Larger	Largest	Large	Least

4.14 MULTIFUNCTIONAL PATCH/PLANAR ANTENNAS

Multifunctional mobile antennas are aimed at realising multiple functional characteristics with a set of antenna elements grouped into a compact aesthetic structure. The multifunction in question, in general, may refer to the following:

- Efficient radiation characteristics (at least at two frequencies), with a single-feed arrangement. For example, such a dual-band antenna can be used for GSM and DCS applications. This structure needs a duplexer for band segregation
- Dual-band antenna using two feeds
- Multipolarisation antenna structure.

The aforesaid multifunctional attributes can be realised only by a set of antenna elements. Typical radiating elements, which can be the parts of such a set, are as follows:

- The classical patch antenna working on the fundamental TM_{01} or higher order modes. They can radiate EM energy with linear or circular polarisation. A high permittivity substrate can be used to reduce the dimensions of the element
- For monopolar radiation, a classical wire monopole can be used. It can be replaced by a "wire patch" antenna, which refers to a small patch antenna whose radiation characteristics are exactly the same as the monopole antenna
- A coplanar version of the "wire patch" antenna can also be used. Associated with a ground-plane, it behaves as a classical patch antenna, but it has smaller dimensions even with air substrate.

The concept of multifunctional antennas leads to *shared antenna systems*. Such systems are becoming increasingly popular in indoor and outdoor wireless communication applications. They imply potential cost saving and reduced environmental impact. Sharing antennas opens new site location opportunities where congested rooftops prevent the addition of new antennas. Finding room for more antennas will become critical for the new 3G networks and future technologies.

A shared antenna system is one that multiple operators can combine their base stations together and apply the combined signal to either multiband or some other combination of dual band units.

Typical example of radiating structures that represent the multifunctional antennas are described in the following sections.

4.15 GPS-DCS ANTENNAS

The GPS function at 1.57 GHz has been realised with a classical patch-radiating element perturbed at opposite corners, so as to provide circular polarisation [4.63]. The substrate used is a dielectric with $\varepsilon_r \approx 2.5$. The DCS function at 1.8 GHz is

obtained with a wire patch element. The multifunction antenna performance is obtained with the two elements superimposed as shown in Figure 4.110.

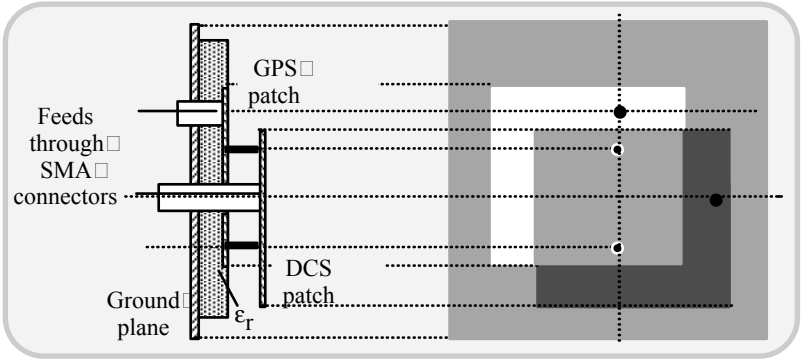


Fig. 4.110 Dual application (GPS-DCS) antenna with a common substrate geometry

The GPS elements are modified suitably so as to be associated with the DCS element. The probe feeding of the DCS element is placed at the centre of the multifunction antenna (since at this locale, the GPS element has a minimum E-field).

GPS, GSM, and FM multifunction antenna

The objective here is to insert in the socket of a classical FM monopole antenna the GPS and GSM functions.

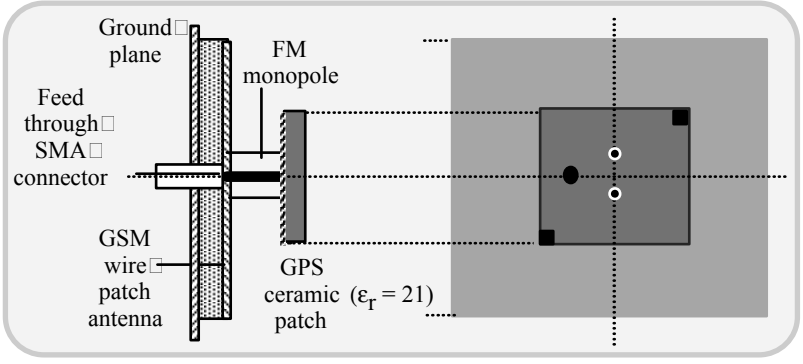


Fig. 4.111 Multifunction antenna to support GPS, GSM and FM communications

Small antenna dimensions are feasible for the GPS application, using a commercialised element with a high permittivity ceramic substrate with ε_r in the

order of 20. The GSM function at 900 MHz is realised by means of a wire patch element constituting a compact antenna structure. The GPS element placed on the top of the GSM element is fed through the "wire" of the wire patch element (Figure 4.111).

GSM-DCS antenna with single feed

Radio communication dual-band (GSM/DCS) antenna for use in automobiles can be made to yield desirable radiation characteristics and small dimensions if the GSM element is placed over a DCS element, as illustrated in Figure 4.112.

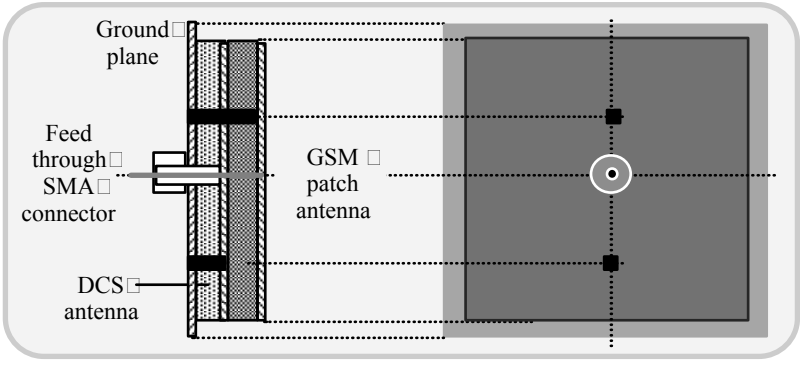


Fig. 4.112 Dual application (GSM-DCS) antenna with a common substrate geometry

Placing the GSM element in proximity to the DCS patch may perturb the radiation pattern required for the DCS function. But in practical designs, it can be restricted so as to remain within a tolerable extent. The GPS element that uses a ceramic element or coplanar wire patch can also be added to this antenna.

GSM-DCS antenna with a dual feed

Another antenna structure that can be used in automobiles comprises of a DCS element placed on the top of a GSM element, as illustrated in Figure 4.113. In comparison with the previous GSM-DCS antenna (of Figure 4.112), the roof of the GSM structure is widened so as to obtain a satisfying bandwidth of the DCS element.

Dual-band element antennas for satellite localisation and aircraft landing aid

For satellite localisations (GPS/USA-1572-1578 MHz; GLONASS/Russian-1591-1627 MHz) and for aircraft landing (MLS) aids, the multifunctional patch antenna elements have been designed.

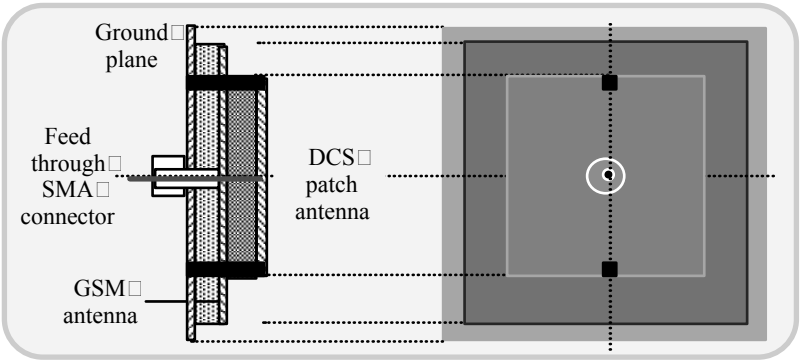


Fig. 4.113 GSM-DCS antenna with dual feed with a common substrate geometry

The radiating structure in Figure 4.111 consists of two patches superimposed (for widening the bandwidth) on a selected substrate (with $\varepsilon_r = 3.2$ and $\tan(\delta) = 1.5 \times 10^{-3}$). The circular polarisation is realised by using two coaxial probes feeding the element through a (– 90°)-coupler. A ground wire located at the centre of the element avoids electrostatic discharge and at the same time, it does not perturb the characteristics of the element. For microwave landing system (MLS) applications at 4.97 - 5.33 GHz, a monopolar element radiation pattern is needed. For this purpose, a circular patch on the dielectric substrate working on the monopolar TM_{10}-mode can be used. Consistent with the frequency of operation, the dimensions of the antenna are of the same order as those of the GPS-GLONASS element.

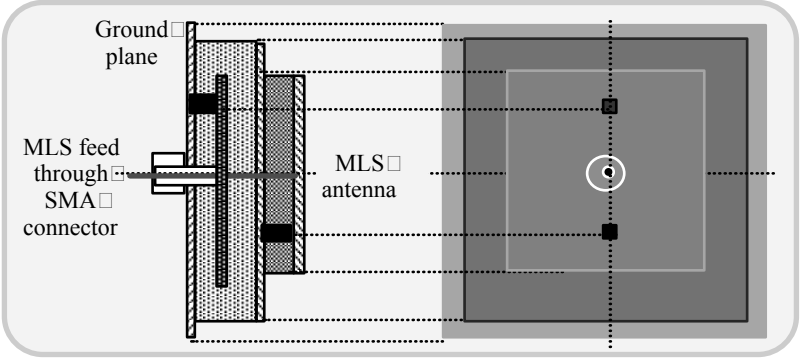


Fig. 4.114 Antenna element for microwave landing system (MLS) applications

4.16 PRINTED ANTENNAS

Compact wireless antennas with multipolarisation and multiband operational features are conceivable with multiport printed planar geometry. Such configurations are based on *strip-slot foam inverted patch* (SSFIP) antenna described in [4.64].

The antenna in question has two layers: A lower layer based on the SSFIP concept and an upper layer, which is a traditional patch antenna, such as a PIFA. The lower structure forms the ground-plane for the upper geometry and a single-port coupling through a single-slot (or through a more complex two-port system) excites it.

The upper layer can be designed with elements (up to 4), each supporting different polarisation and/or frequency. All antenna ports are configured to have a common metallic ground. An example of this antenna is illustrated in Figure 4.115. It refers to a dual-port dual-frequency and single polarisation (per frequency) antenna. The supporting substrate (# 1) of the SSFIP antenna could be free-space (minimising loss and cost). The ground-plane of the upper substrate, on which the upper antenna (a simple patch) is etched, forms the radiating patch of the SSFIP. The upper substrate (substrate #2) is supported by a pair of posts made of semi-rigid coaxial cables adequately placed in order to unperturb the proper function of the SSFIP. Only one of these posts is "active", and forms a transition between a 50-ohm line etched on the lower substrate and the patch-feeding line etched on the upper substrate. The other post forms only a mechanical support and for symmetry reasons. This antenna offers orthogonal polarisations.

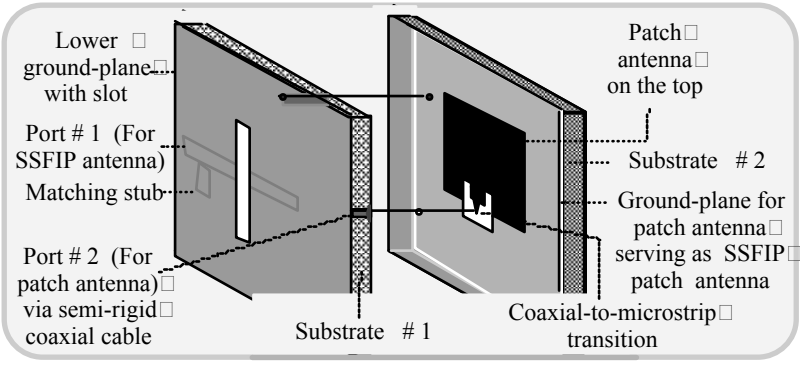


Fig. 4.115 Dual-port, dual-frequency, single polarization (per each frequency) antenna with a patch on top

There are two possible bands of operation feasible with this antenna. For example, band 1 ranging form 1.71 to 1.99 GHz (for DECT plus DCS applications) and band 2 ranging from 2.4 to 2.483 GHz (for WLAN/Bluetooth™ applications); other combinations such as 1.8 GHz/5.7 GHz can also be attempted. The lower SSFIP structure supports the lower band (band 1) and the upper band is realized

through the top/upper layer. The reason is that there is only a limited space on the top of SSFIP so as to accommodate a smaller antenna (at higher frequency band).

The disadvantage of this system is its limited bandwidth (1.5%) of the upper part, stemming from the inherent narrow bandwidth aspects of classical patch antennas. Replacing the patch antenna by a PIFA can overcome this drawback.

An extension of the antenna in question refers to a dual-frequency radiator with polarisation diversity specified for each frequency. This structure depicts a four-port, dual-frequency antenna that enables dual-polarisation per each frequency of excitation.

The relevant concept can be implemented with appropriate patch geometries of the SSFIP. For example, a regular polygonal (such as an octagonal) patch can be mounted on four pods. It gives more space for the upper structure. The top structure could be made of square or some polygonal pair of patches. Both such patches have crossed polarisations (90° between them). Their polarisations are set at 45° rotated relative to SSFIP polarisations. That is, the polarisations of the SSFIP part of the antenna are 0° and 90° respectively with the corresponding polarisations of the upper patch antenna being 45° and 135° [4.65].

4.17 APERTURE-COUPLED PATCH ANTENNAS

Low-profile, dual-circular polarised, aperture-coupled patch antennas can be designed with high port decoupling between the two circular polarisation modes. Such antennas have low cross-polarisation levels (less than – 20dB) and have been adopted for *radio frequency identification* (RFID) systems (intended for identifying remote objects). (For example, they can be used in interrogation tags mounted on objects in automated manufacturing processes or on an automobile passing through a toll-booth on a high-way.) Such operations are done at the ISM band. Relevant modulation on the downlink path (enabling writing data on a tag) and the up-link (for reading data from the tag) corresponds to circular polarisation switching from LHCP to RHCP. The associated antennas at the reader and at the tag sensor should facilitate LHCP/RHCP segregation with adequate cross polarisation isolation (≈ 20 dB).

Typically, such a radiator can be designed with a composite structure made of a lower feed-layer, an upper patch-layer using a single microstrip line that excites two orthogonal coupling slots, with the length of the line between the two slots being $\lambda/4$ (Figure 4.116). Depending on which port is fed, the antenna will produce either an RHCP or an LHCP radiation.

In the structure of Figure 4.116, $\lambda/4$-lines are used (instead of a branch-line coupler, which is commonly adopted in conventional circularly polarised patch-antennas). This structure allows enough space for the RF front-end electronics and circumvents the *decoupling problem*, which can be summarised as follows: There are two main sources of unwanted coupling in the system under consideration. The first is the coupling between the two orthogonal linear modes. If these modes are not excited symmetrically (slot shifted off the centre of the patch) additional modes are excited, which would lead to coupling at the resonance frequency. The second source of coupling can arise from the branch-line coupler, if adopted. In

such a case, signals, which are fed on one input port and reflected by the antenna, return to the other input port. This means that the port decoupling is limited by the impedance match of the antenna and not by the inherent decoupling of the branch-line coupler (which is usually around 30 dB). Hence, when using a branch-line coupler, the requirement on the input reflection of the antenna, for each linear mode, is more severe than the usual – 10 dB. (For desirable performance, at least – 20 dB over the entire bandwidth of operation are required.) One way to avoid this problem is to separate the feed network for each circular polarisation.

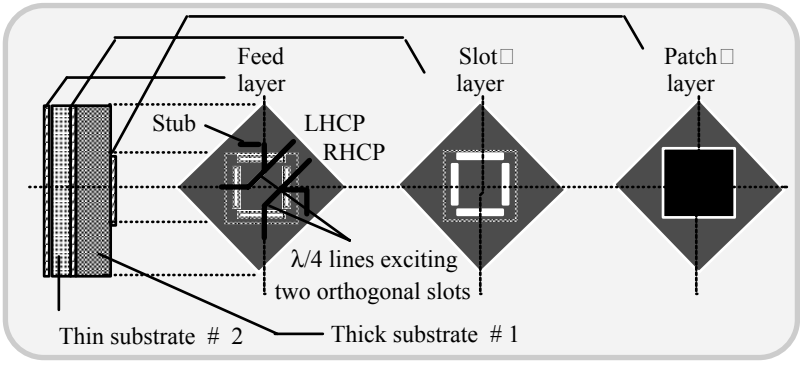


Fig. 4.116 Aperture-coupled patch antenna with separate feeds to yield dual-polarised radiation

The slot aperture in Figure 4.116 consists of two pairs of orthogonal slots. Feeding two orthogonal slots by a λ/4-line, as shown, generates two orthogonal modes with 90^{o} phase difference. Hence the antenna produces both circular polarisations depending on which port is fed.

Another version of aperture-coupled patch antenna refers to a crossed-slot antenna depicted in Figure 4.117. The antenna shown in Figure 4.117 uses a symmetric excitation of the two orthogonal linear modes. This gives a better decoupling compared to the antenna depicted in Figure 4.116. The slots are dogbone-shaped, which make them shorter than a rectangular slot. The signal in each branch is split into equal parts using a reactive power divider. Use of a dielectric with higher permittivity leads to a smaller feed network, but it warrants longer coupling slots. This is critical because, in order to avoid unwanted coupling, a sufficient distance should be kept between the slots on one hand, and between the slots (as well as ensuring the feeding line passing necessarily between them), on the other hand. The two branches of the feed network are connected to a branch-line coupler in order to produce RHCP and LHCP, respectively. The port decoupling is equal to the reflection of the outgoing signals of the branch-line coupler and hence depends on the impedance match of the two excited linear patch modes. The decoupling of the branch-line coupler itself could be considered negligible.

The drawback of the above system is the space-consuming feed network. To overcome this shortcoming, a scheme is presented in [4.66]. It refers to an etched bar antenna as illustrated in Figure 4.118.

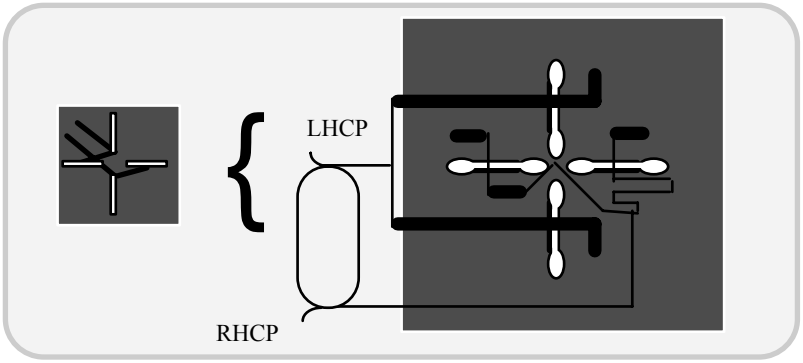


Fig. 4.117 Double-crossed slot-excited patch-antenna with two symmetrically-excited linear modes

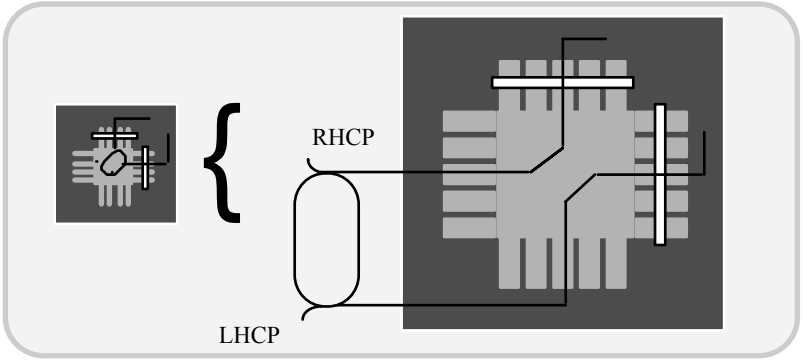


Fig. 4.118 Etched bar antenna: Aperture coupled version with suppressed orthogonal modes

The structure shown in Figure 4.118 is designed to suppress coupling between the two orthogonal linear modes (without using symmetrical coupling slots). With the shape of the patch modified to be cross-shaped, it is then excited at the end of its arms. Small slots are etched in the arms in order to force the patch currents to flow strictly in one direction and hence to suppress the excitation of other than linear modes. Using thicker foam leads to a broader bandwidth but also to longer coupling slots, which would overlap each other. The use of a high permittivity ($\varepsilon_r \approx$ 10) dielectric for the feed circuit reduces the size of the branch-line coupler

enough to place it in the centre of the patch allowing a small and compact feed network.

A simple dual polarisation capability with patch antennas can be realised with two orthogonal non-overlapping slots used to feed a square patch element. However, the normal problems encountered in the relevant designs are as follows:

- Size of the slots constrained by symmetry requirements
- Isolation limitations
- Polarisation purity

Consistent with the objective of dual polarisation implementation discussed earlier, the functional aspect of the relevant design is to realise a decorrelated pair of signals obtained by means of an antenna placed at a single location through two distinct (usually orthogonal) polarisations. The underlying consideration is as follows: When an EM field is incident upon an object, the polarisation of the signal may be altered to some degree. In such events, the receiver can be rotated through 360^o in the vertical plane without any change in signal strength perceived as a result of such depolarising effects.

Decorrelation in dual-polarised antennas is directly related to the performance of the antenna. Optimum decorrelation occurs when the two polarisations are orthogonal (perpendicular or crossed) and the antenna features high isolation between the two ports. Dual-polarised antennas with slant (+ 45^o) and slant (– 45^o)-polarisation, give nearly identical radiation patterns. Since an antenna with a low-profile, small-size and light-weight is preferred, the radiator under discussion is desirable.

A cross-slot structure of a patch antenna that can enable dual polarised radiation is illustrated in Figure 4. 119.

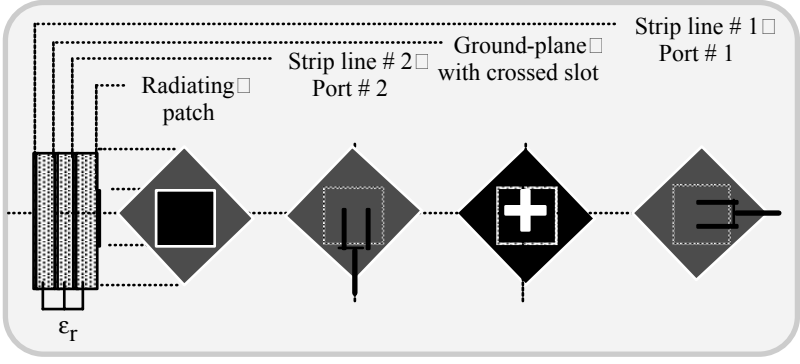


Fig. 4.119 Patch antenna with cross-slot excitation for dual-polarised radiation pattern

Planar inverted F-antenna (PIFA)

In low-profile small antenna realisations, the PIFA enjoys popularity due to its simplicity and compatibility for use in small portable units [4.67]. As indicated in

Chapter 3, the PIFA typically consists of a rectangular planar element, the ground-plane, and a short-circuit plate (of narrower width than that of a shortened side of the planar element). The PIFA is regarded as a linear inverted F-antenna (IFA) with the wire radiator replaced by a plate so as to expand the bandwidth. IFA itself can be regarded as a shunt-driven inverted L-antenna/transmission line with an open-end.

In the perspectives of microstrip line, a PIFA can also viewed as a *short-circuit rectangular microstrip antenna* (SC-RMSA).

The SC-RMSA is a simple rectangular element reasonable to support TM_{100} dominant mode. The length of the rectangular element can be halved by means of a short-circuit plate between the radiator element and the ground-plane at the position where the E-field of the TM_{100} mode is zero. By setting the width of the short-circuit plate narrower than that of the planar element, the effective inductance of the antenna element increases and the resonant frequency becomes lower than that of the structure having the short-circuit plate of same width as the planar element.

As a result, the length of the SC-RMSA can be further reduced. Hence, the ratio of the width of the planar and short-circuit elements decide the design and performance of PIFA.

Thus PIFA, in essence, is an antenna configured on the basis of a transmission line F-antenna that uses a wire element (in lieu of the planar elements of PIFA). A simple PIFA structure is depicted in Figure 4.120.

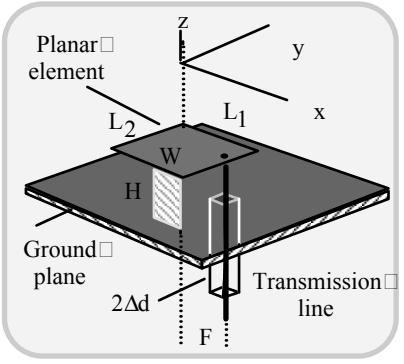


Fig. 4.120 A simple PIFA structure: Geometry and feeding arrangement

4.17.1 PIFA design considerations

Considering the PIFA structure shown in Figure 4.121 the ratio W/L_1 should be set equal to 1 and $L_2 + H$ should be $\lambda/4$ for resonance; further, the condition that $H << \lambda$ would allow negligible fringing of the field in the open-circuited edge of the planar element. The resonant frequency f_o can be expressed as follows:

$$f_o = rf_1 + (1- r)f_2, \qquad \text{for } L_1/L_2 \leq 1$$

$$= r^{\kappa} f_1 + (1-r^{\kappa}) f_2, \text{ for } L_1/L_2 > 1 \quad (4.125)$$

where $r = W/L_1$ and $\kappa = L_1/L_2$; further, f_1 corresponds to wavelength $\lambda_1 = 4 \times (L_2 + H)$ and f_2 refers to λ_2 equal to $4 \times (L_2 + L_1 + H - W)$.

The bandwidth performance of the PIFA is decided by the requirement that VSWR < 2. It increases with an increase in antenna height H and the planar size ratio L_1/L_2. VSWR close to 1.5 is feasible with $W < L_1$. To obtain broadband PIFA, instead of increasing the antenna height, adding a parasitic element and developing a dual resonance could be more effective [4.68].

Problem 4.4

Shown in Figure 4.121 is a typical PIFA mounted on a portable unit with the dimensions as indicated. The antenna is designed for 900 MHz operation. Justify whether the dimensions of the antenna shown would conform to the resonance and bandwidth considerations discussed above.

Assume the following for the PIFA dimensions: $D = 0.09\lambda$, $W = 0.10$ to 0.20λ; $H \approx 0.03\lambda$ and $L_1 = L_2 \approx 0.40$ to 0.50λ (where D, W, H, L_1 and L_2 are indicated in Figure 4.120).

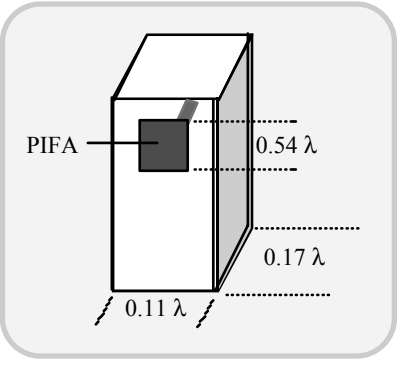


Fig. 4.121 PIFA structure mounted on a portable unit: Approximate dimensions

(*Hint*: See [4.68])

4.17.2 PIFA antenna configurations

Wearable computer systems and miniature wireless portable units warrant almost "featherweight" antennas. Such antennas can be realized as PCB microstrip units. Further, in certain applications, the antenna is required to resonate at two frequencies.

Such dual-band operation on PIFA can be enabled as follows: A U-shaped slot can be etched on a standard PIFA as shown in Figure 4.122. The overall physical length and width ($\ell_1 \times w_1$) specifies the lower frequency operation which can be specified by the formula $f_{lo} = c/[4 \times (w_1 + \ell_1)]$, where $c = 3 \times 10^8$ m/s. The upper resonant frequency can be determined by the formula $f_{ho} = c/[4 \times (w_2 + \ell_2)]$, where w_2 and l_2 are inner dimensions in the U-slot. Another dual-band PIFA can be obtained by stacking two PIFAs of different dimensions with a common shorting pin and appropriate feeds [4.69].
Examples of multiband operations for which PIFA architecture could be conveniently adopted are: (0.9 GHz/GSM, 1.89 GHz/DECT and 2.44 GHz WLAN) or (2.44 GHz WLAN/5.7 GHz WLAN). A triple band PIFA is described in [4.70].

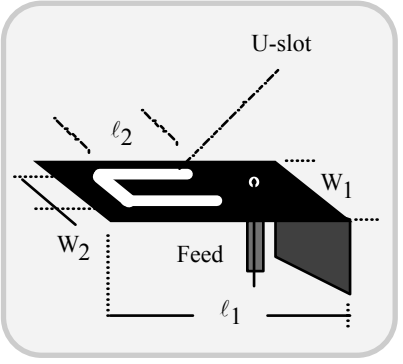


Fig. 4.122 Etched bar antenna: PIFA configuration designed using a PCB

A PIFA with a slot and a capacitive load is introduced in [4.71]. It has a bandwidth of 7% at 940 MHz.

Another dual-band PIFA structure suggested by Viratelle and Langley [4.72] consists of basic inverted-F geometry modified by separating a part of the upper/radiator plate to create a second resonance. A single coaxial feed is used to connect to the inner patch and it enables excitation at both frequency bands. Shorting pins are replaced by strip conductors. Overall VSWR ≤ 1.5 over each band and the azimuthal patterns are near omnidirectional. The PIFA of [4.72] is illustrated in Figure. 4.123. Yet, another interesting modified PIFA has been described in [4.73]. It is designed to offer a bandwidth much higher (10 times more) than the conventional PIFA. It is also very compact with antenna length $\ell < \lambda/8$ and antenna height, $h < 0.01\lambda$. The reduction in length is made feasible by adopting a meandering radiating patch; and, the enhanced bandwidth is obtained by using a chip resistor load replacing the conventional shorting post so that an antenna-height less than 0.01λ is realized.

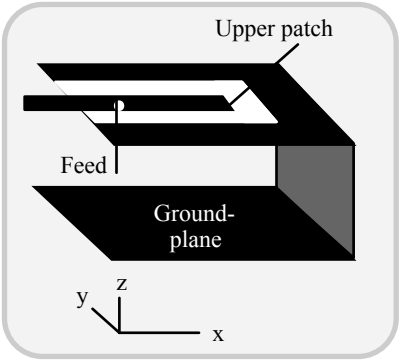


Fig. 4.123 A dual frequency PIFA structure

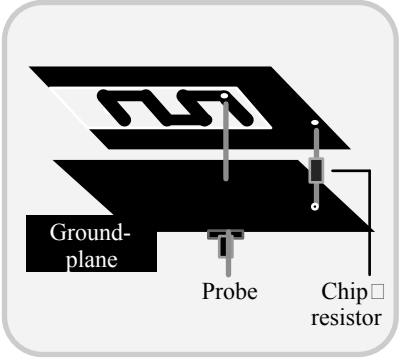


Fig. 4.124 Geometry of meander line printed an inverted F-antenna

Cutting several narrow slits in the rectangular patch, as shown in Figure 4.124, makes the meandering patch. The meandering section increases the electrical length of the patch and resonates at a frequency far below the antenna design frequency. The reduction in patch length by this method corresponds to the reduction done by using capacitive loading [4.74]. The performance aspects of a relevant test antenna of this type as reported in [4.74] can be summarised as follows:

Bandwidth: Use of a chip resistor of value 0 to 5.8 ohms allows a bandwidth range of 0.6 to 11.2 % at 800 MHz operation (The simple PIFA offers a bandwidth of 0.9 %)

Radiation patterns: Measured E-plane and H-plane patterns of the test antenna depict broadside radiation with variation in the E-plane pattern being smoother than that of H-Plane pattern (which has a loaded pattern at $\theta = \pm 90°$ close to ground-plane)

Gain: There is a reduction in gain (~ 6 dB) due to losses in the chip resistor loading.

4.18 ACTIVE PATCH ANTENNAS

Patch antenna as a part of a detection circuit

On wireless radio receivers, zero-IF or direct-conversion demodulation technique has been adopted as an alternative to traditional superheterodyne reception. This technique eliminates RF preselection and bandpass filter stages, thus reducing circuit complexity. The concept of zero-IF detection and a method of introducing a patch antenna configuration as a part of the detection circuit are illustrated in Figure 4.125 [4.75].

The zero-IF detection is based on heterodyning the modulated signal and the unmodulated carrier to produce a zero-IF signal, which contains the amplitude and phase information of the baseband signal. In Figure 4.125(a), the quadrature modulated signal is mixed with the in-phase (I), $[\cos(\omega_c t)]$-component and the quadrature-phase (Q) , $[\sin(\omega_c t)]$-component of the carrier, so that these (in-phase and quadrature-phase) components of the baseband are obtained as outputs. Suitable low-pass filtering eliminates unwanted and other harmonics.

Implementation of the above principle in conjunction with necessary antennas can be done by means of the architecture shown in Figure 4.125(b). There are two patch antennas I and Q, each is rendered active with biased Schottky barrier diodes. The signal power is divided by means of a Wilkinson power divider and Q, I patches are energised with the required phase shift. The inductive elements shown depict the RF chokes and low-pass filters. The capacitive elements depict high-pass filters.

The patch antennas are separated by $\lambda/2$ and are designed to resonate at the design frequency. For broadside incidence of the wave, the RF signal received by the two patches will differ by 90° satisfying the (I, Q) requirement. However, when the wave is offset from the broadside of the active antenna, the phase shifter on one of the LO paths has to be tuned to realise a close (I, Q) baseband components.

The above configuration permits low power consumption, simple circuitry and a good image signal immunity.

Patch antenna mounted with an active device

Another antenna design that can prove to be useful in wireless communication techniques (such as for GSM applications) is the radiating system that combines the classical patch antenna and an active feed system. The purpose of this design is to achieve high efficiency [4.76]. Conventionally high efficiency is obtained *via* synthesising optimum output load at each harmonic by means of an appropriate

matching structure. Alternatively, an active antenna can be used with a feed arrangement directly linked to the transistor as illustrated in Figure. 4.126.

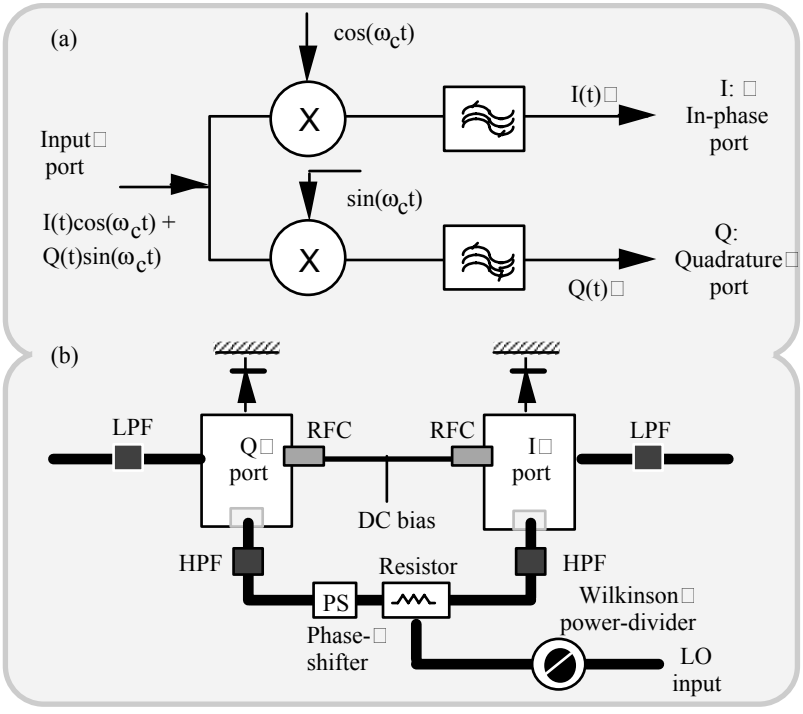


Fig. 4.125 Zero-IF detection active antenna system: (a) Direct-conversion concept and (b) implementation of active antenna

The radiator consists of a high efficiency-active path made of the patch (square, circular, ring) fed directly from the drain (in case of FET, GASFET or from the collector in case of BJT) which works in a high efficiency mode. (Mode E or C-E in BJT or mode F in unipolar devices.) The design of this system accounts for:

- Load-pull concept relevant to having desired load impedances at the fundamental and harmonic frequencies. Such load-impedance selection is dependent on the location of the feed-point on the patch
- Selection of feed-point location specified by the selected transistor

For example, the required load at the fundamental frequency for E-mode operation (ideal switching) is specified by $Z_L = R(1 + jb_c)$, where $R = (k_c/P)[V_{cc} - V_{ce(sat)}]^2$ with b_c and k_c being circuit-specified constants and P is the desired (allowed) output power level; further, V_{cc} is the supply voltage and $V_{ce(sat)}$ is the saturation voltage of the transistor.

Typically, the input impedance of a patch antenna at the harmonics of fundamental frequency is inductive and it can be ascertained *via* the equivalent circuit of the patch antenna shown (when excited by a coaxial feed in Figure 4.127). The details on the concept-design active patch antennas are presented in the following subsection.

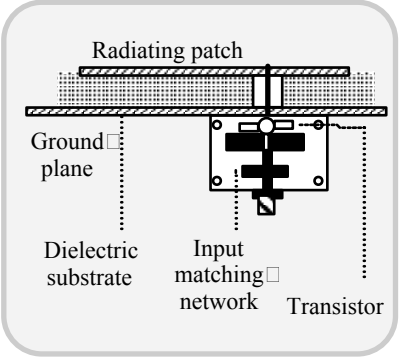


Fig. 4.126 An active patch antenna with a transistor installed as a part of the radiating unit

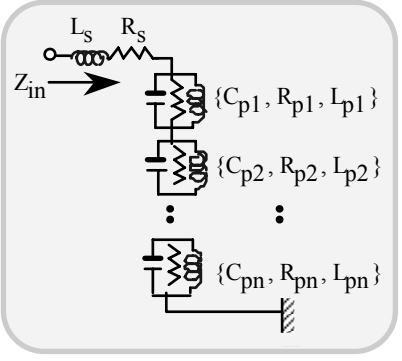


Fig. 4.127 Integrated circuit model of the input impedance of the patch antenna at the fundamental and harmonic frequency. (Subscript indices on p, namely, 0, 1, 2, …, n depict the fundamental and the harmonics) respectively)

4.18.1 High-efficiency amplifiers for integration with wireless antennas

As indicated above the architectures for high-efficiency amplifiers, compatible for integration with wireless antennas rely on harmonic tuning. These are essentially power amplifiers, which facilitate the reduction of unwanted harmonic radiation from the transmitter front-end. The relevant design can be based on one of the following three methods:

- Using an active integrated antenna approach to perform harmonic tuning
- Using a nontraditional periodic microstrip filter
- Using a technique that combines the above two approaches.

The methods enumerated above are aimed at realising a low-cost compact front-end for advanced wireless communication systems without sacrificing system performance. Crucial considerations in the design are efficiency, size, bandwidth and avoidance of EMI between transmitter subsystems. The intended integration refers to stages from the signal input, through (power) amplifiers, to the antenna. Conventionally such a path is designed in discrete subsystems matched to a 50-ohm environment as illustrated in Figure 4.128.

The limitation of the traditional system (shown in Figure 4.129) is that any filter placed at the interconnect stage prior to the antenna to suppress harmonics may prone to be lossy, or may couple with other elements. To overcome this, an alternative approach is indicated in Figure 4.129.

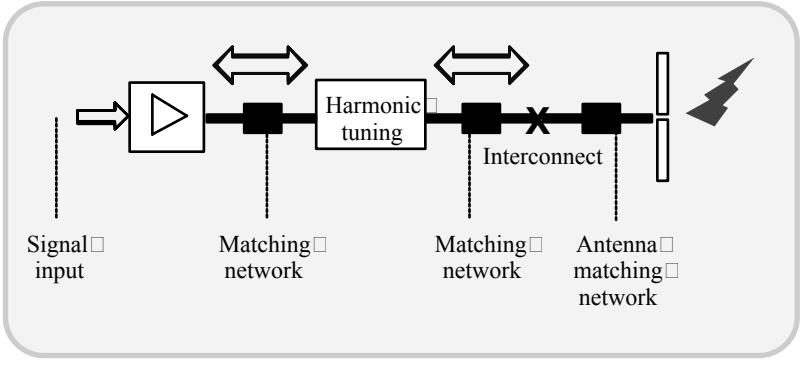


Fig. 4.128 Traditional transmitter front-end with necessary matching units

The active device used in the system of Figure 4.129 corresponds to a power FET or a HEMT posing low output impedance (≈ 100 ohm). This incompatible value of impedance for a 50-ohm system requires design changes (at the expense of power). Another consideration is to increase the efficiency. This is accomplished via harmonic tuning terminating at least second and third order harmonics. Briefly indicated below are the three possible approaches mentioned earlier [4.76].

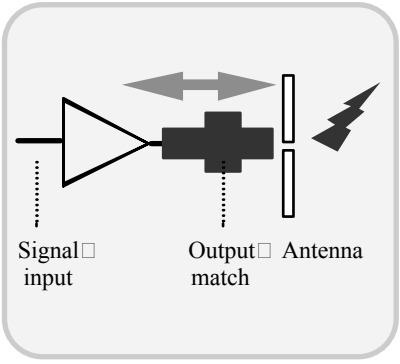


Fig. 4.129 Active integrated antenna transmitter front-end: Active unit is matched to the antenna for maximum power transfer

4.18.2 Active integrated antenna approach

For an antenna whose radiation characteristics are optimally chosen in reference to the intended applications in wireless systems, the design of active integrated antenna be such that the input impedance at an operating frequency and over the harmonics be such that the amplifier performance is maximized. Such input characteristics are dictated by the mode distribution of the patch antenna used. For the impedance constrain under discussion, suggested in [4.76] is the method of modifying the patch surface (with shorting pins) so that the resulting mode profiles optimally facilitate the desired input impedance characteristics.

Any changes on the patch should not, however, modify the radiation characteristics of the patch.

Methods alternative to shorting pins on the patch are notches/slot made on the patch geometry and resorting to appropriate patch geometry (with modifications) such as circular, ring etc.

4.18.3 Periodic-structure approach

The technique indicated above is basically narrowband. For broadband applications, a set of periodic structures (such as wireless etched on the ground-plane of the patch antenna) can be used to tune out the harmonics.

4.18.4 Combined approach

The weakness of the periodic-structure method is that it can be used to tune only even or odd harmonics, and, the active integrated antenna approach lacks in enabling a design that offers proper termination for both even and odd harmonics. Hence, both techniques can be combined to mutually cancel the limitations involved. One possible approach is to use a patch antenna with shorting pins integrated with a periodic structure. The antenna is designed to tune the second harmonic of a power amplifier while the periodic structure can be designed to tune out the third harmonics. A relevant structure is illustrated in Figure 4.130.

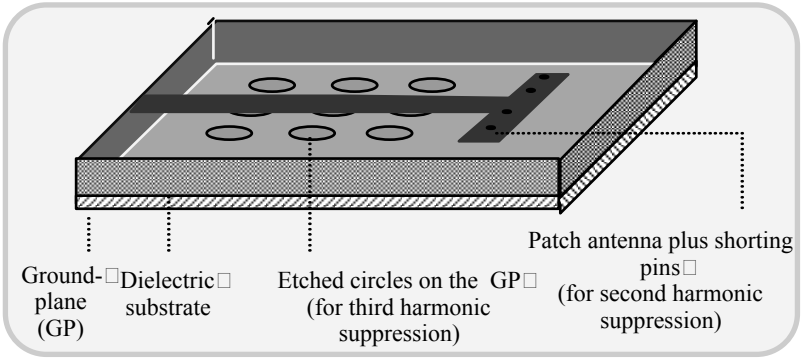


Fig. 4.130 Active patch antenna with harmonic suppression techniques

Cavity-backed printed antenna with resistive loading

For integration into a mobile handset terminal, a small printed antenna can be made with a cavity-backed, probe-fed patch incorporating a strategically-positioned resistive load. The relevant structure is illustrated in Figure 4.131.

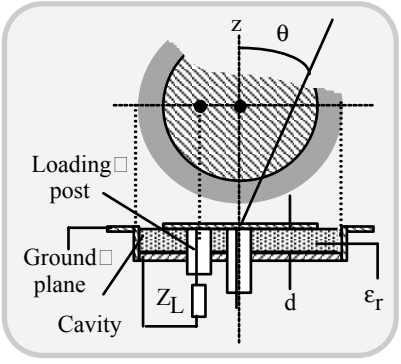


Fig. 4.131 Cavity-backed printed antenna with resistive loading

Cavity-backed patch antennas are integrated within the mechanical support and their performance is similar to that of a simple patch antenna as long as the cavity is large and the walls are electrically far from the edges of the patch. However, when the dimensions are reduced for mobile handset applications, the load (resistive) should be strategically placed to achieve a desired impedance performance as illustrated. This antenna can yield high-Q performance with near-omnidirectional pattern. The impedance bandwidth is governed by loading the cavity by adding a resistor. The radiation patterns are presented in Figure 4.132.

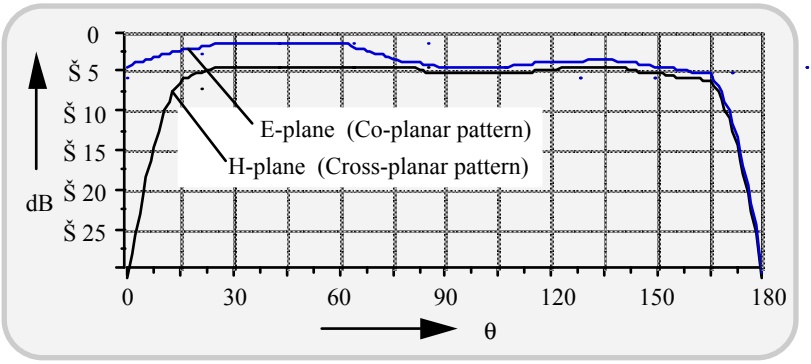


Fig. 4.132 Radiation pattern of cavity-backed, loaded patch antenna

4.19 DIELECTRIC RESONATOR ANTENNAS

A dielectric resonator of compact size can be used to achieve an omnidirectional radiation pattern with the capability for operation over dual bands such as PCS (1750 MHz –1870 MHz) and IMT-2000 (1920MHz – 2170 MHz) bands.

The wideband objective is made feasible by using a dielectric material for the resonator antenna that has a low dielectric constant. A feasible antenna structure is depicted in Figure 4.133. It follows the test antenna described in [4.77].

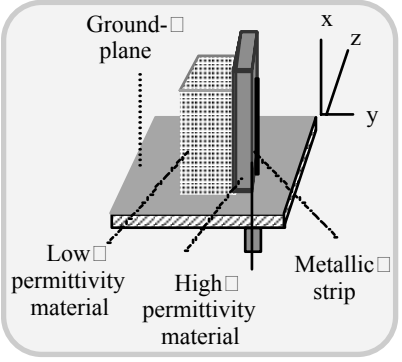


Fig. 4.133 A dielectric resonator antenna made of composite (low and high dielectric constant) materials

The antenna is made of a rectangular resonator of relatively low dielectric constant (such as Al_2O_3 with $\varepsilon_r = 9.3$), a thin high permittivity material (such as $BaTi_9O_{13}$ with $\varepsilon_r = 38$) slab and a conducting strip as shown in Figure 4.133.

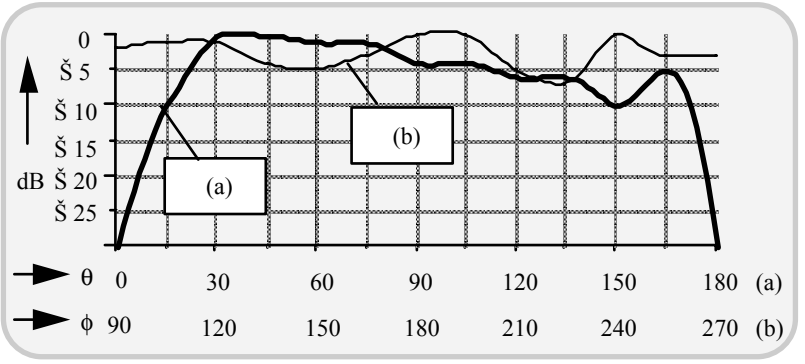


Fig. 4.134 Radiation patterns of a dielectric resonator antenna (a) yz-plane pattern and (b) xy-plane pattern

The high-permittivity thin slab forms a layer on one side of the resonator and the other side of this thin slab is coated with a metallic strip. A coaxial probe extending through the ground-plane connects this strip. The large bandwidth is achieved via the ratio of the dielectric constants of the resonator and the thin slab; the dimensions of metallic strip/coating control the impedance matching.

The radiation patterns of a typical test structure at 1.94 GHz (shown in Figure 4.134) are near omnidirectional. The reason for adopting a rectangular resonator (instead of a circular one), as indicated in [4.77], is to achieve a low cross-polarization level.

4.20 SHORT BACKFIRE (SBF) ANTENNAS

This is a high aperture efficiency antenna, which is small in size, light in weight and simple in configuration. It is therefore preferred in applications such as INMARSAT-M mobile satellite communication systems. The SBF antenna is illustrated in Figure 4.135. It consists of two planar reflectors, a smaller one housed in the layer one with a separation of about $\lambda/2$. This arrangement forms a leaky resonator with radiation beam normal to the small reflector. The antenna is fed by a dipole located between the two planar reflectors. Further, the layer reflector surface has a rim of wall height close to $\lambda/4$.

The SBF antenna offers high gain with low side-lobes (< 20dB). The gain is largely decided by the diameters of the back reflector. This antenna can be modified to include cross-dipole elements in order to obtain circularly polarised radiation. Typical SBF performance parameters are as follows: Gain ≈ 15 dBi and (3-dB)-beamwidth = 34°. The merit of this structure is its simplicity.

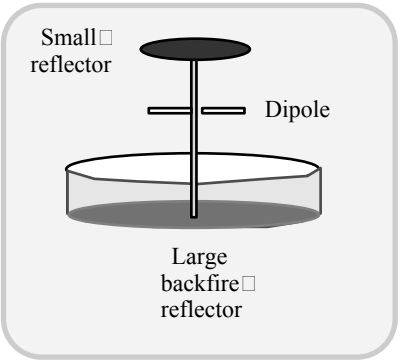


Fig. 4.135 A short backfire antenna: Ratio of the diameters of small and large reflectors is a design compromise for desired radiation characteristics

4.21 DIELECTRIC ANTENNAS

Dielectrics in the form of cylinders, cones, spheres, cone-spheres and in other modified shapes have been found useful structures with unique characteristics. One of the authors of this book has comprehensively addressed these relevant structures in [4.48]. The use of appropriate dielectric antennas for specific wireless communication applications is still an open issue, but has significant potential for designs in miniature forms.

4.22 MOBILE SATELLITE ANTENNAS

A set of four coplanar elements fed with phase difference of 90° can be used to obtain a circular polarization. The elements can be printed on a low permittivity substrate so as to realise a low-profile geometry with reasonable dimensional accuracy. The basic antenna radiates in both directions normal to the antenna plane with LHCP or RHCP (depending on how the antenna is fed).

The basic requirements of an antenna of this type, which are specific to mobile satellite applications (such as for UMTS frequencies), are as follows:

- Circular polarisation
- Near-hemispherical pattern over the operating band
- Easy to manufacture at low cost
- Robust performance and low SAR

The desirable characteristics indicated above are achieved with a structure made of a modified crossed-dipole antenna shown in Figure. 4.136.

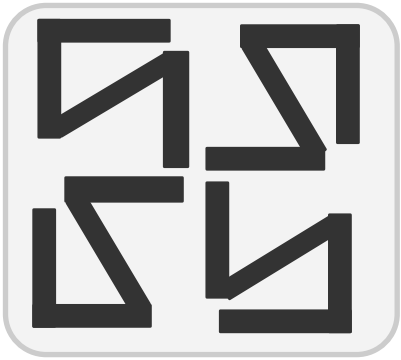


Fig. 4.136 Modified crossed-dipole antenna with a λ/4-spaced finite plane

For most mobile situations, it is desirable to suppress radiation towards the ground. This can be done by facilitating a ground-plane parallel to the plane of the antenna and using a λ/4-array.

4.23 CONCLUDING REMARKS

The significance of this chapter is that, it offers an insight into the fundamental aspects of antenna elements, which are widely adopted as EM radiating structures. In practice, most of these basic structures are used as they are and/or modified for applications in wireless communications. The details furnished in this chapter are relevant considerations gathered from the archival of literature cited (which can be supplanted with additional references listed as [4.78-4.106]). These details are pertinent to classical and applied perspectives of antenna elements. Specifically included is a gamut of topics published in the recent past pertinent to the state-of-the-art radiators used in wireless communication systems.

This chapter is written as a necessary reading companion (with appropriate analytical considerations) so as to grasp the underlying principles of electromagnetics of antenna structures. The subject matter of this chapter will serve as a prelude to the topics of the ensuing chapters.

This page intentionally left blank

REFERENCES

[4.1] Chatterjee, R., *Antenna Theory and Practice*, New Delhi, India: New Age International (P) Ltd. Publishers, 1996

[4.2] Kraus, J. D., *Antennas*, New York, NY: McGraw-Hill Book Co., 1988

[4.3] Balanis, C. A., *Antenna Theory Analysis and Design,* New York, NY: Harper & Row Publishers, 1982

[4.4] Johnson, R. C., and H. Jasik, eds., *Antenna Engineering Handbook*, New York, NY: McGraw-Hill Book Co., 1984

[4.5] Blake, L. V., *Antennas,* Boston, MA: Artech House Publishers, 1984

[4.6] Collin, R. E., and F. J. Zucker, *Antenna Theory,* New York, NY: McGraw-Hill Book Co., 1969

[4.7] Collin, R. E., *Antenna Radiowave Propagation,* New York, NY: McGraw-Hill Book Co., 1985

[4.8] Elliot, R. S., *Antenna Theory and Design,* Englewood Cliff, NJ: Prentice-Hall, Inc., 1981

[4.9] Jordan, E. C., and K. G. Balamain, *Electromagnetic Waves and Radiating Systems,* Englewood Cliff, NJ: Prentice-Hall, Inc., 1968

[4.10] Pozar, D. C., *Antenna Design Using Personal Computers,* Boston, MA: Artech House Publishers, 1985

[4.11] King, R. W. P., *The Theory of Linear Antennas,* Cambridge, MA: Harvard University Press, 1956

[4.12] Lo, Y. T., ed. *Handbook of Antenna Theory and Design,* New York, NY: Van Nostrand Rheinhold, 1987

[4.13] Rudge, A. W., K. Milne, A. D. Oliver and P. Knight eds., *Handbook of Antenna Design,* Stevenage, UK: Peter Peregrinus Ltd, 1983

[4.14] Schelkunoff, S. A., and H. T. Friss, *Antenna Theory and Practice,* New York, NY: John Wiley & Sons, 1952

[4.15] Silver, S., Microwave *Antenna Theory and Design,* New York, NY: McGraw-Hill Book Co., 1949

[4.16] Stutzman, W. L., and G. A. Thiele, *Antenna Theory and Design,* New York, NY: John Wiley & Sons, 1981

[4.17] Wait, J. R., *Antennas and Propagation,* Stevenage, UK: Peter Peregrinus Ltd, 1986

[4.18] Weeks, H., *Antenna Engineering,* New York, NY: McGraw-Hill Book Co., 1968

[4.19] Jasik, W. l., *Antenna Engineering Handbook* New York, NY: McGraw-Hill Book Co., 1961

[4.20] Ramo, S., J. R. Whinnery and T. Van Duzer, *Fields and Waves in Communication Electronics*, New York, NY: John Wiley & Sons, Inc., 1994

[4.21 Balanis, C. A., Pattern distortions due to edge diffractions, *IEEE Trans. Antennas Propagat*., 1970, vol. AP-18, 561-563

[4.22] Legrone, A.H. and G. F. Roberts, Minor lobe suppression in a rectangular horn antenna through the utilization of high impedance choke flange, *IEEE Trans. Antennas Propagat*., 1966, vol. AP-14(1), 102-104

[4.23] Bird, T.S., TE_{11} mode excitation of flanged circular axial coaxial waveguide with an extended center conductor, *IEEE Trans. Antennas Propagat*, December 1987, vol. AP-35(32), 1358-1366

[4.24] Bird, T.S. and G. L. James, Input mismatch of TE_{11} mode coaxial waveguide feeds, *IEEE Trans. Antennas Propagat*., August 1986, vol. AP-34 (8), 1030-1033

[4.25] Rotman, W., A study of single surface corrugated guides, *Proc. IRE*, August 1957, vol. 39, 952-959

[4.26] Neelakantaswamy, P. S., A circular-waveguide primary feed with a corrugated concave disk at its aperture*, Zeitschrift für Elek. Inform-u. Energietechnik*, May 1975, vol. 5(5), 356-361

[4.27] Neelakantaswamy, P. S. and D. K. Banerjee, Radiation characteristics of a circular waveguide aperture with a curved disk, *Arch. Elek. Übertrg*., June 1974, vol. 28, 277-279

[4.28] King, R. W. P., C. W. Harrison, Jr., and D. H. Denton Jr., Transmission line missile antennas, *IRE Trans. Antennas Propagat*., January 1960, vol. 8(1), 88-90

[4.29] Prasad, S., and R. W. P. King, Experimental study on inverted L, J. and related transmission line antennas, *J. Res. NBS*, September/October 1961, vol. 65D

[4.30] Taga, T., and K. Tsunekawa, Performance analysis of a built-in planar inverted-F antenna for 800 MHz portable radio units, *IEEE Trans. Select. Areas Commun*., June 1987, vol. SAC-5(5), 921-929

[4.31] Helmken, H., L. A. Ponce de Leon, and P. S Neelakanta, Radiation characteristics of dipole-excited dielectric disk antenna: Application of field compensation theorem, *Electron. Letts*., July 1991, vol.27(14),1277-1278

[4.32] Chaoli, G., P. S. Neelakanta, V. Ungvichian and P. F. Wahid, Radiation characteristics of an aperture on the surface of a composite medium*, Arch. Elek. Ubertrg.*, 1993, vol. 47(1), 60-61

[4.33] Neelakanta, P. S., *The Handbook of Electromagnetic Materials: Monolithic and Composite Versions and Their Applications*, Boca Raton, FL: CRC Press, 1995

[4.34] Lee, W. C. Y., *Mobile Cellular Telecommunications*, New York, NY: McGraw-Hill, Inc., 1995

[4.35] U.S. Patent 4873527, Oct 1989

[4.36] Young, W. F., B. Belzer, R. Olsen, A field/polarization diversity antenna for portable phones, *Proc. 1998 IEEE-APS Conf. on Antennas and Propagation for Wireless Communications*, Atlanta, Ga, (June 21-26, 1998), 125-128

[4.37] Young, W. F., B. Belzer and R.G. Olsen, A two-element antenna for null suppression in multipath environment, *IEEE Trans. Antennas Propagat*., August 2000 vol. 48(8), , 1161-1174

[4.38] Hansen, W. W., and J. R. Woodyard, A new principle in directional antenna design, *Proc. IRE*, March 1938, vol. 26, 333-345

[4.39] Terada, N. and K. Kagoshima: Compatible mobile antenna for mobile satellite and cellular communication systems, *IEICE Tech. Rep. on Antennas and Propagation*, AP 91-65, 1991

[4.40] Kilgus, C. C., Shaped conical radiation pattern performance of the backfire quadrifiler helix, IEEE Trans. Antennas Propagat., May 1975, vol. AP-23, pp. 392-397

[4.41] Fujimoto, K., A Henderson, K. Hirasura and J. R. James, *Small Antennas*, Herts., UK: Research Studies Press 1987

[4.42] Rumsey, V., *Frequency Independent Antennas*, New York, NY: Academic Press, 1966

[4.43] DuHamel, R. H., and D. E. Isbell, Broadband logarithmically periodic antenna structures, *1957 IRE National Convention Record*, pt. 1, 119-128

[4.44] Silver, S., and W. K. Saunders, The external field produced by a slot in an infinite circular cylinder, *J. Appl. Phys*., February 1950, vol. 21, 153-158

[4.45] Schaffer, H., Die Strahlung der mit H-Wellen angrten oftnen Koaxialleitung, *Arch. Elektr. Übertr*., 1968, vol. 22, 514-518

[4.46] Clarricoats, P. J. B., and A. D. Oliver, *Corrugated Horns for Microwave Antennas*, Piscataway, NJ: IEEE Service Center, 1984

[4.47] Neelakantaswamy, P.S., *Studies on open-ended circular waveguide radiators with modified aperture-ends*, Ph. D. Thesis, Department of Electrical Engineering, Indian Institute of Technology, Madras (India), 1975

[4.48] Chatterjee, R., *Dielectric and Dielectric-Loaded Antennas*, Herts., UK: Research Studies Press Ltd., 1985

[4.49] Love, A. W. (Ed.), *Electromagnetic Horn Antennas*, New York, NY: IEEE Press, 1976

[4.50] Hu, Y. Y., A method of determining phase centers and its applications to electromagnetic horns, *J. Franklin Institute*, January 1961, vol. 271, 31-39

[4.51] Othera, I., and H. Ujiie, Nomographs for phase centres of conical corrugated and TE_{11} mode horns. *IEEE Trans. Antennas Propagat*., November 1975, vol. AP-23(6), 858-859

[4.52] Yamada, Y., T. Nara, S. Kameo, Y. Chatani and H. Abe, " A variable beamwidth corner reflector antenna," *IECE Nat. Conv. Record*, No. 694, 1986

[4.53] James, J. R., and P. S. Hall, *Handbook of Microstrip Antennas*, Vol. I & II, London, UK: Peter Peregrinis, 1989

[4.54] Pozar, D. M., and D. H. Schaubert (Eds.), *Microstrip Antennas: The Analysis and Design of Microstrip Antennas and Arrays*, New York, VY: IEEE Press, 1995

[4.55] Carver, K. R., and J.W. Mink, Microstrip antenna technology, *IEEE Trans. Antennas Propagat.*, 1981, vol.AP. 29(1), 2-24

[4.56] Tulintself, A. N. S.M. Ali and J.A. Kong, "Input impedance of a probe-fed stacked circular microstrip antenna," *IEEE Trans Antenna and Propagation*, 1991, vol. AP-39(3), 1991, 381-390

[4.57] Lubin, Y. and A. Hessel, "Wide-band wide-angle microstrip stacked patch-element phased arrays," *IEEE Trans. Antennas Propagat*., 1991, vol. AP-39(8) 1991, 1062-1070

[4.58] Weinschel, H. D., A cylindrical array of circularly polarized microstrip antennas, *IEEE AP-S Int. Symp. Digest*, 1975, 177-180

[4.59] Zavosh, F. and J.T. Aberatle, Improving the performance of microstrip-patch antenna, *IEEE Antennas Propagat. Mag*., 38(4), 1996, 7-12

[4.60] Waterhouse, R. B., Small microstrip patch antenna, *Electron Lett*., April 1995, vol.31, 604-605

[4.61] Park, I., and R. Mittra, Aperture-coupled small microstrip antenna*, Electron Lett.,* September 1996, vol.32, 1741-1742

[4.62] Wong, K. L., and S.C. Pan, Compact triangular microstrip antenna, *Electron Lett.,* March 1997, vol. 33, 433-434

[4.63] Roy, P. Etudes et réalisations d'antennes imprinmées multifonctions. Application au regroupement des dispoisitifs de radiocommunications sur les mobiles, *Thèse de Doctorat d'Electronique* - n°47-48 – U.E.R. des Sciences-Université de Limoges – Novermber 1998

[4.64] Zürcher, J. F., and F.E. Cardiol, *Broadband Patch Antennas*, Norwood, MA: Artech House Inc, 1995

[4.65] Zürcher, J. F., Q. Xu, A.K. Skrivervik and J. R. Mosig, Dual-frequency, dula-polarization 4-port printed planar antenna. *Microwave Opt. Tech. Lett.,*, October 1999, vol. 23(2), 75-78

[4.66] Habib, L., G. Kossiavas and A. Papiernik, Cross-shaped patch with Etched bars for dual polarization, *Electron Lett*., 1993, vol. 29(10), 916-917

[4.67] Taga, T., K. Tsunekawa and A. Sasaki, Antennas for detachable mobile radio units, *Rev. of the ECL, NTT* (Japan), January 1987, vol. 35(1), 59-85

[4.68] Yokoyama, Y., and N. Gotoh, Dual resonance broadband microstrip antenna, *Proc. Intl. Symposium Antenna and Propagation*, Kyoto, Japan, August 1985, 429-432

[4.69] Barreiros, J., P. Cameirão and C. Peixeiro, Microstrip patch antenna for GSM 1800 handsets, *Proc. 1999 IEEE International Symp., Antenna and Propagat. Soc.,* vol 3., July 11-16, 1999 Orlando, FL, pp. 2074-2077

[4.70] Song, P., P.S. Hall, H., Ghafouri-Shiraz and D. Wake, Triple-band planar inverted F-antenna, *Proc. 1999 IEEE International Symp., Antennas and Propagat. Soc.,* vol 3., July 11-16, 1999 Orlando, FL, pp. 908-911

[4.71] Rowell, C. R., and R.D. Murch, A compact PIFA suitable for dual-frequency 900/1800-MHz operation, *IEEE Trans. Propagat*., April 1998, vol. 46(4), 596-598

[4.72] Viratelle, D., and R.J. Langley, Dual-band PIFA antenna, *Paper # 0114, Session 5A5: Adaptive Antennas*, Millennium Conf. on Antennas and Propagat., 9-14 April 2000, Duvos, Switzerland

[4.73] Wong, K. L., and K. P. Yang, Modified planar inverted F-antenna, *Electron Lett*., January 1998, vol.34(1), 7-8

[4.74] Rowell, C. R., and R.D. Murch, A capacitively loaded PIFA for compact mobile telephone handsets, *IEEE Trans Antenna Propagat*., 1997, vol. AP-45, 837-842

[4.75] Ma, G., P.S. Hall, P. Gardner and M. Hajan, Zero-IF active antenna, *Electron Lett*., January 2001, vol 37(1), 3-4

[4.76] Radisic, V., Y. Qian and T. Itoh, Novel architectures for high-efficiency amplifiers for wireless applications, *IEEE Trans Microwave Theory Tech*., November 1998, vol. 46(11), 1901-1909

[4.77] Moon, J. I., and S.O. Park, Dielectric-resonator antenna for dual-band PCS/IMT-2000, *Electron Lett.,* June 2000, 36(12), 1002-1003

ADDITIONAL REFERENCES

Books on Antennas and Propagation

[4.78] Milligan, T. A., *Modern Antenna Design*, New York, NY: McGraw-Hill Book Co., 1985

[4.79] Wolff, E., *Antenna Analysis*, Norwood, MA: Artech House, 1988

[4.80] Schelkunoff, S. A., *Advanced Antenna Theory*, New York, NY: John Wiley & Sons., 1952

[4.81] Lee, K. F., *Principles of Antenna Theory*, New York, NY: John Wiley & Sons., 1984

[4.82] Maclean, T. S., *Principles of Antennas -Wire and Aperture*, Cambridge, UK: Cambridge Press., 1986

[4.83] Griffiths, J., *Radio Wave Propagation and Antennas*, Englewood Cliffs, NJ: Prentice Hall International., 1987

[4.84] Monser, G., *Antenna Design –A Practical Guide*, New York, NY: McGraw-Hill Book Co., 1996

[4.85] Kuecken, J. A., *Antennas and Transmission Lines*, Indianapolis, IN: Howard Sams, 1969

[4.86] Rulf, B., and G. A. Robertshaw, *Understanding Antennas for Radar, Communications and Avionics*, New York, NY: Van Nostrand Reinhold Co., 1987

[4.87] Johnson, R. C., *Antenna Engineering Handbook*, New York, NY: McGraw-Hill Book Co., 1993

[4.88] Clarricoats, P. J. B. (Ed.), *Advanced Antenna Technology*, London, UK: Microwave Exhibitions and Publications., 1981

Books on Microstrip Antennas and Printed Antennas

[4.89] Bahl, I. J., and P. Bhartia, *Microstrip Antennas*, Norwood, MA: Artech House, Inc., 1980

[4.90] Wolff, I., *Microstrip-Bibliography (1948-1976)*, Aachen, Germany: Verlag H. Wolff, 1979

[4.91] Gupta, K. C., and A. Benella (Eds.), *Micrstrip Antenna Design*, Norwood, MA: Artech House, Inc., 1988

[4.92] Bhartia, K., V. S. Rao and R. S. Tomar, *Millimeter-Wave Microstrip and Printed Circuit Antennas*, Norwood, MA: Artech House, Inc., 1988

[4.93] Edwards, T. C., *Foundations of Microstrip Antennas*, New York, NY: John Wiley & Sons, 1981

[4.94] Zürcher, J. F., and F. Gardiol, *Broadband Patch Antennas*, Norwood, MA: Artech House, Inc., 1995

[4.95] Herscovici, N., *CAD of Aperture-Fed Microstrip Transmission Lines and Antennas*, Norwood, MA: Artech House, Inc., 1996

[4.96] Sainati, R. A., *CAD of Microstrip Antennas for Wireless Applications*, Norwood, MA: Artech House, Inc., 1996

Books on Wireless Communication Antennas

[4.97] Hirasawa, K., and M. Maneishi, *Analysis, Design, and Measurement of Small and Low-Profile Antennas*, Norwood, MA: Artech House, Inc., 1992

[4.98] Fujimoto, K., A. Henderson, K. Hirasawa, and J. R. James, *Small and Low-Antennas*, Herts., UK: Research Studies Press Ltd., 1987

[4.99] Fujimoto, K., and J. R. James, *Analysis, Mobile Antenna Systems Handbook*, Norwood, MA: Artech House, Inc., 1994

[4.100] Okamoto, G. T., *Smart Antenna Systems and Wireless LANS,* , Norwell, MA: Kluwer Academic Publishers., 1999

[4.101] Siwiak, K., *Analysis, Radiowave Propagation and Antennas for Personal Communications*, Norwood, MA: Artech House, Inc., 1995

[4.102] Saunders, S. R., *Antennas and Propagation for Wireless Communication Systems*, New York, NY: John Wiley & Sons., 1999

[4.103] Setian, L., *Antennas with Wireless Communications*, Upper Saddle River, NJ: Prentice Hall PTR, 1999

[4.104] Liberti, Jr. J. C., and T. S. Rappaport *Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications,* Upper Saddle River, NJ: Prentice Hall PTR, 1999

[4.105] Pattan, B., *Robust Modulation Methods and Smart Antennas in Wireless Communications*, Upper Saddle River, NJ: Prentice Hall PTR, 2000

[4.106] Godara, L.C. ed., *Handbook of Antennas in Wireless Communications*, Boca Raton, FL: CRC Press, 2001

CHAPTER 5

Array Antennas in Wireless Communication Systems

5.1 INTRODUCTION

An *antenna array* is formed with a group of two or more antenna elements so as to obtain a desired antenna pattern. In making an array antenna, first the individual antenna elements are chosen for their type (such as half- or quarter-wave dipole). Then they are configured in such a way that their radiation fields interact with each other and produce a resultant pattern, which, in essence, is the vector sum of individual fields due to the elements used [5.1-5.7].

In exciting the antenna elements of an array, the amplitude and phase of the signal impressed on each element is trimmed so that the net radiation pattern is the choice of the design pursued.

An array element can be either *active* or *passive*. The active elements refer to electrically connected (driven) sources (in transmit mode) or sinks (in receive mode). The non-driven (that is, electrically not connected elements) represent passive or *parasitic* elements. Such parasites are excited only through mutual induction with a driven element or another parasitic element.

As indicated above, the purpose of an array is to achieve a desired directivity of the resultant radiation pattern and place the pattern in a desired location of a geographical area. Also, it is possible to let the pattern in a *search-light mode* continuously steering (tracking) over a specified spatial region. Alternatively, the beam can be discretely *switched* over the span of coverage area. Such tracking and switching of the beam can be done electronically through time-dependent excitation of antenna elements with a weighted complex signal.

In the context of cellular communication applications, the array antennas are used for the following purposes:

- In base stations applications, a set of directional antennas are needed for a 120°-sectoral coverage. Such antennas are provisioned either at each corner of a cell or at the cell centre. This arrangement provides higher radiating power in directions where signal power is likely to diminish (as a result of a building, for example, down a street casting shadow in that direction). In some cases a single directional beam may be focused on an area of specific interest or when the area poses poor communication environment. The directional antennas for these applications are conceived *via* array structures. Typically, they are mounted on a

platform meeting the requirement as one-sector support, two-sector support, three-sector support or four-sector support on an *ad hoc* basis. In the simplest form, multiple antennas at the base stations may be used to form multiple beams to cover the whole cell site. Three beams each with a 120°-beamwidth (or six beams each with 60°-beamwidth) can be used for this purpose. The coverage of each beam is then treated as a separate cell, and the frequency assignment is done accordingly. Mobiles are handed to the next beam as they leave the area caused by the current beam. (This is similar to the *hand-off process* when the mobile cross the cell boundary.) The display of typical directional supports used for base station antenna mounting is presented in Figure 5.1.

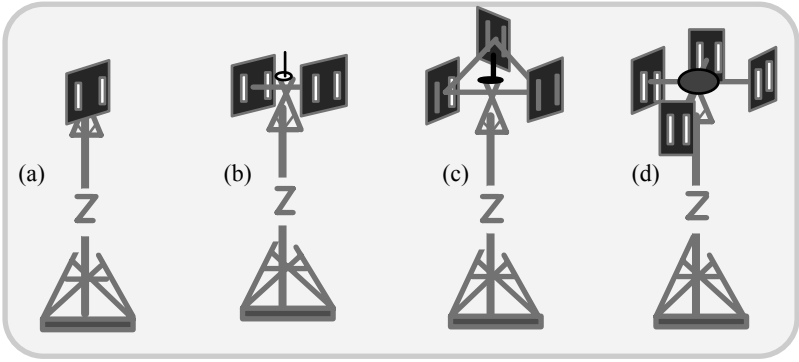


Fig. 5.1 Directional base station antenna supports: (a) One-sector support; (b) two-sector support; (c) three-sector support and (d) four-sector support.

- Traditional base station installations of mobile communication make use of *space-diversity* techniques, which require at least two antennas pointing in the same direction and separated by a distance of 10 to 20 wavelengths. Such spatially-separated antennas provide decorrelated signals, so that in the event of one signal fading heavily, there is a probability the other might not , thus offering a good signal for further processing. Alternative to space-diversity, *polarisation-diversity* can be done through simultaneous reception of two orthogonal polarized signals by means of a dual-polarized antenna
- Another wireless communication antenna configuration that uses the array concept refers to *adaptive antennas*. These arrays are designed to achieve narrow antenna beams in the azimuthal plane so as to obtain a good angular resolution to search/track the position of mobile terminals. When an adaptive array enables the location of each mobile, a set of beams is formed to cover

different mobiles (or group of mobiles). Each such beam can be regarded as a co-channel cell and thus may be able to use the same frequency or code. Considering two time instants T_1 and T_2, how different beams can be formed so as to cover various mobiles adaptively along the directions of mobility is illustrated in Figure 5.2.

Further, the array pattern can be dynamically adjusted to cancel any undesired (interfering) directional sources operating at the same frequency as that of the desired (target) subscriber. This is done by placing nulls in the pattern along those directions of interferences and the main beam is steered in the direction of the desired signal.

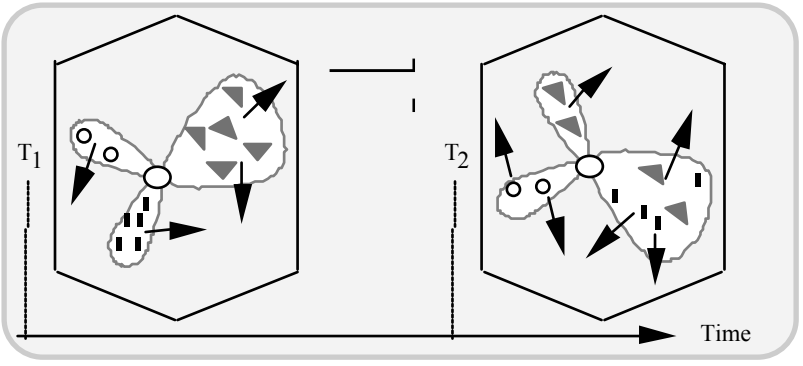


Fig. 5.2 Adaptive coverage of mobile units with the passage of time with antenna beams

Thus, array configurations are used profusely in mobile communication technology for specified reasons. To understand the characteristics and design features of such arrays, relevant theoretical considerations and practical aspects are necessary, as discussed in this chapter.

5.2 THEORY OF ANTENNA ARRAYS

The antenna elements indicated earlier could be a set of linear dipoles or a group of aperture radiators. It is also possible to form an *array of arrays*. To understand the basics of array antennas, the analysis presented in this section covers details on evaluating the radiation field-patterns of various array geometry. Further the *driving point impedance* characteristics and *directive gain* of arrays are defined and determined.

5.2.1 Linear arrays of n isotropic point-sources of equal amplitude and spacing

Figure 5.3 shows a linear array of n isotropic point-sources of equal amplitude and spacing. Though a point-source is fictitious, it offers a convenient concept in the electromagnetic wave theory. This is because of the following reason: In EM radiation problems, it is only the far-field of the antenna (where the electric and magnetic fields vary as 1/r) that is of interest. Hence it can be assumed that, by extrapolating inward along the radii of a sphere, the EM waves appear to originate from a fictitious point-source presumably located at the centre of the observation sphere.

Considering a collinear array of n point-sources (Figure 5.3), the total field E at a large distance in the direction of ϕ is given by:

$$E = 1 + e^{j\psi} + e^{2j\psi} + e^{3j\psi} + \ldots + e^{j(n-1)\psi} \tag{5.1a}$$

where

$$\psi = d_r \cos(\phi) + \delta \tag{5.1b}$$

$$d_r = 2\pi d/\lambda \tag{5.1c}$$

and δ is the phase difference (between the excitations) of adjacent sources.

Suppose the amplitudes of the fields of all the sources are taken as equal to unity and source 1 is considered as the phase-centre. Multiplying equation (5.1a) by $e^{j\phi}$, it follows that,

$$E\, e^{j\phi} = E\,[e^{j\psi} + e^{2j\psi} + e^{3j\psi} + \ldots e^{nj\phi\psi}] \tag{5.2}$$

Subtracting equation (5.2) from equation (5.1a) and dividing by $(1 - e^{jy})$, the result is,

$$E = \frac{1 - e^{jn\psi}}{1 - e^{j\psi}} = e^{j\xi}\,\frac{\sin(n\psi/2)}{\sin(\psi/2)} = \frac{\sin(n\psi/2)}{\sin(\psi/2)}\angle\xi \tag{5.3}$$

where

$$\xi = \left(\frac{n-1}{2}\right)\psi \tag{5.4}$$

If the centre point of the array (instead of source 1) is considered as the phase-centre, then

$$E = \frac{\sin(n\psi/2)}{\sin(\psi/2)} \tag{5.5}$$

Equation (5.1a) may be regarded as the addition of several vectors. The total (resultant) field E can be obtained by a graphical vector addition, as shown in Figure 5.4. The phase of the field is constant wherever E has a finite value but changes by π in those directions for which E = 0 (null directions) (so that equation (5.5) changes sign accordingly).

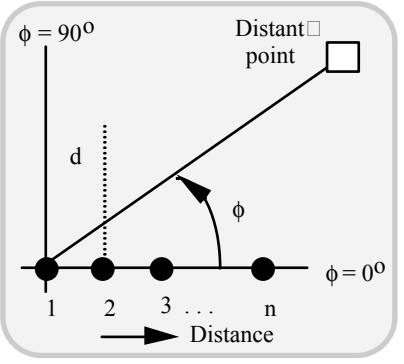


Fig. 5.3 Linear array of n equally-spaced point-sources. (All point sources are assumed to be identical)

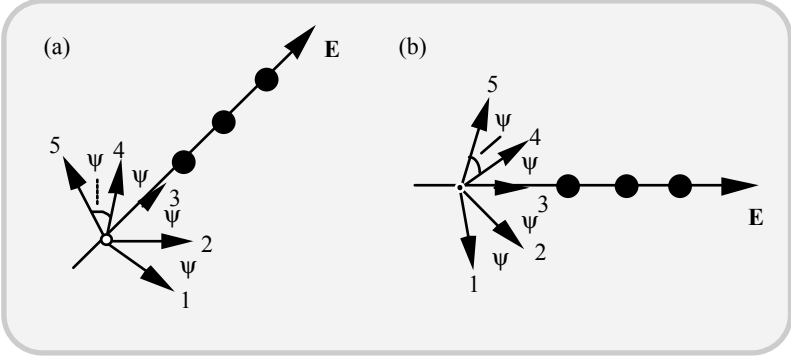


Fig. 5.4 Vector addition of fields to obtain the total field at a large distance from a linear array of n = 5 point-sources of equal amplitude and equal spacing: (a) Source 1 as phase-centre, (b) Source 3 (mid-point) as phase-centre

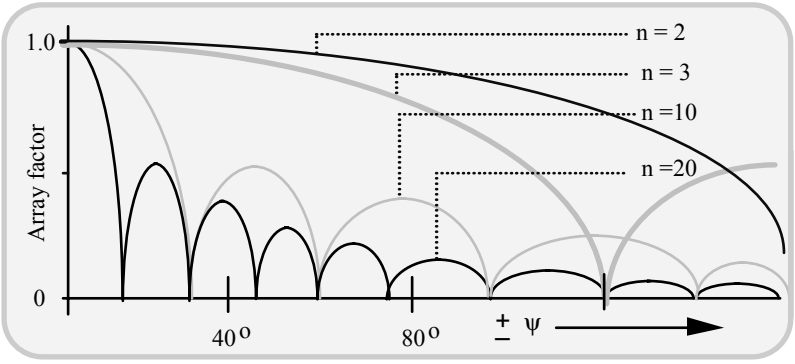


Fig. 5.5 Array factor as a function of ψ for various values of n as per equation (5.7)

When ψ = 0, equations (5.4) and (5.5) are indeterminate, and hence E must be obtained as the limit of equation (5.5) as ψ → 0. Thus, for ψ = 0,

$$E = n \tag{5.6}$$

Equation (5.6) gives the maximum value of E. Therefore, by designating $E_{max} = n$, the normalized value of E is obtained as,

$$E = \left(\frac{1}{n}\right)\frac{\sin(n\psi/2)}{\sin(\psi/2)} \tag{5.7}$$

The factor sin(x)/(x) in equation (5.7) is known as the *array factor*. The field is maximum when ψ = 0. But in some cases, ψ may not become zero. In such cases, the field maximum will correspond to the minimum value of ψ. Figure 5.5 shows the array factor as a function of ± ψ for various values of n.

5.2.2 Linear broadside array of point-sources

In a linear array of n isotropic point-sources of the same amplitude and phase, δ = 0, so that,

$$\psi = d_t\cos(\phi) \tag{5.8}$$

For ψ = 0, ϕ = (2k + 1)π/2, where k = 0, 1, 2, 3, … Further, the field maximum occurs when ϕ = π/2 and 3π/2. The maximum field corresponds to the direction normal to the array. Such an array is called a *broadside array*. For example, Figure 5.6 shows the field and phase patterns of a broadside array of four point-sources in phase, spaced λ/2 from each other.

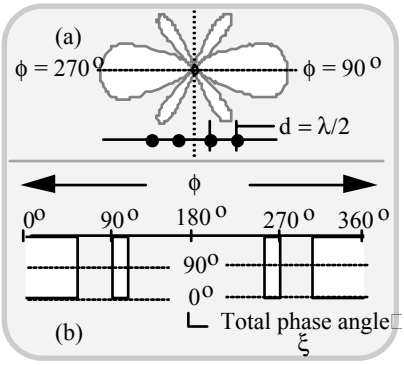


Fig. 5.6 Field and phase patterns of a broadside array of four collinear point-sources (d = λ/2): (a) Field pattern and (b) phase pattern of the array (with the phase-centre located at the mid-point of the array)

5.2.3 Ordinary end-fire array of point-sources

Unlike in a broadside array, the maximum in end-fire arrays occurs at $\phi = 0$, that is, along the direction of collinearly placed array elements. Hence, $\psi = 0$ when $\phi = 0$, and as such, from equation (5.1b), it can be written as,

$$\delta = -d_r \tag{5.9}$$

which means that the phase between adjacent sources is retarded progressively by the same amount as the spacing between sources (expressed in radians).

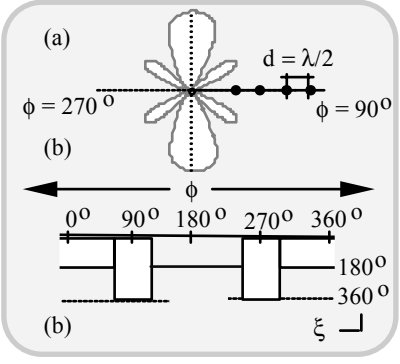


Fig. 5.7 Field and phase patterns of broadside array of four point-sources (d = λ/2): (a) Field pattern; (b) phase pattern with mid-point of array as phase-centre. (ξ denotes the total phase angle)

For example, considering an end-fire array of four point-sources with $d = \lambda/2$, $\delta = -d_r = -2\pi d/\lambda = -2\pi/2 = -\pi$. Figure 5.7 shows the field and phase patterns of such an end-fire array.

Comparing the broadside pattern of Figure 5.6 and the end-fire pattern of Figure 5.7 (both corresponding to identical collinear array of four elements), it can be seen that the shapes of the patterns are the same in both cases; however, the orientation of the patterns are distinctly disposed with respect to the collinear direction of the elements. The orientation of the pattern is decided by the relative phase of the signals exciting the elements. In Figures 5.6 and 5.7, the phase considerations are such that the resulting patterns are broadside and end-fire types, respectively.

5.2.4 End-fire array with increased directivity

Hansen and Woodyard [5.8] have shown that a higher directivity could be obtained by making,

$$\delta = -(d_r + \pi/n) \tag{5.10}$$

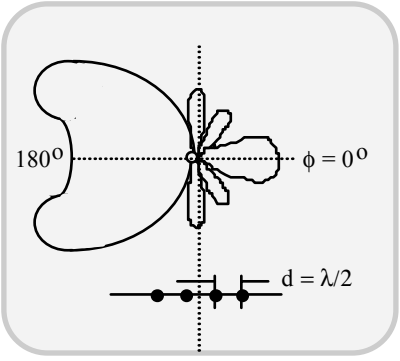


Fig. 5.8 Field pattern of an end-fire array of four point sources with ($d = \lambda/2$) and ($\delta = -5\pi/4$)

The condition specified by equation (5.10) is known as the *Hansen and Woodyard's condition* of increased directivity. For example, for a linear array of four point-sources with $d = \lambda/2$, equation (5.10) leads to $\delta = -(\pi + \pi/4) = -5\pi/4$. The corresponding field pattern of such an array is shown in Figure 5.8. In this pattern, the main lobe is considerably sharper. (However the extent of the back lobe is much larger. That is, the reduced radiated power along the narrower end-fire beam manifests as the back and/or minor lobes in the field pattern).

To realize a better pattern with increased directivity, the spacing d should be reduced and s should be increased so that $|\psi|$ is about π/n at $\phi = 180°$. For example, if $n = 10$ and $d = \lambda/4$, then $\delta = -0.6\pi$. Figure 5.9 shows the field pattern of such an

array, which definitely has increased directivity and reduced back lobe, (but carries an infestation of several minor lobes).

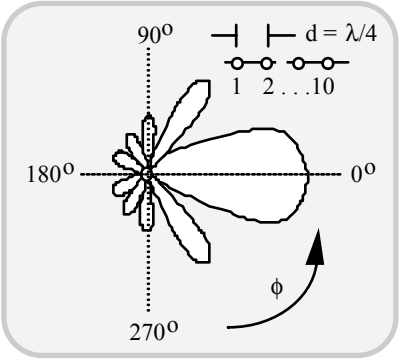


Fig. 5.9 Field pattern of end-fire array of ten point-sources with (d = λ/4). The pattern exhibits increased directivity, reduced back lobe but enhanced number of minor lobes

5.2.5 Array with maximum field in an arbitrary direction

Should a linear array of n point-sources have a maximum in some of arbitrary direction ϕ_1 (not equal to $k\pi/2$, where k = 0, 1, 2 or 3), then, the required condition is given by,

$$d_r \cos(\phi_1) + \delta \equiv 0 \tag{5.11}$$

By specifying the value of d_r and knowing ϕ_1, δ can be found out.

5.2.6 Direction of nulls and maxima for arrays of n isotropic point-sources of equal amplitude and spacing

Nulls

From equation (5.3) it can be seen that the null directions for an array of n isotropic point-sources (of equal amplitude and spacing) occur when E = 0; that is, when,

$$e^{jn\phi} = 1 \tag{5.12}$$

provided

$$e^{j\phi} \neq 1 \tag{5.13}$$

Equation (5.12) requires that,

$$n\psi = \pm 2K\pi \tag{5.14}$$

where K =1, 2, 3, … . This leads to the following relation:

$$\psi = d_r \cos(\phi_0) + \delta \equiv \pm \frac{2K\pi}{n} \tag{5.15}$$

or

$$\phi_o = \cos^{-1}\left[\left(\pm\frac{2K\pi}{n} - \delta\right) \times \frac{1}{d_r}\right] \tag{5.16}$$

where ϕ_0 gives the direction of the pattern nulls. (*Note*: The condition imposed by equation (5.13) requires that the values of K= mn should be excluded, where m = 1, 2, 3, …)

In a *broadside array,* $\delta = 0$, so that ϕ_o can be specified by the following relation:

$$\phi_o = \cos^{-1}\left[\pm\frac{2K\pi}{nd_r}\right] = \cos^{-1}\left(\pm\frac{K\lambda}{nd}\right) \tag{5.17}$$

If γ_o is the complementary angle of ϕ_0, then $\gamma_o = (\pi/2 - \phi_0)$, so that

$$\gamma_o = \sin^{-1}\left(\pm\frac{K\lambda}{nd}\right) \tag{5.18}$$

For a long array for which nd >> kλ, the following approximation is valid:

$$\gamma_o \approx \pm\frac{K\lambda}{nd} \tag{5.19}$$

And, for an ordinary end-fire array $\delta = -d_r$, so that from equation (5.16),

$$\cos(\phi_o) - 1 = \pm \frac{2K\pi}{nd_r} \tag{5.20}$$

or

$$\frac{\phi_o}{2} = \sin^{-1}\left\{\pm\sqrt{\frac{K\pi}{nd_r}}\right\} \tag{5.21}$$

Further, for a long array for which nd >> Kλ

$$\phi_o \approx \pm\sqrt{\frac{2K\lambda}{nd}} \tag{5.22}$$

Hence, for a long ordinary end-fire array, the total beamwidth of the main lobe between first nulls is approximately given by:

$$2\phi_{o1} \approx \pm\sqrt{\frac{2\lambda}{nd}} \tag{5.23}$$

For *end-fire arrays with increased directivity* (as proposed by the Hansen and Woodyard condition), equation (5.16) becomes

$$d_r(\cos\phi_0 - 1) - \frac{\pi}{n} = \pm 2\frac{K\pi}{n} \tag{5.24}$$

That is,

$$\frac{\phi_o}{2} = \sin^{-1}\left\{\pm\sqrt{\frac{\pi}{2nd_r}(2K - 1)}\right\} \tag{5.25}$$

or

$$\phi_o = 2\sin^{-1}\left\{\pm\sqrt{\frac{\lambda}{4nd}(2K - 1)}\right\} \tag{5.26}$$

For a long array for which nd >> Kλ, equation (5.26) simplifies to the following approximate relation:

$$\phi_o \approx \pm\sqrt{\frac{\lambda}{nd}(2K - 1)} \tag{5.27}$$

Hence, the total beamwidth of the main lobe between first nulls for a long end-fire array with increased directivity is given by,

$$2\phi_{o1} \approx \pm 2\sqrt{\frac{\lambda}{nd}} \tag{5.28}$$

which is $1/\sqrt{2}$ (or 71.7 percent) of the beamwidth of the ordinary end-fire array. Figure. 5.10 shows the beamwidth between the first-nulls in the pattern of different arrays of n isotropic point-sources of equal amplitude as a function of nd/λ.

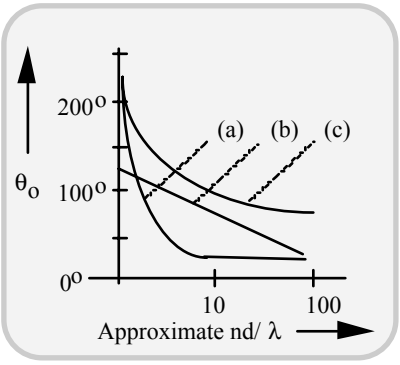


Fig. 5.10 Beamwidth between first nulls (θ_o) *versus* the array length (nd/λ) for arrays of n isotropic point-sources of equal amplitude: (a) Broadside array; (b) End-fire array with increased directivity and (c) ordinary end-fire array.

The maxima

The major lobe maximum usually occurs when $\psi = 0$, for a broadside or for an ordinary end-fire array. In the case of broadside arrays, this corresponds to $\phi = 90°$ and 270°. For an ordinary end-fire array, the major-lobe points along $\phi = 0°$ and 180°. For the end-fire array with increased directivity, the main lobe maximum occurs at $\psi = \pm \pi/n$ with the main lobe at 0° or 180°.

The maxima of the minor lobes are situated between the first and higher order nulls. These maxima occur approximately whenever the numerator of equation (5.7) is a maximum, that is, when,

$$\sin(n\psi/2) = 1 \tag{5.29}$$

The numerator $\sin(n\psi/2)$ of equation (5.7) varies as a function of ψ more rapidly than the denominator $\sin(n\psi/2)$. This is especially true whenever n is large. This can be seen in Figure 5.11. Hence, the maxima occur approximately when $\sin(n\psi/2) = 1$, which requires that

$$\frac{n\psi}{2} = \pm(2K+1)\frac{\pi}{2} \tag{5.30}$$

where K = 1, 2, 3, ….

Substituting equation (5.30) in equation (5.1b), it follows that,

$$d_r \cos(\phi_m) + \delta = \frac{\pm(2K+1)\pi}{n} \tag{5.31}$$

or

$$\phi_m \approx \cos^{-1}\left\{\left[\frac{\pm(2K+1)\pi}{n} - \delta\right]\frac{1}{d_r}\right\} \tag{5.32}$$

where ϕ_m gives the direction of the minor lobe maxima.

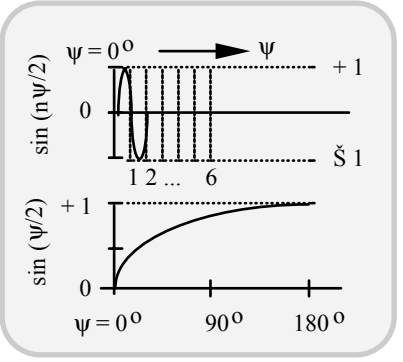


Fig. 5.11 sin ($n\psi/2$) and sin ($\psi/2$) *versus* ψ for n = 0. (Points (1,2, …, 6) represent the first maximum for broadside or ordinary end-fire array, first maximum for end-fire with increased directivity, first null, approximate maximum of first minor lobe, second null, and approximate maximum of second minor lobe respectively)

For a broadside array, $\delta = 0$, so that equation (5.32) becomes,

$$\phi_m \approx \cos^{-1}\left[\frac{\pm(2K+1)\lambda}{2nd}\right] \tag{5.33}$$

And, in reference to an ordinary end-fire array, $\delta = -d_r$ so that

$$\phi_m \approx \cos^{-1}\left[\frac{\pm(2K+1)\lambda}{2nd} + 1\right] \tag{5.34}$$

Considering an end-fire array with increased directivity, $\delta = -(d_r + \pi/n)$ and as such,

$$\phi_m \approx \cos^{-1}\left\{\frac{\lambda}{2nd}\left[1 \pm (2K+1)\right] + 1\right\} \tag{5.35}$$

Since $\sin(n\psi/2)$ is approximately equal to unity at the maximum of a minor lobe, the relative amplitude of a minor lobe maximum E_{mL} is given by,

$$E_{mL} \approx \frac{1}{n\sin(\psi/2)} \tag{5.36}$$

By expressing ψ explicitly in terms of K and n,

$$E_{mL} \approx \frac{1}{n\sin[(2K+1)\pi/2n]} \tag{5.37}$$

When n >> K, that is, for the first few minor lobes of an array of a large number of sources,

$$E_{mL} \approx \frac{2}{(2K+1)\pi} \tag{5.38}$$

In a broadside ordinary end-fire array, the major lobe maximum is unity with the relative amplitudes of the maximum and the following first five minor lobes being; 1, 0.22, 0.13, 0.09, 0.07 and 0.06 respectively. For an end-fire array with increased directivity, the maximum for $\phi = 0$ and n = 20 occurs at $\psi = \pi/20 = 9°$; at this value the array factor is equal to 0.63. Therefore, the relative amplitudes for such an array become 1, 0.35, 0.21, 0.14, 0.11 and 0.09.

The maximum value of the smallest minor lobe occurs for (2K + 1) = n. Hence,

$$\sin\left[\frac{(2K+1)\pi}{2n}\right] \approx 1 \tag{5.39}$$

and

$$E_{mL} \approx \frac{1}{n} \tag{5.40}$$

5.2.7 Two isotropic point-sources of unequal amplitude and any phase difference

Suppose two sources 1 and 2 are situated on the x-axis at a distance d as shown in Figure 5.12(a). If source 1 has a field E_o at a large distance r, and the field from source 2 at the same distance is aE_o, then from Figure 5.12(b), the magnitude and phase angle of the total field E is given by:

$$E = \left\{E_o\sqrt{[1+a\times\cos(\psi)]^2 + (a^2)\sin^2(\psi)}\right\}$$
$$\angle\tan^{-1}\left(\frac{a\times\sin(\psi)}{1+a\times\cos(\psi)}\right) \tag{5.41}$$

where $\psi = d_r\cos(\phi) + \delta$, and the phase angle of E is referred to source 1.

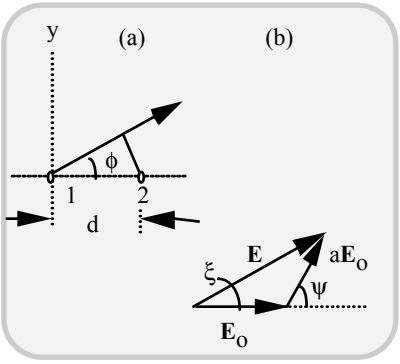


Fig. 5.12 Two isotropic sources of unequal amplitude and arbitrary phase: (a) Geometry, (b) Vector addition of field components

5.2.8 Nonisotropic similar point-sources and the principle of pattern multiplication

While an isotropic point-source radiates equally in all directions, a non-isotropic point-source does not do so. Two nonisotropic point-sources are said to be "similar" if the variation of the amplitudes and phases of their radiation fields with absolute angle ϕ is the same. Further, the patterns must not only be the same shape but also be oriented in the same direction. If the maximum amplitudes are equal, then the sources are not only "similar" but are also "identical". The total field of an array made of nonisotropic sources can be ascertained by *the principle of pattern multiplication,* which can be stated as follows:

> *"The total field pattern of an array of nonisotropic similar sources is the product of the individual source pattern of an array of isotropic point sources each located at the phase-centre of the individual source and having the same relative amplitude and phase, and the total phase pattern referred to the phase centre of the array is the sum of the phase patterns of the individual source and the array of isotropic point sources".*

In symbolic form the above principle can be expressed as follows:

$$E = f(\theta, \phi)F(\theta, \phi)\angle f_p(\theta, \phi) + F_p(\theta, \phi) \qquad (5.42)$$

where

$f(\theta, \phi)$	Field pattern of individual source
$f_p(\theta, \phi)$	Phase pattern of individual source
$F(\theta, \phi)$	Field pattern of array of isotropic source
$F_p(\theta, \phi)$	Phase pattern of array of isotropic source

In equation (5.42), the patterns are expressed as functions of both polar angles θ and ϕ to indicate that the principle of pattern multiplication applies to space patterns as well as to two-dimensional cases.

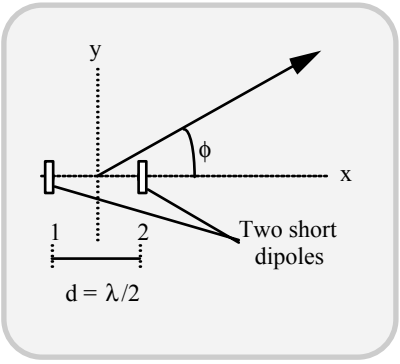


Fig. 5.13 Array of two non-isotropic point-source representation of an array of two short dipoles

For example, consider an array of two non-isotropic point sources as shown in Figure 5.13. Let each individual source pattern be given by

$$E_o = \cos(\phi) \tag{5.43}$$

This source pattern is produced by short dipoles oriented parallel to the y-axis and spaced $\lambda/2$ from each other as shown in Fig. 5.14(a). By the principle of pattern multiplication, the total (normalized) field is given by,

$$E = [\cos(\phi)] \times \{\cos[\frac{\pi}{2}\cos(\phi)]\} \tag{5.44}$$

Figures 5.14(a) and 5.14(b) show the element and array pattern respectively; and, Figure 5.14(c) shows the resultant pattern.

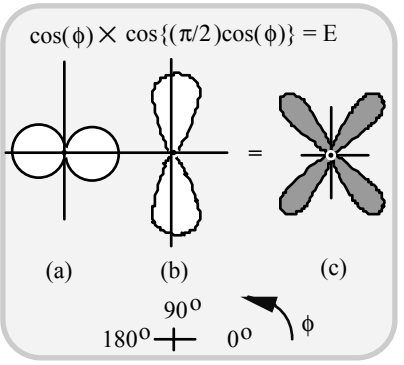


Fig. 5.14 Pattern multiplication for array of two non-isotropic similar point-sources. (a) Individual pattern; (b) array pattern and (c) total (resultant) pattern

5.2.9 Array of non-isotropic and dissimilar point sources

If the sources are dissimilar, the principle of pattern multiplication cannot be applied. On the other hand, the fields of the sources should be added at each pair of angles (θ, ϕ) for which the total field is calculated. Thus, for two dissimilar sources 1 and 2 situated on the *x*-axis with source 1 at the origin, and the sources being separated by a distance d as shown in Figure 5. 15, the total field is given by,

$$E = E_1 + E_2$$

$$= E_o \left\{ \begin{matrix} \left[f(\phi) + aF(\phi)\cos(\psi) \right]^2 + \\ \left[aF(\phi)\sin(\psi) \right]^2 \end{matrix} \right\}^{1/2} \angle\chi \qquad (5.45)$$

where $\angle\chi = \angle f_p(\phi) + \tan^{-1}\left(\dfrac{aF(\phi)\sin(\psi)}{f(\phi) + aF(\phi)\cos(\psi)} \right)$ and the field from source 1 is:

$$E_1 = E_o f(\phi) \angle f_p(\phi) \qquad (5.46)$$

Further, the field from source 2 is

$$E_2 = aE_o F(\phi) \angle [F_p(\phi) + d_r \cos(\phi) + \delta] \qquad (5.47)$$

where

E_o	Constant (representing an arbitrary amplitude)
a	Ratio of maximum amplitude of source 2 to source1 ($0 \le a \le 1$)

ψ	$d_r\cos(\phi) + \delta - f_p(\phi) + F_p(\phi)$
δ	Relative phase of source 2 with respect to source 1
$f(\phi)$	Relative field pattern of source 1
$f_p(\phi)$	Phase pattern of source 1
$F(\phi)$	Relative field pattern of source 2
$F_p(\phi)$	Phase pattern of source 2

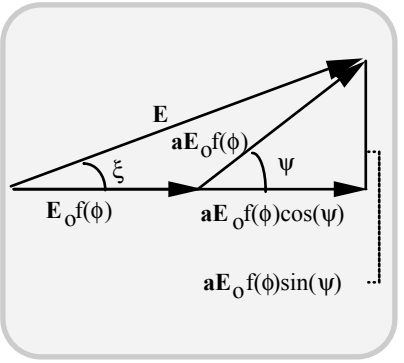


Fig. 5.15 Vector addition of the fields of two non-isotropic dissimilar point sources

The vector addition of the fields is shown in Figure. 5.15, where explicitly,

$$\xi = \xi^1 + f_{p1}(\phi) \tag{5.48}$$

with

$$\xi^1 = \tan^{-1}\left\{\frac{aF(\phi)\sin(\psi)}{f(\phi)aF(\phi)\cos(\psi)}\right\} \tag{5.49}$$

As an example, consider two non-isotropic dissimilar point-sources 1 and 2 having fields E_1 and E_2 given by

$$E_1 = \cos(\phi)\angle 0 \tag{5.50}$$

and

$$E_2 = \sin(\phi)\angle\psi \tag{5.51}$$

where $\psi = d_r\cos(\phi) + \delta$.

Suppose $d = \lambda/4$ and $\delta = \pi/2$. Then, it follows that

$$\psi = \frac{\pi}{2}[\cos(\phi) + 1] \tag{5.52}$$

Figure 5.16 shows the two individual patterns and the total pattern.

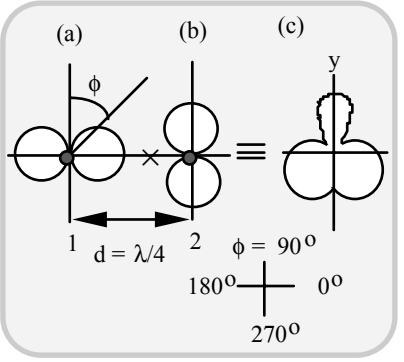


Fig. 5.16 Two non-isotropic dissimilar point sources ($\delta = 90°$). (a) and (b): Individual patterns and (c) total (resultant) pattern

5.3 LINEAR BROADSIDE ARRAYS WITH NONUNIFORM AMPLITUDE DISTRIBUTION

Linear broadside arrays can also be constructed with nonuniform current amplitude distributions across the elements as shown in Figure 5.17.

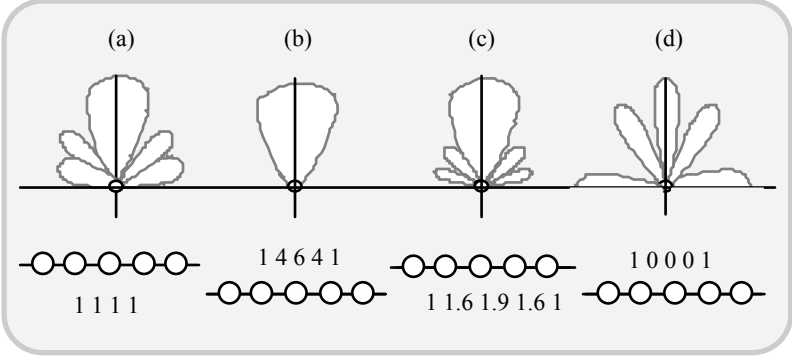


Fig. 5.17 Normalized field patterns of broadside arrays of five isotropic point sources spaced $\lambda/2$ apart. (a) Uniform distribution; (b) binomial distribution; (c) optimum distribution, and (d) edge distribution

There are typically four types of nonuniform amplitude distributions prescribed for broadside arrays. They are as follows:

- Uniform distribution
- Binomial distribution
- Edge distribution
- Optimum distribution

As an example, consider a broadside array of five isotropic point-sources with λ/2-spacing and excited with current distributions as indicated above. Figure 5.17 shows the normalised field patterns for the four types of nonuniform distribution from which the following can be observed. The binomial distribution gives no side lobes and the uniform distribution leads to the minimum beamwidth of the major lobe; but the corresponding pattern has large number of side lobes. This results from the intense edge field. On the other hand, the optimum distribution gives the compromise between the major lobe beamwidth and the side lobe levels. The optimum distribution indicated is known as the *Dolph-Tchebyscheff distribution*, which is considered in detail in the following section.

5.3.1 Linear arrays with optimum or Dolph-Tchebyscheff distribution

Consider a linear array of even number n_e or odd number n_o of isotropic point-sources of uniform spacing d and of the same phase as shown in Figure 5. 18. The individual sources have amplitudes A_0, A_1, A_2, …, etc., with the amplitude distribution being symmetrical about the centre of the array.

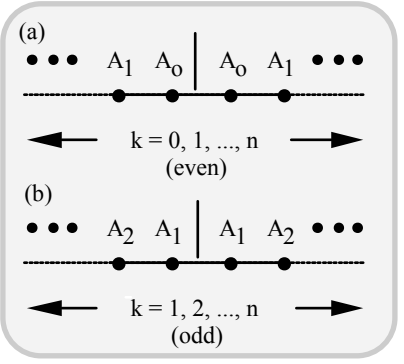


Fig. 5.18 Linear broadside arrays of n isotropic sources with uniform spacing for (a) n being even, and (b) n being odd

For an even number of elements, the total field E_{ne} can be specified as follows:

$$E_{ne} = 2A_0\cos(\psi/2) + 2A_1\cos(3\psi/2) + \ldots + 2A_k\cos\left[\frac{(n_e - 1)}{2}\psi\right] \quad (5.53)$$

where

$$\psi = \left(\frac{2\pi d}{\lambda}\right)\sin(\theta) = d_r \sin(\theta) \quad (5.54)$$

The total field E_{ne} can be rewritten as follows:

$$E_{ne} = 2\sum_{k=0}^{k=N-1} A_k \cos\left[\frac{(2k+1)}{2}\psi\right] \quad (5.55)$$

where $N = n_e/2$.

For an odd number of elements, the total field E_{ne} can be specified as follows:

$$E_{no} = 2A_0 + 2A_1\cos(\psi) + 2A_2\cos(2\psi) + \ldots + 2A_k\cos\left[\frac{(n_o - 1)}{2}\psi\right]$$

$$= 2\sum_{k=0}^{k=N} A_k \cos\left(2k\frac{\psi}{2}\right) \quad (5.56)$$

where $N = (n_o - 1)/2$. The series represented in equations (5.55) and (5.56) are truncated Fourier series with N finite number of terms. (Both equations (5.55) and (5.56) are polynomials of degree $(n_e - 1)$ and $(n_o - 1)$ respectively.)

In the Dolph-Tchebyscheff distribution, it can be shown that the coefficient of the pattern series can be uniquely determined so as to produce a pattern of minimum bandwidth for a specified side-lobe level. The procedure is as follows: Suppose a broadside array is of interest, in which case $\delta = 0$. Consider the following series:

$$\cos(m\psi/2) = \cos^m\left[\frac{\psi}{2}\right] - \frac{m(m-1)}{2!}\cos^{(m-2)}\left[\frac{\psi}{2}\right]\sin^2\left[\frac{\psi}{2}\right] + \frac{m(m-1)(m-2)(m-3)}{4!}\cos^{(m-4)}\left[\frac{\psi}{2}\right]\sin^4\left[\frac{\psi}{2}\right] \quad (5.57)$$

which is obtained by expanding $\cos(m\psi/2) = \mathrm{Re}[\{\cos(\psi/2) + j\sin(\psi/2)\}^m]$. Now, with the substitution of $x = \cos(\psi/2)$, $\sin^2(\psi/2) = [1 - \cos^2(\psi/2)] = (1 - x^2)$ and $T_m(x) = \cos(m\psi/2)$, the following polynomials in x can be elucidated:

$$\left.\begin{aligned}
T_0(x) &= 1 \\
T_1(x) &= x \\
T_2(x) &= 2x^2 - 1 \\
T_3(x) &= 4x^3 - 3x \\
T_4(x) &= 8x^4 - 8x^2 + 1 \\
T_5(x) &= 16x^5 - 20x^3 + 5x \\
T_6(x) &= 32x^6 - 48x^4 + 18x^2 - 1 \\
T_7(x) &= 64x^7 - 112x^5 + 56x^3 - 7x, \text{ etc.}
\end{aligned}\right\} \qquad (5.58)$$

Here, $T_m(x)$, m = 0,1,2, … are known as *Tchebyscheff (Chebyshev) polynomials*, and,

$$T_m(x) = 0 \text{ when } \cos(m\psi/2) = 0 \text{ or } (m\psi/2) = (2k - 1)\pi/2 \qquad (5.59)$$

Or

$$x = x' = \cos[(2k - 1)\pi/2m] \qquad (5.60)$$

The Tchebyscheff polynomials $T_m(x)$ as functions of x, for various values of m, are tabulated in [5.9]. (*Note*: The name Tchebyscheff is also spelt as *Chebyshev* in the literature.) From the associated plots of $T_m(x)$ *versus* x, indicated in [5.9], the following characteristics can be observed:

- All the $T_m(x)$ polynomials pass through the point (1, 1)
- For values of x in the range $(-1 \le x \le +1)$, all the polynomials lie between ordinate value of + 1 and – 1. Further, all roots occur in this range, and all maximum or minimum values in this range are ± 1.

Using the Tchebyscheff polynomials (T-polynomial), a linear broadside array can be optimized as follows: The underlying principle assumes the field pattern of a linear broadside array of sources as a polynomial of degree equal to the number of sources less 1. This array polynomial is then identically equated to the Tchebyscheff polynomial of the same degree. For example, for an array of 6 sources, the array polynomial (which is of degree 5), is equated to the T-polynomial of degree 5 (as illustrated in Figure 5.19).

Let the ratio of the main lobe maximum to the minor lobe level be equal to R. Then the point (x_0, R) on the $T_5(x)$ polynomial would correspond to the main-lobe maximum, while all the minor lobes have a maximum value of 1. Further, the roots of the polynomial correspond to the nulls of the field pattern. An important property of the T-polynomial is that, once the ratio R is specified, the beamwidth to the first null ($x = x_1'$) is minimised; then, with a specified beamwidth, the ratio R is maximised.

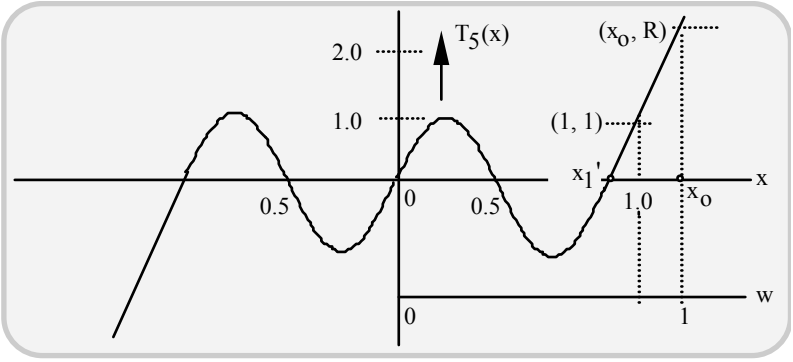


Fig. 5.19 Tchebyscheff polynomial of 5th degree. The transformed w-scale is also indicated

The field polynomials specified by equation (5.55) or (5.56), are equated to $T_{n-1}(x)$, where n is the number of sources n_e or n_o. Suppose m = (n – 1) and,

$$T_m(x) = T_{n-1}(x_0) = R \tag{5.61}$$

For R > 1 it can be seen from Figure 5.19 that $x_0 > 1$. However, in order to avoid the constrain stipulated by $x = \cos(\phi/2)$, $|x| < 1.$, a new variable w can be introduced, where

$$w = x/x_0 \tag{5.62}$$

Now, w can be set equal to $\cos(\psi/2)$ so that $-1 \le w \le +1$. Hence, the polynomial of equation (5.55) or 5.(56) can be expressed as a polynomial in w or in terms of (x/x_0). Equating the array polynomial E_n to the Tchbyscheff polynomial $T_{n-1}(x)$, it follows that,

$$E_n(x/x_0) \equiv T_{n-1}(x) \tag{5.63}$$

By matching term by term, the coefficients of the array polynomial can be obtained from equation (5.63). Hence, the amplitude distribution can be ascertained.

To prove that the Tchebyscheff polynomial gives the optimum distribution, consider any other polynomial $P_5(x)$ of degree 5 that passes through (x_0, R) as well as the highest root x'_1 in Figure 5.19. This polynomial lies between – 1 and + 1 for all smaller values of x. If $P_5(x)$ is less than $\pm$ 1 at its maxima and minima, then $P_5(x)$ would give a smaller side-lobe level for this beamwidth, and $T_5(x)$ will not be optimum. Since $P_5(x)$ lies between + 1 and – 1 in the range $-x'_1 \leq x \leq +x'_1$; it must intersect the curve $T_5(x)$ in at least (5 + 1) = 6 points, including (x_0, R). The polynomials of the same degree m = 5, which intersect (m + 1) ≡ (5 + 1) = 6 points must be of the same order. Therefore,

$$P_5(x) = T_5(x) \tag{5.64}$$

which implies that $T_5(x)$ is the optimum polynomial.

Example 5.1
Design a linear broadside array consisting of eight isotropic point-sources with optimal distribution. Assume the first side lobe level as – 26 dB below the main beam.

Solution
The design in question warrants the use of 8 in-phase point sources spaced λ/2 apart with a side-lobe level of 26 dB below the main-lobe level and the array should produce minimum beamwidth first nulls. This can be achieved with an appropriate amplitude distribution across the antenna point source elements.

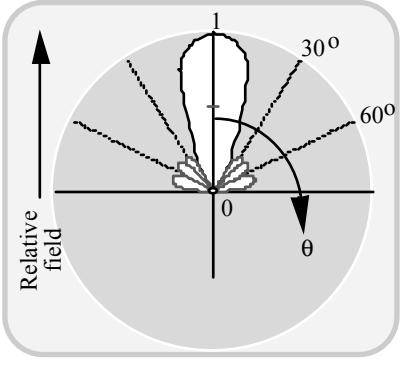


Fig. 5.20 Relative field pattern of broadside array of eight isotropic point-sources spaced λ/2 apart, with optimum Tchebyscheff distribution

Since $20\log_{10}(R) = 26$ dB, R = 20 and the required T-polynomial of degree (m – 1) with m = 8 is $T_7(x)$. That is, $T_7(x_0) = 20$. With relevant substitutions, it follows that,

$$x_o = (1/2)[1+(R^2 - 1)^{½}]^{1/m} + R - [(R^2 - 1)^{1/2}]^{1/m} \quad (5.65)$$
$$= 1.15$$

Further, with $w = x/ x_0$ and $m = 8$ in equation (5.53), it follows that

$$E_{m=8} = 64A_3(x/x_0)^7 + (16\ A_2 - 112A_3)\ (x/x_0)^5 + \ldots \infty \quad (5.66)$$

Further, from equation (5.58), $T_7(x) = (64x^7 - 112x^5 + 56x^3 - 7x)$, which, when identically equated to E_8 term-by-term, yields matched terms of (x^7) as follows: $64A_3/x_0^7 \equiv 64$; and, substituting $x_0 = 1.15$, $A_3 = 2.66$.

Likewise, equating the other coefficients of the terms of like degree, the following results are obtained: $A_2 = 4.56$; $A_1 = 6.82$ and $A_0 = 8.25$. The corresponding relative amplitudes of eight sources are: {1, 1.7, 2.6, 3.1,2.6, 1.7, 1}. Using these relative amplitudes and the other relations, namely $\psi/2 = d_r\sin(\theta)/2$, $\cos(\psi/2) = w = (x/ x_0)$, so that $x = x_0\cos[d_r\sin(\theta)]/2$, the field pattern $E_8(x)$ is obtained as depicted in Figure. 5.20.

5.4 PLANAR AND VOLUME ARRAYS

A linear array is intended to steer the beam in one dimension either by varying the phase difference (δ) between the adjacent elements or by changing the frequency, or by varying the amplitudes of the signal exciting the elements. A two-dimensional extension of linear array concept leads to what is known as the *planar array*. Such a planar array can steer the beam in two dimensions.

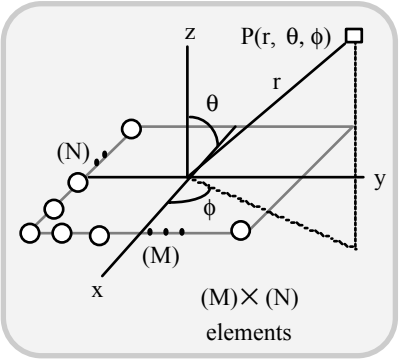


Fig. 5.21 Planar array of (M) × (N) -elements placed in the horizontal xy-plane

Consider a *planar array* consisting of M elements along the x-direction and N elements along the y-direction in the z-plane as shown in Figure 5.21. Let the

$(mn)^{th}$ element be excited by a source $V_{mn}\exp\{j(m\psi_x + n\psi_y)\}$ where the steering phases ψ_x and ψ_y are the incremental phase shifts between elements in the x- and y -directions respectively. Then the array factor of the planar array of (M × N) elements is given by:

$$S(T_x, T_y) = \sum_{m=1}^{M-1}\sum_{n=1}^{N-1} V_{mn}\exp\left\{j\frac{2\pi mb}{\lambda}\left(T_x + \frac{\psi_x}{2\pi b/\lambda}\right) + \frac{2\pi nd}{\lambda}\left(T_y + \frac{\psi_y}{2\pi d/\lambda}\right)\right\} \tag{5.67}$$

where $T_x = [\sin(\theta)][\cos(\phi)]$; $T_y = [\sin(\theta)][\sin(\phi)]$; b is the spacing between elements in the x-direction; and, d denotes the spacing between elements in the y-direction.

If the array elements are excited uniformly in amplitude, so that $V_{mn} = V_{oo}\exp\{j(m\psi_x + n\psi_y)\}$, then the array factor becomes,

$$S_o(T_x, T_y) = \sum_{m=1}^{M-1}\exp\left\{j\frac{2\pi mb}{\lambda}\left(T_x + \frac{\psi_x}{2\pi b/\lambda}\right)\right\} \times \sum_{n=1}^{N-1}\exp\left\{j\frac{2\pi nd}{\lambda}\left(T_y + \frac{\psi_y}{2\pi d/\lambda}\right)\right\}$$

$$= S_x S_y \tag{5.68}$$

where

$$S_x = \sum_{m=1}^{M-1}\exp\left\{j\frac{2\pi mb}{\lambda}\left(T_x + \frac{\psi_x}{2\pi b/\lambda}\right)\right\} \tag{5.69}$$

and it corresponds to the array factor of linear array of M elements in the x-direction. Likewise,

$$S_y = \sum_{n=1}^{N-1}\exp\left\{j\frac{2\pi nd}{\lambda}\left(T_y + \frac{\psi_y}{2\pi d/\lambda}\right)\right\} \tag{5.70}$$

which refers to the array factor of linear array of N elements in the y-direction.

The total field of field of the planar array is obtained by pattern multiplication principle:

$$E = E_e(T_x, T_y) \times [S_x \times S_y] = E_e[\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi)] \times [S_x \times S_y] \tag{5.71}$$

where $E_e(T_x, T_y)$ depicts the far-field pattern of the individual antenna element of the array.

From equation (5.70), it can be seen that by properly varying the steering phases, namelyψ_x, and ψ_y, and/or by appropriately changing the amplitude V_{mn} of the elements, the main beam of the antenna can be made to scan the whole space.

A *volume array* (also known as a *three-dimensional array*) is a logical extension of a planar array, which can be used for a three-dimensional coverage (Figure 5.22). In practice, they can be configured conformally on three-dimensional surfaces such as cube, cone, sphere or cylinder. The number of elements and the way they are excited enable characteristic radiation patterns that are compatible for different applications. Volume arrays find extensive applications in rockets, missiles, aircraft and other airborne vehicles.

The array factor of a volume array is given by,

$$S(\alpha,\beta,\theta) = \sum_{m=1}^{M-1}\sum_{n=1}^{N-1}\sum_{p=1}^{P-1} A_{mnp} \qquad (5.72)$$

where $A_{mnp} = a_{mnp}\exp\{jk[ps_x\cos(\alpha) + ms_y\cos(\beta) + ns_z\cos(\theta)]\}$ and (M, N and P) denote the number of elements.

Further, (s_x, s_y and s_z) represent the spacing between consecutive elements in the x-,y- and z-directions respectively; and (α, β and θ) are, respectively, the angles made by the radius vector (drawn from the origin to the distant point) with respect to the x-, y- and z-axes.

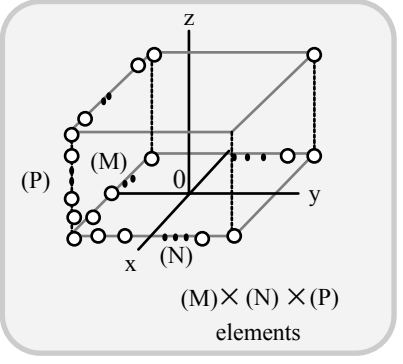


Fig. 5.22 Three-dimensional array made of {M × N × P} set of elements in the three orthogonal directions

The coefficient a_{mnp} depicts the excitation current in the $(mnp)^{th}$ element. Suppose this excitation current is explicitly taken as,

$$a_{mnp} = A_{mnp}\exp\left\{-jk\left(m\psi_x + n\psi_y + p\psi_z\right)\right\} \tag{5.73}$$

where ψ_x, ψ_y and ψ_z are the progressive phase shifts in the x-, y- and z-directions respectively. Then,

$$S(\alpha,\beta,\theta) = \sum_{m=1}^{M-1}\sum_{n=1}^{N-1}\sum_{p=1}^{P-1} A_{mnp}\exp\left\{\begin{matrix}-jk\left(m\psi_x + n\psi_y + p\psi_z\right) + \\ jk\left(\begin{matrix}ms_x\cos(\alpha) + ns_y\cos(\beta) + \\ ps_z\cos(\theta)\end{matrix}\right)\end{matrix}\right\} \tag{5.74}$$

The array factor $S(\alpha, \beta, \theta)$ given by equation (5.74) when multiplied by the element factor $E(\theta, \phi)$ gives the total radiation pattern of the array. If the array consists of a large number of elements, usually the array factor is more directional than the element pattern. Hence the main beam would essentially be determined by the array factor.

For an equally-spaced uniform cubic array shown in Figure 5.22, the array factor $S(\alpha, \beta\ \theta)$ can be written in the form:

$$S(\alpha, \beta, \theta) = S_x \times S_y \times S_z \tag{5.75}$$

where S_x, S_y, and S_z are the array factors of the linear arrays in the x-, y- and z-directions respectively.

The array factor given by equation (5.67) for a planar array has a periodicity along T_x and T_y axes. As such, the shape of the main beam and the associated side-lobes will be repeated every λ/b and λ/d intervals in the T_x and T_y axes, respectively. Such repetitions of the main lobe are designated as *grating lobes*, which lead to more than one principal beam plus a set of associated secondary maxima in the radiation pattern. The grating lobes can also exist in the case of cubic arrays.

Arrays can also be designed on surfaces which are non-planar. Such arrays are called *conformal arrays*, and the word "conformal" signifies that the array conforms to the shape of the object on which the array is to be placed. The conformal arrays are very useful for surface-mounted applications in aircrafts, missiles, rockets and ships, which have unique surface geometry that are optimised to meet the aero- or hydro-dynamic constrains.

5.5 FEED TECHNIQUES FOR ARRAY ANTENNAS

There are three possible ways of feeding the elements of an antenna array with the signal energy. They are as follows;

(a) *End-feeding*: Here, the feeder enters the bottom of the antenna; the power to radiators is tapped off the cable as it ascends the antenna column. (Figure 5.23(a))
(b) *Center-feeding*: In this case, the main cable enters the array at its mid-height and splits into one descending and one ascending feeder within the column from which power is tapped off to the radiators in sequence. This method is illustrated in Figure 5.23(b)
(c) *Parallel-feeding*: This feed arrangement has feed cables split regularly and form a fanlike arrangement spreading over the height of the column as depicted in Figure 5.23(c). It is readily seen that all the radiating elements are thus fed in parallel; or, more importantly, the electrical path lengths to the radiators from the common feeder are all essentially equal.

In end- as well as centre-feeding, the taps off the feed lines are arranged to be at intervals of one wavelength at the design frequency (so that all the radiators are correctly phased at this frequency). End- and centre-feeding are invariably used in slot-arrays so as to minimise mechanical size, whereas parallel-feeding is almost always exploited in broadband panel arrays.

The feed arrangements shown in Figure 5.23 are suitable for arrays made of discrete dipoles, or for patch elements etched on a panel or for slots cut on a surface. In designing the feed arrangements that accommodate the discrete array elements, sufficient space should be provided so as to facilitate the feeding cables on the structure.

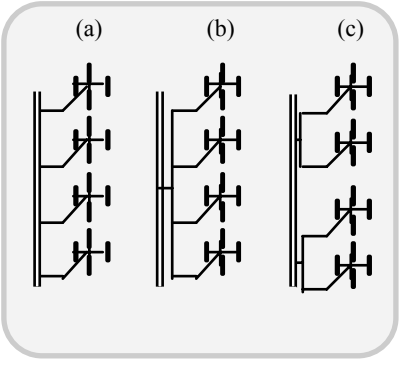


Fig. 5.23 Feeding techniques for antenna arrays: (a) End-fed array; (b) centre-fed array and (c) parallel-fed array

Mechanically, the feed systems facilitated on panel arrays are relatively straight forward since there is enough space to accommodate semi-flexible coaxial feeders formed into cable harnesses.

With slot-arrays, the slotted external cylinder would commonly form the outer conductor of a transmission line. The slots on this external cylinder are

inductively coupled to an internal tabular conductor located coaxially within the outer cylinder. This inner conductor of the line forms a centre-feed arrangement; and, a third conductor located inside the lower half of the inner conductor may be needed to form a triaxial line connecting to the common feed-point at the centre of the antenna.

5.5.1 Vertical radiation patterns of arrays *versus* feed arrangements

As regards to the vertical radiation pattern of the array *vis-à-vis* the type of feed arrangement adopted, the following can be observed. Intuitively, it follows that the relative phases of the radiating elements would vary with frequency more in the end-fed system than in the centre-fed system (as a result of relative physical disposition/placement of the antenna elements). Therefore, it can be expected that the phase variation will be even more with a parallel-feed arrangement. As a result, the associated vertical radiation patterns will significantly exhibit distinct frequency-dependent variations in comparison with the patterns of the arrays of the other three schemes of feed system indicated above.

5.6 ARRAYS WITH PARASITIC ELEMENTS

In the foregoing sections, it was assumed that all radiation elements are driven, meaning that each radiating element is supplied with signal power by means of a transmission line feed system. However, directional arrays can also be designed with elements in which currents (of required amplitude and phase) can be induced by the EM fields of a proximally located driven element.

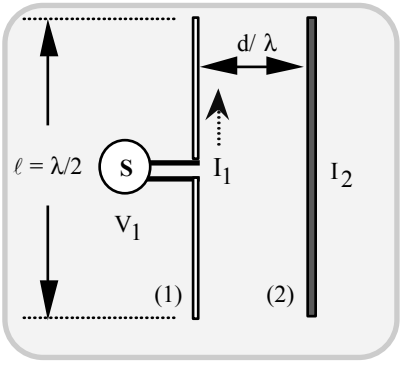


Fig. 5.24 An array made of one driven element (1) driving a parasitic element (2)

The elements, which are not connected to the source (and depend on the energisation from an active or driven element), are known as *parasitic elements*. Considering an array in free-space made of one driven dipole and a parasitic element as shown in Figure.5.24, the relevant analysis due to Brown [5.10] in evaluating the radiation characteristics is based on the following circuit relations for the elements:

$$\begin{bmatrix} V_1 \\ 0 \end{bmatrix} = \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \tag{5.76}$$

where V_1 is the voltage applied to the active element-1 (which drives a current I_1 into the element); I_2 is the induced current in the parasitic element-2. Further, Z_{11} and Z_{22} are the self-impedance of elements-1 and -2 and $\{Z_{12}, Z_{21}\}$ is the set of mutual impedance between these elements. Specifying these impedance in their complex form made of a real part $\{R_{11}, R_{12}, R_{21}$ or $R_{22}\}$ and a corresponding reactive part $\{X_{11}, X_{12}, X_{21}$ or $X_{22}\}$, it follows from equation (5.76) that $I_2 = I_1|Z_{11}/Z_{22}|\angle\xi$ where $\xi = [\pi + \arctan(X_{22}/R_{12}) - \arctan(X_{22}/R_{22})]$.

The electric field intensity at a large distance from the array as a function of the azimuthal angle ϕ can be written (in terms of $d_r = 2\pi/\lambda$) as,

$$E(\phi) = k\{I_1 + I_2 \angle[d_r\cos(\phi)]\}$$

$$= kI_1\{1 + |Z_{11}/Z_{22}| \angle[\xi + d_r\cos(\phi)]\} \tag{5.77}$$

Further, the driving-point impedance Z_1 of the driven element is specified as follows:

$$Z_1 = Z_{11} - (Z_{12}^2/Z_{22})\cos(\psi) + R_{1L} \tag{5.78}$$

whose real part can be written as

$$R_1 = R_{11} - (R_{12}^2/R_{22})\cos(\psi) + R_{1L} \tag{5.79}$$

where $\psi = [2 \times \arctan(X_{22}/R_{12}) - \arctan(X_{22}/R_{22})]$ and R_{1L} depicts the effective loss resistance, if any is present. Hence, the corresponding power input to the active element can be written as $P = I_1^2 R_1$. Therefore, $I_1 = (P/R_1)^{1/2}$, which can be substituted in equation (5.77) so that,

$$E(\phi) = k[P/\, R_{11} + R_{1L} - (R_{12}^2/R_{22})\cos(\psi)\,]^{1/2} \times [1 + |Z_{11}/Z_{22}| \angle\{\xi + d_r\cos(\phi)\}$$

$$\tag{5.80}$$

Suppose, a single dipole is considered. Its electric field intensity at the same distance as the parasitic element under discussion can be written as,

$$E_s(\phi) = kI_o - k[P/(R_{oo}+R_{oL})]^{1/2} \tag{5.81}$$

where R_{oo} is the self-resistance and R_{oL} is the loss-resistance of the single λ/2-element.

Now, the gain of the array can be specified with respect to a single, λ/2-antenna excited with the same power. Since $R_{oo} = R_{11}$ and letting $R_{oL} = R_{1L}$, it follows that,

$$G(\phi) = \{(R_{11} + R_{1L}) / [\, R_{11} + R_{1L} - |\, R_{12}^2 / R_{22} \,|\cos(\psi)]\} \times$$

$$[1 + |Z_{12}/Z_{22}| \angle(\xi + d_r \cos\phi)] \qquad (5.82)$$

The magnitude of the current in the parasitic element and its phase relation to the current in the driven element depends on "tuning". That is, by adjusting the length of the parasitic element, it can be made to pose an inductive or capacitive reactance. Hence, when the λ/2 parasitic element is tuned to be inductive (with its length adjusted to be larger than the resonant length), it will act as a *reflector*, so that the beam pattern will be directed away from it towards the active element.

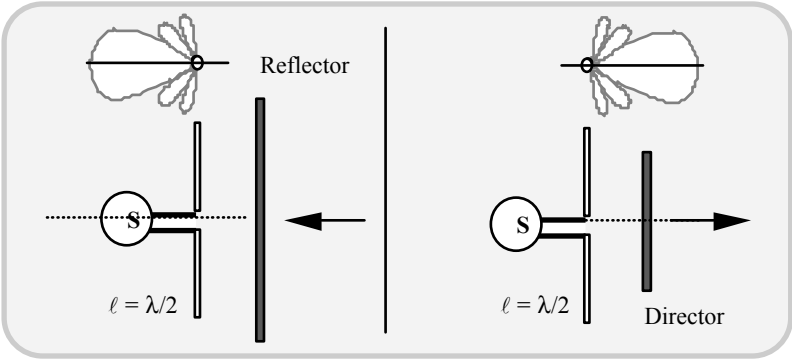


Fig. 5.25 "Reflector" and "director" roles of a parasitic element driven by an active element

When the parasitic element is tuned to be capacitive (by adjusting its length to be shorter than the resonant length), this element will act as a "director" and let the radiation beam be directed along its location with respect to the disposition of the active element. The reflector and director profiles of the parasitic element in an array are illustrated in Figure 5.25.

A typical horizontal plane pattern for a 3-element array constituted by a driven element plus a director and a reflector is shown in Figure. 5.26. In general a parasitic array with closely spaced elements has a small driving-point radiation resistance and relatively narrow beamwidth.

In summary, for a given length of the active element, the parasitic element would behave as a reflector, if its length is greater than that of the active element.

Otherwise, its role becomes a director. In Figure 5.26, the array is constituted by the three elements – an active radiator, a reflecting parasite and a directing parasite. The result will be an end-fire beam as illustrated. An array can be designed with more than three elements. That is, by using one actively excited element plus a number of reflectors and directors (constituting a fishbone structure), extremely narrow beams can be realised.

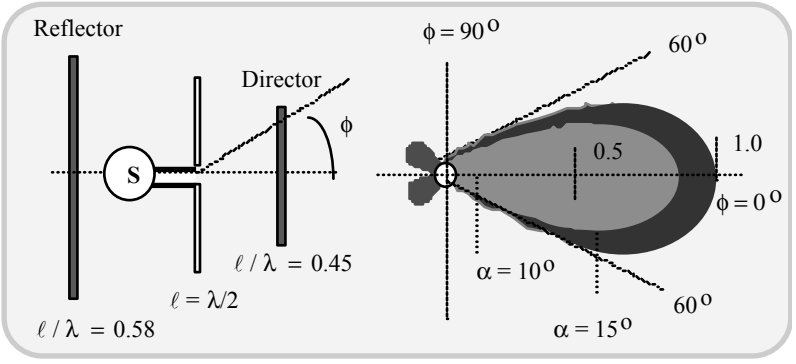


Fig. 5.26 (a) 3-element array made of a driven element plus a director and reflector and (b) relative horizontal plane field patterns at different elevation angles, α

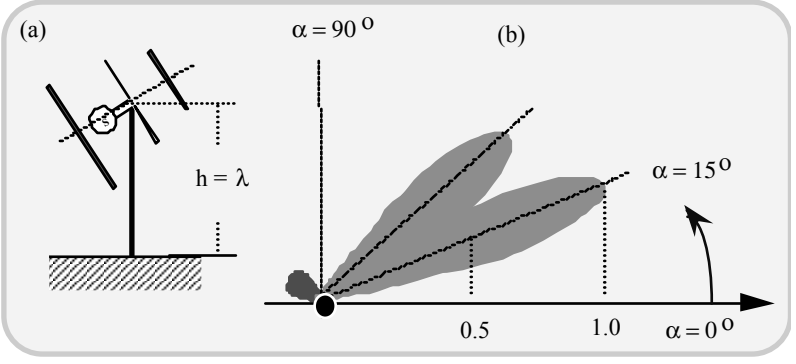


Fig. 5.27 (a) A 3-element parasitic antenna placed at a height of h = l above the ground and (b) the vertical pattern of the array (for ϕ = 0)

The radiation pattern of an array could be influenced by the proximity of the array structure to the ground surface. For example, an array, such as the one depicted in Figure 5. 27, has a vertical pattern that depends on the height of the array structure above the ground surface. Assuming that an array is placed at a height h = λ above

the ground-plane, the approximate vertical pattern for $\phi = 0$ is presented in Figure 5.27 for comparison against the pattern obtained in the absence of the ground-plane.

The radiation pattern in the vertical plane of the antenna under discussion could be further altered, if the size of the ground-plane is finite. This arises as a result of currents induced at the edges of such a ground-plane and would cause secondary radiation patterns that would influence the array pattern to an extent as decided by the height of the array above the ground-plane. This consideration is important in mobile systems in which the array can be located in a space-constrained region with a finite-sized ground-plane (such as in airborne vehicles or in an automobile).

5.6.1 Yagi-Uda array

A typical end-fire array that uses parasitic elements is well-known as the *Yagi-Uda array* [5.11, 5.12]. The practical structure of a Yagi-Uda antenna is conceived with multiple directors and it yields end-fire directivity at the expense of bandwidth. The lengths of the reflector and the directors, as well as their relative spacing, are for the most part a compromise with such factors as gain, input impedance, front-to-back radiation ratio, magnitude and extent of minor lobes and the bandwidth. A total antenna length of 6λ has been indicated as a reasonable limit (though arrays of thirty or forty elements have been built as a "fish-bone" structure). Essentially, the compactness (size-constraint), small side lobe level and desired end-fire directivity (beamwidth) are primary design factors in choosing the overall Yagi-Uda antenna configuration.

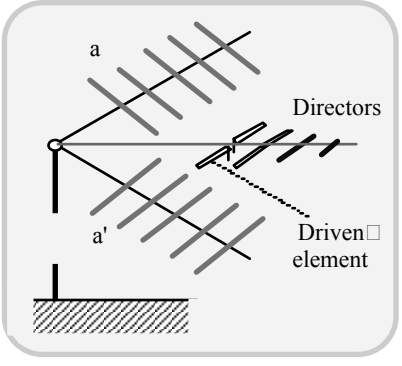


Fig. 5.28 Corner reflector based Yagi-Uda array (a-a′: Corner reflector structure)

TheYagi-Uda structure can be extended to realise the following types of arrays as well:

- *Arrays of Yagi-Uda arrays*: The gain from a single Yagi-Uda array is often limited by length to a value of about 30 (15 dB). Hence, arrays of such arrays have been suggested to achieve further gain. Ideally, two such arrays should provide a total gain of 18 dB and four arrays should provide 21 dB. However, the structural design and compactness should be concurrent factors in the design of these configurations
- *Yagi-Uda arrays plus a corner-reflector*: The corner reflector is inherently a wideband antenna capable of yielding a substantial gain. Considering the Yagi-Uda array, the effectiveness of its reflector parasite is reduced, if the element spacing is increased. (Such an increase in the d/λ is inevitable at lower wavelength (or higher frequency) operations curtailing the wideband capability of the Yagi-Uda array.) To alleviate this problem, a corner-reflector plus Yagi-Uda hybrid structure has been conceived (and widely adopted in UHF TV applications). It is illustrated in Figure 5.28

6 *Yagi-Uda array providing circular polarization*: The Yagi-Uda structure can be modified to get circularly-polarised radiation. For this purpose the Yagi-Uda array can be constructed with each of its elements made of a pair of crossed (orthogonal) elements mounted on the same boom as illustrated in Figure 5.29.

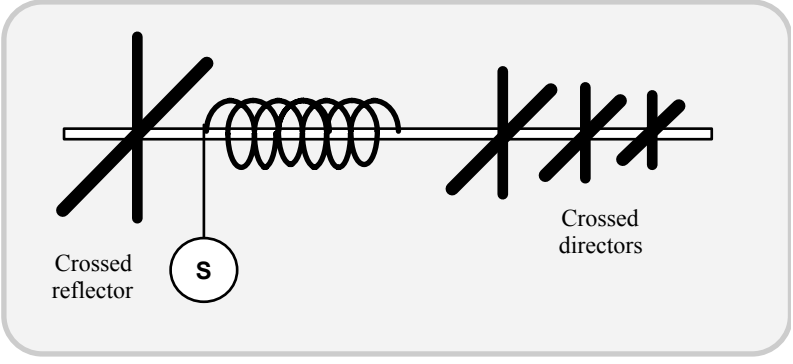


Fig. 5.29 Monofilar axial-mode helical antenna adopted as an active element in a Yagi-Uda array for circular polarisation

In general, a well-designed Yagi-Uda antenna provides increased directive gain as the length of the boom is increased supporting additional elements. Further, as shown in Figure 5.29, while the directors and the reflector are crossed dipoles, the active element is an axial mode helical antenna that symmetrically excites the crossed-dipole elements.

The arrangement shown in Figure 5.29 facilitates a directive, circularly-polarised, end-fire radiation along the boom axis. In Table 5.1, some guidelines to achieve the required gain, as a function of the boom-length and number of elements required, are presented. The gain (expressed in dBd) refers relative to a dipole antenna.

Table 5.1 Gain *versus* boom length and number of elements in typical Yagi-Uda antennas

Gain (dBd)	Boom length in wavelengths	Number of elements
3.0-4.5	0.10-0.25	2
4.5-6.5	0.15-0.35	3
6.0-7.5	0.30-0.50	4
7.0-8.5	0.50-0.80	5
8.5-10.0	0.80-1.20	6
9.5-10.5	1.20-1.50	7
10.0-11.0	1.50-2.00	8
11.0-12.0	2.00-2.50	9-10
12.0-13.0	2.50-3.50	11-13
13.0-14.0	3.50-4.50	14-18

5.6.2 Planar Yagi antenna-like array

Described in [5.13] is a planar array developed on the basis of the Yagi-Uda array concept. It is an array structure capable of tilting the beam and beamforming as a result of integration of its passive elements with an active device (such as a *high electron-mobility transistor*, HEMT).

The configuration of a planar Yagi antenna is illustrated in Figure 5.30. A typical design of the array shown (in Figure 5.30) at X-band frequencies, yields E- and H-plane patterns as illustrated in Figure 5.31 It consists of an active patch excited by an HEMT-based source and three directors. The patch is fabricated on a dielectric substrate and necessary isolation from ground-plane.

The design of the array accounts for relevant conditions for the self-oscillation of the active device. The array dimensions are specified as per conventional dipole Yagi-Uda array design. That is, the antenna sizing (dimensions of the elements and their spacings relative to the operating wavelength) is done as per passive Yagi-Uda arrays.

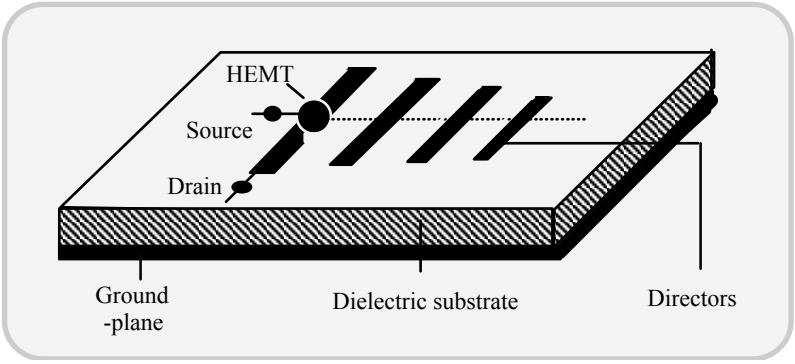


Fig. 5.30 A planar Yagi-Uda array antenna with an active device (HEMT) integration

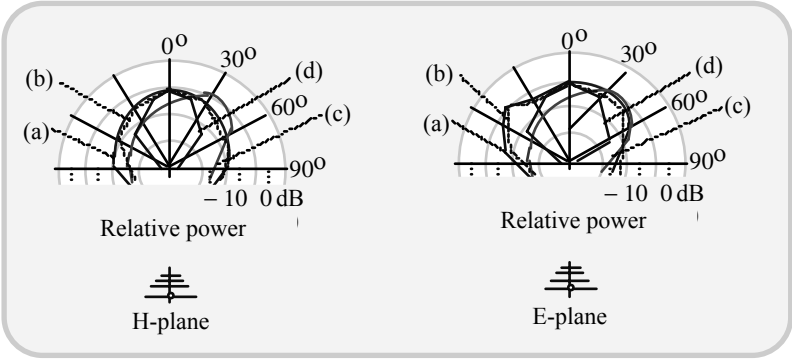


Fig. 5.31 Radiation patterns of a patch Yagi-Uda antenna designed with: (a) No directors; (b) single director; (c) two directors and (d) three directors

The general characteristics of the type of antenna shown in Figure 5.30 are as follows:

- With the active antenna alone, the 3-dB beamwidth is about 100° in the H-plane and about 70° in the E-plane. The beam becomes narrower in the H-plane with the addition of directors. (For example, the 3-dB beamwidth is about 85° with a single director and reduces to about 55° and 45° with two and three directors.) In the E-plane, however, there is no significant change in 3-dB beamwidth with the addition of directors

- In the H-plane, the addition of directors let the beam to tilt from the broadside direction to an extent of about 30°
- Cross-polarization level is about 20 dB down from the maximum in both planes
- The director elements also play a role in controlling the oscillation frequency. In the test antenna used in [5.13], a 50-MHz change per parasitic element used is reported.

5.6.3 Slot/aperture arrays

An array antenna can be conceived with a set of slots or apertures on a surface, with appropriate EM excitation. As discussed in Chapter 4 such slots and/or apertures can be realised on a flat, cylindrical or spherical surface. Normally, they are formed conformal to the body surface. For example, the nose cone of a missile/rocket system, an aircraft body, etc. are locales where such conformal arrays of slots/apertures have been popularly adopted.

The slots/apertures are excited by a compatible transmission or resonator system backing slot/aperture set. Examples of traditional slots/aperture arrays are as follows:

Slot-array on a waveguide

Shown in Figure 5.32 is a longitudinal slot on the broadside wall of a rectangular waveguide supporting the dominant mode TE_{10}. The current flowing along the longitudinal direction on the broadside upper wall is given by:

$$\mathbf{J}_S \propto \mathbf{u}_x \cos\left(\frac{\pi x}{a}\right) - j\frac{a\beta}{\pi}\mathbf{u}_z \sin\left(\frac{\pi x}{a}\right) \tag{5.83}$$

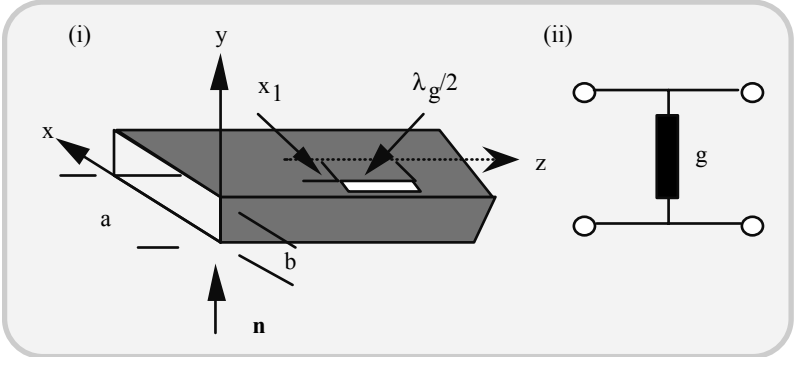


Fig. 5.32 (i) A longitudinal $\lambda_o/2$-slot cut on the broad side wall of a rectangular waveguide and (ii) its equivalent circuit.

At the centre where x = a/2, the current is entirely in the z-direction. Hence, a narrow slot formed at the centre does not cut any current flow lines and is not, therefore, excited. However, if the slot is placed away from the centre, say at x = x_1, current flow lines are intercepted and the slot is excited. This excitation increases with increasing x_1. The relevant normalized conductance of the slot is given by:

$$g = 2.09 \frac{\lambda_g}{\lambda_o} \frac{a}{b} \cos^2\left(\frac{\pi\lambda_o}{2\lambda_g}\right) \sin^2\left(\frac{\pi x_1}{a}\right) \qquad (5.84)$$

where λ_g is the guide wavelength and λ_0 is the free-space wavelength; and g acts as a shunt conductance across the waveguide.

There are two possible types of arrays of slots on a waveguide, namely, the *resonant* and the *non-resonant* array. The resonant array is designed on the broadside wall and uses a slot spacing of $\lambda_g/2$, where λ_g is the guide wavelength. The non-resonant array can be designed to have a main lobe at any angle with respect to the normal to the broadside wall of a guide, but in the plane that contains the array axis and the normal **n**.

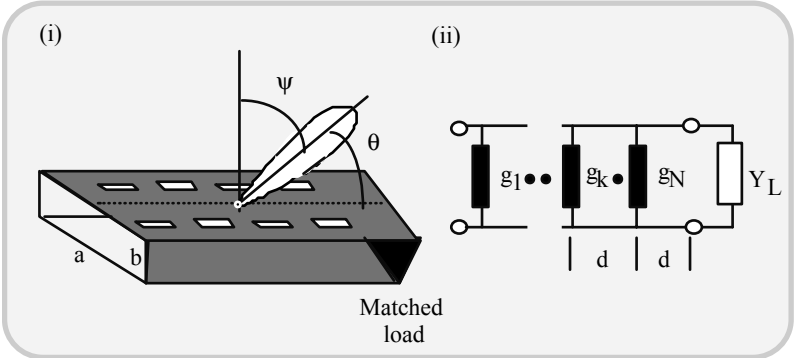


Fig. 5.33 (i) An array of resonant slots on a waveguide and (ii) its equivalent circuit

Using a resonant array, to achieve radiation in the broadside direction, the slots must all be excited in-phase, which can be done by making the spacing between consecutive slots equal to $\lambda_g/2$. In order to avoid more than one main lobe in the visible space (or sector of interest), the spacing must be less than λ_o for a broadside array. However, in a waveguide $\lambda_g > \lambda_o$. Therefore, the spacing cannot be $\lambda_g/2$. This difficulty is overcome in the following manner. From waveguide theory, it is known that the surface current in the x-direction is of opposite sign on adjacent sides of the centre line of the broadside wall of the rectangular waveguide. So, the slot excitation can be reversed in sign by off-setting it on the adjacent side of the

centre line. Hence a slot-array, as shown in Figure 5.33(i), where every other slot is on the opposite side of the centre line, and with a slot spacing of $\lambda_g/2$, can be designed. Figure 5.33(ii) represents the equivalent circuit of such an array, and $g_e = \sum_{n=1,\ldots N}(g_n)$ is the equivalent shunt conductance of the array.

If a_n is the desired excitation coefficient for the n^{th}-element of the array, for example, with excitation coefficients corresponding to Tchebysheff (or some other form of distribution), then the associated conductance g_n is given by,

$$g_n = K a_n^2 \tag{5.85}$$

where K is chosen such that $g_e = 1$; that is,

$$K \sum_{n=1}^{N} a_n^2 = 1 \tag{5.86}$$

This condition will enable the array matched to the input of the waveguide for maximum radiated power. Once g_n is specified, the slot off-set x_1 can be found for the n^{th} slot by using equation (5.84).

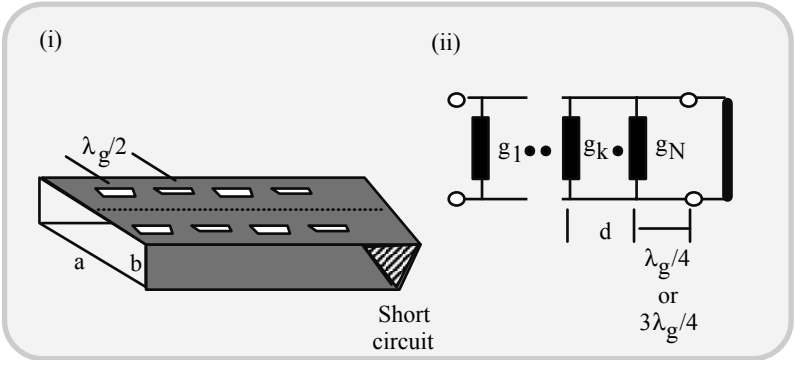


Fig. 5.34 (i) Non-resonant slot-array on a waveguide and (ii) its equivalent circuit

The broadside resonant array formed on the broadside wall of a waveguide is a narrow-band radiator. The reason is as follows: The spacing between the short-circuit and the last-slot (designed as $\lambda_g/4$) would change with frequency. As such, as the frequency changes, it will not emulate any longer an open-circuit condition. Therefore the short-circuit will reflect the incident EM energy with an appreciable amplitude. A *non-resonant array* is usually designed for operations other than broadside radiation. It consists of resonant longitudinal broadside wall slots, and the waveguide is terminated in a normalized load Y_L. The array and its equivalent circuit are shown in Figure 5.34.

Let d be the slot spacing and x_n the off-set measured from the centre line x = a/2. Then the location of the n^{th} slot is given by,

$$\mathbf{r}_n = nd\mathbf{u}_z + x_n\mathbf{u}_x \tag{5.87}$$

Let the excitation of the n^{th} slot be a_n (not including the possible 180° phase change due to off-set on the opposite side of the centre). Then, the array factor (F) is given by,

$$F = \sum_{n=1}^{N} a_n \exp\left(jk_o\mathbf{r}\cdot\mathbf{r}_n - j\beta nd\right) \tag{5.88}$$

for the case of the slots all off-set on one side of the centre line (case1); and,

$$F = \sum_{n=1}^{N} a_n \exp\left(jk_o\mathbf{r}.\mathbf{r}_n + jn\pi - j\beta nd\right) \tag{5.89}$$

for the case where every other slot is displaced on adjacent sides of the centre line (case 2). In equations (5.88) and (5.89) $\beta = 2\pi/\lambda_g$ and $\mathbf{r}$ is given by:

$$\mathbf{r} = \mathbf{u}_x\sin(\theta)\cos(\phi) + \mathbf{u}_y\sin(\theta)\sin(\phi) + \mathbf{u}_z\cos(\theta) \tag{5.90}$$

where θ is the polar angle and ϕ is the azimuthal angle in spherical polar co-ordinates. To have maximum radiation in the direction $\phi = \pi/2$, $\theta = (\pi/2 - \psi)$, the relevant condition is as follows:

$$[k_0 d\sin(\psi) - \beta d] = 2m\pi \qquad \text{(for case 1)} \tag{5.91}$$

and

$$[k_0 d\sin(\psi) - \beta d] + \pi = 2m\pi \qquad \text{(for case 2)} \tag{5.92}$$

where m is an integer.

From equations (5.91) and (5.92), it follows that,

$$d = \frac{m\lambda_o\lambda_g}{\lambda_g\sin(\psi)\ \check{S}\ \lambda_o} \qquad \text{(for case 1)} \tag{5.93}$$

and

$$d = \frac{(2m-1)\lambda_o\lambda_g}{2(\lambda_g\sin(\psi)\ \check{S}\ \lambda_o)} \qquad \text{(for case 2)} \tag{5.94}$$

The above relations should be satisfied such that only one main lobe occurs in the visible space (or sector of interest), that is, no second-order beams should occur in that space.

Many solutions for d are possible depending on the relative value of λ_g/λ_0; But $\lambda_g/2\lambda_0$ is preferred so that the guide will not be too dispersive, because high dispersion will make the radiation very frequency sensitive. For $\lambda_g < 2\lambda_0$ the value of d must be chosen so that

$$\frac{\lambda_o\lambda_g}{(\lambda_g + \lambda_o)} < d < \frac{2\lambda_o\lambda_g}{(\lambda_g + \lambda_o)} \tag{5.95}$$

and

$$\sin(\psi) = \frac{\lambda_o}{\lambda_g} - \frac{\lambda_o}{d} \tag{5.96}$$

which will give a single main lobe in any direction between $-\pi/2$ and $\sin^{-1}[(2\lambda_0 - \lambda_g)/\lambda_g]$ with all of the slots off-set on one side of the centre line.

In the case of alternate slots being off-set on opposite sides of the centre line, if λ_g is chosen to be less than $2\lambda_0$, then,

$$\frac{\lambda_o\lambda_g}{(\lambda_g + \lambda_o)} < 2d < \frac{3\lambda_o\lambda_g}{(\lambda_g + \lambda_o)} \tag{5.97}$$

and

$$\sin(\psi) = \frac{\lambda_o}{\lambda_g} - \frac{\lambda_o}{2d} \tag{5.98}$$

Here the angle of radiation is limited to the range between $(-\pi/2)$ and $\sin^{-1}[(2\lambda_0 - \lambda_g)/3\lambda_g]$.

For angles greater than the angle ψ given by equation (5.98), it is not possible to get a single lobe in the forward direction without having another main lobe appearing at some other angle.

The fact that the main lobe depends on the relative value of λ_g/λ_0 for a given value of d can be made use of in scanning the lobe by changing the frequency. A highly dispersive waveguide is useful for frequency scanning, and this can be done by the use of waveguides bent into a sinusoidal shape so that the electrical length between the slots is increased.

Slot/notch arrays

To combat against serious fluctuations and fading of received signals in mobile telephones used in urban areas, apart from the spatial diversity technique, another proposal indicated in [5.14] refers to what is known as *the (flat) energy density antenna* system. This is made of two antennas mutually perpendicular to each other. Typical versions refer to:

- Monopole and a crossed-slot antenna combination
- Ring patch antenna comprised of a center probe terminated into a circular disk with a few matching posts that radiate the same pattern of a monopole antenna over a wide frequency range.

For mobile telephones, however, flat (conformal) designs are preferred. This can be achieved by assembly notches made on to a circular disk constituting an array.

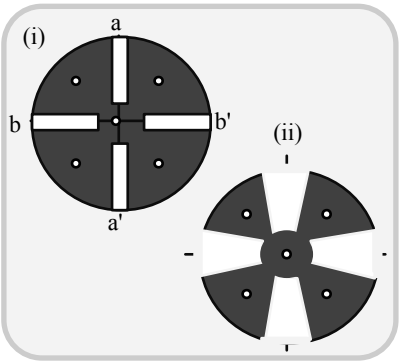


Fig. 5.35 Notch antenna on a circular disk. (i) Slit type and (ii) sectoral type

Described in [5.14] are two versions, with slit-type notches cut in a circular disk as illustrated in Figure 5.35. The four slots or notches cut in the disk such that they do not disturb the dominant currents from the centre feed point.

5.7 MICROSTRIP PATCH ANTENNA ARRAYS

These antenna arrays have been developed to increase the directivity with the patch elements fed by a single line or multiple lines as illustrated in Figure 5.36. Typically, in wireless applications a microstrip antenna array could be designed to yield downward-tilted beams for 840-940 MHz narrow-band applications. As depicted in Figure 5.37, the tilted beams are required for a confined illumination of the region of interest.

Microstrip patch antenna configurations (like patch elements) follow multiple variations in patch shape, array geometries, feeding techniques and substrate characteristics. This flexibility in form and other features has led to arrays that are generally classified as: series-fed, corporate-fed, scanning and polarisation-

specific arrays. The series-fed arrays are simple to fabricate and no soldering to elements is needed. It is used mainly for achieving fixed beams. Corporate-fed arrays are used where more control is needed for the excitation or when shaped beams are required. Details on scanning and polarisation-specific antennas are presented in the next chapter.

Microstrip patch antenna array designs can be envisaged *via* CAD techniques.

Choice of proper substrate material (dielectric) enables designs that lead to compact structures as well as they lend themselves for easy fabrication. For example use of polymethacrylate foam offers low-loss performance and it allows direct etching on the foam.

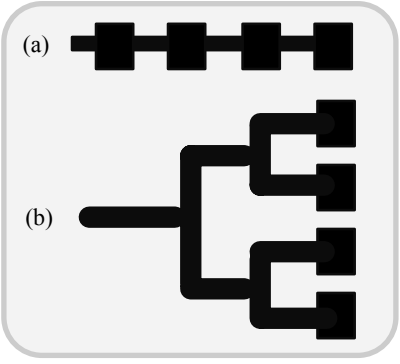


Fig. 5.36 Feed arrangements for microstrip patch antennas: (a) Single-line feed and (b) multiple-line feed. (The multiple-line feed facilitates electronic scanning)

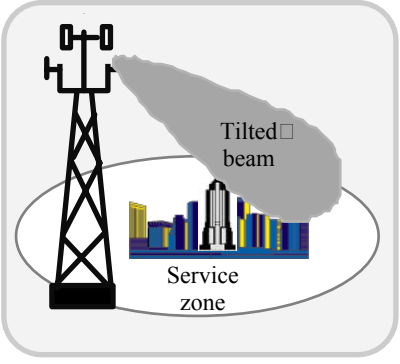


Fig. 5.37 Illumination of a service area with a tilted beam from a tower-mounted array

The major advantage of a microstrip array design is that it can be made compact, and therefore allows easy mounting. Also, compact arrays occupy less space and may not present a visibly protruding (objectionable) structure. Further, light-weight structures are easy for rigid fastening on the towers against wind forces. The small-size arrays can be conveniently covered by radome units for protection against rain and snow.

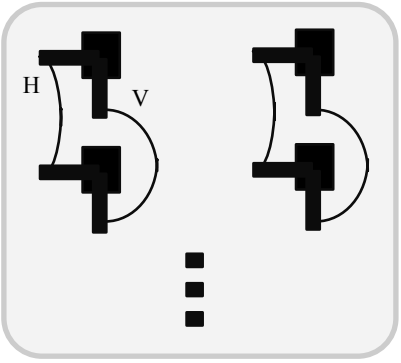


Fig. 5.38 Polarisation diversity patch antenna concept. (V and H are feeding ports)

Polarisation diversity radiation can also be accomplished *via* patch array antennas. As illustrated in Figure 5.38, such an array could be made with circular patches.
In the structure illustrated in Figure 5.38, each patch has two feeding ports, one for H-polarisation and the other for V-polarisation kept orthogonal to each other. These two receiving polarisations facilitate polarisation diversity reception.

5.8 PHASED ARRAYS

A *phased array* refers to an array of multiple elements with each element being capable of being excited independently with variable amplitudes and phases, so as to control or steer the beam direction and the beam shape (including the side lobes).

The phased array is categorized as follows:

- Frequency scanning array
- Retro-array
- Adaptive array

In the *frequency scanning array*, the required phase change is achieved by varying the frequency. It is a simple strategy, since no phase control at the elements is warranted. A *retro-array* is one in which the incoming signal is reflected back to the source end automatically. Also, known as *retro-reflector*, each element reradiates a signal, which is the conjugate of the received signal. (It emulates the role of a passive square-corner reflector antenna.)

A retrodirective array represents as a set of antennas, which reflect any incident signal back towards the source without *a priori* knowledge of the source location. From the antenna point of view, this array has an omnidirectional coverage while maintaining a high level of antenna gain simultaneously. Such characteristics are desirable in self-steering, search and rescue, noninvasive identification systems and interrogator-transponder systems.

The retrodirective radiation is realized as follows: Each element of the array is set to radiate an outgoing wave whose phase is conjugate to that of the incoming signal (relative to a common reference). A popular design refers to *Van Atta array* where transmission lines of equal lengths connect the conjugated elements (of a symmetric array).

The *adaptive arrays* are designed to be aware of their environment so as to adjust to it in a desirable fashion. That is, an adaptive array can automatically steer its beam towards a desired signal while steering a null towards an undesired or interfering signal. With the advent of modern digital technology, the output of each element in an adaptive array can be sampled, digitised and processed (by a DSP architecture) to accomplish several sophisticated tasks associated with the basic functions of a simple adaptive array. Such arrays are known as *smart antennas* widely being adopted in mobile communication systems. A comprehensive discussion on smart antennas is presented in Chapter 6.

5.8.1 Practical considerations in designing microstrip antenna arrays

As indicated earlier, a microstrip antenna consists of a patch on a PCB or substrate backed by a conducting ground-plane with an appropriate RF connector. Such antennas can be manufactured quite inexpensively and reasonably priced with the closely controlled substrate materials widely available in the market [5.15].

The beam patterns conceivable with microstrip antennas are: Pencil beams, fan beams and omnidirectional coverage; the polarisations obtained are linear (horizontal/vertical) and circular (RHCP/LHCP) polarisations. Further the patches can be made conformal to a flat or curved surface.

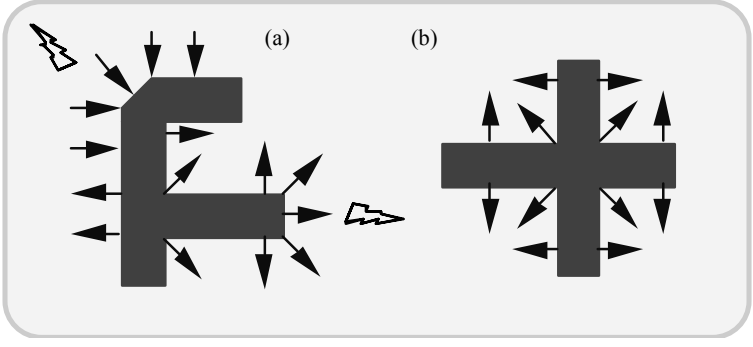


Fig. 5.39 Discontinuities in microstrip traces (a) Asymmetric corner T-section with abrupt-end discontinuities leading to radiation conditions; and (b) symmetric cross/junction discontinuity with no radiations due to symmetry

A straight microstrip trace is a "good" transmission line in supporting non-radiating guided-wave propagation. Should there be some sort of discontinuity in the transmission path, this structure will radiate. That is, at the microstrip discontinuities, the field fringes to an unbalanced condition leading to radiation. Typical asymmetrical fringe fields induced at some microstrip discontinuities (enabling radiation) are illustrated in Figure. 5.39. Such radiation will be enhanced if the strip-width is increased and/or thick substrates of low permittivity are adopted.

Any discontinuity in the transmission-line system represents admittance and a microstrip radiator array is therefore a system of lined up admittances with mutual coupling. The extent of mutual coupling controls the radiation resistances of the coupled elements; and it has implications on the matching, aperture illumination and side-lobe performance of the array.

5.8.2 Linear microstrip arrays

These are used for fan-shaped coverage and constitute the basic blocks for planar arrays. There are three major feed methods for linear arrays, as shown in Figure 5.40. They are: *Corporate feed*, *traveling-wave feed* and *resonant-feed* techniques. The corporate feed offers broadband performance and the overall bandwidth is limited by the bandwidth of individual radiators. The feed is large and expensive. It uses a power divider to energize the array elements. Corporate feeds are used in phased arrays where phase shifters incorporated in the transmission lines steer the beam direction electrically.

The traveling-wave feed has a single transmission line feeding all the radiators and the line is terminated with an absorptive load. Along the length of the line the energy decays. Therefore, in order to get a symmetrical power distribution along the radiating elements, necessary weighting/radiator-coupling is done asymmetrically.

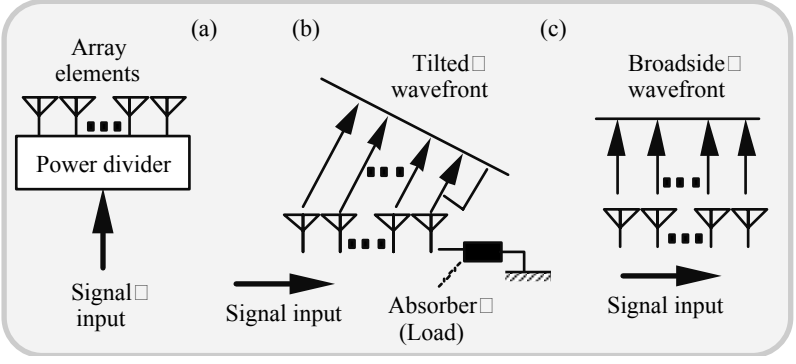


Fig. 5.40 Linear array feeds: (a) Corporate, (b) traveling wave and (c) resonant feeds

By keeping the distance between the radiators not an exact multiple of $\lambda_g/2$, the reflections from the radiators will be set to cancel each other. Good matching is dictated by tight radiator coupling factors and bandwidth of operation is limited by permitted beam-pointing variation.

In the resonant arrays, radiators are spaced deliberately a multiple of $\lambda_g/2$ apart so as to produce a broadside beam. These coupling factors are symmetrical. The bandwidth is limited only by matching requirements.

Example 5.2
Prepare a design outline on a linear traveling-wave array to yield the following performance characteristics:

- H-plane beamwidth ≈ 10°
- H-plane sidelobe level ≈ (less than –22 dB)
- H-plane beam boresight: 30° towards feed

Solution
The linear array under discussion can be conceived by a configuration due to James and Wilson [5.16].

Description of the array
The array configuration under consideration can be made of a central feed with stubs attached. The stubs are made $\lambda/2$ long so that the radiation fields seen at either end are in phase. The stub spacing can be set at $\lambda/2$, but should be trimmed for exact beam-pointing (boresight) direction. The design of this antenna warrants data on stub radiation conductance and corresponding mutual coupling conditions.

5.9 ARRAY TECHNIQUES FOR BEAMFORMING/SCANNING

In recent years beamforming/scanning technology has become useful in space-based communications antennas and in radar systems. There are two ways of beamforming in practice, *via* lenses and by using circuits.

The lens-based beamforming/scanning technique uses dielectric and waveguide lenses; the *Ruze lens*, the *bootlace concept* and *the Rotman, R-2R, R-KR* and *objective lenses*.

The circuit type beamformer/scanner are based on the so-called *Blass* and *Butler matrices*. A brief summary of these beamforming/scanning techniques is presented below.

5.9.1 Lens-based beamformers/scanners

Dielectric lenses

Lens antennas are equivalent to parabolic reflectors in the sense that, they (like parabolic reflectors) convert a spherical phase-front from the feed (assumed as a point source) to a plane phase-front at an aperture. Since the lens geometry directs the EM wave through itself, any feed-specified aperture-blockage (normally encountered in parabolic reflector system) will be absent in lens-based beamforming.

The dielectric lens design is decided by the geometry of the lens (the contours of front and back lens surfaces) and by the dielectric permittivity. The dielectric constant (or refractive index) of the lens can either be chosen to be uniform or nonuniform. In the case of nonuniform design, the dielectric constant varies from the centre to the edge of the lens following a certain law. Typical example is the *Luneberg lens*, which has a refractive index n that varies as a function of the radius r of a dielectric sphere as per the law $n(r) = (2 - r^2)^{1/2}$. There are other types of dielectric lenses with different refractive index profiles chosen on *ad hoc* basis. Dielectric lens antennas can also be constructed with artificial dielectrics.

Waveguide lenses

This type of lens is made with a set of waveguide elements as illustrated in Figure 5.41. The physical lengths of the waveguides are determined by the condition that the electrical path-length of an off beam-axis ray from a source (at the focus) to a plane perpendicular to the axis of the lens is equal to the path-length of the central (beam-axial) ray. Use of titanium-based waveguide structures has enabled light-weight versions of waveguide lenses compatible for multibeam communication antennas in the Ku-band or higher.

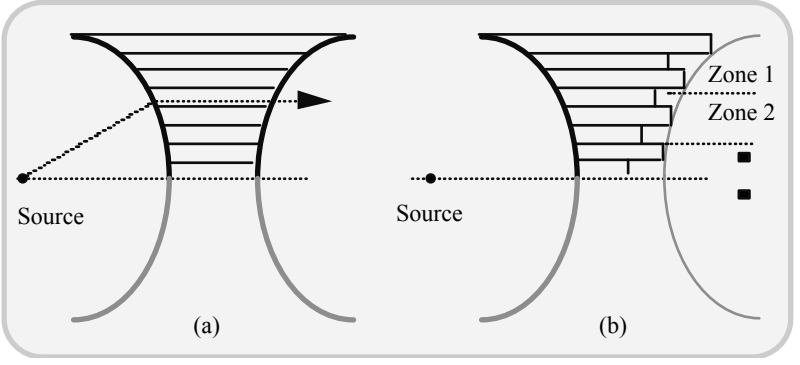


Fig. 5.41 Waveguide lenses: (a) Conventional and (b) zoned type

The arrangement of waveguides that enables lens-like beamforming is illustrated in Figure 5.41(a). In order to avoid bulky and thicker edges of the lens (so as to reduce the size and weight), waveguide-lens can be "zoned" of its contour. (*Zoning* is a method of introducing discontinuities on the lens contour.) Ray path-lengths through any two adjacent zones are designed to differ from each other by exactly 360° (electrical) degrees (or a whole multiple thereof), so that a plane of constant-phase front results on the collimated side of the lens.

In reference to lens antennas, the so-called *lens formulas* that have been developed are intended to specify the profiles of the lens surfaces (non-zoned or zoned) for a given index of refraction, n. Such formulas apply for both for $n > 1$ or $n < 1$.

5.9.2 Bootlace lens concept and Rotman lens

This type of lens consists of a set of transmission lines (supporting TEM mode), which interconnect a pick-up array of small radiators on the lens primary (inner) surface with a similar array of radiators on the secondary (or outer) contour of the lens.

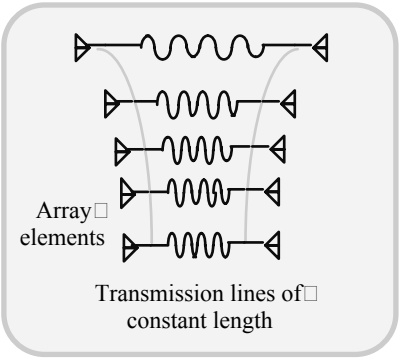


Fig. 5.42 Bootlace lens formed by a set of constant-length transmission lines supporting TEM mode

The bootlace lens configuration is practically nondispersive. That is, no cutoff wavelength of the waveguide elements comes into play. There are limitations, however, posed by impedance mismatch between elements and by restricted bandwidths of radiation elements. The TEM mode lines commonly used are coaxial lines; and, the radiating elements adopted are small dipoles or horn radiators situated on inner and outer surfaces as illustrated in Figure. 5.42.

In the Rotman lens, the bootlace design is adopted with the electrical path-length across the lens surfaces being not necessarily parallel to the lens axis. (On the other hand, in the so-called *Ruze lens*, the electrical path-length across the lens surface is parallel to the lens axis.)

Rotman lens is used to form multiple beams for antenna arrays. A typical arrangement is shown in Figure 5.43. The path-length difference of beam deflection is set such that the beam direction remains invariant with frequency. In Figure 5.43, the Rotman lens feeds a linear array that uses TEM propagation between flat parallel plates.

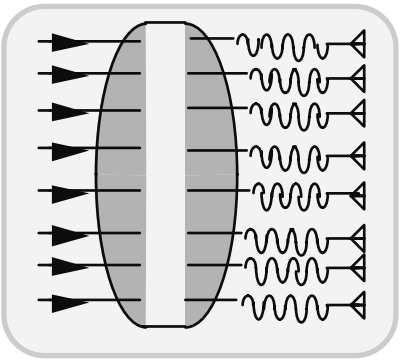


Fig. 5.43 Rotman lens-fed linear array

5.9.3 Circuit-specified beamformers

Beam scanning of the antenna array requires circuit-specified beamformers, which use couplers, power-splitter and transmission lines. Here the phase-shift produced is decided by the length of the transmission lines and by the aperture amplitude distributions (controlled by the power-splitter ratios). The two commonly adopted beamformers are known as the *Blass matrix* and the *Butler matrix*, which are briefly described below.

The Blass matrix

This architecture consists of a number of travelling-wave feed-lines connected to a linear array through another set of lines. The required number of amplifiers need not be binary. It can be reduced to equal number of active ports. This arrangement requires a coupler at every cross-point as shown in Figure 5.44.

In the matrix shown in Figure 5.44, the vertical lines are fed at the bottom and RF energy is partially coupled to every horizontal line as per the coupling coefficient of a directional coupler placed at the node of intersection. The EM energy admitted to the coupler from the vertical lines into the horizontal lines is such that the flow is upward from left to right. The input and output port numbers correspond to the beam and antenna element number respectively.

The Blass matrix can be implemented with waveguides, strip-lines or by a compatible TEM-line technology. The underlying principle is as follows: A signal applied at a beam port, travels along a feed-line to the end-termination. At each cross-over point a small signal is coupled into each element line that excites the corresponding radiating element. The path difference between the input and each

element controls the radiated beam direction. This travelling wave-based operation produces the input under matched conditions, and the set of beams formed will scan with the frequency. The major design factor of the associated network refers to selecting appropriate phase and coupling values so as to achieve a desired set of beams. The Blass matrix can be conceived to form a planar (two-dimensional) multiple beam microstrip patch array with a typical scan angle of ± 60° for an aperture size of about 15λ.

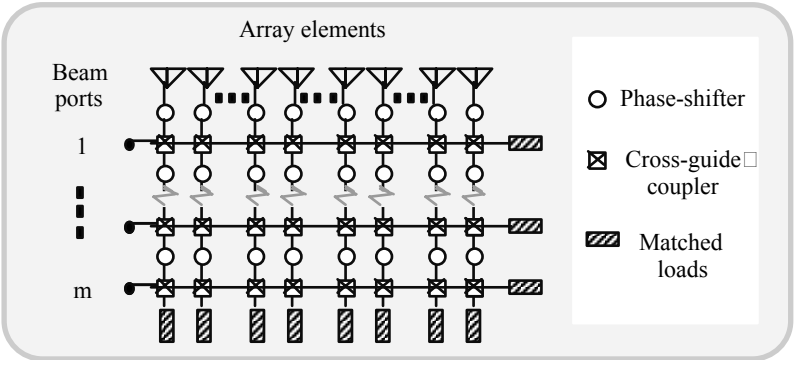


Fig. 5.44 The Blass matrix

The Butler matrix

The underlying concept of Butler matrix for beamforming is as follows: The design enables a signal input at any one of the transmit (TX) or receive (RX) ports to produce equal amplitudes at all the antenna ports and a linear phase progression from port to port. The phase slope depends on which TX/RX port is used. When the antenna ports are connected in sequence to an equally-spaced linear array, N beams are formed, one for each TX/RX port.

By externally combining two or more of the TX/RX ports of the Butler matrix, the antenna pattern can be moved, broadened, or the sidelobe level can be improved in certain directions.

The above linear array concept can be extended to a two dimensional array by having a Butler matrix behind each vertical array and then connecting each horizontal row of ports to a Butler matrix. In this fashion, an 8×8 array would have 64 beams. By combining groups of TX/RX ports from the 64 Butler matrix ports, it is possible to get beams with the desired shape and sidelobe levels.

Thus, the Butler matrix is a circuit-specified beamformer made of interconnected, fixed phase-shift sections and 3-dB couplers as shown in Figure 5.46. In this architecture, the output of each element is sampled, digitised and processed (*via* digital signal processing) to accomplish several functions of an adaptive array. Specifically, in the so-called smart antennas, application of this version of beamforming is becoming more popular.

The arrows in Figure 5.45 show how a signal fed into a port divides to excite the radiating elements with equal and linearly varying phase. This matrix allows a good input VSWR and beam isolation. More details on this architecture are furnished in Chapter 6 dealing with smart antennas.

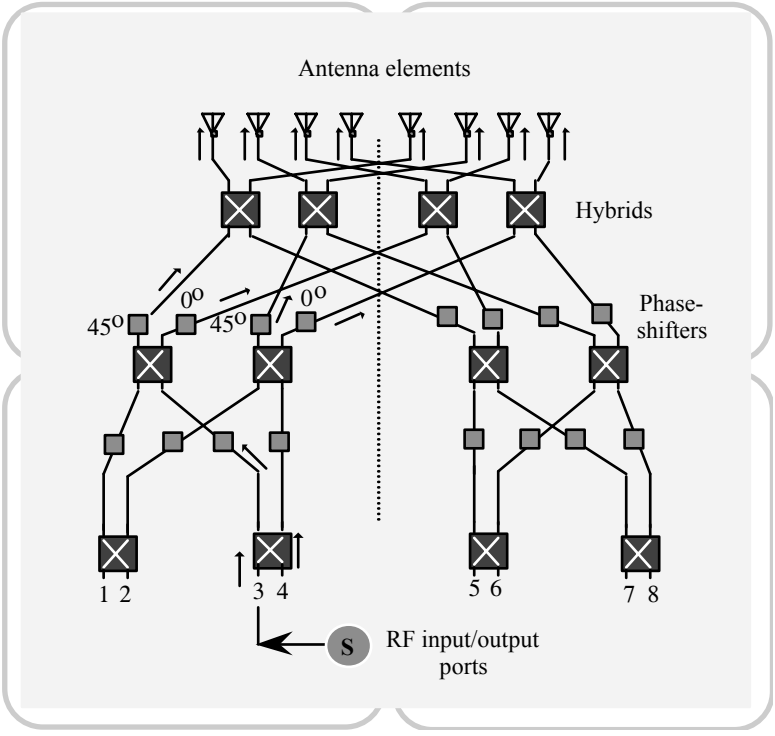


Fig. 5.45 A Butler matrix configured to excite eight elements

5.10 ARRAY ANTENNAS IN WIRELESS COMMUNICATIONS

Array antennas have been judiciously adopted for different applications in wireless communication systems on an *ad hoc* basis. The following is a compendium of typical applications.

5.10.1 Base-station applications

Triangular array

This is a typical base-station radiator formed by three dipole elements mounted on the poles as a triangular array. This configuration is designed to provide a desired pattern-shaping at UHF bands used in mobile telephone systems. The arrangement of array elements is illustrated in Figure 5. 46.

The discrete dipoles can be substituted with patch or slot antenna systems. Functionally, the triangular disposition of such radiating elements of the array is of main consideration in facilitating beam patterns that provide the required sectoral coverage.

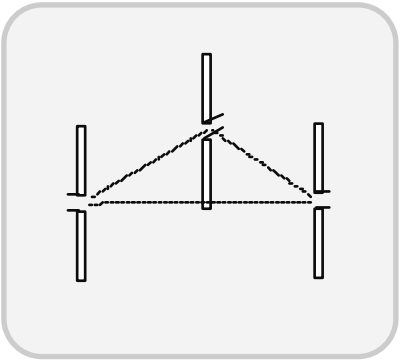


Fig. 5.46 Triangular array of dipoles for base station applications

Corner reflector array

The corner reflector-based array (Figure 5.47) is prescribed in base-station applications in order to achieve a desired sector beamwidth as well as enable beam-tilting (under frequency reuse situations). Further, relevant designs allow low sidelobe patterns. For example, a set of printed dipoles backed by a corner reflector in each 120°-sector can provide almost an omnidirectional horizontal beam and its vertical pattern would exhibit a desired tilt.

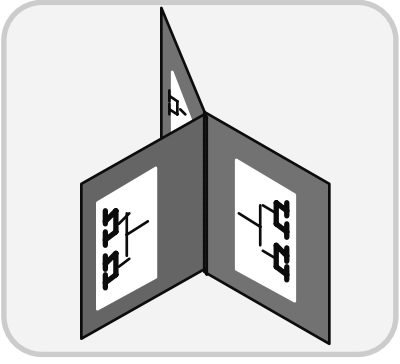


Fig. 5.47 Array with corner-reflector and sector-wise placed on printed dipoles

Typically corner reflector-based arrays are formed with 16 printed dipole elements and the beamwidth is controlled by the choice of the corner angle. Such arrays are used in the frequency band of 870-940 MHz. Further, in realising sector beams, corner reflector-based Yagi-Uda structures that are compatible for dual-frequency designs (at 900MHz and 1.5GHz) have also been developed as described in [5.17].

Microstrip patch array antenna

A set of microstrip patches constituting an array of radiating elements has been conceived to yield beams of narrow beamwidth required in base-station applications. Such an array provides a downward tilt beam in the 810-940 MHz band compatible for frequency reuse situations.

Pattern diversity antennas

Typically, a set of four dipole elements configured as an array can perform *pattern diversity* with each element. This type of array produces an omnidirectional pattern. Such arrays are made as base-station compatible structures. They are compact with element space being small. A typical arrangement is shown in Figure 5.48.

In the pattern diversity array, the omnidirectional pattern is synthesised with the 180°-sector coverage of each antenna of the array. In reference to Figure 5.48, the spacing (S-to-N or E-to-W) between the antennas is typically kept as 6λ. This yields a correlation coefficient less than 0.2 in urban environments. The (N-S) and (N-E) pairs are fed each *via* hybrid circuits (H) as illustrated.

Being compact, this version of array can be easily mounted on a tower or can be located on the open terrace of a tall building without being seen conspicuously.

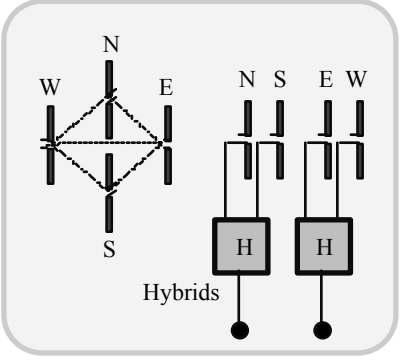


Fig. 5.48 Pattern diversity array providing omnidirectional pattern. (N: North; S: South; E: East and, W: West)

Polarisation diversity array

These compact configurations, using microstrip patch antenna elements, have been designed to offer a pattern of controllable beamwidth (obtained by adjusting the element spacing) and polarisation diversity. Typically, each element is a circular disk microstrip antenna with two feeding ports for diversity reception, which are orthogonal to each other. The correlation coefficient realisable is less than 0.2 in urban ambients. In this structure, the average received power in each port could differ significantly, since the reception is largely vertically polarised. As such, only when a compact or small volume of installation space is warranted, this array is advocated. The beam coverage is specified for 60°-sectors.

5.10.2 Array antennas in mobile units

The mobile units in the modern context of wireless telecommunications refer to portable, automobile-mounted, maritime-specific, locomotive-based and airborne vehicle supported versions.

In the portable devices (such as cell-phone, handheld computer etc.), the compactness is of primary concern and therefore, use of multiple antenna elements (constituting an array) is rather limited. Mostly, single-element antennas are widely used in portable units. Plain monopoles (telescopic, ribbon-type, and inductance loaded), dipole derivatives (sleeved, and planar inverted-L or -F), normal-mode helical structure, slotted antennas, patches and loops in different forms constitute the plethora of antennas conceived for portable unit applications, as discussed in Chapter 4.

The use of array antennas in automobiles is also limited so as to avoid high wind resistance and drag. Simple array structures such as four dipoles each housed in a quadrant corner-reflector forming UHF-band radiating system with angular diversity, have been developed. Further, to achieve polarisation diversity, cross-dipoles arrayed perpendicularly are used. Alternatively, an array formed by a loop and a monopole has been devised to provide polarisation diversity.

An array made of a circular patch with a parasitic element is also used in mobile telephone (800 MHz band) applications offering low-profile geometry and vertical as well as horizontal polarisations. Such an array is self-resonant and broadband.

In trains, the antenna structure used is kept on the roof-top. It operates at a frequency of 150 MHz (and communicates with the nearby stations). To avoid wind resistance, this structure is usually compact and is in the form of a single-element inverted structure offering low drag. (Use of array is not warranted.)

Maritime systems use array structures on a larger scale. Here, compactness is not a constraint. End-fire patterns are often required with circular cross polarisation. Hence, the following types of array structures are widely deployed.

- Axial-mode helical antenna with rear-end reflector yielding a wideband circular polarisation beam with a gain of 8-15 dBi
- Yagi-Uda array of cross-dipoles forming a end-fire structure with circular polarisation pattern and a gain of 8-15 dBi

- Quad-helix array made of four-turn helix giving high aperture efficiency and a pattern of circular polarisation
- A two-dimensional array of 16 cross-dipoles to give a pattern of circular polarisation and a medium gain of about 15 dBi
- Array made of short backfire antennas to realise high-aperture efficiency, low sidelobe levels and wideband operations. Either orthogonal or circular polarisation can be facilitated.

For airborne vehicle applications, the compactness and streamlined geometry are crucial. Slant/bent elements conformal to the vehicle surface offer dragless structures. Mostly, such single-element structures are used for various aeronautical communication applications. However, microstrip antennas in two planes mounted on the top of fuselage have been conceived to form a phased array for airplane telephones offering a gain of about 12 dBi.

Arrays also find applications in the satellite systems developed for wireless communication. Typically, a spiral element set constituting a thin array (facilitated with mechanical tracking) yields a medium gain of 15 dBi over a directional pattern. This structure is adopted in INMARSAT-M, Mobilesat, ETS-V and MSAT-X systems.

Likewise, a phased array made of microstrip antennas has been developed for electronic tracking in ETS-V and MSAT-X systems. The associated gain is about 15 dBi.

A patch array made of four-elements has been conceived as transportable antenna for INMARSAT-C systems.

Other arrays used in the satellite wireless systems refer to the following: A dual patch array (of microstrip antenna elements) mounted on the lid of a briefcase terminal and used for low-speed data transmission in ETS-V systems. Another microstrip patch array (with a high gain of 40 dBi) has been designed for Ku bands (12 GHz) for TV and base-station (BS) reception on a vehicle. This array is facilitated with mechanical tracking.

5.11 CONCLUDING REMARKS

The existence of several types of discrete radiators has permitted to conceive and use different versions of antenna arrays. From the gamut of such structures, however, only some have been widely chosen and used in wireless communication systems. The reasons are as follows: The compactness required in portable units, specific radiation characteristics and specified coverage warranted and restricted frequency bands of the applications. However, array antennas in wireless communication systems have been rendered “intelligent” so as to form beams that search and direct at blind sources. Such optimised structures are known as smart antennas, which are exclusively addressed in the next chapter.

This page intentionally left blank

REFERENCES

[5.1] Chatterjee, R., *Antenna Theory and Practice*, New Delhi, India: New Age International (P) Ltd. Publishers, 1996

[5.2] Collin, R. E., and F. J. Zucker, *Antenna Theory,* New York, NY: McGraw-Hill Book Co., 1969

[5.3] Balanis, C. A., *Antenna Theory Analysis and Design,* New York, NY: Harper & Row Publishers, 1982

[5.4] Kraus, J. D., *Antennas*, New York, NY: McGraw-Hill Book Co., 1988

[5.5] Jordan, E. C., and K. G. Balamain, *Electromagnetic Waves and Radiating Systems,* Englewood Cliff, NJ: Prentice-Hall, Inc., 1968

[5.6] Jasik, W. l., *Antenna Engineering Handbook* New York, NY: McGraw-Hill Book Co., 1961

[5.7] Schelkunoff, S. A., and H. T. Friss, *Antenna Theory and Practice,* New York, NY: John Wiley & Sons, 1952

[5.8] Hansen, W.W., and J. R. Woodyard, A new principle in directional antenna design, *Proc. IRE*, March 1938, vol. 26, 333-345

[5.9] Abromowitz, M. and I. Stegun (Eds.), *Handbook of Mathematical Functions with Functions, Graphs, and Mathematical Tables,* New York, NY: Dover Publications, Inc., 1970

[5.10] Brown, G. H., Directional antennas, *Proc. IRE*, January 1937, vol.25, 78-145

[5.11] Yagi, H., Beam transmission of ultra short waves, *Proc IRE*, June 1928, vol.16, 715-740

[5.12] Uda, S., On the wireless beam of short electric waves, *JIEE (Japan*), No.452, 273-282 (March 1926), and No. 472, 1209-1219 (November, 1927)

[5.13] Murata, M., K. Li, and T. Matsui, Planar active Yagi-like antenna, *Electron. Letts*., November 2000, vol. 36(23), 1912-1913

[5.14] Arai, H., H. Iwatshita, N. Toki and N.Goto, A flat-energy density antenna system for mobile telephone, *IEEE Trans. Veh. Technol*., May 1991, vol 40(2), 483-486

[5.15] Hill, R., A practical guide to the design of microstrip antenna arrays, *Microwave J*., February 2001, 166-180

[5.16] James, J. R., and G. J. Wilson, New design techniques for microstrip antenna arrays, *Proc. 5th Euro. Microwave Conf*., September 1975, 102-106

[5.17] Suzuki, T. and K. Kogoshima, Corner reflector antenna with same beamwidth in two frequency bands, *IEICE Trans*., 1992, vol. J75-B-II, No.12, 950-956

CHAPTER 6

Smart Antennas for Wireless Networks

6.1 INTRODUCTION

In the realm of modern wireless networks, the term "smart antennas" defines an antenna array combined with an assortment of units used for adaptive signal-processing carried out in space and time. The considerations, which led to the development and evolution of such smart antennas, refer to certain major impediments faced in realising high performance characteristics in wireless communications. For example, the co-channel interference, intersymbol interference and signal-fading conditions encountered in wireless systems have placed constraints on the performance aspects of modern wireless communication strategies as indicated in earlier chapters. The counteracting measures conceived thereof form the basis for the proliferation of various smart antenna designs and their implementations [6.1-6.10].

The importance of going for smart antenna applications in wireless communication systems can be appreciated by explicitly identifying the possible performance impairments associated with such systems:

- *Co-channel interference*: This refers to the interference caused by other users operating in a designated EM spectrum and it constrains the number of users serviced by the system. (This is known as the *system capacity scalability*)
- *Fading effects*: These effects are due to multipaths experienced by reflected components of the electromagnetic wave supporting the radio signals. They lead to high bit error rates (BERs) and eventual limited data rates of transmissions.

Both undesirable influences indicated above can be mitigated to a large extent by resorting to the use of smart antenna principles. The underlying considerations are as follows: The desired signal from the target subscriber as well as the co-channel interference components (due to other users) may arrive simultaneously at the receiving end from different directions. Then the smart antennas exploit the inherently associated temporal differences among such components so as to reduce the interference involved and thereby augmenting the channel capacity. Likewise, the multipath components of the signal from a target subscriber may also arrive

from different directions. As such, appropriate spatial processing of the received signal at different elements of the array can be used to attenuate the multipath components (relative to the desired signal), leading to a reduced intersymbol interference and fading influences.

The spatial-processing mentioned above refers to the sector-signal received at the (spatially-disposed) elements of the array. Such signals are individually "processed", so as to reduce optimally the influence of path-difference effects in the combined signal extracted from the array. Spatial signal-processing offers a degree of freedom and flexibility in the system design. Further, as indicated above, array signal-processing would allow countermeasures on co-channel and multipath specified performance degradations. Relevant efforts lead to improved overall system performance identified *via* range extension, capacity enhancement, high data-rate feasibility, and better BER profiles.

There are two major perspectives that govern smart antenna design and implementation: The first one refers to assessing the complexity of spatial domain processing and the second one specifies the hardware features of the antenna elements, which constitute the array. Such features in effect designate the "real-estate" considerations of the array together with the size, weight and power-handling attributes of the constituent elements.

The design of smart antennas relies on what is known as the *channel models* adopted, which depict the spatiotemporal characteristics of the signals supported on the channels. In reference to macro- and/or microcells, the prevailing channel models refer to three categories, namely, *models based on statistical considerations*, *measurement-based models* and *site-specific models*, which are described in the following section.

The performance of smart antennas eventually depends on the type of spatial-processing techniques used. In reference to digital cellular systems (such as, IS-95, IS-136 and GSM), the underlying processing techniques are distinct with relative merits and demerits. However, the scope of spatial-processing in all such cases is directed at achieving the same goals as indicated above, namely, range extension, interference suppression and capacity increase. Generally, the temporal- and spatial-processing in such systems may also include spatial and temporal equalisations and joint optimisation of transreceive array units.

The adoption of smart antennas has been a major leap in the evolution of wireless communication systems. Relevant to the systems in vogue, smart antennas are integrated with conventional site equipment deployed *in situ*. These antennas function largely as an adaptive spatial filter that enables increased signal quality, capacity, and/or range over traditional antennas. The spatial-filter implementation calls for a mixed-signal handling pertinent to active RF and digital signals. This requires a cohesive use of RF and digital electronics as an integral part of the system.

Further, since the multi-user cellular communication systems are prone to interference and poor voice quality (resulting from omnidirectional transreceive profile of the RF signal), advanced techniques are imminent to improve the quality of service and boost up the traffic capacity. Any such attempt, however, should be at an affordable budget. The scope of smart antenna implementation, in essence,

covers these system requirements. In the real estate of wireless systems, smart antennas, as mentioned before, are realised by using a collection of multiple antennas supplanted by intelligent signal-processing techniques. Specifically, they are designed to perform the following: The antenna system smartly attempts to form and adjust the beam patterns so as to provide optimal gain consistent with the user's location; also, it simultaneously identifies and minimises the interference stemming from other users. This function, in essence, refers to resorting to spatial dimensioning in signal-processing. The ultimate goal of the associated endeavour aims at improving flexibility, cost-saving, quality and capacity of the system as a whole. The capacity improvement explicitly allows more people to share the same air-interface.

Specific to indoor environments, the propagation losses of RF channels could be even more severe and multipath effects intense. To combat against such situations, it may be required to increase the transmission power in order to cover the area specified by the cell zones. Smart antennas are considered as possible ways to obviate the associated problems (especially in millimetre wave transmissions). Such smart antenna deployments are expected to offer options such as range extension, improved immunity to flat-fading, interference suppression and the so-called *spatial division multiple access* (SDMA).

Use of antenna arrays in smart antenna designs of wireless communication systems improves reliability and capacity in three possible ways:

- *Diversity-combining techniques*, which aggregate the signals from multiple antennas in a way that mitigates multipath fading (as indicated earlier)
- *Adaptive beamforming method* (*via* antenna arrays) that provides capacity improvement through interference reduction. It also mitigates multipath fading. That is, adaptive arrays cancel or coherently combine multipath components of the desired signal and null the interfering signals that have different directions of arrival (with respect to the desired signal)
- Using *switched fixed beams*, which allow coarser pattern control (than the adaptive arrays), but can still provide capacity improvement. Relevant principle refers to using two or more fixed beams that permit diversity reception.

Adaptive and switched-beam antennas are referred to as smart antennas because of the dynamic system intelligence required for their operation [6.8]. There are also other smart antenna systems known as *polarisation sensitive adaptive arrays*. In their applications, the system matches the polarisation of a desired signal and it nulls an interferer that has the same direction of arrival as the desired signal but of different wave polarisation.

It is also possible to combine both space and polarisation reuse concepts. Such designs enable better link quality, augmented system capacity and improved reliability. Also, the power consumption at the handheld units can be reduced with an increase of battery life [6.7, 6.9].

6.1.1 System aspects of smart antenna technology

The smart antenna technology with its associated spatial-processing is considered as the last frontier in battling against the impairments perceived by wireless communication technology [6.10]. The advantages expected with smart antenna systems can be summarised as follows:

- Higher sensitivity in receptions
- Possibility of implementing SDMA
- Interference cancellation in uplink and downlink functions
- Mitigation against multiple fading.

State-of-the-art smart-antenna designs and implementation-schemes *vis-à-vis* the modern wireless technology, have certain specific considerations, which can be enumerated as follows:

- Smart antennas are systems of significant complexity
- They are, in general, expensive
- They provide additional degrees of freedom for the radio network control and planning
- Smart antenna configurations can be optimised in terms of channel characteristics and receiver structures
- The associated spatial-processing technology includes: network planning, radio network management, receiver circuit aspects and processing algorithms
- Radio network performance through smart antennas can be improved by combining different spatial-domain processing (such as beamforming, sectorisation and spatial diversity) and temporal-domain processing
- The network planning is often specific to the base station's real estate
- System integration efforts of smart antenna-based receivers are required to take into consideration of expected channel conditions, user mobility and offered traffic.

This chapter is written to provide an overview of spatial channel models and smart antenna techniques. Further, the concept of an antenna array that includes the principle of adaptive processing is described in the light of smart antenna considerations.

6.2 CHANNEL MODELS

Knowledge of spatial properties of propagation channels over which the wireless communication transmissions are supported is crucial towards the design, implementation and evaluation of antenna arrays adopted as smart radiating systems in modern wireless communications.

Such knowledge is relevant, in general, to any system of communication links established between a stationary base station and a set of mobile units. The

channel model in those circumstances essentially refers to assessing the profile of multipath propagations involved as illustrated in Figure 6.1.

The multipath propagation model of the channels, as depicted in Figure 6.1, largely depends on the statistics of the geographical dispositions of the mobile units (MU) relative to the base station (BS) and the scatterers (such as buildings, B) in the vicinity.

In reference to a receiver representing a single sensor, its response can be modelled in terms of the narrowband statistics of time-varying (fading) signal power perceived at the sensor. However, if a conglomerated set of multisensors is used, the resultant signal vector should be characterised by the various multipaths, the associated delays and the *angles of arrival* (AoA) of the signal in each path [6.11].

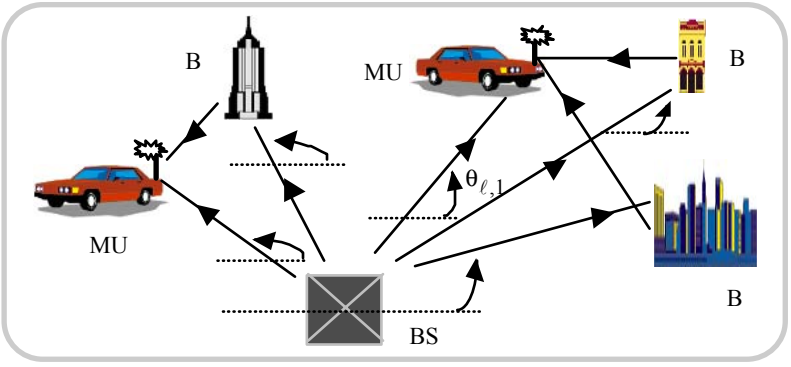


Fig. 6.1 Multipath propagation channel model. (BS: Base Station; MU: Mobile Unit, B: Buildings)

With reference to Figure 6.1, the following notations apply in modelling the multipath propagation channels:

Mobile unit index:	k
Signal-component index at this k^{th} mobile:	ℓ
Amplitude of the signal component:	$A_{\ell,k}$
Carrier-phase of the signal component:	$\phi_{\ell,k}$
Delay associated with the signal component:	$\tau_{\ell,k}$
Angle-of-arrival (AoA) of the signal component:	$\theta_{\ell,k}$
Doppler shift associated with the signal component:	f_d

Since the aforesaid components, in general, could be time-varying, an RF channel depicting the multipath propagations illustrated in Figure 6.1 is a time-variant channel; its baseband (complex) envelope can be represented, for example for MU1, by a *vector channel impulse response* given by [6.12]:

$$\mathbf{h}_1(t,\tau) = \sum_{t=0}^{L(t)-1} \{A_{\ell,1}(t) \times \exp[j\phi_\ell(t)] \times \mathbf{a}[\theta_\ell(t)] \times \delta[t - \tau_{\ell,1}(t)]\} \qquad (6.1)$$

where L(t) is number of multipath components, $\mathbf{a}[\bullet]$ is the *array response vector*, which is a function of AoA and the array geometry. The other parameters are defined above.

Explicitly, considering an arbitrary two-dimensional array geometry such as the one shown in Figure 6.2, the array response vector is given by,

$$\mathbf{a}[\theta_\ell(t)] = \begin{bmatrix} \exp(-j\psi_{\ell,1}) \\ \exp(-j\psi_{\ell,2}) \\ \bullet \\ \bullet \\ \bullet \\ \exp(-j\psi_{\ell,m}) \end{bmatrix} \qquad (6.2)$$

where $\Psi_{\ell,i}(t) = \beta\{\mathbf{x}_i \cos[\theta_\ell(t)] + \mathbf{y}_i \sin[\theta_\ell(t)]\}$, $\beta = 2\pi/\lambda$ and λ is the wavelength of the carrier.

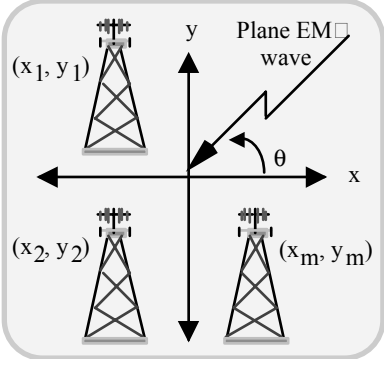


Fig 6.2 Two-dimensional antenna array with the elements positioned arbitrarily in the different quadrants of the Cartesian co-ordinate system

The spatial channel response depicts the vector sum of the multipath components, each of which is characterised by its own amplitude, phase and AoA. The amplitude $A_{\ell,k}$ is commonly modelled as a Rayleigh-distributed random variable. However, a generalised non-Rayleigh amplitude distribution (such as the Nakagami probability density function) could also be applicable for a specified environment. The phase shift $\phi_{\ell,k}$ is uniformly distributed. Further, the statistics of angle spread of the channel is decided by the antenna heights and the type of environment.

Typically, the wireless communication ambient is classified as *macrocell* and *microcell* environments. In the macrocell, it is assumed that the scatterers surrounding the mobile station are more or less of the some height or higher than the mobile unit. This assumption leads to the signal arriving at the MU (after bouncing from the surrounding scatterers) to be received as an omnidirectional reception from all directions. Hence, it is reasonable to presume that the distribution of AoA at the MU is uniform.

In contrast, considering the BS, where the antenna deployed is normally higher than the surrounding scatterers, the signal component scattered in the vicinity of the MU and reaching the BS are constrained within a narrow angular regime. As such, the distribution of the AoA at the BS is no longer uniform over 0, 2π.

The so-called microcell environment refers to a situation in which the BS antenna is mounted at the same height as the surrounding objects. As a result, the multipath AoA spread would increase.

Consistent with the general channel propagation scenario described above, there have been a number of other channel models evolved and described in the literature to represent the varying features of the environments in question. The following is a summary of relevant models [6.4].

6.2.1 Lee's model

This refers to evenly spaced "effective" scatterers on a circular ring about a mobile unit as shown in Figure 6.3. (Each of the "effective" scatterers in this model may cohesively depict the effect of many scatterers within the region.)

Lee's model predicts the correlation coefficient of signals between any two elements of the array separated by d metres. The extent of correlation determines the performance of spatial diversity methods pursued. Greater spread of beam coverage and larger element spacing would reduce the correlation and increase the diversity gains. Typically the radius of scatterers involved may range from 100 to 200 wavelengths.

Assuming discrete AoAs, the correlation coefficient ρ as a function of d, θ_o, R and D (Figure 6.3), is specified as follows:

$$\rho(d, \theta_o, R, D) = (1/N) \sum_{i=0}^{N-1} \exp[-j2\pi d \cos(\theta_o + \theta_i)] \qquad (6.3)$$

where N is the number of effective scatterers and θ_i is the angular separation between scatterers placed within the line-of-sight (LoS) as illustrated in Figure 6.3. This discrete angular spacing is given by,

$$\theta_i = (R/D)\sin(2\pi i/N), \quad i = 0, 1, 2, \ldots, (N-1) \tag{6.4}$$

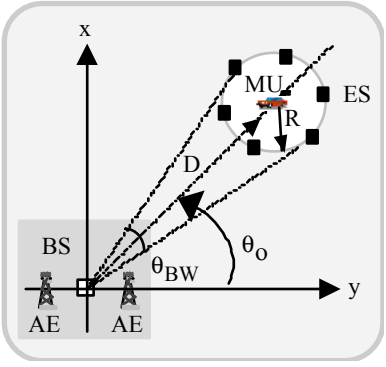


Fig. 6.3 Lee's model (ES: Effective scatterers; MU: Mobile unit, BS: Base station and AE: Array element)

In the above model, the scatterers are assumed to be stationary with respect to the MU. However, if an attribute of relative velocity to the ring scatterers is given, the corresponding Doppler shift can be included in terms of v/R, where v is the velocity of the vehicle, and it can be explicitly specified in the description of the channel model.

6.2.2 A model of discretely disposed, uniform set of evenly-spread scatterers

This model evaluates the correlation coefficient of N uniform scatterers evenly disposed over an AoA range within a narrow beamwidth, as illustrated in Figure 6.4.

The relevant correlation coefficient is given by:

$$\rho(d, \theta_o, \theta_{BW}) = (1/N) \sum_{i=n_1}^{n_2} \exp[-j2\pi d \cos(\theta_o + \theta_i)] \tag{6.5}$$

where $n_1 = -(N-1)/2$, $n_2 = +(N-1)/2$ and $\theta_i = i\theta_{BW}/(N-1)$ with $i = n_1, \ldots, n_2$. This model again represents a discrete AoA version and the correlation predicted falls off more quickly than that of Lee's model.

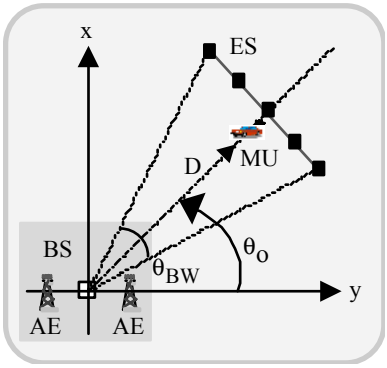


Fig. 6.4 A set of discretely dispersed and evenly spaced, uniform scatterers. (ES: Effective scatterers; BS: Base station)

6.2.3 Macrocell model

This is a geometry-specified, single-bounce circular model, which assumes that the scatterers lie within a circular ring around the mobile as shown in Figure 6.5.

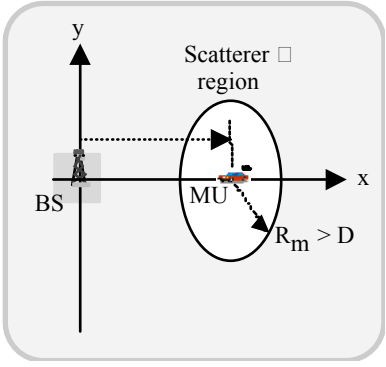


Fig. 6.5 Microcell model of a wireless channel. (BS: Base station; MU: Mobile unit)

This model is intended to describe macrocell environments where antenna heights are relatively large. It enables the elucidation of the joint angle-of-arrival and *time-of-arrival* (ToA) density function at both base station and mobile unit.

6.2.4 Microcell wideband model

This is again a geometry-based model with the scatterers dispersed in an elliptical region as shown in Figure 6.6. In this model, the scatterers are assumed to be

uniformly disposed within an ellipse, whose foci correspond to the BS and the MU. This model is intended for microcell environments where antenna heights are relatively low. It enables ascertaining information on AoA, ToA, joint AoA and ToA, Doppler-shift and signal-amplitude details.

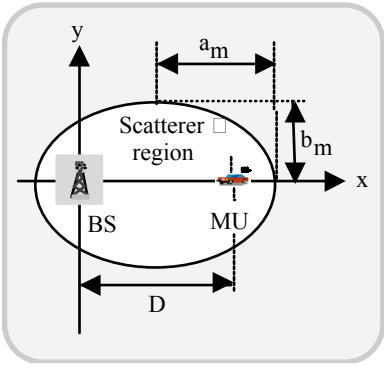


Fig. 6.6 Microcell wideband model of wireless channel: The BS and MU are the foci of the elliptical region

6.2.5 Gaussian, wide-sense stationary, uncorrelated scattering (GWSSUS) model

This model refers to a set of arbitrary scattering clusters in the wireless communication ambient as shown in Figure 6.7.

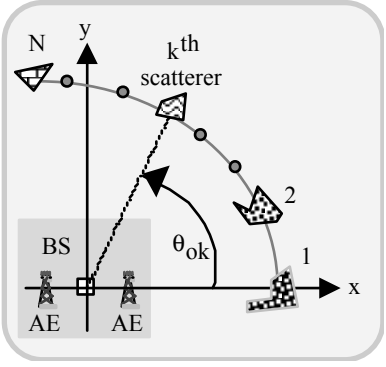


Fig. 6.7 Wireless channel model based on random cluster of scatterers: Each scatterer is a conglomeration (cluster) of smaller scatterers

In the ambient of scatterers illustrated in Figure 6.7, there are N scatterers, which are grouped into clusters in space, such that the delay differences within each cluster are not resolvable within the passband of transmission signal bandwidth. This model provides an explicit analytical description for the array covariance matrix.

6.2.6 Gaussian angle of arrival model

This is a special case of the GWSSUS model with a single cluster depicting a dominant scatterer (Figure 6.8).

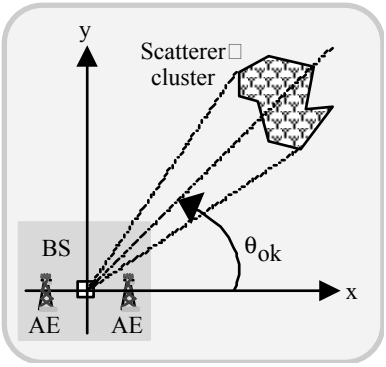


Fig. 6.8 Gaussian angle of arrival model specific to a single cluster of scatterers: The scatterer is a single dominant entity

The AoA statistics of this model is assumed as being Gaussian-distributed about some nominal (mean) angle. Further, this model corresponds to a narrowband channel and it is used in the analytical descriptions of the array covariance matrix.

6.2.7 Time-varying vector channel model (Rayleigh's model)

This model is illustrated in Figure 6.9. It assumes the signal energy leaving the region of a mobile as Rayleigh-faded and the prevailing angle spread is essentially attributed to the presence of dominant reflectors coexisting in an ambient of local scatterers indicated in Figure 6.9. Thus, this model explicitly relates to Rayleigh fading and offers theoretical characterisation of relevant spatial correlation properties.

Rayleigh fading takes place whenever there are multiple indirect paths between transmitter and receiver and no distinct dominant path, such as an LoS path. Normally, this model depicts a worst-case scenario.

Rayleigh fading can be dealt with analytically, providing insights into performance characteristics that can be used in difficult environments, such as downtown urban settings and crowded metropolitan bazaar streets. A wealth of knowledge is available on Rayleigh statistics from the radar cross-section studies

developed to address the so-called Swerling-Marcum models to describe the radar echo statistics from complex targets [6.13]

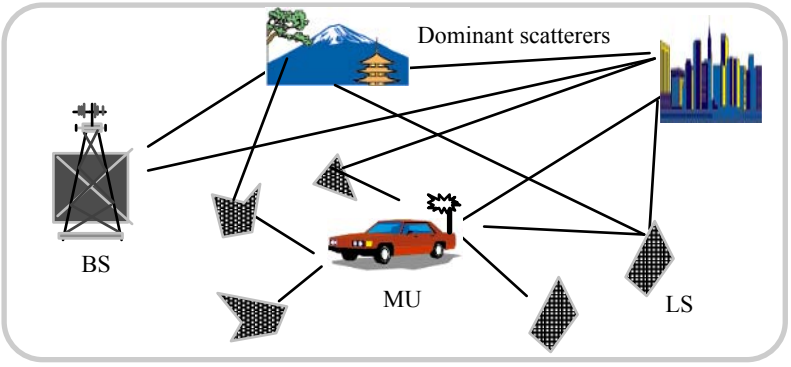


Fig. 6.9 Rayleigh model pertinent to the coexisting ambient of local scatterers (LS) and dominant scatterers

6.2.8 Typical urban (TU) model

This is used as a simulation model for GSM, DCS1800, and PCS1900 systems. Its time-domain properties are similar to the GSM-TU defined in GSM 05.05. Typically, relevant modelling includes a large number (about 120) of scatterers randomly placed within a practical radius (of 1 km) about the mobile unit. The received signal is determined by brute-force using the data on the location of each of the scatterers and the time-varying location of the mobile. In this model, the AoA statistics are approximately Gaussian.

6.2.9 Bad urban (BU) model

This model, as illustrated in Figure 6.10, is again adopted as a simulation model for GSM, DCS1800, and PCS1900 systems.

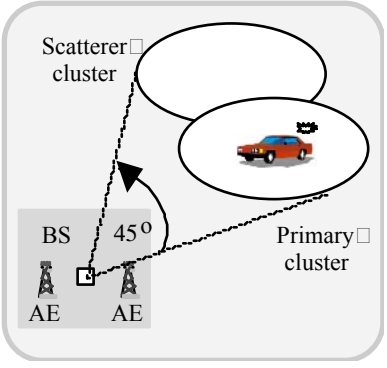


Fig. 6.10 Bad urban vector channel environment with a pair of primary and a secondary cluster

The BU model accounts for large reflectors that are not in the vicinity of the mobile. It is identical to the typical urban model with the addition of a second scatterer cluster offset 45° from the first.

The presence of the second cluster results in an increased angle spread, which in turn reduces the off-diagonal elements of the array covariance matrix. Further, the presence of the second cluster also causes an increase in the delay spread.

6.2.10 Uniform sectored distribution model

This model assumes that scatterers are dispersed over a sector formed by the angle θ_{BW} and a radial range of ΔR (centred about the mobile). The magnitude and phase associated with each scatterer are selected at random from a uniform distribution of [0, 1] and [0, 2π], respectively. This model is illustrated in Figure 6.11.

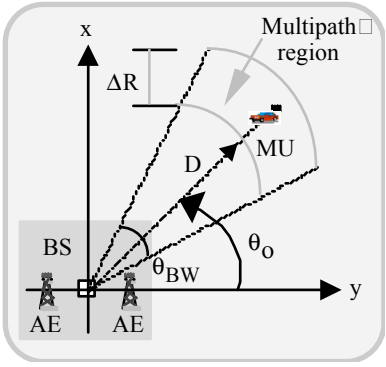


Fig. 6.11 Uniform sectored distribution geometry. (BS: Base station; MU: Mobile unit)

6.2.11 Modified Saleh-Valenzuela's model

This is an extension to the so-called *Saleh-Valenzuela model* [6.14]. Originally, in [6.14], Saleh and Valenzuela developed a multipath channel model for the indoor environment based on the clustering phenomenon observed in experimental data. In their model, the clustering is attributed as a result of multipath components arriving at the antenna in groups; it was observed that both the clusters and the rays within the cluster decayed in amplitude with time. The modified version of Saleh-Valenzuela model applied to outdoor cellular wireless communication channels assumes the inclusion of AoA information in the channel model. Further, the model assumes that the time and the angle are statistically independent.

6.2.12 Extended tap delay-time model

This is a wideband channel model. It is an extension of traditional statistical tap-delay-line model and includes AoA information. This model is composed of specifying the channel impulse function in terms of W taps, each associated with a

time-delay τ_w, a complex amplitude α_w and AoA, θ_w. Using measured data, the joint distribution of the model parameters are empirically specified.

6.2.13 Spatiotemporal model

This is also known as *Lu, Lo and Litva model*. It depicts a channel description of multipath propagation based on the distribution of the scatterers in a set of elliptical subregions as shown in Figure 6.12. Each subregion corresponds to a range of excess delay time involved.

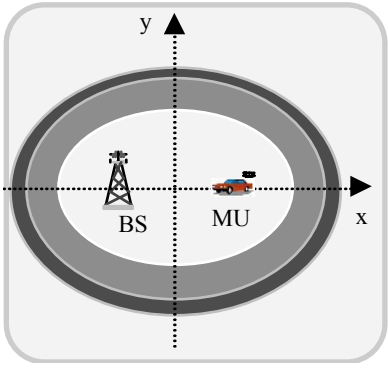


Fig 6.12 Elliptical subregions, each characterised by a unique value of excess delay

6.2.14 Measurement-based model

This model characterises the propagation environment in terms of a set of scattering points. Relevant parameters are decided from the data collected *via* measurements. Hence, a transreceive system function is used to describe the time-variant impulse response of the model.

6.2.15 Ray-tracing model

This is a deterministic model evolved *via* geometrical theory of ray-tracing attributed to reflection, diffraction and scattering phenomena associated with a wireless channel environment. This model (unlike the statistical models described above) is *site-specific* and uses deterministic data on building architectures, their dispositions, path-losses of the electromagnetic wave etc. This model, however, becomes highly cumbersome when the associated data-base is large.

The spatial channel models described above include the AoA properties of the channel, which are often characterised by the angle spread. The angle spread has a major impact on the correlation observed between the pairs of elements in the array. These correlation values specify the received signal vector covariance matrix, which is known to determine the performance of linear combining arrays. In general, the higher the angle spread, the lower the correlation observed between the pair of elements in the array. The various spatial channel models provide

different extents of angle spreads and hence are capable of predicting different levels of system performance.

6.3 SMART ARRAYS: ANTENNA AND DIVERSITY GAINS

As indicated earlier, the smart antenna techniques (based on antenna array technology) were conceived primarily to overcome the following three major impairments in the performance of wireless communication systems:

- Multipath fading resulting from different paths traversed by a transmitted signal arriving at a receiver. (As a consequence, the received signal is a superposition of components with varying extents of phase-difference (due to multipath traverse) posing a fading of intensity at the receiver antenna)
- Delay spread stemming from difference in propagation delays associated with multipath traversing of the signal
- Co-channel interference arising from the cellular systems dividing the available frequency channels with a frequency reuse scheme in which each cell uses a specific channel set, which is reused in another cell.

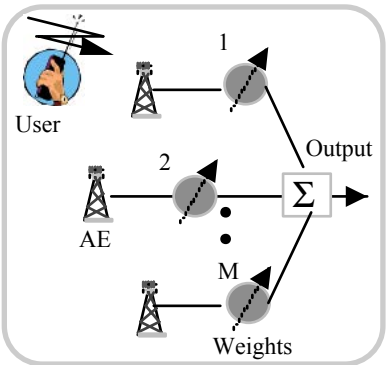


Fig. 6.13 Antenna array technology constituted by a set of M radiators and array signal processing elements

In order that a greater coverage and a larger capacity at each base station are achieved (despite of the aforesaid inevitable impairments), the array technology was ushered into wireless cellular systems. Basically, the relevant technique deploys a set of M-element antenna elements constituting an array at the base station (Figure 6.13).

The design of an array antenna, in general, is aimed at realising an enhanced *antenna gain* of M together with a *diversity gain* against multipath fading. The antenna gain is defined as the reduction in required received signal power for a

specified (average) output signal-to-noise ratio (SNR) and this SNR is independent of the environment.

The diversity gain is specified only under a multipath fading environment. It refers to the reduction in required average output SNR for a specified BER with the fading being present. It depends on the correlation of signal fading among the antenna elements. There are three techniques for provisioning diversity gain (through realising a low correlation between antenna elements). They are as follows:

- Space diversity
- Polarisation diversity
- Angle diversity

Space diversity requires the antenna elements to be spatially separated far enough so that a low correlation of signal reception prevails between them. The required separation is decided by the angular spread, which refers to the angular sector over which the signal arrives at the receiver antenna.

For handheld sets with a number of scattering objects surrounding it, the angular spread is typically 360°. In this case, a λ/4-spacing of the antennas will be sufficient to achieve low correlation. Similar considerations hold for BS antennas in indoor systems. However, for outdoor BS systems with tall tower structures (designed to avoid ground clutter), the angular spread could be about a few degrees (or somewhat higher in urban areas). In such cases, a horizontal separation of 10 to 20 wavelengths may be required to obtain a reasonable extent of low correlation coefficient. This would, however, make the antenna array size *versus* the real estate an issue in practical implementations.

For *polarisation diversity*, the horizontal and vertical polarisations are utilised. These (orthogonal) polarisations pose a low correlation and size-wise, the antennas can have a small profile. But the maximum diversity gain is limited to at most, 2 when the polarisation diversity is adopted. Normally, for tall BS antenna structures, the azimuthal polarisation is about 6 to 10 dB weaker than the vertical polarisation. As a result, the diversity gain is further reduced in practice.

Angle diversity refers to using narrow beams adjacently. The antenna profile in this case is also small and the narrow beams render the low fading correlation. But, a small angular spread of the adjacent beam would make the received signal strength to be about 10 dB weaker than the strongest beam. This may eventually cause a reduced diversity gain.

The three diversity considerations can be summarised in terms of their relative performance profiles. For this purpose, the four configurations illustrated in Figure 6.14 can be considered. Figure 6.14 (a) depicts a multielement (M) spatial diversity arrangement with approximately 7λ-separation between the antenna elements so as to yield 120° azimuthal sector coverage. Each element offers 120°/M horizontal beamwidth and the vertical beamwidth is small (about 8°-10°). In essence, the array yields a horizontal fan-beam of over 120°-sectoral coverage. Figure 6.14 (b) illustrates a dual-polarised antenna array system with the array elements being closely spaced. (That is, no spatial diversity gain is

facilitated.) The antenna element is comprised of 45° slant polarisation rotations, which are more common than horizontal-vertical arrangement.

Figure 6.14 (c) is also a dual polarisation antenna array with the elements separated sufficiently apart, so that spatial diversity gain is also realised concurrent to polarisation diversity gain.

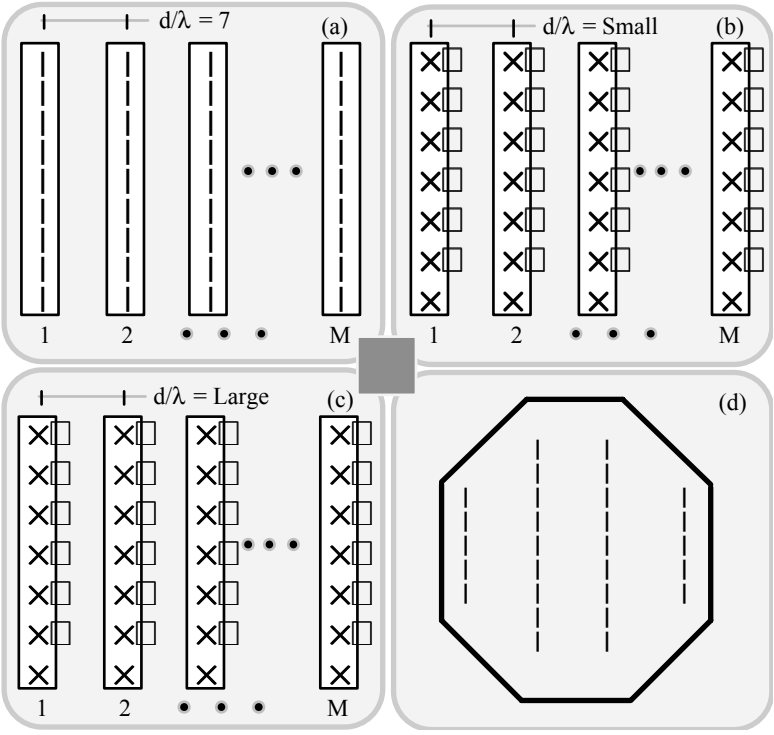


Fig. 6.14 Diversity options with different antenna array structures. (a) Space diversity; (b) polarisation diversity; (c) space and polarisation diversity; and (d) angular diversity

The angular diversity can be achieved by a set of closely spaced (λ/2) vertically-polarised antennas as shown in Figure 6.14 (d). This array can be configured in a small geometrical profile. Performance-wise the combination of space and polarisation diversity arrangement (Figure 6.14c) would offer a relatively higher diversity gain than other structures. Typical performance aspects of the structures of Figure 6.14 are compared in Table 6.1.

With reference to smart antenna applications under low-fading correlation, the diversity gain is realised at the base stations by resorting to a *selection diversity criterion*. That is, the highest signal power or the maximum SNR computed with

the received signals (weighted and combined) from a system of antennas (at least two) is deployed.

Diversity considerations can also be applied to handsets. Normally, a handset faces an infestation of scatterers around it. The antenna spacing required in such cases is about $\lambda/4$ for space diversity under low-fading correlation. It means that, multiple spatial diversity antennas are compatible for use within the handset.

Table. 6.1 Typical relative performance characteristics of antenna arrays under different diversity options

	Typical parameters		
Array type	Element spacing	Operational conditions	Diversity gain
Space diversity, Figure 6.14 (a)	~ 7λ	f = 1900 MHz Number of elements: 4 BER: 10^{-2}	4.2
Polarisation diversity, Figure 6.14(b)	~ λ		2.9
Space plus polarisation diversity, Figure 6.14 (c)	~ 7λ		4.4
Angle diversity, Figure 6.14 (d)	~ $\lambda/2$		1.1

Dual polarisation can also be facilitated in such systems. The limitation of such systems could, however, stem from the cost and power consumption involved in the receiver electronics at each antenna, whereas size considerations would form secondary limiting factors.

6.3.1 Diversity combining techniques

The diversity combining schemes indicated above allow antenna arrays to combine judiciously the diverse signals so as to cope with the fading channel characteristics. There are three methods adopted towards implementing diversity combining techniques. They are as follows:

- Selection diversity combining
- Maximal ratio combining
- Equal gain combining.

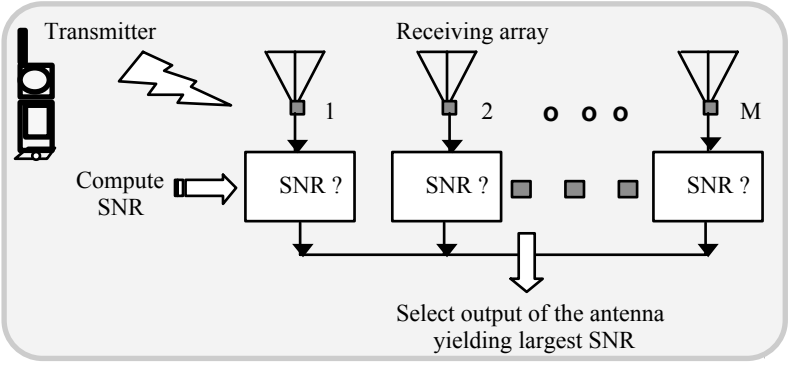


Fig. 6.15 Selection diversity combining scheme based on maximum output SNR criterion

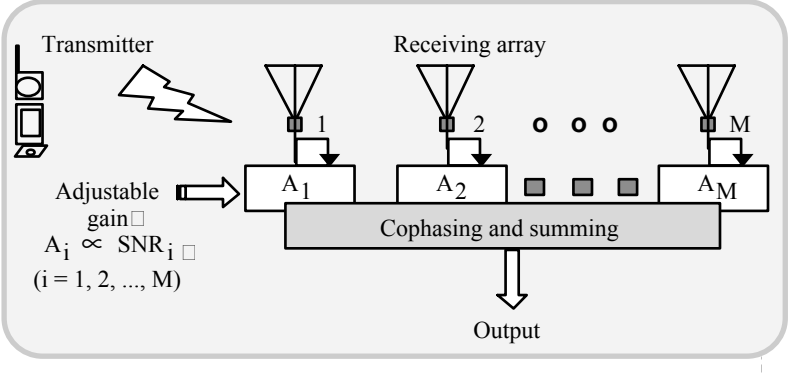


Fig. 6.16 Maximal-ratio combining method based on weighted in-phase combining of branch signals of the array

The selection diversity combining is illustrated in Figure 6.15. It involves a set of M antennas and the (output) signal is acquired from an antenna of this array such that it yields the largest signal-to-noise (SNR) at the output. The larger the number of M, the higher is the scope of better signal selection. This system is much similar to matched filter designs.

The so-called *maximal-combining technique* takes advantage of all the diversity branches of the array. As shown in Figure 6.16, all M branches of the array are weighted with their respective signal-to-noise ratios. The branches are then co-phased prior to summing in order to ensure that all branches are added in phase so as to realise maximum diversity gain. This method is more sophisticated than the selection diversity scheme.

The equal-gain combining scheme is a variation of maximal-ratio combining. It involves setting the gains of all the branches to the same value and kept invariant. Then, the output is obtained by co-phasing and summing the outputs of all the branches (Figure 6.17).

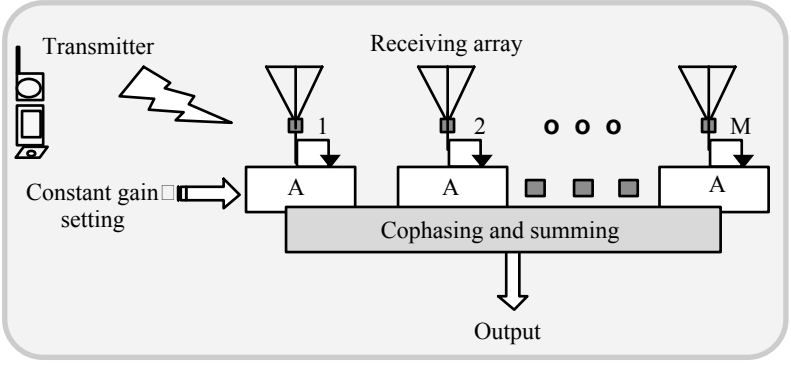


Fig. 6.17 Equal-gain combining scheme: The output corresponds to in-phase summing of equally weighted branch signals of the array

6.3.2 Types of smart antennas

As mentioned before, smart antennas are essentially "adaptive" antenna arrays, which have emerged specifically to enhance the performance of virtually any CDMA system, including IS-95, IMT-2000 and wideband CDMA. These structures are constituted by appropriate beamforming networks that are supplemented by multi-user spatial-processing systems. A typical smart antenna configuration is illustrated in Figure 6.18.

A version of smart antennas refers to those which can change patterns dynamically, adjusting to noise, interference, and multipath conditions as they track mobile users. The smart antenna technology, in essence, uses an intelligent combination of the signals received at multiple antenna elements at the base station (BS) to mitigate the effects of interference-based performance impediments. Such

signal combination is done efficiently and accurately with a diligent knowledge of the propagation channel that exists between the mobile and each of the elements.

The adaptive array concept of smart antenna developments used in modern wireless communications is essentially based on the wealth of knowledge acquired through phased-array technology developed for radar applications. As illustrated earlier in Figure 6.13, the relevant system is comprised of a set of antenna elements arranged in space. The output of each element is multiplied by a complex weight and combined *via* various summing techniques presented in Figures 6.15-6.17. The resultant output of the array is therefore decided by the individual radiation pattern of each element, the complex (amplitude and phase) weighting applied and by the spatial (geographical) disposition of the elements themselves.

Further, the phased-array output could be adaptive when the weights imposed vary with time so as to exploit and improve the receiver performance against the time-varying channel attributes resulting from the mobility of the user. The adaptive manoeuvring of the weights is based on an algorithm that optimises some metric of the system performance. Practically, this algorithm is implemented *via* digital signal-processing (DSP). The implementation refers to estimating the desired weights (through DSP) and applying them in complex baseband to a sampled version of the signal from each of the elements.

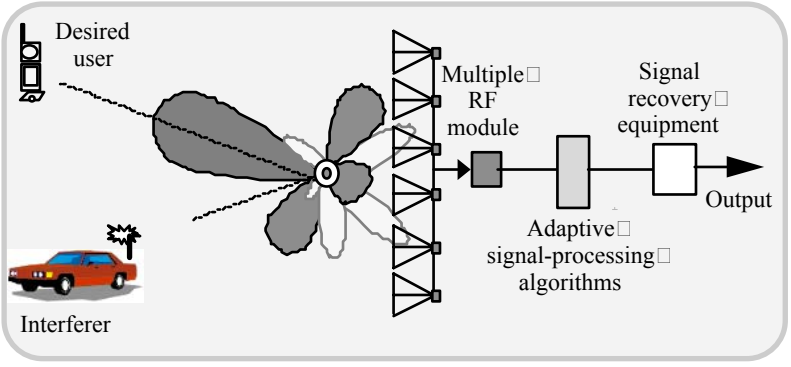


Fig. 6.18 Smart antenna configuration and its associated signal handling sections

The concept of adaptive phased array could be used both for transmit and receive purposes in view of the reciprocity considerations applied to channels and antenna characteristics. However, applying this concept to downlink channels (constrained by a prescribed accuracy) would rather be challenging. The reason is as follows: In the downlink, should an adaptive phased-array be deployed (at the receiving end), it would involve choosing the best transmit antenna at any given instant and choosing the amplitudes and phases of the signals transmitted so that they combine in-phase at the receiver. However, this process would pose difficulties in availing information about the downlink channel as a result of the differences of time and frequency that prevail between the incoming and outgoing channels.

The major aspect of adaptive arrays used in wireless systems is their performance in multipath environments versus line-of-sight (LoS) condition. Classical studies in phased-array systems mainly address the LoS environments. The objective in such efforts refer to realising a null in the radiation pattern in the direction of interference (or undesired) signal received, and the maximum of the radiation pattern is steered towards the direction of desired signal arrival. Choice of the appropriate number of array elements and phase of the signal supported at each element enables the control of steering the null and/or the pattern maximum towards a desired direction. Algorithms, like MUSIC and ESPRIT have been developed to determine the direction of arrival of rays. In general, an array with M antenna elements can form up to (M – 1) nulls to cancel up to (M – 1) interferers. Relevant near-LoS specified algorithms can be used for adaptive antenna designs pertinent to mobile radio base stations operating in flat rural environments deploying many (for example, eight) tall antenna structures. Here, the number of antennas used is usually in excess of the number of arriving-signal rays (since multipath arrivals are assumed to be absent). As such, the associated arrivals conform to a small number of angles of arrivals. The algorithms, therefore, are specified to express the array response in terms of such small number of angles of arrival rather than the received signal phase at each element.

Considering the multipath situation, the signals arrive from each user *via* the directions of multiple-path angles of arrival. As such, it would not be possible to form an antenna pattern with a beam in the direction of each arriving path of the desired signal and achieve nulls in the directions of interfering signals. This is because, the number of nulls required will be much greater than the number of antenna elements themselves; and synthesising an array to yield such a pattern is not possible. Further, in order to achieve a specified diversity gain, the antennas should be spaced several wavelengths apart. This would result in many *grating lobes* (that is, repeated set of lobes in the field of view of interest). Also, if dual polarisation is adopted, the resulting field pattern would be different in different polarisation planes. In essence, the task of synthesising a desired field pattern through the antenna array becomes fruitless. But, regardless the number of paths a signal may traverse, the result is a specified set of amplitude and phase at each antenna for each signal. That is, there exists an array response corresponding to each signal. As such, the performance of an array is decided by the number of signals rather than by the number of paths traversed. In other words, the array processing would essentially refer to an analysis in the signal space domain and not in the angular domain. This consideration is valid as long as the delay spread is small. Otherwise, the delayed signals have to be treated as separate signals. Thus, an adaptive array (with the signal-processing deliberated in the signal space) can facilitate nulls in the pattern towards (M – 1) interferers independent of LoS or a multipath specified environment.

Hence, such adaptive arrays in multipath environments have the ability to cancel interferers independent of the angle of arrival. That is, an interferer proximally located in the vicinity of a mobile unit (posing a small angular separation from the mobile unit) can be resolved. Such resolution (and separation) of closely-spaced signals is not, however, feasible in LoS-specified environments.

In general, with adaptive array processing, the number of signals that can be separated increases with the number of receiving antennas, the angular spread and the density of the multipath reflection prevailing within the angular spread.

Should delay spread be significant, as mentioned above, the array could treat the delayed versions of the signals as discrete signals. In such cases, an adaptive array with M antennas can cancel (M – 1) delayed signals over any delay involved. However, temporal equalisers are typically needed in such cases in combination with the spatial array processing to counteract any temporal distortion.

6.4 TRACKING AND SWITCHED BEAM ARRAY TECHNIQUES

In the foregoing sections, the art of producing a smart antenna concept at the base station so as to enhance both communication channel capacity and the quality of the links (in terms of reduced interference and/or fading effects) was discussed. It was seen that the relevant approach conforms to generating an appropriate beam pattern that has maximum gain along the direction of the target subscriber in a given cell or a sector. Thus the smart antenna, functionally is a spatial filter. It is a selective, bandpass (gating) system, which has pass- and stop-bands (gates) created along the directions of the (desired) signal and the (undesired) interferers respectively.

Now, the question is how to track the target mobiles. There are two possible approaches that have been advocated [6.6]. The first scheme refers to tracking the target user with steerable beams; and, in the second method, a beam that locates the (spatial) location of the target user is selected out of a set of previously designated (fixed) beams.

Further, in either case, the information on the desired beam pattern is made available to each mobile subscriber served by the base station both in the transmitting and receiving modes; and, this beam-pattern detail is ascertained adaptively with the mobility of the individual user concerned. That is, the desired beam pattern is (dynamically) obtained via the antenna arrays.

The associated (participating) lobe of the antenna array pattern is narrow, depicting a gain much higher than conventional antennas. In general, the traditional antenna creates a single lobe that covers the entire cell or sector. A smart antenna on the other hand, uses a number of narrow lobes to cover the same cell/sector. The choice of antenna elements that constitute the smart antenna array should conform to the following characteristics:

- Extremely low cross-polarisation levels (theoretically, null in the principal planes)
- Devoid of or reduced effects pertinent to the interaction between radiating elements and corporate feeding network (so that the shift in resonant frequency, radiation from the feeding network, etc. are minimal)
- Feasibility to support dual-polarisation capability.

As indicated above, smart antennas may vary from switched-beam configuration to fully adaptive arrays. Since the switched-beam arrays use beamforming techniques

that yield multiple, fixed and simultaneously available lobe pattern, they should have high gain and low sidelobes or controlled beamwidth.

The adaptive array schemes, on the other hand, use adaptive beamforming techniques and bear the ability to reject interfering signals having a direction of arrival different from that of a desired signal. The multipolarised version of adaptive arrays can reject interfering signals having polarisation states that differ from that of the desired signal. (Such polarisation-specified rejection is feasible even if the desired and undesired signals have the same direction of arrival.)

6.5 FIXED BEAMFORMING STRATEGIES

With regard to the class of cellular telephone applications of smart antennas that need fixed beams (covering an angular sector), there are several beamforming techniques. Relevant techniques commonly make use of the Butler- or Blass matrix systems.

Briefly, the Butler matrix, as described in Chapter 5, is a beamforming network that uses a combination of 90^{o}–hybrids and phase shifters. For example, a (8×8) Butler matrix performs a special fast Fourier transform (FFT) and provides 2^3 orthogonal beams. When used with a linear array, the Butler matrix produces beams, which overlap at about 3.9 dB below the beam maxima. This matrix can cover a sector up to 360^{o}, depending upon the pattern and the spacing of the individual elements. Then, a dedicated transmitter and/or receiver can use each beam. Alternatively, the transmitter and/or receiver can be selected for use with an appropriate beam (on an *ad hoc* basis) by means of an RF switch. The Butler matrix can also be used to steer the beam of a circular array by exciting the beam ports with specified amplitude/phase weights.

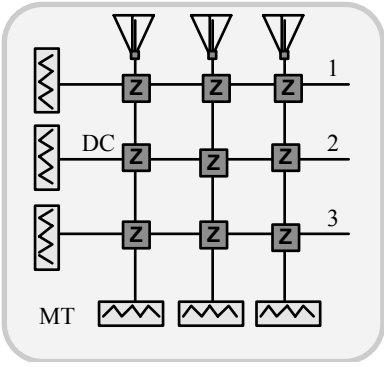


Fig. 6.19 Example of a Blass matrix: A 3-element configuration for producing a broadside beam. (MT: Matched termination. DC: Directional coupler)

The Blass matrix uses transmission lines and directional couplers to form beams *via* time-delay considerations. It is suitable for broadside applications. Considering the three-element Blass matrix illustrated in Figure 6.19, port 2 provides equal

delays to all the elements resulting in a broadside beam. The other two ports facilitate progressive time delays between the elements and produce beams, which are off the broadside direction. The Blass matrix is lossy because of the resistive termination and can be designed with any number of elements.

As indicated in Chapter 5, fixed beams can also be formed using lens antennas such as Luneberg or Rotman lenses with multiple feeds.

6.6 ARRAY PROCESSING THROUGH BEAMFORMING

The essence *of array processing* in cellular communications refers to beamforming techniques, essentially for the downlink of transmissions. The objective is to selectively include (or constructively combine) signals supporting the same information so that the receiver gets the desired signal, whereas the undesired interference and noise can be nulled or suppressed.

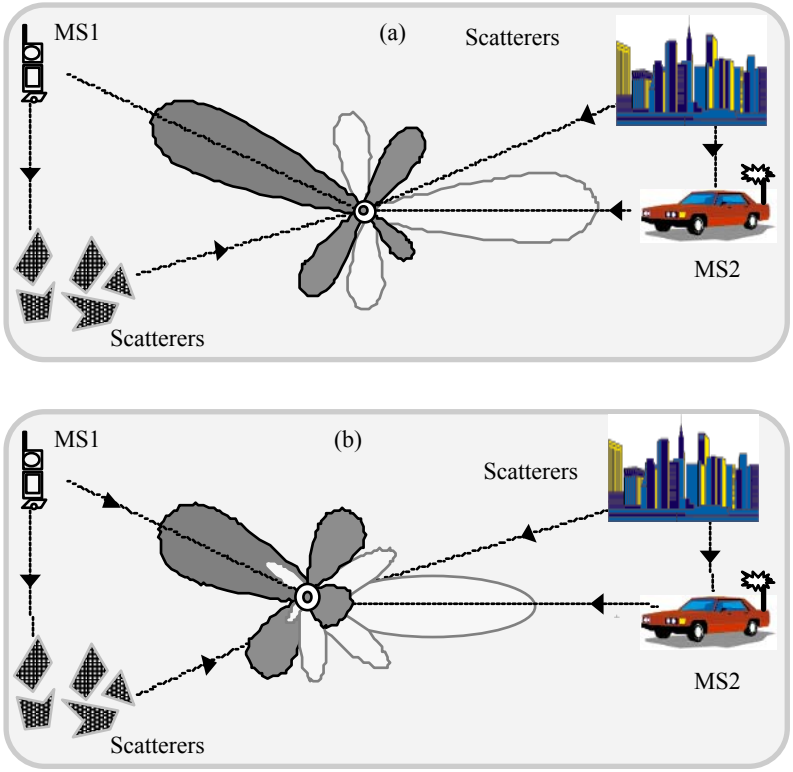


Fig. 6.20 Beamforming techniques: (a) DoA/LoS-based method and (b) spatial-signature specified strategy

The beamforming technique as mentioned before, could be based on the angular *direction of arrival* (DoA), (also known synonymously as the line-of-sight, LoS) of the signals or on the *spatial signature* of the signals. The former computes the

downlink weight vectors based on the array response vectors, while the latter users the spatial signature of the signals.

The two schemes are illustrated in Figures 6.20(a) and 20(b). In Figure 6.20(a), mobile stations 1 and 2 (MS1 and MS2) are served based on the LoS signals received from them. In Figure 6.20(b) such service is rendered by analysing the characteristics of the spatial signatures in the signals from MS1 and MS2 taking into account the undesired components due to the scatterers in the vicinity.

6.6.1 Basic beamforming algorithms

DoA/LoS-based beamforming algorithms

This approach is adopted in DoA/LoS-specified applications and is based on two methods, namely,

- Dominant DoA/LoS specified technique
- Pseusoinverse DoA/LoS specified technique

The first approach captures the spatial signatures of the signals and determines the DoAs of the signal using a subspace-based computation (like MUSIC or ESPRIT), and the amplitudes associated with the DoAs are estimated. Then, the DoA corresponding to the maximum amplitude is selected and its array response vector is chosen as the weight vector for that link to carry out further processing.

Considering, a set of M array elements, the received signal x(t) at an antenna can be written as a snapshot vector model given by,

$$\mathbf{x}(t) = \alpha_1 \mathbf{a}(\theta_1) s_1(t) + \sum_{\ell=2}^{M} \alpha_\ell \mathbf{a}(\theta_\ell) s_1(t) + \mathbf{n}(t) \qquad (6.6)$$

where the first term is the direct-path component α_ℓ, refers to the complex phase and amplitude of the ℓ^{th} element or multicomponent, and θ_ℓ is the angle of arrival of the ℓ^{th} component. Further, α_ℓ depicts the associated spatial-signature.

The second approach, namely, the pseudoinverse DoA algorithm is similar to the previous one except that a pseudoinverse of the array response vectors of all the DoAs is ascertained (instead of just that for the DoA of the desired user). This method places nulls in all DoAs except for the desired user so as to minimise interference. For example, suppose the k^{th} mobile unit has one direct path signal and causes a multipath signal with DoAs, $\theta_1^{(k)}$ and $\theta_2^{(k)}$ respectively. The corresponding weight vectors are designed as follows:

Suppose there are two independent sources 1 and 2. Designating their weight vectors by $\mathbf{w}_1$ and $\mathbf{w}_2$ respectively, the orthogonality conditions, namely $\mathbf{w}_1 \perp \left\{ \mathbf{a}\left(\theta_2^{(1)}\right), \mathbf{a}\left(\theta_1^{(2)}\right), \mathbf{a}\left(\theta_2^{(2)}\right) \right\}$ and the identity $\mathbf{a}^{\mathbf{T}}\left(\theta_1^{(1)}\right)\mathbf{w}_1 \equiv 1$ will then render the received signal exactly equal to the desired one from source 1. That is, even if both sources transmit co-channel signals, mobile user 1 would receive only

the desired signal (from source 1). This is enabled by the fact that the choice of $\mathbf{w}_1$ is such that the transmission pattern of the antenna array has nulls in all directions of arrivals except at $\theta_1^{(1)}$. Likewise, $\mathbf{w}_2$ can be designed such that the pattern has its nulls in all the estimated DoAs except at $\theta_2^{(2)}$, in which case the user 2 will receive the desired signal from source 2 only. (In the above discussion, $\mathbf{a}^{\mathbf{T}}(.)$ denotes the transpose of the vector matrix **a**.)

Spatial signature-based algorithms

In these techniques, the spatial signatures of the links are captured and used to generate the weight vectors. Relevant approaches are as follows:

- Complex-conjugate spatial signature algorithm
- Pseudoinverse complex-conjugate spatial signature algorithm

In the first method, the main objective is to maximise the signal power or the signal-to-noise ratio. It does not attempt to null out any directional interference. The relevant technique involves taking the complex conjugate of the spatial signature.

The second approach is similar to the above except that the weight vectors are obtained by taking the pseudoinverse of the spatial signature. That is, the pseudoinverse weight vector $\mathbf{w}_i$ is specified on the basis of the spatial signature of the sources such that $\mathbf{w}_i \perp \mathbf{a}_k$ for $i \neq k$ and $\mathbf{a}_i^T \mathbf{w}_i = 1$.

In this approach, the weight vector adjusts the relative phase and amplitude of the direct path and multipath component so that the signals exactly cancel out at the location of the signal of interest. For example, if there are two discrete sources 1 and 2, the corresponding weight vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ are such that $\mathbf{a}_1^T \mathbf{w}_1 = 0$ and $\mathbf{a}_1^T \mathbf{w}_2 = 0$, corresponding to the pattern of source 1. However, the array pattern relevant to source 2 does not imply placing nulls at all the DoAs associated with source 1. All it requires is that $\mathbf{a}_1^T\left(\theta_1^{(1)}\right)\mathbf{w}_2 \equiv -\alpha_2 \mathbf{a}_T\left(\theta_2^{(2)}\right)\mathbf{w}_2$ not warranting $\mathbf{w}_2$ to be not orthogonal to both $\mathbf{a}\left(\theta_1^{(1)}\right)$ and $\mathbf{a}\left(\theta_2^{(2)}\right)$. That is, $\mathbf{w}_2$ is designed to control the phase and amplitude of the two paths such that they exactly cancel out at the location of mobile user 1.

The smart system concept, as above, can be accomplished by resorting to one of the two methods, namely, the tracking and switched-beam processing schemes mentioned before. The first method refers to adaptively tracking each subscriber in a given cell with an individual beam pattern as the target subscriber moves within that cell. The second technique is a method of selecting a beam pattern for each subscriber (out of a number of predesigned patterns) who is attributed with a selection algorithm in reference to the base station.

Functionally, considering the adaptive tracking strategy, the beam structure adapts to the (changing) RF signal environment and directs beams to the *bona fide* signals (from the target subscriber), depressing the pattern in the direction of the interferers. Such systems are generally more DSP-intensive than the switched-beam systems and are more expensive. The analytical description of adaptive arrays is also mathematically intensive.

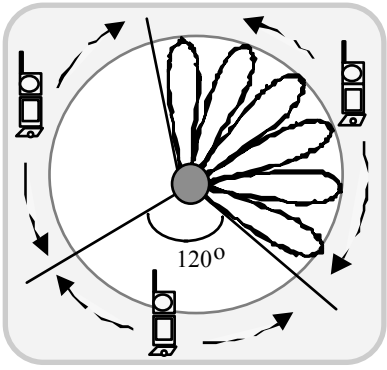


Fig. 6.21 Directional beams for a smart antenna

The switched-beam array generates a multiplicity of juxtaposing beams whose output may be switched to a receiver or a bank of receivers. The cell is therefore blanketed with a cluster of contiguous beams. Relevant systems deploy a fixed set of relatively narrow azimuthal beams, which are generally narrower than those used in a 120º-sectorised system*.

The RF output to the beams is either an RF or baseband signal that is digitally processed so as to conform to the sector in which the communicating mobile may be located. The cellular region is divided into three sectors (each with a coverage of 120º) and each sector is served by an array of radiating elements fed by a beam-forming network, which ideally forms independent beams.
Assuming a set of six beams in each 120º sector (with each beam having a nominal beamwidth of 20º in azimuth), a display of the beams is presented in Figure 6.21.

**Sectorisation* refers to cell sectorisation for the purpose of increasing the capacity of a cellular system in the presence of limited spectrum. The cell is divided, for example, into 120º sectors, which implies that each sector is served by a directional antenna with a nominal beamwidth (in azimuth) of 120º. Smart antennas are basically an extension of beam sectorization in which the sector coverage in supplanted by multiple beams. This is achieved by the use of array structures, and the number of beams in the sector (for example, 120º) is a function of array extent.

6.6.2 Adaptive array configurations

The adaptive array nominally modifies its antenna pattern structure so as to place a directional beam in the direction of a *bona fide* signal (from a target subscriber). Thus, it maximises the SNR and nulls in the direction of the interferers.

Relevant to this genre of adaptive array of smart antennas, the underlying versions of implementation are known as the *Applebaum* and the *Widrow configurations*. These are illustrated in Figure 6.22

In the Applebaum configuration, the weights are adjusted dynamically on the basis of instantaneous output. On the other hand, the weights are formed in the Widrow version by computing the difference between the array output and a reference signal (which is a replica of the desired signal). The former requires prior knowledge of potential source locations. This is facilitated by means of a *search-light (tracking) mode* through steering the vector information.

The Widrow genre requires some *a priori* knowledge of the desired signal (The spatial signal location is not, however, required). Hence, an error signal is generated. It consists mainly of interference components.

Using the so-called *least-mean-square* (LMS) algorithm, the error signal is cross-correlated with the incident input samples to obtain weighted updates for the weights in the elements. Minimisation of the error corresponds to minimising antenna response to these signals, which results in pattern nulls in the direction of interferers.

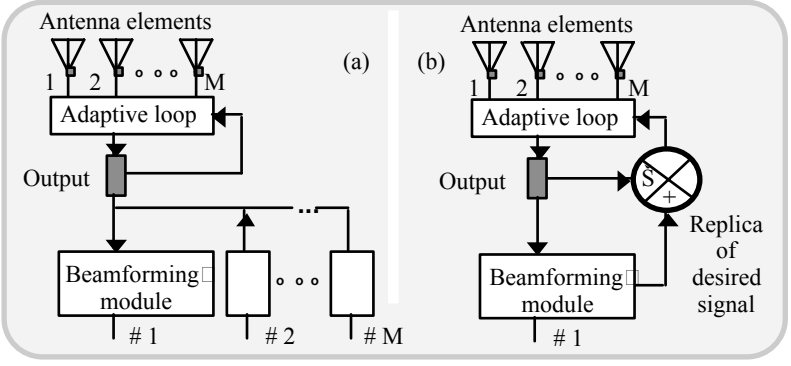


Fig. 6.22 Adaptive array configurations: (a) Applebaum version and (b) Widrow version

6.6.3 Switched-beam array configuration

This array technology, as indicated before, uses a beamforming network to provide fixed directional beams with output ports corresponding to fixed-beam positions. Output is sampled either simultaneously or sequentially switched to ascertain beam activity. Handoff between beams is warranted whenever the users go through a tangential motion.

A generic version of the Butler matrix that can be used as a switched beamforming network is illustrated in Figure 6.23. There are two antenna elements separated by a distance d. A quadrature hybrid drives these two elements with specific phase/amplitude relations as indicated. The elements are identically energised in amplitude but are differentially phased. The resulting beam has a peak right to the boresight. The beam offset angle $\theta_1 = \arcsin(-\lambda / 4d)$.

The Butler matrix, in essence, enables a feed system for an array to produce a multiple number of beams in a sector. The feed system has a number of input ports with each input port exciting the array so as to produce one of the multiple beams, all offset from each other by a finite angle.

As illustrated in Figure 6.23, a switching matrix is used to address a particular user. In essence, the system relies on forming virtual beam fixed in space with an array factor weighted by an appropriate element factor. Typically, a Butler labyrinth of the type shown in Figure 6.23 can facilitate this beamforming.

Illustrated in Figure 6.24(a) are beamforming matrices that use a combination of hybrid junctions and fixed-phase shifters to achieve the desired results. The feed system in Figure 6.24(a) consists of a hybrid junction such as a waveguide magic-tee and a (90°)- fixed-phase shifter.

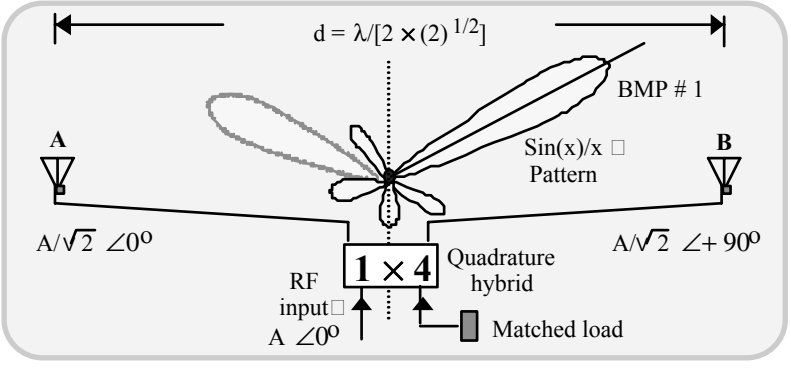


Fig. 6.23 A basic two-beam Butler matrix. (BMP # 1: Beam formed with input at port # 1)

This configuration produces the quadrature hybrid formalism depicted in Figure 6.24(b), where the hybrid junction has the property that ports 1 and 4 as well as ports 2 and 3 are uncoupled. Also, the transmissions from port 1 to ports 2 and 3 are equal, while the transmissions from port 4 to ports 2 and 3 differ in phase by 180°.

Consider now the effect of exciting port 1. The signal delivered to element B in the array will be (– 90°) out-of-phase with respect to the signal delivered to

element A. Hence, the beam is formed in a direction that will make the path length ℓ (from element A), ($\lambda_o/4$)-longer than that from the element B. (This is done deliberately in order to compensate for the (– 90°)-phasing of element B, so that the relevant beam direction will then be 45° left of the centre line.)

When port 4 is excited, the phase of element B will be advanced by 90° relative to that of element A and thus produces a beam at an angle of 45° on the right side of the centre line. Since ports 1 and 4 are uncoupled, these beams are independent and may exist separately or simultaneously.

The principle involved in this two-element array may be extended to arrays with many elements and will result in a number of independent uncoupled input ports, each of which produces a single beam using the full gain capability of the array. The number of elements in the array must equal a power of 2 in order to construct the beamforming matrix using hybrid junctions.

Another way of realising the functionality of a quadrature hybrid is by using directional couplers as illustrated in Figure 6.25. An example of Butler matrix that facilitates a beam which can be switched to four positions by exciting the input appropriately, is illustrated in Figure 6.26. It uses four quadrature hybrids plus two fixed phase-shifters.

In general, the number of beams is equal to number of elements used. Each beam has a beamwidth (in degrees) equal to $[50/(N-1)] \times (\lambda/d)$, where N is the number of elements and λ is the wavelength. Further, the location of the beams θ (that specifies the angle off the bore-sight) is decided by the following relationship:

$$\sin(\theta) = (\pi/Nd)[k - (1/2)], \quad k = 1, 2, \ldots, (N/2) \tag{6.7}$$

where k is known as the *beam number*.

The far-field intensity of the array is given by the well-known product of the *array factor* $F(\phi)$ and the *element factor* $E(\phi)$. Explicitly, $F(\phi)$ is given by

$$F(\phi) = \sin(N\phi/2)/N\sin(\phi/2) \tag{6.8}$$

with $\phi = (2\pi d/\lambda)\sin(\theta) - \delta$ where δ denotes the progressive phase difference generated by the matrix. It is specified for a beam number k as $(2k-1)\pi/N$. Butler arrays can be built to have any power of two beams: 2, 2^2, 2^3, 2^4, 2^5, and so on.

In short, the principle of the Butler matrix in beamforming consists of the following: A signal input at any one of the transmitter/receiver ports should produce equal amplitudes at all of the antenna ports, and a linear phase progression from port-to-port. The phase slope depends on which transmitter/receiver port is used. When the antenna ports are connected in sequence to an equally-spaced linear array, N beams are formed, one for each transmitter/receiver (TX/RX) port.

Combining externally the ports of two or more of the Butler matrix TX/RX ports, it is possible to move or broaden the antenna pattern. Also, the sidelobe levels can be improved selectively in preferred directions. Adaptive nulling and beam-shaping permit the antenna patterns to "adapt" to the signal environment. In

smart antenna applications, this adaptive nulling and beam-shaping is accomplished on dynamic (time-dependent) basis.

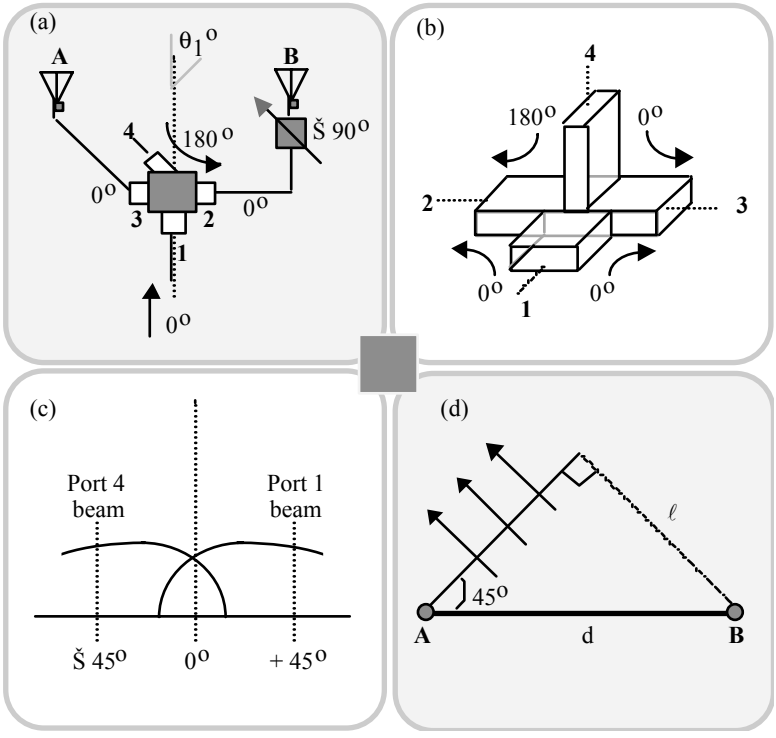


Fig. 6.24 Principle of a quadrature hybrid realised using a magic-T: (a) A two-element array; (b) magic-T; (c) beam *versus* excitation ports; and (d) delay path (ℓ) improvised to form port 4 beam

A two-dimensional array based on a Butler matrix can be obtained by incorporating a Butler matrix incorporated with each vertical array and then connecting each horizontal row of ports to a Butler matrix. Thus, an 8×8 matrix would lead to 64 beams. Again, by combining groups of TX/RX ports from the 64 Butler matrix ports, it is possible to obtain beams with acutely desired shapes and sidelobe characteristics.

In obtaining Butler matrix-based array architectures, a set of hybrid-matrix amplifiers (HMAs) are used. A HMA corresponds to a parallel set of amplifiers whose inputs are fed and the outputs are combined using multi-port matrices made of hybrid couplers. Connecting hybrid matrices back-to-back enables creating information paths that are segregated at input and output ports (that is, for example,

considering a 4 × 4 matrix, the signal entering the port input # 1 would exit at output port # 4 etc.).

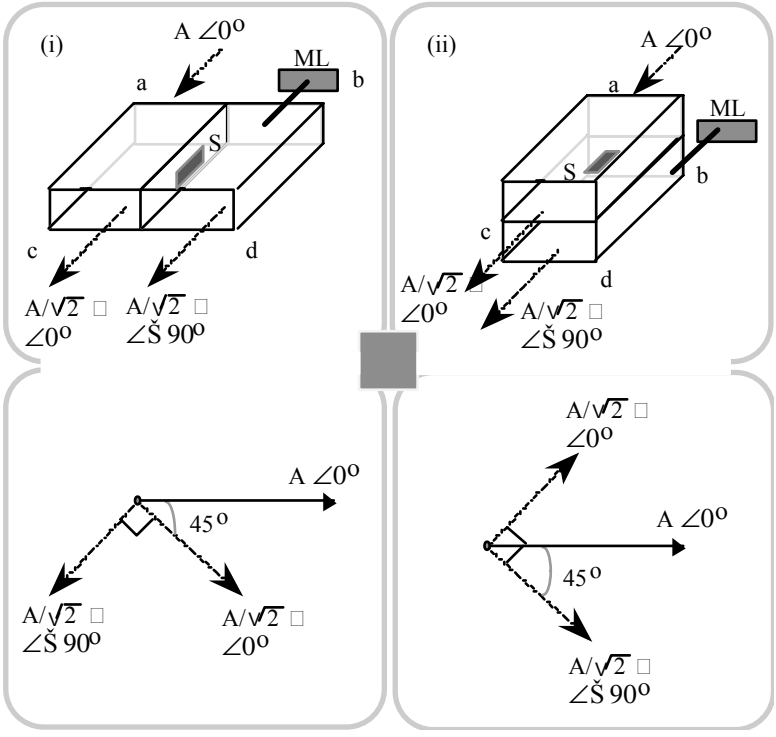


Fig. 6.25 Directional coupler-based quadrature hybrid realisations: (i) Side-wall coupled hybrid; (ii) top-wall coupled hybrid; S: Slot-coupling at the interspaced wall between the waveguides; and, ML: Matched load

Hence, at the nodes depicting the connections between the matrices where the amplifiers are located, the signals are distributed evenly in amplitude and with a specific phase relationship on the input port signals. By provisioning matched amplifiers between the two matrices, the signals on all paths are amplified.

The HMAs carry the ability to apply all amplifier assets to any or all of the input/output ports. This is significantly useful in applications where the transmitter has multiple antennas and the HMA allows the antennas to share the ouput optimally. (At the same time, the signal traffic for each sector or beam is maintained independently.) This optimal sharing concept facilitates cost-effective systems. The HMAs also provide a high dynamic range with redundant amplifiers.

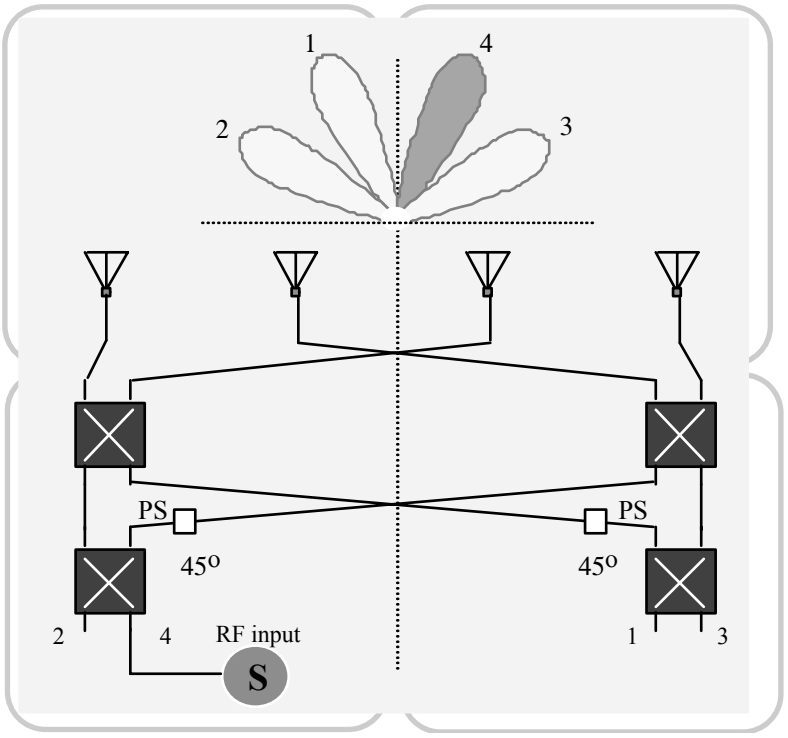


Fig. 6.26 Synthesis of a four-position, switchable beam using a Butler matrix. (PS: Phase shifter)

Example 6. 1
Analyse the beamforming matrix for the four-element array of Figure 6.27.

Solution
In reference to Figure 6.27, the matrix uses 90° phase-lag hybrid junctions with the transmission properties.

By tracing the signal from the four input ports to the array elements, one can verify that the following aperture relative phase distributions are established:

Port 1:	0°	– 135°	– 270°	– 405°
Port 2:	0°	– 45°	– 90°	– 135°
Port 3:	0°	+ 45°	+ 90°	+ 135°
Port 4:	0°	+ 135°	+ 270°	+ 405°

The system thus produces four separate beams, as illustrated in Figure 6.27.

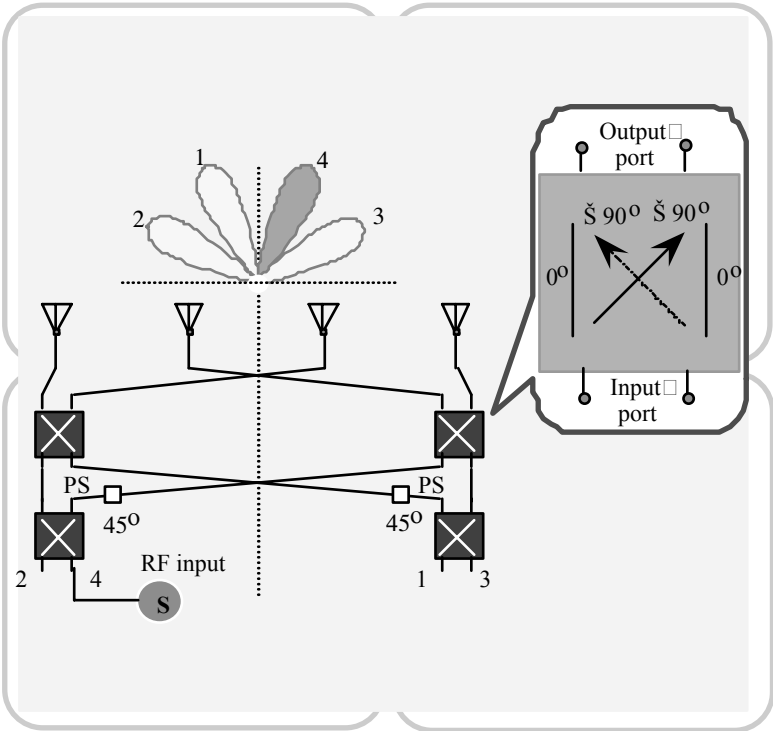


Fig. 6.27 Illustration of the role of quadrature hybrid in the four-element Butler matrix: The hybrid function offers a 90° phase-lag transmission relationship

Example 6.2
Design a Butler matrix for use in a linear array to produce eight switchable spot-beams over a sector of 120°.

Solution
Topology
A conceivable beamforming matrix of the Butler type for use in a linear array so as to produce eight contiguous beams over a specified sector, is illustrated in Figure 6.28. Since eight beams are required, there should be eight antenna elements. That is, N = 8. Correspondingly, the Butler matrix should have eight output ports and

eight (switchable) input ports. This conforms to a standard (N × N) Butler matrix, with $N = 2^j$, where j is an integer.

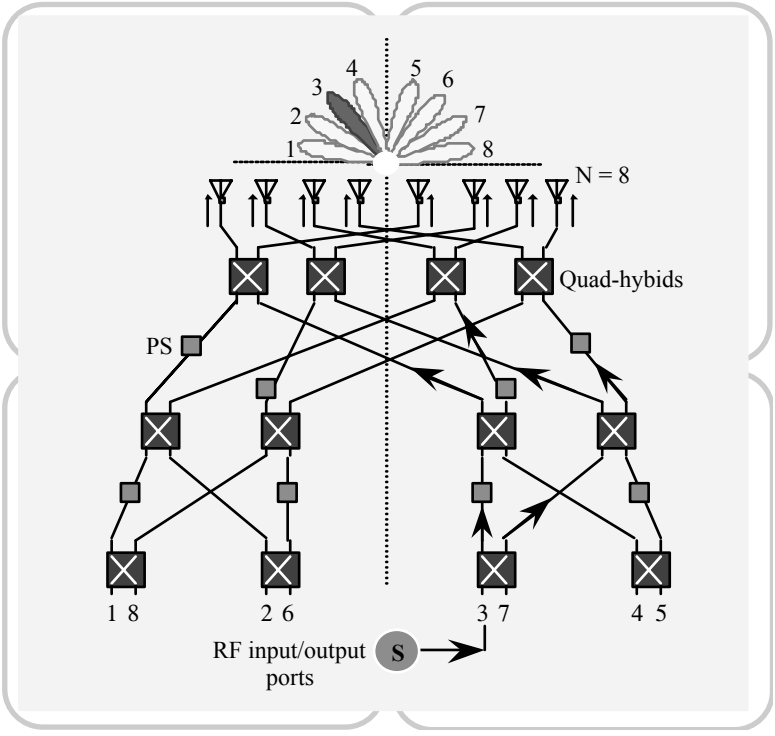


Fig. 6.28 Topology of a Butler matrix to obtain eight switchable beams

The number (n) of quad-hybrids or directional couplers required depends on the number of beams (N) to be produced. For a linear array,

$$n = (N/2)\log_2(N) = Nj/2 \tag{6.9}$$

Hence, with N = 8, n = 4 × 3 = 12. The number (m) of phase-shifters needed is given by:

$$m = (N/2)[\log_2(N) - 1] \tag{6.10}$$

Therefore, with N = 8, m = 4 × 2 = 8. Though the values of n and m could be significant with large values of N, the number of such devices needed for wireless/cellular applications is rather modest.

Beamforming

For an antenna element spacing of $\lambda/2$, the equally-spaced beam positions (with N = 8) are given by:

$$\theta_1 = \arcsin[(2k - 1)/N] \qquad (6.11)$$

with k = 0, 1, 2, …, 7 and θ_1 is the angle off boresight. Explicitly, these beam-pointing angles about the boresight are as follows: $\theta_1 = \pm 7.18°, \pm 22.02°, \pm 48.59°$ and $\pm 61.05°$.

This Butler array is uniformly illuminated leading to a pattern shape, sin(x)/x. Such patterns are spaced to be orthogonal, and cross-over is – 3.9 dB below the first sidelobe level of – 13.2 dB. Further, with quadrature hybrids, this scheme will be devoid of a boresight beam. (If a boresight beam is desired, the quadrature hybrids can be replaced by – 180° hybrids.)

Matrix labyrinth

The compatible Butler matrix for the design under consideration is illustrated in Figure 6.28.

The antenna gain of the pattern of a linear array is given (to the first approximation) by,

$$G = 40000\eta/\phi\theta \qquad (6.12)$$

where η is the antenna efficiency and (θ, ϕ), in degrees denote the beamwidths in azimuth and elevation respectively. The elevational beam for the array of the type illustrated in Figure 6.28 is about 85°, which can, however, be reduced by gauging the antenna elements.

In the configuration of the Butler array presented in Figure 6.28, for a signal injected at port 3, the signal bifurcation along labyrinth is traced. It eventually yields equal outputs at the eight antenna elements (confirming uniform illumination), so as to produce a sin(x)/(x) beam at position 3.

Merits and demerits of Butler array based arrays

In view of the discussions presented above, the Butler array trade-offs can be summarised as follows:

Advantages: The array construction with Butler matrices leads to a gain of the whole aperture. It offers isolation between adjacent beams. Further, a narrow horizontal beam capability can be achieved. The array can be shaped into a cylindrical form factor and this form factor could be conformal to the suface.

Disadvantages: The conceivable limitations of this system include the following: Cross-overs in feed network when more than 4 ports are required. There

are also inevitable feed network losses. Lastly, the wind-loading on the structure could be substantial.

Smart antenna implementation

Suppose a 120°-coverage is sought in switched-beam mode operation. This can be accomplished with four beams per sector, each beam with a beamwidth of about 30°. The relevant array is made of four elements and driven by a beamforming network, such as a Butler matrix. The scheme of implementation is illustrated in Figure 6.29 along with the scalloping (shell-like) geometry of the virtual beams, which form – 39 dB intersection divots for beam orthogonality.

Operationwise, a set of directional couplers samples the incoming signal from the mobile units. The sampling is done by means of a signal sampler. Relevant data gathered activate a programmable switch to "switch" the beam to the location of the incoming signal from a desired mobile subscriber and this signal is forwarded to the receiver. The system shown in Figure 6.29 also includes its operation through the use of duplexers, which segregate the transreceive signals.

In the implementation of switched-beam smart antennas, the signal-processing involved can be done either in the analog or digital domain. In the digital scheme, the RF signal is down-converted to the intermediate frequency and the baseband signal is recovered. Analog-to-digital conversion of this baseband signal yields digital information for (digital) beamforming *via* the software. The analog and digital implementation allows realising a broad range of wireless technologies, including FDMA, TDMA, and CDMA systems to be fullfilled.

The notable benefits of smart antennas stem from their potential to provide enhanced range and reduced underlying costs during the deployment stage and increased long-term capacity as well as robust link performance once the system is built. The enhanced range performance (in terms of coverage and extension) arises as a result of the associated design aimed at EM energy penetration through obstacles (like buildings) and filling of holes. Further smart antenna systems can form a distinct beam for each subscriber on the uplink and downlink, minimising the impact of noise and interference for each subscriber and the base station.

The range extension enables a mobile to operate further from the base station without increasing the uplink power transmitted by the mobile unit or the downlink station transmitter. Using an M-element array, one can achieve an SNR improvement, for example, equal to $10\log_{10}(M)$ dB in an AWGN (additive white Gaussian noise) environment (with no interference or multipath prevailing). This SNR improvement is relative to the situation of using a single element. In a typical example, a smart antenna using an 8-element array (in a CDMA system at an operational frequency of 1920 MHz), could provide a 60 % range improvement with no increase in transmitter power. Alternatively, it can offer pre-channel noise performance at the receiver an 9-dB excess link margin as compared to conventional sectoring.

6.7 SPACE DIVISION MULTIPLE ACCESS (SDMA) TECHNIQUE

The SDMA is specified for smart antennas used in multipath environments. It refers to the wireless ambient where the received multipath components of each

transmitted signal arriving from all directions to the base station antenna are presumably uncorrelated; a suppression of interfering signals is carried out to extract the desired (transmitted) signal.

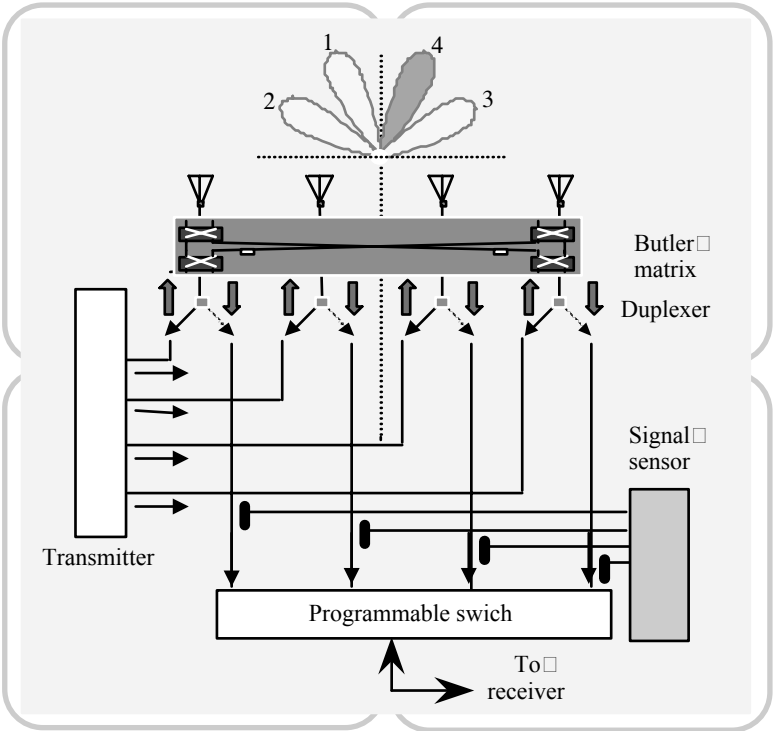


Fig. 6.29 Switched-beam smart antenna implementation

A smart antenna system can be conceived to match the SDMA considerations by means of a number of highly isolated antenna elements for the transmission and reception of multiple signals with independent fading characteristic [6.5].

6.8 CONCLUDING REMARKS

Emphasis has been placed in recent times to augment the performance of wireless communication systems implemented to match the growing number of cellular telephone users. In such efforts the smart antennas play a crucial role. They depict radiating systems conceived as adaptive arrays. Classically, the underlying

concepts were explored in depth in radar applications. The pertinent knowledge gained thereof has been appropriately adopted in the development of systems governing modern wireless communication technology. The unique fading environment and characteristics of the wireless communications have been duly taken into account in the relevant designs.

Decision to deploy a smart antenna system relies on the following design perspectives: (i) a justification on the system cost to achieve increased capacity; (ii) mounting requirements and zoning issues dictating the size and weight of the overall structure: (iii) wind-loading and zoning issues specifying the form factor required and (iv) other system performance specifications such as, horizontal- and vertical pattern-shapes, antenna gain, polarisation, interference and fading, isolation considerations, bandwidth requirements (consistent with the frequency band of operation) and port requirements.

Simply, going for a smart system would render an antenna appear to be "smart", functionally intelligent and operationally effective.

With the growing trend in wireless communication systems (of both outdoor and indoor applications), the smart antenna would find a substantial niche in the technology. The present chapter provides a prelude into existing strategies and paves the path for some futuristic insights.

REFERENCES

[6.1] Liberti, J. C., and T. S. Rappaport, *Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications*, Upper Saddle River, NJ: Prentice Hall PTR, 1999

[6.2] Pattan, B., *Robust Modulation Methods and Smart Antennas in Wireless Communications,* Upper Saddle River, NJ: Prentice Hall PTR, 2000

[6.3] Goldsmith, A., Smart antennas*, IEEE Trans. Personal Commun*., February 1998, vol. 5(1), 9

[6.4] Ertel, R. B., P. Cardieri, K. W. Sowerby, T. S. Rappaport and J. H. Reed, Overview of spatial channel models for antenna array communication systems*, IEEE Trans. Personal Commun*., February 1998, vol. 5(1), 10-22

[6.5] Winters, J. H., Smart antennas for wireless systems*, IEEE Trans. Personal Commun*., February 1998, vol. 5(1), 23-27

[6.6] Choi, S., D. Shim and T. K. Sarkar, A comparison of tracking-beam arrays and switching-beam arrays operating in a CDMA mobile communication channel*, IEEE Antenna Propagat. Magazine*, December 1999, vol. 41(6), 10-22

[6.7] Chryssomallis, M., Smart antennas*, IEEE Antenna Propagat. Magazine*, June 2000, vol. 42(3), 129-136

[6.8] Vaughan, R. G., On optimum combining at the mobile*, IEEE Trans. Veh. Technol*., November 1988, vol. VT-37(4), 181-188

[6.9] Dietrich, C. B., W. L. Stutzman, B. K. Kim and K Dietze, Smart antennas in wireless communications, *IEEE Antenna Propagat. Magazine*, October 2000, vol. 42(5), 142-151

[6.10] Boukalov, A. O., and S. G. Haggman, System aspects of smart antenna technology in cellular wireless communications – An overview, *IEEE Trans. Microwave Theory Tech.*, June 2000, vol. 48(6), 919-929

[6.11] Kavak, A., Torlak, M., Vogel, W. J. and G. Xu, Vector channels for smart antennas - Measurements, statistical models and directional properties in outdoor environments*, IEEE Trans. Microwave Theory Tech.*, June 2000 vol. 48(6), 930-937

[6.12] Rappaport, T. S., *Wireless Communications: Principles and Practice,* Upper Saddle River, NJ: Prentice Hall PTR

[6.13] Neelakanta, P. S., D. De Groff and S. Sudhakar, Low-altitude target detection by coastline operated marine radar, *IEEE Trans. Aerospace Electronic Syst.,* January 1992, vol. 28(1), 217-222

[6.14] Saleh, A. A. M., and R. A. Valenzuela, A statistical model for indoor multipath propagation, *IEEE J. Select. Areas Commun*., February 1987, vol. SAC-5(2), 128-137

CHAPTER 7

Antennas for Indoor Wireless Communications

7.1 INTRODUCTION

Another domain of interest in modern wireless communications refers to indoor radio systems for voice or data transmissions within a building such as an apartment, an office, a warehouse, a factory, a hospital, a convention centre, a closed auditorium etc. Such a *personal communication system* (PCS) is attractive since it offers unlimited user mobility, "free from cords". Typically *wireless LAN* (WLAN), wireless data communication and systems like *Bluetooth*™ have been conceived with their scope of development in indoor applications.

The design and implementation of antennas in such systems call for unique considerations in view of the fact that indoor channels differ from the traditional mobile radio channels in two aspects namely, the *interference pattern* and the *fading rate*. For example, an interference stemming from a noncommunication electronic system operating inside the building and affecting the indoor wireless communication channels could be highly sporadic and nondeterministic. Furthermore, the dynamic range of fading experienced by a mobile unit (such as a handheld unit) inside a building is often smaller than that experienced by a moving vehicle in an urban setting. There are also a variety of secondary effects in the vicinity of an indoor operating unit, which could significantly influence the radio channels. For example, motion of people, doors being opened and closed etc. may affect the propagation characteristics of wireless channels as slowly varying functions of time. Further, the local features of indoor building architecture, construction materials used, furnishings done as well as the coexistence of other communication systems (like Bluetooth and WLAN) would strongly influence indoor wireless propagations. Also, the polarisation of EM waves and frequency of operation may distinctly specify certain unique indoor radio channel characteristics. In order to identify the specific type of antenna for an indoor wireless service, knowledge of spatial and temporal statistics of signal attenuation, multipath delay-spread and impulse response of the indoor radio channel involved is essential. Pertinent modelling carried out as reported in the literature [7.1-7.2] is based on information gathered from RF signal propagation measurements within buildings of various types - large, medium and small. Such measurements done at large, refer to different civil engineering premises (specified by structural and material aspects) with the inclusion of typical indoor furniture and appliances. Further, these measurements address the range of an approved EM spectrum intended for indoor radio applications with the associated modulations and polarisation. The propagation delay measurements envisaged correspond to

assessing the transit time of a radar-like impulse transmitted and its echo (reflection) received. In the following section a review on the indoor propagation characteristics of EM waves is presented consistent with the information on measured data collected and reported in the literature.

7.2 INDOOR AMBIENT *VERSUS* EM WAVE PROPAGATION

As indicated above, the indoor EM wave propagation models are largely empirical based on measured data, logistical heuristics and civil engineering aspects of building interiors. Saleh and Valenzuela, [7.3], for example, have reported results on indoor measurements on propagation between two vertically polarised omnidirectional antennas located on the same floor of a medium-sized office building. Measurements made refer to using 10 ns, 1.5-GHz, radar-like pulses. The signal data gathered correspond to averaging the square-law detected pulse response at the receiver obtained by sweeping the frequency of the transmitted pulse. Using this technique, the multipath components of the received EM wave have been assessed within a resolution of 5 ns.

The relevant results indicate that the indoor channel is *quasistatic*, in the sense that, the statistics of the channel's impulse response vary slowly with time and are independent of transmitting and receiving antenna polarisation, whenever there is no line-of-sight (LoS) path between them. It is further reported that a maximum multipath delay-spread of the order of 100-200 ns exists within rooms and about 300 ns of delay-spread in lengthy hallways. Further, the measured rms delay-spread within such rooms observed has a median of 25 ns and a maximum of 50 ns. The signal attenuation (with no line-of-sight between the transreceive path) is found to vary over a range of 60-dB and obeys a long distance power-law with an exponent (η) varying between 3 and 4. The corresponding *path loss* (PL) is given by:

$$\log_{10}[PL(d)] = k\,\eta \log_{10}(d/d_O) \qquad (7.1)$$

with the constant of proportionality k that can be chosen appropriately with respect to a standard reference distance, d_O.

The measurement-based multipath indoor-channel model described in [7.3] assumes that the rays arrive in clusters and prescribes a statistical description of the amplitude and phase conditions of such clustered rays received. Essentially, the amplitudes of received ray are regarded as independent Rayleigh random variables (with variances that decay exponentially with cluster-delay as well as with the ray-delay within a cluster). In addition, the corresponding phase angles are assumed to be independent uniform random variables over $[0, 2\pi]$. The clusters (and the rays within a cluster) form a Poisson arrival process with different rates; further, they have exponentially distributed interarrival times. The formation of clusters itself is related to the building structure, whereas the rays within the cluster are formed by multiple reflections from objects in the vicinity of the transmitter and receiver.

Another indoor radio channel model due to Rappaport [7.4] is based on measurements carried out at 1300 MHz in five factory buildings (and an extended measurement done subsequently in other types of buildings as well). Relevant

multipath delays inferred range form 40 to 800 ns; the mean multipath delay and rms delay-spread values correspond to a range of 30 to 300 ns with median values of 96 ns and 105 ns respectively for LoS and obstructed paths. Further, assessed delay-spreads have been found to be uncorrelated to transreceiver separation but are functions of factory inventory, building construction materials, age of the building, locations of the walls and ceiling heights. Typically, for example, the measurements done in a food-processing factory (that manufactures dry-goods and devoid of extensive metal objects dispersed in the ambient) indicate the rms delay-spread of values being only half of those observed in factories producing metal products.

Further, it is surmised that newer buildings, which incorporate steel beams and steel reinforced concrete in the building structure, can pose stronger multipath signals than older structures, which use wood and brick for perimeter walls. The data collected by Rappaport [7.4] lead to a statistical model for the indoor radio channels and accounts for both specular reflections from walls and ceilings as well as random scattering from inventory contents and equipment. Using the considerations reported in [7.4] a statistical model is empirically derived; and a computer code called *Simulation of Indoor Radio Channel Impulse-response Model* (SIRCIM) that recreates realistic samples of indoor-channel measurements has been developed. Later versions also include coverage between floors specified by building blueprints [7.5].

In gathering inferential heuristics from field measurements, there are specific physical issues to be recognised on distinct EM transmissions and reflections (from walls, floors and ceilings) pertinent to buildings of different types, as discussed below.

Considering residential houses, there are fewer occupants as compared to an office building or a commercial area. The latter includes manufacturing floors, shopping malls, storage places, and transportation stations. Often, people inside their homes are either sitting down or standing up. Furthermore, since it is not a standard practice to buy new furniture or move furniture around the house frequently, home furniture and appliances should be viewed as fixed reflectors of electromagnetic (EM) waves depicted by a few constant factors imbedded in the statistical models.

In residential indoor settings, PCs, laptop computers, printers, and other computer devices and equipment are not typically used while in motion. Subsequently, the only potential mobile receivers or transmitters in residential areas are typically cordless and wireless telephones and personal digital assistants (PDAs). Further, within the home, it is unlikely that a person would continuously walk during an entire call. If they do so, their movements are most likely to be characterised by repeated patterns of short tautological paces. Doppler spreads associated with the use of wireless units caused by holding these mobile units or reflecting their signals, are smaller in residential homes when compared to other indoor-propagation environments.

In fact, Doppler spreads in homes will hardly reach the maximum frequency recommended by the PCS Joint Technical Committee for indoor pedestrian-communication environments. This frequency is about 9.6 Hz for those services operating at 2 GHz. By examining the nature of human occupancy and movements,

it is safe to assume negligible Doppler spreads in a typical residential home over long periods of daytime, and zero Doppler spread during the night.

The residential homes represent the smallest sites for indoor communications. For short-range connectivity and by the virtue that they are contained in the same home, there is modest path loss due to the distance travelled in free-space at home compared to outdoor propagation. If the loss in the first metre is 30 dB, and a free-space loss of 6 dB per octave is factored in, then a house that is 24 m long will attenuate the signal by 54 dB, as the signal propagates from one end of the home to the other. The signal-power attenuation may also occur as a result of losses associated with signals travelling through exterior and interior building structures. Residential homes are typically wooden framed with low ceilings designed as single-family units with one or two stories. The interior walls are usually covered with a thin layer of plaster inside cardboard*. The exterior frame is often filled with insulation and covered by layers of plywood and wooden or aluminium siding or brick. In contrast to indoor-office areas, houses are not usually built using metallic studs and concrete frames and their floors often do not contain large amount of metal and concrete. There are almost none, or a very few, large-sized metal objects or cubicles, for which there is no application or use at home. In [7.6], typical partition losses measured in an office building are reported. Further, comprehensive tables pertinent to measured (average) signal-loss on indoor radio channels gathered by various researchers for radio paths obstructed by common building material are given in [7.2].

In reference to residential (home) indoor communications, the associated losses could possibly be up to a maximum of 7 dB per interior partition. Signals penetrating exterior walls would be attenuated to such an extent to be considered as a distressing sources of interference to nearby homes. Conversely, it is maintained that, unlike commercial areas, residential homes have several glass windows that may remain open most of the summer and spring seasons. Signals escaping from one house into a neighbouring house situated in very close proximity (through open windows) could be a worrisome source of interference unless the protocols (such as the frequency-hopped spread-spectrum (FHSS) communication) adopted are different in the two houses. More typical and aggressive electromagnetic interference (EMI) in short-range communications in residential areas can stem from microwave ovens, garage door openers etc. as will be detailed later in this chapter.

Statistical modelling and computer simulation of indoor radio channels is presented comprehensively in [7.7-7.9].

Denoting the EM path loss (PL) associated with the indoor wireless transmissions in a home environment occurring within the same floor as L_p, it can be written as:

---

*_Note_: The description of the home interior architecture described here applies to the households in North America and Europe. The civil engineering is very much different in subtropical/tropical countries (such as in Asia/Africa) where brick-walled, cement-plastered houses supported by reinforced concrete pillars and beams are more common.

$$L_p = L_o + 10\eta \log_{10}(d) + \sum_{m=1}^{N} P_m \qquad (7.2)$$

where the exponent η is between 2 and 3 for most of the underlying short-range applications, L_o is the drop in power in decibels at 1 metre, and d is the overall distance between two communicating devices. The summation term in equation (7.2) represents the drop in power due to path obstruction by N home interior walls, and each of the terms under summation can be considered equal, inasmuch as the interior walls do not vary much in texture and thickness.

Figure 7.1 shows an example of short-range connectivity setting that includes a laptop, PC, and a printer, each in a separate room. With the room doors open, the PC has a LoS to the printer indicated as (a). In this case, only the first term of the path-loss in equation (7.2) applies. Conversely, the laptop and the printer can only communicate either through an obstructed LoS or *via* a set of multipaths. The transmitted signal power drops by 6 dB per octave, and considering a typical room and hallway sites in residential homes, the transmitted signal along path (b) obstructed by a plywood wall will be received at the printer with higher power than the signal along multipath cluster (c). The latter loses power due to the longer distance travelled and also due to the attenuation incurred from the reflections against the interior walls.

It should be mentioned that there are several other possible multipaths in a specific home structure (such as that depicted in Figure 7.1), which could link the printer to the two computers through reflections from walls, floors, and ceilings. More so, many of those multipaths, after a certain number of reflections and transmissions through walls, may become sufficiently attenuated and their effects on signal fading are therefore negligible. It is also important to point out that Figure 7.1 shows only an example depicting a cluster of specular multipaths. This is because, at approximately 2-GHz centre frequency, most surfaces of reflection in homes are relatively smooth (as specified by Rayleigh criterion). Therefore, the scattering media that produce diffused reflections are minimal. As such, presence of diffused multipaths for short-range connectivity applications in homes can be ignored.

Corresponding to the bit duration adopted in wireless data transmissions, the spatial length covered in that duration is approximately 30 m long. Therefore, frequency-selective fading responsible for intersymbol interference would occur only when the path differences inside a residential home are significant portions of 30 m. For small buildings the maximum delays could be in the order of 100 ns, (which may even be smaller when specifically considering residential homes). Not counting the loss due to reflections, the spatial length of 30 m (indicated above) may cause an approximate 30-dB drop in power relative to a signal that is received one metre away from the transmitter. This drop is significant and may render any resolvable multipath to be weaker than the first signal arrival. As a result, the so-called *flat-fading channel* primarily governs signal propagation inside residential homes, where all effective multipaths arrive within the information symbol.

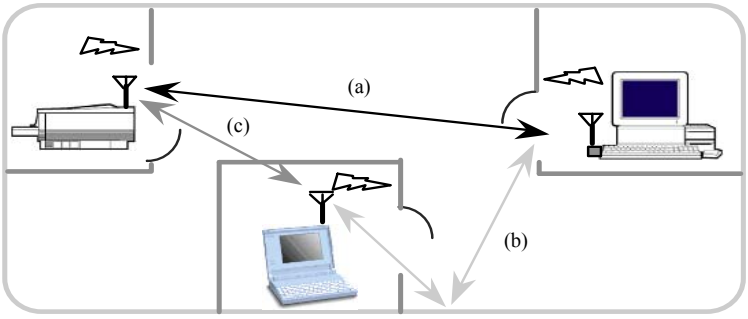


Fig. 7.1 Multipath conditions in typical home-based indoor RF transmissions: (a) LoS path; (b) scattered/reflected path and (c) an obstructed path

An indoor wireless channel can be described by Rician fading, where the probability density function (pdf) of received signal envelope x is given by:

$$f_X(x) = (x/\sigma^2)\exp[-(x^2 + A^2)/2\sigma^2]I_o(Ax/\sigma^2) \qquad (7.3)$$

where $I_O(.)$ is the modified Bessel function of the first kind and zeroth order; and the parameter $0 \leq A \leq 1$ denotes the peak value of specular radio signals. It represents the superposition of the dominant LoS and time-invariant scattered signals reflected from the walls, ceilings, and other stationary objects. The parameter σ^2 represents the average power of the signal received over paths that vary with time due to people wandering in the house. When $K = (A^2/2\sigma^2)$ becomes infinity, the channel is Gaussian; whereas $K = 0$ defines the so-called Rayleigh channel. A value of 3 dB for K is typical in modelling indoor radio-channel amplitude fluctuations.

As indicated before, due to the sparse nature of human occupancy in residential homes and considering small home sizes, the associated Doppler shifts in the EM waves and the corresponding delay-spread could be minimal. This renders the wireless indoor-communication channels to be either almost stationary or (very) slowly varying Rician flat-fading channels.

Further, when the transmitter and receiver are not in mobile states and in the absence of pedestrian movements, the communication channels (such as in Figure 7.1) are characterised by a constant, time-invariant impulse response with zero Doppler shifts. In this case, the channel is deterministic and its multipath characteristics remain constant over a long period of time.

Now, considering the transmitter-receiver antenna dispositions in an indoor environment, and depending on the locations of such antennas, the multipath components may add constructively or destructively at the receiving end. As a result, the received signals may be modified by a constant factor or remain in a state of deep fade. The locales, which face destructive interference, correspond to

"blind-spots" and any receiver placed at these locations would experience strong fading conditions.

Therefore, (in applications like Bluetooth) it is necessary to ascertain such blind spots as well as to know the optimum locations to place the devices so that the associated antennas will not be positioned at locales of blind-spots. Further, the directivity of these antennas can be specified with reference to optimum transreceive locations.

Hence, the home-based EM propagation environment can be classified in the perspectives of three scenarios. In the first case, suppose there is a laptop computer in each room of a house that communicates with a fixed printer located in a specific room. These locations should be specified in modelling the most likely propagation scenarios. That is, they can be depicted as stationary dispositions devoid of Doppler shifts pertinent to a setting wherein most of the house occupants are either away or assumed to be frozen at "stand-still". In this context, with no randomness induced in the communication channel, the factor K can be approximated by a positive infinity guaranteeing a specular LoS connectivity.

In the second type of propagation environment, the transmitter and receiver are stationary, but it is assumed that people within the house are moving. This may cause the communication channel to be slowly time- varying. In such cases, the multipath components can be divided into two different categories. The first category is comprised of the dominant LoS signal (if it exists), in addition to all time-invariant scattered signals reflected from the stationary objects such as walls, ceilings, furniture, and appliances. The other category consists of multipaths whose first- or higher-order reflection patterns include at least one path higher-order that bounces from a nonstationary source in the indoor site. The lifetime of multipath signals belonging to the first category is very long and may last until the signal transmission is terminated. On the contrary, the multipath signals in the second category is evanescent and may cease to exist shortly after they became established. The contribution of the first category of multipaths to a signal-fading environment is deterministic with zero Doppler shift; whereas, the effect of the second multipath category is stochastic and changes the signal correlation and frequency characteristics. The combined categories yield a Rician fading described by equation (7.3) with σ assuming nonzero values.

The third type of propagation environment refers to a scenario where either the receiver and/or transmitter could be in motion. This situation indeed would introduce Doppler effects on the transmitted signal with frequency changing up to approximately 10 Hz. The movement, as well as displacement of the antennas inside the house, may then create a new LoS or obstruct a LoS that is already existing (as illustrated in Figure 7.2). Further, the EM scattering profiles near the transmitters and receivers may change, consequently giving rise to random scattering with the statistics of a Rician pdf.

The correlation properties of a fading envelope can be determined in terms of relative powers contributed by different scatterers (and arrive as multipath components) and their corresponding angles of arrival. In isotropic scattering, the power is uniformly distributed over 360^{o}. This leads to a U-shaped spectrum shown in Figure 7.3, which is known as the *classic Doppler spectrum*. While the classic spectrum is often assumed in outdoor propagation, a flat Doppler spectrum is

shown to be characteristic of indoor communications. In the flat Doppler spectrum, the uniformity is encountered across frequencies rather than over the angles of arrivals. This type of spectrum results from random movements of scattering elements in the area of communication path. A flat Doppler spectrum is also shown in Figure 7.3 (for comparison), where the impulse represents the presence of the LoS signal path that can be associated with the classic and flat Doppler spectra. The impulse location at zero frequency indicates, among other possibilities, the orthogonality between the direction of motion of the receiver and the line connecting the receiver and transmitting antennas.

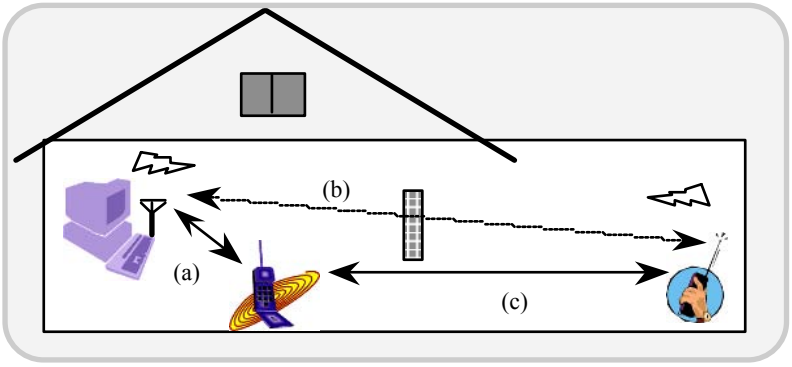


Fig. 7.2 Multipath conditions resulting from the mobility of the transreceive units inside a house. (a) A LoS path; (b) a path that is prone to obstructions and (c) a path that may or may not be of LoS type and it is dependent on the mobility of the users on both ends

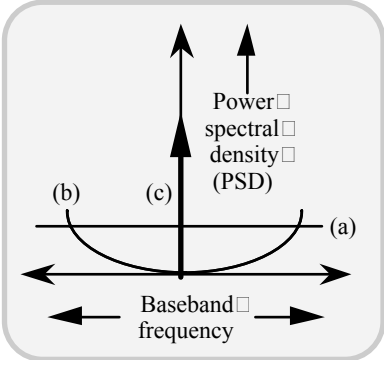


Fig. 7.3 U-shaped and flat-Doppler spectra. (a) Flat-Doppler spectrum; (b) U-shaped spectrum and (c) impulse depicting the fixed (LoS-specific) component

Considering the standards (like that of Bluetooth) assigned for short-range voice and data transfers through low-power RF transmission and reception, the associated EM propagation may involve reflections and transmissions through solid, non-metal objects being present in the premises. This happens in a nominal link range that may stretch from 10 cm to 10 m (with possible extension up to 100 m by increasing the transmitted power).

Suppose an access point and a mobile host (MH) are on different floors. Then, the direct ray between them has to pass through the intervening floors. With reference to office buildings, they are invariably constructed with concrete floors and drop ceilings of acoustical material supported by a metal frame. The acoustical material has a low dielectric constant and is readily penetrated by an incident EM wave. However, the space between the drop ceiling and the floor above may contain an irregular collection of supporting beams, ventilation ducts, lighting fixtures, pipes, etc., all of which would scatter and attenuate the EM wave. Although it is difficult to characterise the scattering, an effective floor loss can be inferred from measurements made with antennas situated on different floors of a building. For reinforced concrete or precast concrete floors, transmission loss has been measured as 10 dB or more. Floors constructed of concrete poured over corrugated steel panels show much greater loss. In this case, the received signal may, in fact, be associated with other paths involving diffraction outside the building or in stairwells and elevator shafts, rather than transmission through the floor itself. The direct ray and two diffraction paths are suggested in Figure 7.4, assuming a simple geometry. Here a building is shown in cross-section with the transmitter located in one floor and the receiver on a lower floor.

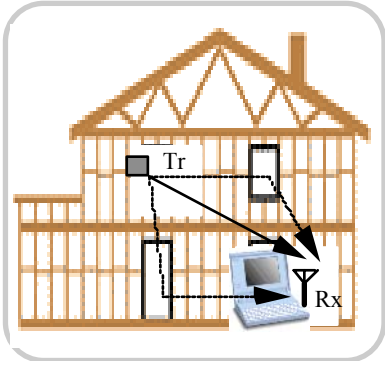


Fig. 7.4 Cross-section of a building showing the direct path and two other paths involving double diffraction between a subscriber and an access point located on different floors

The direct path from the transmitter to the receiver may pass through walls and the intervening floors. As mentioned above, floors of the building in question can be assumed to be constructed of precast concrete panels (with the underside of the

floor panels being the ceiling of the floor below), and without suspended ducts, light fixtures etc. In such situations, the floor loss could be somewhat low, and the direct path signal can be ascertained by the free-space path gain reduced by the transmission loss through the walls and floors. In addition, rays can exit the building through large windows lining the hallway, diffract down along the building face, and diffract again back into the lower floor through the windows and eventually illuminate the receiver.

Considering the path of diffraction out of and back in through the windows, the path gain for these rays can be specified solely due to diffraction (with no reflections) plus any reductions specified by the transmission loss through the windows and the walls. A typical signal level computed for the direct path (assuming a 7-dB transmission loss through the floor) and the sum of the signal computed for the two diffraction paths, using relevant absorbing screen diffraction coefficient, is shown in Figure 7.5. The initial, rapid decrease of signal in Figure 7.5 arises from the transmission loss through the floor, while the slow decrease seen later, results from diffracted signals, which suffer an extensive loss in case of a single floor separation but vary only to a limited extent with increased floor separations.

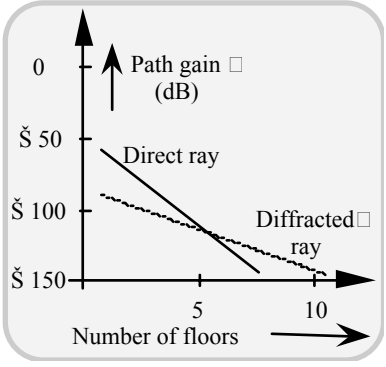


Fig. 7.5 Typical plots depicting the path gain (in dB) *versus* number of floors in reference to direct and diffracted rays inside a building

In addition to the above considerations, scattered furniture and/or other absorbing materials being present in the path of an EM wave propagation could influence the path loss/gain characteristics. Typically, between the floor and ceiling the propagation is governed by Fresnel zone considerations, as illustrated in Figure. 7.6.

A detailed study pertinent to the hostile indoor environment, path loss, and delay-spread using the ray-tracing method is also available in the literature. For example, Obayashi and Zander [7.10] have elaborated a *body-shadowing model* for indoor wireless communications using ray-tracing and imaging methods.

In short, the propagation of radio waves inside a building is a complex process. It depends significantly on the specific architecture and unique contents of

the indoor environment (such as in a house or an office, or in a factory), as well as on the topography involved (such as LoS, or obstructed transmissions). Further, as mentioned before, the statistics of indoor radio channel may vary with time due to motion of people and equipment.

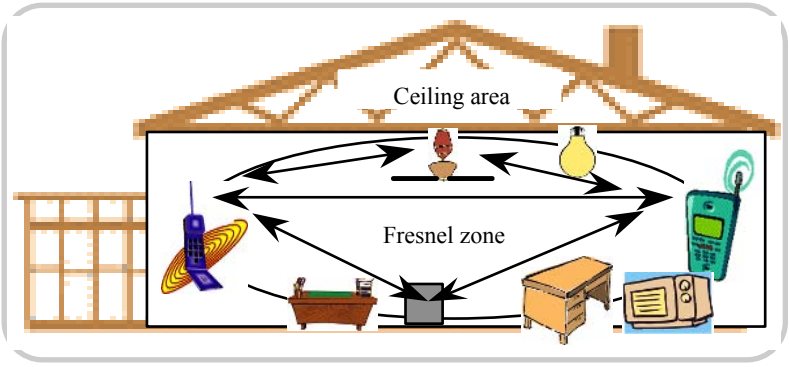


Fig. 7.6 Fresnel zone for propagation between a subscriber and an access point located on the same floor of a building illustrating the propagation through the unobstructed space between furniture and ceiling fixtures

In reference to a factory environment, for example, the radio channels can be characterised by high path losses and large variation in losses. The reported values of the path loss exponent (η) in such cases are 1.5 to 2.0 for LoS and 2.5 to 3.0 for obstructed channels. Typically, the statistics of overall path loss is log-normal distributed with an extent of about 7 dB standard deviation in a factory ambient; whereas, in an office environment, the signal envelope is largely Rician-distributed (with K equal to about 2-3 dB), as indicated earlier.

Whenever the motion within a building causes intense signal fading, it compares well with the Rician distribution with K approximately equal to 7 dB and 10 dB for office and factory environments respectively. If the degree of motion is small (as in a home/small office environment), the fading is very slow with the Doppler bandwidth being about 4 to 5 Hz. In such cases, the channel becomes quasistatic. In general, the indoor fading statistics can be regarded as statistically nonstationary, or, at least they possess wide-sense stationary characteristics exhibited for periods up to 2 s. Since many digital symbols transmitted correspond to this period at practical data rates, the type of fading involved is, therefore, *quasi-wide-sense stationary*.

7.3 INDOOR ANTENNAS: UNDERLYING CONCEPTS

Indoor antennas are developed to meet the demanding style and performance needs of today's wireless handset market. Specific requirements include a rugged, but of compact/small-size design, increased sensitivity, and impedance matching without additional circuitry. The antenna design of choice for indoor handset components is directed to maximise ground-plane efficiency and obtain adequate sensitivity in the horizontal and vertical polarisations. A desirable feature of such an antenna is

that it should offer favourable gain in all planes. Commensurate with these considerations, the generic structures of indoor antennas that can be conceived on the basis of various elements and array geometry discussed in the earlier chapters can be identified in the following sections.

The compatible antenna types for indoor wireless applications are dipole, flat-panel, and microstrip structures. Other feasible versions in the ISM band are: multiple element dipoles, Yagi-Uda arrays, parabolic dishes, and slotted antennas. However, the more complex antennas are less likely to have uses in indoor systems either because of cost and form-factor, or due to their radiation patterns being strongly directional, which would tend to be less suitable in certain applications. An outline of basic structures of antennas compatible for indoor applications is narrated below and more specific types will be described in a later section.

Meander-line antenna

This represents a miniaturised-printed antenna that might fit along the edge of a laptop computer. Its small size is accomplished by embedding the wire structure on a dielectric substrate.

Key benefits of this type of antenna are its simple configuration, a potentially *low specific-absorption rate* (SAR), as well as being inexpensive.

Low profile planar antenna

This refers to a microstrip antenna, with a low-profile structure that is easily adaptable for planar integration. It has a wide bandwidth with a hemispherical radiation pattern and poses low VSWR characteristics. Further, it can be made compact and lightweight with features of a simple configuration.

Microstrip antennas, as discussed in Chapter 4, are simply copper traces/patterns on substrates or PCBs with a ground-plane backing. Making such traces devoid of any unwanted radiating components, they are turned into useful antenna structures.

The crucial aspect of small antenna designs, such as those needed in indoor wireless systems, refers to trimming the radiation just for the coverage area and not spilling the EM energy in those sectors which are not of interest. Otherwise, there is a wastage of radiated power, as well as, in those sectors where radiation is unnecessarily spilt over, there would be interference.

Combination of externally placed and internally located dual antennas

An externally mounted structure plus an internally built-in antenna combination can serve as an indoor antenna as follows: The internal antenna is used as the primary antenna and the external (such as a retractable whip) antenna is used in conjunction, whenever higher performance is needed.

The external retractable antenna might, for example, extend out of the lid of a laptop computer. Alternatively, an external side-mounted antenna could be attached to the base of the computer. The internal antenna can be deactivated when the whip antenna is extended. Major features of this type of antenna could be: low cost of manufacturing, feasibility of snap-on installation, reduced SAR, and improved performance for low cell-site density applications.

Half-wave coaxial dipole antenna

This structure carries a simple, ground-independent design with minimal fixture and user effects. Also, it can offer favourable gain (≈ 1 dBi), low VSWR (< 2.0) and may prove to be cost-effective.

Slender dipole antennas are cylinders, with the signal usually fed from the bottom. Simple dipole structures are often made out of short sections of standard coaxial cable for easy matching implementation.

The elevation patterns of a dipole antenna transmits best from the side of the antenna. As indicated in Chapter 4, the length of a dipole antenna is related to the wavelength of the signal supported. Half- and quarter-wave dipoles are commonly used. Dipole antennas are available in various form-factors. The versions of λ/2-dipole radiators available on the market correspond to snap-in surface-mounted packages and swivel antennas built with a SMA connector.

Quarter-wave stubby antenna

Quarter-wave stubby antennas are ground-plane sensitive structures with omnidirectional pattern. They provide acceptable gain (0 dBi), and are also cost-effective.

Flat-panel antenna

As discussed in Chapter 4, flat-panel antennas refer to small metal patches, which are usually square or rectangular. They are strongly directional, with their radiation pattern therefore being not so ideal for handheld devices. On the other hand, they can be made very small for the ISM band and can be mounted directly onto PCBs, both of which help to reduce costs.

A popular form of flat panel antenna is the planar inverted F-antenna (PIFA), which is named for its resemblance to the capital letter F on its side as mentioned in Chapter 4. It has a flat panel as far away from the ground as possible and is fed by two contacts, which form the arms of the F geometry.

A PIFA antenna can be fabricated as a microstrip antenna using the trace on a PCB as shown in Figure 7.7.

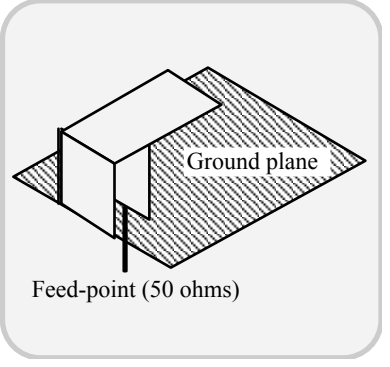


Fig. 7.7 A typical PIFA structure. (The feed-point is a SMA connection tuned for 50-ohm impedance matching)

For short-range (10 m) wireless connectivity (such as Bluetooth applications), the PIFA and patch style antennas are compatible and can be designed as custom-built and/or as off-the-shelf units to meet specific applications and space requirements.

7.4 INDOOR ANTENNA CHARACTERISTICS

The parameters and considerations enumerated below play a dominant role in specifying as well as elucidating the performance aspects of practical indoor antennas. In addition to those indicated, size/weight considerations, the form factor of the antenna and cost effectiveness are other design aspects that can be stipulated as regard to indoor antennas.

Voltage standing wave ratio

A crucial performance consideration of antennas in general, refers to the voltage-standing-wave ratio (VSWR), which as mentioned in earlier chapters, is an indicator of reflected power at the antenna input. More specifically, VSWR is related to the amount of power that reaches the antenna versus the amount that is reflected back into the radio unit. A typical VSWR performance of an indoor-specific antenna is illustrated in Figure 7.8.

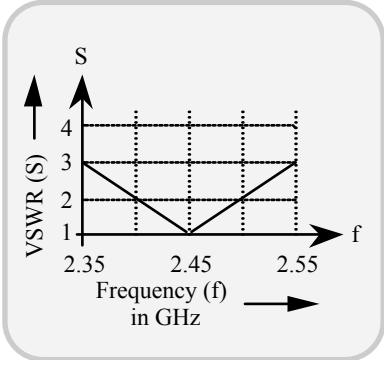


Fig. 7.8 An example of VSWR *versus* frequency of a typical indoor antenna operating in the ISM band

Antenna placement in the radio unit

Another design perspective of indoor antennas is concerned with the disposition of antenna location in a radio unit. In such placement issues, due emphasis should be given to LoS considerations versus the anticipations on how the antenna will be spatially oriented (in the regular usage of the radio unit). Further, the antenna should be located outside of any shielded space of the unit. Also, any adjacent metal surfaces proximal to the antenna should be avoided.

Antenna polarisation

To maximise power reception, the polarisation of transmitter antenna should match the polarisation of the receiver antenna. Most antennas used in portable communication units are linearly polarised. However for indoor applications, circular polarisation is preferable, as will be explained in a later section.

Antenna radiation patterns

The radiation pattern specifies the plot of power radiated from an antenna in a specified plane - the two principal orthogonal planes, namely, the vertical and horizontal (azimuthal) planes.

For stationary end-entities of an indoor radio-link, the most desired result would be a pattern which is omnidirectional in the azimuthal plane (and with a limited coverage in the vertical plane). This corresponds to a typical doughnut-shaped three-dimensional pattern. The associated coverage can link any two such end units regardless of their relative placement in the coverage area. This is akin to almost a broadcast mode of operation of one unit with respect to any other unit to which a wireless connectivity is sought. Further, as the end units are assumed to lie in the same horizontal level, the vertical coverage should be kept minimal in order to conserve the power. In general, the doughnut-shaped pattern with a narrow figure-of-eight cross-section can meet the desired pattern coverage and it can be realised with a simple dipole, as discussed in Chapter 4.

In certain specific circumstances, where a point-to-point (LoS) connectivity between any two stationary end-units is warranted, directional patterns are more appropriate.

When mobility of the units comes into the picture, the patterns of the associated units should be consistent with the reliable transreceive communication sought as well as any other specific constrains. Such patterns could be then asymmetrical or one-sided (so as to avoid, for example, the user exposure and excessive SAR implications with respect to cellular/cordless phones).

In general, the pattern specifications of indoor antennas should be consistent with all the associated parameters such as backlobe and sidelobe levels, pattern nulls, half-power bandwidth, desirable plane(s) of coverage, polarisation and asymmetric/skewed aspects of the pattern, in addition to the desired main-lobe characteristics.

7.5 INDOOR WIRELESS COMMUNICATION SYSTEMS

In order to deploy an antenna in the indoor applications, it is crucial to know the various wireless services adopted as indoor wireless systems. In modern perspective, there are four major wireless communication systems applicable to indoor applications for which the antenna requirements could be analysed. These systems are as follows:

- Cordless wireless telephone
- Two-way radio (deployed in indoor and outdoor applications)
- WLAN
- Bluetooth

In order to understand the specific aspects of each of these systems *vis-à-vis* antenna requirements, an outline on the overall system aspects plus relevant descriptions in each case are presented below.

7.5.1 Cordless wireless telephone

A cordless phone is a combination of a conventional telephone and a radio transmitter/receiver. It has two major parts:

- The *base* attached through the phone jack to a standard phone wire connection. It receives the incoming call through the phone line and converts it for FM broadcasting
- The *handset* receives the radio (FM) signal broadcast from the base, demodulates it and gives it to the speaker allowing the subscriber to hear. When the subscriber talks, the handset broadcasts the talked message by a second FM signal so that the base picks it up, demodulates it and sends it to the phone line to the other subscriber.

The designated carrier frequencies for the cordless phone generation are:

- 43-50 MHz (10-25 channels; range - 300 m)
- 900 MHz (20 to 60 channels; range - 200 m)
- 2.4 GHz (50 to 100 channels; range-unspecified)

The reasons for using antennas in cordless phone systems are two-fold: One for the base unit and the other for the handset. Relevant design considerations and descriptions are presented later.

Two-way radio systems

These are point-to-point mobile phones capable of medium-range (2 miles) and long-range (5 miles) coverage feasibility. They are intended for "neighbourhood communication", outdoor-to-outdoor and/or indoor/outdoor to outdoor/indoor environments. They operate as 14/32 channel radios on VHF/UHF, *family radio service* (FRS). The operating frequencies are as follows:

- VHF high-band 150-174 MHz
- UHF bands
  - A 409-450 MHz
  - D 450-480 MHz
  - E 470-490 MHz

The channel spacing is 25/30 kHz. The modulation corresponds to FM with the maximum frequency deviation of ± 5 kHz. The receiving system is based on the double conversion superheterodyne principle. No licence is needed to operate the radio.

The medium-range radio uses a simple flexible dipole with an omnidirectional azimuthal pattern. The long-range set needs a base station with an

antenna mast. Typically, the RF energy handled by this antenna structure is up to 500 watts (into a 50-ohm impedance) in the VHF band. The gain is typically 5 dB. The pattern, again, is omnidirectional in the azimuth. The VSWR is about 1.5.

7.5.2 Wireless LAN (WLAN)

Specified by IEEE 802.11, the proliferation of high-performance portable computers combined with end-user mobility plus a connectivity to other communication (wireline) networks and service providers in a local area environment refers to the *wireless LAN* (WLAN) technology [7.11]. It is a flexible data communication system implemented as an extension to or as an alternative for, a wired LAN within a building or a campus. Using radio frequency (RF) technology, WLANs transmit and receive data over the air, minimising the need for wired connections. Thus, WLANs combine data connectivity with user mobility; and, they enable movable LANs through simplified configuration. With the advent of WLAN, users have the ability to operate their portable computers globally while remaining connected to legacy LANs on a wireline and to service providers. The intention to deploy wireless LAN largely resides with the computer-specified mobility, which can be categorised as follows:

- The first type refers to a simple physical relocation/wandering of computer that retains its network connectivity at either end. For example, a person roaming around with the PDA in operation
- The second version is an *ad hoc* LAN set up so as to allow a group of people to share data/information via computers.

In WLAN environments, several computers under mobile conditions may remain interconnected as well as connected to conventional networking and service providers.

As indicated in Chapter 1, the physical media advocated for WLAN are the infrared (IR) and radio frequency (RF) channels. Relevant to antenna deployments, only the RF channels are of concern presently. RF LANs are typically implemented in the industrial, scientific, and medical (ISM) frequency bands 902-928 MHz, 2400-2483 MHz and 5725-5850 MHz.

The front-end entities at the transreceive units mediating the EM propagation across the physical medium of the radio channel are the antennas. Their design and implementation should match and be consistent with the WLAN specifications. The relevant ISM band EM radiation/reception handled at a portable WLAN product by an antenna must comply with the pattern and band selectivity criteria so as not to infringe any spectral etiquette. Though WLAN can be installed without having to worry about licensing on the frequency, the extent of power radiated should be within the regulatory constraints. This constraint is pertinent due to increased demand for RF products. In the United States, for example, it is necessary to implement spectrum spreading (that is, apply spread-spectrum technique in the transmissions) for operation in the ISM bands. Another design constraint is to confine the emitted spectrum to a band, necessitating selective amplification at desired carrier frequencies, frequency-conversion using

precision local oscillators, and the use of selective components. The antenna structures come under the relevant system integration considerations.

Further, RF systems must cope with ambient noise that is either naturally present (for example, atmospheric noise) or is a man-made (such as those EMI emitted by microwave ovens, copiers, laser printers, and other heavy electrical machinery). In addition, RF LANs operating in the ISM frequency ranges may suffer interference from amateur radio operators.

Operating RF LANs indoors introduces additional problems caused by multipath propagation, Rician/Rayleigh fading, and absorption. As discussed earlier, many materials used in civil engineering could be lossy to RF transmissions passing through them resulting in absorption. This would lead to a reduced coverage within rooms. The coverage range further depends on obstacles in the room that may reflect, diffract and/or scatter the EM energy of the RF radiation. The resulting multipath propagation would severely affect the RF signal received. A method of alleviating such destructive interference of the signal composed of the multipath components is to use directional beam antennas so as to realise LoS propagation as much as possible between the transmitter and the receiver. Another technique is to go for antenna diversity reception.

As discussed before, in indoor RF ambient, the signal fading is specified by the Rician/Rayleigh pdf. It corresponds to the statistics of signal fluctuations caused by the movement of an end-entity (such as a laptop) whose overall dimensions are comparable to the wavelength used and implicitly govern the dimension of the antenna used.

The *medium access control* (MAC) protocol for wireless LAN should meet harsh wireless ambient with strong fading channels, where channel characteristics could change over very short distances resulting in unreliable communication and unfair channel access due to capture effects. Further, unlike in legacy-wired LAN, the carrier sensing in WLAN warrants a larger time (30 to 50 s), which may correspond to a significant part of the packet transmission time.

The WLAN has a unique cellular structure. Dividing the associated region of coverage into cells facilitates the required wireless coverage in a building. Each cell has an *access point* (AP) called a *mobile support station* (MSS) or " base station", which is connected to some wired network (such as a legacy LAN). The wandering mobile users are called *mobile hosts* (MHs). The MSS performs the function of channel allocation and provides connectivity to existing wired networks, as illustrated in Figure 7.9.

The difficulties encountered in implementing the WLAN layouts are:

- Possibility of dynamic user relocation between the cells
- Acknowledgement by the nodes of the movements of users between the cells.

There are two modes of communication links in WLAN: *Uplink* depicting the mobile-to-MSS communication and *downlink* depicting the reverse communication. The *downlink* communication often conserves major bandwidth (about 75%), since the data passage is more intense from the MSS than between the mobile hosts. The above considerations have been focused in the development of IEEE 802.11 standards towards an appropriate MAC protocol.

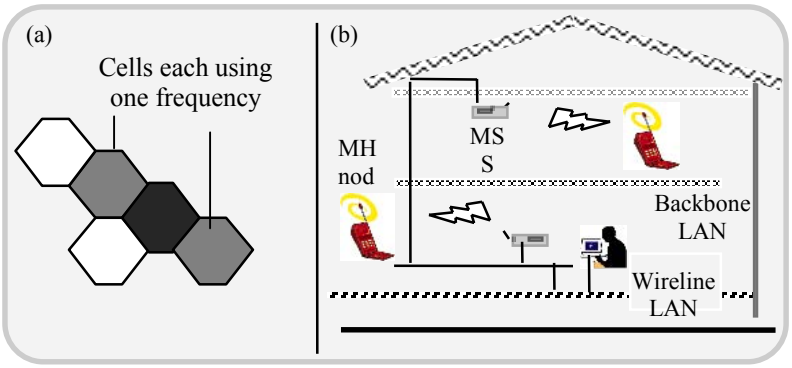


Fig. 7.9 WLAN: (a) Cellular structure and (b) layout of infrastructure mode of operation. (*Note*: Each room is considered as a cell)

The physical media layer of the WLAN architecture operates as a mediator of physical connectivity between the wireless LAN connected entities. With wireless LANs, users can access shared information without looking for a place to plug in, and network managers can set up or augment networks without installing or moving wires. There are two types of WLANs, namely, *ad hoc* (peer-to-peer) and *infrastructure* layouts.

In the infrastructure mode of operation illustrated in Figure 7.9(b), the communication links prevail between the mobile hosts and the access point (MSS). The corresponding antenna system should facilitate the best transreceive conditions between such communicating units.

Another WLAN architecture is the *ad hoc* network depicted in Figure. 7.10. Here, the communication links are between the host units with no specific access point.

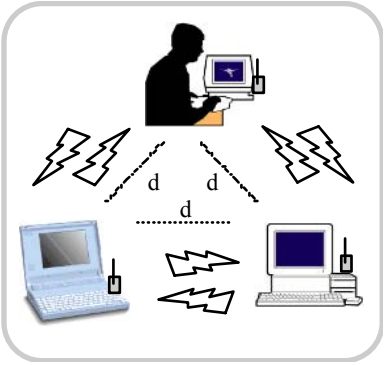


Fig. 7.10 An *ad hoc* WLAN configuration. (The separation (d) between nodes can be up to 500 feet)

In the case of *ad hoc* WLAN, the antenna system should enable transreceive links between the hosts with desirable quality across the wireless medium involved. The WLAN, as an *ad hoc* (peer-to-peer) network, is the simplest configuration. It contains a group of portable computers with wireless adapters. These portable computers have the same workgroup name, or a password (as applicable). Any time, two or more wireless adapters within access range of each other can establish an independent network. These on-demand networks typically require no administration or preconfiguration.

The access points can extend the range of *ad hoc* networks by acting as repeaters, effectively doubling the distance between wireless PCs (Figure 7.11). These access points may be located at strategic points in a room, for example, in the ceiling. Therefore, the antenna structure deployed at such access points should comply with such placements and surroundings so that a desirable pattern for adequate serving the participating units is foreseen.

With reference to infrastructure network, there are two versions of it. These can be described as an *infrastructure WLAN* and an *overlapping infrastructure WLAN*. In an infrastructure network, multiple access points link the WLAN to the wired network and allow users to efficiently share network resources (Figure 7.12). The access points not only provide communication with the wired network but also mediate wireless network traffic in the immediate neighbourhood. Multiple access points can provide wireless coverage for an entire building or campus.

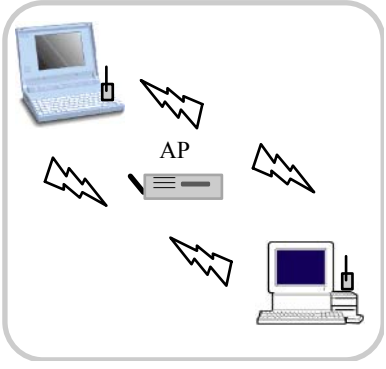


Fig. 7.11 Extended ad hoc network: Separation between the nodes are doubled as a result of the interposed access point (AP)

Wireless communication, in general, is limited by how far the radiated signals are reliably received for a given power output. The cellular architecture in WLANs is known as *microcells*. Using the microcells has the purpose similar to that conceived in cellular telephone system - that is, to extend the range of wireless connectivity.

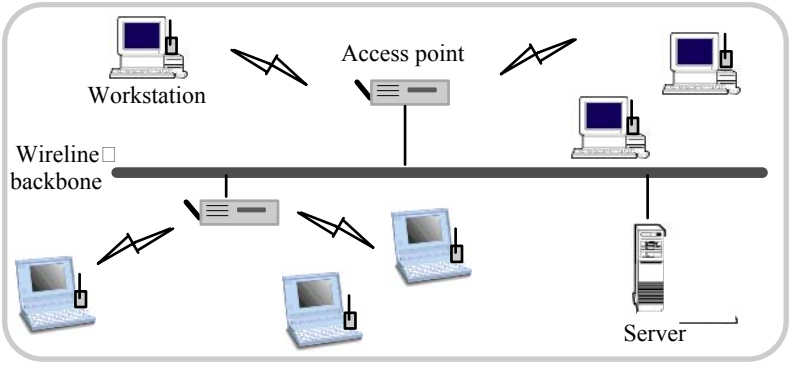


Fig. 7.12 Infrastructure network: The hosts carry the wireless LAN PCMCIA cards

At any instant of time, a mobile PC equipped with a WLAN adapter can be associated with a single access point and its microcell, or an area of coverage. Individual microcells may overlap to allow continuous communication within a wired network (Figure 7.13). They handle low power signals and hand off users as they roam through a given geographic area.

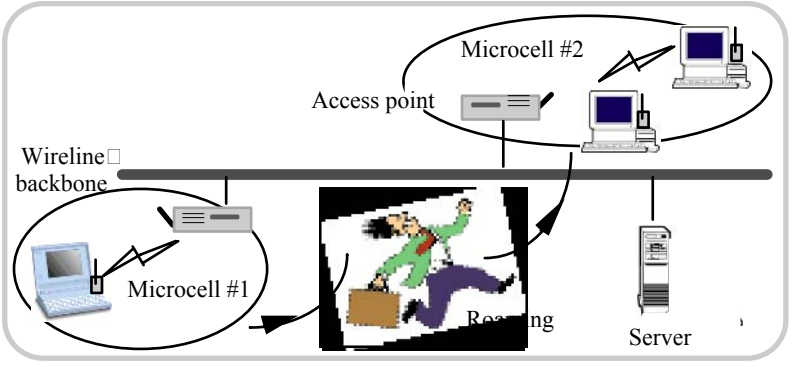


Fig. 7.13 Microcells and roaming: The roaming WLAN is facilitated into the existing LAN backbone

Spread-spectrum (SS) technology, used by most wireless LAN systems, is a wideband radio frequency technique designed to trade off bandwidth efficiency for reliability, integrity, and security. As indicated in Chapter 1, there are two types of spread-spectrum technology, namely, frequency-hopping and direct sequence. Frequency-hopping spread-spectrum (FHSS) uses a narrowband carrier that changes frequency in a pattern known to both transmitter and receiver. Properly

synchronised, the net effect is to maintain a single logical channel. To an unintended receiver, FHSS appears to be a short-duration impulse noise.

In either *ad hoc* or infrastructure implementation of WLAN, the indoor propagation model should be such that the reflections from the ceiling and floor of the building must be duly accounted for. Using multiple image concepts applied to transmitting antenna placed in the ceiling and/or floor, relevant modelling can be done. Transmission around bends and into the sides is analogous to propagation in a high-rise building environment and may be studied using *ray-tracing techniques* developed for use with a two-dimensional building database. When the ceiling is curved rather than flat, more rigorous ray techniques should be deployed to account for the effect of the curvature on the divergence of the ray tubes.

In reference to WLAN, the simplest ray approximations used to predict the propagation characteristics (within buildings) refer to computing the path loss evaluated by considering only the direct ray from the base station AP/MSS to the MH. In this approach, wall and floor losses are added to the free-space loss associated with the direct ray. More general ray-tracing models apply to two or three dimensions for predictions that include both reflection and transmission at the walls, ceilings and floors. The two-dimensional codes are used for coverage over single floor and neglect reflections at ceiling and floors. Because of multiple reflections at the walls, ray-tracing procedure, in general, leads to a binary tree that must be constructed to find all the contributing rays. In three dimensions, the rays that can penetrate the floors and ceiling and reach other floors as well as getting reflected from the floors and ceilings should be duly accounted for.

Generally, indoor wireless communication should allow easy set-up, reconfigurability, portability and mobility for phones, terminals and computers connected within the network. In view of the transmission reliability, it should be noted that high bit-rate transmissions are not easy to achieve with wireless indoor communications, as compared to the wired connections, as a result of multipath propagation effects. The resulting delay-spread experienced at the receiver may also result in intersymbol interference whenever the delay-spread exceeds 1/10 of a symbol period.

A higher bit rate can be realised by employing signal-processing and architectural modification of the system components [7.12]. The later approach involves reconfiguration of transmitter-receiver separation or elimination of reflected rays. The delay-spread can be minimised in a number of ways, as reported in the literature. Among those techniques are: 1) using circular polarisation instead of linear polarisation with directive antennas [7.13, 7.14]. (Reversing the handedness of a circularly polarised reflected wave and employing similar transmit and receive antennas would help rejection of multipath components emerging after a single reflection. Since channel degradation is primarily caused by singly-reflected waves, this approach would effectively suppress the multipaths); 2) using a distributed antenna system [7.15], in which case, the user is always, close to an antenna. This may lead to installation of a base-station antenna to the ceiling of every room within the premises; and, 3) using directive antennas facilitating almost a LoS transmission. With narrower antenna beams, most of the signal delay paths are eliminated.

Driessen [7.16] has reported a successful implementation of this approach by using directional antennas with 15°-beamwidth to reduce the multipath components. A composite notch-wire antenna for polarisation diversity reception in an indoor base-station system has been reported by Kuga et al. [7.17]. Multiple-beam antennas can also be used to illuminate multiple receiving terminals in an indoor environment. Such *radiation diversity* can significantly reduce multipath transmission due to narrower beamwidths. Body-shadowing effects can also be significantly reduced or eliminated by employing similar receiver antennas and using space diversity for the indoor base station.

In addition to the radio channel environment and antenna perspectives described for WLANs, another indoor propagation consideration that requires specific deliberation refers to the Bluetooth system. In the following section, the Bluetooth technology is revisited and its associated indoor propagation channel characteristics, as well as the prevalence of possible EMI/RFI ambient, are reviewed comprehensively.

7.5.3 Bluetooth technology

The Bluetooth system [7.18-7.20], as elaborated in Chapter 1, refers to a default radio interface that allows handheld electronic devices to be rapidly interconnected into *ad hoc* networks. It has been conceived as a class of short-range radio technology, which enables users to rapidly interconnect electronic devices such as cellular phones, palm devices or notebook computers. The emergence of Bluetooth as a default radio interface in these devices provides an opportunity to turn them from being stand-alone tools into networked equipment. Building Bluetooth *ad hoc* networks also represents, however, a number of new challenges, partly stemming from the fact that Bluetooth was originally developed for single-hop wireless connections. Bluetooth allows large numbers of *piconets* to form a *scatternet* using designated nodes that participate in multiple piconets. That is, Bluetooth is a short -range radio technology (operating in the unlicensed ISM band at 2.4 GHz) and its terminal units are networked into piconets. There is one Bluetooth device in each piconet that acts as the *master*, which can have any number of *slaves* out of which up to seven can be active simultaneously. Being a master or a slave is only a perception of logical state — any Bluetooth unit can be either a master or a slave.

Bluetooth, in short, denotes a low-cost, short-range radio link, which facilitates *ad hoc* communication environments, and is intended to enable users to connect to a wide range of computing and telecommunications devices without the need to buy, carry, or connect cables. In general, Bluetooth can be used in phones, pagers, modems, LAN-access devices, headsets, as well as notebook, desktop, and handheld computers.

As mentioned earlier, Bluetooth uses the frequency-hopped spread-spectrum (FHSS) technique (at 2.4 GHz ISM band) to combat potential interlopers from baby monitors, garage-door openers, cordless phones, and microwave ovens, all of which operate in the same frequency band. Bluetooth has built-in security, and supports isochronous and asynchronous services and easy integration of transmission-control protocol/Internet for networking. A Bluetooth unit can participate in more than one piconet at any time but it can be a master in only one piconet. A unit that participates in multiple piconets can serve as a bridge, thus

allowing the piconets to form a larger network. A set of piconets that are all interconnected by such bridging units is referred to as a *scatternet network* (Figure 7.14).

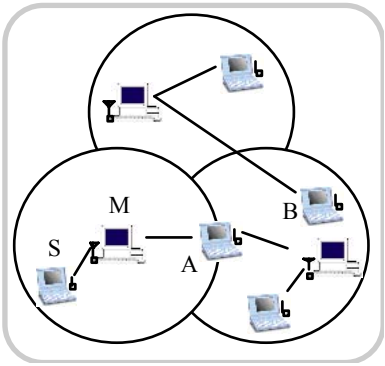


Fig. 7.14 A scatternet. (M: Master; S: Slave; A: Slave in two piconets; and, B: Slave in one piconet and master in another)

Since a Bluetooth unit can transmit or receive in only one piconet at a time, bridging units must switch between piconets on a time division basis. Due to the need for synchronising its radio from one piconet to another and perform the necessary signalling, a Bluetooth unit necessarily loses some time duration while switching, which represents an important performance constraint in setting up scatternets.

A given set of physical nodes can be arranged in scatternets in a number of alternative ways within the hard constraint represented by physical distance. By grouping nodes into piconets and by selecting the bridging nodes one can significantly impact on the resulting performance parameters, such as the attainable throughput. Knowing the scatternet topology that optimises performance would be of prime importance for any Bluetooth implementations. However, the complexity of the problem seems to prohibit an analytical approach and the large number of free parameters makes it problematic to use numerical optimisation through simulation.

Bluetooth has a *slotted MAC protocol*, where master-slave and slave-master communications take turns. Packets can be one, three or five slots long. As a result, the amount of allocation and goodput seen in the two directions and the total overhead depend on the asymmetry of traffic on a Bluetooth link.

In summary, Bluetooth is a universal, short-range wireless communication system operating in the 2.4 GHz Industrial Scientific and Medical (ISM) band. It allows small portable devices to connect and communicate together in an *ad hoc* architecture. However, Bluetooth profile specifications define more stationary LAN access profiles compatible for Bluetooth environment. In addition to conventional LAN access, this profile can be used to secure connectivity access to several other wired network entities like printers and/or noncommunication items,

such as home entertainment system, refrigerators etc. In this extended profile operation, the access point (AP) is the master of the piconet. This feature makes it feasible to implement novel adaptive antennas in a Bluetooth environment, as will be indicated later.

Bluetooth has three different link types. One is reserved for control purposes and the other two for communication needs. Both real-time, constant-rate, *synchronous connection-oriented* (SCO) links as well as, variable-rate, *asynchronous connection-less* (ACL) packet-data links (with ARQ) are supported by the Bluetooth systems.

Further, a Bluetooth system uses binary *Gaussian frequency-shift keying* (GFSK) modulation with a modulation index in the range 0.28 to 0.35 and the modulation bit rate being 1 Mbps. Frequency-hopping over 79 carrier frequencies is applied between subsequent bursts. Different piconets are not synchronised in time, but they are distinguished by independent hopping sequences that are defined by a Bluetooth identification (ID) number and the clock of the master. Within a piconet, the transmission time is divided into slots of 625 μs each. Downlink transmissions (master-to-slave) are set at even slots and uplinks are placed in the odd slots[*]. The Bluetooth operation is, therefore, *time-division duplex* (TDD). The specification of a Bluetooth system is summarised in Table 7.1.

The frequency-hopping rate is at 1600 per second (that is once per slot). The access code is 72 bit long for piconet identification and synchronisation purposes. A 54-bit header decides the connected slave and packet properties. The payload field is variable.

Thus, Bluetooth forms an interesting, "top-notch" technology that is being ushered in and enthusiastically promoted so as to enable devices of all kinds -from laptops and cell phones, to personal digital assistants and household appliances - with the capability of communicating and interoperating with one another in the wireless medium.

Table 7.1 Basic Bluetooth specifications: RF and modulation parameters

Parameter	Specifications
Radio frequency	2.40-2.48 GHz
Multiple access	TDMA/TDD: Maximum: 8 units
Frequency-hopping (FH)	1600 hops/s
FH timing, one slot length	625 μs
FH timing, TDD guard time	220 μs

[*] Bluetooth timing is explained and illustrated in Appendix 7.1

FH channels	79 (or 23) 1 MHz channels
Modulation	GFSK with bandwidth-time product (WT_B) equal to 0.5 Modulation index = 0.35
Bandwidth	1 MHz (– 20 dB); and 220 kHz (– 3dB)
Bit rate	1 Mbps
Receiver sensitivity	– 70 dBm
Transmitter power Class 1 Class 3	 20 dBm (100 mW) 0 dBm (1 mW)

Taking a tour into the emerging vista of Bluetooth indicates that the underlying concept of this technology is a technique that could enable a transparent wireless RF communication between electronic devices. As indicated above, this mode of communication can offer a set of advantages encompassing a viable voice/data access point, no cumbersome cabling and creation of an "as-you-please" *ad hoc* wireless network. Bluetooth technology is characterised by personal connectivity of enabled devices, spontaneous creation of disposable networks and low cost, low power indoor wireless transmission of short ranges supporting 721 kbps or 432.6 kbps over asynchronous and synchronous links, respectively. With Bluetooth initiatives placing their standards on wireless communications, an associated problem that looms in the horizon in the form of EMI is attributed to electromagnetic emanations from high power ISM devices, such as microwave ovens. Since Bluetooth is largely conceived as an indoor, *ad hoc* mobile communication facility and, inasmuch as the microwave oven is an indispensable domestic appliance, the question is how to resolve the compatibility issues on the persisting EMI on the Bluetooth interface caused by microwave leakage from the oven.

Since Bluetooth is a short range (< 10m) radio technology, which operates in the 2.4 GHz ISM band and, inasmuch as, the ISM band is free for use, there are many sources of interference, for example, microwave ovens, WLAN etc. In addition, Bluetooth uses low power in the range of 1mW-100mW. To avoid interference, the physical layer supports FH-CDMA transmissions. Further, Bluetooth uses 79 carrier frequencies and as per the regulated prescription, the signal bandwidth is limited to 1 MHz. The nominal bit rate is 1 Mbps and the sensitivity level (the input level for which the raw BER specification is met) and co-channel interference resistance are specified as –70 dBm and 11-14 dB respectively (for BER < 0.001).

Higher layer packets are fragmented into Bluetooth physical layer packets. Each packet has 126 bits of control information and a maximum payload of 2745 bits. Packet transmission takes place on one hop frequency and can take multiple slots. Full duplex communication is achieved by implementing TDD. The master exercises the access control. A communication is possible only between the master and a slave. The master transmits in an odd numbered slot and the corresponding slave responds in the next slot. 1/3- or 2/3-FEC can be used for data. In addition, an 1-bit acknowledgement indication (ARQN) is implemented. If the data is received incorrectly, it is retransmitted at the next opportunity.

Bluetooth environment modelling

A typical approach towards modelling a Bluetooth ambient refers to a room environment filled with a set of Bluetooth devices randomly dispersed (in the spatial locales of the room) constituting M piconets. Such dispersion may also depict the overlap of piconets. Since Bluetooth operates in the ISM 2.45 GHz band, as indicated before, there could be many sources of interference other than Bluetooth devices, such as a microwave oven and IEEE 802.11 WLANs. A piconet would experience interference when a frequency hit occurs, that is, when the transmission frequency of a packet in a piconet matches with that being used in one or more other piconets for some overlapping duration.

Since the devices in the piconets are spatially distributed, the *interfering piconet* does not represent a fixed location or a device but a random device in that piconet transmitting at a given time. A physically stationary environment (with little or infrequent movements) can be assumed so that the channel can be regarded as quasi-static and quasi-wide-sense stationary. With this stationarity condition for the *duration of a frequency hit* on the single frequency, constant interference power can be assumed to be received from the individual interferers. This model can predict the outage probability and mean outage duration as functions of the *signal-to-interference ratio* (SIR). These entities, in general, would monotonically increase with the number of the piconets in the room.

Any attempt to improve the performance of Bluetooth deployment in a given indoor environment (at the physical layer level) should correspond to a judicious choice of antenna structure. This choice should reflect appropriate polarisation of the radiation involved so as to improve the SIR of the links even in harsh environments (such as an industrial floor) and in the presence of other ISM devices or WLAN coexisting in the same ambient.

Pertinent to the Bluetooth implementation considerations briefly reviewed above, the *in situ* aspects of performance of indoor wireless communications (at the physical layer level) are largely decided by the prevailing channel and propagation characteristics of the indoor ambient This is much akin to the WLAN system described earlier. In summary, relevant radio channel profiles are explicitly governed by the following:

- Characteristics of room/building interiors (size and architecture) wherein the piconets are formed on an *ad hoc* basis and distributed arbitrarily in space
- Interior furnishing and structural/material characteristics

- Coexistence of other wireless communication systems such as WLAN
- Coexistence of other noncommunication ISM-band devices such as microwave ovens.

The antenna system conceived for Bluetooth applications should facilitate a high-performance of the radio links established in the indoor channels having the above attributes. The underlying aspects of the properties and modelling of a Bluetooth environment are reviewed below, when considering antenna deployment in indoor applications.

The antenna performance for indoor applications (such as Bluetooth) should conform to effective short-range transfer of control and data signals between the end-entities, which are designated to operate at low RF power levels. The associated voice and data transmission quality relies not only on the measures taken at a higher layer (such as CRC, FEC), but also on the proper design of physical layer connectivity that endures the effects of multipath and Doppler shifts caused by indoor reflections, mobility and interference sources. As discussed earlier, indoor wireless communication is essentially specified by the physical channel response and signal-fading governed by the indoor civil engineering (architecture and materials), human occupancy and furnishing plus the existence of other electronic devices.

The associated multipath profile signature, delay-spread and transreceive path loss decide the spatial and/or temporal variations on the correlation of signals among multiple antennas, *direction-of-arrivals* (DoA) and multipath angle-spread. Relevant details are essential in designing a smart antenna system for the short-range connectivity sought. This connectivity profile, in a global sense, should include the entire range of indoor devices – laptop and desktop computers, printers, scanners, pagers, and cordless, wired and wireless telephones. Connectivity among these devices should also explicitly specify the shared resources, such as scanners and printers, transfer of e-mail, Internet, and video data among wireline and wireless units. Further the networking must indicate the mobility of pagers, wireless and cordless phones, as well as the immobility of PCs and the portability of laptop computers. Configurations such as *home local-area networks* (HLANs), that use telephone lines as a tool for connectivity are often faced with difficult challenges in their performance. This is due to variations in the voltage levels and impedance of the lines, poor shielding, and improper line termination. Furthermore, the use of phone lines, in general, would limit the mobility of wirelessly-linked end-entities and may not totally lend itself to future integration of all electrical and electronic devices at home for wireless interconnection. This consideration is important in view of the fact that home appliances could soon be equipped with wireless transmission systems (*via* Bluetooth) to communicate household information to PCs or any other wireless devices.

Short-range wireless connectivity between computer and telecommunication devices at home is dependent on the emerging novelties of achieving reliable robust physical layer links. As more and more devices are deployed, this technology must be evaluated for adequacy and appropriateness of physical layer

connectivity achieved *via* appropriate antenna usage concomitant to the signal impairments present.

In contrast to IrDA, which requires that the receiver be within an angular span of 30^o of a direct LoS with the transmitter, Bluetooth is omnidirectional. That is, the transmitted signal is radiated with equal power in all directions, implementing a point-to-multipoint propagation. In addition, Bluetooth offers 10 times the range of IrDA devices. In general, Bluetooth and WLAN, at physical layer level, should be designed for non-LoS transreceiver operation through walls and/or other solid objects with only moderate attenuation.

Coexistence of WLAN and Bluetooth under a single roof

Modern application profiles of WLAN and Bluetooth may lead to their deployments in the same indoor environment. Typical such coexistence situations may cause mutual interference conditions. The feasible scenarios of jointly using the WLAN and Bluetooth systems under the same roof are elaborated in Appendix 7.2. In developing indoor wireless communication antenna systems, relevant details on such coexistence (deployment and performance considerations of WLAN and Bluetooth) could be useful and pertinent.

7.6 INDOOR WIRELESS ANTENNA DESIGN CONSIDERATIONS

A desirable solution for reliable and robust transreceive performance at physical layer level in indoor wireless applications refers to an indoor station aiming the available (low power) EM energy towards an access point operating at a short range. For example, such an antenna in a Bluetooth piconet should provide minimal interference for other Bluetooth devices in the vicinity and enable at the same time maximum data transfer capability between the communicating units.

Taking into account compactness and low-profile considerations, the wireless antenna developed for indoor application (such as for Bluetooth) are small, made almost "invisible" and often sheltered as an "embedded" structure in the radio unit. Conceivably, they could be a single passive resonating structure at the front-end of the transreceive system or an actively embedded unit (of the front-end) or it could be a "smart" array with adaptive characteristics as regard to their radiation pattern beamforming and control networks. Efficient directive patterns with minimal sidelobes could be considered as the design targets for these antennas. Ultimately, the performance of the system, which uses such antennas should be of low cost and offer good quality of service (QoS) with low connection time, and reduced transmission power (so as to achieve an extended battery life). Also, considering exclusive environments, where Bluetooth-enabled non-communication devices (such as ovens, refrigerators, sauna heaters etc.) may exist, the antenna in question should easily be implemented while at the same time considering the feasibility of mounting it on conventional communication devices like telephone, laptop, PC, PDA, printer etc.

7.6.1 Traditional antennas for indoor applications: System-specific aspects

In view of different wireless communication systems developed and deployed for indoor applications, the antennas to be used in the relevant radio channels are system-specific. That is, in selecting and designing an appropriate antenna for any

given system (such as WLAN or Bluetooth), one has to give due considerations to the system requirements, associated protocols and performance specifications detailed for the application in question. There could be several options (with or without smart considerations) in the choice of antenna configurations for a given application, as described in earlier chapters.

Among the plethora of such choices, the simplest approach of antenna deployment in an indoor wireless system (as indicated earlier in this chapter) is to use basic short dipole/monopole structures suitably mounted on the devices so as to realise an omnidirectional azimuthal pattern. This, however, provides coverage between any transreceive system without any directional discrimination. The disadvantage of this technique is the poor link quality that may prevail when the transreceive path is obstructed or shadowed. Also, the omnidirectional coverage could cause interference with other wireless systems operating in the vicinity; or, the antenna in question may pick up interference from the sources like microwave oven etc. that may operate in nearby areas. Nevertheless, because of their simplicity, dipole-based antennas have been popularly used in indoor wireless communication applications. Some of the practical versions, which are system-specific, are described below.

7.6.2 Cordless phone antennas

Consistent with the description of cordless phones presented earlier, there are two antennas in this system — one is at the base and the other is a part of the handset. In either case, a telescopic-type dipole could be adopted. Further, the base station antenna can be made tiltable and/or of swivelling type for optimum link performance. Typically a telescopic $\lambda/4$-monopole is convenient for handling both at the base and at the handset at the VHF/UHF band of the cordless telephone. Alternatively, a ribbon-type element ($\lambda/4$-monopole) of push-in/push-out type can be used. This is thin, flexible and convenient for handling and housing in compact units. To reduce the length less than $\lambda/4$ (especially at VHF/UHF bands), an inductive-loading can be used for a self-resonant operation. However, its bandwidth would be lower than a monopole.

Lengths in excess of $\lambda/4$ in a monopole structure (with L = 3 $\lambda/8$, 5 $\lambda/8$ and 7 $\lambda/8$) when used (say in base units) could provide higher gain than $\lambda/4$ monopole. The relevant input resistance is, however, greater than 50 ohm and inductive. Therefore, a matching circuit is necessary. The radiation patterns of such antennas are upward tilted.

A ($\lambda/2$) sleeve-antenna is also compatible for handsets of cordless phones. It has less body effects than a $\lambda/4$-monopole and, inherently, it needs no balanced-to-unbalanced (*balun*) transformation.

In UHF band operations, a PIFA either mounted on or built into the radio unit can be used for portable handsets. Normal-mode helical antenna (NMHA) shorter than a $\lambda/4$-monopole can be designed for self-resonant operation at UHF/VHF bands. It is structurally flexible. A combination of monopole and a PIFA structure can yield diversity with an element spacing of 0.1 λ of UHF band operations offering a low correlation factor.

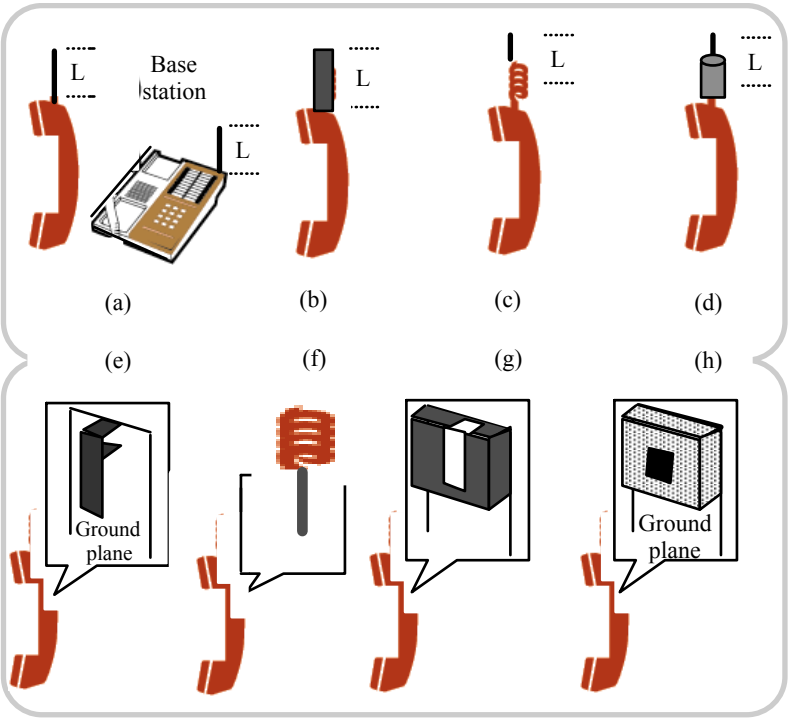


Fig. 7.15 Cordless phone antennas: (a) Monopole antennas (L = λ/4, 3 λ/8, 5 λ/8 and 7 λ/8) — Fixed, telescopic, tiltable and/or swivelling type; (b) λ/4 ribbon push-in/push-out type; (c) inductor-loaded; (d) sleeve type; (e) PIFA; (f) NMHA; (g) bent-slot type; and (h) square-patch type

Specific to portable handsets, the slot antenna can also be implemented via slots facilitated on the top and sides of unit. With the length of the slot being almost λ/2, a bandwidth of 3% to 8% can be realised in self-resonant mode. This antenna offers semidiversity in performance.

A bent-slot on the side of portable units (of length close to λ/2) can again be used as a self-resonant, UHF band antenna.

Patch structure, in the form of a square patch (of side λ/4) with one end shorted, can be used as a narrow band self-resonant radiator (with vertical polarisation) in potable (handset) units.

Various structures of base and/or handset antennas in cordless phones described above are sketched in Figure 7.15.

7.6.3 Antenna for two-way radios: Indoor deployment considerations

As indicated before, the two-way radios are used both in outdoor as well as in indoor wireless ambient. These are hand held transreceivers (and may be supplemented by a base station in long-range applications). These radio units are

much similar to cordless phones and hence, their antenna requirements would approximately refer to those of cordless phones and/or cellular phones with necessary scaling of the antenna dimensions to match the FRS band. The long-distance version of the two-way radios may require a high-power base station with an antenna erected outdoor as a mast.

7.6.4 Antenna for 2.4 GHz ISM band

The antenna design for 2.4 GHz ISM applications, such as Bluetooth-enabled devices, wireless data-control, WLAN etc. is affected by the lack of space (physical volume) in the radio unit. The device/unit that houses a 2.4 GHz transreceive circuit should be designed to provide space for the circuit, the associated package shielding and production tolerances. Such an antenna should be conceived with a predesign decision whether it has to serve as an omnidirectional or high directivity radiator. The directional consideration should be linked to (i) the capacity, (ii) EMI/RFI ambient (SIR), (iii) spectrum sharing rules or etiquette profile, (iv) range *versus* power (10 mw/100 mw) and (v) power-control provisions consistent with the QoS requirements.

A simple omnidirectional antenna (in the azimuthal plane) in compact form can be realised for 2.4 GHz applications as a standard version or as a custom-built type to match, or as a tailor-made version to suit the mounting feasibility in WLAN and/or Bluetooth-enabled services.

The standard versions can be of a *swivel type* or a *snap-in structure* or a *surface-mounted* unit. The swivel radiator is a ground–plane independent $\lambda/2$ dipole made for minimal space considerations. The associated design could include (i) mounting the antenna in a standard SMA connector placed either on the casing or directly on the PCB; and, (ii) the other version can be soldered directly on the PCB so as to allow the radiator be placed outside the casing.

Swivelling allows the radiator to be a flexible structure rotatable over 360^{o} so as to enable optimum *in situ* transreceive performance. Further, the antenna can be locked in any desired angular disposition. In any Bluetooth-enabled desktop PCs, printers, diverse instruments, network APs and in any development prototypes, use of such antennas could be simple and mechanically robust.

The snap-in version also can offer a high-performance low-profile use at 2.4 GHz. It is again a $\lambda/2$-dipole directly assembled on the PCB without soldering. It is ground-plane independent and ensures stable performance. It can be trimmed to suit the applications such as in all the Bluetooth-enabled devices mentioned for swivel radiators.

To realise miniature radiators, surface-mounting technology can be resorted at 2.4 GHz. Here, the size and/or weight are prime considerations. The SMA antennas at this frequency can be made as small as 15 mm $\times$ 10 mm $\times$ 4 mm in size. It can be manufactured through an automated assembly. It requires no special matching units. In incorporating this type of antenna, the ground-plane of the device (being Bluetooth-enabled) itself can be used as an active part of the antenna. Therefore, selection and placement of this antenna in the unit warrant careful design and engineering considerations. The typical Bluetooth-enable devices where surface-mounted antennas can find applications are as follows: Mobile phones, PDAs, headsets, laptops, peripheral equipment and PC cards.

The gain and VSWR of SMD antennas are platform-dependent. Such antennas can be designed for nominal impedance of 50 ohm.

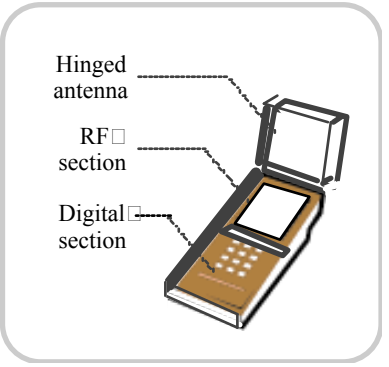


Fig. 7.16 Hinged flat-panel antenna: Hinged movements allow optimal RF transmission and reception

The dipole whip (swivel and ground-independent type) is the simplest form, which offers an adjustable operation with high gain. Its mounting *via* standard connector or reverse-polarity connector (meeting the design requirements and/or compliance with Part 15 FCC regulations) offers convenience. The nominal operational characteristics of such antenna are as follows:

Frequency range	2.4 – 2.5 GHz, monopole (λ/4) structure
Impedance	50 ohm nominal
Gain	1 to 3 dBi
Radiation pattern	Omnidirectional
VSWR	< 2:1 maximum
Polarisation	Vertical

The whip can be made of flexible polyurethane and the swivel mount could be made of hard polycarbonate. The connector is typically an SMA plug and if needed a TNC plug and/or reverse polarity SMA or TNC-plug, reverse thread SMA and MCX plug connectors can be used. This antenna can be designed to operate over a range of – 20° C to + 65° C.

In systems like Bluetooth-enabled devices, unless the physical orientation of the device is specific, the Bluetooth itself should appear as a directionless system. Further, should the device itself be shielded for RFI, then the Bluetooth antenna must be kept on the exterior of the device, housed in a dielectric plastic case.

In the envisioned Bluetooth operational scenario, the relevant antennas being considered and deployed are dipole, flat-panel and microstrip versions. In specific circumstances where a directional deployment is warranted, use of arrays

may not be ruled out. Typically, monopole structures fed *via* coaxial cable or central-fed dipole of dimensions $\lambda/4$ to $\lambda/2$ can be used in Bluetooth applications with snap-in surface-mounted or swivel-facilitated fixtures through appropriate connectors.

The flat-panel radiator (based on patch antennas) is an alternative antenna for Bluetooth devices and it can be directly mounted on the PCB. The PIFA type antenna is compatible for this application. A conceivable design for Bluetooth or WLAN applications in the form of a hinged polarisation diverse operation is described below.

The flat-panel antenna that can form a part of the PCB card of a Bluetooth or WLAN PCMCIA card is illustrated in Figure. 7.16. It can include polarisation diversity so as to minimise fading and optimise performance in a dynamic EM ambient. The antenna is built in a rotating/hinged structure linked to a panel that may support the RF and digital electronics

The flat-panel antenna could be made polarisation-diverse due to the following reasons: The Bluetooth and/or WLAN systems operate indoor where the relevant LoS of RF links could be severely obstructed and exhibit fading, as discussed earlier. Therefore, to compensate for the resulting fading, spatial or polarisation diversity can be adopted, which can provide about 9-dB link-margin against fading as compared to a nondiverse system. For unobstructed links, spatial diversity may become ineffective, should the transmitter and receiving antennas become cross-polarised (due to their portability). Therefore, it is preferable to use polarisation diversity instead of spatial diversity.

The planar assembly of the antenna housed in the hinged structure with almost 135^{o} swivelling (in E-plane) is therefore designed for polarisation diversity. In the normal disposition of the panel set vertical, the antenna radiates or receives almost omnidirectionally. A swing of $\pm$ 45^{o} about the vertical is facilitated to account for the adaptability to indoor EM environment of the operating area in order to make the null off from the desired direction of transreception.

The (polarisation-diverse) antenna geometry could be conceived with two orthogonal folded-slot radiators encased in a plastic housing. The slots are half wavelength folded types and remain within the allowable antenna footprint of about 2" $\times$ 1.5". This structure, as depicted in Figure 7.17, has an electrically small ground-plane and proximally located orthogonal slots. Relevant microstrip feedings from the edge of the panel to the antennas can be optimised for VSWR ($<$ 2:1) over the operating frequency of 2.40 to 2.48 GHz. The electrical connections from the edge of the panel to the RF front-end are realised *via* two flexible coaxial cables.

Radiation patterns in the elevational and azimuthal planes of this antenna would approximately correspond to classical dipole patterns with the null axis orthogonal to the length of the slot. In practice, the azimuthal pattern may be skewed as a result of the plastic housing [7.21].

Another form of dual-polarised antenna can be realised with *coplanar waveguide* (CPW). These CPW antennas are similar to microstrip antennas. Both are configured as open-ended transmission lines with cross-sectional expansion.

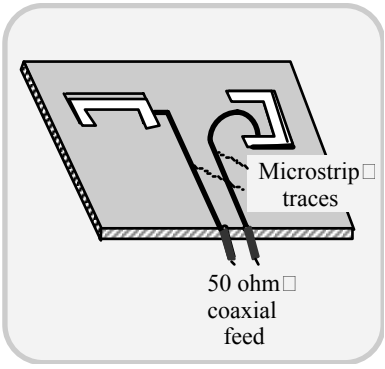


Fig. 7.17 Polarisation-diverse antenna: Two orthogonal microstrip traces constitute the radiating elements

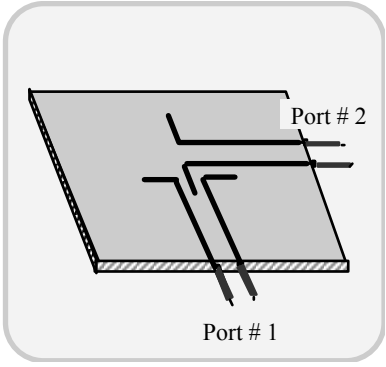


Fig. 7.18 Planar CPW coupled/patch (dual-polarised) antenna with two orthogonal radiating elements

In the microstrip antennas, the feed-lines are microstrip lines and the EM coupling is achieved through an aperture in the common ground-plane. The separation of radiating elements and the feeding network may warrant multilayer structures with separate substrates. Relevant design and making the antenna are complex due to the exact alignment of substrates needed.

In CPW structures, on the other hand, the planar antenna is electromagnetically coupled to the feed lines on a single-layer substrate. An example of a dual polarised antenna based on CPW concept is illustrated in Figure 7.18.

7.6.5 PC-card antennas for 2.4 GHz ISM band applications

The design criteria for PC-card antennas specified for 2.4 GHz ISM band applications can be enumerated as follows. Compact size with low profile, low cost, 80 MHz bandwidth with sharp cut-off, vertical null in the antenna pattern, good efficiency, flexible impedance matching to CMOS low-noise amplifiers

(LNAs), circular polarisation and possible differential operation. Relevant antennas conceived for WLAN and /or Bluetooth applications and meeting the above requirements can be based on PIFA/dual PIFA configuration for reduced complexity.

Most of such antennas are designed to provide standard 50-ohm impedance. However, use of a CMOS low-noise amplifier (LNA) at the front-end may warrant an impedance match condition compatible with the input impedance of the LNA used. Further, if filtering is done at the LNA stage, the relevant function can be implemented as a part of the antenna-to-LNA matching section. Also, in a conceivable design, a differential mode of operation can be introduced so that the antenna provides a differential feed to a (differential) LNA.

Further, the attempt to realise a null in the vertical direction of the radiation pattern is helpful in reducing intersymbol interface as indicated in [7.22]. In addition, if circular polarisation is facilitated, it would ensure antifading performance as a result of the obtained diversity.

The two possible antenna configurations for the above applications are [7.23] as follows:

- Standard PIFA
- Dual PIFA

The geometry of a standard PIFA is presented in Figure 7.19. One end of the PIFA is fully short-circuited and the physical length of the PIFA is about $\lambda/4$ (equal to 19 mm at 2.4 GHz).

The antenna is probe-fed and its ground-plane is the conventional PCB copper backing. The dielectric constant of a standard PCB is 2.4. The probe-feed can also be substituted by a microstrip feed for direct connection to the RF front-end. With reference to the placement of the PIFA on the PCB card (of typical dimensions 85 mm $\times$ 55 mm) from the edge as shown in Figure 7.19, the associated offset distance (d_1) would influence the gain of the antenna. It is reported in [7.23] that the larger the offset length from the edge the higher is the gain.

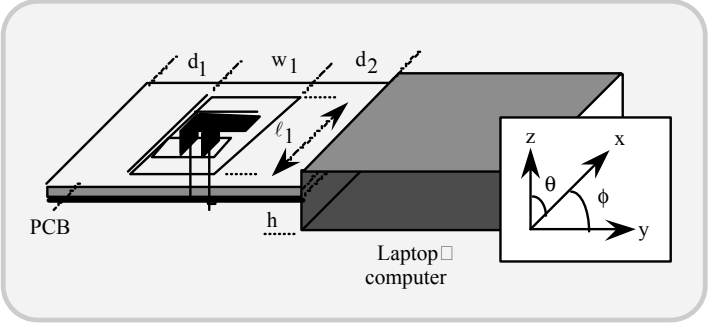


Fig. 7.19 Standard PIFA based PCB mounted antenna for laptop computers

Performance-wise, the studies addressed in [7.23] indicate the following inferences in reference to standard PIFA antennas:

Bandwidth 80 MHz with $S_{11} \leq -10$ dB

Radiation patterns: E-field patterns - (Measured in yz- and xz-planes)

- No nulls in the vertical direction
- The presence of the (metallic) laptop skews the yz-plane pattern towards the laptop and reduces backlobe in both planes
- Gain of the antenna increases with the offset distance (d_1) from the PCB edge

7.6.6 Dual PIFA configurations

This structure consists of two stacked PIFAs placed back-to-back with a common short pin for both PIFA located in the middle of the PIFA; a feed-plate is used in [7.23] with a coaxial feed, as illustrated in Figure 7.20. The performance aspects of a typical dual PIFA can be summarised as follows:

Bandwidth 80 MHz with $S_{11} \leq -10$ dB

Radiation patterns: E field patterns - (Measured in yz- and xz-planes)

- There is a null with vertical direction in both planes
- The presence of the laptop skews the yz-plane pattern

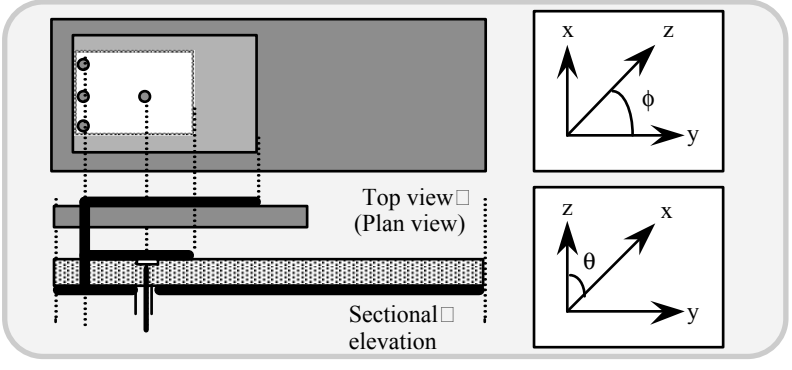


Fig. 7.20 Dual PIFA with two-stacked inverted F radiators of different dimensions

7.6.7 Dual-band antenna for 2.4 GHz and 5.7 GHz indoor wireless systems

A dual-band compact antenna structure for 2.4 GHz and 5.7 GHz WLAN applications can be designed using a folded monopole/loop configuration. Essentially, such an antenna should exhibit a double-resonance. Dual-band resonance can be achieved *via* a monopole helical and the PIFA structures concept used in cellular base station applications, and can be extended to obtain indoor dual-band antenna structures.

Specifically, a simple structure would refer to a folded monopole/loop antenna shown in Figure 7.21, which can be etched onto an electrically thin dielectric substrate and mounted on a ground-plane. The dimensions can be optimised for matched performance at the two dual bands in questions. Appropriate dual-feed arrangements should be incorporated in the antenna system to ensure that a matched operation at the resonances prevails. Typical dimensional relations of this folded monopole/loop, when compared to a standard folded monopole antenna, are approximately as follows [7.24]: $h_2 \approx 2\ h_1$; $s_2 \approx 2.5 s_1$; $\ell_2 \approx 2.5\ \ell_1$ and $d_2 \approx 2.5 d_1$, where the subscripts 2 and 1 depict the folded monopole/loop and standard folded monopole respectively.

These dimensions should, however, be optimised/trimmed to match the performance requirements. Specifically, by controlling the spacing between the traces, the design can be optimised to work at two frequencies without any external matching circuitry. Further, this antenna can yield a good linearly-polarised omnidirectional pattern in the horizontal plane. The substrate dimension and dielectric constant are other design parameters that can be used for system optimisation.

The expected radiation patterns are linearly polarised and omnidirectional in the xy- plane at both resonant bands. However, in the xz-plane some distortion in the pattern may arise as a result of asymmetry in the antenna geometry along this plane. As the technology moves forward towards eventual integration, an antenna on-board a chip is possible as part of the Bluetooth radio chip. Such an antenna, for example, can be implemented as a four-armed, spiral microstrip on the top surface of Bluetooth chip's ceramic package. It can accompany a RF filter printed on the top of the package for the purpose of improving the characteristics of the device's transceiver.

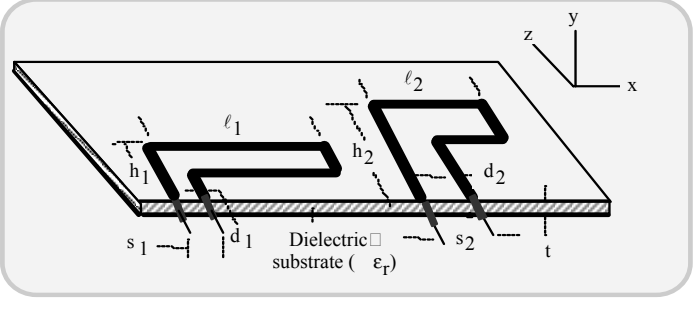


Fig. 7.21 Dual-band indoor antenna for WLAN applications made of microstrip traces

The cost and space savings of such antennas are obvious; however, because the chip could block the radio signal above it, the chip antenna can only radiate in a hemisphere on one side of the chip. However, this might be seen as an advantage for some products such as headsets, where SAR-related health concerns have been expressed. Users may be able to put up with the inconvenience of only being able to connect to devices on one side of their body for the sake of the assurance that radiation levels will be reduced yet further below the already existing low power levels of Bluetooth. There are some devices, which will require more complete coverage. In such cases, even a radio unit made of a chip with an on-board antenna may typically require an alternative external antenna.

The characteristics of an antenna, in general, could significantly be affected by the surrounding ground-planes. Shielding from castings and components may also affect antenna radiation patterns; the feed to the antenna specifies the reflection losses. The combination of these factors dictates the optimum radio performance. Hence, it is crucial to take the geometry, profile, type and mounting considerations of the antenna structure into account while designing PCBs for radio units.

7.6.8 Smart antennas for Bluetooth applications

The adaptive (smart) array concepts can also be pursued for applications in Bluetooth antenna designs [7.25]. Such an antenna belonging to a Bluetooth radio unit can be used to aim a pattern at an access point (AP) or a printer or a similar Bluetooth-enabled device in the piconet. The associated adaptive directivity enables minimal interference to other Bluetooth devices in the piconet placed in the vicinity. Thus, a high performance in data transfer can be realised as a result of low interference.

The adaptive array in question can be designed with radiating elements configured to form a cross pattern, as shown in Figure 7.22. These elements are nonuniformly excited and the array factor of this structure can be specified in terms of the element separation (d).

Considering the xy-plane of the array matrix illustrated (Figure 7.22), the position vector of any $(m, n)^{th}$ element with respect to the origin can be written as:

$$\mathbf{r}'_{mn} = \mathbf{x}'_{mn}\,x + \mathbf{y}'_{mn}\,y \qquad (7.4)$$

The angle (θ, ϕ) between the plane of the array and an incident ray (specified by a unit vector $\mathbf{r}$) can be written in terms of the following dot product:

$$\mathbf{r} \cdot \mathbf{r}'_{mn} = x'_{mn}\sin(\theta)\cos(\phi) + y'_{mn}\sin(\theta)\sin(\phi) \qquad (7.5)$$

The corresponding phase difference, as a result of element separation, is equal to $(2\pi/\lambda) \times \mathbf{r} \cdot \mathbf{r}'_{mn} = \xi_{mn}$ with λ being the wavelength of operation.

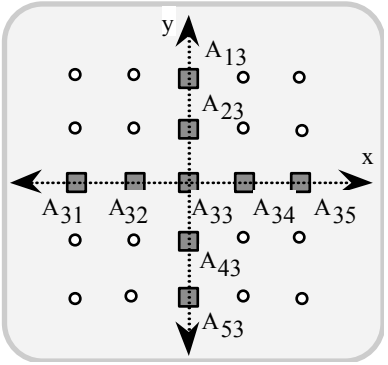


Fig. 7.22 Cross-array matrix for the smart-antenna realisation: An example of a 9-element array

The beam-steering is enabled by the appropriate amplitude and by phase-shifting the RF excitation of the elements. Suppose A_{mn} denotes the relative amplitude of the $(m, n)^{th}$ element and α_{mn} is the phase of the current feeding that element. Hence, the resulting array factor (AF) can be written as follows:

$$AF(\theta, \phi) = \sum_{n=1}^{N} \sum_{m=1}^{M} A_{mn} \exp(j\alpha_{mn}) \exp(j\xi_{mn}) \qquad (7.6)$$

where M = N = 5 depicting the size of the array matrix of Figure 7.22 and α_{mn} is given by:

$$\alpha_{mn} = -(2\pi/\lambda)[x'_{mn} \sin(\theta_o)\cos(\phi_o) + y'_{mn} \sin(\theta_o)\sin(\phi_o)] \qquad (7.7)$$

where (θ_O, ϕ_O) denote the angular co-ordinates of the main beam-pointing direction. For a symmetrical pencil beam, the centre element at the origin of the matrix is excited with $A_{33} = 1$ and its phase-shift $\theta_O = \phi_O = 0^o$; further, $A_{31} = A_{13} = A_{53} = A_{35}$; and, $A_{32} = A_{23} = A_{34} = A_{43}$. Hence, the beam can be steered by controlling (θ_O, ϕ_O) and the sidelobe level and beamwidth are dictated by the relative amplitude of the current level.

For Bluetooth applications, the sidelobe levels should be kept low in order to avoid interference and the beamwidth should be designed for pencil-beam so as to point precisely at the access-point device. In [7.25], the current amplitude distribution is adopted as a compromise between binomial and Dolph-Chebyshev (DC) distributions. This strategy offers a (– 37.5)-dB sidelobe level and the half-power beamwidth is same as that obtainable with Dolph-Chebyshev distributions.

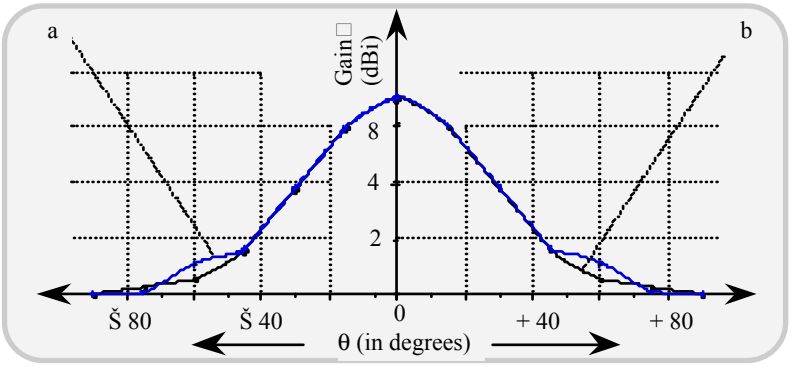


Fig. 7.23 Radiation pattern of 9-element cross-array for E- and H-planes (a: E-plane pattern and b: H-plane pattern)

Since the parallel feeding of the elements of the matrix is feasible, the phase scanning can be performed using a microprocessor as shown in Figure 7.24. Relevant smart phase-scanning implementation shown consists of a microprocessor-controlled bus, which applies control information to each element with appropriate current level (L) and phase (ϕ) through gain and phase shifters. A Wilkinson power combiner/divider network links the array matrix to the transreceiver at the RF front-end.

The phase-shift of each element can be controlled by a 16-bit word, the eight most significant bits of which can be used to select the state of each phase-shifter and variable gain amplifier. The bits of the address field decide the number of devices controlled.

For the 9-element array of Figure 7.24, for example, with 4-bit phase-shifters (with 180º, 90º, 45º and 22.5º phase-bits), a control network with relevant biasing can be designed as discussed in [7.25]. The use of a microprocessor makes the system adaptive with an intelligent algorithm digitally controlling the phase-shifter and variable gain amplifier exciting the beamforming array.

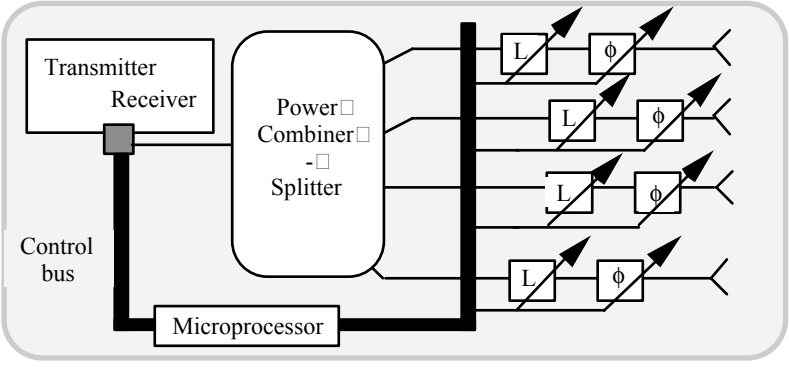


Fig. 7.24 Block diagram of the intelligent phase-scanning control. (L: Variable gain control; ϕ: Variable phase control)

Operational considerations

With reference to the smart array system described above, the array offers several time-slots to one slave, which is assumed to exist at a certain locale in the room. When the data transfer is accomplished, the master aims the beam to another direction where the next slave is located. This procedure is continuously repeated on a round-robin basis. An example of this is illustrated in Figures 7.25. Here, the true bearing to slave # 1 is 135^o from master and 030^o to slave # 2 respectively; and, the elevation to both slaves is 45^o. Figure. 7.25 shows a bird-eye view of the directions of the beams.

Figure 7.26 illustrates the usage of time-slots in the transmissions of packets. In the specific example given in [7.25], slave # 1 is using ACL packets that occupy 3 slots. Slave # 2 is using 1-slot packets. Master points towards slave # 1 through beamforming in the direction of 135^o during the time-slots 0-3 and 6-9 and 12-15 and so on. Likewise, during the time-slots 4-5, 10-11, 16-17 etc., the master point towards slave # 2 through beamforming in the direction of 30^o. Slaves in general, can be mobile and may change their position during the connection and master can adaptively alternate the direction of the beam. Due to fast frequency-hopping, the speed of such direction change could be high. However the slaves must remain within the range of the Bluetooth piconet, so that the distance cannot increase indefinitely. The range is constrained by the available link budget. It is limited by the prescribed (70-dBm) receiver-sensitivity and by the transmitter power of 0 and 20 dBm (corresponding to the ranges of 10 m and 1000 m respectively).

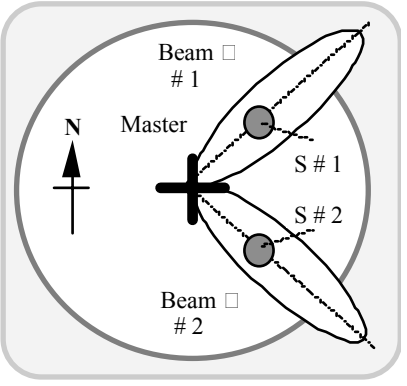


Fig. 7.25 Beamforming strategy in relation to master and slave dispositions

Thus, the antennas play a crucial role in the link budget considerations. This is illustrated by the following example.

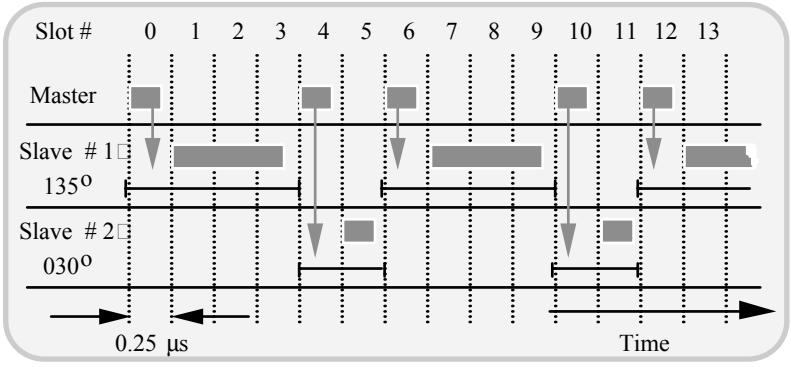


Fig. 7.26 Timing and beam-pointing direction schedule for Bluetooth packets

---

Example 7.1
Suppose a Bluetooth system is designed with an antenna with a gain of – 3dB yielding a range of 15m. Determine the enhanced range with a directional antenna (such as the matrix array discussed above) with a directional gain of 17 dB.

Solution

Consider the Friis formula for the received power, namely, $P_r = (\lambda/4\pi r)^2 G_t G_r P_t$ where r is the transreceive separation, G_t and G_r are transmitter and receiver antenna gains respectively and P_t is the transmitted power. Taking $G_t = G_r = G$, for the same transmitted power P_t, r will increase to 157 m with G = 17dB as compared to 15 m when G = – 3dB.

Problem 7.1
For the 9-element matrix illustrated in Figure 7.22, the proposed current amplitude distribution indicated in [7.25] is a compromise between binomial and Dolph-Chebyshev (DC) distributions. This mixed approach has been prescribed to realise a design value of sidelobe level equal to – 37.5dB (instead of – 30dB, which would have otherwise resulted, if the DC distribution alone was adopted).

The coefficients used in [7.25] are: $A_{31} = A_{13} = A_{53} = A_{35} = 0.25$; and $A_{32} = A_{23} = A_{34} = A_{43} = 0.75$, and A_{33} corresponding to adding a value of 1/12 to each amplitude to a normalised binomial distribution (except that of central element). Prove that these values conform to the design compromise adopted in [7.25].

7.6.9 Polarisation-switched antennas for indoor applications

One of the authors of this book has proposed elsewhere [7.26] a scheme of superimposing random polarisation switching of the frequency-hopped CDMA packets transmitted, for example, in Bluetooth systems. That is, in the conventional FH/CDMA, each hopped frequency supports a packet and the polarisation of EM wave transmitted is kept solely in one type of linear polarisation (either vertical, V or horizontal, H) for all the packets transmitted. Furthermore, no specification has been indicated and/or proposed whether the polarisation of individual hopped-frequency transmission should be consistently kept invariant or it could be changed.

Hence, in [7.26] a scheme it is proposed to randomly switch the polarisation (between V and H) in a Bernoulli fashion of the hopped frequency sequences, so as to realise the following advantages:

- By random switching of the polarisations (between V and H) of the FH packet transmissions, an inherent polarisation diversity is infused. As a result, a maximum of 3-dB improvement in SIR can be expected
- Further, a polarisation (V or H) can be selectively chosen over a certain duration so as to achieve improved RFI/EMI immunity. This can be illustrated by considering the following example: As described in [7.26, 7.27], Bluetooth links are often impaired by the unintentional leakage of 2.4 GHz EM energy from an ISM device like a microwave oven in the vicinity. It is shown in [7.27] that such an EMI/RFI from the microwave oven is largely vertically polarised. As such, it is proposed in [7.26] that by making a microwave oven Bluetooth-enabled, it can play the role of a master in the piconet and it can instruct the other Bluetooth (slave) links to perform the following: The command will be to switch slaves' polarisation of transmissions to horizontal, so that these transmissions will be minimally offended by the vertically polarised (unintentional) leakage from the oven. Once the oven is turned off, the master (oven) can instruct the other Bluetooth devices to revert back to either vertical or randomly-switched V and H polarisation.

A detailed account of such polarisation switching *vis-à-vis* Bluetooth system protocol and operation is presented below. As mentioned before, it should be noted that the prevailing Bluetooth specifications place no constraint on the polarisation of the FH packets transmitted/received at the physical layer level.

The basis for the method under discussion is that the polarisation switching could reduce the effects of fading and other propagation impairments prevailing in indoor environments. Theoretically, it can be expected that an improved performance versus EMI environment in the CDMA transmissions could be of 3 dB, when the two orthogonal polarisations are adopted randomly. This results from the heuristics of proneness of EM propagation otherwise set on a single polarisation.

This polarisation control can be applied on Bluetooth links to defeat the EMI/RFI effects. That is, the Bluetooth-enabled microwave oven can perform a polarisation control instead of power-control so as to operate under a desirable SIR following the schedule described below:

- Whenever the microwave oven is not turned on, the Bluetooth transmissions in the vicinity could impose a random polarisation-switching on the hopped-frequency events with a symmetric Bernoulli statistics (50% vertical polarisation and 50% horizontal polarisation). This would improve the transreceive performance of CDMA transmissions, which could otherwise suffer more propagation-related impairments in the indoor environments than the Bluetooth operation with a single polarisation.
- Whenever the microwave-oven is turned on (posing, unintentional leakage of microwave radiation prevailing in the surrounding), the Bluetooth communications in the vicinity could resort to an asymmetric Bernoulli process of random polarisation-switching imposed on their hopped-frequency transmissions. The asymmetry proposed refers to adopting largely horizontal polarisations (say, 90%) and the other 10% being vertical polarisation. As mentioned before, this predominance of horizontal polarisation imposed on the Bluetooth transmissions will prescribe diversity against the EM waves of the microwave leakage from the oven (which are significantly vertically polarised, as stated earlier).

Now, the question is how to implement the proposed scheme. It involves three major technological considerations, as indicated below:

- A hardware requirement to facilitate orthogonal polarisations on the hopped-frequency EM waves of the Bluetooth transreceivers
- Compatible antenna structures on the transreceivers to allow the required (-switched) polarisations
- A scheme to alert the Bluetooth units to impose predominant horizontal polarisations on their transmissions, whenever there is a microwave oven or a similar interference source operating in the vicinity.

The above first and the second requirements could be essentially implemented by resorting to the relevant concept of diode-switching facilitated on a pair of orthogonal antenna structures. That is, the antennas when switched, could radiate (or receive) vertically or horizontally polarised waves; such a switching can be randomly done with triggers applied through steering diodes attached to the orthogonal antenna structures. (Alternatively, the transreceive antennas could be

configured as circularly-polarised structures, so that they have the capability to transmit or receive either vertically or horizontally polarised waves.)

The third design consideration of alerting the Bluetooth units to assume transmissions mostly in horizontally polarised plane can be done as follows: First, the microwave oven can be made Bluetooth-enabled. This would allow the oven whenever switched on, to function as a master and alert the other Bluetooth units (slaves) in the vicinity to resort to sending (their) transmissions predominantly in the horizontal polarisation via appropriate steering through diode-triggering, done at the antenna structure. This method of adopting a Bluetooth-enabled microwave oven is similar to the technique proposed in [7.27] except that in the present case the "master" (oven) communicates to "slaves" to control their polarisation of transmissions. This is different from the scheme of [7.27], in the sense that in [7.27], the relevant command refers to a power control (and not a polarisation control). In other words, in [7.27] the Bluetooth units are "advised" by the (Bluetooth–enabled) oven to increase their carrier power to achieve a higher carrier-to-interference ratio, so as to mitigate the EMI arising from the unintentional microwave leakage from the oven. The scheme of a polarisation control (in lieu of the power-control strategy of [7.27]) could be justified in terms of the relevant heuristics presented below.

As indicated earlier, the Bluetooth communications are specified to use frequency-hopped spread-spectrum transmissions. Further, the Bluetooth radio chip uses the GFSK modulation (with a modulation index between 0.28 and 0.35 in order to balance the effect of in-band interference between bit "0" and bit "1") and a WT_B factor equal to 0.5 on the Gaussian filter (where WT_B is the *time-bandwidth product* of the filter-response). The time-domain response g(t) of the Gaussian filter can be specified by a transfer function, which in reference to a rectangular pulse of unit amplitude and duration T_B (centred at the origin) at the input, can be written as:

$$g(t) = \frac{1}{2}\left[\operatorname{erfc}\left(\pi\sqrt{\frac{2}{\log_e(2)}}\,WT_B\left(\frac{t}{T_B}-\frac{1}{2}\right)\right) - \operatorname{erfc}\left(\pi\sqrt{\frac{2}{\log_e(2)}}\,WT_B\left(\frac{t}{T_B}+\frac{1}{2}\right)\right)\right] \qquad (7.8)$$

In essence, this pulse response of the filter enables a trimming of the spectrum of the rectangular pulse at the GMSK modulator. Consequently, the waveform of the pulse can be specified in terms of the (dimensionless) time-bandwidth product WT_B which plays the role of a design parameter.

A rectangular pulse passing through a Gaussian filter would spread out of the bit duration time (namely, $-T_B/2$ to $+T_B/2$), resulting in a *degradation factor* (ν), which can be defined as follows: It is the ratio between the pulse energy at the output of a Gaussian filter (over the interval $-T_B/2$ to $+T_B/2$) and the energy of the rectangular pulse representing the original bit at the input. That is,

$$\nu = (1/N_o) \int_{-T_B/2}^{+T_B/2} [g(t)\, dt] / [E_b/N_o] \qquad (7.9)$$

The corresponding bit error probability or bit error rate (BER) P_e in reference to the GFSK scheme, can be written in a general form (to specify the BER) as follows [7.26]:

$$P_e = \text{BER}$$
$$= Q(a,b) - (1/2)\exp[-(a^2 + b^2)/2] \times I_o(ab) \qquad (7.10)$$

where

$$a = \{[(\nu E_b/N_o)/2][1 - (1 - |\rho|^2)^{1/2}]\}^{1/2}$$
$$b = \{[(\nu E_b/N_o)/2][1 + (1 - |\rho|^2)^{1/2}]\}^{1/2}$$
$$|\rho| = |\sin(\pi t \Delta f)/(\pi t \Delta f)| = |\sin(\pi h)/(\pi h)|$$

with h representing the modulation index; further, Q(a,b) is the Marcum's Q-function; $I_o(a, b)$ is the modified Bessel function of order zero; and, E_b/N_o is the signal-energy to noise-power ratio.

Using the expression for the BER of equation (7.10), the frame error rate (FER) in ACL packets of Bluetooth transmissions can be ascertained for different formats of the packet structures indicated in Appendix 7.1.

In reference to AUX1 packet (described in Appendix 7.1), there are no FEC and CRC. It means that no error correction and/or detection is facilitated. Therefore, the whole packet may fail whenever a single bit error occurs in the packet. For DH1, DH3, and DH5 packets, a set of 16 CRC bits are included per packet. But these CRC bits would enable only an error-checking (through the ARQ scheme), and they do not offer any error correction. Hence, both AUX as well as DH packets do not have the capability for error corrections. Therefore, such packets will be discarded if a single bit error occurs, and the corresponding frame error probability (FER) can be written as:

$$\text{FER}_{\text{GFSK}}(\text{AUX1/DH packet}) = 1 - (1 - P_e)^n \qquad (7.11)$$

where n is the number of bits in a AUX1 or a DH packet

Considering a DM packet, the information payload plus the CRC bits are encoded with a rate 2/3 FEC, which adds 5 parity bits to every 10-bit segment of the codeword as illustrated in Figure 7.27. This code can correct all the single bit errors in each codeword. Since the encoder operates with information segments of length 10, a set of tail bits (with value zero) may have to be appended after the CRC bit. Further, the total number of bits to be encoded (that is, payload header, user data, CRC, and tail bits), must be a multiple of 10.

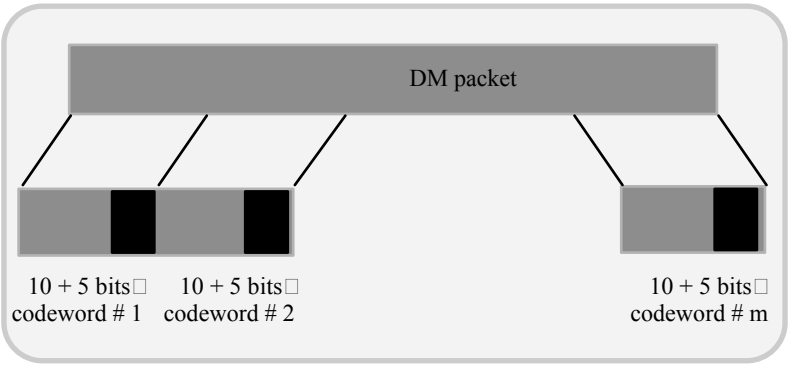


Fig. 7.27 The codeword section in a DM packet. (Each codeword is made of 15 bits)

The probability of error in the DM packets (with each codeword having the capability to correct the single bit errors), is given by:

$$P_c = 1 - \sum_{k=0}^{1} \binom{15}{k} P_e^k (1 - P_e)^{15-k} \tag{7.12}$$

where P_e is the probability of bit error rate, P_c is the probability of codeword error rate. The corresponding probability of frame error rate is given by:

$$FER_{GFSK}(\text{DM packet}) = 1 - (1 - P_c)^m \tag{7.13}$$

with m being the number of codewords in a packet.

Now, the FER_{GFSK} performance of Bluetooth communications in the presence of unintentional microwave leakage from an oven can be considered: As discussed by Neelakanta and Sivaraks in [7.26] and elaborated in [7.27], the causative mechanism behind the relevant FER_{GFSK} is as follows: A packet under transmission may fall into a hopped-frequency range coinciding in the duration in which the microwave oven leakage may prevail. This situation is decided by the considerations relevant to the E_b/N_o ratio of the GFSK discussed above.

Suppose a random polarisation-switching is imposed on the frequency-hopped transmissions of Bluetooth, as proposed. Since the microwave leakage is predominantly of vertical polarisation, in the event of a coincidence of the spectral window of microwave leakage and a Bluetooth hopped-frequency spectrum takes place, a frame error could be anticipated only when the polarisations of both electromagnetic waves (of microwave leakage and Bluetooth transmission) are identical. If a brute-forced, polarisation-switching is adopted for the Bluetooth transmissions (so as to be different from microwave oven leakage), the resulting FER will be reduced by a weighting factor (< 1) achieved through this polarisation

diversity. Hence, the resultant FER of Bluetooth transmissions can be specified as $FER_{GFSK} \times FER_P$ where FER_p is the polarisation-imposed weighting factor.

For the proposed scheme of using orthogonal polarisation-switching superimposed on the frequency-hopped CDMA packet transmissions, the following considerations on implementation strategy are suggested, commensurate with the existing scope of Bluetooth specifications.

If the proposed polarisation-switching is to be designed, its major circuit (to control the polarisation-switching) should be consistent with the Bluetooth baseband/radio specifications. Accordingly, the following two feasible approaches are indicated:

- *Synchronous polarisation-switching*: In this case, the transmitter and the receiver units will be hopped in frequency and switched to identical polarisations synchronously. This scheme can be improvised by considering the output of the frequency-hop selection process and extending it through an algorithm to include the polarisation-switching (selection).

  For example, the vertical polarisation can be applied whenever an odd-number designated, random frequency channel hopping occurs; horizontal polarisation can be applied whenever an even-number designated, randomly-hopped, frequency channel is used. The existing frequency-hopping scheme of Bluetooth technology is illustrated in Figure 7.28 (a) while the modified scheme to include polarisation-switching is depicted in Figure 7.28(b).

  A block diagram of a Bluetooth radio using frequency-hopping superimposed with polarisation-switching and GFSK modulation is illustrated in Figure 7.29.

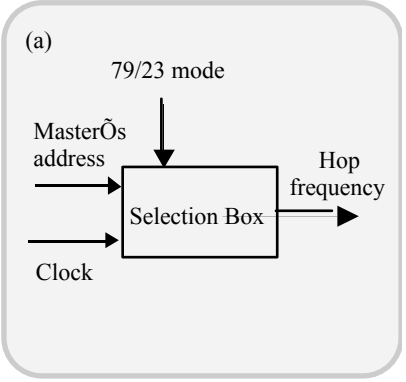


Fig. 7.28(a) Existing frequency hop selection scheme based on Bluetooth specification V1.0b

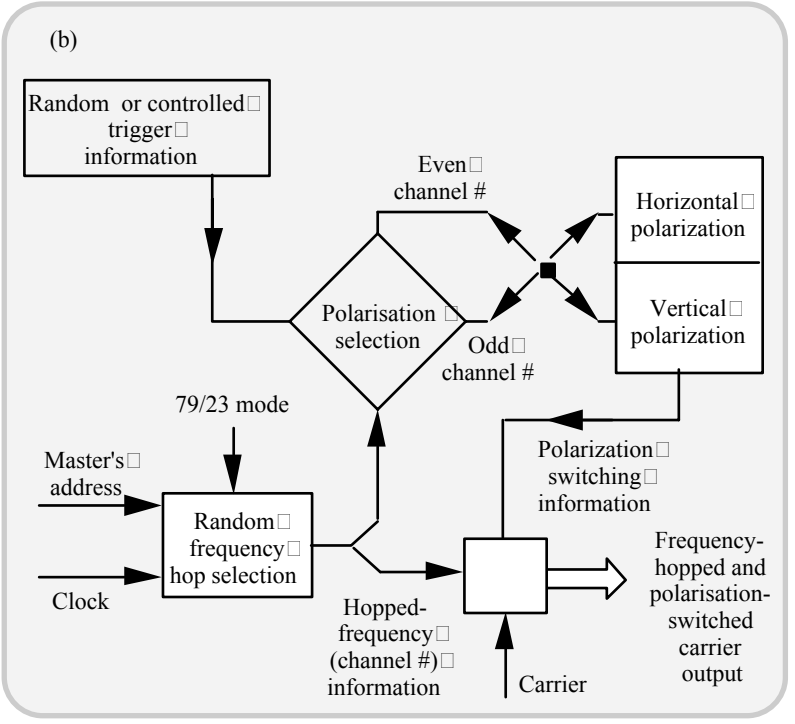


Fig. 7.28(b) Proposed (modified) hop selection scheme superimposed with polarisation-switching

- *Unsynchronous polarisation-switching*: In this scheme, only the transmitter unit will perform the added polarisation-switching (as per some Bernoulli random process schedule). The receiver unit would then be facilitated with either an orthogonal set of antennas or a circularly polarised antenna. The polarisation diversity will dictate a decision unit to choose appropriate polarisation of the received signal in order to get the maximum carrier-to-interference ratio.

Common to both schemes described above, the polarisation selection should be done on packet-by-packet basis (similar to the frequency-hop selection). Further, as described earlier, a strategy to overcome the influence of EMI due to unintentional leakage from a microwave oven is to make the oven itself Bluetooth-enabled. This technique will enable the following: Whenever the oven is on, it would alert the nearby (other) Bluetooth units to increase their power levels of transmissions [7.27]. That is, the microwave oven would transmit a "warning signal" to other Bluetooth devices in the vicinity to increase their power level so as to alleviate the

interference from the leakage of EM energy from the oven. However, a scheme of carrier power enhancement, as illustrated in Figure 7.30 may cause interference at other Bluetooth devices in the adjacent piconets (as a result of the enhanced carrier level envisaged).

On the contrary, the polarisation-switching by a command from a master (such as a Bluetooth-enabled microwave oven) would not cause interference effects on adjacent piconets.

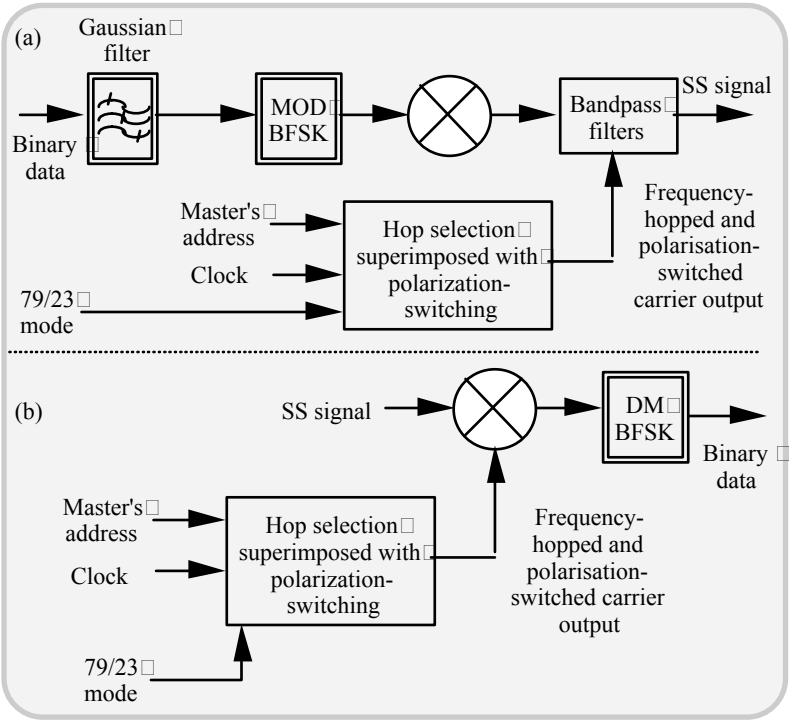


Fig. 7.29 Bluetooth units with polarisation-switching plus frequency hopping of packet transmission capability. (a) Transmitter and (b) receiver

Simulation studies performed and reported in [7.26] ascertain the resultant FER in typical Bluetooth systems improvised with the polarisation-switching proposal indicated above. The chosen example refers to microwave spectral windows of EM energy leakage from a microwave oven that could possibly interfere with Bluetooth transmissions in the vicinity. Relevant FERP computed with random polarisation-switching (as per a Bernoulli process) addresses two statistical scenarios. Case 1: This refers to a symmetric Bernoulli process in which the random imposition of vertical and horizontal polarisation are proportioned as

50% - 50%; and, in Case 2, the horizontal polarisation is made dominant (90%) against the vertical polarisation (10%).

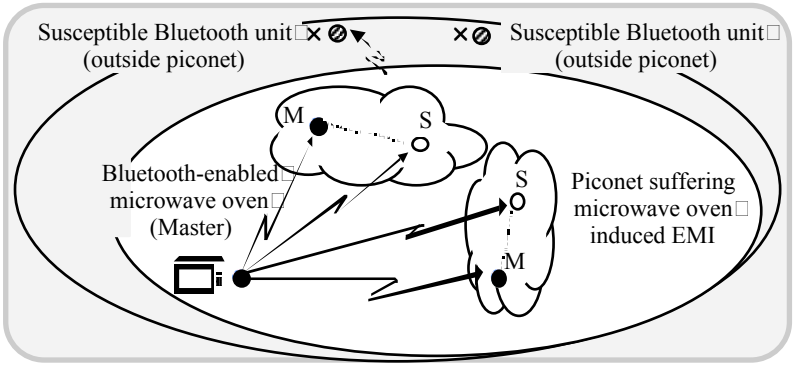


Fig. 7.30 Possibility of Bluetooth units becoming susceptible to interference from the adjacent piconets facilitated with higher carrier levels (so as to mitigate the microwave oven-dictated EMI as per [7.27]).

The simulated results gathered and indicated in [7.26], reveal a reduction in the resultant FER of Bluetooth packet transmissions (especially at low E_b/N_o) as a result of the random polarisation-switching adopted. Thus, the scheme of random polarisation-switching advocated on frequency-hopped packet transmissions of the Bluetooth offers a viable strategy to mitigate the invasion of EMI/RFI from devices like microwave ovens.

Further, this polarisation-switching method can be regarded as better than the power-control strategy proposed by Neelakanta and Sivaraks in [7.26]. It not only saves the Bluetooth transmissions from external RFI, but also renders these transmissions to cast minimal influence on adjacent piconets (unlike the power control strategy). It also takes advantage of polarisation-diversity specified performance improvements of indoor wireless transmissions.

The scheme described above can be further supplemented with circularly-polarised passive reflector deployments in indoor environments, as suggested in [7.28]. In general, the underlying concept can be extended to any CDMA transmissions (either hopped-frequency or direct-sequence version) and can be adopted for both outdoor cellular and/or other indoor wireless technology.

7.6.10 Implementation of switched-polarisation antenna system

To implement the polarisation-switching scheme in the technique described above, there should be two orthogonal (V and H), linearly-polarised antennas; steering of the signal between them (through a random algorithm such as the Bernoulli process) can be done *via* steering diodes. A relevant outline is described below:

A pair of antennas, each set orthogonal to the other, can radiate/receive in both vertical and horizontal planes. For example, a simple pair of CPW antennas

described in Chapter 4 can be adopted. Each antenna should be an independent part to which the signal can be directed through a feed arrangement. When the signal is steered to an appropriate antenna, the EM wave supporting that signal will be radiated in a polarisation set by that antenna. Such steering of signal to V and H antennas can be done via a pair of (steering) diodes. For example, beam-lead PIN diodes can be used for this purpose. They can be introduced at a common intersection of the CPW antenna. Appropriate biasing is needed to let the steered diode to go into ON state. Also necessary impedance corrections may be needed to counter the reactive impedance of ON/OFF state biased diodes [7.29].

Problem 7.2
Design a concept configuration of polarisation-switching on a pair of folded slot radiators described earlier (Figure 7.17). Indicate a method of facilitating the steering diodes at the feeding point with necessary bias arrangements and matching requirements. Use, for example, MACOM 4P461 beam-lead diodes.

(*Hint*: see [7.29])

7.6.11 Circularly-polarised antennas for FH/CDMA based indoor wireless communications

In the indoor radio network applications, it has been observed [7.30] that circularly-polarised transmissions in transreceive links can significantly mitigate the multipath fading caused by reflections from walls, ceiling, and floor. This results from energies of single-bounce reflected waves do not appear at the receivers, thereby causing significant amplitude fade and delay-spread reduction. Hence, use of circular polarisation is suggested as an antimultipath-fading scheme in LoS indoor radio communications. Relevant conclusions, summarised in [7.30], are as follows:

- Amplitude fade level could be marginally reduced by 7-11 dB with circular polarisation in comparison to using linear polarisation
- Circular polarisation transreception also shows better polarisation diversity performance
- With circular polarisation, delay-spread is reduced to (approximately) half its value. This reduction could increase further with larger transreceive separations.

The antimultipath-fading scheme via circular polarisation can be profitably used in Bluetooth and/or WLAN implementations. Relevant practical implementation of circularly-polarised transmissions in indoor wireless applications is as follows. Suppose the transreceive system of the indoor wireless operation (such as the

Bluetooth) envisages the use of circularly-polarised antennas; and, a switching scheme is implemented such that the frequency-hopped packets are sent on randomly specified (Bernoulli process) right-handed (RHCP) or left-handed (LHCP) circular polarisations. Implementation of this scheme would lead to the following desirable performance aspects:

- Use of circular polarisation in general offers antimultipath-fading on EM transmissions as indicated earlier.
- When randomly switched between LHCP and RHCP on FH/CDMA, the packets would provide a maximum of 3dB additional improvement to SIR.

The practical way of switching the polarisation (between LHCP and RHCP) using steering diodes can be based on a scheme described in [7.29]. Relevant details are as follows: A circularly-polarised antenna with switchable polarisation-sense can be deigned with a basic structure of a rectangular patch antenna excited by a diagonal slot fed by a microstrip line (Figure 7.31).

The generated mode TM_{10} and TM_{01} are linear orthogonal modes resonant at separate frequencies. To obtain a circular polarisation (RHCP in Figure 7.31), they are excited in equal amplitude and 90° out of phase at the operation frequency (f_o) by adjusting the aspect ratio L_p/W_p. The excitation process is similar to the cross-shaped aperture proposed in [7.29], where the 3-dB fractional bandwidth of the axial ratio indicated is 2.5%. A high permittivity substrate ($\varepsilon_{r1} = 9.8$) is used for the feeding line to optimise the coupling with the slot. The stub length L_s is adjusted to optimise the impedance bandwidth. In Figure 7.31, the axial ratio (AR) and the return loss appear in defining the CP bandwidth. This CP-specific bandwidth explicitly refers to the frequency band where AR > 0.7, that is, AR (in dB) > – 3 dB, at broadside. A 5% CP bandwidth is obtained around f_o with a return loss smaller than – 13 dB in the overall frequency band. The measured gain is 6 dB.

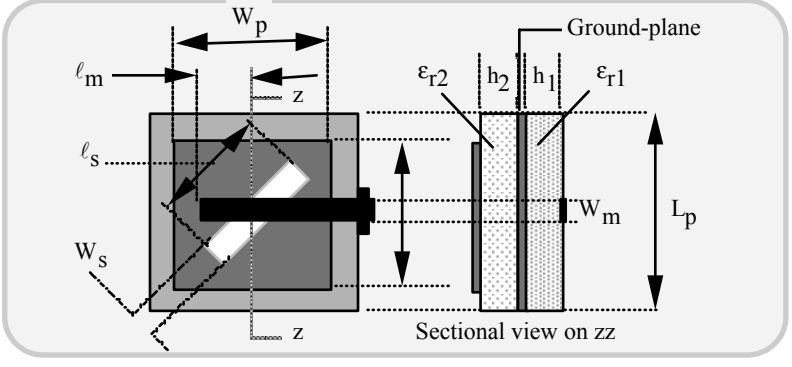


Fig. 7.31 Configuration of RHCP patch antenna excited by single diagonal slot. (W_p =13.1 mm, L_p = 15.0 mm, L_s = 13.4 mm, W_s = 0.3 mm, L_m = 4.8 mm, W_m = 0.6 mm)

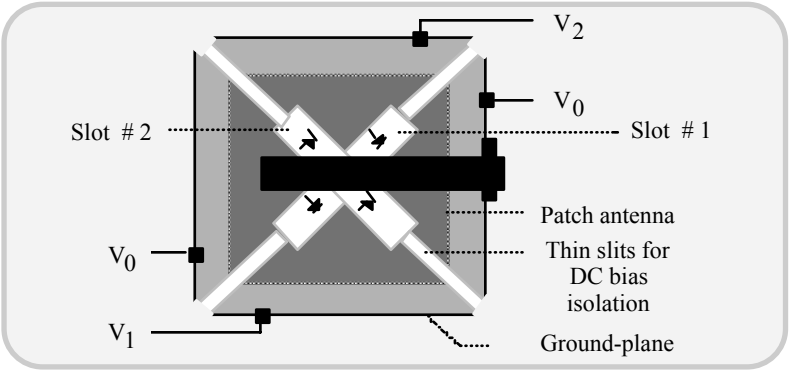


Fig. 7.32 Circularly-polarised patch antenna with switchable polarisation sense. Slot 1 is ON and Slot 2 is OFF for $V_1 = V_o$ and $V_2 = 0$ (RHCP); Slot 1 is OFF and Slot 2 is ON for $V_1 = 0$ and $V_2 = V_o$ (RHCP). (W_p = 13.5 mm, L_p = 18.0 mm, L_s = 13.4 mm, W_s = 0.3 mm, L_m = 4.6 mm, W_m = 0.6 mm)

In an improved structure, an identical slot can be added along the other diagonal with two short circuits included near the intersection with the first slot. The purpose of this antenna is to emphasise the effect of an added short-circuited slot on the CP performance, that is, to control the effectiveness of the final switchable antenna including two slots and pin diodes (Figure 7.32). The short circuits are made with 0.1 mm-wide strips, which model ideal forward biased pin diodes (ON-state), located at 3 mm from the slot intersection. Provided that the lengths of the antenna and the slot were tuned (L_p = 16.9 mm, L_s = 15.4 mm), a 4% CP bandwidth is kept with good return-loss performances. Nevertheless, the addition of a second slot yields a 300-MHz drop in the central CP frequency.

7.6.12 Circularly-polarised patch antenna with switchable polarisation sense using PIN diode switching

Considering a practical switchable antenna as depicted in Figure 7.32, the pairs of beam-lead pin diodes (such as MACOM4P461) are inserted in each coupling slot at 3 mm from the intersection. By switching ON a pair of diodes while the other pair being OFF, one can expect to select either sense of CP. The bias isolation is facilitated by thin slits (100 μm) covered by metal-insulator-metal capacitors that separate the ground-plane into four regions. Each region is DC biased, as shown. In Figure 7.32, the RHCP is selected by the states ($V_1 = V_o$ = 10 V, V_2 = 0 V), while the LHCP is obtained for: V_1 = 0 V, $V_2 = V_o$ = 10 V. The forward current in the PIN diodes is fixed at 10 mA by a biasing resistor of 1000 ohm. Since PIN diodes may not be always ideal and could present a reactive impedance for both bias states, the patch length should be adjusted as necessary (to L_p = 18.0 mm) in order to obtain CP in a 5% bandwidth for both polarisation senses around 5 GHz. Owing

to fabrication tolerances, the CP bandwidths may not be exactly centred at the same frequency and the CP purity could become lower (with AR < 0.8) as compared to the passive structures. Nevertheless, a common 4% CP bandwidth could be obtained with good matching ($S_{11} < -10$dB) on the overall band. Typical measured gain is 5.5 dB.

The aforesaid scheme can further be improved by adopting a polarisation diversity with passive reflectors in indoor radio channels, as proposed by Kajiwara [7.28]. It is shown in [7.28] that the use of passive reflectors could reduce the outage probability of an indoor radio channel experienced at an access point or an indoor base unit. A relevant concept can be combined with the LHCP/RHCP switched transreceive system described above. Such a system can provide for diversity gain through the improvised passive reflectors.

Problem 7.3

Develop a predesign proposal on adopting circularly-polarised antenna schemes in an infrastructure WLAN consisting of a clustered set of PCs situated in two floors of a building and strategically located mobile support stations. Assume that the wireless communications are FHSS-based. The design should elaborate improvising the following:

Use of circular polarisation antennas with LHCP/RHCP switching capability.

Random switching of polarisation (between LHCP and RHCP) should correspond to a Bernoulli process, synchronously implemented at the transmitter and receiver.

Include passive reflectors in the room to reduce the outage probability at the access points. (Suggest numerical values on a hypothetical system consistent with WLAN specifications)

7.6.13 Smart antenna for high capacity indoor wireless systems

Multiuser wireless communication systems are prone to interference resulting from omnidirectional transmission and reception of RF signals. A method of improving the performance through better quality of service and enhance traffic capacity at an affordable cost is to use a smart antenna system. In such systems, the use of multiple antenna patterns to point at user location and launch beam-nulls in the direction of interference specific to an indoor environment requires meeting challenges posed by intense propagation losses and multipath fading especially in millimetre wave band systems. The key components of design options in such systems are:

- Range extension
- Improved immunity to flat-fading
- Interference suppression
- Spatial division multiple access

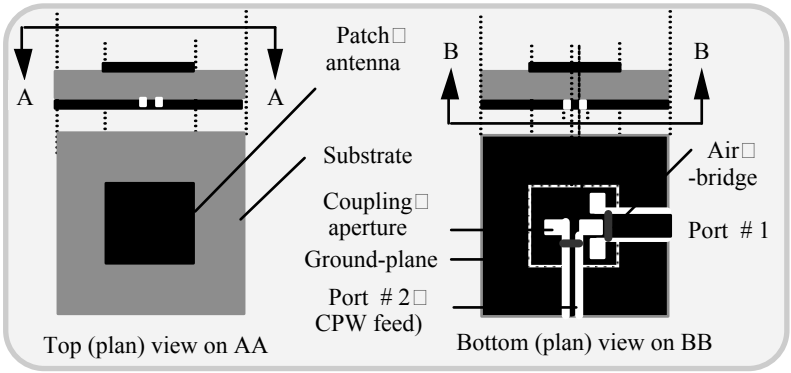


Fig. 7.33 Planar CPW-coupled patch antenna

An appropriate array structure, compact in size, can be realised with planar antennas fed by electromagnetically-coupled CPW on single layer substrate for exciting dual polarisation. This configuration uses a cross-slot coupling a patch antenna. That is, a rectangular patch antenna is one side of the substrate, while a cross-slot fed by a coplanar line is arranged opposite to the patch in the ground-plane on the other side of the substrate. The building block of the array, namely the radiation plus feed system indicated above, is shown in Figure 7.33.

It is indicated in [7.31] that a typical patch radiator made for 35 GHz operation with a dielectric of relative permittivity of 9.9 and thickness 0.254 mm offers a good matching with $S_{11} = -24$dB and $S_{22} = -11$dB (at 35 GHz) and a bandwidth of 700 MHz for VSWR ≤ 2. Dual polarisation and compactness of this antenna makes it a viable candidate for smart antenna array designs.

A typical CPW feeding of two elements is illustrated in Figure 7.34.

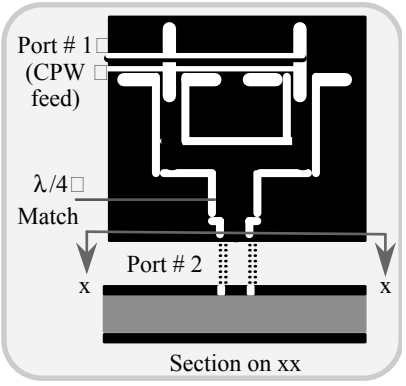


Fig. 7.34 CPW feeding arrangement for two elements

7.6.14 Smart indoor antenna for PCS receivers

To increase the capacity and improve the QoS considerations, *space-division multiple access* (SDMA) has been proposed for wireless communications. In an optimal SDMA implementation, the smart antenna design refers to a multiport antenna system with substantial port-to-port isolation and a digital signal processor supporting an adaptive algorithm. The multiport antenna can be integrated with transreceive units of indoor applications as described in [7.32].

The implementation of a smart antenna in SDMA environments generally consists of using a number of highly isolated antenna elements (for transmit and receive purposes) with independent fading characteristics. In the design proposed in [7.32], a set of elements constitutes a multiport antenna, and each element is subjected to both polarisation and space diversity schemes. Thereby, the antenna elements can be brought to close proximity (contrary to the general constraint of isolation imposed in SMDA strategies) within a radius of $\lambda/2$ at the design frequency. Hence, a significant reduction in size can be realised while maintaining low mutual coupling.

The antenna indicated in [7.32] uses multiple corner reflectors to divide an antenna array into a number of sectors as illustrated in Figure 7.35. The corner reflectors provide a shield for antenna elements in one sector from being strongly influenced by elements in other sectors while maintaining a compact antenna structure. With the utilisation of these reflectors, the antenna array is capable of producing multiple independent beams to individually receive uncorrelated signals arriving from different directions, and whereby pattern diversity is accomplished.

To further increase the number of antenna elements contained in the system, polarisation diversity is applied to isolated sectors of the antenna structure. Two orthogonally-polarised antenna elements are located closely together in each sector of the antenna. With orthogonal polarisations, the elemental antennas in each sector are able to generate orthogonally-polarised beams that are aiming at the same direction without coupling to each other. As a result, the antenna can radiate or receive multiple signals having independent fading and different planes of polarisation.

In SDMA schemes the conventional beamforming antenna, used is usually a linear or two-dimensional array of radiating elements, and requires a minimum element spacing of half-wavelength to attain substantial isolation between elements. The antenna design proposed in [7.32] overcomes this major problem of the conventional array by allowing elemental antennas to be in close proximity while maintaining the low mutual coupling. Further, a dual-polarisation, multidirectional antenna system possessing the ability to radiate or receive RF energy with various planes of polarisation in different directions is facilitated.

The multiport test antenna described in [7.32] consists of twelve elemental antennas mounted upon a round ground-plane conductor, six reflecting planes, each having a shape of one quarter of a circle and are radially disposed about the centre of the ground-plane conductor. Both the ground-plane conductor and the reflecting planes have radii of half a wavelength at the operating frequency of 1.7 GHz. The entire antenna structure thus has the shape of a hemisphere with six 60^{o} sectors.

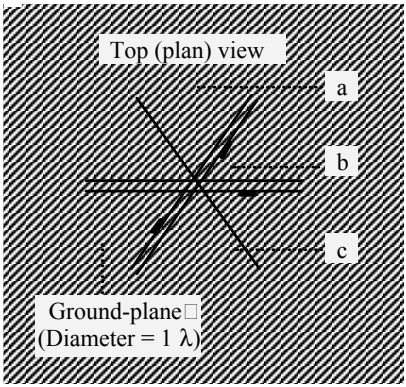


Fig. 7.35 Multiport antenna: (a) Monopole with top-hat and ground capacitive-loading and fed *via* ground-plane; (b) horizontal central-fed loop (patch) supported mid-way above the ground-plane and fed through a transmission line soldered on one face of the reflecting surface and (c) 60^{o} corner reflecting planar surfaces

The six reflecting planes act as corner reflectors for the twelve radiating elements in corresponding sectors and provide a shield for the elements in one sector from being substantially affected by those in other sectors. Each identical sector contains two types of radiating elements mounted adjacent to the corner of the reflector. The first elemental antenna is responsive to RF (radio frequency) energy having horizontal polarisation, while the second elemental antenna is responsive to RF energy having vertical polarisation.

The first elemental antenna in each sector comprises a horizontal centre-fed loop type patch antenna mounted on the 60^{o}-corner reflector. The loop antenna is supported midway above the ground-plane and coupled to a RF feed by a transmission line soldered on one of the reflecting planes. To implement this configuration, an L-shaped microstrip line, with a gap on the ground-plane side of the line, is inversely mounted on each corner reflector. Consequently, all the RF feeds for the loop antennas are placed on the bottom side of the ground plane to make the installation of the entire antenna structure easier.

The electrically small loop antenna in this connection between the two reflecting planes has a low radiation resistance as well as a series inductance. This combination of impedance components can be matched to 50 ohms with a combination of the gap size adjustment and the overlap length adjustment. The gap size of the loop antenna controls a shunt capacitive susceptance, whereas the overlap length of the conductor strip of the loop controls a series capacitive reactance. Hence, the gap size and the overlap length are adjusted to provide approximately 50-ohm input impedance with zero reactance at the operating frequency.

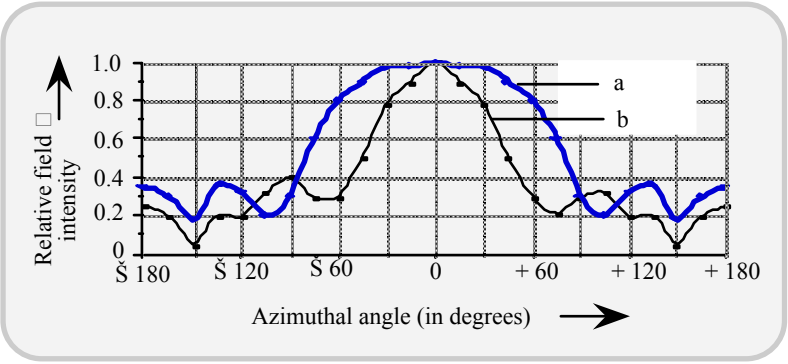


Fig. 7.36 Radiation characteristics: (a) Pattern of the loop and (b) pattern of the monopole

The loop antenna is fed by the centre gap to yield a polarisation that is completely horizontal and not coupled to the vertical antenna element. The second elemental antenna in each sector comprises a vertical monopole antenna disposed on top of the ground plane conductor. The monopole antenna is positioned a distance from the loop antenna and coupled to a RF feed located on the bottom side of the ground plane. From input impedance considerations, an electrically small electric monopole of this type may be treated as a series resistance, a large capacitive reactance and a small inductive reactance.

The series resistance of the monopole is smaller than 50 ohms and varies approximately as the square of the operating frequency. By placing a capacitive hat or plate on top of the monopole, the resistance can be raised and the series capacitive reactance can be reduced so that the inductive reactance will dominate. A capacitive component can now be placed at the base of the monopole to perform the well-known L match. This will raise the input resistance to approximately 50 ohms and tune out the inductive reactance of the top loaded monopole at the operating frequency.

In some applications, it may be desirable to have a large back lobe for both elemental antennas so as to enhance the antennas capability of transmitting or receiving signals in different directions. This can be achieved by simply lowering the height of each corner reflector, and therefore the height of the entire antenna structure. However, there is a trade off between the size of the back lobe generated and the elemental antenna isolation. The loop of the test antenna described in [7.32] has a return-loss as low as –27dB at 1.7 GHz and its impedance bandwidths at 3dB and 10dB return loss points are in excess of 500 to 200 MHz respectively. It is further reported in [7.32] that the associated monopole also exhibits low return-loss (less than –28 dB at 1.7 GHz) with 3 dB and 10 dB return loss bandwidths being 500 MHz and 200 MHz respectively. The aforesaid low return-loss considerations allow the design for an easy input match for 50-ohm standards.

The loop antennas used have individual directional and horizontal patterns limited by the associated corner reflector. The monopole's azimuthal pattern is also directional; but narrower than that of the loop. The side and backlobes of both loop and the monopole are reported to be small. Further, the loop and the monopole in the same sector pose a good isolation (33.6 dB at 1.7 GHz); adjacent sector isolation averages to 17.3 dB at 1.7GHz. The radiation patterns of the test antenna under discussion are presented in Figure 7.36.

7.7 CONCLUDING REMARKS

Perhaps, it is for the first time in the history of antennas that EM radiating structures are so widely deployed in indoor environments, thanks to the developments in wireless communications. Adjunct to voice communication, the data and broadband transmissions have joined the list of such indoor RF transreceive systems. Relevant antennas being conceived, designed and deployed are adding to the existing list. New configurations are suggested and tested. Yet, the basic concepts of the elements adapted and/or extended into arrays have not materially altered. It is left to the imagination of antenna design engineers to come up with a structure that can find a foreseeable application in the indoor wireless systems. Thus, there exists an immense scope to ponder over many possible designs for this sector of antennas.

For example, the broadband applications alone may warrant specific design requirements in conceiving appropriate structures. Hence, presented in the next chapter (Chapter 8) is a variety of descriptions on broadband antennas for wireless communications including that of indoor systems.

REFERENCES

[7.1] Bertoni, H. L., *Radio Propagation for Wireless Communications*, Upper Saddle River, NJ: Prentice Hall PTR, 2000

[7.2] Rappaport, T. S., *Wireless Communications*: *Principles and Practice*, Upper Saddle River, NJ: Prentice Hall PTR, 1996

[7.3] Saleh, A. A. M., and R. A. Valenzuela, A statistical model for indoor multipath propagation, *IEEE J. Select. Areas Commun.,* February 1987, vol. SAC-5(2), 128-137

[7.4] Rappaport, T. S., Characterization of UHF multipath radio channels in factory buildings,, *IEEE Trans. Antennas Propagat.,* August 1989, vol. 37(8), 1058-1069

[7.5] Sedel, S. Y., and T. S. Rappaport, 914 MHz path loss prediction models for indoor wireless communications in multifloored buildings, *IEEE Trans. Antennas Propagat.,* February 1992, vol. 40(2), 207-217

[7.6] Morrow, R., Site-specific engineering for indoor communications, *Appl. Microwave & Wireless,* March 1999, 30-39

[7.7] Ganesh, R., and K. Pahlavan, *Statistical modeling and computer simulation of indoor radio channel,* Proc. IEEE, June 1991, vol. 138, 153-161

[7.8] H. Hashemi, H., The indoor radio propagation channel, *Proc. IEEE*, July 1993, vol. 81(7), 943-968

[7.9] Karnik, A., and A. Kumar, Performance analysis of the Bluetooth physical layer, *Proc. ICPWC 2000*, 70-74

[7.10] Obeyashi, S., and J. Zander, A body-shadowing model for indoor radio communication environments, *IEEE Trans. Antennas Propagat.,* June 1998, vol. 46(6), 920-927

[7.11] Davis, T., and C. R. McGuffin, *Wireless Local Area Networks*, New York, NY: McGraw-Hill, Inc., 1995

[7.12] Crimini, L. J., Performance studies for high-speed indoor wireless communications, *Wireless Personal Commun.,* April 1996, vol. 14, 441-448

[7.13] Manabe, T., Y. Miura, and T. Ihara, Effects of antenna directivity and polarization on multipath propagation characteristics at 60 GHz, *IEEE J. Select Areas Commun.,* 1995, vol. 2(1-2), 67-84

[7.14] Manabe, T., K. Sata, H. Mazuwaza, K. Taira, T. Ihara, Y. Kasashima and K. Yamaki, Polarization dependence of multipath propagation and high-speed

transmission characteristics of indoor millimeter wave channel at 60 GHz, *IEEE Trans. Veh. Technol.,* May 1995, vol. 24, 268-273

[7.15] Saleh, A. A. M., A. J. Rustako, and R. S. Roman, Distributed antenna for indoor radio communications, *IEEE Trans. Commun.,* December 1987, vol. COM-35, 1245-1251

[7.16] Driessen, P. F., Gigabit indoor wireless systems with directional antennas, *IEEE Trans. Commun.,* August 1996, vol. 44, 1034-1043

[7.17] Kuga, N., H. Arai and N. Goto, A notch-wire composite antenna for polarization diversity reception, *IEEE Trans. Antennas Propagat.,* June 1998, vol. 46, 902-906

[7.18] Muller, N. J., *Bluetooth Demystified*, New York, NY: McGraw Hill, Inc., 2001

[7.19] Bray L., and C. F. Sturman, *Bluetooth – Connect without Cables*, Upper Saddle River, NJ: Prentice Hall PTR, 2001

[7.20] Miller, B. A., and C. Bisdikian, *Bluetooth Revealed*, Upper Saddle River, NJ: Prentice Hall PTR, 2001

[7.21] Bergmann, S. A., and H. W. Arnold, Polarisation diversity in portable communications environment, *Electron. Letts.*, May 1986, vol. 22(11), 609-610

[7.22] Guo, Y.J., A. Paez, R. A. Sadeghzadeh, and S. K. Barton, A circular patch antenna for radio LAN's, *IEEE Trans. Antenna and Propagat.,* January 1997, vol.45(1), 177-178

[7.23] Ng, A. K., J. Lau and R.D. Murch, 2.4 GHz ISM band antenna for PC cards, *Proc. IEEE Int. Symp. 1999, Antenna and Propagation Soc.*, July 11-16, 1999, Orlando Florida, USA, vol.3, 2066-2069

[7.24] Lee, E., Hall, P. S., and P. Gardner, Dual band folded monopole/loop antenna for terrestrial communication system, *Electron. Letts*, 2000, vol.36(24), 1990-1991

[7.25] Salonen, P., L. Sydänheimo, M. Keskilammi and M. Kivikoski, A novel antenna solution for Bluetooth access point, *Proc. IEEE Wireless Communication and Networking Conf.*, Chicago, IL, USA, Sept 2000, 23-28

[7.26] Neelakanta, P. S., and J. Sivaraks, A novel method to mitigate microwave-oven dictated EMI on Bluetooth communications, *Microwave J.*, July 2001, vol. 44(7), 70-88

[7.27] Neelakanta, P. S., J. Sivaraks and C. Thammakorannonta, Bluetooth-enabled microwave ovens for EMI compatibility communications, *Microwave J.*, January 2001, vol. 44(1), 138-151

[7.28] Kajiwara, A., Circular polarization diversity with passive reflectors in indoor radio channels, *IEEE Trans Veh Technol.,* May 2000, vol. 49(3), 778-782

[7.29] Boti, M., Dussopt, L., and J. M. Laheurte, Circularly polarised antenna with swichable polarisation sense, *Electron Letts*., August 2000, vol. 36(18), 1518-1519

[7.30] Kajiwara, A., Line-of-sight indoor radio communication using circular polarised waves, *IEEE Trans Veh. Technol*., August 1995, vol. 44(3), 487-493

[7.31] Hettak, K., and G.Y. Delisle, Smart antenna for capacity enhancement in indoor wireless communications at millimetre waves, *Proc. IEEE 51st Vehicular Technology Conf.* VTC 2000, Tokyo, vol.3, 2152-2156

[7.32] Tung, E. and R.H. Johnson, A multiport antenna for indoor PCS smart receiver, *Proc. 1999 IEEE Int. Antenna and Propagation Soc. Symp*., vol.4, 2436-2439

[7.33] Sivaraks, J., *A Study on Bluetooth and Mobile IP/CDPD Technologies: Performance Considerations and EMI/RFI Implications,* Ph. D. Dissertation, May 2001, Department of Electrical Engineering, Florida Atlantic University, Boca Raton, Florida, USA

APPENDIX 7.1

Characteristics of Bluetooth Packets

7A1.1 TIMING CONSIDERATIONS

As described in the text of this chapter, the transreceive time of communication in a Bluetooth link operating in a given piconet is divided into slots, each with a nominal length of 625μs. The purpose of these slots is as follows:

The downlink transmissions (from the master-to-slave) are designated to start only at the beginning of even slot numbers; and the uplink transmissions can start at the beginning of odd slot numbers, but only immediately after a downlink burst sent to the respective slave. Hence a *time-division duplex* (TDD) scheme is implemented with polling for uplink traffic. Further a guard time or switch time between two different hops is 220 μs; thus in a single-slot communication the effective time is only (625 – 220) μs = 405 μs. Multi-slot ACL packet are also provided for gaining the high data rate performance. Multi-slot ACL packets reserve 3 or 5 consecutive slots. The frequency hopping rate is usually 1600/s (namely once per slot), but within multi-slot bursts no hopping occurs and for the entire time can be used for more effective communication without guard time.

Bluetooth packets consist of a 72-bit access code for piconet identification and synchronisation purposes, a 54 gross bit header determining the connected slave and other properties of the packet, and a payload of variable length. A received access code is accepted if the number of matching bits in the 64 sync-word bits of the access code exceeds a certain threshold.

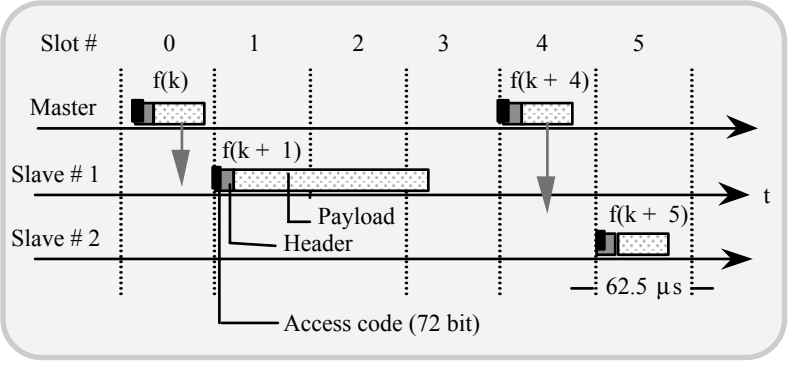


Fig. 7A1.1 Bluetooth timing: Slave # 1 transmits a 3-slots long DM3 packet and slave # 2 transmits DM1 packet of 1-slot long

The *asynchronous packet data connections* (ACL) packets of Bluetooth are detailed in Table 7A1.1.

Header and ACL burst payload are accepted if the header and payload CRC match. If any part of burst fails, the entire packet has to be retransmitted. SCO bursts have no payload CRC and may therefore contain residual bit errors. Table 7A1.1 presents some important properties of the available packet types for asynchronous packet data connections (ACL) and for real-time, constant rate connections (SCO). In DM packets, the payload is FEC coded with a simple 2/3 rate block code, while DH packet payloads are uncoded. After transmission of only partially filled ACL packets, nothing is transmitted in the remaining time.

Table 7A1.1 Properties of Bluetooth ACL packets for different link types

Link type	Packet type	Payload FEC code rate	User payload [bytes]	Burst length (in μs)	Used slots
Control ACL	NULL		0	126	1
	POLL		0	126	1
	DM1	2/3	0-17	171-366	1
	DM3	2/3	0-121	186-1626	3
	DM5	2/3	0-224	186-2871	5
	DH1	1	0-27	150-366	1
	DH3	1	0-183	158-1622	3
	DH5	1	0-339	158-2870	5
SCO	HV1	1/3	10	366	1
	HV3	2/3	20	366	1
	HV5	1	30	366	1

Table 7A1.2 ACL packets of Bluetooth and their characteristics

Packet type	User payload (bytes)	Pay-load header	Maximum overall packet size (bits	FEC	CRC (16 bits)
DM1	0-17	1	240	2/3	Yes
DH1	0-27	1	240	No	Yes
DM3	0-121	2	1500	2/3	Yes
DH3	0-183	2	1496	No	Yes
DM5	0-224	2	2736	2/3	Yes
DH5	0-339	2	2744	No	Yes
AUX1	0-29	1	240	No	No

ACL: Asynchronous connectionless
FEC: Forward error correction
CRC: Cyclic redundancy check

DM: Data-medium rate (1, 3 and 5 depict time-slot lengths)
DH: Data-high rate (1, 3 and 5 depict time-slot lengths)

All SCO link types provide transparent synchronous bi-directional transmission with 64 kb/s in each direction. For this rate, a robust CVSD (continuous variable slope delta) modulated voice coding scheme is defined within the Bluetooth system. The difference between the HV packet types is due to the payload FEC encoding, see Table 7A1.2.

One HV1 link occupies half the capacity, and a HV3 link occupies one third. The maximum packet lengths in Table 7A1.2 are at least 220μs shorter than the respective multiple of the slot duration to allow for synchronised-tuning to the next FH channel.

This page intentionally left blank

APPENDIX 7.2

Multiple Indoor Wireless Transmissions

7A2.1 COEXISTENCE OF MULTIPLE, INDOOR WIRELESS SERVICES

While it is anticipated that WLAN would become a legitimate enterprise technology in 2000s, there have also been forecasts on the so-called *multi-tenant unit* (MTU) broadband access. This can facilitate high-speed Internet access to tenants in multiple tenant locales such as residential apartments, offices, condominiums, hotels, airports and other public facilities. Here lies a niche for the deployment of *building local exchange carriers* (BLECs) to combine and interoperate WLAN, Bluetooth™ etc. and coexist with legacy LANs so as to facilitate a wireline/wireless MTU access, including high-speed Internet etc.

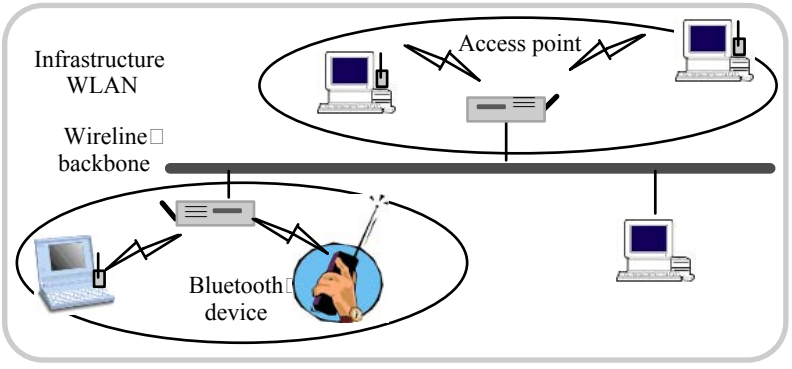


Fig. 7A2.1(a) Integration of Bluetooth and WLAN: Integration in infrastructure WLAN

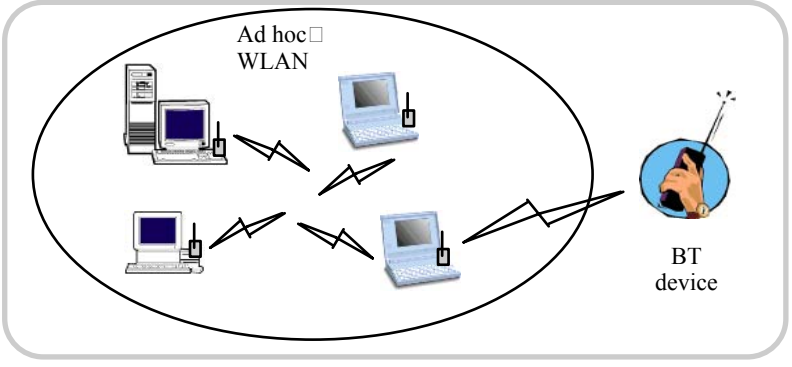


Fig. 7A2.1(b) Integration of Bluetooth and WLAN: Interconnection in *ad hoc* WLAN

The hybrid and coexistence facilitation of WLAN plus Bluetooth in the indoor environments would again warrant antenna features and implementation considerations on the coexistence environment under discussion. The feasible aspects of relevant system integration in WLAN/LAN communications along with Bluetooth operation are described below:

The scope of developing viable methods for system integration of Bluetooth, WLAN and legacy LAN transmissions refers to a master control approach for interoperation purposes. The interoperation between the Bluetooth™ devices and the wireless LAN could be in two possible interconnected scenarios, as indicated in Figures 7A2.1(a) and 7A2.1(b). The possible techniques for interoperation between WLAN and Bluetooth as depicted in Figure 7A2.1, refer to:

- Bluetooth-to-WLAN interoperation via traditional TCP/IP (as per Bluetooth Specification 1.0B in Part K:9 LAN Access Profile)
- Bluetooth-to-Ethernet/WLAN(*Ad Hoc* Mode) interoperation via SDP (future version) and bridge concept
- Bluetooth-to-WLAN interoperation through Infrastructure Mode.

The following are relevant details for such system integration efforts:

Bluetooth-to-WLAN: Traditional TCP/IP based implementation
To understand the steps involved, the following acronyms are first furnished.

TCP:	*Transport Control Protocol* that refers to the part of Internet protocol (IP) suite intended for providing reliable data transfer
PPP:	*Point-to-Point Protocol* of IP suite, which provides a standard way of transporting datagrams from many other protocols over point-to-point links
RFCOMM:	This is a protocol for RS-232 serial cable emulation in Bluetooth systems
L2CAP:	*Logical Link Control* and *Adaptation Protocol*, which is implemented by a layer (called logical link control and Adaptation, L2CAP) in the Bluetooth stack. (This layer provides segmentation and reassembly services to allow large packets to pass across Bluetooth links and facilitates multiplexing for higher layer protocols and services)
LMP:	*Link Management Protocol* implemented by the link manager (LM), which handles configuration of the Bluetooth baseband links
SDP:	*Service Discovery Protocol* of Bluetooth that allows a client to discover the devices offered by a server
LLC:	*Logical Link Control* is released by the IEEE 802 committee for using in conjunction with the medium access standard to provide the data link layer service to network layer protocol

MAC: *Media Access Control* is another sublayer of data link layer that is defined by the IEEE 802 committee

PHY: *Physical Layer* depicting the lowest layer of the OSI seven-layer model

The procedure indicated below refers to integrating WLAN/LAN via TCP/IP with PPP over RFCOMM from the Bluetooth unit to an access point (AP), which emits WLAN or LAN frames of the IP packets are shown in Figure 7A2.2.

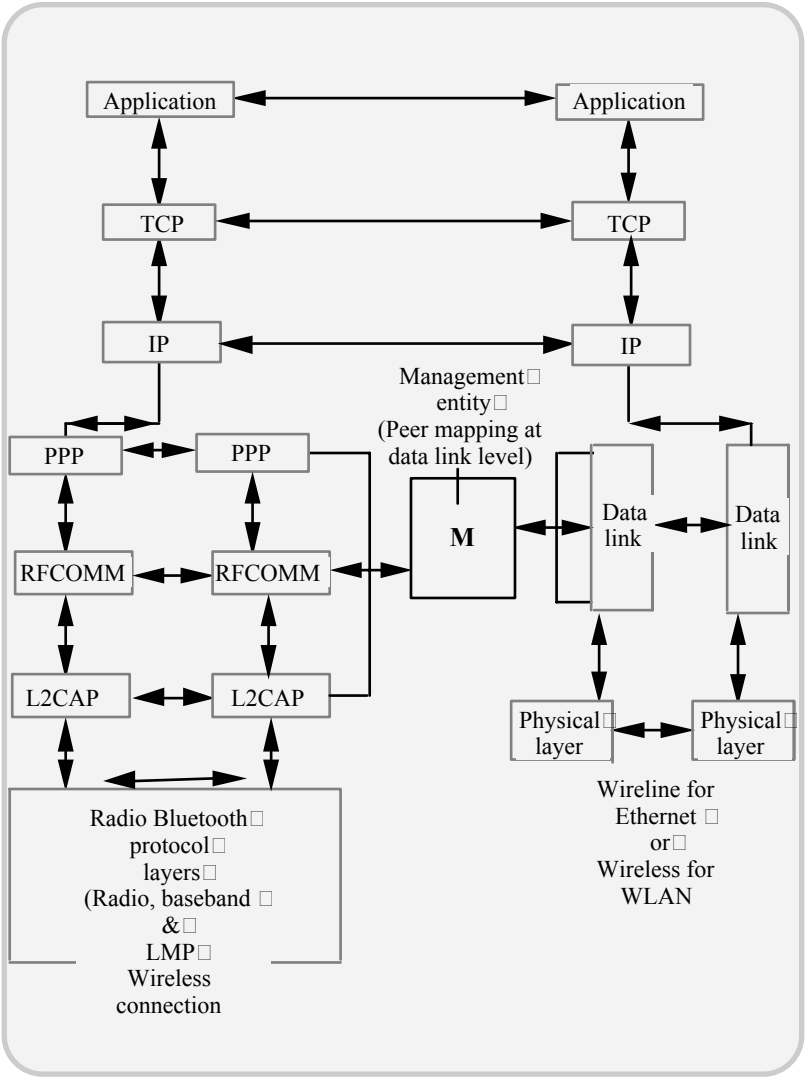


Fig. 7A2.2 Bluetooth-to-Ethernet/WLAN: Traditional implementation under TCP/IP environment

- When a Bluetooth data terminal comes within a specified range with another Bluetooth terminal Exchange Point. Then the LMP, Baseband, and Radio Channel in the Bluetooth protocol will be established automatically
- The LAN access application in the Bluetooth device would then send a command to L2CAP to establish a connection with the Exchange Point (after the RFCOMM and PPP (Point-to-Point) complete the connection establishment function). Here, the RFCOMM refers to the emulation of a serial port and the PPP is the protocol that allows the serial port emulation at the IP node in the Internet, such as the conventional modem used in while connecting to the Internet.)
- Next, the Management Entity would map between PPP, RFCOMM, L2CAP in Bluetooth architecture as well as the whole Data Link layer (LLC and MAC) of Ethernet/Wireless LAN; hence, the IP, TCP, and higher layer applications of both Bluetooth and WLAN can communicate with each other
- This approach is, however, limited by the bit rate on RFCOMM as specified by the maximum allowed bit rate on a serial port. It means that there is a limit on the speed of interoperation. Moreover, the Management Entity could be more complicated than the bridge concept (presented in the next section) specified to map between two different MAC layers
- The merit of this technique is that it is simple to implement since the PPP/RFCOMM operation in this profile is similar to PPP operation in normal dial-up networking, except that the modem uses AT commands to establish the link whereas the Bluetooth uses the RFCOMM
- In traditional sense, this technique may not be quite compatible for Bluetooth and WLAN interoperations based on PPP/RFCOMM. A Bluetooth device requires automatic service discovery. With the increasing number of Internet services and different types of networking that may exist, the Bluetooth device may have to reconfigure its attributes wherever it is moving to a new network.

Bluetooth-to-Ethernet/WLAN (ad hoc mode) interoperation via SDP (future version) and bridge concept

Here the integration of WLAN/LAN is done via bridging/reframing Bluetooth's L2CAP frames into a WLAN or LAN frame. This concept is based on future IEEE 802.15 (Wireless Personnel Area Network) and Bluetooth Version 1.1 Specifications, where the L2CAP packet is treated as a kind of MAC frame. Consequently, the interoperation between L2CAP and other MAC could be based on the bridge concept. Moreover, the SDP concept can be substituted for

PPP/RFCOMM protocols so as to avoid the limitations of RFCOMM based technique indicated in the previous section.

- This concept allows the SDP applications on Bluetooth and the WLAN to communicate with each other through the exchange of L2CAP packets and WLAN MAC frames
- A "bridge" is used just to repacketise the L2CAP packet into a WLAN MAC frame
- Bluetooth™ SDP is presumably compatible with the LAN/WLAN SDP. (Note: LAN/WLAN SDP does not exist currently)
- Further, the LLC protocol could be available only in Bluetooth™ specification: and, as indicated above SDP has to be installed as the new protocol for the WLAN workstation.

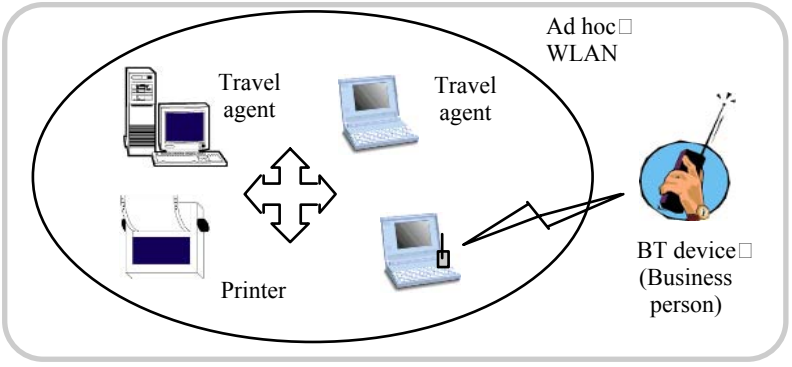


Fig. 7A2.3 Example scenario of concept # 2

An example of realising this technique is as follows: Suppose a business executive on travel carries a Bluetooth-enabled PDA for storage of information on flight schedules, rental car etc. It is assumed that this person visits a travel agent's office, which has a cluster of laptops supported on an *ad hoc* wireless LAN; and, these computers are Bluetooth-enabled. Suppose there is printer in that office (on that *ad hoc* network), which is not, however, Bluetooth-enabled. As such, the business executive cannot directly link the PDA to this printer. But he/she may connect the PDA to one of the laptops *via* the Bluetooth mode of operation and hence could get access to the printer in question. This way of accessing is automatic in the sense that the PDA will automatically discover the printer (including its characteristics and service abilities), even though the PDA is now totally on a foreign network. This automatic process of accessing and discovering the printer (or any other non-Bluetooth enabled end-entity) is not, however, feasible, if the Bluetooth and WLAN interoperation is based on PPP/RFCOMM protocol.

The above concept is illustrated in Figure 7.A2.3 (SDP* and LLC*: They do not exist currently. There are potentials for their specifications and implementations, in future.)

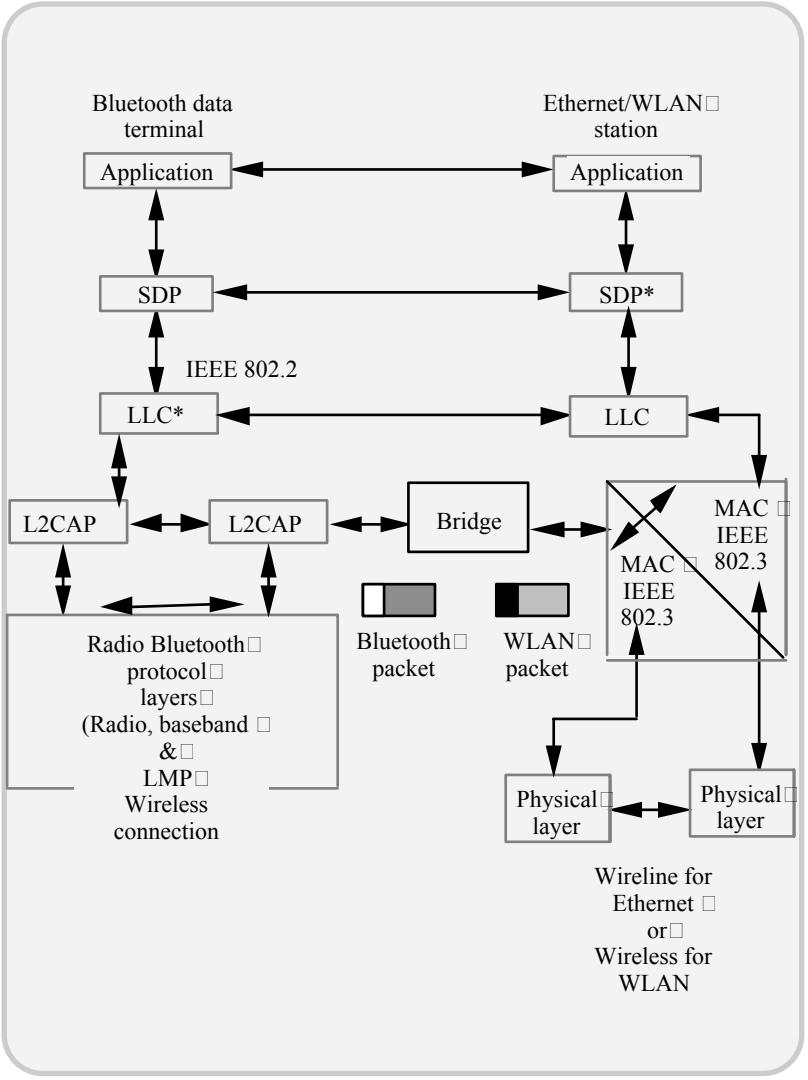


Fig. 7A2.4 Bluetooth-to-Ethernet/WLAN (*ad hoc* mode) interoperation *via* SDP (future version) and bridge concept

Bluetooth-to-WLAN interoperation through infrastructure mode via SDP (future version) and cascaded bridge concept

Similar to concept # 2, this technique could also be an extended version for interoperation of LAN, WLAN and Bluetooth networks. Here, the access point (AP) in the WLAN infrastructure mode could be specified as a bridge between the wireless station and the wireline Ethernet network. Consequently, a cascaded bridge concept applies towards the interoperation between Bluetooth network and the WLAN infrastructure mode.

In all the feasible co-operative/interoperative scenarios of WLAN and Bluetooth plus legacy LANs, it is imperative to note that compatible antenna designs are warranted in all WLAN units, Bluetooth and access points on *ad hoc* basis. Traditional voice/data communication devices (such as telephone, computers, PDAs etc.) as well as other electronic appliances, which are Bluetooth-enabled, would require unique antenna configurations and implementations. Frequency of operation, fixed or switched polarisation and single or multiple (for diversity) deployments in these devices call for strategic application considerations and tactical design perspectives [7.33].

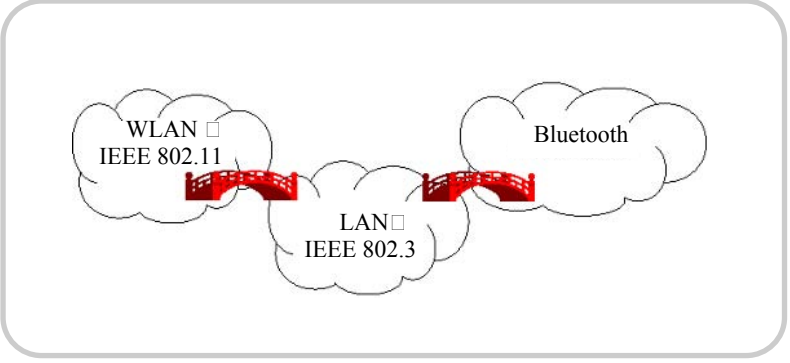


Fig. 7A2.5 Bridging of Bluetooth and WLAN with a legacy wireline LAN (*Note*: SDP * and LLC do not exist currently. There is potential for their future specification and implementation [7.33])

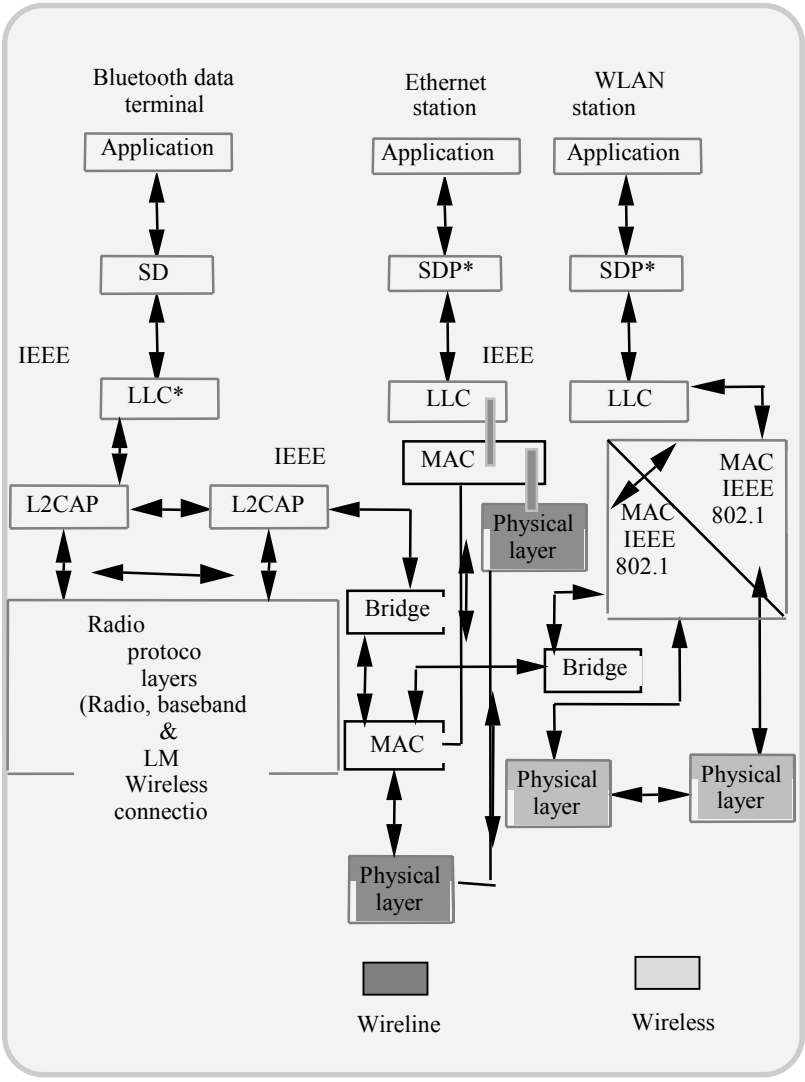


Fig. 7A2.6 Bluetooth-to-WLAN (Infrastructure Mode) interoperation *via* SDP (future version) and cascaded bridge concept

CHAPTER 8

Broadband Wireless Communication Systems and Antennas

8.1 INTRODUCTION

In the wireless scenario, there has been an increasing demand for advanced information services, largely posed by bandwidth-intensive transmissions such as high-speed access and multimedia systems. High-speed in network access, whether be it *via* wireline or wireless, is much needed for novel applications like *data broadcasting* and *multicasting* strategies.

The digital cellular telephony being extended into 3G-service profiles is required to facilitate high-speed access to the Internet and multimedia services. Relevant considerations have become service and technology drivers for modern broadband wireless system-designs, including the associated antenna structures. The wideband outdoor wireless systems, in general, encompass the relevant terrestrial access networks and satellite-based systems. The emerging 3G-communication systems focus on multimedia information services over Internet protocols (IP) and IP/ATM over satellite communications (SATCOM) networks.

On the terrestrial side, providing wideband service to mobile users and enabling high-speed access to homes through radio ports are the part of blooming wireless communication trends. Eventually, the conceived *wireless local loop* (WLL) access links are aimed for broadband services and expected to serve the customers probably with Gbps-rate/broadband information flow. The feasibility of wireless distribution of digital TV (*in lieu* of contemporary wireline- or cable-based TV programmes) is another challenge faced in the terrestrial application of wireless technology. There also exist broadband wireless ambient which are concerned with supporting relevant services through indoor systems, such as wireless LAN (WLAN). High data-rate WLANs require exclusive allocation of wide bandwidths. The FCC in the United States and the equivalent regulatory bodies elsewhere assign these frequencies for such applications. Over the past decade or so, two major standardisation efforts of WLANs have emerged: The IEEE 802.11 standard in the ISM bands (910MHz, 2.4 GHz, and 5.8 GHz) and the European HIPERLAN (*high performance radio LAN*) standard in the 5 GHz band. Both standards deem to support multimedia traffic — the IEEE 802.11 standard at data rates up to 2 Mbps and the HIPERLAN standard at data rates up to 23.5 Mbps.

Data-PCS is a license-free band for new high data-rate LAN for the emerging wireless communication systems. Data rate of 10 Mbps is designated for

PCS and 20 Mbps is specified for other emerging services. FCC has allotted 10 MHz bandwidth for data in the 2.39-2.40 GHz band with necessary spectrum etiquette constrains in such cutting-edge technologies.
Millimetre (mm) wave technology is another avenue indicated for broadband wireless transmissions. The relevant frequency bands reside above 40 GHz. Three available bands, namely, 47.6-47.8 GHz, 59-64 GHz, and 76-77 GHz are specified for unlicensed systems. The 59-64 GHz allocation has been indicated for computer-to-computer high data-rate (5 Gbps) transmissions.

8.2 BROADBAND WIRELESS LOCAL ACCESS

The Working Group 802.16 in the IEEE 802 LAN/MAN Standard Committee was set up as a part of the *broadband wireless access* (BWA) initiative. The objective of BWA systems is to utilise base stations for providing broadband data to business or homes as an alternative to wired "last-mile/first-mile" access loops (that traditionally use fiber-links, telephone-lines or cable connections). (In a broad sense, the base stations can be either terrestrial, or in orbit, or mounted on airplanes or dirigibles in the stratosphere.)

In the conceived BWA, the customer terminal can support two-way communications for Internet access, digital video, telephony and other services. Blocks of microwave and/or millimetre wave spectrum (of the size of about 1 GHz) will be made available for BWA networks. Typically, 24 GHz, 28 GHz, 31 GHz or 40 GHz are the frequency allocations contemplated worldwide. It is expected that BWA will be more cost-effective than the wired counter-parts. In the United States, 24 GHz and 38 GHz systems are on commercial trials since 1998. The following is a summary of relevant considerations [8.1 – 8.5].

8.2.1 Local multipoint distribution service

Local multipoint distribution service (LMDS) is a wireless, two-way broadband technology designed to allow network integrators and communication service providers to quickly and inexpensively bring a wide range of high-end, quality services to homes and businesses. Services that can use LMDS technology include high-speed Internet access, real-time multimedia, file transfers, remote-access to corporate local area networks, interactive video, video-on-demand (VoD), video conferencing, and telephony among other potential applications. In the United States, the FCC considers the LMDS to introduce competition in the telecommunication marketplace, through the use of 1.3 GHz of RF-spectrum for transmitting voice, video and fast-data to and from homes and businesses. In terms of existing LMDS technology, this approximately translates to a 1-Gbps digital data pipeline. (Canada already has 3 GHz of spectrum set aside for LMDS and is actively setting up systems around the country. Likewise, many other developing countries view this technology as a way to bypass the expensive implementation of cable or fiber optics and leapfrog into the twenty-first century.)

In the past, the telecommunication technologies focused their attention mostly in the lower side of the RF spectrum. This is mainly due to the ability of low-frequency signals to reach longer distances and penetrate obstacles. LMDS, however, uses low-powered, microwave/millimetre wave regime of frequency (25 - 31 GHz) signals that are characteristically useful only over a short distance because

of associated rain/snow-attenuation properties and other quasioptic propagation artifacts. As such, LMDS systems are designed cellular over short line-of-sight distances. These cells are typically spaced 4-5 kilometers (2.5 - 3.1 miles) apart. Thus, LMDS cell layouts have to bear the real-estate cost of building base stations. Further, direct line-of-sight between the transmitter and the receivers within the foot-print of the cell is necessary so as to avoid shadow regions and interference between signals. Field studies indicate that a single transmitter would reach only slightly more than 60% of the homes in a cell. With overlapping cells and repeaters, however, that number can be increased to almost 85% of homes.

Since LMDS signals are supported on microwave/millimetre wave transmissions, they are attenuated by rain due to molecular absorption. This could influence the cell-size covered by the service rendered. The loss in signal strength should be corrected either by increasing the transmitted power when it rains, or by restricting the cell size. Intervening vegetation can also cause signal-loss, but overlapping cells and roof-mounted antennas generally may overcome this problem.

For the purpose of BWA services, the FCC auctioned the RF spectra in 1998 for ground-based wireless networks to deliver the full range of broadband services that can be deployed quickly and inexpensively in point-to-point and point-to-multipoint configurations to residential and commercial customers. Relevant details on the evolution of LMDS specifications are presented in Appendix 8.1.

Now, why LMDS?

The concept of LMDS blossomed due to the following considerations:

- To realize high-speed/broadband access at premises that cannot be covered (accessed) cost-effectively by wirelines (due to distance and/or expensive capital expense on low subscriber-density in the region)
- LMDS offers high-speed/broadband dedicated links between high-density nodes in a network
- Relevant network installations can take place concurrent to customer premises equipment (CPE) installation. This allows the service providers not to invest *a priori* and wait for revenue
- Further, LMDS carries the following merits:
    - Low entry/deployment cost expended towards last-mile solutions
    - *Ad hoc* installation (Easy and speedy)
    - Fast revenue recovery
    - Demand-based build-out and scalable architecture employing open industry standards ensuring services and coverage areas that can be easily expanded in par with customer demands
    - No stranded capital when customers churn (resulting in cost-shift from fixed to variable components)

* Feasibility for cohabitation with cellular service providers in the area for cost effective OAM (operation, administration and maintenance)

LMDS – its network architecture

An LMDS setup has a central facility with a fibre-linked PSTN and Internet connections. It refers to either a CO-based or analog fibre-based architecture as distinguished in Table 8.1.

Table 8.1 CO-based and analog fibre-based LMDS architectures

CO-sited BS architecture	Analog fibre-based architecture
CO and BS are at the same site. This is the most commonly used version	BS indoor equipment is connected to multiple microwave outdoor systems
Indoor digital equipment connects the network infrastructure to roof-top microwave system (plus transreceive antennas)	Analog fiber is used for indoor to outdoor interconnection
Transreceive antennas provide 90-, 45-, 22.5- or 15-degree sectoral coverage. (Omnidirectional coverage of the cell-site is obtained through divided 4, 8, 12, or 24 sectors)	Outdoor units use tower-installed antennas

LMDS relays signal *via* line-of-sight (LoS) microwave links, which in turn pass the signal along to hubs, located on rooftops or as stand-alone towers, for point-to-multipoint (PMP) transport to the end-site. As such, a LMDS network architecture has four parts:

Network operation center (NOC)	Contains a network management system (NMS) for interconnecting NOCs. The network management equipment for managing the regions of customer network are housed in the NOC. Multiple NOCs can also be used and interconnected.
Fibre-based infrastructure	Includes SONET/OC (12/3) and DS-3 links, central office (CO) equipment, ATM/IP switches, and interconnection facility with the Internet and PSTN.

Base station (BS)	Consists of fibre-termination network interfacing with a wireless system, antennas and units of wireless traffic management, authentication billing etc.
Customer premises equipment (CPE)	Made of vendor-specific units such as antennas, transreceiver and interfaces for customer premises end-entities.

Wireline-to-wireless transition

The conversion from fibered-infrastructure to a wireless-infrastructure happens at the base stations. It includes necessary interface for fibre-termination, modulation and demodulation functions, and microwave transmission and reception equipment. Local switching can also be done at the base station. If local switching is present then customers communicating within the same base station can communicate with each other without entering into the fibre infrastructure.

The customer premises equipment may vary widely from vendor to vendor. But in general, all configurations include indoor digital equipment plus modulation circuits and outdoor-mounted microwave equipment (plus the antenna unit). The customer premise equipment may attach to a network using TDMA, FDMA or CDMA-based access methods. Different CPE require different configurations to run the full range of telecommunications traffic from DS-0, POTS, 10BaseT, unstructured DS-1, structured DS-1, the frame relay, ATM25 serial ATM over T1, DS-3, OC-3 and OC-1. Furthermore, the customer premise locations can range anywhere from malls to residential locales. Relevant two base-station architectural options are those indicated in Table 8.1.

Radio frequency planning

In radio frequency planning of LMDS, the networks typically use multiple sectors of microwave systems. That is, transmit and receive antennas are arranged to provide service over sectors, each covering 90, 45, 30, 22.5 or 15 degree beamwidth; and, in this idealized omnidirectional, azimuthal coverage, the included area around the cell is divided into 4, 8, 12, 16, 24 sectors. Alternative architectures include connecting a base station indoor unit to multiple remote microwave transmission and reception systems with analog fibre-interconnection between the indoor data unit and the outdoor data unit. There are also manufacturers who have introduced certain proprietary approaches in LMDS network architectures. (For example, there is a technology where the base station is airborne (set in an aircraft/drone). It transmits signals from overhead, but this technique is sophisticated and involves inevitable ranging from air traffic control.)

While coming up with architectures for LMDS, a standard issue that is considered is the point-to-multipoint (PMP) communication. In this context, a relevant question that arises is whether the PMP is actually required at all. PMP allows multiple microwave paths, thus requiring the spectrum and capacity to be shared, as needed. So it requires higher bandwidth than the PTP (point-to-point)

transreceivers. A new model that is ramping up quickly is the *invisible fibre unit* (IFU). In this technique, two IFU's are setup in a LoS link and are placed back-to-back with other links. Thus, the IFU link has both transmit and receive links between source and destination.

Receiver design
In the LMDS service, the CPE has an outdoor unit with transmitter and receiver antenna and an indoor unit, which in turn, communicates with subscribers equipment such as telephones and PCs. The indoor unit accepts the signal from the outdoor unit, demodulates and demultiplexes it and then interfaces the signal with the connected subscriber equipment. The downstream *intermediate frequency* (IF) in LMDS is the same as the satellite IF (namely, 950-2050 MHz). A major design issue of a receiver lies in achieving a large frequency acquisition range in the carrier recovery loop.

Wireless link and access optionsIn general, for any wireless upstream link, there can be three access methodologies: TDMA, FDMA and *code-division multiple access* (CDMA). Specific to LMDS, there are two wireless links and access options, namely, FDMA and TDMA options, which are summarised in Table 8.2.

Table 8.2 Wireless links and access options: TDMA *versus* FDMA issues

Issues	TDMA	FDMA
User data	Slots used to support the data bursts. Slots not requested during idle (no data) duration	FDMA link is always on regardless of bursty or idle conditions
Wireless MAC	MAC efficiency is about 65-90 % depending on burstiness	No MAC is used
Customer premises mix	Delay sensitive traffics are allowed to pass through with higher priority in both TDMA and FDMA systems	
Channel efficiency	About 88 % due to preamble and ranging	-
FEC	75-85 %	91 %
Maximum data rate	Linked to burstiness *via* fairness algorithms	

In the downstream direction (from base station to customer premises), most companies supply TDM (time division multiplexed) streams either to a particular user or shared among various user sites.

The FDMA scheme allows a fixed bandwidth, or a bandwidth varying slowly over time. If the user requirement is a constant bandwidth (a dedicated one) and expecting continuous availability, like a wireless DS-3 or a multiple structured DS-1 connection, the FDMA access links are appropriate.

FDMA links terminate in a dedicated FDMA demodulator, which as it should be, is in the base station. When the customer does not have a intense upstream traffic and just needs a 10base T port, adopting the TDMA makes sense for practical reasons. In short, the choice is based on customer requirements and system design.

In the FDMA option, the following are relevant:

- Downstream from BS to customer is TDMMultiple user sites share the downstream connectionEach user site has a distinct frequency to tune in.

In the TDMA option, the following are relevant:

- Downstream from BS to customer is TDMMultiple user sites share both downstream and upstream channelsEach user site has a distinct time slot that bears the data.

The CDMA supports significantly smaller number of users than a TDMA. There are two classes of CDMA that are available, one is known as the *orthogonal CDMA* (O-CDMA) and the other is called the *non-orthogonal CDMA*. In practice, systems may often use a combination of the two. O-CDMA is said to have identical capacity to TDMA.

It allocates channels using a mutually orthogonal spreading (encoded) sequence (of data) to each channel. The other class of CDMA (namely, the non-orthogonal CDMA) is a pseudorandom sequence CDMA. In this techniques all the users may interfere with each other and the capacity is, therefore, limited by the extent of interference that a channel is prepared to tolerate. Both CDMA and TDMA may have certain case-based advantages to be considered in a specific system.

Space division multiple access (SDMA)

As described in Chapter 6, smart antennas are used in the base stations. Such antennas are based on adaptive array concepts and cover a sector adaptively, unlike the fixed beam antennas. With the help of multiple RF sensors, the beam-maximum can be moved blindly in the direction of the user.

This can be done dynamically (either horizontally or vertically). The smart antennas implement what is known as the *space division multiple access* (SDMA). Since the users in the TDM mode of operation sequentially use the channel (in time), it is well suited for the SDMA, (whereas with CDMA, requirement on simultaneous access makes it complicated for SDMA applications).

LMDS modulation schemes

Modulation schemes can tune the data rate to some extent. Low-density modulation allows greater distance at a given power, but sacrifices data throughput rates. The

LMDS, however, utilizes QPSK modulation. Therefore, it can realise about 1.8 Gbps of raw capacity. (In contrast even though the MMDS bandwidth is five times more, it can give only 1 Gbps for its downstream links with its 64-QAM scheme.)

Recently, broadband developers are looking into advanced coding methods to achieve efficient use of bandwidth. Hence, considerations of using coding techniques such as OFDM (*orthogonal frequency division multiplexing*) for LMDS have been put forward. Another new coding scheme, known as the *frequency-domain reciprocal modulation* (FDRM), has also been proposed as an alternative to ODFM. Further suggested is the use of *turbo-product codes* for LMDS applications. By using such turbo codes, it is expected that the wireless original equipment manufacturing (OEM) vendors/developers can cut the number of base stations necessary for LMDS Internet access. This would potentially reduce the effects of rain-fade common to such broadband wireless systems. A summary of modulation schemes prescribed for LMDS is as follows: (It includes the future standard on TDMA. The term CBR is an acronym for *constant bit rate*.)

TDMA	Modulation	
	64-QAM (A future standard)	
FDMA	Modulation	Bandwidth (in MHz)*
	BPSK	2.8
	DQPSK	1.4
	QPSK	1.4
	8-PSK	0.8
	4-QPSK	1.4
	16-PSK	0.6
	64-QAM	0.4

(* Specified for 2-Mbps rate CBR connection)

Data rate

The data rate varies for the various modulation schemes indicated above. The maximum data rate for a given type of modulation is specified by the *bandwidth spectrum efficiency* defined below. The TDMA band does not use the 64-QAM modulation. For the other modulations, it has reduced data rates. *Bits-per-second per hertz* (bps/Hz) is adopted as the metric for *data rate capacity* in both TDMA and FDMA accesses.

Spectrum efficiency
This is the figure-of-merit specified for different modulation schemes and is expressed in bps/Hz. Listed below are the spectrum efficiencies for different modulation schemes pertinent to FDMA.

Modulation	Spectrum efficiency (for FDMA)
4-QPSK	1.5 bps/Hz
16-PSK	3.5 bps/Hz
64-QAM	5.0 bps/Hz

Data rate capacity/FDMA
This refers to the number of cell sites within the system multiplied by the capacity per cell site.

Cell site capacity/FDMA
This is the number of sectors within the cell-site multiplied by the capacity per sector.

Data rate capacity/TDMA
TDMA has reduced data rate capacity (of about 80 % of that of FDMA).

Maximum number of customer premises sites
For the FDMA scheme, assuming a "x" MHz spectrum with a reuse frequency of "r", the LMDS system provides x/r MHz usable spectrum per sector. If one assumes the downlink spectrum to be "d" times the uplink spectrum, the downlink will have $d \times (x/r)/(d + 1)$ spectrum and the uplink would have a $(x/r)/(d + 1)$ spectrum. For a channel bandwidth of "b", the maximum number of customer premises equipment will be $(x/r)/((d + 1) \times b)$.

Considering the TDMA, for a given $(x/r)/(d + 1)$ spectrum, if one assumes about 16 DS-0 connections are possible with 1 MHz spectral bandwidth, then the total number of simultaneous users is equal to $16 \times (x/r)/((d+1) \times b)$. Further, if the connectivity concentration over the entire sector and cell are assumed to be in the ratio (1 : s), then the total connections will be $s \times 16 \times (x/r)/[(d+1) \times b]$, which is quite high when compared to what is feasible with FDMA.

CPE network interface in WLL
This interface essentially refers to the RF-to-baseband section of the home/office based end-entities. Its typical deployment is illustrated in Figure 8.1. The access is facilitated by means of a roof or side-wall installed antenna, which constitutes an air-to-RF interface. A network interface unit (NIU) is used to convert the RF signal and distribute it to an appropriate CPE, as shown.

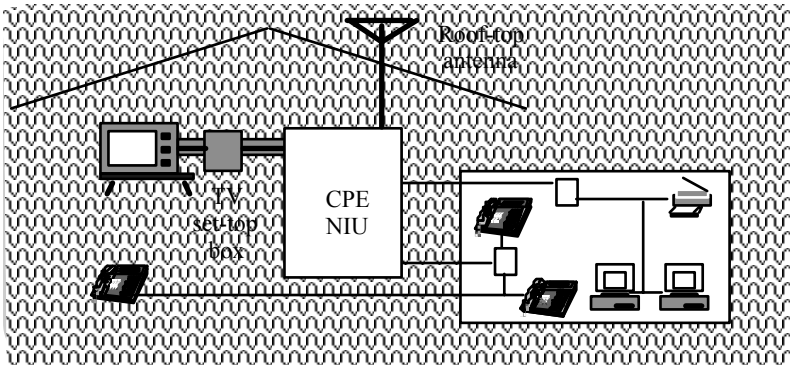


Fig. 8.1 Customer premises equipment (CPE) and WLL /RF interface. (NIU: Network interface unit)

Potential applications of LMDS and the related economics

LMDS applications have been largely centred around the transmission of video. With the recent surge of interest in the Internet and the accompanying demand for bandwidth, fast-data transport appears to be an added potential of using this technology.

Typically, with 1.3 GHz of spectrum, LMDS can provide a pipeline to support a variety of services. In practice, the residential subscribers pay about $30 per month for video, but businesses regularly pay over $1000/month for a high-speed T1 (1.544 Mbps) line from phone companies. Using only the 850 MHz unrestricted bandwidth, along with a modulation scheme such as QPSK, well over (100 × T1) equivalent lines can be accommodated in an LMDS cell (even without splitting the cells into separate sectors). If the telcos charge half the charges on such (100 × T1)-line equivalent service, they can generate $50,000 in revenue per month per cell!

Further, by using polarisation diversity (that is by adopting horizontal and vertical polarised sectors in a cell), LMDS providers can re-use the bandwidth twice as much and thereby multiply the number of T1-equivalents available. A typical commercial LMDS system can potentially provide a staggering downlink throughput of 51.84 - 155.52 Mbps and a return link of 1.544 Mbps (T1). This capacity translates into phenomenal infrastructure to provide "full service network" packages of integrated voice, video and high-speed data services. (Actual service carrying capacity would depend on how much of the spectrum is allocated to video *versus* voice and data applications.)

Assuming that 1 GHz-wide spectrum is available, an all-video system could provide up to 288 channels of digital broadcast quality (MPEG-2) television plus on-demand video services. Fast data/file transfers, high-speed Internet access, PCS backhaul, local-loop bypass, digital video, digital radio, work-at-home telecommuting, and telemedicine are other viable revenue-earning applications. And, as a matter of fact, they all could be provisioned within the same cell.

Transreceive hubs and cell sites (base stations) must be established up front for LMDS, but once the real estate is completed and the equipment is in place, additional costs will incur only when the new customers sign in. The largest fixed expense (capital cost) associated with building out LMDS cells will probably be the cost of subscriber equipment, not the transmission and infrastructure equipment itself. (By contrast, when installing wireline networks, the majority of costs are borne by the telco before the first-paying customer is turned on.) That is, LMDS can thrive with a low customer penetration rate; whereas, the wireline systems require high penetration rates in order to offset the cost of the initial capital required building them.

Armed with the experience gained through digital broadcast satellites, switched broadband, hybrid fibre-coax systems and other forms of wireless cable services in the past, implementing LMDS becomes conceivably feasible. Further, advances in the technologies pertinent to gallium arsenide (GaAs) integrated circuits, digital signal-processors, video-compression techniques and advanced modulation systems could supplant improvements in the performance and reduction in the cost of LMDS systems. Also, the rewriting of telecommunication regulations has scrapped many of the barriers that prevented companies from entering new businesses such as LMDS.

The demand for bandwidth, which used to be confined to large corporations, governments and universities, is now available to the consumer level, thanks to the proliferation of the Internet; and, tying this bandwidth quest through LMDA is logical and timely. Further, all the aforesaid considerations have combined to create a need for a technology that has access to a huge amount of bandwidth and can be deployed for low up-front costs – the LMDS meets this need.

LMDS can be actively pursued in an effort to get more commitment to deploy revenue-yielding services in the residential market. Vendor equipment is already available to provide services to commercial building and *multi-dwelling unit* (MDU) buildings. However, the rural providers want to look at solutions for single-family dwelling units.

For commercial/residential subscribers in the urban/suburban areas as well as for sparsely populated rural customers, a telco can offer a wide variety of services beyond the traditional telephone offerings (including long-distance), cellular telephony, Internet access, and wireless cable television. LMDS can be installed and maintained for distance-learning applications covering several rural schools (elementary and secondary) and/or several university campuses. The remote and rural medical clinics can also profitably make use of the LMDS.

In addition, with an integrated network planning, the telcos will have more efficient use of access network and the economies of scale in operations. For example, the LMDS equipment can be used off an ATM-based switching platform, so that it will allow the company to overlay its circuit-based network with ATM. This will offer new and future ATM-based feature-added services. Such ATM-LMDS system can yield an opportunity to expand as a facility-based broadband wireless *competitive local exchange carrier* (CLEC). It is both a promising option and an alternative in the ever changing and competitive industry environment.

Summarizing, the LMDS works on a fixed-wireless paradigm and has the following advantages:

- Its entry and setup costs are small; that is, the setup investment is low and expansion can always be opted on demand. It means revenue grows concurrent to service extended without latency
- The system can be setup with ease and speed. Equipment are made portable and can be installed without much of erection-time
- Equipment can be setup only after customer signs up (This is different from wired systems because for wired access, a complete infrastructure has to be built up-front even before the customers show up). Thus the build out becomes "demand-based"
- Cost of upgrading can be substantially less, as there are no other infrastructure overheads other than the end-equipment. Once the equipment is designed, upgrading is simple
- Because of localised equipment, there is less overhead in changing the transmission equipment and/or tracing the faults
- Once the basic infrastructure is well placed and organised, the reliability of service is assured
- Bandwidth reuse (because of the cell structure used and possible dual polarisation schemes) enhances the system capacity
- Network management, maintenance and operation costs can be relatively less.

Network planning

Network planning for LMDS includes cell design, issues related to planning the frequency, data transmission speeds, cell reuse considerations and reuse optimisation as summarised below:

Cell design issues

Basic considerations towards designing a LMDS cell are as follows:

Cell size selection:	Based on the desired reliability level
Cell overlap:	An issue that has to be taken into consideration while designing the cells
Subscriber penetration:	The number of subscribers with prescribed minimum signal level to achieve robust service
Number of cells:	The number of cells in a sector is dependent on the cell size decided
Traffic capacity:	Based on the traffic capacity of the area, the cell size and properties are fixed
Quality of service:	Cell overlaps that exceed the allowed normal can affect the quality of service
Link budget:	Estimation of the maximum distance that a user can be located from the cell while still achieving acceptable service reliability.
Capital cost per cell:	Used to estimate the network capital requirement.

Frequency planning

The channel spacing that is usable by the operators in Europe is 112, 56, 28, 14, 7, and 3.5 MHz. (These are obtained by successive division of 112 by 2.0.) The capacity in upstream and downstream usually differs because, even if the bandwidth allocated is same, the physical layer functions of both channels are different. Therefore, even if the bandwidth is equally distributed among the upstream and downstream channels, it is not possible to get the same capacity. Hence, physical layer issues such as channel coding and filtering have to be taken duly into consideration when planning the channels, if equal capacity for down and up links is desired.

Reuse schemes

An important issue that can substantially change the speed of transmission and utilisation of bandwidth is the *frequency reuse* consideration. This refers to how effectively the frequencies can be reused in a given geographical area. First possibility is to use a hexagonal cellular pattern (like the traditional mobile cells). The frequency allocation scheme requires three times the bandwidth allocated to one cell.

Another possibility is to use rectangular cells. Each quadrant of the cell is labeled with a digit, which indicates the frequency or group frequencies used in that sector. The frequency reuse pattern reduces the bandwidth requirements by two with the use of two orthogonal polarisations. This is the initial state, after optimization the distribution is made only with two colours. Antenna sectoring within a cell has advantage of reducing the maintenance costs.

The techniques to optimize frequency reuse are as follows:

- Maximisation of isolation between adjacent sectors by use of polarisation.
- Maximisation of the directivity of the cell antennas by sectoring the distribution system. Minimisation of cross-polarisation and multipath.

LMDS promises a wireless alternative to fibre and coaxial cables It has the potential to replace the existing wired networks, it may prove to be the easiest way to deliver high-speed data and two-way video service. Its capability of handling thousands of voice channels with the existing bandwidth makes it a good contestant in the voice industry. With current industry trends, that are tending to merge the telecommunication and the networking industries, LMDS seems to be a solution that suits all their needs. For the recent digital TV world, LMDS is a boon of a choice considering the fact that LMDS was designed with digital TV broadcast in mind.

LMDS and MMDS … a summary

LMDS is a newly envisioned, stationary (fixed) broadband wireless access technology designed to attract a mass subscriber marketplace. It provides for local multipoint distribution services. It uses millimetre frequencies at 24 GHz and above. LMDS has the potential for facilitating cheaper in-building, wideband

communication links than fibre or copper. The wireless technology such as LMDS allows wireless data for the enterprise both cheap and ubiquitous.

Several vendors have commenced offering LMDS products today; however, this technology is still in its development stages. In order to use this technology for production level enterprise-wide, the public-shared networks namely, network service providers (NSPs) and the Internet service providers (ISPs) should first acquire a wireless infrastructure for widespread public service offering. (The internal Intranet solutions using LMDS are available currently.)

MMDS is another technology for wireless access - particularly useful for the Internet. It enables a multi-channel multipoint distribution service. With MMDS, a transmitting tower placed at a high elevation can reach customers over the horizon to about 35-mile radius. The users of this service install receiving dishes on the side or roof of the building.

Presently, in the United States, the network service providers such as Sprint and MCI WorldCom have been considering MMDS for use in reaching the local customers without negotiating access agreements with regional Bell operating companies. As a viable technology, MMDS is contemplated to be ready to take off the ground in 2002.

In comparison, LMDS has a smaller radius of coverage and is more expensive to deploy than MMDS. It is expected that MMDS may go ahead in the market first due to economic and technologic considerations.

8.2.2 WLL based on Wideband CDMA

Wideband CDMA (W-CDMA) technology has been selected for WLL service to provide primarily ISDN-like data rates to subscribers. It offers a smooth transition into the *third generation* (3G) wireless technology and its air interface is an interim TTA standard, which specifies a CDMA-based protocol operating in the 2.30-2.40 GHz band. This standard allows two types of bandwidth, namely, 5 and 10 MHz.

Typically, a W-CDMA system may consist of a cell in which a mobile station and a radio base station operate over a 5 MHz broadband channel. This channel supports specific transmission speeds assigned for voice, the Internet and real-time video signals. A video encoder is used to compress transmission by measuring the changes from image-to-image of slow sequences and rapid sequences – in a way that uses the radio channel flexibly and with optimal quality.

3G-wireless technology

The goals of the third generation (3G) technology include a wide range of operating wireless environments pertinent to high-mobility (vehicular), low-mobility (pedestrian) and almost-stationary (indoor) scenarios. The evolution of wireless technology from (2G)-to-(2.5G)-to-(3G) transitions refers to graceful migrations in the following systems:

- Low data-rate (around 10 kbps) systems (such as GSM, ISM-136/D-AMPS and IS-95/CDMA) adopted for PCS (indoor) and cellular (outdoor) applications

- High data-rate (10 Mbps or above) systems (such as Hiper-LAN, *ad hoc* networks, IEEE 802.11 etc.) designated for indoor wireless LAN applications.

In such migrations, the objectives of 3G-wireless systems include the following considerations:

- Achieving a wide range of transmission rates (1.2 kbps to 2 Mbps)
- Supporting different transmission modes (namely, circuit voice, circuit data and packet data)
- Offering a wide range of services with QoS control capabilities
- Remaining compatible (and interoperable) with 2G/2.5G systems
- Possessing interoperable capability with different wireless core networks.

The CDMA2000 is designed to be backward compatible with IS-95. It has a flexible design that allows for variety of deployment options. In general, it corresponds to (N × cdma2000) where N = 1, 3, 6, 9, or 12. (When N =1, the system is compatible with IS-95 on 1-to-1 basis.) The air interface of CDMA2000 consists of:

- Tower-to-3G-terminal
  - Multi-carrier (N = 3, 6, 9, 12) forward-link
  - Direct spread forward- or reverse-link
- Tower-to-2G-terminal
  - Single carrier link
  - IS-95 channel (in reverse link with N = 1)

In conceiving 3G-systems, a number of possible *radio transmission technologies* (RTTs) have been proposed to ITU. They are based on TDMA, hybrid TDMA/CDMA and CDMA. The associated RF channels of CDMA 2000 are characterized below.

Notably, the North American cdma2000 (based on IS-95), W-CDMA of ARIB/Japan and UTRA W-CDMA of ETSI/Europe are best known proposals. (With minor differences in the type and properties of the codes used, cdma2000 and W-CDMA are similar to each other.)

The associated RF channels of CDMA 2000 are characterized as follows:

Channel	RF bandwidth (MHz)	Chip rate (Mcps)
IS-95 (N = 1)	1.25 MHz	1.2288 Mcps
Multi-carrier (N = 3, 6, 9, 12) (Forward channels only)	(N + 1) × 1.25 MHz	N × 1.2288 Mcps
Direct spreading (N = 3, 6, 9, 12) (Forward and reverse channels)	(N + 1) × 1.25 MHz	N × 1.2288 Mcps

CDMA system

In a CDMA system, all calls are handled in carrier waves, "stacked" one above the other. In order to be able to separate the calls, each call is assigned a unique code (“chip code”). In this method, each bit in the original message is given a code signal that varies, for example, 125 times faster than the original signal, resulting in a signal that is 125 times broader in frequency. This process is known as “spreading”.

The code is repeated during the entire call. The receiver can now separate its particular message by applying the code again. That is, the code is able to "neutralize itself" (“despreading”) so that the digital signal of this particular message once again becomes a slowly varying signal, while the other calls (which have different codes) remain broadband (that is remain “spread”). The underlying principle of CDMA is known as *spread spectrum* (SS) technique.

The receiver filter is designed to cut out the narrow-frequency bit containing the particular call and "taps" the compressed call 125 times stronger than the other calls, which are noticeable only as background sound. In cdma2000 the chip codes adopted are: (i) Walsh orthogonal code; (ii) complex short PN (pseudorandom)-code (quad-code or pilot code) and (iii) long PN-code.

The CDMA technology conforms to the emerging features in the wireless technology industry and as indicated above, it is specified as the target technology for the 3G-system. As subscriber growth continues to increase in the United States, in Europe and in the other parts of the world, carriers and infrastructure providers are facing a tremendous challenge in addressing bandwidth problems associated with exponential growth capacity. Hence, the next-generation wireless standard

(3G) has been conceived and expected to address the capacity issue as well as the disparate wireless standards.

There are two technologies that compete for entry into the next-generation wireless communication being nurtured *via* 3G-considerations. These are CDMA and GSM. Currently, the GSM has a large market share and the 3G-technology itself has GSM roots from Europe. It is expected, however, that 3G-options may bring together diverse wireless standards like CDMA, TDMA and PDC. As a result, considerable progress is taking place both in the GSM and CDMA markets. Notional GSM may remain dominant for voice and data transmissions up to about 100 kbps and the W-CDMA and/or cdma2000 access techniques will serve as complements to GSM for applications involving high-speed data transmissions.

The GSM system is specified for two sections, namely, network and accessing. In this architecture, the TDMA standard comes into play only in the access section and may influence a small part of the network structure. Hence, it has been observed that it will be completely within scope for operators to introduce broadband W-CDMA services in existing GSM networks.

In practice, the collective set of 3G-technologies is referred to as *3G-services*. IMT-2000 is the global body coordinating the global standard and spectrum allocation for 3G-services through the World Radio Congress (WRC). The major challenge ahead for the carriers (telecommunication providers) for entry into 3G-service lies in establishing a pragmatic path from the current 2G-services to 2.5G- and then to 3G-services. Most of the European carriers are already moving from their current infrastructure of 2G- into 2.5G-systems. Some carriers, like British Telecom (BT) and Vodafone, have relevant services complete and operational in Europe. In the United States, Sprint has been conducting a series of testing of their GPRS infrastructure.

The carriers around the world pursue two different ways to migrate from 2.5- to 3G-services. In Europe and Asia, the carriers are adopting the GPRS or EDGE route, which will provide high-speed connectivity to customers. GPRS promises to provide “always-on connection” with a practical speed of 19.2 kbps. It uses packet technology and allows operators to maximize the efficient use of their networks by allowing multiple data subscribers to share channels on the network. Just as short messaging services take advantage of spare network capacity to deliver brief text messages, packet-based technologies are an extension of the same concept. The carriers are eventually expected to move to third generation technology (UMTS) through the first decade of 2002.

In the United States and in the countries that support CDMA, the carrier will take an alternative path. They will adopt the cdma2000-1x standard, known as IS-2000, CDMA's 2.5G-technology. Two of the carriers in the United States (namely, Verizon Wireless and Sprint PCS), are already in the process of deploying these services with the help of Qualcomm and Lucent. The cdma2000-1x provides a data rate of 153 kbps suitable for a streaming video application.

The cdma2000, as mentioned above, is a spread-spectrum technology originally developed by Qualcomm. The first phase of cdma2000-1x is required to provide capability of 144 kbps data rate and the second phase is expected to offer more than 1.5 Mbps of data rate.

In Japan carriers like NTT-DoCoMo has already started working on the 3G-technologies. The much successful I-mode service uses packet data network (PDC-P). The carriers are expected to move to W-CDMA standard in 2002. Hence, the two prime candidates for the 3G-standard will be Wideband-CDMA and Broadband-CDMA (otherwise known as the cdma2000).

The eventual winning standard is expected to offer global, seamless communication. The 3G-service will also support the newly upgraded IPv6 scheme inasmuch as, many device and infrastructure vendors have raised concerns about the number of IP addresses that the current Internet protocol can support.

The hype for broadband wireless access also stems from the developments focused on multimedia networks for the future of mobile telephony. A relevant system is again being considered on broadband CDMA (W-CDMA/cdma2000) and the radio transmissions will be able to handle up to 2 Mbps. Such multimedia transmissions may support a large subscriber base for ecstatic applications ("Killer Apps") like video games. Ericsson and NTT of Japan have shown potential interests in the growth of multimedia wireless systems. Further, broadband multimedia wireless access can make the long-considered, science-fictional concept of surfing the Internet *via* mobile networks, or gaining access to video on demand (intended to watch a movie of one's choice, at any time), a reality. NTT-DoCoMo is creating a mobile network to handle these services in addition to its W-CDMA technique for radio transmission.

Demands on the next generation of mobile telephony, regardless of solution (GSM or CDMA), will be pertinent to transmitting extensive information within a very short time. In such situations, it is obvious that transmission of data at the traditional speeds of 9 or 13 kbps *via* mobile-network will be totally incompatible. As such, test systems have been developed to operate at higher bit rates starting at 64 kbps. For example, the General Packet Radio Service (GPRS) has been conceived to provide GSM with 115 kbps and bit rate enhancement efforts have been addressed for the American digital TDMA standard (D-AMPS).

A rate of transmission at 384 kbps (corresponding to the minimum requirement for multimedia) is expected to reach in GSM with existing carrier widths and upgraded modulation techniques. Furthermore, a giant-step towards 2-Mbps speeds for broadband CDMA applications will become an interesting alternative. This would, however, require broadening the carrier wave spectrum. When carrier waves are broadened, in fact, some problems involved with CDMA are eliminated, along with certain advantages of TDMA. The fading properties of CDMA will be improved, for example, while the TDMA technique may experience some problems with equalisers and high-peak effects in the mobile.

The W-CDMA differs significantly from the narrow band Q-CDMA, which operates at the transmission rate of 1.25 MHz and introduced in the United States as standard IS 95. (The Q-CDMA system is primarily intended for voice transmissions and does not carry any relative merit when compared with existing mobile systems.)

The broadband CDMA concept will require a completely new architecture for radio base stations. Until now, all base stations have been channel-oriented, or divided among a number of transceivers. The new system will use a function-oriented architecture whereby process resources are managed in a pool to facilitate

dynamic broadband. In W-CDMA, all calls are mixed together and identified with the help of codes. The system makes it possible to flexibly distribute capacity with variable channel widths. Further, W-CDMA requires a transport network able to handle this type of variable channel width. It uses the ATM technique, which is essentially a circuit-oriented, fixed packet-size connection. ATM, when used between the exchange and base stations, will be a highly economical alternative for transmitting information and the technique will provide substantial flexibility for operators. ATM is also a standard for high-speed, broadband-integrated (voice, data and video) transmissions as indicated in Chapter 1.

In summary, the WLL based on W-CDMA/cdma2000 is intended to support a variety of user requirements of a WLL namely, voice service, voice band data service (providing access to the Internet via dial-up modem), fax service, the Internet (data) access and ISDN service, in addition to the wideband services. For the W-CDMA/cdma2000 systems, the antennas that have been conceived have different forms. They support, for example, vertical and dual polarization, adjustable electrical down-tilt, and multiband operation with required bandwidth in each band. Smart antennas have also emerged for broadband applications for both fixed ands mobile systems. The objective of using smart antennas in such applications is to obviate the limitations posed by CDMA systems. Discussed below are details on antennas that are compatible for broadband applications.

8.3 BROADBAND ANTENNAS FOR WIRELESS SYSTEMS

As discussed in the previous sections, the emerging systems of wireless communications in the modern context are pushed into regimes of higher frequencies (1 GHz to 100 GHz) of operation with support capabilities of tens of Mbps to Gbps data rates. Corresponding to these high bit rates, the bandwidths of operation are also being stretched to their limits. Hence, the radio systems including antenna structures are required to facilitate these wide bandwidth (or large data rate) transmissions at gigahertz frequencies. Specific design concerns and possible geometry of pertinent antennas are presented and discussed in the following subsections.

The role of antennas in mobile communication systems is to enable a radio channel between two stations, at least one of which could be moving. In order that a base station antenna communicates with many mobile stations simultaneously, multiple channels are needed. This requires wideband operation and a function for segregating and/or combining the channels. That is, a multichannel transreceive system co-use, implicitly dictates wideband application requirements. For example, the base stations of a Japanese cellular system using the 800 MHz band utilise a single antenna for both transmitting and receiving. Hence the required bandwidth of operation of that antenna is in excess of about 7% (within the specified VSWR being less than 1.5). Moreover, when the antenna is shared by several systems, (for example, an analog mobile telephone and a digital land mobile telephone), then even a wider antenna frequency bandwidth is required. In such cases, still complying with the radio regulatory allocations, the frequency bandwidth of the antenna is stretched to about 17%.

When referencing to the antennas of mobile satellite systems, almost all the existing and future designs for GEO systems use the L-band (1.6/1.5 GHz) for

links between the satellite and mobiles. The required frequency band at this L-band operation is about 8% so as to cover both transmitting and receiving channels. In systems like IRIDIUM and ODYSSEY, the S- and L-band allocation by WARC 92 require frequency bandwidth of about 5%. To accommodate this bandwidth requirement, the narrow band antenna elements (such as the patch antenna) should be modified appropriately as described in the following sections.

8.3.1 Broadband antennas: Bandwidth considerations

The development of wireless communication systems associated with new frequencies and bandwidths requires compact, lightweight antennas. UMTS, for example, is an interesting system supporting these new activities. Further, with today's synthesised radios and the FCC frequency re-farming, there is again an increased need for wideband antennas.

Wideband antenna requirements can be viewed in two perspectives as discussed above. The first one refers to outdoor broadband access for the home and mobile users and the other is the indoor wireless environment.

The use of new frequencies (such as millimetre waveband) to accommodate the broad bandwidth, as a fractional percentage over a specified carrier, dictates an EM environment with unique propagation characteristics. The associated attenuation, delay and other fading considerations play a vital role in deciding the *signal-to-interference ratio* (SIR). Further, conceiving suitable antenna structures in such cases, as means of transreceive enablers at the physical layer level, is a crucial design consideration towards a reliable service maintaining an acceptable quality.

Not only the conceived antennas operate efficiently at these prescribed high frequencies, but they should also provide a wide passband accommodating sufficient and required tuning-in bandwidth at that operating band.

In essence, a broadband or a wideband antenna should have an acceptable frequency response over a wide frequency range. The acceptability criterion is set by the VSWR that the antenna offers to the system to which it is connected. The limit on VSWR is typically 2:1 for the transmission supported. At higher bands of frequency, the SWR limitation is lower. Considering the upper and lower bounds of the bandwidth, namely f_H and f_L, a *beamwidth factor* (BWF) can be defined as follows:

$$\text{BWF} = [f_H - f_L)/f_o] \times 100\ \% \qquad (8.1)$$

where $f_o = (f_H \times f_L)^{1/2}$ defines the centre (resonant) frequency of the antenna.

Thus, in practice, the *bandwidth* refers to a frequency range over which the antenna remains resonant at the rated VSWR. As mentioned in earlier chapters, in the cellular phone technology the VSWR is around 1.5:1 at 900 MHz. As a general rule, antennas in such systems are designed to have a bandwidth of approximately 4% of the tuned frequency. For example, an antenna tuned to 155 MHz would have a bandwidth of about 6 MHz, indicating that the antenna can be tuned from 152 to 158 MHz band and still remain under 1.5:1 VSWR performance.

The wideband operation means achieving a bandwidth much in excess of the "4% rule". For example, considering a midfrequency of operation at 2 GHz, the

task of realising a bandwidth from 1.5 GHz to 2.55 GHz (meaning 52% for the VSWR being less than 2) becomes the wideband design objective of such antennas.

8.4 WIDEBAND TECHNIQUES IN WIRELESS ANTENNA DESIGNS

In wireless communications, as discussed in Chapter 4, the microstrip patch antennas used allow low profile and compact designs. However, they suffer from narrow bandwidth characteristics.

Typically, a standard patch antenna offers only 1 to 2% bandwidth. Therefore, for wideband applications, techniques are needed to beef up the bandwidth of patch antennas. Some possible methods are as follows:

- Designing the antenna with low unloaded, Q
- Implementing a double resonance (stagger-tuning) over the bandwidth of interest
- Adopting a wideband impedance-matching network
- Appropriate arranging of the elements of an array antenna.

The aforesaid techniques are briefly discussed below.

8.4.1 Patch antenna with low, unloaded Q substrate

For a dominant mode-excited patch antenna, its equivalent circuit corresponds to a parallel resonant circuit (viewed at any signal terminal). The constituent reactive elements of this equivalent resonant circuit can be derived on the basis of a (resonant) *cavity model*. This cavity model, in essence, treats the region between two parallel conductor planes (consisting of the patch radiator and the ground-plane) as a resonant cavity.

By developing an appropriate EM wave equation inside the cavity and applying relevant boundary conditions, it can be shown that the electromagnetic fields in the cavity region are completely determined as a function of the input current (at the terminals) [8.6]. Once the terminal currents are known, the total radiation field can be calculated as a superposition of the contributions from all the modes (as described by Huyghen's principle) relevant to the magnetic wall along the periphery of the patch (regardless of its shape).

The associated powers in a patch antenna refer to the total power radiated and the total power dissipated in the conductor as well as *via* dielectric losses. Hence, for a lossy (dissipative) resonant structure depicting an antenna, its unloaded Q (namely, the Q_o) is related to the relative bandwidth B_r (for an input VSWR < ρ). It is given by:

$$Q_o B_r = \left[(\beta\rho - 1)(1 - \beta/\rho)\right]^{1/2} \qquad (8.2)$$

where β is the coupling coefficient defined by G_o/G' (where G_o represents the conductance of the transmission line terminated with the lossy resonator and, G' is the conductance of the patch antenna). The relative bandwidth (B_r) refers to the

bandwidth normalised with respect to the resonant frequency. The maximum value of $Q_o B_r$ is given by:

$$(Q_o B_r)_{max} = (\rho^2 - 1)/2\rho \qquad (8.3)$$

which is realised under the condition that β is equal to β_o given by $(\rho^2 - 1)/2\rho$.

The unloaded Q (namely, Q_o) can be functionally represented by a relation $Q_o = F_1(\sqrt{s}/t, \varepsilon_r)$ at the resonance frequency of f_o, which is specified by the following relation: $\sqrt{s} f_o = F_2(\sqrt{s}/t, \varepsilon_r)$, where s is the area of the patch, and ε_r and t are the substrate dielectric constant and thickness respectively.

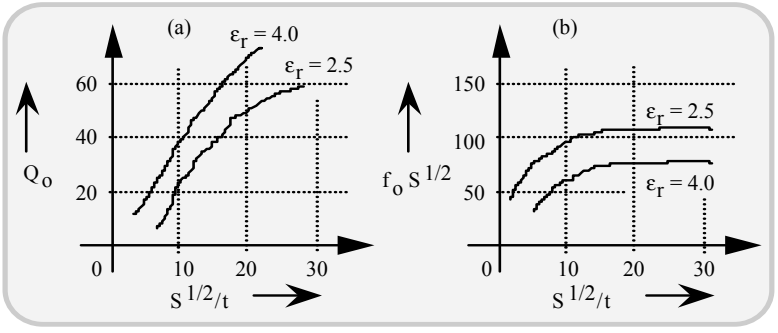


Fig. 8.2 (a) Q_o *versus* $\sqrt{s}/t$ and (b) $f_2\sqrt{s}$ *versus* $\sqrt{s}/t$ for a square patch antenna (where f_2 is the second-order resonant frequency)

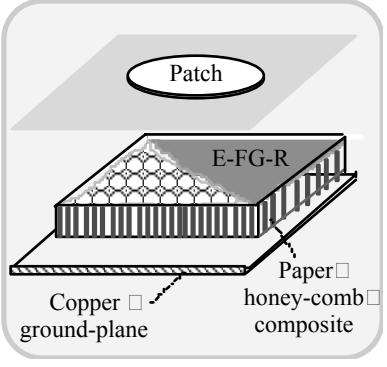


Fig. 8.3 Wideband antenna with thick substrate made of artificial dielectric substrate with $\varepsilon_r < 1$. (E-FG-R: Epoxy fibre-glass reinforcement)

Design charts based on measurements can be used to explicitly specify the empirical functions F_1 and F_2. Typically such design charts (for rectangular/square and/or circular patch antennas) indicate a monotonic increase of Q_o with respect to $\sqrt{s/t}$ with ε_r being the parameter. Typical variations of Q_o and $f_2\sqrt{s}$ as functions of $\sqrt{s/t}$ for a square-patch are shown in Figure 8.2 (where f_2 is the second-order resonant frequency). From Figure 8.2(a), it can be observed that Q_o can be reduced by using a thick substrate made of a low dielectric constant material. A method of realising a lightweight structure with a thick substrate and having a low ε_r is to use an *artificial dielectric* [8.7].

An example of artificial dielectric refers to a paper honey-comb structure constituting a composite material that can emulate a dielectric medium with a low relative permittivity ($\varepsilon_r \rightarrow 1$). Normally, the core of such an artificial dielectric substrate is sandwiched between epoxy fiber-glass for reinforcement as illustrated in Figure 8.3. One of the authors of this book has described in [8.7] various types of composite dielectrics and some of which could be judiciously synthesised for the test structure under discussion for use as the substrate material (*in lieu* of the paper-honey comb composite illustrated in Figure 8.3).

A typical patch antenna structure, as sketched in Figure 8.3, can yield a bandwidth of about 10 %.

8.4.2 Vertically stacked patches

These refer to using multiple patches stacked vertically to improve the bandwidth performance of microstrip antennas. For example, a two-layer configuration can be adopted with the microstrip feed-line etched on the lower layer (of alumina substrate), and a resonant patch located on the top layer (upon a polyguide substrate). Also, two layers of polyguide can be used, in which case, it would allow two vertically stacked patches as illustrated in Figure 8.4.

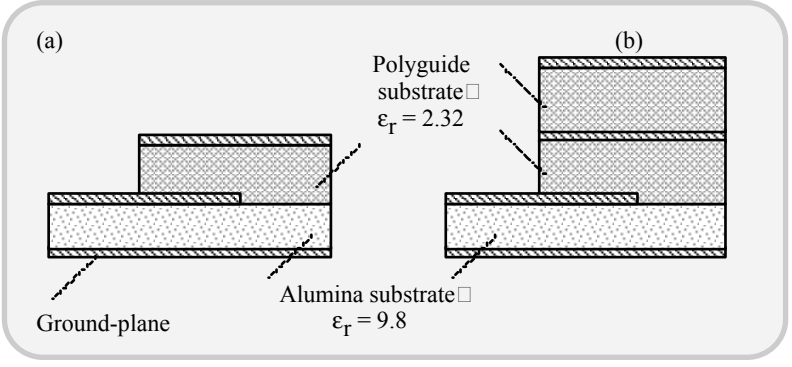


Fig. 8.4 Broadband patch antennas: (a) Two-layer structure and (b) three-layer structure

The antenna of Figure 8.4 yields a bandwidth of 18% for 10-dB return loss. Typical examples are discussed in [8.8, 8.9]. In the relevant antenna structures using multilayer dielectric as illustrated in Figure 8.4, use of high permittivity alumina layer allows easy integration with circuits.

The multilayer concept has been used by Anandan and Nair [8.10] to realise a compact broadband microstrip antenna with parasitic elements. This system consists of a number of parasites, which are gas-coupled to a driven-patch. A rectangular patch is considered in [8.10] as an example with RT Duriod as the substrate material ($\varepsilon_r = 2.2$).

The active patch is fed from its nonradiating edge *via* 50-ohm coupling so as to resonate at 870 MHz. With the addition of parasites gas-coupled to the nonradiating edges, a bandwidth performance of 6% (with VSWR < 2.0) is realized over the frequency range of 854 MHz to 898 MHz. The geometry of the antenna described in [8.10] is as shown in Figure 8.5.

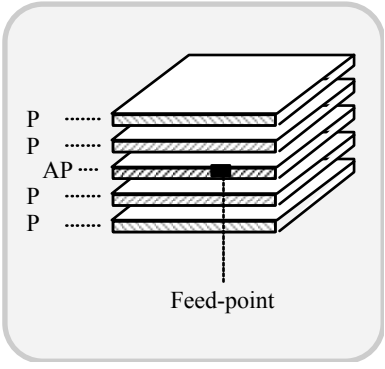


Fig. 8.5 Broadband patch antenna with gas-coupled parasites (AP: Driven-patch and P: parasitic elements)

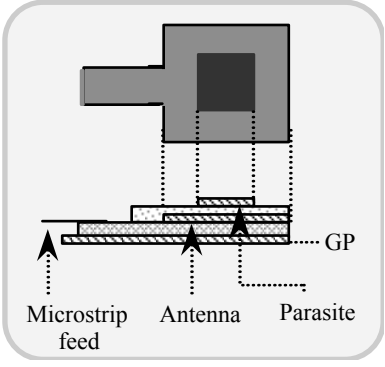


Fig. 8.6 Radiator patch plus parasite based broadband antenna with the parasitic element functioning as a director

The underlying concept behind obtaining increased bandwidth *via* overlaying a parasitic element that acts as a director is explained in [8.11]. The structure is shown in Figure 8.6.

8.4.3 Single-plane multiple patches antenna

These wideband antennas configure multiple resonators facilitated in the same plane. The key concept is the same as for vertically stacked patches, namely, improvising stagger-tuned resonances, which can lead to wider bandwidth of operation. That is, a set of patches can be designed such that, each patch is tuned at a centre-frequency and the resonant curves of these patches successively overlap, thereby juxtapositioning the tuned frequencies in a staggered fashion. Hence, the overall resonance curve is rendered broadband and it has a rippled top with a ripple factor that can be designed to a specification. The relevant design is similar to synthesising a bandpass filter with a prescribed extent of ripple factor.

Further, by locating a capacitively-coupled, quarter-wave version of short-circuited parasitic patches at the radiating edges (of the central patches) will enable doubling the bandwidth of a single, rectangular patch.

An example of this structure is described in [8.12] where the antenna characteristics are explained in terms of an antiphase mode of a pair of coupled resonators; it is shown that the bandwidth improvement is independent of the coupling capacitance. The relevant structure is illustrated in Figure 8.7.

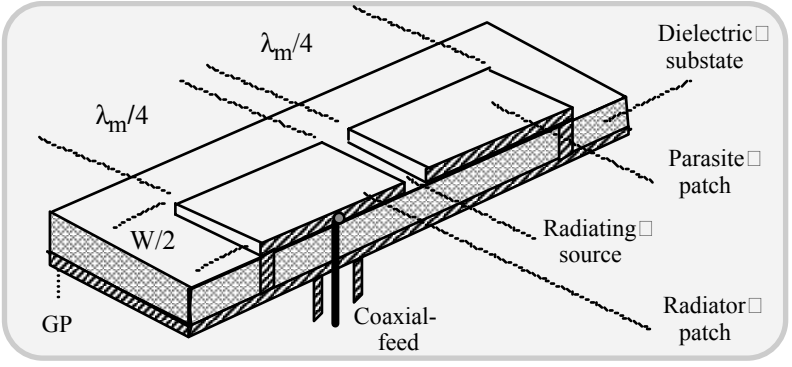


Fig. 8.7 Antenna made of an excited patch plus parasites, all set on a single plane

The $\lambda_m/4$-rectangular patch with a short circuit along one edge (as shown in Figure 8.7) has a parasitic excitation *via* capacitive coupling with the radiating edge. Near the resonant frequency, the input impedance of a patch antenna corresponds to a parallel GLC circuit. Considering the radiation between the two patches of the structure shown in Figure 8.7, the radiation conductance G can be replaced by admittance Y_e corresponding to the patch edge discontinuity and the radiating slot. This added admittance leads to multiple resonants leading to a wideband

performance. The test structure of Figure 8.7 can also be modified to yield circularly polarized radiation by means of the geometry shown in Figure 8.8.

The third method of increasing the bandwidth of a microstrip antenna is by incorporating a wideband impedance matching network. This is based on the fact that for conventional microstrip patches, the bandwidth limitation is imposed by reflection loss and not by radiation characteristics. That is, the restricted bandwidth observed is essentially an impedance-specified bandwidth. Therefore, bandwidth improvement can be accomplished (by a factor of 2 or 3, for example) *via* a suitably designed impedance matching network. Relevant details are presented in [8.13].

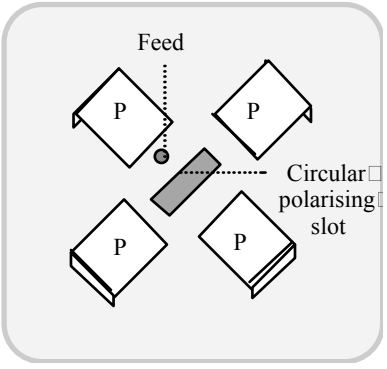


Fig. 8.8 Circularly polarised antenna made of multiple parasitic patches on a single plane

Assuming that no higher-order modes are excited on a microstrip resonator, an RLC circuit can be prescribed to depict the resonating structure whose bandwidth (when the feeder impedance is equal to the resonant input impedance), is given by:

$$\mathrm{BW} = \frac{(S-1)}{\sqrt{S}} \times \frac{1}{Q} \times 100\% \qquad (8.4)$$

where S is the specified maximum VSWR and Q is the *quality factor* of the resonator.

The principle of impedance-matching refers to match as best as possible a frequency-dependent load impedance to the source (or feeder) over a prescribed frequency range. The bandwidth and degree of match are, however, contradictory requirements. The maximum bandwidth that can be achieved for a specified degree of match depends on the type of load. For a RLC resonant circuit, this maximum bandwidth ($\mathrm{BW_m}$) is given by:

$$\mathrm{BW_m} = \frac{\pi}{\ln[(S+1)/(S-1)]} \times \frac{1}{Q} \times 100\% \qquad (8.5)$$

An impedance-matching network is essentially a bandpass filter. The reactive part of the load acts as the first resonator of the filter. The bandwidth obtained will be maximum for an ideal bandpass characteristic, that is, when a constant return-loss within the passband (and a 0-dB return-loss elsewhere) are realised.
In practice, a Tchebyscheff filter of finite order can be synthesised to approximately realize an ideal filter, and, the bandwidth improvement *via* matching can be specified by a factor F, given by,

$$F = BW_m/BW = \pi\sqrt{S}/\{(S-1)\ln[(S+1)/(S-1)]\} \qquad (8.6)$$

Appropriate electromagnetic coupling can be adopted to realise the impedance matching conditions. For example, Pozar and Kaufman [8.14] recommend a *proximity coupling* technique. Improvement in bandwidth can also be obtained by means of magnetic coupling of microstrip dipoles using parasitic metallic straps attached to the feeding line [8.15].

8.5 INDOOR BROADBAND WIRELESS ANTENNAS

A strategy to support 100 Mbps rates of multimedia applications in ATM platforms, where wireless systems conceived for indoor deployment refer to using millimetre wave bands (20-60GHz). The target system is a microcellular ATM LAN with transport bit rates up to 160 Mbps per microcell. For use in such systems, small, lightweight and affordable antennas are required. Typically, such antennas should provide for omnidirectional pattern in the horizontal plane and beam-tilt directivity in the vertical plane [8.16].

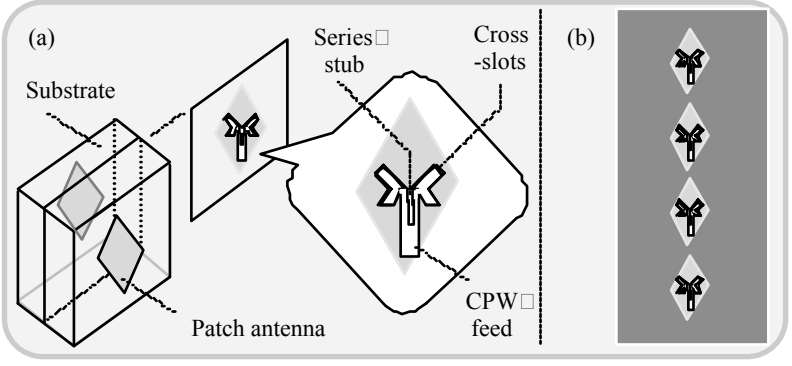


Fig. 8.9 Basic configuration of dual-polarised, single, omnidirectional and array antennas. (a) Element excitation and (b) array configuration

Specific to the indoor environment, the multipath effects have crucial impact on the system performance. In such cases, polarisation diversity can offer a mitigatory solution on the detrimental fading effects arising from multipath phenomenon as

discussed in Chapter 7. Further, if millimetre (mm) wave transmissions are deployed, the associated propagation attenuation would be significantly severe calling for more transmission power so as to cover the whole extent of the cell zones involved. Also, for the wideband transmissions supported at the mm-bands, the multipath effects become even more severe.
Therefore, to cope with the situation, as above, it becomes necessary to use dual-polarised schemes (conceived through array systems and configured for polarisation diversity) so as to improve the immunity against flat-fading. For example, a dual-polarised coplanar waveguide (CPW) fed antenna is illustrated in Figure 8.9. It consists of a back-to-back square-patch antenna excited by a CPW *via* cross-slots on the ground plane. It can produce an omnidirectional pattern with dual polarisation.

In configuring the antenna system of Figure 8.9, a select group of basic antenna elements described in Chapter 4 can be used. They can be arranged as a linear array as indicated in Figure 8.9. The relative permittivity of the dielectric used as the substrate should be high (≈ 10). Further, a thin substrate would allow easy integration of the antenna with monolithic circuits. The patterns of this category of antennas are nearly omnidirectional both in H- and E-planes. The matching requirement for VSWR ≤ 2 can be satisfied over a bandwidth of about 3% around the resonant frequency.

8.5.1 An angular diversity antenna system for broadband WLAN

For broadband WLAN applications a circular array of *bent stacked slot antennas* (BSSA) described by Fassetta and Sibille [8.17] can also be used. Such an antenna enables an *angular diversity*. This diversity again can provide mitigation *vis-à-vis* multipath propagation and intersymbol interference that may occur in indoor wireless communications.

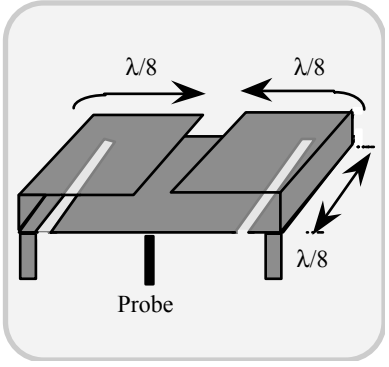


Fig. 8.10 BSSA geometry. (The antenna is placed above the ground-plane and is fed by a probe. Two vertical metallic strips provide support to the antenna)

The relevant compact and low profile BSSA for a circular array implementation is illustrated in Figure 8.10. The radiating element is intended for mounting on a ground-plane with a probe and two vertical metallic strips. Hence, it acts as a monopole (on a ground-plane) with a main-lobe in the end-fire direction. There are two λ/2-slots (slightly differing in their lengths) stacked below the radiator to provide two resonant frequencies (high and low) so as to enable a high bandwidth of about 30%. The planar structure is comprised of three pieces of lengths λ/8, λ/4 and λ/8 (in a bent form), totalling λ/2. This geometry allows a compact design. The feeding is done via coaxial line through a SMA connector at the ground-plane with a wire probe connected to the lower metallic plane of the stack.

To achieve angular diversity (in six states), a set of six elements can be disposed symmetrically on the periphery of a circle. The operation consists in breaking the circular symmetry of the array by connecting two BSSAs to the source and the others are connected to a pure inductive reactance. These inductively loaded radiating elements would play the role of reflectors as per the Yagi-Uda antenna array principle; and, the reflectors constitute proximity-coupled and passively-excited parasites. Beam-switching is enabled by circular permutation of the source/reactive loading configuration. The results presented in [8.17] show the 3-dB beamwidths of the array in azimuth and elevation as 15^{o} and 50^{o} respectively and a directivity of 7.7 dBi (as compared to 4 dBi of a single radiator). Further, the design illustrated in [8.17] refers to a 5.2 GHz operation. The backlobe of the array is reported to be less than –10 dB. Relevant version of the array is compatible for bandwidth requirements specified for the HIPERLAN.

8.5.2 Bowtie-patch antenna/arrays for broadband indoor wireless communications

A *microstrip bowtie antenna* (and an array) based on the design of equilateral triangular patches realised with a CPW feed are described in [8.18]. It is a compatible structure for 5.5 GHz, 20 Mbps transreceive C-band radio modems used in in-door applications.

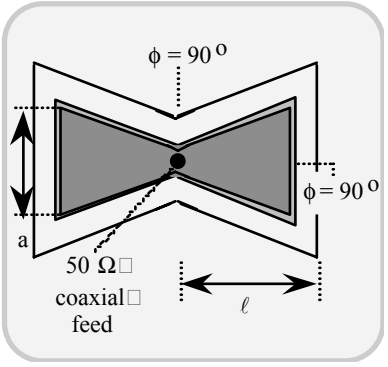


Fig. 8.11 Coplanar waveguide bowtie antenna with a coaxial feed arrangement

In radio modem based in-door communications, apart from easy set-up, reconfigurability, portability and mobility for phones, terminals and/or computers connected within the network, transmission reliability should be answered against the possible delay-spreads experienced at the receiver (as a result of multipath propagation). This is of concern especially when such delay-spread exceeds $(1/10)^{th}$ of the symbol period (resulting in intersymbol interference). Under this situation, an array-based on *coplanar-waveguide bowtie antenna* (CPWBA) can be designed for relevant broadband radio modem applications.

In a CPWBA structure, the antenna corresponds to a microstrip structure having an open-ended transmission line with a cross-sectional expansion. A typical layout of a CPW bowtie antenna fed by a 50-ohm SMA coaxial feed set on a dielectric substrate is illustrated in Figure 8.11.

The CPWBA consists of a finite ground-plane and a pair of equiangular microstrip patches. The resonant frequencies of such a patch for z-independent TM modes satisfy a perfect magnetic-wall boundary condition given by:

$$f_{mn} = \frac{2\ell}{3a\sqrt{e_r}}\left(m^2 + mn + n^2\right)^{1/2} \qquad (8.7)$$

where m and n are the orders of various resonant modes, ε_r is the relative dielectric constant of the substrate, and a and ℓ are the dimensions shown in Figure 8.10. If the conductor thickness t is taken into account, a in equation (8.7) is modified to $(a + t/\sqrt{\varepsilon_r})$.

A variety of arrays can be configured using the CPWBA as illustrated in Figure 8.12.

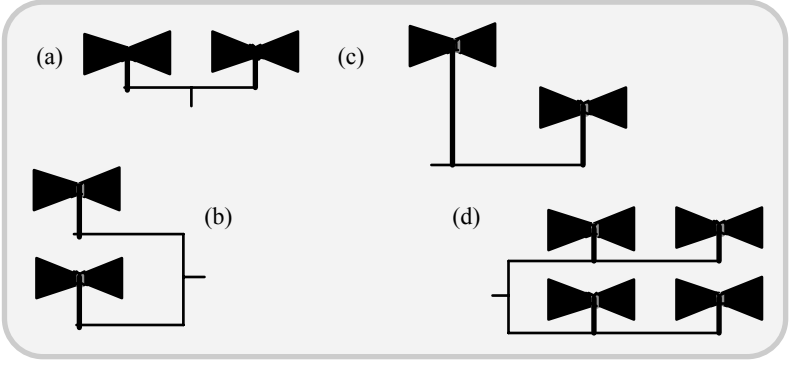


Fig. 8.12 CPWBA-specific arrays: (a) (1 × 2) array; (b) (2 × 1) array; (c) offset (2 × 1) array and (d) 2 × 2 array

Several other bowtie-patch antenna configurations are also possible considering their suitability with respect to broadband operation, adjustable beamwidth, and multiple-beam (radiation diversity) capability for use in broadband in-door wireless communications.

Further, adopting a finite-ground CPWBA has certain favourable features: Such structures are compact, allow easy integration with active devices, have wide bandwidth (about 10%), and pose desirable radiation characteristics.

The CPWBA can be coaxial-fed from its apex and matched to the operation band. The resonant slot-length of this antenna is around three times that of the guide wavelength. The realisable gain for the CPWBA antenna is about 8 dB in the operations such as at C-band.

Corporate microstrip feed networks matched to the antenna input impedance can be used in the realisation of (1×2), (2×1), and (2×2)-element MBA arrays. Typical gains for such arrays may vary around 14 dB (with 1×2-element) and 18 dB (with 2×2-element). The 2:1 VSWR bandwidths could lie in the range 9 % to 11 %. Corresponding beamwidths may vary between 15^o and 85^o, which permit desirable multipath minimisation. The radiation pattern can be optimised by adjusting the configuration, mainly with the spacing between the elements. This is achieved either by a (1×2)-element array or by an offset design of a (2×1)-element array, which can provide higher gains and slightly increased bandwidths.

8.5.3 Broadband antenna with polarisation diversity for WLAN applications

As discussed earlier, polarisation diversity offers a mitigatory solution against signal fading in the complex indoor environment due to multipath effects. A compatible antenna can be designed by a modified (printed) strip monopole plus a hybrid feed system illustrated in Figure 8.12.

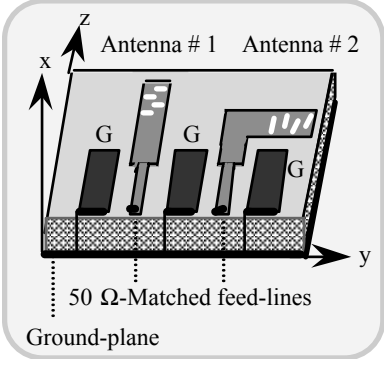


Fig. 8.13 Modified printed strip monopole antenna with a hybrid feed system

The two antennas, #1 and #2, in Figure 8.13, are two separated strip monopoles, which have interdigitised slits. These meandered-slits are kept orthogonal to each other to achieve polarisation diversity. The accompanying feed system is a hybrid version. It consists of a CPW having a 50-ohm input impedance connected to a 50-ohm microstrip line. The antenna is fed by the CPW cascaded with 50-ohm microstrip line-feed connected to the front-end RF circuits.

The size of the ground-plane of CPW is made much smaller than the antenna size so that the effect of this finite ground-plane on impedance-matching is rendered capacitive. This capacitive reactance provides conjugate compensation for the inductive reactance introduced by the interdigital section of the antenna. This helps towards necessary impedance-matching requirements and allows broader bandwidth of operation. The bandwidth is controlled by the number of interdigital meanders implemented and, by proper choice of the size of the ground-plane sections in the CPW. In [8.19], it is shown that a wideband bandwidth as large as 16 % is feasible at 2.4 GHz operation with a WLAN PCMCIA card.

The test antenna described in [8.19] has the antenna #1 designed for 2293-2580 MHz operation with five interdigitised slits offering a 10-dB return-loss specified impedance-bandwidth of 418 MHz (16.8 %); antenna #2 has six interdigitised slits enabling a 10-dB return-loss specified impedance-bandwidth equal to 645 MHz (26 %). The relevant ground-plane dimensions of the CPW are as follows: For antenna #1: $(0.04 \times 0.096)\lambda_o$ either sides. For antenna #2: $(0.04 \times 0.096)\lambda_o$ and $(0.06 \times 0.096)\lambda_o$ on its sides. Here λ_o is 12.5 cm corresponding to the frequency of 2.4 GHz

The measured E- and H-plane patterns for the test antennas of [8.19] indicate the following: For antenna #1, the E-plane (xy-plane) pattern is almost omnidirectional and the H-plane (xz-plane) pattern shows a null at $\theta = 0$ implying a dipole pattern with null axis parallel to the zy-direction. For antenna #2, the above considerations are just reversed, with almost omnidirectional H-plane (xz-plane) pattern and the E-plane (xy-plane) having a null at $\phi = 0$ implying, again, a dipole pattern with null axis parallel to the y-axis.

8.6 FRACTAL ANTENNAS

As well known, antennas are “tuned devices” and hence are essentially narrowband devices. Their radiation characteristics are dependent on the ratio of antenna size to operating wavelength. This resonant nature of an antenna of fixed size greatly influences the antenna parameters such as gain, input impedance, pattern shape and minor-lobe levels as a function of frequency. These parameters show significant variations when the operating frequency is changed. For example, even with a simple linear dipole of a given length, each time when the frequency is doubled, several sidelobes would appear distributing the radiated power into space.

Another implication on the practical size of an antenna is that, a minimum size (relative to wavelength) is required to operate efficiently. However, considering antenna requirements in wireless communication systems, compact size and unique braoadband requirements are needed in some applications as discussed earlier. A method of achieving this performance consideration refers to designing the so-called *fractal antennas* and *arrays*, which can counterpoise the

size-related issues *versus* frequency in antenna designs. Further, such structures offer a rich variety of geometrical shapes with some interesting properties. The underlying principle behind fractal antennas is explained below.

Fractals are conceptually geometrical shapes exhibiting *self-similarity* and *fractional dimension*. An object is said to be self-similar when it is entirely constructed (replicated) by many copies of itself at different scales. The construction of an ideal fractal shape corresponds to having an infinite number of copies of itself with a characteristic scale factor of 3 relating the sizes of them all. It can be mathematically proved that such an object has a dimension of D ~ 2.72 means that, a fractal shape is half-way between a 3D- and a 2D-object. Hence it corresponds to an object with a surface that fills a 3D-space better than any classical Euclidean surface.

The fractal design of antennas conforms to a self-similar structure (which contains many replications of itself at several scales), and operates in an identical manner over a stretch of several wavelengths. That is, the antenna preserves similar radiation characteristics through several bands of frequency. Also, the space-filling properties of some fractal shapes allow the fractally-shaped small antennas to consume small surrounding space; hence the concept of fractal antennas and arrays leads to the possibility of multifrequency operations and antenna size reduction feasibility.

In the 1970s Mandelbrot [8.20] introduced the term “fractal"” to describe a family of geometrical objects that defy the traditional rules of the Euclidean geometry. At the same time, however, such weird fractal shapes are, in fact, among the most common forms in nature. This “weird” geometrical consideration is beneficially adopted in fractal antenna developments.

Using the fractal theory, the fractal multiband antennas developed are due to Puente et al. [8.21]. Subsequently, the potential use of fractal theory to design multifrequency/multiband arrays was also indicated by Puente [8.22]. Eventually, the Universitat Politècnica de Catalunya applied for the patent [8.23] on fractal and multifractal antennas that retain the same radiation parameters over several frequency bands.

In reference to fractal radiators, a *fractal element antenna* can be defined as a structure, which is shaped in a fractal fashion, either through bending or shaping a volume, or introducing holes. Popular fractal antennas that were developed correspond to the following types: Sierpinski-sieve monopole [8.24, 8.25], fractal-loop antenna [8.26], resonant fractal-loop, fractal-tree antenna and fractal-dipole. Corresponding fractal shapes that have been adopted in the antenna designs are the Sierpinski triangle (sieve), Mandelbrot tree, Koch curve, Koch island etc.

With the above geometrical attributes, the fractal antennas can be designed such that, even with the size shrunk from two to four, the radiators adequately retain invariant radiation properties. It means that a fractal antenna is inherently broadband.

An example of a Koch curve dipole is illustrated in Figure 8.14. The Koch dipole shown is made of the geometry of replicating self-similar squares. Because of the unique geometrical structure, this fractal dipole, (as well as other fractal antennas) do not have the same impedance characteristics as their conventional counterparts. For example, by transforming the input impedance to 95 ohms for the

third iteration Koch square curve, an extremely wideband match can be achieved – from 5 GHz to 20 GHz.

Fractal concepts and configurations have also been extended to develop arrays. Specifically, such arrays are conceived to realise certain advantages over the conventional arrays such as broadband/multiband operation, low sidelobe levels, desirable form-factors and size.

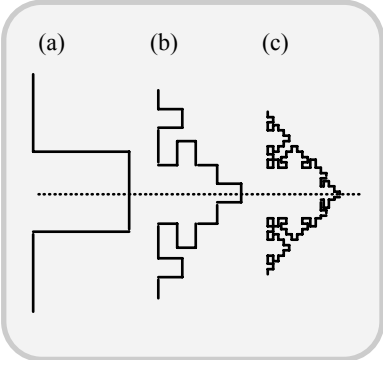


Fig. 8.14 Koch dipole based on the fractal Koch square geometry

8.7 CONCLUDING REMARKS

As indicated in this chapter, special considerations and novel design efforts are required to achieve a broadband performance with wireless communication antennas. The needs for such antennas are yet to be seen in a broader scale since the associated 3G-systems are still in their growing stage. The broadband performance concurrently specifies the use of the high-end microwave spectrum and/or mm wave frequencies in the wireless transmissions. It means that distributed effects will be more pronounced and the dimensions will become more critical. Further, the choice of dielectrics will be more constrained in view of dispersion and lossy consideration [8.7].

The basic systems described in this chapter, however, outline the scope to realise feasible structures and the design directives. There are a number of conceivable broadband antenna designs. For example, an interesting design to obtain large bandwidth characteristics using a rectangular microstrip-fed triangular patch placed in a circular slot is described in [8.27].

REFERENCES

[8.1] Sari, H.., Broadband radio access to homes and businesses, *IEEE Computer Networks*, 1999, vol 31, 379-393

[8.2] Calhoun, G., *Wireless Access and Local Telephone Network*, Boston, MA: Artech House, 1992

[8.3] Papazian, P. P., G. A Hufford, R. J Achatz, and R. Hoffman, Study of the LMDS radio channel, *IEEE Trans. Broadcasting*, June 1997, vol. 43(2), 175-184

[8.4] Yoon, Y. K., and M. Ulema: A wireless local loop system based on wideband CDMA technology, *IEEE Commn. Magazine*, October 1999, vol. 37(10), 128-135

[8.5] S. Sheng and R. W. Brodersen, *Low-Power CMOS Wireless Communications – A Wideband CDMA System Design*, Norwell, MA: Kluwer Academic Publishers, 2001

[8.6] Hirasawa, K., and M. Haneishi eds., *Analysis, Design and Measurement of Small and Low-profile Antennas*, Artech House, Boston, MA: 1992 pp. 102

[8.7] Neelakanta, P. S., *Handbook of Electromagnetic Materials*, CRC Press, Boca Raton, FL: CRC Press, 1995

[8.8] Hall, P. S., C. Wood and C. Garrett, Wide bandwidth microstrip antennas for circuit integration, *Electron Letts*.,1979, vol. 15(4), 459-460

[8.9] Sabban, A., A new broadband stacked two-layer microstrip antenna, *1983 IEEE AP-S Int. Symp. Antenna and Propagation Digest*, 63-66

[8.10] Anandan, C. K., and K.G. Nair, Compact broadband microstrip antenna, *Electron Letts*., 1986, vol 22(20), 1064-1065

[8.11] Dubost, G., and G. Beauquet, Patch antenna bandwidth increased by means of a director, *Electron. Letts*., 1986, vol. 22, 1345-1347

[8.12] Wood, C., Improved bandwidth of microstrip antennas using parasitic elements, *Proc. IEE*, 1980, vol. 127, Pt. H, 231-234

[8.13] Pues, H. F., and A. R. Van de Capelle, Wideband impedance-matched microstrip resonator antenna," *Proc. IEE Second Int. Conf. Antennas and Propagation*, Part.1, 1981, 402-405

[8.14] Pozar, D. M., and B. Kaufman, Increasing the bandwidth of a microstrip antenna by proximity coupling, *Electron. Letts*., April 1987, vol. 23(8), 368-369

[8.15] Katehi, P. B., and N. G. Alexópoulos, A bandwidth enhancement method for microstrip antennas, *IEEE Trans. Antennas Propagat*., Jan 1987, vol. AP-35(1), 5-12

[8.16] H. Iwasaki, A back-to-back rectangular patch antenna fed by a CPW, *IEEE Trans. Antennas Propagat.*, October 1998, vol. 46(10), 1527-1530

[8.17] Fassetta, S., and A. Sibille, Switched angular diversity BSSA array antenna for WLAN, *Electron. Letts*., April 2000, vol. 36(8), 702-703

[8.18] Uysal, S., M. S. Leong and C. H. Ng, Bowtie patch antennas and simple arrays for wireless indoor communications, *IEEE Trans. Microwave Theory Tech*., June 1999, vol. 47(6), 738-745

[8.19] S.T. Fang, A novel polarization diversity antenna for WLAN applications, *Proc. IEEE Antennas and Propagation Soc. Int. Symp. 2000*, vol.1, 282-285

[8.20] B.B. Mandelbrot, B. B., *The Fractal Geometry of Nature*, New York, NY: W.H. Freeman and Company, 1983

[8.21] Puente, C., J. Romeu, R.Pous, X. Garcia, F. Benítez, Fractal multiband antenna based on the Sierpinski gasket, *Electronics Letts.,* January 1996, vol.32(1), 1-2

[8.22] Puente, C. and R. Pous, Fractal dsign of multiband and low side-lobe arrays, *IEEE Trans. Antennas and Propagat*., May 1996, vol.44(5), 730-739

[8.23] Puente, C., R. Pous, J. Romeu, X. García, Antenas Fractales o Multifractales, *Invention Patent, nº: P-9501019*. Presented at the Oficina Española de Patentes y Marcas. Owner: Universitat Politècnica de Catalunya, May 1995

[8.24] Song, C. T. P., P. S. Hall, H. Ghafouri-Shiraz and D. Wake, Sierpinski monopole antenna with controlled band spacing and input impedance, *Electronics Letts*., 1999, vol.35(13), 1036-1037

[8.25] Song, C. T. P., P. S. Hall, H. Ghafouri-Shiraz and D. Wake, Fractal stacked monopole with very wide bandwidth, *Electronics Letts.,* 1999, vol.35(12), 945-946

[8.26] Song, C. T. P., P. S. Hall, H. Ghafouri-Shiraz and D. Wake, Multi-circular loop monopole antenna, *Electronics Letts*., 2000, vol.36(5), 391-393

[8.27] Jang, Y. W., Characteristics of a large bandwidth rectangular microstrip-fed inserted triangular patch in a circular slot antenna, *Microwave J*., May 2002, vol. 45(5), 288-298

APPENDIX 8.1

Details on LMDS Evolution

A8.1 Salient specifications on LMDS

Spectrum

- Block A [27.5 - 28.35, 29.10 - 29.25 GHz, 31.075 - 31.225 GHz]
- Block B [31.00 - 31075 and 31.225 - 31.300]

Services

- Voice/data
- VoD/real-time video
- Local exchange telephone service
- Internet access
- Other broadband services

Range

3 – 5 miles (Cell size is, however, *rain-fade* limited)

Locales designated for wireless local access

- Homes
- Business premises
- Schools/libraries
- Healthcare providers
- Rural communities

Network

LMDS architecture is cellular-like. But the services are for fixed end-entities and not for mobile units. The networking involves placing "hub" transreceivers on towers spaced a few kilometers apart. Each hub is at the center of a cell, serving customers with small roof-top antennas. The hubs could be connected with optical fibers and/or wireless backhaul links. Both ATM and IP transport can be supported. Point-to-point and point-to-multipoint configuration are possible.

Extended LMDS bandwidth

2.2-GHz bandwidth for interactive data in several channels. (For example, 550 lines at 4 Mbps each.)

Frequency reuse

Feasible since the (28 to 31)-GHz links can support short ranges. If required, polarisation diversity can be used.

Other international deployments

Canada	1-GHz BW/28GHz spectrum LMCS (*Local multipoint communication system*)
UK	40 GHz, *multipoint video distribution system* (MVDS)
Other countries	Argentina, France, Korea, the Philippines, Romania, Venezuela, Poland etc.

Standardisation

In the United States, standardisation of BWA technology is done by National Wireless Electronic Systems Test-bed (N-WEST)/US Commerce Department Agencies (NIST and NTIA) and participating industries at large. Further, the BWA Study Group's projected studies towards BWA standardisation are as follows:

Wireless links	Use of microwave and millimetre wave spectra (around 30 GHz) (with a general scope for using 10 to 66 GHz)
Spectrum	(Licensed)
Locale	Metropolitan in scale
Provisioning	By public networks to paying customers
Architecture	Point-to-multipoint transmissions between tower-mounted and roof-top installed antennas
Service:	Heterogeneous traffic and QoS-specified applications pertinent to data, video and voice
Bandwidth type	Broadband (> 2 Mbps)

The LMDS concept was initially motivated by digital TV applications. Standardising efforts on digital TV was first initiated in Europe with the establishment of the *digital video broadcasting* (DVB) project by the European broadcasting union. The technical specifications given by the DVB project were passed over to the European Telecommunications Standard Institute (ETSI) for adoption and publication of standards. At that stage the focus was on using microwave transmission for digital TV transmissions. The DVB then suggested the relevant standard for short-range millimeter wave radio systems. It initially called the system as *multipoint video distribution* system. Another international body called the *digital audio video council* (DAVIC), consisting of major network operators, service providers and consumer electronics, telecommunications and computer industries also has contributed its input to evolving LMDS. (But DAVIC, as such, is not a part of any official standard-making body.)

DVB specifications

In order that LMDS benefits from the mass-market of broadcasting satellites, specifications for LMDS downlink channel have been kept same as those of "direct to home" satellite services. Both LMDS and direct-home-satellite systems use the QPSK modulation and concatenated *forward error correction* (FEC) coding scheme with a convolution inner-code and a Reed-Soloman outer-code. The transmission frame is based on MPEG-2 transport data stream. The outer-code supports 188 information bytes. It has a block length of 204 bytes and can correct up to 8-byte errors per each block. This code is obtained by shortening the RS-(255, 239) Reed-Solomon code. A convolution interleaver (with interleaving depth of I = 12) is inserted between inner and outer encoders. This is done in order to distribute uniformly the errors, which occur, by bursts at the video output in the receiver.

The input data bytes in the interleaver are fed in a cyclic fashion to 12 parallel branches made of simple, first-in-first-out (FIFO) shift-registers. Delays starting from 0 are increased by multiples of 17 with the second branch having a 17-byte delay and so on. Suppose a convolution interleaver of length N and depth I comprises of I branches. Then the I^{th} branch includes a delay of $(I - 1) \times (N/I)$ units.

The output switch in the interleaver moves cyclically with the input switch; the deinterleaver has also the same structure except for the reverse order of delays. The DVB specifications are written to specify all the transmission and receive functions as well as system parameters, except for the symbol rate of modem operation. (This is because no frequency planning was readily available.)

DAVIC specifications

DAVIC specification for LMDS are basically the same as the DVB specification (except for an option of certain, so-called *alpha values* for channel filtering). It also refers to either QPSK or 16-QAM for modulation. DAVIC further defines future extensions. Along with the MPEG-2 scheme used for video broadcasting, a mapping function to ATM data in the down-stream channel is also made. (Two 187-byte packets are formed when 3 control bytes are appended to 7 consecutive 53 bytes ATM channel.)

The specification on the return channel is primarily due to DAVIC because DVB is more interested in broadcast services in its first phase. The return channel that has been specified by DAVIC for LMDS is a multiple-access channel. It uses the TDMA. The MAC protocol allocates time-slots to different users. Each user can transmit only if a time slot is allotted. These time slots as per the specification are made of 68 byte each that includes 4-byte preamble and 1-byte guard fields. The remaining 63 bytes include 53 bytes of information plus 10 bytes for parity-check fields. This allows each time slot to support an ATM cell. Error protection on the upstream channel is not as efficient it as on downstream channel. But necessary compensation is made at the design of transmit and receive functions.

The MAC protocol is used to allocate resources to various user terminals. Both downstream and the upstream frames are encapsulated as one ATM cell. Each frame on the downstream includes two slots. There is a frame-start slot followed by a random-access slot. The upstream frame has three slots namely, the

polling-response, the contention and the reserved-time slots. The polling-response slots are used to response to a poll message. The contention slots are shared and utilized by more than one terminal. They may result in collision and the contention when a collision occurs can be resolved (by waiting for a random amount of time before retransmitting). Reserved time slots are reserved for use by the terminal. The terminal transmits on these slots whenever it has data; and whenever it does not have any data, it transmits an empty cell. The MAC protocol has also an option for a combination of circuit-mode reservation for constant bit rate (CBR) services and a dynamic reservation for the variable bit rate (VBR) and unspecified bit rate (UBR) services. Polls are periodically repeated at intervals of less that or equal to 2 seconds. If a new user comes in, the system listens to a downstream channel to find any message being sent to it. If it does not find the message in 2 seconds, then it switches to the next downstream channel and listens again. This polling search goes on till the terminal finds a message transmitted to it [8.1-8.5].

On March 11, 1999 the IEEE 802 Executive Committee approved the BWA Project Authorization Request (PAR) and charted the 802.16 Working Group on BWA to develop the air interface. The next PAR is directed at BWA system coexistence. (N-WEST Web site: http://nwest.nist.gov)

ABBREVIATIONS AND ACRONYMS

A	
ACF	Autocorrelation function
ACL	Asynchronous connectionless
ADF	Automatic direction finder
AM	Amplitude modulation
AMPS	Advanced Mobile Phone Service
AoA	Angle-of-arrival
ANSI	American National Standards Institute
AP	Access point
ARQ	Automatic return request
ARQN	Automatic return request indication
ATM	Asynchronous transfer mode
AWGN	Additive white Gaussian noise
B	
BER	Bit error rate
BJT	Bipolar junction transistor
BS	Base station
BSSA	Bent-stacked slot-antenna
BU-model	Bad urban model
BWA	Broadband wireless access
BWF	Bandwidth factor
C	
CAD	Computer-aided design
CB	Citizen band
CBR	Constant bit rate
CCITT	Consultative Committee on International Telegraphy and Telephony
CDMA	Code-division multiple access
CDPD	Cellular digital packet data
C/I ratio (CIR)	Carrier-to-interference ratio
CLEC	Competitive local exchange carrier
CMOS	Complementary metal-oxide semiconductor
CO	Central office
CP	Circular polarisation
CPE	Customer premises equipment
CPW	Coplanar waveguide
CPWBA	Coplanar waveguide bow-tie antenna

CRC	Cyclic redundant code
CVSD	Continuous variable slope data
D	
dB	Decibel
dBi	Decibel gain with reference to isotropic antenna
dBd	Decibel gain with reference to dipole antenna
DAVIC	Digital Audio -Video Council
DC	Direct current
DC distribution	Dolph-Chebyshev distribution
DCS-1800	Digital Communication System-1800
DECT	Digital European/Enhanced Cordless Telephone
DH rate	Data high rate
DM rate	Data medium rate
DoA	Direction-of-arrival
DSP	Digital signal processing
DS-SS	Digital sequence spread spectrum
DS-x	Digital signal level x (x = 1, 2, …, etc.)
D/U ratio	Desired-to-undesired signal ratio
DVB	Digital video broadcasting
E	
EDGE	Enhanced Data Rates for GSM Evaluation
EIA	Electronic Industries Association
EIRP	Effective isotropic radiated power
emf	Electromotive force
EM	Electromagnetic
EMI	Electromagnetic interference
E_o/N ratio	Bit energy-to-noise ratio
ETACS	European Total Access Communication System
ETSI	European Telecommunication Standard Institute
F	
FCC	Federal Communication Commission
FDD	Frequency division duplex
FDMA	Frequency division multiple access
FDRM	Frequency domain reciprocal modulation
FEA	Fractal element antenna
FEC	Forward error correction
FET	Field-effect transistor

FFT	Fast Fourier transform
FH	Frequency hopping
FH-CDMA	Frequency-hopped CDMA
FHSS	Frequency-hopped spread-spectrum
FM	Frequency modulation
FSK	Frequency-shift keying
G	
GFSK	Gaussian frequency-shift keying
GMSK	Gaussian minimum shift keying
GPRS	General Packet Radio Service
GSM	Group Special Mobile
GTD	Geometric theory of diffraction
GWSSUS	Gaussian wide-sense stationary uncorrelated scattering model
H	
HCMTS	High Capacity Mobile Telephone System
HEMT	High electron-mobility transistor
HIPERLAN	High Performance Radio LAN
HMA	Hybrid matrix amplifier
HV-link	Header value-(link)
I	
IFU	Invisible fibre unit
ILEC	Incumbent local exchange carrier
IMT-2000	International Mobile Telecommunication-2000
IMTS	Improved Mobile Telephone Service
IP	Internet Protocol
irLAN	Infrared LAN
IR	Infrared
IRP	Isotropic radiated power
IRPA	International Radiation Protection Agency
ISM band	Industrial, scientific and medical band
IS-95	EIA Interim Standard for US CDMA
ISDN	Integrated services digital network
ITU	International Telecommunication Union
ITU-T	Telecommunication Standardisation Sector for ITU
JK	
JDC	Japanese Digital Cellular (Pacific Digital Cellular)

L	
LAN	Local area networking
LEO	Low earth orbit
LHCP	Left-handed circular polarisation
LLC	Logical link control
LMDS	Local Multipoint Distribution Service
LMS	Least-mean square
LNA	Low noise amplifier
LoS	Line-of-sight
LW	Long-wave
M	
MAC	Medium access control
MAN	Metropolitan area networking
MH	Mobile host
MLS	Microwave landing system
mm wave	Millimetre wave
mmf	Magnetomotive force
MMCB patch antenna	Multi-probe cavity-backed patch antenna
MMDS	Multichannel Multipoint Distribution System
MPEG	Motion Picture Experts Group
MS	Mobile station
MSO	Multiple service operator
MTSO	Mobile telephone switching office
MU	Mobile unit
MVDS	Mobile video distribution service
MW	Medium-wave
N	
NADC	North American Digital Cellular
NIU	Network interface unit
NMHA	Normal mode helical antenna
NMT-450	Nordic Mobile Telephone-450
NMS	Network management system
NOC	Network operation centre
NTT	Nippon Telephone Telegraph
O	
OAM	Operation, administration and maintenance
OC x	Optical carrier level x (x = 1, 3, 12, … etc.)
OEM	Original equipment manufacturer
OFDM	Orthogonal frequency division multiplexing

P	
PAN	Personal area networking
PC	Personal computer
PCB	Printed circuit board
PCM	Pulse code modulation
PCS	Personal communication system
pdf	Probability density function
PDA	Personal digital assistant
PIFA	Planar inverted-F antenna
PL	Path-loss
PMP	Point-to-multipoint
PN	Pseudorandom noise
POTS	Plain old telephone system
PPP	Point-to-point
PSK	Phase-shift keying
PSTN	Public switched telephone network
Q	
QAM	Quadrature amplitude modulation
QoS	Quality-of-service
QPSK	Quadrature PSK
R	
RCS	Radar cross-section
RF	Radio frequency
RFI	Radio frequency interference
RFID	Radio frequency identification
RHCP	Right-handed circular-polarisation
rms (value)	Root mean-squared (value)
RX	Receiver
S	
SAR	Specific absorption rate
SATCOM	Satellite communication
SBF antenna	Short backfire antenna
SCO	Synchronous connection-oriented
SC-RMSA	Short-circuit rectangular microstrip antenna
SDMA	Space-division multiplex access
SFCP antenna	Single-fed circularly-polarised antenna
SIR	Signal-to-interference ratio
SNR	Signal-to-noise ratio
SONET	Synchronous Optical Network
SS	Spread spectrum
SSFIP antenna	Strip slot foam inverted patch antenna
SW	Short-wave

T	
TACS	Total Access Communication System
TCP/IP	Transmission Control Protocol/Internet Protocol
TDD	Time-division duplex
TDMA	Time-division multiplex access
TE	Transverse electric
Telco	Telecommunication company (Service provider)
TEM	Transverse electromagnetic
TIA	Telecommunications Industries Association
TM	Transverse magnetic
ToA	Time-of-arrival
TU-model	Typical urban model
TX	Transmitter
U	
UHF	Ultra-high frequency
UMTS	Universal Mobile Telecommunication System
V	
VHF	Very-high frequency
VoD	Video-on-demand
VSWR	Voltage standing-wave ratio
W	
WAN	Wide area networking
WAP	Wireless application protocol
WCDMA	Wideband CDMA
WLAN	Wireless LAN
WLL	Wireless local-loop
WRC	World Radio Congress
WWW	World Wide Web
XYZ	
-	-
Numerics	
2G/2.5G/3G	Second generation, etc.

SUBJECT INDEX