THIRD EDITION

Antenna Theory
and Design

Warren L. Stutzman   Gary A. Thiele

Antenna Theory and Design

Antenna Theory and Design

THIRD EDITION

Warren L. Stutzman
Professor Emeritus
Virginia Tech

Gary A. Thiele
Professor Emeritus
University of Dayton

The top cover art is the 100-m radio telescope of the National Radio Astronomy Observatory in Green Bank, West Virginia, USA; photo taken by Warren L. Stutzman. The bottom cover art is the NASA 70-m deep space communication antenna at Tidbinbilla, Australia; photo taken by Gary A. Thiele.

This book was set in Times by MPS Limited and printed and bound by Courier Westford, Inc. The cover was printed by Courier Westford, Inc.

This book is printed on acid free paper. ♾

Founded in 1807, John Wiley & Sons, Inc. has been a valued source of knowledge and understanding for more than 200 years, helping people around the world meet their needs and fulfill their aspirations. Our company is built on a foundation of principles that include responsibility to the communities we serve and where we live and work. In 2008, we launched a Corporate Citizenship Initiative, a global effort to address the environmental, social, economic, and ethical challenges we face in our business. Among the issues we are addressing are carbon impact, paper specifications and procurement, ethical conduct within our business and among our vendors, and community and charitable support. For more information, please visit our website: www.wiley.com/go/citizenship.

Evaluation copies are provided to qualified academics and professionals for review purposes only, for use in their courses during the next academic year. These copies are licensed and may not be sold or transferred to a third party. Upon completion of the review period, please return the evaluation copy to Wiley. Return instructions and a free of charge return mailing label are available at www.wiley.com/go/returnlabel. If you have chosen to adopt this textbook for use in your course, please accept this book as your complimentary desk copy. Outside of the United States, please contact your local sales representative.

This book is dedicated to our wives,
Claudia and Jo Ann
and to our children,
Darren and Dana
Eric, Scott, and Brad
and to our students.

Preface

The first edition of *Antenna Theory and Design* was published in 1981 and the second in 1998. This three-decade span has seen major expansions of antenna application areas, mainly in wireless communications. Along with technology advances has come public awareness of the presence and importance of antennas. Base station antennas are now visible in all urban and rural areas. Wireless telephones such as smart phones have put antennas into the consumer's hand. Personal and fleet vehicles have many wireless features such as satellite navigation (e.g., GPS), satellite and broadcast radio, cellular telephone, remote keyless entry, wireless toll tags, remote tire pressure monitoring, and many more.

This third edition has 16 chapters, which is an increase of 4 chapters over the second edition, representing both introduction of new material and reorganization of previous topics. The third edition has several new sections of material specifically addressing modern wireless applications for antennas. Chapter 11 on low-profile antennas and personal communication antennas includes an expanded treatment of microstrip antennas and arrays, and new material on leaky wave antennas, fundamental limits on antenna size, antennas for compact devices, radio frequency identification (RFID) antennas, dielectric resonator antennas, near fields for large antennas, effects of the human body on antenna performance, and biohazards of RF exposure.

Chapter 12 on terminal and base station antennas has new material on satellite terminal antennas, base station antennas, vehicular antennas, smart antennas, and adaptive antennas.

The book organization has changed slightly to reflect new emphasis on wireless applications. The new Chapter 1 has a greatly expanded discussion of the history of wireless communications with emphasis on antennas. This provides a setting for the applications for antennas and a motivation for pursuing the fundamentals needed to design antennas for specific applications. Chapter 1 also includes a detailed overview of antenna parameters and antenna types. Chapter 4 on system aspects allows students to evaluate antennas as used in systems earlier in the book than in previous editions. Chapter 4 includes an introduction to array antennas, and the detailed treatment of arrays is now in Chapter 8 and includes several practical aspects of antenna arrays such as feed networks and wideband phased arrays.

The number of problems at the end of the chapters has increased from 504 in the second edition to 591 in this third edition. Many of the new problems include simulation aspects to encourage the student to perform high-level quantitative investigations of antennas. Problems involving open-ended projects have also been added.

This book is a textbook and finds its widest use in the college classroom. Thus, the primary purpose is to emphasize the understanding of principles and the development of techniques for examining and designing antenna systems. The book is used by practicing engineers as well as students. This is because of the applied nature of the material and because it provides a coherent development of basic topics that are directly useable for analyzing practical antennas. Handbooks with data on many antennas are available. But, handbooks and special topics books typically do not provide sufficient background for full understanding of antennas, which can lead to the dangerous practice of "formula lifting" that can end in inappropriate use of a formula.

Antenna Theory and Design covers antennas from three perspectives: antenna fundamentals, antenna techniques, and the design of popular antennas. The first four chapters stress antenna fundamentals. It is assumed that the student has no previous

exposure to antennas, and thus all basic physical and mathematical principles are presented early, in Chapters 1, 2, and 3. This includes Maxwell's equations along with a physical explanation of how antennas radiate. The four types of antenna elements are treated in the following places in the book:

> Electrically small antennas—Sec. 3.1
> Resonant antennas—Chapters 3 and 6
> Broadband antennas—Chapter 7
> Aperture antennas—Chapter 9

Also, arrays are introduced in Sec. 3.5 and covered in detail in Chapter 8. The use of antennas as devices in systems is covered in Chapter 4. The synthesis of arrays and continuous antennas for low side lobe or shaped main beam applications is presented in Chapter 10. Chapters 11 and 12 treat low-profile and terminal antennas. Antenna measurements are treated in Chapter 13. Chapters 14 to 16 introduce computational electromagnetic methods (CEM) for evaluating simple antenna elements as well as large complex antenna systems. There are now a wide variety of antenna-specific computer codes and general mathematics applications packages available for antenna computations (see Appendix G).

The level assumed for this book is that of a typical senior in electrical engineering. It is assumed the reader has had a basic course in electromagnetics, such as commonly required of electrical engineering and physics majors. Essential mathematics includes complex numbers, trigonometry, vector algebra, and coordinate systems (rectangular, cylindrical, and spherical). Vector calculus is used at various points, and scalar integration is frequently used.

SUGGESTIONS FOR ACADEMIC COURSES

This book can be readily adapted to various academic programs at both introductory and advanced levels. The first course in antennas is usually either a one-semester senior elective or an entry level graduate course. For a one-semester introductory course, Chapters 1 to 7 and some of 8 and 9 are covered. For a master's level course, parts of Chapters 8 and 9 can be added. Topics from later chapters such as synthesis (Chapter 10), low-profile antennas (Chapter 11), and terminal antennas (Chapter 12) can be treated along with the early chapters. For example, PIFA antennas in Sec. 11.6.3 are a good fit with wire antennas of Chapter 6.

A second course can focus on advanced design, synthesis, and computational methods using Chapters 10 to 16. Alternatively, a second course can specialize on computational methods using Chapters 14 to 16.

LEARNING APPROACH

This text adopts a general approach to the treatment of antenna types and specific antennas. After a general discussion of the principles, design formulas are derived and explained using sample calculations and detailed examples. Often some details are left as student exercises or problems at the end of the chapter. This is done to keep the amount of material from overwhelming the reader. For the same reason, while many antennas for wireless applications are presented, it is not an exhaustive treatment that is better left to handbooks and focused-topic books. As appropriate, data are presented using numerical or experimental models or computations based on theoretical formulations. In addition, empirical formulas are often presented for easy evaluation of performance parameters.

SIMULATIONS AS A LEARNING AID

Literature describing how to write antenna computational programs and commercial simulation packages for antennas are now widely available (see Appendix H.10). Therefore, throughout this book results from numerical simulations are presented, often in association with analytical calculations to explore the accuracy of approximate formulas. Many problems at the end of each chapter, or parts of problems, require the use of simulations to obtain numerical values. Most include the notation "*Simulation*." These problems greatly enhance the learning experience and give confidence in using simulations. Of course, these problems can be skipped or postponed until computational methods are treated in Chapters 14 to 16. Alternatively, computational material can be covered as needed while going through the early chapters. In any case, the reader should obtain at least one simulation package and use it while studying antennas. Many commercial codes have excellent user interfaces that handle three-dimensional geometries and also have well-organized outputs. The first choice for a code would be a moment method-based, wire antenna simulation package. There is further discussion of computing tools for antennas and specific suggestions of available simulation packages given in Appendix G.

LEARNING RESOURCE MATERIALS

Several features are included to aid in learning and in preparation for further self-study. There are many terms that are specific to the field of antennas, and the terms used in this book follow the IEEE standard definitions. Numbered literature references are found at the end of each chapter for further reading. In addition, the bibliography in Appendix H lists nearly all antenna books published in English, arranged by technical topic. Included are the many books that have appeared since the second edition, mostly on specific antenna topics. Citations to these books are made in the text using the following format: "[H.x.y: Author, page number]". The other appendices give information on the radio spectrum, material data, and many mathematical relations used in antenna analysis.

ACKNOWLEDGMENTS

materials. Thanks are also extended to the University of Dayton ECE Department for their support, to Mr. Kuan Chang Pan of the University of Dayton for assisting in the preparation of more than six chapters of the Solutions Manual, and to the MathWorks Book Program for the use of MATLAB during the writing of this third edition.

Finally we recognize our wives, Claudia and Jo Ann, for enduring countless hours of neglect during the preparation of the three editions.

Warren L. Stutzman
Professor Emeritus
Virginia Tech Antenna Group
ECE Department, Virginia Tech
Blacksburg, Virginia

Gary A. Thiele
Professor Emeritus
ECE Department
University of Dayton
Dayton, Ohio

Contents

Chapter 1

Introduction

Fundamental principles will be presented in this textbook in a manner that students can readily apply to the analysis and design of antennas. The physics principles of electromagnetism are used to understand how antennas operate. Antenna design problems are solved using the physics principles together with mathematics and computational methods, but the emphasis is on reducing the physics and mathematics to an understandable level. Before developing mathematical tools for designing antennas in subsequent chapters, this chapter provides a background to the field of antennas along with a glimpse of the basic principles, terminology, and application areas. We begin with a brief history of antennas and related technology developments.

References to specific literature works are found at the end of each chapter in which the citation occurs. In addition, a list of antenna books in Appendix H is grouped by their topical coverage to aid the student in locating further details. Citations to books in Appendix H are of the following form: [H.x.y: Author, page nos.]. The IEEE definitions of antenna terms [H.2: IEEE] are followed closely in this book.

1.1 THE HISTORY OF ANTENNAS

Antennas are mainly found in communication applications, with most uses in *telecommunications*, which means distant communications and has roots in the Greek word "tele" meaning "at a distance." References 1 through 6 and [1–6; H.8.2: Visser, Chaps. 1 and 2] are general treatments on the history of technology related to antennas. A highly recommended work is Sarkar, Mailloux, Oliner, Salazar-Palma, and Sengupta [2], which provides a chronological summary of historical milestones in electricity, magnetism, light, and wireless communications, including antennas, and provides comprehensive discussions of key wireless-related technology developments.

1.1.1 Overview of the History of Communications

Antennas are most widely used in the field of communications, but electronic communication, including wireless communication, is a recent development in human history. In fact, of the 4.5 billion years of the Earth's existence, modern humans have been present for only 200,000 years, whereas electronic communication is less than 200 years old. Ancient long-distance communication employed couriers to deliver messages. Early messages used pictures and symbols. The Greek alphabet was developed around 500 BC. The word "alphabet" comes from the first two letters of the Greek alphabet, "alpha" and "beta." [3] Couriers were used extensively by the Romans to deliver hand-written messages. Message delivery was reinvented in the 19th century in North America with horseback riders (e.g., the Pony Express), who delivered messages coast to coast in ten

days, which was fast for the era. Carrier pigeons were also used to transport written messages.

Table 1-1 gives the historical timeline of significant events in communications with emphasis on wireless communications. Communication during the pre-modern civilization period used optical communications such as flags for line-of-sight and smoke signals for non-line-of-sight, as well as devices such as drums for non-line-of-sight acoustical communications. Signal flags are still in use today for ship-to-ship communications. Drum beats traveled through the ground for up to 10 km. [3]. Smoke signals in the daylight and fire signals at night were used extensively in antiquity. It is very interesting to note that all of these early long-distance communication mechanisms made use of digital techniques. The dawn of electronic communication in 1844 was also digital via Morse code over the first telegraph transmission line from Baltimore to Washington, DC, carrying the now famous message: "What hath God wrought?" The telegraph, invented by Samuel Morse (1791–1872), used "dots" and "dashes" generated by hand using a telegraph key and received by a operator listening at the other end for audible sounds. The

Table 1-1 Timeline of Significant Events in the History of Antennas and Communications

Pre-modern civilization (up to 2 million years ago)
   Optical communications: Smoke signals, flags
   Acoustical communications: Drums

1844	Telegraph—The beginning of electronic communication
	Samuel Morse
1864	Maxwell's equations—Principles of radio waves and the electromagnetic spectrum
	James Clerk Maxwell
1866	First lasting transatlantic telegraph cable
1876	Telephone—Wireline analog communication over long distance
	Alexander Bell
1887	First Antenna
	Heinrich Hertz
1897	First practical wireless (radio) systems
	Guglielmo Marconi
1901	First transatlantic radio
	Guglielmo Marconi
1920	First broadcast radio station

World War II Development of radar; horn, reflector, and array antennas

1950s	Broadcast television in wide use
1960s	Satellite communications and fiber optics
1980s	Wireless reinvented with widespread use of cellular telephones

Electromagnetics Pioneer	Antenna Pioneer	Wireless Pioneer
James Clerk Maxwell	*Heinrich Hertz*	*Guglielmo Marconi*

first underwater telegraph cable in 1851 spanned the English Channel. An interesting side note is that the cable suddenly stopped functioning due to a fisherman who had snagged the cable, pulled it up, and hacked it apart, thinking it was a sea serpent. [3, p. 32] After a few failed attempts at a transatlantic telegraph cable, the first lasting connection was completed in 1866, but had initial usage rates of $1 per letter! With the invention of the telephone by Alexander Graham Bell (1847–1922) in 1876, electronic communication began a long period of analog communication before digital communication returned, rather recently.

The mathematical and physical foundations for electromagnetic waves were presented by James Clerk Maxwell in 1864 and are now referred to as Maxwell's equations. Wireless (radio) telegraphy was made possible with the inventions of antennas and radio systems in 1887 by Heinrich Hertz and in 1897 by Guglielmo Marconi, respectively. In 1901, Marconi performed the first transatlantic radio transmission. During the early part of the 20th century there were several advances in electronics that enabled more sophisticated radio systems. This culminated during World War II with development of microwave devices, enabling, for example, radar systems. A new form of tethered communication returned in the 1960s with the maturing of optical fiber cable. Finally, massive deployment of radio communications was launched in the 1980s, which turned out to be the reinvention of wireless communications about a century after its first deployments. We now give some details on key science and technology milestones in the history of communications, with emphasis on antennas.

1.1.2   The Significant Contributions to the Understanding of Electromagnetic Waves

Electromagnetic waves, which enable wireless communication, are composed of electric and magnetic fields. Electric and magnetic fields are associated with electric and magnetic forces which are two of the four forces of nature,[1] with gravitational and nuclear (weak and strong) being the other two. Common forces arising from, for example, friction or spring tension are actually manifestations of electric and magnetic forces. The early investigations of electromagnetics treated electricity and magnetism as separate unrelated phenomena, which we now know is true only in the static case.

In 1600, English physicist and physician to Queen Elizabeth I, William Gilbert (1544–1603) published a book on electricity and magnetism. In 1772, English chemist and physicist Sir Henry Cavendish (1731–1810) derived the now well-known inverse distance square law of force between electric charges, which French physicist Charles Augustin Coulomb (1736–1806) experimentally verified in 1785, and now the law bears his name. The invention of the electric battery by Italian physicist Alessandro Volta (1745–1827) in 1800 provided the first continuous source of electricity, which was the key to subsequent experimental discoveries. In 1819, Danish physicist Hans Oersted (1777–1851) accidentally observed in a classroom demonstration to his students that a compass needle is deflected when brought near a current-carrying wire, laying the foundation for *electromagnetism*. Oersted did not fully realize that the current produced a magnetic field which interacted with the magnetic field of the compass needle, but French mathematician and physicist Andre Marie Ampere (1775–1836) did. A week after Oersted announced his results, Ampere confirmed the findings and introduced the "right-hand rule," which gives the direction of the magnetic field as that of the curled fingers when the thumb is along the current direction. Ampere also showed that parallel current-carrying wires attract (repel) if

---

[1] Physicists often include magnetic forces with electric forces because they arise from electrical charge motion.

the current directions are the same (opposite). Ampere postulated that magnetic materials consist of aligned tiny electric current loops, which was shown to be true decades later.

The full connection between electricity and magnetism was finally established in 1831 by English physicist and chemist Michael Faraday (1791–1867) when he discovered electrical induction (i.e., creation of an electric field from a changing magnetic field) and transformer action; he constructed the first electric generator to provide steady current without using a battery. Faraday formulated "lines of force" based on his observations of iron filing alignments along the magnet fields from a permanent magnet. Faraday used lines of force, now called field lines, to visualize field direction and intensity (lines are more closely spaced lines for stronger fields), and he even believed that fields were physically real. Field theory finally explained "action-at-a-distance" in electric, magnetic, and gravitational phenomena. Maxwell later further refined the use of fields.

In 1842, Joseph Henry (1797–1878) of Princeton University performed perhaps the first radiation experiment. He "threw a spark" in a circuit in an upper room and observed that needles were magnetized by the current in a receiving circuit located in the cellar. This experiment was extended to a distance of over a kilometer. Henry also detected lightning flashes with a vertical wire on the roof of his house. These experiments marked the beginning of wire antennas.

Based on his observations in 1875 that telegraph key closures radiate, Thomas Edison patented a communication system in 1885 that employed top-loaded, vertical antennas.

The theoretical foundations for antennas rest on Maxwell's equations, which Scottish mathematician and physicist James Clerk Maxwell (1831–1879)[2] presented before the Royal Society in 1864, which unify electric and magnetic fields into a single theory of electromagnetism. Maxwell used the laws of Ampere, Gauss, and Faraday and introduced displacement current to form a single set of electromagnetic theory equations. Maxwell also predicted that light is explained by electromagnetics and that light and electromagnetic disturbances both travel at the same speed; he actually derived a value for the speed of light that was very close to what is now known to be correct, 3×10^8 m/s. However, Gustav Kirchhoff (1824–1887) was the first to discover that electrical signals propagate at the speed of light in 1857. Maxwell introduced the notion that electromagnetic phenomena reside in the space around electrified bodies as well as in the bodies. Although Maxwell incorrectly believed in the existence of an *ether* material in space that supports action-at-a-distance, and he rejected the notion that electricity has a particle nature, his equations have endured. In spite of Maxwell being a capable experimentalist, he performed no measurements to validate his theories, nor did he even propose an experiment. He published the two-volume set *Treatise on Electricity and Magnetism* in 1873, but it was hard for his followers to understand. Interestingly, Maxwell developed the basics for color photography and made the first color photograph in 1861. He died in 1879 at the age of 48 from stomach cancer before his theories were experimentally verified by Hertz.

Maxwell's work was not immediately accepted, probably because it was customary at the time for physicists to connect their work to that of Newton by including mechanical laws, which Maxwell did not do; also, unlike Newton, Maxwell was not an aggressive self-promoter. [9] Many physicists, in fact, held out for decades that a mechanical substructure would be found for Maxwell's theory, but when relativity and quantum theories took hold, the quest for a unified physics foundation was abandoned. [10, p. 100] Albert Einstein's (1879–1955) special theory of relativity in 1905 showed that Newtonian mechanics requires modification when relativistic speeds (i.e., approaching the speed of light) are involved, whereas Maxwell's equations already are relativistic and require no modification for cases of charges moving at relativistic speeds.

---

[2] References focusing on Maxwell's life are [7–9].

The importance of the contributions of several of the pioneers in electromagnetism, which enabled wireless technology to eventually flourish, has been recognized by naming units of electrical quantities after these giants:

Unit	Honoree	Quantity	Symbol
Ampere	Andre Ampere	Current	A
Coulomb	Charles Coulomb	Charge	C
Farad	Michael Faraday	Capacitance	F
Henry	Joseph Henry	Inductance	H
Hertz	Heinrich Hertz	Frequency	Hz
Volt	Alessandro Volta	Voltage (emf)	V

1.1.3   Key Developments in Communication Technology

In 1872, American dentist Mahlon Loomis (1826–1886) obtained the first patent on wireless telegraphy. He performed crude demonstrations using kite wires 200 m long at two mountain-top sites in the Blue Ridge Mountains of Virginia separated by 22 km. By grounding one end, a disturbance of the atmospheric electricity caused a deflection in a current meter at the other end. [2, p. 69 and p. 251] But Loomis believed that he was controlling a conducting layer in the lower atmosphere, so the basic premise of his patent was flawed.

In 1884, John Henry Poynting (1852–1914) showed that Maxwell's theory can be used to predict that energy flows through empty space calculated using the electric and magnetic fields. Heaviside independently at nearly the same time came upon the same discovery. [1, p. 313]

The German physicist Heinrich Hertz (1857–1894)[3] was able to verify experimentally in 1886 (published in 1887) Maxwell's claim that electromagnetic actions propagate through air at the speed of light. Hertz received his Ph.D. at the University of Berlin in 1880 studying under Gustav Kirchhoff, who is well known to students of electrical engineering through his contributions to circuit theory (he also discovered and coined the term for blackbody radiation). Hertz also studied under Hermann Helmholtz (1821–1894), who has been credited with the concept of conservation of energy and who suggested that Hertz look for radiation above the visible region. Hertz built what can be considered to be the first radio system, which included the first dipole antenna and first loop antenna, as shown in Fig. 1-1.[4] The primary source of electrical disturbances was a spark gap generator consisting of two metal plates in the same plane, each with a wire connected to an induction coil. This early antenna is similar to the capacitor-plate dipole antenna, which will be described in Sec. 3.1, and was called a "Hertzian dipole." Hertz discovered that electrical disturbances could be detected via visually observed sparks with a single wire loop of the proper dimensions for resonance and an air gap for the sparks to occur. By moving the detector between the transmitter and a large conducting plate acting as a reflector, he created standing waves and calculated the wavelength as 6 m, corresponding to 50 MHz. Hertz's understanding of electromagnetic waves was so thorough that he varied the loop perimeter dimension in increments to maximize the induced spark length, thereby tuning the loop to half-wavelength resonance at a perimeter of 4.3 m, corresponding to a wavelength of about 8.6 m for a resonant frequency of about 35 MHz. [11,12]

---

[3] Articles on Hertz's life and contributions are found in [11–13].
[4] Although Henry used wire antennas more than 40 years earlier, Hertz was the first to fully understand their operation and use.
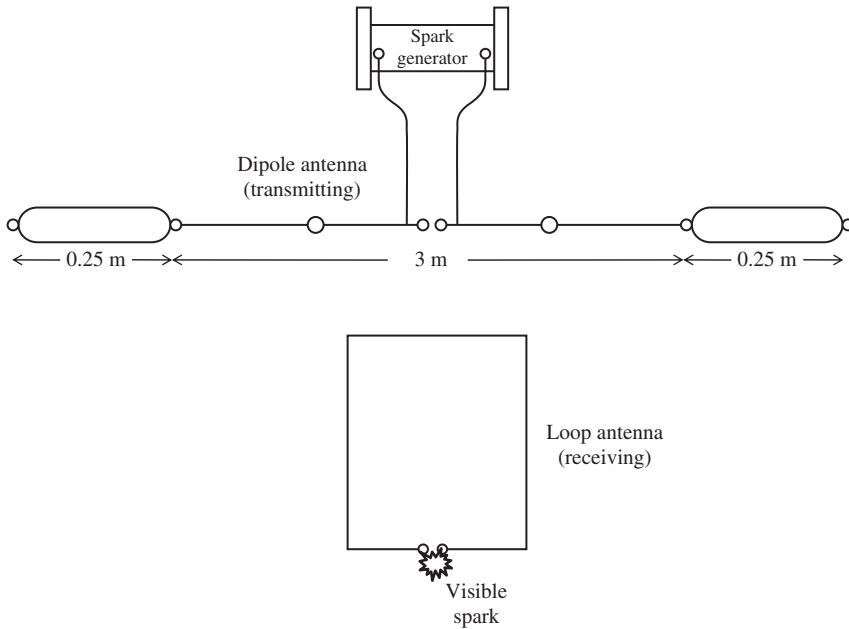
Figure 1-1 The first complete radio system, built and tested by Heinrich Hertz in 1886. It includes a dipole source antenna and a loop receiving antenna. Both antennas were the first of their kind.

Hertz was the first to demonstrate conclusively the existence of electromagnetic waves. With this laboratory radio system, Hertz made several key discoveries, including standing waves, reflection, refraction, and polarization. Based on Maxwell's prediction that light is an electromagnetic wave, Hertz used ray optics principles, well known at the time, to design a more directive antenna. This was necessary because he experimented indoors and realized that he needed a directive antenna to reduce signal reflections off the walls. In 1888, he constructed a parabolic cylinder reflector antenna from a sheet of zinc; see Fig. 1-2. It was fed with a dipole along the focal line and operated at 455 MHz. Although the aperture extent was only 1.8 wavelengths (1.2 m) in the horizontal dimension and only 3 wavelengths (2 m) in the vertical dimension, this was sufficient to produce a narrow directional beam that was 35° wide and 80° in the vertical plane. [14] Hertz published what is probably the first book on antennas, *Electric Waves* [H.3: Hertz], in 1893. He also investigated the skin effect and the photoelectric effect. Hertz died of complications from a dental abscess at the early age of 36. [15]

In 1888, English physicist and electrical engineer Oliver Heaviside (1850–1925), who did not complete college, extended Maxwell's work. He reduced the 20 equations of Maxwell that were in scalar form and introduced vector notation, yielding the compact set of 4 equations we use today. In his three-volume work, *Electromagnetic Theory*, he predicted the existence of the charged layer in the upper atmosphere, called the *ionosphere* or *Heaviside layer*, which is responsible for long-distance communications via reflection from the layer. He also proposed inductive loading on transmission lines to produce distortionless signal propagation, making practical long-distance wire communication possible. Heaviside is often called the father of telecommunication.

English physicist Oliver Lodge (1851–1940) in the 1890s conducted experiments similar to those of Hertz and Marconi but at higher frequencies. His U.S. Patent 609,154 of Aug. 16, 1898, includes implementations of "Hertzian-wave telegraphy," loaded dipoles, biconical antennas (see Sec. 7.4), dielectric lenses (see Sec. 9.8), and even frequency division multiplexing. He also coined the term "impedance" and correctly
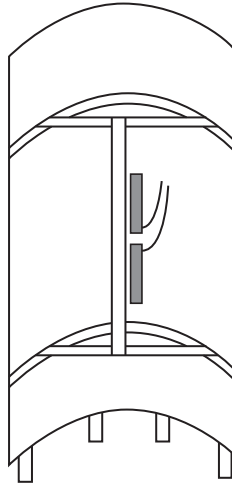
Figure 1-2 The 455-MHz cylindrical-parabolic reflector antenna invented by Hertz in 1888. The parabolic shape in the horizontal direction has an aperture width of 1.2 m and the linear shape in the vertical direction is 2 m long.

suggested that the sun emits radio waves. After 1910, Lodge shifted his efforts from technology to other areas, such as communicating with the dead.

Indian physicist J. Bose (1858–1937) performed extensive investigations of phenomena and devices at microwave and millimeter wave frequencies up to 60 GHz. In 1896, he developed the first pyramidal horn antenna (see Sec. 9.4.3) which he called a collecting funnel; it was an extension of an open-ended waveguide radiator (see Sec. 9.2.2) that Lodge developed. He also investigated polarization and polarizers.

1.1.4 Long-Distance Wireless Communications

Guglielmo Marconi (1874–1937), an Italian electrical engineer, studied the work of Hertz and concluded that electromagnetic waves could be used for telegraph purposes. After first experimenting with detecting lightning over kilometer distances, he built a microwave parabolic cylinder like the one used by Hertz in Fig. 1-2 in 1895 for code transmissions over very short distances at 1.2 GHz. He then reduced the frequency of operation and employed monopole antennas. Like most of the antenna designers of his day, Marconi did not use theory to guide his experiments, but instead proceeded largely on an empirical basis. He did not fully understand why lower frequencies were better for long-distance communication, and he did not know that his monopole operated best for a quarter wavelength height. But, the ability to communicate great distances both at night and during the day resulted in HF-band and lower frequencies being used heavily for decades. Failing to obtain support from the Italian government for his work, Marconi moved to England in 1896 and teamed with the British Post Office in subsequent experiments. In 1901, Marconi received British Patent Number 7777 (often referred to as the "four sevens patent") for a tunable radio system, allowing for multiple systems to operate on different frequencies. Contrary to popular belief at the time, Marconi believed that radio signals would follow the curvature of the earth at low frequencies. To prove this dramatically, he set up a 3,500 km link with the transmitter in Poldhu in Cornwall, England, and receiver in St. Johns, Newfoundland, Canada. The transmitting antenna for the first transatlantic radio communication in 1901 consisted of a spark transmitter connected between the ground and a system of 54 fanned out wires 1 meter apart at the
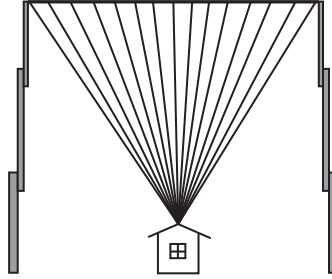
Figure 1-3 The fan monopole transmitting antenna used by Marconi at 500 kHz in the first transatlantic radio link.

top and supported on 45.7-m tall masts; see Fig. 1-3. The antenna resembled a variation of a planar monocone antenna (see Sec. 7.4). The receiving antenna was a 122-m long copper wire suspended by a kite. The exact operating frequency remains a subject of controversy. Reproductions of the experiment (scaled up in frequency) and numerical simulations indicate that Marconi used a frequency of 500 kHz. [2, Chap. 12] The Morse code letter "s" (three dots) was sent repeatedly, and Marconi logged three successful receptions in his laboratory notebook. Although many of the elements of his success were known before his work, he is credited with inventing *wireless telegraphy* because he combined the many elements into a commercially feasible system. Reliable transatlantic wireless telegraphy service began in 1907. The Marconi Wireless Telegraph Company of America started in 1899 and evolved into the Radio Corporation of America (RCA) in 1919. As a historical note, the passenger ship SS *Titanic* that sank the night of April 14, 1912, had a Marconi wireless set on board and transmitted distress calls. But the radio operator on duty on a nearby ship, the *California*, was asleep at the time of the sinking and distress calls from the *Titanic* were not received.[5] A distant ship, the *Carpathia*, did receive the distress call and arrived four hours later at the *Titanic* to rescue 706 survivors of the 2,233 people aboard.

The term wireless was used until 1920 when broadcast radio began. The popular meaning of the term radio remains in use today for the application of broadcast AM and FM. With the widespread use of cellular telephones, the term wireless has regained popularity and now is applied to a variety of untethered communication systems that do not use wires. The term *radio* may have evolved from the word radii because the waves propagate radially outward from the transmitter [2, p. 107], or may simply be a contraction derived from electromagnetic radiation. The IRE (Institute of Radio Engineers) officially adopted the term radio in 1913.

The Russian physicist Alexander Popov (1859–1905) also recognized the importance of Hertz's discovery of radio waves and began working on ways of receiving radio waves a year before Marconi. He is sometimes credited with using the first antenna in the first radio system by sending a signal over a 3-mile ship-to-shore path in 1897. Popov used the first elevated wire antenna, which Marconi also used. The elevated wire antenna was the primary antenna used for the next three decades. Although Popov was first, it was Marconi who developed radio commercially and also pioneered transoceanic radio communication. Marconi is considered to be the father of wireless.

Canadian-American physicist Reginald Fessenden (1866–1932) in 1906 demonstrated modulation onto a carrier by sending and receiving music, thus creating *wireless telephony* and *amplitude modulation (AM)*. Fessenden obtained 500 patents but was not a

---

[5] A very readable reference with details on the equipment Marconi used is [H.8.2: Visser, Chap. 2].

serious challenge to Thomas Edison's world record 1093 patents. Fessenden patented the *heterodyne* method of reception in 1902 and *superhetrodyne* reception in 1905.

Efficient signal detection technology developments occurred in the early 1900s, starting with patents by English electrical engineer John Fleming (1849–1945) and American Lee De Forest (1873–1961) on the *vacuum tube diode* (referred to as electric valve then) in 1904 and 1905, respectively. In 1906, De Forest patented the *triode tube* ("Audion") which remained the primary means of detection in radios until the appearance of the transistor decades later and enabled wide-spread implementation of wireless voice communication. American electrical engineer Edwin Armstrong (1890–1954) invented the *superhetrodyne radio receiver* using eight tubes in 1918, which permitted tuning of a radio receiver by merely rotating a dial. Armstrong also invented *frequency modulation* (*FM*), which became popular after World War II. Although he earned millions of dollars from his patents, Armstrong had considerable business and legal troubles and committed suicide in 1954.

In 1912, the IRE was formed from the merger of the Wireless Institute and the Society of Radio Engineers.[6] The importance of antennas is punctuated by the fact that the first article of the first issue of the journal *Proceedings of the IRE* was on an antenna topic, how to measure radiation resistance. [16] The term *antenna* was introduced by Marconi and has its origins in the wooden poles used to support wire antennas; the Italian word for pole is antenna. [H.8.2: Visser, p. 31; 15] The term *aerial* was also popular and remains in use. The term *antennae* is sometimes encountered in the literature and is the plural of antenna in Latin. The term antenna is also used for the "feelers" on the heads of insects and crustaceans such as lobsters.

Antenna developments in the early part of the 20th century were limited by the availability of signal generators, which for the most part operated in the low MHz range and below. Resonant length antennas (e.g., a half-wavelength dipole) were physically large and of low directivity. The need for secure communication in World War I provided motivation to aggressively pursue directive antennas, which are more easily realized at higher frequencies in the HF range (3 to 30 MHz, often called *short waves*) and above. In 1916, Marconi developed a cylindrical parabolic reflector with a monopole feed at 150 MHz.

Commercial broadcast radio by Westinghouse Company started in 1920 in Pittsburgh, Pennsylvania, at KDKA-AM when continuous programming of speech and music intended for the public was launched. [17, Chap. 15] Westinghouse sold home-use AM receivers starting in 1921. The use of directional broadcast antennas was introduced in 1932 in the United States. FM broadcasting began in 1940 in the United States and operated in the 42 to 50 MHz band until 1945. Popularity of FM broadcast waned in the mid 1950s due to rapid growth of broadcast *television* (*TV*), but revived in the late 1950s. Regular television broadcasting with black and white pictures began in England in 1939 and in the United States in 1941, with most growth in the 1948 to 1952 period. By the early 1960s all the networks were broadcasting color TV programming using the RCA tri-color system.

1.1.5   The Modern Era of Wireless

Mobile telephone, invented by Bell Telephone Company, was introduced into New York City police cars in 1924. In 1926, the *Yagi antenna* was developed in Japan (see Sec. 6.3). In 1934, commercial telephone microwave links began and used parabolic reflector antennas.

---

[6] The IRE merged with the American Institute of Electrical Engineers (AIEE) to form the Institute of Electrical and Electronic Engineers (IEEE) in 1963.

Microwave antennas operating at low GHz frequencies were investigated before 1900 by Bose and Lodge, but did not find applications until the 1930s when signal generators at these frequencies appeared. The term *microwaves* (now defined as 0.3 to 300 GHz frequencies) was first used in 1932. [18] The first commercial microwave radio telephone system in 1934 was operated between England and France at 1.8 GHz. The *cavity magnetron* and *reflex klystron* microwave sources were developed in 1937 and 1939, respectively. The magnetron is used for high-power microwave generation such as in microwave ovens. With microwave sources available and new applications stimulated by World War II military needs, horn antennas (see Sec. 9.4) and reflector antennas (see Sec. 9.6) invented many years earlier finally found widespread use and remain popular today for both communication and the new application, *radar* (see Sec. 4.6).[7] A U.S. government sponsored think tank for microwave technology development called MIT Radiation Laboratory operated during WW II and resulted in publication of 28 volumes, including 1 volume on antennas [H.3: Silver], mainly aperture antennas. Also, practical array antennas (see Chap. 8) started a long development process which continues today.

In the post WW II period, three major wireless application areas grew to maturity. Full mechanical scanning of reflector antennas mainly for radar was gradually replaced by feed-only mechanical scan and phased arrays with no mechanical moving parts other than mechanical phase shifters. [5] Broadcast television exploded in the 1948 to 1952 period much like broadcast radio in the early 1920s. Satellite communications began in the 1960s and now provide both wide-bandwidth interconnections of large nodes for voice, video, and data, as well as user links like mobile satellite telephones and satellite mobile broadcast radio. Finally, cellular terrestrial land-mobile radio began in the 1980s and spawned many innovations in small antennas for handsets (Chap. 11). Also, two completely new types of antennas were introduced, frequency independent antennas and microstrip antennas (see Table 1-4, Chap. 7, and Sec. 11.2). Beginning in the 1960s, computational methods for electromagnetic problems began the transition from a research area to an easy-to-use tool for the practicing engineer; see Chaps. 14 through 16. Now many commercial software packages are available for simulating complex antenna systems. Currently, antenna design is evolving from a purely device level approach in two directions: design in concert with full system requirements and integration of new materials. Antennas via wireless are a part of the growing deployments of mixed-mode communications which combines wire, coaxial cable, fiber optical cable, the Internet, and wireless.

1.2   WHAT IS AN ANTENNA AND WHEN IS IT USED?

1.2.1   What Is an Antenna?

Electrical signals are carried between points in one of two ways: via transmission line or through empty space using antennas at the terminals. A *transmission line* confines the electrical signals and the energy of the associated electromagnetic waves to the region near, or inside, the transmission line. This is also the situation for conventional circuits where no energy appears distant from the circuit. Transmission lines often use a balanced system of conductors or a metallic enclosure to confine the energy to either entirely internal to the transmission line or very nearby. An antenna has the opposite purpose—to encourage electrical signals to reach large distances from the antenna: to *radiate*. For example, a good transmitting antenna will produce power densities that are detectable at great distances from the source. The IEEE defines an antenna as "that part of a transmitting or receiving system that is designed to radiate or to receive electromagnetic

---

[7] See [H.8.2: Visser, Sec. 2.4] for a historical discussion of radar.

waves." [Sec. H.2: "IEEE Standard Definitions of Terms for Antenna"] A transmission line requires a guiding structure (typically at least one conductor), whereas an antenna requires no guiding structure. Examples of RF (radio frequency) transmission lines are coaxial cables, parallel-wire lines, and hollow pipe waveguides (see Fig. 6-31). It is appropriate to view an antenna as a transducer that converts a guided (or bound) wave on a transmission line to a free-space electromagnetic wave (for the transmitting case) or vice versa (for the receiving case).

   For an antenna to be efficient, it must have a physical extent that is at least an appreciable fraction of a wavelength at the operating frequency. Table 1-2 summarizes the electromagnetic spectrum showing common frequency bands and the corresponding wavelength values; more details on the spectrum are found in Appendix A. Conventional circuits that are smaller than 1 meter in extent, for example, are much smaller than a wavelength in size for operating frequencies up to many MHz and thus do not radiate. Conventional circuit analysis uses a lumped element model where the elements (resistors, capacitors, and inductors) are assumed to have all fields confined to the element extent. Transmission lines, on the other hand, are often a sizable fraction of a wavelength and

Table 1-2 The Electromagnetic Spectrum

Band Designation	Frequency	Wavelength	Example Uses
ELF (Extremely Low Frequency)	3 to 30 Hz	100 to 10 Mm	
SLF (Super Low Frequency)	30 to 300 Hz	10 to 1 Mm	Power lines
ULF (Ultra Low Frequency)	300 to 3 kHz	1 Mm to 100 km	
VLF (Very Low Frequency)	3 to 30 kHz	100 to 10 km	Submarine comm.
LF (Low Frequency)	30 to 300 kHz	10 to 1 km	RFID
MF (Medium Frequency)	300 kHz to 3 MHz	1 km to 100 m	AM broadcast
HF (High Frequency)	3 to 30 MHz	100 to 10 m	Shortwave broadcast
VHF (Very High Frequency)	30 to 300 MHz	10 to 1 m	FM and TV broadcast
UHF (Ultra High Frequency)	300 MHz to 3 GHz	1 m to 10 cm	TV, WLAN, GPS, Microwave ovens
SHF (Super High Frequency)	3 to 30 GHz	10 to 1 cm	Radar, WLAN, Satellite comm.
EHF (Extremely High Frequency)	30 to 300 GHz	10 to 1 mm	Radar, Radio astronomy, Point-to-point high rate data links, Satellite comm.
Microwaves	1 to 300 GHz	30 cm to 1 mm	
Millimeter waves	30 to 300 GHz	10 to 1 mm	
Submillimeter waves	>300 GHz	<1 mm	

Frequency-to-Wavelength Conversions

$$\lambda = c/f$$
$$\lambda(m) = 3 \times 10^8/f(Hz)$$
$$\lambda(m) = 300/f(MHz)$$
$$\lambda(cm) = 30/f(GHz)$$

require what is called distributed element analysis, whereby the electrical characteristics are expressed per unit length quantities such as Farads/meter. It should be mentioned that radiation can be unintended, creating interference. The design of high-speed digital circuits and high-frequency devices includes techniques for minimizing radiation (see Sec. 4.7).

1.2.2   When Is an Antenna Used?

In contrast to a transmission line, an antenna requires no guiding structure. However, the waves are radiated in all directions away from a transmitting antenna, albeit not uniformly. For a transmitter-receiver spacing of r, the power loss of a signal traveling in a transmission line is proportional to $(e^{-\alpha r})^2$ where α is the attenuation constant of the transmission line, whereas for antennas used in a line-of-sight configuration the power loss is proportional to $1/r^2$. This prompts the question of how to choose between using a cable or an antenna, which is the same as asking whether a transmission line or a radio should be used. Many factors other than propagation loss enter into this decision. Generally speaking, transmission lines are used at low frequencies and for short distances. But high frequencies are popular for communication applications because of the available wide bandwidth. So as the distance or frequency increases, the signal losses and costs of using transmission lines increase, and the decision is often to use antennas. For long distances, radio communication using transmit and receive antennas is more appropriate and much less expensive than using a transmission line.

Antennas *must* be used in several applications. For example, mobile communications involving aircraft, spacecraft, ships, or land vehicles require antennas. Antennas are also popular in broadcast situations where one transmit terminal can serve an unlimited number of receivers, which can be mobile (e.g., car radio). Non-broadcast radio applications such as municipal radio (police, fire, rescue), amateur radio, and wireless personal communications also require antennas. There are also many non-communication applications for antennas. These include remote sensing and industrial applications. Remote sensing systems are either active (e.g., radar) or passive (e.g., radiometry) and receive scattered energy or inherent emissions from objects, respectively; see Secs. 4.5 and 4.6. The received signals are processed to infer information about the objects or scenes. Industrial examples are cooking and drying with microwaves.

Other factors that influence the choice of the type of transmission system include historical reasons, security, and reliability. Telephone companies provided wireline communications before radio technology was available, and telephone infrastructure continues to be used, although consumers continue to shift to wireless personal communications. Telephone companies employ microwave radio and fiber optic transmission lines for long-distance telephone calls. Satellite radio links are used for international telephone calls and for distributing television program material to affiliate stations. Also, many millions of homes receive television programs from direct broadcast satellites. Very small aperture terminals (VSAT) are widely used in private data networks to interconnect, for example, gas stations and retail stores via satellite, mainly for credit card authorization. Although transmission lines inherently offer more security than radio, radio links can be secured with coding techniques.

Reliability is an additional factor impacting the decision between wired and wireless interconnectivity. Radio signals are affected by environmental conditions such as structures along the signal path, the ionosphere, and weather. Furthermore, interference is always a threat to radio systems. On the other hand, cables are vulnerable to being damaged, most often by storms (for elevated cables) and by accidentally digging up buried cables. All of these factors must be examined together with the costs associated with using transmission lines or antennas. Cable-based systems usually require the purchase or lease of land right-of-way, which can be expensive. Radio-based systems only have costs associated with antenna sites and benefit from improved reliability and

reduced equipment costs as technology advances. Cable and radio communication systems will continue to be used in the future with the choice depending on the specific application.

The next two sections of this first chapter provide a basic understanding of how antennas operate and an overview of the types of antennas encountered in practice.

1.3   HOW ANTENNAS RADIATE

In this section, we explain the basic principles of radiation using some examples explained with simplified physics in preparation for the full mathematical development of antennas in the next chapter. The mathematical foundation is needed for subsequent analysis and design of antennas. Radiation is a disturbance of the electromagnetic fields that propagates away from the source of the disturbance so that the total power in the wave does not decay with distance. This disturbance is created by a time-varying current source, which has an accelerated charge distribution associated with it. We, therefore, begin our discussion of radiation with a single accelerated charge.

Consider the single electric charge of Fig. 1-4 moving with constant velocity in the z-direction. We know from electrostatic principles that the electric field (often called the Coulomb field) of a charge terminates either on an oppositely signed charge or at infinity. Thus, prior to arrival at point A ($t < 0$), the static electric field lines extend radially away from the charge to infinity and move with the charge. When the charge reaches point A at time $t = 0$ the charge begins to be accelerated (i.e., velocity is increased) until reaching
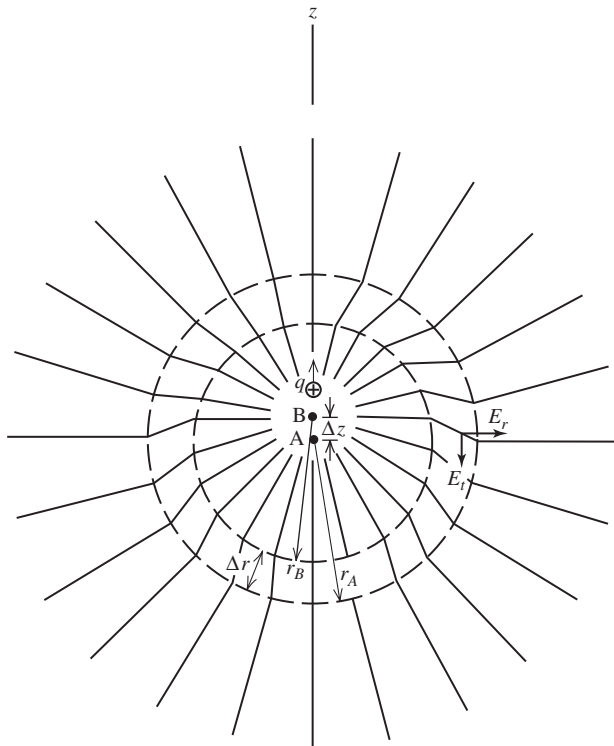
Figure 1-4 Illustration of how an accelerated charged particle radiates. Charge q moves with constant velocity in the $+z$-direction until it reaches point A (time $t = 0$) after which it accelerates to point B (time $t = \Delta t$) and then maintains its velocity. The electric field lines shown here are for a time r_B/c after the charge passed point B.

point B at time $t = \Delta t$ where it continues on at the acquired velocity for all time $t > \Delta t$. The radial field lines outside the circle of radius r_A in Fig. 1-4 originated when the charge was at point A. The circle of radius r_B is centered on point B, which is the charge position at the end of the acceleration period Δt and the electric field lines extend radially away from point B interior to the circle or radius r_B. The distance between the circles is the distance light would travel in time Δt, or $\Delta r = r_B - r_A = \Delta t/c$. Because the charge moves slowly compared to the speed of light (c), $\Delta z \ll \Delta r$ and the circles are nearly concentric. The distance Δz in Fig. 1-4 is shown large relative to Δr for clarity. The electric field lines in the Δr region are joined together because of the required continuity of electric field lines in the absence of charges. This region is obviously one of disturbed field structure and was caused by acceleration of the charge which ended a time r_B/c earlier than the instant represented in Fig. 1-4. This disturbance propagates outward at the speed of light. The electric field lines in the Δr region have a "kink" in them which introduces a transverse component E_t in addition to a radial component. As we shall see, the presence of the transverse field component is a characteristic of radiated fields, and the disturbance propagates to infinity.

The important physical principle in this example is that accelerated charges radiate. Acceleration, in general, occurs through change of speed or change of direction of a charge. For antennas the primary mechanism of radiation is via charges (i.e., charge disturbances) moving back and forth on a wire, or oscillating in response to an oscillating generator. Charge direction is reversed at the ends of the wire, which is an acceleration and results in radiation.

The fact that radiation is a disturbance is directly analogous to a transient wave created by a stone dropped into a calm lake where the disturbance of the lake surface continues to propagate radially away from the impact point long after the stone has disappeared. Another useful analogy of wave behavior, credited to Leonardo da Vinci (1452–1519), is the waves moving across the tops of a wind-swept field of wheat while the base of the stalks remain stationary. [H.3: Schelkunoff and Friis, p. 10] Accelerated charges are necessary for creating radiation, but not for sustaining the wave. Once a wave is started, the electric field lines that close on themselves (because no charges are present for the lines to terminate on) and propagate outward away from the original source at the speed of light. The helpful mechanical analogy here is that of a rope with one end fixed and a kink is generated at the other end by a flick of the wrist, creating a propagating kink.

The directional properties of radiation are evident in the accelerated charge example. The disturbance in Fig. 1-4 is maximum in a direction perpendicular to the charge acceleration direction where the transverse electric field component is maximum. We will show in the next chapter that maximum radiation occurs perpendicular to a straight wire antenna which carries the accelerated charges.

The transient excitation example of a single accelerated charge is the most elementary radiation case. The common situation is steady-state (usually sinusoidal) charge oscillations. The frequency of the charge oscillations determines the frequency of the radiated wave. This is illustrated with a pair of equal but stationary, oppositely signed charges separated by a distance much smaller than the wavelength of the oscillation frequency of the charges. That is, the two charges oscillate in amplitude sinusoidally and in phase, other than the opposite sign difference. Conceptually, the oscillation of charges back and forth between the ends occurs by an oscillating uniform current called an ideal dipole. The fields created by this oscillating charge dipole for a few instants of time are shown in Fig. 1-5. [H.3: Hertz, pp. 144–145; H.3: Schelkunoff and Friis, Sec. 4.17] A wire dipole antenna is shown also to indicate that the radiation is the same as that of a small dipole antenna to be discussed in Sec. 2.3. The dipole provides a current path to allow charge distributions to alternate back and forth as shown later for the ideal dipole of Fig. 3-2. In Fig. 1-5, an oscillating current of frequency f (and period $T = 1/f$) was turned on a quarter period before $t = 0$. The upward-flowing current creates an excess of charges on the upper half of the dipole and a deficit of charges on the lower half.

(a) $t = 0$

(b) $t = T/4$

(c) $t = T/2$

$\longleftarrow$ $\lambda/2$ $\longrightarrow$
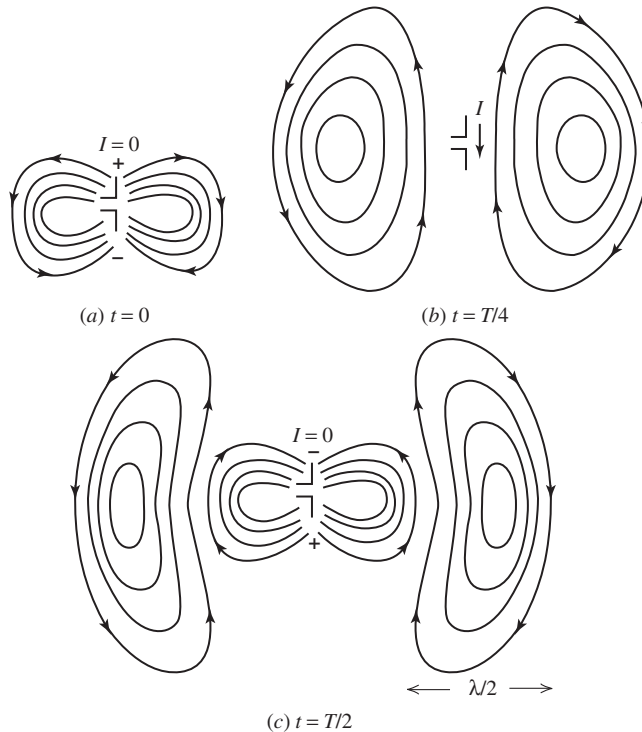
Figure 1-5 Fields of an oscillating charge dipole for various instants of time. Electric field lines are shown on one side of the dipole but exist everywhere around the dipole. The oscillations are of frequency f with a period of $T = 1/f$. The separation between the equal, opposite-sign charges is much smaller than the wavelength of the oscillations. A dipole antenna is shown to indicate that the radiation is the same as that for an ideal dipole antenna.

Peak charge build-up occurs at $t = 0$ as shown in Fig. 1-5a and produces a voltage between the dipole halves. The positive charges on the top are attracted to the negative charges on the bottom half of the dipole, creating a current. The current is maximum at $t = T/4$ as shown in Fig. 1-5b, at which time the charges have been neutralized and there are no longer charges for termination of electric field lines, which form closed loops near the dipole. During the next quarter cycle, negative charges accumulate at the top end of the dipole as shown in Fig. 1-5c. Near the dipole, the fields are most intense normal to the oscillating charges on the dipole, just as we found with the single accelerated charge. As time progresses, the electric field lines detach from the dipole, forming closed loops in space, and spread outward from the dipole as time progresses. Viewed in terms of current, the conduction current on the antenna converts to a displacement current in space, consisting of longitudinal fields near the antenna and solenoidal (as created by wire loops) fields away from the antenna. Thus, continuity is satisfied. This process continues, producing radiation via electric field components that are transverse to the radial direction and propagate to large distances from the antenna. This example illustrates the general principle that if charges are accelerated back and forth (i.e., oscillate), a regular disturbance is created and radiation is continuous. Antennas are designed to support charge oscillations.

This oscillating dipole example assumes that the charge separation distance is small compared to a wavelength. For dipoles that are not of infinitesimal length, radiation occurs continuously along the wire [19]; dipoles of various lengths are addressed in Sec. 6.1. Another important simplification assumed in the oscillating dipole model is that
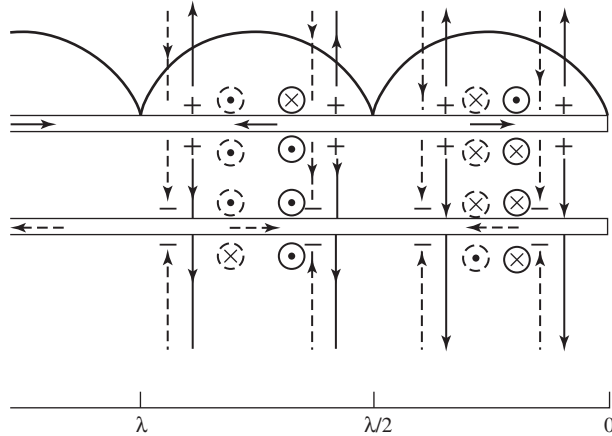
(a) Open-circuited transmission line showing currents, charges, and fields. The electric fields are indicated with lines and the magnetic fields with arrow heads and tails, solid (dashed) for those arising from the top (bottom) wire. (Note: The separation between the wires is larger than for a normal transmission line to permit showing the interior fields.)
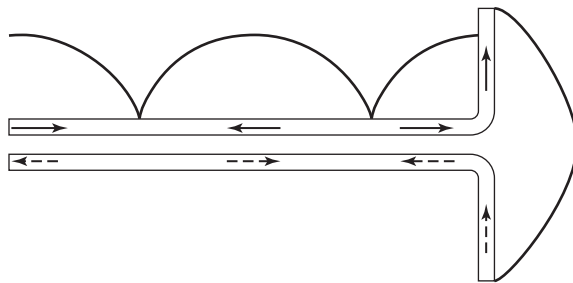
(b) Peak currents and fields of a half-wavelength dipole created by bending out the ends of the transmission line.

Figure 1-6 Evolution of a dipole antenna from an open-circuited transmission line.

charges actually move to the dipole ends. For finite length antennas, the charges barely move at all and are only displaced enough to support the disturbance created by the generator [20]; again the wind waves in the wheat field help to visualize the situation. Another helpful analogy is a row of collapsing dominoes, where the dominoes fall at a speed much less than the wave propagating across the tops of the dominoes while the bases of the dominoes remain stationary. However, it is convenient to treat charge disturbances resulting in an effective current as if charges actually move along the length of the antenna.[8] The same assumption is also employed in conventional circuit analysis.

We will see in Sec. 2.3 that the mathematical solution of the oscillating dipole produces the property required for successful radiation: that fields die off with distance as $1/r$. In contrast, the electric fields of an electrostatic dipole die off as $1/r^3$. The time-space behavior of fields from an antenna are revisited in Chap. 15.

The key to understanding how an antenna radiates lies in Maxwell's discovery that a transmission line acts as a guide for waves associated with the fields surrounding the transmission line. An antenna is a structure that encourages waves to extend out into the space surrounding the antenna rather than being bound to it. We exploit this concept

---

[8] From this point forward, for simplification, we will adopt the traditional model for antenna currents as that of charge transport along the antenna, where the current is understood to be an effective current representing charge displacements from equilibrium transferred to neighboring charges.

in the next example that begins with the open-circuited transmission line of Fig. 1-6*a*. The open-circuit termination presents large discontinuity that creates reflections which interact with the incident wave, producing a standing wave pattern with a zero current magnitude at the wire end and nulls every half wavelength from the end. The currents are in opposite directions on the transmission line wires, as indicated by arrows in Fig. 1-6*a*. In transmission lines, the conductors guide the waves and the power resides in the region surrounding the conductors as manifested by the electric and magnetic fields shown in Fig. 1-6*a*. The electric fields originate from or terminate on charges on the wires and are perpendicular to the wires, and the magnetic fields encircle the wires. Note that all fields reinforce between the wires and cancel elsewhere. This is true for a wire spacing that is much smaller than a wavelength, which is usually the case. If the ends of the wires are bent outward as shown in Fig. 1-6*b*, the reinforced fields between the wires are exposed to open space. Note that the currents on the vertical wire halves, which are each a quarter-wavelength long in this case, are no longer opposed as with the transmission line, but are both upwardly directed for this instant of time. In reality, the dipole current envelopes are approximately sinusoidal as shown, but the transmission line currents are not pure standing waves because the impedance presented by the antenna is a better match to the characteristic impedance of the transmission line compared to the open circuit discontinuity of the open circuited transmission line. Fig. 1-6*b* shows the peak envelope of the current. As time proceeds, the current oscillates and disturbances are created which propagate away from the wire, much as for the accelerated single charge case. The field structure for this half-wave dipole resembles that of the ideal dipole of Fig. 1-5. Wire antennas will be discussed in detail in Chap. 6.

1.4 THE FOUR ANTENNA TYPES

Most antennas are reciprocal devices and behave the same on transmit as receive, so antennas are treated as transmitting or receiving as appropriate for the best understanding of the particular situation. In the receiving mode, antennas act to collect incoming waves and direct them to a common feed point where a transmission line is attached. In some cases, antennas focus radio waves just as lenses focus optical waves. In all cases, antennas have directional characteristics; that is, electromagnetic power density is radiated from a transmitting antenna with varying intensity around the antenna.

In this section, we introduce the parameters used to evaluate antennas and then discuss the four types of antennas. Antenna performance parameters are listed in Table 1-3 and will be defined and developed in more detail in the next chapter. The *radiation pattern* (or simply, *pattern*) gives the variation of radiation with angle around an antenna when the antenna is transmitting. When receiving, the antenna responds to an incoming wave from a given direction according to the pattern value in that direction. The typical pattern in Fig. 1-7 shows the pattern main beam and side lobes. Radiation is quantified by noting the value of power density S at a fixed distance r from the antenna. A directive antenna has a single narrow main beam, which is the example shown in Fig. 1-7, and finds application in point-to-point communications. In some applications, the shape of the main beam is important. In others, an omnidirectional pattern with constant radiation in one plane is used in broadcast situations to achieve uniform coverage in all directions in the horizontal plane around the antenna.

An antenna is essentially a spatial amplifier, and *directivity* expresses how much greater the peak radiation level is for an antenna than it would be if all the radiated power were distributed uniformly around the antenna. Fig. 1-7 shows the radiation pattern of a real antenna compared to an *isotropic* spatial distribution; also see later Fig. 2-14. The radiation enhancement in preferred directions that can be achieved by an antenna is evident. Gain G is directivity reduced by the losses on the antenna.

The third parameter of *polarization* describes the vector nature of electric fields radiated by an antenna. The figure traced out with time by the tip of the instantaneous

Table 1-3 Antenna Performance Parameters

- **Radiation Pattern**, $F(\theta, \phi)$: Angular variation of radiation around the antenna, including:
  Directive single or multiple narrow beams
  Omnidirectional (uniform radiation in one plane)
  Shaped main beam
- **Directivity, D**: Ratio of power density in the direction of the pattern peak to the average power density at the same distance from the antenna.
- **Gain, G (or radiation efficiency e_r, where $G = e_r D$)**: Directivity reduced by the losses on the antenna.
- **Polarization**: The figure traced out with time by the instantaneous electric field vector associated with the radiation from an antenna when transmitting. Antenna polarizations: Linear, Circular, Elliptical
- **Impedance, Z_A**: The input impedance at the antenna terminals.
- **Bandwidth**: Range of frequencies over which important performance parameters are acceptable.
- **Scanning**: Movement of the radiation pattern in space. Scanning is accomplished by mechanical movement or by electronic means such as adjustment of antenna current phase.
- **System Considerations**: Mechanical considerations (size, weight, aerodynamics, vibration, positioning accuracy), environmental aspects (effects of wind, rain, temperature, altitude), scattering/radar cross section, esthetic appearance.
- **Special Considerations for transmitting antennas**: Power handling, intermodulation, radiation hazards.
- **Special Considerations for receiving antennas**: Noise.
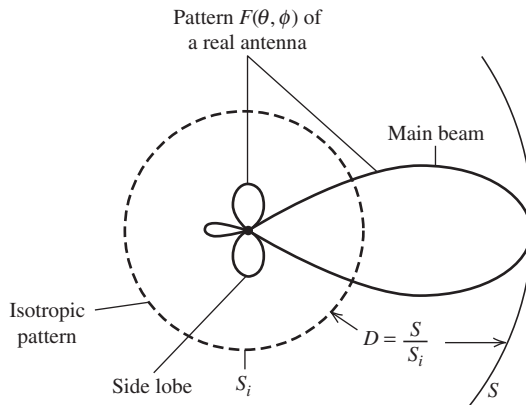
Figure 1-7 Illustration of radiation pattern $F(\theta, \phi)$ and directivity D. The power densities at the same distance are S and S_i for the isotropic and real antennas, respectively.

electric field vector determines the polarization of the wave. A straight-wire transmitting antenna produces a wave with linear polarization parallel to the wire. Another popular polarization is circular. In general, polarization is elliptical. A *dual-polarized antenna* enables the doubling of communication capacity by carrying separate information on orthogonal polarizations over the same physical link on the same frequency.

The *input impedance* of an antenna is the ratio of the voltage to current at the antenna terminals. The usual goal is to match antenna input impedance to the characteristic impedance of the connecting transmission line. *Bandwidth* is the range of frequencies with acceptable antenna performance as measured by one or more of the performance parameters; see Sec. 7.1 for commonly used definitions of bandwidth. Finally, it is often desired to scan the main beam of an antenna over a region of space. This can be accomplished by mechanically rotating the entire antenna, by electronic scanning, or by a combination of mechanical and electronic scanning.

There are tradeoffs among parameter values. Usually performance cannot be improved significantly for one parameter without sacrificing one or more of the other parameter levels. This is the antenna design challenge.

Antennas can be divided into four basic types by their major performance characteristics. These antenna types are introduced so that the common features can be grasped early in the study of antennas. When you encounter a new antenna, try to determine which type it is. The four antenna types are listed in Table 1-4 in the order that they are commonly used by frequency range. Electrically small antennas are used at VHF frequencies and below. Resonant antennas are mainly used from HF to low GHz frequencies. Broadband antennas are mainly used from VHF to middle GHz frequencies. Aperture antennas are mainly used at UHF and above. This discussion serves as an overview and should be referred to from time to time as your knowledge of antennas builds.

Electrically small (or simply, *small*) *antennas* are much less than a wavelength in extent. They are simple in structure and their properties are not sensitive to construction details. The vertical monopole used for AM reception on cars is a good example. It is about 0.003 λ long and has a pattern that is nearly omnidirectional in the horizontal plane. This is often a desirable property, but its low input resistance and high input reactance are serious disadvantages. Also, small antennas are inefficient because of significant ohmic losses on the structure.

Resonant antennas are popular when a simple structure with good input impedance over a narrow band of frequencies is needed. It has a broad main beam and low or moderate (a few dB) gain. The half-wave dipole is a prominent example.

Many applications require an antenna that operates over a wide frequency range. A *broadband antenna* has acceptable performance as measured with one or more parameters (pattern, gain, and/or impedance) over a 2:1 bandwidth ratio of upper to lower operating frequency. A broadband antenna is characterized by an active region. Propagating (or traveling) waves originate at the feed point and travel without radiation to the active region where most of the power is radiated. A broadband antenna with circular geometry has an active region where the circumference is one wavelength and produces circular polarization. An example is the spiral antenna illustrated in, Fig. 1-8 which can have a 20:1 bandwidth. A broadband antenna made up of linear elements or straight edges has an active region where the elements are about a half-wavelength in extent and produces linearly polarized radiation parallel to the linear elements. Because only a portion of a broadband antenna is responsible for radiation at a given frequency, the gain is low. But it may be an advantage in many applications to have gain that is nearly constant with frequency, although low. Also, the traveling wave nature of a broadband antenna means that it has a real-valued input impedance that is easily matched to the connecting transmission line.

Aperture antennas have an opening through which propagating electromagnetic waves flow. The horn antenna shown in Fig. 1-8 is a good example; it acts as a "funnel" for waves. The aperture is usually several wavelengths long in one or more dimensions. The pattern usually has a narrow main beam, leading to high gain. Bandwidth is moderate (as much as 2:1). One property that can be a disadvantage is that the pattern main beam narrows with increasing frequency for a fixed physical aperture size.

Table 1-4 Types of Antennas

- **Electrically Small Antennas:** The extent of the antenna structure is much less than a wavelength λ.
  Properties:
    Low directivity
    Low input resistance
    High input reactance
    Low radiation efficiency

Examples:

Short dipole   Small loop

$\leftarrow\ \ll\lambda\ \rightarrow$ $\leftarrow\ \ll\lambda\ \rightarrow$
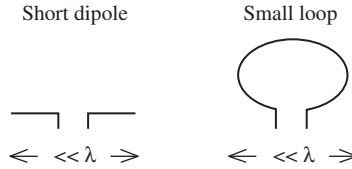
- **Resonant Antennas:** The antenna operates well at a single or selected narrow frequency bands.
  Properties:
    Low to moderate gain
    Real input impedance
    Narrow bandwidth

Examples:

Half-wave dipole   Microstrip patch   Yagi

$\longleftarrow\ \sim\frac{\lambda}{2}\ \longrightarrow$ $\leftarrow\ \sim\frac{\lambda}{2}\ \rightarrow$ $\sim\frac{\lambda}{2}$
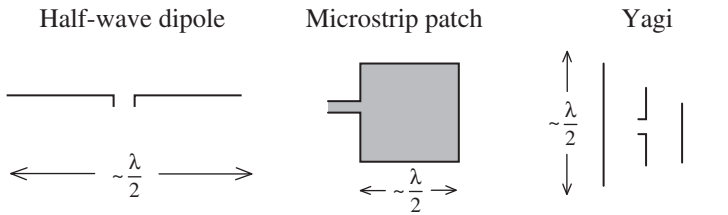
- **Broadband Antennas:** The pattern, gain, and impedance remain acceptable and are nearly constant over a wide frequency range.
  Characterized by an active region with a circumference of one wavelength or an extent of a half-wavelength, which relocates on the antenna as frequency changes.
  Properties:
    Low to moderate gain
    Constant gain
    Real input impedance
    Wide bandwidth

Examples:

Spiral   Log Periodic Dipole Array
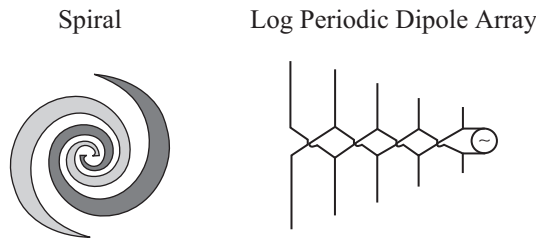
<div align="center">**Table 1-4** Continued</div>

- **Aperture Antennas:** Has a physical aperture (opening) through which waves flow.
  Properties:
    High gain
    Gain increases with frequency
    Moderate bandwidth

Examples:
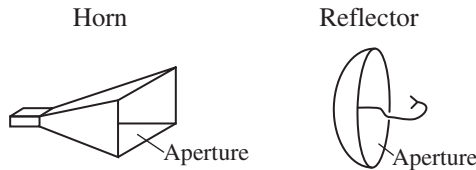
A common design situation is the search to find an antenna to achieve a desired gain value. The list below provides a guide in this selection process.[9]

- **5 dB or less**
  Electrically small antennas
  Loops
  Dipoles/monopoles
- **5 dB to 8 dB**
  Microstrip patches
  Planar frequency-independent antennas (e.g. spirals)
- **8 dB to 15 dB**
  Yagi-Uda
  Helix (axial mode)
  Log periodic dipole array
- **15 dB and more**
  Aperture antennas
    Horns
    Reflectors

Also, multiple elements can be used in an array configuration to increase gain.

What is the future for antenna engineering? An interesting aspect of the field of antennas is that new antenna innovations generally do not replace existing antenna types, as is often the case in electronics and computers. Long-used antennas will find new applications along with the latest innovative solutions. Fundamental physics limitations prevent unlimited miniaturization of conventionally designed antennas, which is often the case in electronic devices through higher scale integration. But the demand is for more pervasive communications to enable each person and device to be free of wire connections. An often quoted goal is be able to reach any person or device worldwide without connections. A "wireless" society is possible only through the use of antennas. Indeed, the future of antenna engineering is very bright. Sec. 4.5 will give a technical overview of antennas in wireless communications.

---

[9] Gain in dB is 10 log of the gain as a power ratio.

REFERENCES

1. Sir Edmund Whittaker, *A History of the Theories of Aether and Electricity*, Vol. 1: *The Classical Theories*, Harper Torchbooks, New York, 1960. First published by Thomas Nelson and Sons, 1910.
2. T. Sarkar, R. Mailloux, A. Oliner, M. Salazar-Palma, and D. Sengupta, *History of Wireless*, Wiley, 2006.
3. E. Larsen, *Telecommunications—A History*, Frederick Muller Ltd., London, 1977.
4. Isaac Asimov, *Asimov's Biographical Encyclopedia of Science and Technology*, revised 2nd edition, Doubleday, Garden City, New York, 1982.
5. Jack Ramsay, "Highlights of Antenna History," *IEEE Ant. & Prop. Soc. Newsletter*, pp. 8–20, Dec. 1981. (Reprinted from IEEE Communications Magazine, Sept. 1981.)
6. A. D. Olver, "A Century of Antenna Development," IEE Conference Publication No. 274, Vol. I, 1987.
7. L. Campbell and W. Garnett, *The Life of James Clerk Maxwell*, 1882. Available from www .sonnettsoftware.com.
8. James Rautio, "In Search of Maxwell," *Microwave J.*, Vol. 49, pp. 76–88, July 2006.
9. James Rautio, "Twenty Three Years: The Acceptance of Maxwell's Theory," *Microwave J.*, Vol. 51, pp. 104–116, July 2008.
10. Albert Einstein, *Out of My Later Years*, Wisdom Library, 1950.
11. John D. Kraus, "Heinrich Hertz—Theorist and Experimenter," *IEEE Trans. on Micro. Th. and Tech.*, Vol. 36, pp. 824–829, May 1988.
12. J. H. Bryant, "The First Century of Microwaves—1886 to 1986," *IEEE Trans. on Micro. Th. and Tech.*, Vol. 36, pp. 830–858, May 1988.
13. C. Susskind, "Heinrich Hertz: A Short Life," *IEEE Trans. on Micro. Th. and Tech.*, Vol. 36, pp. 802–805, May 1988.
14. J. Ramsay," Microwave Antenna and Waveguide Techniques before 1900," *Proc. IRE*, Vol. 46, pp. 405–415, Feb. 1958.
15. G. Pelosi, S. Selleri, and B. Valotti, "Antennae," *IEEE Vehicular Technology Soc. Newsletter*, Vol. 42, pp. 14–16, Aug. 2000.
16. M. I. Pupin, "A Discussion on Experimental Tests of the Radiation Law for Radio Oscillators," *Proc. of the Institute of Radio Engineers*, Vol. 1, pp. 3–13, 1913.
17. L. Lichty and M. Topping, *American Broadcasting: A Source Book on the History of Radio and Television*, Hastings House, New York, 1975.
18. G. Pelosi, "The Birth of the Term Microwaves," *Proceedings of IEEE*, Vol. 84, p. 326, Feb. 1996.
19. E. K. Miller, "The Incremental Far Field and Degrees of Freedom of the Sinusoidal Current Filament," *IEEE Trans. Ant. and Prop. Mag.*, Vol. 49, pp. 13–20, Aug. 2007.
20. H. E. Green, "A Comment on Charge Motion on a Dipole Antenna," *IEEE Trans. Ant. and Prop. Mag.*, Vol. 50, pp. 174–175, Aug. 2008.

PROBLEMS

1.3-1   Show that the decay of power density for a wave radiated by an antenna ($1/r^2$) results from the increase in the area of a sphere centered on the antenna and the fact that the total radiated power is conserved.

1.3-2   Consider an ideal dipole of length $L \ll \lambda$. If charge carriers actually moved along the full length of the wire in half cycle, $T/2$, show that the speed of the charges is much less than the speed of light (i.e., $\nu \ll c$).

1.4-1   *Project: Antenna hunt.* (a) Locate one representative antenna for each of the four major antenna type categories. Preferably the antennas will be in your community and be actively used. Take a photograph, or make a sketch, of the antenna and its surroundings. If you cannot find an example in your community, you can use other sources such as a catalog, magazine, or the Internet. (b) Prepare a brief report that shows each of the four antennas and includes information about the antenna such as the type of antenna, operating frequency, purpose and use, and any other information you can determine.

Chapter 2

Antenna Fundamentals

The overview of the principles of electromagnetics and the application to antennas presented in the first chapter is followed in the first two sections of this chapter with a presentation of the fundamental principles of electromagnetics in a mathematical form suitable for direct application to antenna analysis and design. The principles are applied in Sec. 2.3 to the important classical problem of an ideal dipole. The remaining sections in the chapter introduce the student to the basic antenna quantities of pattern, directivity and gain, impedance, efficiency, and polarization.

2.1 FUNDAMENTALS OF ELECTROMAGNETICS

The formulations of electromagnetics for antenna problems are in vector form and usually expressed in spherical coordinates. This is because electromagnetic fields, unlike scalar fields such as sound waves, have polarization, and vectors are a convenient way to represent polarization, which can be simplistically described as the orientation of the electric field. Spherical coordinates are required because antennas radiate in all directions (i.e., three dimensions) and the fields are expressed as a function of the spherical coordinate angles θ and ϕ around the antenna. Although antenna problems involve vectors and spherical coordinates, nonlinear conditions rarely arise as they often do in mechanical engineering problems. In fact, applications of electromagnetics are primarily to linear, homogeneous, isotropic systems, thus avoiding several mathematical complications.

This section is a concise review of material treated in electromagnetic fields courses in electrical engineering or physics. Any basic electromagnetics textbook can be consulted for more details.

The fundamental electromagnetic equations in the time domain are[1]

$$\nabla \times \boldsymbol{\mathscr{E}} = -\frac{\partial \boldsymbol{\mathscr{B}}}{\partial t} \tag{2-1}$$

$$\nabla \times \boldsymbol{\mathscr{H}} = \frac{\partial}{\partial t}\boldsymbol{\mathscr{D}} + \boldsymbol{\mathscr{I}}_{\mathbf{T}} \tag{2-2}$$

$$\nabla \cdot \boldsymbol{\mathscr{D}} = \rho_T(t) \tag{2-3}$$

$$\nabla \cdot \boldsymbol{\mathscr{B}} = 0 \tag{2-4}$$

$$\nabla \cdot \boldsymbol{\mathscr{I}}_{\mathbf{T}} = -\frac{\partial}{\partial t}\rho_T(t) \tag{2-5}$$

---

[1] Time-varying quantities will be denoted with script quantities, for example, $\boldsymbol{\mathscr{E}} = \boldsymbol{\mathscr{E}}(x, y, z, t)$.

The first four of these differential equations are frequently referred to as Maxwell's equations and the last as the continuity equation. The curl equations together with the continuity equation are equivalent to the curl and divergence equations. In time-varying field problems, the curl equations with the continuity equation is the most convenient formulation. Each of these differential equations has an integral counterpart.

If the sources $\rho_T(t)$ and $\mathscr{J}_\mathbf{T}(t)$ vary sinusoidally with time at radian frequency ω, the fields will also vary sinusoidally and are frequently called time-harmonic fields. The fundamental electromagnetic equations and their solutions are considerably simplified if phasor fields are introduced as follows[2]:

$$\mathscr{E} = \mathrm{Re}(\mathbf{E}e^{j\omega t}), \quad \mathscr{H} = \mathrm{Re}(\mathbf{H}e^{j\omega t}), \quad \text{etc.} \tag{2-6}$$

where phasor quantities $\mathbf{E}$, $\mathbf{H}$, $\mathbf{D}$, $\mathbf{B}$, ρ_T, and $\mathbf{J_T}$ are complex-valued functions of spatial coordinates only (i.e., time dependence is not shown). Using the phasor definitions of the electromagnetic quantities from (2-6) in (2-1) to (2-5) and eliminating the $e^{j\omega t}$ factors that appear on both sides of the equations yields

$$\nabla \times \mathbf{E} = -j\omega\mathbf{B} \tag{2-7}$$

$$\nabla \times \mathbf{H} = j\omega\mathbf{D} + \mathbf{J_T} \tag{2-8}$$

$$\nabla \cdot \mathbf{D} = \rho_T \tag{2-9}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{2-10}$$

$$\nabla \cdot \mathbf{J_T} = -j\omega\rho_T \tag{2-11}$$

The time derivatives in (2-l) to (2-5) have been replaced by a $j\omega$ factor in (2-7) to (2-11) and time-varying electromagnetic quantities have been replaced by their phasor counterpart. This process is similar to the solution of circuit equations where the time-dependent differential equations are Laplace-transformed and the time derivatives are thus replaced by $j\omega$ (or s). Eqs. (2-7) to (2-10) are often referred to as the time-harmonic form of Maxwell's equations because they apply to sinusoidally varying (i.e., time-harmonic) fields.

If more than one frequency is present, the time-varying forms of the electromagnetic quantities can be found by inverse transforms after (2-7) to (2-11) have been solved for the phasor quantities as a function of radian frequency ω. This is again analogous to the procedure used to solve circuit problems. Fortunately, this is not usually necessary in antenna problems because the bandwidth of the signals is usually very small. In the typical case, a carrier frequency is accompanied by some form of modulation giving a spread of frequencies around the carrier. For analysis purposes, we use a single frequency equal to the carrier frequency. Thus, unless otherwise noted, subsequent material in this book will assume time-harmonic fields.

The total current density $\mathbf{J_T}$ is composed of an impressed, or source, current $\mathbf{J}$ and a conduction current density term $\sigma\mathbf{E}$, which occurs in response to the impressed current:

$$\mathbf{J_T} = \sigma\mathbf{E} + \mathbf{J} \tag{2-12}$$

The role played by the impressed current density is that of a known quantity. It is quite frequently an assumed current density on an antenna, but as far as the field equations are concerned, it is a known function. The current density $\sigma\mathbf{E}$ is a current density flowing on a nearby conductor due to the fields created by source $\mathbf{J}$ and can be computed after the field

---

[2] The student is cautioned that some authors use $e^{-j\omega t}$, which leads to sign differences in subsequent developments.

equations are solved for **E**. In addition to conductivity σ, a material is further charac-terized by permittivity ε and permeability μ, where[3]

$$\mathbf{D} = \varepsilon\mathbf{E} \tag{2-13}$$

and

$$\mathbf{B} = \mu\mathbf{H} \tag{2-14}$$

We now rewrite the field equations in preparation for their solution. Substituting (2-12) and (2-13) into (2-8) gives

$$\nabla \times \mathbf{H} = j\omega\left(\varepsilon + \frac{\sigma}{j\omega}\right)\mathbf{E} + \mathbf{J} = j\omega\varepsilon'\,\mathbf{E} + \mathbf{J} \tag{2-15}$$

where we have defined $\varepsilon' = \varepsilon - j(\sigma/\omega)$. For antenna problems, we are usually solving for the fields in air surrounding an antenna where $\sigma = 0$ and $\varepsilon' = \varepsilon$. We therefore use ε instead of ε' in subsequent developments. However, if the conductivity is nonzero, ε can be replaced by $\varepsilon' = \varepsilon - j(\sigma/\omega)$. Note also that **E** and **H** are the fields of primary interest in antennas.[4] They are referred to as electric and magnetic fields and have units of V/m and A/m, respectively.

Let ρ be the source charge corresponding to the source current density **J**. Then using (2-12) to (2-14) in (2-7) and (2-9) to (2-11), and repeating (2-15) permits Maxwell's equations to be expressed in terms of **E** and **H** (see Prob. 2.1-2 for (2-18)):

$$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H} \tag{2-16}$$

$$\nabla \times \mathbf{H} = j\omega\varepsilon\mathbf{E} + \mathbf{J} \tag{2-17}$$

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon} \tag{2-18}$$

$$\nabla \cdot \mathbf{H} = 0 \tag{2-19}$$

$$\nabla \cdot \mathbf{J} = -j\omega\rho \tag{2-20}$$

These are the time-harmonic electromagnetic field equations with source current density **J** and source charge density ρ shown explicitly. Sometimes, it is convenient to introduce a fictitious magnetic current density **M**. Then (2-16) becomes

$$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H} - \mathbf{M} \tag{2-21}$$

Magnetic currents are useful as equivalent sources that replace complicated electric fields.

The solution of the fundamental electromagnetic equations is not complete until the boundary conditions are satisfied. A sufficient set of boundary conditions in the time-harmonic form is

$$\hat{\mathbf{n}} \times (\mathbf{H_2} - \mathbf{H}_1) = \mathbf{J}_s \tag{2-22}$$

$$(\mathbf{E_2} - \mathbf{E_1}) \times \hat{\mathbf{n}} = \mathbf{M}_s \tag{2-23}$$

where the electric and magnetic surface currents $\mathbf{J}_s$ and $\mathbf{M}_s$ flow on the boundary between two homogeneous media with constitutive parameters ε_1, μ_1, and σ_1, and ε_2, μ_2, σ_2. $\mathbf{M}_s$ is

---

[3] In general, ε and μ can be complex, but in most antenna problems they can be approximated as real constants.
[4] Sometimes fields other than **E** and **H** are used as the two fundamental field quantities, such as **E** and **B**. [1]

zero unless an equivalent magnetic current sheet is used. The unit normal to the boundary surface $\hat{\mathbf{n}}$ is directed from Medium 1 into Medium 2. The cross products with the unit normal form the tangential components to the boundary, and these equations can be written as

$$H_{\tan 2} = H_{\tan 1} + J_s \tag{2-24}$$

$$E_{\tan 2} = E_{\tan 1} + M_s \tag{2-25}$$

These boundary conditions are derived from the integral form of (2-17) and (2-21). If one side is a perfect electrical conductor, the boundary conditions become

$$H_{\tan} = J_s \tag{2-26}$$

$$E_{\tan} = 0 \tag{2-27}$$

The tangential boundary conditions on the magnetic field intensity are illustrated in Fig. 2-1 for the general case and for the case where Medium 1 is a perfect conductor. It is important to note that all field quantities in the boundary condition equations are evaluated at the boundary and the equations apply to each point along the boundary.

Also derivable from Maxwell's curl equations is a conservation of power equation, or Poynting's theorem. Consider a volume v bounded by a closed surface s. The complex power P_s delivered by the sources in v equals the sum of the power P_f flowing out of s, the time-average power $P_{d_{av}}$ dissipated in v, plus the time-average stored power in v:

$$P_s = P_f + P_{d_{av}} + j2\omega(W_{m_{av}} - W_{e_{av}}) \tag{2-28}$$

The complex power flowing out through closed surface s is found from

$$P_f = \frac{1}{2} \oiint_s \mathbf{E} \times \mathbf{H^*} \cdot d\mathbf{s} \tag{2-29}$$

where $d\mathbf{s} = ds\hat{\mathbf{n}}$ and $\hat{\mathbf{n}}$ is the unit normal to the surface directed *out* from the surface. Note that $\mathbf{E}$ and $\mathbf{H}$ are peak phasors, not rms, leading to 1/2 in power expressions. The integrand inside this integral is defined as the *Poynting vector:*

$$\mathbf{S} = \tfrac{1}{2}\mathbf{E} \times \mathbf{H^*} \tag{2-30}$$

which is a power density with units of W/m^2.

The time-average dissipated power in volume v bounded by closed surface s is

$$P_{d_{av}} = \frac{1}{2} \iiint_v \sigma|E|^2 dv \tag{2-31}$$

The time-average stored magnetic energy is

$$W_{m_{av}} = \frac{1}{2} \iiint_v \frac{1}{2}\mu|H|^2 dv \tag{2-32}$$

The time-average stored electric energy is

$$W_{e_{av}} = \frac{1}{2} \iiint_v \frac{1}{2}\varepsilon|E|^2 dv \tag{2-33}$$

If the source power is not known explicitly, it can be calculated from the volume current density as follows:
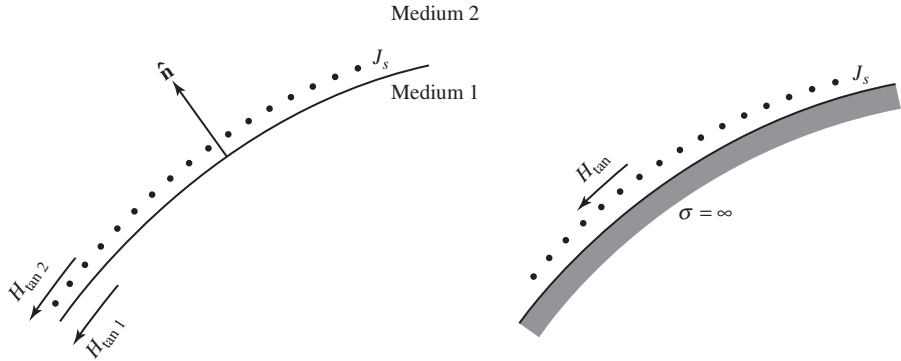
(*a*) General case.   (*b*) One medium a perfect conductor.

Figure 2-1 Magnetic field intensity boundary condition.

$$P_s = -\frac{1}{2} \iiint\limits_{v} \mathbf{E} \cdot \mathbf{J}^* \, dv \tag{2-34}$$

If magnetic current density is present, the term $\mathbf{H}^* \cdot \mathbf{M}$ is added to the integrand in this equation.

From (2-29), we see that the integral of the complex Poynting vector $\frac{1}{2}\mathbf{E} \times \mathbf{H}^*$ over a closed surface s gives the total complex power flowing out through the surface s. It is assumed that the complex Poynting vector represents the complex power density in watts per square meter at a point. Then the complex power through any surface s (not necessarily closed) can be found by integrating the complex Poynting vector over that surface. We are particularly interested in real power, which is the real component of the complex power that represents the electric and magnetic field intensities being in-phase. The real power flowing through surface s is

$$P = \mathrm{Re}\left(\iint\limits_{s} \mathbf{S} \cdot d\mathbf{s} \right) = \frac{1}{2}\mathrm{Re}\left(\iint\limits_{s} \mathbf{E} \times \mathbf{H}^* \cdot d\mathbf{s} \right) \tag{2-35}$$

The reference direction for this average power flow is that of the specified unit normal $\hat{\mathbf{n}}$ contained in $d\mathbf{s} = ds\hat{\mathbf{n}}$.

2.2   SOLUTION OF MAXWELL'S EQUATIONS FOR RADIATION PROBLEMS

This section develops procedures for finding fields radiated by an antenna based on Maxwell's equations. Subsequent antenna analysis in this book begins with these basic relations, and it is usually not necessary to return to Maxwell's equations.

The antenna problem consists of solving for the fields that are created by an impressed current distribution $\mathbf{J}$. In the simplest approach, this current distribution is obtained during the solution process. How to obtain the current distribution will be discussed at various points in the book, but for the moment suppose we have the current distribution and wish to determine the fields $\mathbf{E}$ and $\mathbf{H}$. As mentioned in the previous section, we need only work with the two curl equations of Maxwell's equations as given by (2-16) and (2-17). These are two coupled, linear, first-order differential equations. They are coupled because the unknown functions $\mathbf{E}$ and $\mathbf{H}$ appear in both equations. Thus, these equations

must be solved simultaneously. In order to simplify the solution for **E** and **H** with a given **J**, we introduce the scalar and vector potential functions Φ and **A**.

The vector potential is introduced by noting from (2-19) that the divergence of **H** is zero. Therefore, the vector field **H** has only circulation; for this reason, it is often called a solenoidal field. Because it possesses only a circulation, it can be represented by the curl of some other vector function as follows:

$$\boxed{\mathbf{H} = \frac{1}{\mu}\boldsymbol{\nabla} \times \mathbf{A}} \tag{2-36}$$

where **A** is the *(magnetic) vector potential.* To be more precise, (2-36) is possible because it satisfies (2-19) identically; that is, from (C-9) $\boldsymbol{\nabla} \cdot \boldsymbol{\nabla} \times \mathbf{A} \equiv 0$ for any **A**. The curl of **A** is defined by (2-36), but its divergence needs to be specified for a complete definition of **A**.

The scalar potential is introduced by substituting (2-36) into (2-16), which gives

$$\boldsymbol{\nabla} \times (\mathbf{E} + j\omega\mathbf{A}) = 0 \tag{2-37}$$

The expression in parentheses is an electric field, and because its curl is zero, it is a conservative field and behaves as a static electric field. The *(electric) scalar potential* Φ is defined from

$$\mathbf{E} + j\omega\mathbf{A} = -\boldsymbol{\nabla}\Phi \tag{2-38}$$

because this definition satisfies (2-37) identically; that is, from (C-10) $\boldsymbol{\nabla} \times \boldsymbol{\nabla}\Phi \equiv 0$ for any Φ. Solving (2-38) for the total electric field gives

$$\mathbf{E} = -j\omega\mathbf{A} - \boldsymbol{\nabla}\Phi \tag{2-39}$$

which may be a familiar result.

The field **E** and **H** are now expressed in terms of potential functions by (2-36) and (2-39). If we knew the potential functions, then the fields could be obtained.

The solution for the potential functions begins by substituting (2-36) into (2-17) to give

$$\boldsymbol{\nabla} \times \mathbf{H} = \frac{1}{\mu}\boldsymbol{\nabla} \times \boldsymbol{\nabla} \times \mathbf{A} = j\omega\varepsilon\mathbf{E} + \mathbf{J} \tag{2-40}$$

Using the following vector identity, from (C-17),

$$\boldsymbol{\nabla} \times \boldsymbol{\nabla} \times \mathbf{A} \equiv \boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \mathbf{A}) - \nabla^2\mathbf{A} \tag{2-41}$$

and (2-39) in (2-40) yields

$$\boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \mathbf{A}) - \nabla^2\mathbf{A} = j\omega\mu\varepsilon(-j\omega\mathbf{A} - \boldsymbol{\nabla}\Phi) + \mu\mathbf{J} \tag{2-42}$$

or

$$\nabla^2\mathbf{A} + \omega^2\mu\varepsilon\mathbf{A} - \boldsymbol{\nabla}(j\omega\mu\varepsilon\Phi + \boldsymbol{\nabla} \cdot \mathbf{A}) = -\mu\mathbf{J} \tag{2-43}$$

As we mentioned previously, the divergence of **A** is yet to be specified. A convenient choice would be one that eliminates the third term of (2-43). It is the Lorentz condition (but more properly attributed to L. Lorenz rather than H. Lorentz [2]):

$$\boldsymbol{\nabla} \cdot \mathbf{A} = -j\omega\mu\varepsilon\Phi. \tag{2-44}$$

Then (2-43) reduces to

$$\boxed{\nabla^2\mathbf{A} + \omega^2\mu\varepsilon\mathbf{A} = -\mu\mathbf{J}} \tag{2-45}$$

The choice of (2-44) leads to a decoupling of variables: that is, (2-45) involves **A** and not Φ. This is the *vector wave equation*. It is a differential equation that can be solved for **A** after the impressed current **J** is specified. The fields are then found from (2-36) and from

$$\boxed{\mathbf{E} = -j\omega\mathbf{A} - j\frac{\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{A})}{\omega\mu\varepsilon}} \tag{2-46}$$

where this equation was obtained from (2-39) and (2-44). Notice that only a knowledge of **A** is required. A more cumbersome approach would be to solve the scalar wave equation

$$\boldsymbol{\nabla}^2\Phi + \omega^2\mu\varepsilon\Phi = -\frac{\rho}{\varepsilon} \tag{2-47}$$

in addition to the vector wave equation. It is left as a problem to-derive (2-47). If this approach is used, **E** is found from (2-39). Note that ρ in (2-47) is related to **J** in (2-45) by the continuity equation of (2-20).

The vector wave equation (2-45) is solved by forming three scalar equations. This begins by decomposing **A** into rectangular components using (C-18):

$$\boldsymbol{\nabla}^2\mathbf{A} = \hat{\mathbf{x}}\boldsymbol{\nabla}^2A_x + \hat{\mathbf{y}}\boldsymbol{\nabla}^2A_y + \hat{\mathbf{z}}\boldsymbol{\nabla}^2A_z \tag{2-48}$$

Rectangular components are used because the unit vectors in rectangular components can be factored out of the Laplacian, $\boldsymbol{\nabla}^2(\ )$, since they are not themselves functions of coordinates. This feature is unique to the rectangular coordinate system. Although **A** is always decomposed into rectangular components, the Laplacian of each component of **A** is expressed in a coordinate system appropriate to the geometry of the problem. The solution proceeds by substituting (2-48) into (2-45) and equating rectangular components:

$$\begin{aligned}
\boldsymbol{\nabla}^2A_x + \beta^2 A_x &= -\mu J_x \\
\boldsymbol{\nabla}^2A_y + \beta^2 A_y &= -\mu J_y \\
\boldsymbol{\nabla}^2A_z + \beta^2 A_z &= -\mu J_z
\end{aligned} \tag{2-49}$$

where $\beta^2 = \omega^2\mu\varepsilon$. The real-valued constant

$$\beta = \omega\sqrt{\mu\varepsilon} \tag{2-50}$$

is recognized as the phase constant for a propagating wave. It is a measure of the shift in phase per unit distance, of wave travel. By definition a wave shifts 2π radians of phase for one full cycle, i.e., one wavelength, so

$$\beta = \frac{2\pi}{\lambda} \tag{2-51}$$

The velocity[5] of electromagnetic waves is actually contained in (2-50) as

$$c = \frac{1}{\sqrt{\mu\varepsilon}}$$

---

[5] It is common to use the term velocity in this context and it is understood that the magnitude is intended without direction being indicated; what is meant is speed.

Maxwell found this result, concluding correctly that the velocity is a finite constant and that this result also applies to light because light is an electromagnetic wave. Combining (2-50) and (2-52) gives

$$\beta = \frac{\omega}{c} = \frac{2\pi f}{c} \tag{2-53}$$

and again with (2-51)

$$c = f\lambda \tag{2-54}$$

In free space (and in air), $c = 1/\sqrt{\mu_0 \varepsilon_0} \approx 3 \times 10^8$ m/s where μ_0 and ε_0 are the permeability and permittivity of free space. This relationship is used to compute wavelength from frequency for all electromagnetic waves. Hertz experimentally determined the velocity of light in his laboratory by measuring distance between adjacent nulls of the standing wave (which is a half wavelength) created by reflections off a sheet of metal to find the wavelength value and then using (2-54).

In the time domain, the non-infinite propagation velocity is responsible for the delay in arrival time of waves traveling distance Δr, the so-called *retarded time* given by $\Delta t = \Delta r/c$. The corresponding phase change is found from the fraction of the period T $(=1/f)$ of the time delay:

$\Delta phase = \beta \Delta r = \frac{2\pi f}{c}(c\Delta t) = 2\pi \frac{\Delta t}{T}$. For example, a propagation distance of $\Delta r = \lambda/4$ corresponds to a retarded time of $\Delta t = \frac{\Delta r}{c} = \frac{\lambda}{4}\frac{1}{f\lambda} = \frac{T}{4}$ and a phase increment of $\Delta phase = 2\pi \frac{T/4}{T} = \frac{\pi}{2} = 90°$.

The three equations in (2-49) are identical in form. After solving one of these equations, the other two are easily solved. We first find the solution for a *point source*. This unit impulse response solution can then be used to form a general solution by viewing an arbitrary source as a collection of point sources. The differential equation for a point source is

$$\nabla^2 \psi + \beta^2 \psi = -\delta(x)\,\delta(y)\,\delta(z) \tag{2-55}$$

where ψ is the response to a point source at the origin, and $\delta(\ )$ is the unit impulse function, or Dirac delta function (see Appendix F.l). In spite of the fact that the point source is of infinitesimal extent, its associated current has a direction. This is because in solving practical problems, the point source represents a small subdivision of current that does have a direction. If the point source current is taken as z-directed, then

$$\psi = A_z \tag{2-56}$$

Because the point source is zero everywhere except at the origin, (2-55) becomes

$$\nabla^2 \psi + \beta^2 \psi = 0 \tag{2-57}$$

away from the origin.

This is the complex *scalar wave equation* or *Helmholtz equation.* Because of spherical symmetry, the Laplacian is written in spherical coordinates and ψ has only radial dependence. The two solutions to (2-57) are $e^{-j\beta r}/r$ and $e^{+j\beta r}/r$. These correspond to waves propagating radially outward and inward, respectively. The physically meaningful solution is the one for waves traveling away from the point source. Evaluating the constant of proportionality (see Prob. 2.2-3), we have for the point source solution:

$$\psi = \frac{e^{-j\beta r}}{4\pi r} \tag{2-58}$$

This is the solution to (2-55) and is the magnitude and phase variation with distance r away from a point source located at the origin. If the source were positioned at an arbitrary location, we must compute the distance R between the source location and observation point P (see Fig. 2-2). Then

$$\psi = \frac{e^{-j\beta R}}{4\pi R} \tag{2-59}$$

The point source serves as a starting point for the ideal dipole antenna solution, which is discussed in the next section.

For an arbitrary z-directed current density, the vector potential is also z-directed. If we consider the source to be a collection of point sources weighted by the distribution J_z, the response A_z is a sum of the point source responses of (2-59). This is expressed by the integral over the source volume v':

$$A_z = \iiint\limits_{v'} \mu J_z \frac{e^{-j\beta R}}{4\pi R}\, dv' \tag{2-60}$$

Similar equations hold for the x- and y-components. The total solution is then the sum of all components, which is

$$\mathbf{A} = \iiint\limits_{v'} \mu \mathbf{J} \frac{e^{-j\beta R}}{4\pi R}\, dv' \tag{2-61}$$

This is the solution to the vector wave equation (2-45). The geometry is shown in Fig. 2-2. The coordinate system shown is used to describe both the source point and field point. $\mathbf{r}'$ is the vector from the coordinate origin to the source point, and $\mathbf{r}_p$ is the vector from the coordinate origin to the field point P. The vector $\mathbf{R}$ is the vector from the source point to the field point and is given *by* $\mathbf{r}_p - \mathbf{r}'$. This notation is standard and will be used here.

We can summarize rather simply the procedure for finding the fields generated by a current distribution $\mathbf{J}$. First, $\mathbf{A}$ is found from (2-61). The $\mathbf{H}$ field is found from (2-36). The $\mathbf{E}$ field can be found from (2-46), but frequently it is simpler to find $\mathbf{E}$ from (2-17) as

$$\mathbf{E} = \frac{1}{j\omega\varepsilon}(\nabla \times \mathbf{H} - \mathbf{J}) \tag{2-62}$$

in the source region, or from

$$\mathbf{E} = \frac{1}{j\omega\varepsilon} \nabla \times \mathbf{H} \tag{2-63}$$

if the field point is removed in distance from the source; that is, if $\mathbf{J} = 0$ at point P.
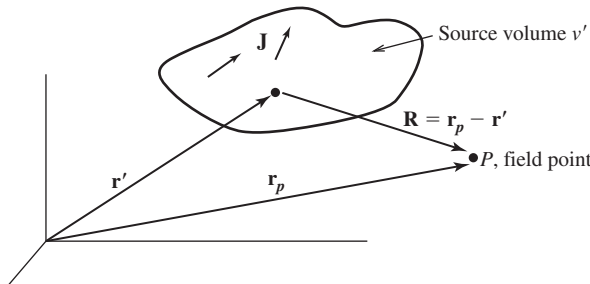
Figure 2-2 Vectors used to solve radiation problems.

2.3   THE IDEAL DIPOLE

The principles presented in the previous section are used in this section to find the fields of an infinitesimal element of current. We shall use the term **ideal dipole** for a uniform amplitude current that is electrically small with $\Delta z \ll \lambda$. It is ideal in the sense that the current is uniform in both magnitude and phase over the radiating element extent. Such a discontinuous current is difficult to realize in practice; practical realizations that approximate the ideal dipole are presented in Sec. 3.1. The term *current element* is often used for the ideal dipole to describe its application as a section of a larger current associated with an actual antenna. Thus, any practical antenna can first be decomposed into filaments of continuous current that are then subdivided into ideal dipoles. The fields from the antenna are then found by summing contributions from the ideal dipoles. Other terms used for the ideal dipole are *Hertzian electric dipole, electric dipole, infinitesimal dipole*, and *doublet*. An electrically small, center-fed wire antenna has a current distribution that tapers to zero from the center to the ends of the wire. This *short dipole* antenna has the same pattern as an ideal dipole and is discussed further in Secs. 2.7 and 3.1.

Consider an element of current of length Δz along the z-axis centered on the coordinate origin. It is of constant amplitude I. In this case, the volume integral of (2-61) for vector potential reduces to the one-dimensional integral[6]

$$\mathbf{A} = \hat{\mathbf{z}}\mu I \int_{-\Delta z/2}^{\Delta z/2} \frac{e^{-j\beta R}}{4\pi R}\, dz' \tag{2-64}$$

The length Δz is very small compared to a wavelength and to the distance R; see Fig. 2-3. Because Δz is very small, the distance R from points on the current element to the field point P approximately equals the distance r from the origin to the field point. Substituting r for R in (2-64) and integrating gives
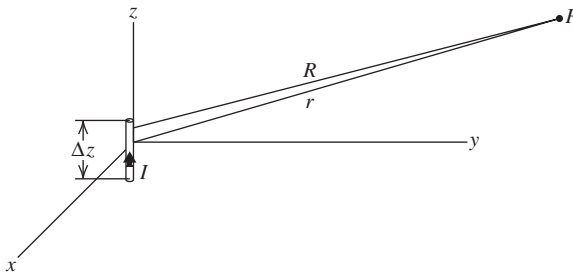
Figure 2-3 The ideal dipole. The current I is uniform, $\Delta z \ll$ l, and $R \approx r$.

---

[6] The result in (2-64) could also be obtained by representing the current density on the dipole as

$$\mathbf{J} = I\,\delta(x')\,\delta(y')\hat{\mathbf{z}} \quad \text{for} \quad -\frac{\Delta z}{2} < z' < \frac{\Delta z}{2}$$

Substituting this into (2-61) yields

$$\mathbf{A} = \hat{\mathbf{z}}\mu I \int_{-\infty}^{\infty}\delta(x')dx' \int_{-\infty}^{\infty}\delta(y')\,dy' \int_{-\Delta z/2}^{\Delta z/2}\frac{e^{-j\beta R}}{4\pi R}\,dz'$$

from which (2-64) follows.

$$\mathbf{A} = \frac{\mu I e^{-j\beta r}}{4\pi r} \Delta z \,\hat{\mathbf{z}} \tag{2-65}$$

This is exactly true for a point current element and is approximately true for a small ($\Delta z \ll \lambda$ and $\Delta z \ll R$) but finite uniform current element. The vector potential A_z for a point source was also derived in the previous section; see (2-58) in which $I\Delta z = 1$. For many current sources, we can readily make the substitution of r for R in the denominator of the integrand in (2-64), but usually cannot make the same substitution in the exponent. However, in the case of a very small source, we can use r for R in both the denominator and exponent.

We are now ready to calculate the electromagnetic fields created by the ideal dipole. The magnetic field is found from (2-36) as

$$\mathbf{H} = \frac{1}{\mu} \nabla \times \mathbf{A} = \frac{1}{\mu} \nabla \times (A_z \hat{\mathbf{z}}) \tag{2-66}$$

If we apply the vector identity (C-16), the preceding equation becomes

$$\mathbf{H} = \frac{1}{\mu} (\nabla A_z) \times \hat{\mathbf{z}} + \frac{1}{\mu} A_z (\nabla \times \hat{\mathbf{z}}) = \frac{1}{\mu} (\nabla A_z) \times \hat{\mathbf{z}} \tag{2-67}$$

The second term is zero because the curl of a constant vector is zero. Substituting (2-65) into (2-67), we have

$$\mathbf{H} = \nabla \left(\frac{I \,\Delta z e^{-j\beta r}}{4\pi r} \right) \times \hat{\mathbf{z}} \tag{2-68}$$

Applying the gradient in spherical coordinates from (C-33) gives

$$\begin{aligned}
\mathbf{H} &= \frac{I\Delta z}{4\pi} \frac{\partial}{\partial r} \left(\frac{e^{-j\beta r}}{r} \right) \hat{\mathbf{r}} \times \hat{\mathbf{z}} \\
&= \frac{I\Delta z}{4\pi} \left[\frac{-j\beta e^{-j\beta r}}{r} - \frac{e^{-j\beta r}}{r^2} \right] \hat{\mathbf{r}} \times \hat{\mathbf{z}}
\end{aligned} \tag{2-69}$$

From (C-3), we have

$$\hat{\mathbf{r}} \times \hat{\mathbf{z}} = \hat{\mathbf{r}} \times (\hat{\mathbf{r}} \cos\theta - \hat{\boldsymbol{\theta}} \sin\theta) = -\hat{\boldsymbol{\phi}} \sin\theta \tag{2-70}$$

Substituting (2-70) into (2-69) gives

$$\mathbf{H} = \frac{I\,\Delta z}{4\pi} \left[\frac{j\beta}{r} + \frac{1}{r^2} \right] e^{-j\beta r} \sin\theta \,\hat{\boldsymbol{\phi}} \tag{2-71}$$

The electric field can be obtained from (2-63) as

$$\begin{aligned}
\mathbf{E} &= \frac{I\,\Delta z}{4\pi} \left[\frac{j\omega\mu}{r} + \sqrt{\frac{\mu}{\varepsilon}} \frac{1}{r^2} + \frac{1}{j\omega\varepsilon r^3} \right] e^{-j\beta r} \sin\theta \,\hat{\boldsymbol{\theta}} \\
&\quad + \frac{I\,\Delta z}{2\pi} \left[\sqrt{\frac{\mu}{\varepsilon}} \frac{1}{r^2} + \frac{1}{j\omega\varepsilon} \frac{1}{r^3} \right] e^{-j\beta r} \cos\theta \,\hat{\mathbf{r}}
\end{aligned} \tag{2-72}$$

Eqs. (2-71) and (2-72) can be written as

$$\mathbf{H} = \frac{I \, \Delta z}{4\pi} j\beta \left(1 + \frac{1}{j\beta r} \right) \frac{e^{-j\beta r}}{r} \sin \theta \, \hat{\boldsymbol{\phi}} \tag{2-73a}$$

$$\mathbf{E} = \frac{I \, \Delta z}{4\pi} j\omega\mu \left[1 + \frac{1}{j\beta r} - \frac{1}{(\beta r)^2} \right] \frac{e^{-j\beta r}}{r} \sin \theta \, \hat{\boldsymbol{\theta}}$$

$$+ \frac{I \, \Delta z}{2\pi} \eta \left[\frac{1}{r} - j\frac{1}{\beta r^2} \right] \frac{e^{-j\beta r}}{r} \cos \theta \, \hat{\mathbf{r}} \tag{2-73b}$$

Even the simplest of antennas, the ideal dipole, has a complicated field structure at distances close to the antenna where $1/r^2$ and $1/r^3$ terms are not negligible. The electric field expression of (2-73b) has transverse (θ) and radial (r) components. The associated power density near the ideal dipole has a real-valued (dissipative) outward-directed radial (r) component and an imaginary-valued (reactive) transverse (θ) component; see Prob. 2.3-3 and (2-79) below. If βr is large (i.e., $\beta r \gg 1$, or $r \gg \lambda$ because $\beta = 2\pi/\lambda$), then all terms having inverse powers of $j\beta r$ are small compared to unity, and (2-73) reduces to:

$$\mathbf{E} = \frac{I\Delta z}{4\pi} j\omega\mu \frac{e^{-j\beta r}}{r} \sin \theta \, \hat{\boldsymbol{\theta}} \tag{2.74a}$$

$$\mathbf{H} = \frac{I\Delta z}{4\pi} j\beta \frac{e^{-j\beta r}}{r} \sin \theta \, \hat{\boldsymbol{\phi}} \tag{2.74b}$$

These are the fields at large distances from the dipole that are called the *radiation fields of the ideal dipole*. The ratio of these electric and magnetic field components is

$$\frac{E_\theta}{H_\phi} = \frac{\omega\mu}{\beta} = \frac{\omega\mu}{\omega\sqrt{\mu\varepsilon}} = \sqrt{\frac{\mu}{\varepsilon}} = \eta \tag{2-75}$$

where $\eta = \sqrt{\mu/\varepsilon}$ is the *intrinsic impedance* of the medium (for free space $\eta_0 = 376.7 \, \Omega \approx 120\pi \, \Omega$). This is a property of plane waves. Also, as we shall see, at large distances from any antenna the fields are related in this manner.

Using the fields of (2-74) in (2-30) gives an expression for the complex power density flowing out of a sphere of radius r surrounding the ideal dipole:

$$\mathbf{S} = \frac{1}{2} \mathbf{E} \times \mathbf{H}^*$$

$$= \frac{1}{2} \left(\frac{I\Delta z}{4\pi} \right)^2 j\omega\mu \frac{e^{-j\beta r}}{r} \sin \theta \, \hat{\boldsymbol{\theta}} \times (-j\beta) \frac{e^{+j\beta r}}{r} \sin \theta \, \hat{\boldsymbol{\phi}} \tag{2-76}$$

$$= \frac{1}{2} \left(\frac{I\Delta z}{4\pi} \right)^2 \omega\mu\beta \frac{\sin^2 \theta}{r^2} \hat{\mathbf{r}}$$

which is real-valued and radially directed, both characteristics of radiation. The total power flowing out through a sphere of radius r surrounding the ideal dipole using (2-29) is

$$P_f = \iint \mathbf{S} \cdot d\mathbf{s} = \frac{1}{2} \left(\frac{I\Delta z}{4\pi} \right)^2 \omega\mu\beta \int_0^{2\pi} d\phi \int_0^{\pi} \sin^3\theta \, d\theta$$

$$= \frac{1}{2} \left(\frac{I\Delta z}{4\pi} \right)^2 \omega\mu\beta 2\pi \frac{4}{3} \qquad (2\text{-}77)$$

$$= \frac{\omega\mu\beta}{12\pi} (I\Delta z)^2$$

This is a real quantity, and real power indicates dissipated power. It is dissipated in the sense that it travels away from the source, never to return. In fact, the average power going out through a sphere of radius r can be found as indicated in (2-35) by taking the real part of (2-77), which leaves it unchanged. This power expression is independent of r, and thus if we integrate over a sphere of larger radius, we still have the same total power streaming through it. We refer to this type of power as *radiated power*.

The general field expressions of (2-73) are valid at any distance from an ideal dipole and are important in some applications and for understanding the input impedance of a dipole. For distances so close to the dipole that $\beta r \ll 1$, or $r \ll \lambda$, only the dominant terms with the largest inverse powers of r need be retained in each component of (2-73):

$$\mathbf{H}^{\text{nf}} = \frac{I\Delta z e^{-j\beta r}}{4\pi r^2} \sin\theta \, \hat{\boldsymbol{\phi}} \qquad (2\text{-}78a)$$

$$\mathbf{E}^{\text{nf}} = -j\eta \frac{I\Delta z}{4\pi\beta} \frac{e^{-j\beta r}}{r^3} \sin\theta \, \hat{\boldsymbol{\theta}} - j\eta \frac{I\Delta z}{2\pi\beta} \frac{e^{-j\beta r}}{r^3} \cos\theta \, \hat{\mathbf{r}} \qquad (2\text{-}78b)$$

These are referred to as the *near fields* of the antenna. Actually, the magnetic field of (2-78a) which varies as $1/r^2$ is that of a short, steady or slowly oscillating current, that is, an induction field. The electric fields of (2-78b) vary as $1/r^3$ and are those of an electrostatic or quasi-static dipole with charges of $+q$ and $-q$ spaced Δz apart. Note that the electric field components E_θ^{nf} and E_r^{nf} are in-phase, but are in phase-quadrature with the magnetic field H_ϕ^{nf}, indicating reactive power. This can be shown directly using these near fields in the complex Poynting vector expression of (2-30):

$$\mathbf{S}^{nf} = \frac{1}{2} [E_\theta^{nf} H_\phi^{nf*} \hat{\mathbf{r}} - E_r^{nf} H_\phi^{nf*} \hat{\boldsymbol{\theta}}]$$

$$= -\frac{j\eta}{2\beta} \left(\frac{I\Delta z}{4\pi} \right)^2 \frac{1}{r^5} (\sin^2\theta \, \hat{\mathbf{r}} - \sin 2\theta \, \hat{\boldsymbol{\theta}}) \qquad (2\text{-}79)$$

Note that this power density vector is imaginary and therefore has no time-average radial power flow. The radiation fields, in contrast, are in-phase giving a real-valued Poynting vector that is radially directed; see (2-74) and (2-76). The imaginary power density corresponds to standing waves, rather than traveling waves associated with radiation, and indicate stored energy as in a reactive device. The quadrature phase relationship between the electric and magnetic field components of (2-78) indicates that energy is interchanged between these fields with time. That is, at one instant of time the electric fields are strong near the dipole close to maximum charge regions, and a quarter-period later energy is stored in the magnetic field, primarily close to the center of the dipole where the current is maximum.

The imaginary power density in the near field is manifested by a reactive component of the antenna input impedance. The real part of the input impedance represents radiation if

ohmic losses on the antenna structure can be neglected. Antenna impedance will be discussed further in Sec. 2.6. The power density associated with radiation exists everywhere and passes through the near field. The radiated power density from (2-76) and the near-field power density from (2-79) are both maximum for $\theta = 90°$. The distance for which the maximum radiated and reactive powers are equal for the ideal dipole is $r = \lambda/2\pi$; see Prob. 2.3-5. That is, interior to this radius reactive power dominates. This region is sometimes referred to as the *radiansphere.*

At large distances ($r \gg \lambda/2\pi$) from an antenna, called the *far-field region*, all power is radiated power. The far field is further characterized by the fact that the angular distribution around the antenna (i.e., the radiation pattern) is independent of distance from the antenna. Field regions and the distance away from an antenna where the far field begins are discussed further in Sec. 2.4.3

2.4   RADIATION PATTERNS

We briefly introduced the concept of a radiation pattern in Sec. 1.4 as a description of the angular variation of radiation level around an antenna. The purpose of an antenna is to transmit, or receive, signals preferentially in various directions. An antenna's radiation pattern is perhaps its most important characteristic. In this section, we present several definitions associated with patterns and develop the procedural steps for calculating radiation patterns.

2.4.1   Radiation Pattern Basics

A *radiation pattern* (also called an *antenna pattern)* is a graphical representation of the radiation (far-field) properties of an antenna; a general example radiation pattern is shown in Fig. 1-7. We saw from the analysis in the previous section that the radiation fields from a transmitting ideal dipole antenna vary inversely with distance, e.g., $1/r$. This is a general property for antennas of all types; that is, the radiated fields observed far from any antenna decay with distance as a spherical wave. The variation with observation angles (θ, ϕ), however, depends on the size and construction details of the antenna. In fact, the relationship of an antenna to its radiation pattern forms a large part of antenna investigations.

A helpful way to understand the concept of radiation pattern is to visualize a spherical ball of bread dough. The ball represents the hypothetical isotropic radiation pattern (see Fig. 1-7) that radiates equally in all directions. Squeezing on the dough ball distorts the shape, representing more radiation in some directions and less in others. The total power radiated in all directions is same, but the power density (in W/m^2) at points on a distant spherical surface varies with direction in proportion to the square of the pattern (square is used because the pattern is a field quantity, unless otherwise noted).

Radiation patterns are determined from the calculated or measured fields radiated from the antenna. We illustrate with the ideal dipole shown in Fig. 2-4. The length and orientation of the field vectors in Fig. 2-4a follow from (2-74) and are shown for an instant of time for which the fields are peak. An electric field probe antenna moved over the sphere surface and oriented parallel to E_θ will have an output proportional to $\sin \theta$; see Fig. 13-4. The relative variation of both E_θ and H_ϕ over the sphere is $\sin \theta$, as displayed in Fig. 2-4b. This radiation pattern graphically represents unit-less, normalized values over the surface of a sphere of radius r which is in the far field. Any plane containing the z-axis has the same radiation pattern because there is no ϕ-variation in the fields, which is noted by the constant variation in the xy-plane shown in Fig. 2-4c. A pattern taken in one of the planes through the z-axis is called an *E-plane pattern* because it contains the electric vector. A pattern taken in a plane perpendicular to an *E*-plane and cutting across the test antenna (the xy-plane in this case) is called an *H-plane pattern* because it contains the
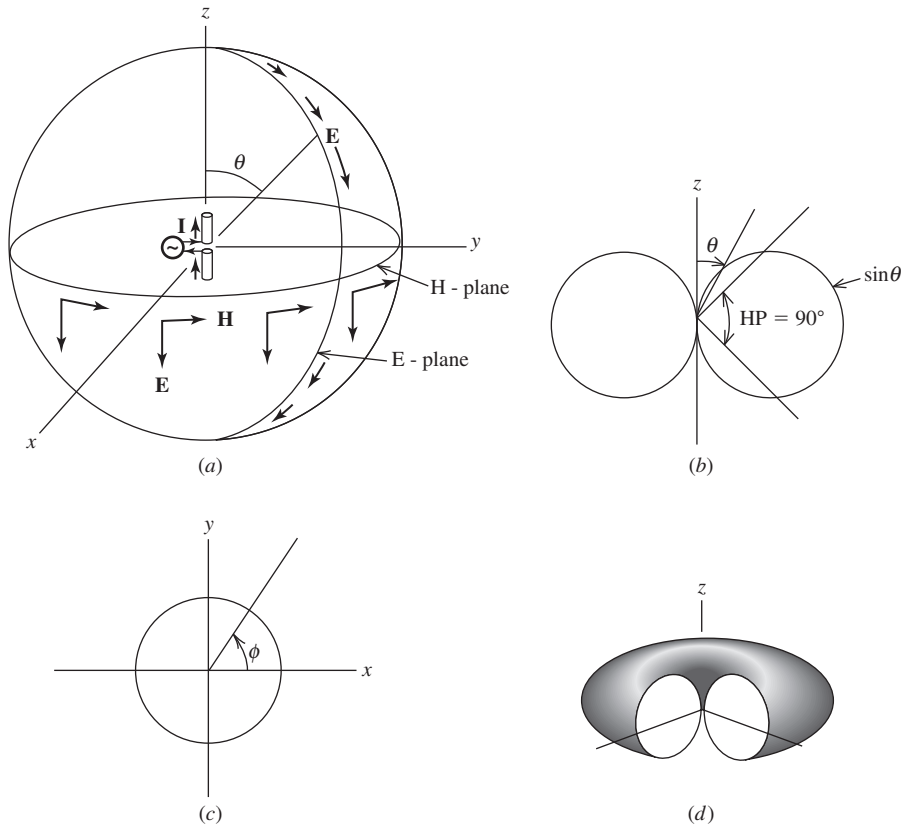
Figure 2-4 Radiation from an ideal dipole, (*a*) Field components. (*b*) *E*-plane radiation pattern polar plot of $|E_\theta|$ or $|H_\phi|$. (*c*) *H*-plane radiation pattern polar plot of $|E_\theta|$ or $|H_\phi|$. (*d*) Three-dimensional plot of radiation pattern.

magnetic field H_ϕ. The *E*- and *H*-plane patterns, in general for linearly polarized antenna, are referred to as *principal plane patterns*. The *E*- and *H*-plane patterns for the ideal dipole shown in Fig. 2-4*b* and *c*, respectively, are polar plots in which the distance from the origin to the curve is proportional to the field intensity; they are often referred to as a *polar pattern*. For patterns that have a single unidirectional beam, the principal planes are taken to be two orthogonal planes passing through the beam peak.

The complete 3D pattern for the ideal dipole is shown in isometric view with a slice removed in Fig. 2-4*d*. This solid radiation pattern resembles a "doughnut." It is referred to as an *omnidirectional pattern* because it is uniform in one plane, the *xy*-plane. Omnidirectional antennas are very popular in ground-based applications where the omnidirectional-pattern plane is horizontal (parallel to the ground). When encountering new antennas the reader should attempt to visualize the complete pattern in three dimensions.

2.4.2   Radiation from Line Currents

Radiation patterns in general can be calculated in a manner similar to that used for the ideal dipole if the current distribution on the antenna is known. This is done by first finding the vector potential given in (2-61). As a simple example, consider a filament of current along the *z*-axis and located near the origin. Many antennas can be modeled by
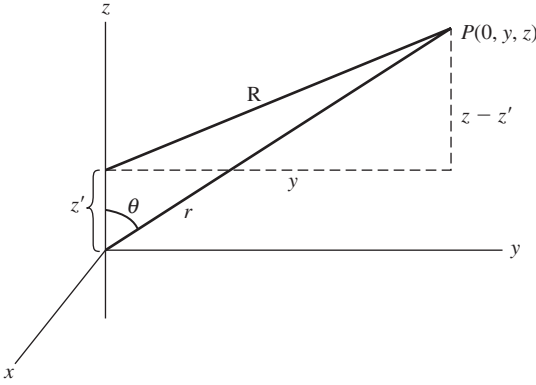
Figure 2-5 Geometry used for field calculations of a line source along the z-axis.

this **line source**; straight wire antennas are good examples. In this case, the vector potential has only a z-component and the vector potential integral is one-dimensional[7]:

$$A_z = \mu \int I(z') \frac{e^{-j\beta R}}{4\pi R} dz' \tag{2-80}$$

Due to the symmetry of the source, we expect that the radiation fields will not vary with ϕ. This is because as an observer moves around the source such that r and z are constant, the appearance of the source remains the same; thus, its radiation fields are also unchanged. Therefore, for simplicity we will confine the observation point to a fixed ϕ in the yz-plane ($\phi = 90°$) as shown in Fig. 2-5. From Fig. 2-5 we see that

$$r^2 = y^2 + z^2 \tag{2-81}$$
$$z = r\cos\theta \tag{2-82}$$
$$y = r\sin\theta \tag{2-83}$$

If we apply the general geometry of Fig. 2-2 to this case, $\mathbf{r}_p = \mathbf{r} = y\hat{\mathbf{y}} + z\hat{\mathbf{z}}$ and $\mathbf{r}' = z'\hat{\mathbf{z}}$ lead to $\mathbf{R} = \mathbf{r}_p - \mathbf{r}' = y\hat{\mathbf{y}} + (z - z')\hat{\mathbf{z}}$, and then

$$R = \sqrt{y^2 + (z - z')^2} = \sqrt{y^2 + z^2 - 2zz' + (z')^2} \tag{2-84}$$

Substituting (2-81) and (2-82) into (2-83), to put all field point coordinates into the spherical coordinate system, gives

$$R = \{r^2 + [-2r\cos\theta\, z' + (z')^2]\}^{1/2} \tag{2-85}$$

In order to develop approximate expressions for R, we expand (2-85) using the binomial theorem (F-4):

$$R = (r^2)^{1/2} + \frac{1}{2}(r^2)^{-1/2}[-2r\cos\theta\, z' + (z')^2] + \frac{\frac{1}{2}\left(-\frac{1}{2}\right)}{2}(r^2)^{-3/2}$$

$$\cdot [-2r\cos\theta\, z' + (z')^2]^2 + \cdots$$

$$= r - z'\cos\theta + \frac{(z')^2 \sin^2\theta}{2r} + \frac{(z')^3 \sin^2\theta\cos\theta}{2r^2} + \cdots \tag{2-86}$$

---

[7] This result could also be obtained by using $J_z(r') = I(z')\,\delta(x')\,\delta(y')$ in (2-60) where $dv' = dx'\, dy'\, dz'$.

Successive terms in this series decrease as the power of z' increases if z' is small compared to r. This expression for R is used in the radiation integral (2-80) to different degrees of approximation. In the denominator (which affects only the amplitude), we let

$$R \approx r \tag{2-87}$$

We can do this because in the far field r is very large compared to the antenna size, so $r \gg z' \geq z' \cos \theta$. In the phase term $-\beta R$, we must be more accurate when computing the distance from points along the line source to the observation point.

The integral (2-80) sums the contributions from all the points along the line source. Although the amplitude of waves due to each source point is essentially the same, the phase can be different if the path length differences are a sizable fraction of a wavelength. We, therefore, include the first two terms of the series in (2-86) for the R in the numerator of (2-80), giving

$$R \approx r - z' \cos \theta \tag{2-88}$$

Using the far-field approximations (2-87) and (2-88) in (2-80) yields

$$A_z = \mu \int I(z') \frac{e^{-j\beta(r - z'\cos\theta)}}{4\pi r} dz' = \mu \frac{e^{-j\beta r}}{4\pi r} \int I(z') e^{j\beta z'\cos\theta} dz' \tag{2-89}$$

where the integral is over the extent of the line source. This may be recognized as a Fourier-transform-type integral; see Sec. 5.3. [H.3: Silver, Sec. 6.3] The Fourier transform relationship between the antenna current distribution and its radiation pattern will be revisited for linear antennas in Sec. 5.3 and for aperture antennas in Sec. 9.1. The integral is referred to as the *radiation integral* or *diffraction integral*.

The magnetic field for the line source is found using (2-36):

$$\mathbf{H} = \frac{1}{\mu} \nabla \times \mathbf{A} = \frac{1}{\mu} \nabla \times (A_z \hat{\mathbf{z}})$$

$$= \frac{1}{\mu} \nabla \times (-A_z \sin\theta \, \hat{\boldsymbol{\theta}} + A_z \cos\theta \, \hat{\mathbf{r}}) \tag{2-90}$$

where (C-3) was used. Since A_z is a function of r and θ, the curl in spherical coordinates, as given by (C-35), leads to

$$\mathbf{H} = \hat{\boldsymbol{\phi}} \frac{1}{\mu} \frac{1}{r} \left[\frac{\partial}{\partial t} (-r A_z \sin\theta) - \frac{\partial}{\partial \theta} (A_z \cos\theta) \right] \tag{2-91}$$

Substitution of (2-89) into the above gives

$$\begin{aligned}
\mathbf{H} &= \hat{\boldsymbol{\phi}} \frac{1}{\mu} \left\{ \mu \frac{-\sin\theta}{4\pi r} \int I(z') e^{j\beta z'\cos\theta} dz' \frac{\partial}{\partial r} e^{-j\beta r} - \frac{e^{-j\beta r}}{4\pi r^2} \frac{\partial}{\partial \theta} \right. \\
&\quad \left. \cdot \left[\mu \cos\theta \int I(z') e^{j\beta z'\cos\theta} dz' \right] \right\} \\
&= \hat{\boldsymbol{\phi}} \frac{1}{\mu} \frac{e^{-j\beta r}}{4\pi r} \left\{ j\beta \mu \sin\theta \int I(z') e^{j\beta z'\cos\theta} dz' \right. \\
&\quad \left. - \frac{1}{r} \frac{\partial}{\partial \theta} \left[\mu \cos\theta \int I(z') e^{j\beta z'\cos\theta} dz' \right] \right\}
\end{aligned} \tag{2-92}$$

The ratio of the first term to the second term is of the order βr. If $\beta r \gg 1$, the second term is small compared to the first and can be neglected, as we did for the far-field approximation of the ideal dipole. Thus, (2-92) becomes

$$\mathbf{H} = \hat{\boldsymbol{\phi}}\frac{j\beta}{\mu}\sin\theta\,\mu\frac{e^{-j\beta r}}{4\pi r}\int I(z')e^{j\beta z'\cos\theta}dz' = \frac{j\beta}{\mu}\sin\theta\,A_z\hat{\boldsymbol{\phi}} \tag{2-93}$$

The electric field is found from (2-46), which is

$$\mathbf{E} = -j\omega\mathbf{A} - j\frac{\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{A})}{\omega\mu\varepsilon} \tag{2-94}$$

Using (2-80) in (2-81) and retaining only the r^{-1} term (and assuming $\beta r \gg 1$) lead to the far-field approximation

$$\mathbf{E} = -j\omega A_\theta\hat{\boldsymbol{\theta}} = j\omega\sin\theta\,A_z\hat{\boldsymbol{\theta}} \tag{2-95}$$

Note that this result is the portion of the first term of (2-94) that is transverse to $\hat{\mathbf{r}}$ because $-j\omega\mathbf{A} = -j\omega(-A_z\sin\theta\,\hat{\boldsymbol{\theta}} + A_z\cos\theta\,\hat{\mathbf{r}})$. This is an important general result that is not restricted to line sources.

The radiation fields from a z-directed line source (any z-directed current source in general) are H_ϕ and E_θ, and are found from (2-93) and (2-95). The only remaining problem is to calculate A_z, which is given by (2-60) in general and by (2-89) for z-directed line sources. The calculation of A_z is the focal point of antenna analysis. We will return to this topic after pausing to further examine the characteristics of the far-field region.

The ratio of the radiation field components in (2-93) and (2-95) yields

$$E_\theta = \frac{\omega\mu}{\beta}H_\phi = \frac{\omega\mu}{\omega\sqrt{\mu\varepsilon}}H_\phi = \eta H_\phi \tag{2-96}$$

where $\eta = \sqrt{\mu/\varepsilon}$ is the intrinsic impedance of the medium. Thus, the radiation fields are perpendicular to each other and to the direction of propagation $\hat{\mathbf{r}}$ and their magnitudes are related by (2-96). These are the familiar properties of a plane wave. They also hold for the general form of a "transverse electromagnetic (TEM) wave" that has both the electric and magnetic fields transverse to the direction of propagation. In general, radiation from a finite antenna is a special case of a TEM wave, called a "spherical wave," that propagates radially outward from the antenna and the radiation fields have no radial components. Spherical wave behavior is also characterized by the $e^{-j\beta r}/4\pi r$ factor in the field expressions; see (2-93). The $e^{-j\beta r}$ phase factor indicates a traveling wave propagating radially outward from the origin and the $1/r$ magnitude dependence leads to constant power flow just as with the infinitesimal dipole. In fact, the radiation fields of all antennas of finite extent display this dependence with distance from the antenna.

Another way to view radiation field behavior is to note that spherical waves appear to an observer in the far field to be a plane wave. This "local plane wave behavior" occurs because the radius of curvature of the spherical wave is so large that the phase front is nearly planar over a local region.

2.4.3 Far-Field Conditions and Field Regions

The results for the line current from the previous section are easily generalized to an arbitrary, but finite-sized, antenna. In the far field of an antenna, the fields exhibit local plane wave behavior and have $1/r$ magnitude dependence. In this section, we develop the conditions for determining the minimum distance from an antenna for far-field behavior. This begins with a geometric interpretation for far-field approximations.

If parallel lines (or rays) are drawn from each point on a line current as shown in Fig. 2-6, the distance R to the far field is geometrically related to r by (2-88), which was derived by neglecting high-order terms in the expression for R in (2-86). The parallel ray
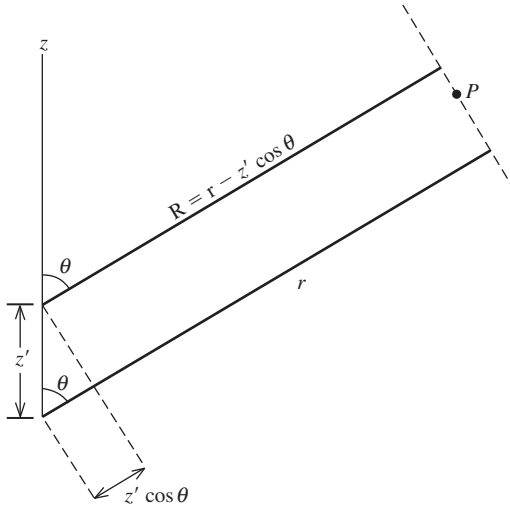
Figure 2-6 Parallel ray approximation for far-field calculations of a line source.

assumption is exact only when the observation point is at infinity, but it is a good approximation in the far field. Radiation calculations often start by assuming parallel rays and then determining R for the phase by geometrical techniques. From the general source shown in Fig. 2-7, we see that

$$R = r - r' \cos \alpha \tag{2-97}$$

Using the definition of dot product, we have

$$R = r - r' \frac{\mathbf{r} \cdot \mathbf{r}'}{rr'}$$

or
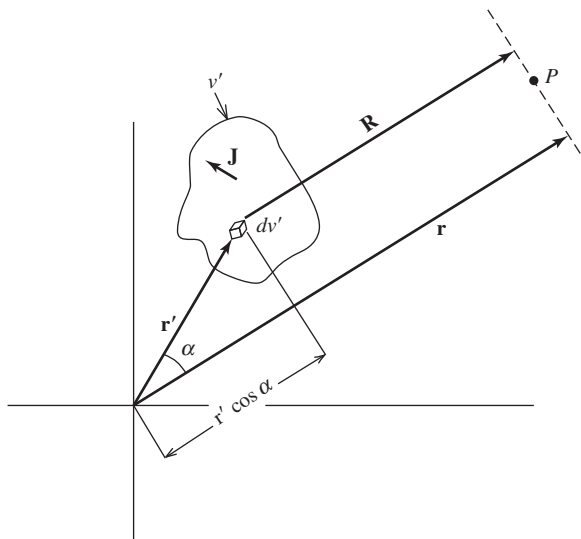
Figure 2-7 Parallel ray approximation for far-field calculations of a general source.

$$R = r - \hat{\mathbf{r}} \cdot \mathbf{r}'$$

$$(2\text{-}98)$$

This is a general approximation to R for the phase factor in the radiation integral. Notice that if $\mathbf{r}' = z' \, \hat{\mathbf{z}}$, as for line sources along the z-axis, (2-98) reduces to (2-86).

The definition of the distance from the source where the far field begins is where errors due to the parallel ray approximation become insignificant. The distance where the far field begins, r_{ff}, is taken to be that value of r for which the path length deviation due to neglecting the third term of (2-86) is a sixteenth of a wavelength. This corresponds to a phase error of $2\pi/\lambda \cdot \lambda/16 = \pi/8 \text{ rad} = 22.5°$.

If D is the length of the line source, r_{ff} is found by equating the maximum value of the third term of (2-86), which occurs for $z' = D/2$ and $\theta = 90°$, to a sixteenth of a wavelength, or $(D/2)^2/2r_{ff} = \lambda/16$. Solving for r_{ff} gives

$$r_{ff} = \frac{2D^2}{\lambda}$$

$$(2\text{-}99)$$

The far-field region is $r \geq r_{ff}$ and r_{ff} is called the *far-field distance*, or Rayleigh distance. The general far field condition of (2-99) is not sufficiently large if the antenna size is small compared to a wavelength. Additional considerations are necessary and will be developed after first discussing the field regions in more detail.

The region interior to the far field, called the *near field*, is divided into the *reactive near field* where the reactive fields dominate over the radiation (real-valued) fields and the *radiating near field* where the radiation fields dominate over the reactive fields. For the ideal dipole, the boundary between the reactive and radiating near field regions occurs for $r = \lambda/2\pi$ which is the radius of the radiansphere introduced at the end of Sec. 2.3. This result follows from the general field expressions for the ideal dipole in (2-73). There are terms with radial dependencies of $1/r$, $1/r^2$, and $1/r^3$. With increasing distance from the dipole, first the $1/r^3$ term diminishes compared to other terms, then the $1/r^2$ term diminishes, and only the $1/r$ term is significant in the far field. The real and imaginary parts of the vector components are equal for $\beta r = 1$, or $r = \lambda/2\pi = 0.16\lambda$, defining the reactive-radiating near field boundary; see Prob. 2.4-3. We found the same result in Prob. 2.3-5 by examining the power density expression. The conditions on the derivation of the ideal dipole fields in Sec. 2.3 were that $\lambda \gg \Delta z$ and $r \gg \Delta z$. The ratio of radiated to reactive power density is $1/(\beta r)^3$, as derived in Prob. 2.3-5. So, if $\beta r = 10$ the radiated power density will be 1000 times (30 dB) greater than the reactive power density. Thus, $r = 10/\beta = 1.6\lambda$ forms the far-field boundary for an ideal dipole. It turns out that the boundary value of $\beta r = 10$ also corresponds to the point where the wave impedance (ratio of transverse field components) converges to the intrinsic impedance of free space, η, indicating that a free space wave is well formed; see Prob. 2.4-4. A more conservative condition of $r = 5\lambda$ arises based on the error in magnitude of the E_θ component of (2-73b) being 30 dB down compared to the far-field approximation field of (2-74a); see Prob. 2.4-5. [3]. Fig. 2-8a summarizes the field regions for the ideal dipole as: the reactive near field extends from $r = 0$ to 0.16λ, the radiating near field from 0.16λ to 5λ, and the far field is beyond 5λ. These results also apply to any *electrically small antenna*, which is defined as fitting inside a sphere of radius a equal to or less than 0.16λ.

The reactive-radiating near field boundary changes for radiating sources other than an ideal dipole. For a line source, this region extends to a distance $0.62\sqrt{D^3/\lambda}$ from the antenna (see Prob. 2.4-6). [H.8.8: Walter, (2-109)] As noted in (2-99), the far field starts at $2D^2/\lambda$. An additional far-field distance condition is that r be greater than D, which was mentioned in association with the approximation $R \approx r$ of (2-87) for use in the

(a) Small antennas ($a \ll \lambda$).

(b) Large antennas (D > 2.5λ).
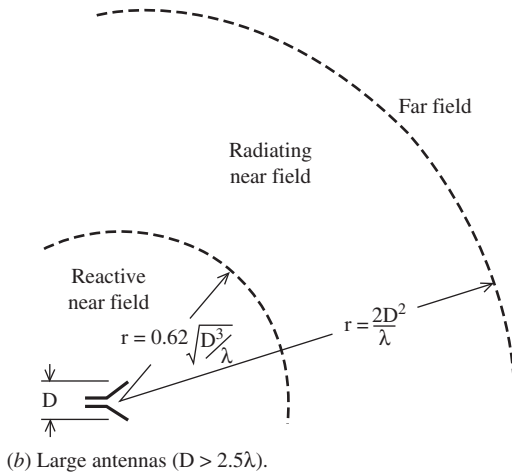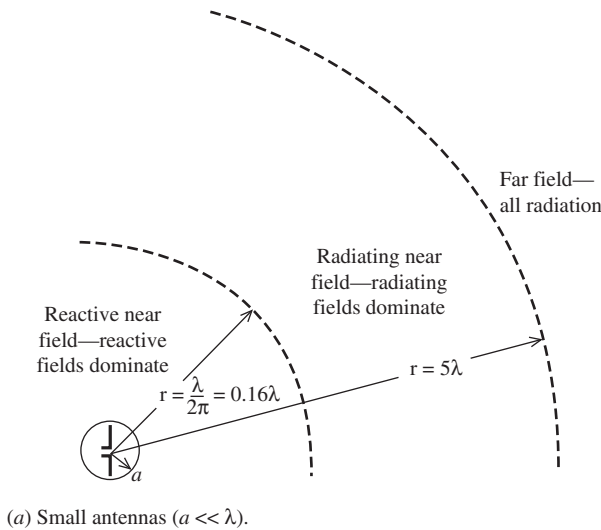
Figure 2-8 Summary of field regions (not to scale).

magnitude dependence. Specifically we require r to be greater than 5 D, which is based on the maximum error in the distance to the far field being 10% see Prob. 2.4-7. In addition, the condition $r \gg \lambda$ follows from $\beta r = (2\pi r/\lambda) \gg 1$ which was used to reduce (2-92) to (2-93). Often this limit is chosen to be $r > 1.6\lambda$ based on reasoning similar to that used for the ideal dipole. The conditions on the far-field distance for line sources of length D are summarized as:

$$r > \frac{2D^2}{\lambda} \qquad\qquad\qquad\qquad (2\text{-}100a)$$

$$r > 5D \qquad \textit{far-field conditions} \qquad (2\text{-}100b)$$

$$r > 1.6\lambda \qquad\qquad\qquad\qquad (2\text{-}100c)$$

The far-field distance criterion of $2D^2/\lambda$ is usually the limiting condition for antennas operating in the UHF region and above. At lower frequencies, where the antenna can be small compared to the operating wavelength, the far-field distance may have to be greater

than $2D^2/\lambda$ in order to satisfy all conditions in (2-100); see Prob. 2.4-8. We will return to the topic of far-field distance in Sec. 13.2.

The radiated fields (i.e., in the far field) are found by summing up all contributions due to tiny subdivisions of the current distributions (i.e., integrating). The resulting radiation pattern is independent of distance from the antenna, r, and varies only with angle around the antenna. For example, the $\sin\theta$ pattern of an ideal dipole is valid anywhere in its far field. This result is a consequence of approximating rays from parts of the source as being parallel. Then distance in the far field is irrelevant and only angles matter. A further general principle for all antennas is that in the far field the only field components present are transverse to the radial direction (i.e., the θ- and ϕ- components).

Fig. 2-8b summarizes the field regions for antennas that are not electrically small. The reactive near field extends from $r = 0$ to $0.62\sqrt{D^3/\lambda}$. The far field lies at distances beyond the largest of $2D^2/\lambda$, $5D$, or 1.6λ. The radiating near field lies between the reactive near field and the far-field distance. The $5D$ and $2D^2/\lambda$ criteria are equal for $D = 2.5\lambda$, so in the context of far-field distance an antenna is considered large if $D > 2.5\lambda$.

2.4.4   Steps in the Evaluation of Radiation Fields

The derivation for the fields radiated by a line source in Sec. 2.4.2 can be generalized for application to any antenna. Fortunately, the derivation itself need not be repeated each time an antenna is analyzed. That is, it is not necessary to return to Maxwell's equation with each new antenna system. Instead, we work from the results of the line source and its generalizations, which can be reduced to the three-step procedure detailed below.

1. **Find A.** Select a coordinate system most compatible with the geometry of the antenna, using the notation of Fig. 2-2. In general, use (2-61) with $R \approx r$ in the magnitude factor and the parallel ray approximation of (2-98) for determining phase differences over the antenna. These yield

$$\mathbf{A} = \mu \frac{e^{-j\beta r}}{4\pi r} \iiint_{v'} \mathbf{J} e^{j\beta \hat{\mathbf{r}} \cdot \mathbf{r}'} \, dv' \qquad (2\text{-}101)$$

   For z-directed volumetric sources,

$$\mathbf{A} = \hat{\mathbf{z}}\mu \frac{e^{-j\beta r}}{4\pi r} \iiint_{v'} J_z e^{j\beta \hat{\mathbf{r}} \cdot \mathbf{r}'} \, dv' \qquad (2\text{-}102)$$

   For z-directed line sources on the z-axis,

$$\mathbf{A} = \hat{\mathbf{z}}\mu \frac{e^{-j\beta r}}{4\pi r} \int I(z') e^{j\beta z' \cos\theta} \, dz' \qquad (2\text{-}103)$$

   which is (2-89).

2. **Find E**. Use the component of

$$\mathbf{E} = -j\omega\mathbf{A} \qquad (2\text{-}104)$$

   which is transverse to the direction of propagation $\hat{\mathbf{r}}$. This is expressed formally as

$$\mathbf{E} = -j\omega\mathbf{A} - (-j\omega\mathbf{A} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} = -j\omega(A_\theta\hat{\boldsymbol{\theta}} + A_\phi\hat{\boldsymbol{\phi}}) \qquad (2\text{-}105)$$

   which arises from the component of $\mathbf{A}$ tangent to the far-field sphere. For z-directed sources,

$$\mathbf{E} = j\omega \sin\theta \, A_z \hat{\boldsymbol{\theta}} \qquad (2\text{-}106)$$

   which is (2-95).

3. **Find H.** Use the plane wave relation.

$$\mathbf{H} = \frac{1}{\eta}\hat{\mathbf{r}} \times \mathbf{E}. \tag{2-107}$$

This equation expresses the fact that in the far field, the directions of **E** and **H** are perpendicular to each other and to the direction of propagation, and also that their magnitudes are related by η. For z-directed sources,

$$H_\phi = \frac{E_\theta}{\eta} \tag{2-108}$$

which is (2-75).

The most difficult step is the first, calculating the radiation integral. This topic will be discussed many times throughout the book, but to immediately develop an appreciation for the process, we will present an example. This uniform line source example also provides a specific setting for introducing general radiation pattern concepts and definitions.

EXAMPLE 2-1   *The Uniform Line Source*

The **uniform line source** is a line source for which the current is constant along its extent. If we use a z-directed uniform line source centered on the origin and along the z-axis, the current is

$$I(z') = \begin{cases} I_o & x' = 0, \quad y = 0, \quad |z'| \leq \dfrac{L}{2} \\[2mm] 0 & \textit{elsewhere} \end{cases} \tag{2-109}$$

where L is the length of the line source; see Fig. 2-9. We first find A_z from (2-103) as follows:

$$\begin{aligned} A_z &= \mu \frac{e^{-j\beta r}}{4\pi r} \int_{-L/2}^{L/2} I_o e^{j\beta z' \cos\theta}\, dz' \\[2mm] &= \mu \frac{e^{-j\beta r}}{4\pi r} I_o \left[\frac{e^{j\beta(L/2)\cos\theta} - e^{-j\beta(L/2)\cos\theta}}{j\beta\cos\theta}\right] \\[2mm] &= \mu \frac{I_o L e^{-j\beta r}}{4\pi r} \frac{\sin\left[(\beta L/2)\cos\theta\right]}{(\beta L/2)\cos\theta}. \end{aligned} \tag{2-110}$$

(a) Antenna geometry.   (b) Current distribution.
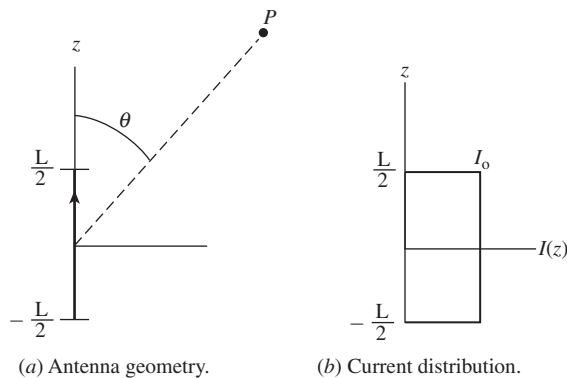
Figure 2-9 The uniform line source (Example 2-1).

The electric field from (2-106) is then

$$\mathbf{E} = j\omega \sin\theta\, A_z \hat{\boldsymbol{\theta}} = \frac{j\omega\mu I_o L e^{-j\beta r}}{4\pi r} \sin\theta \frac{\sin\left[(\beta L/2)\cos\theta\right]}{(\beta L/2)\cos\theta}\hat{\boldsymbol{\theta}} \qquad (2\text{-}111)$$

The magnetic field is simply found from this using $H_\phi = E_\theta/\eta$.

2.4.5   Radiation Pattern Definitions

Since the radiation pattern is the variation of the radiated electric field over a sphere centered on the antenna, r is constant and we have only θ and ϕ variations of the field. It is convenient to normalize the field expression such that its maximum value is unity. This is accomplished as follows for a z-directed source that has only a θ-component of $\mathbf{E}$:

$$F(\theta,\phi) = \frac{E_\theta}{E_\theta(\text{max})} \qquad (2\text{-}112)$$

where $F(\theta,\phi)$ is the *normalized field pattern* and $E_\theta(\text{max})$ is the maximum value of the magnitude of E_θ over a sphere of radius r. Here the pattern is real-valued and the magnitude refers to taking the sign to be positive. Also in this case, the pattern is relative to linear polarization in the θ-orientation. Patterns, in general, can be referenced to any polarization state; see Sec. 13.4 for more details.

In general, E_θ can be complex-valued and, therefore, so can $F(\theta,\phi)$. In this case, the phase is usually set to zero at the same point the magnitude is normalized to unity. This is appropriate because we are only interested in relative phase behavior. The pattern magnitude is obtained by taking the absolute value of (2-112).

An element of current on the z-axis has a normalized field pattern from (2-74a) of

$$F(\theta) = \frac{(I\Delta z/4\pi)j\omega\mu(e^{-j\beta r}/r)\sin\theta}{(I\Delta z/4\pi)j\omega\mu(e^{-j\beta r}/r)} = \sin\theta \qquad (2\text{-}113)$$

and there is no ϕ variation.

The normalized field pattern for the uniform fine source is from (2-111) in (2-112)

$$F(\theta) = \sin\theta \frac{\sin\left[(\beta L/2)\cos\theta\right]}{(\beta L/2)\cos\theta} \qquad (2\text{-}114)$$

and again there is no ϕ variation. The second factor of this expression is the function $\sin(u)/u$ and we will encounter it frequently. It has a maximum value of unity at $u = 0$; this corresponds to $\theta = 90°$, where $u = (\beta L/2)\cos\theta$. Substituting $\theta = 90°$ in (2-114) gives unity and we see that $F(\theta)$ is properly normalized.

In general, a normalized field pattern can be written as the product

$$\boxed{F(\theta,\phi) = g(\theta,\phi)f(\theta,\phi)} \qquad (2\text{-}115)$$

where $g(\theta,\phi)$ is the *element factor* and $f(\theta,\phi)$ the *pattern factor*. The pattern factor comes from the integral over the current and is strictly due to the distribution of current in space. The element factor is the pattern of an infinitesimal current element in the current distribution. For example, we found for a z-directed current element that $F(\theta) = \sin\theta$. This is, obviously, also the element factor, so

$$g(\theta) = \sin\theta \qquad (2\text{-}116)$$

for a z-directed current element. Actually, this factor originates from (2-95) and can be interpreted as the projection of the current element in the θ-direction. In other words,

at $\theta = 90°$ we see the maximum length of the current, whereas at $\theta = 0°$ or $180°$ we see the end view of an infinitesimal current that yields no radiation. The $\sin \theta$ factor expresses the fraction of the size of the current as seen from the observation angle θ. On the other hand, the pattern factor $f(\theta, \phi)$ represents the integrated effect of radiation contributions from the current distribution, which can be treated as being made up of many current elements. The pattern value in a specific direction is then found by summing the parallel rays from each current element to the far field with the magnitude and phase of each included. The radiation integral of (2-101) sums the far-field contributions from the current elements and when normalized yields the pattern factor. Antenna analysis is usually easier to understand by considering the antenna to be transmitting as we have here. However, most antennas are reciprocal and thus their radiation properties are identical when used for reception; see Sec. 13.1.

A typical pattern is shown in Fig. 2-10 as a polar plot and in 3D. The rays from various parts of an antenna arrive in the far field with different magnitude and phase due to variations
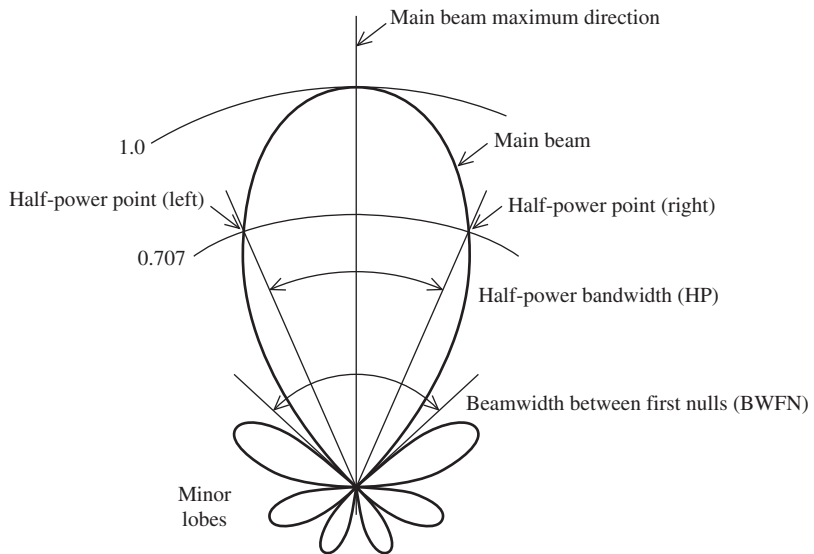
(*a*) Polar plot in relative field intensity.
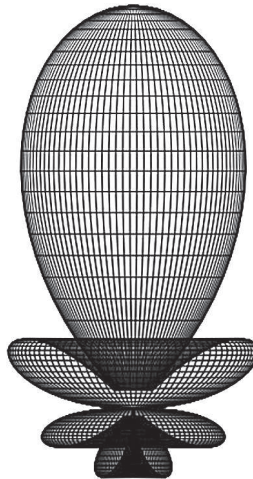
(*b*) Three-dimensional view.

Figure 2-10 A typical radiation pattern.

in the current where the ray originated on the antenna and due to phase changes arising from path length differences to the far field. These rays interfere, as computed through the radiation integral and produce a "lobing" effect.

The radiation lobe containing the direction of maximum radiation shows the angle with the highest radiation intensity and is caused by the fact that the rays from various parts of the antenna arrive in the far field more nearly in-phase than they do for other directions. For a source with constant phase, all rays arrive in-phase in the direction normal to the antenna and the pattern is maximum there. For the ideal dipole, we have said that the source is so small that there are essentially no phase differences for rays along the source and thus the pattern factor is unity.

For the z-directed uniform line source pattern of (2-114), we can identify the factors as

$$g(\theta) = \sin\theta \tag{2-117}$$

and

$$f(\theta) = \frac{\sin\left[(\beta L/2)\cos\theta\right]}{(\beta L/2)\cos\theta} \tag{2-118}$$

For long line sources ($L \gg \lambda$), the pattern factor of (2-118) is much sharper than the element factor $\sin\theta$, and the total pattern is approximately that of (2-118); that is, $F(\theta) \approx f(\theta)$. Hence, in many cases we need only work with $f(\theta)$, which is obtained from (2-103). If we allow the beam to be scanned (this will be discussed in Sec. 2.4.6), the element factor becomes important as the pattern maximum approaches the z-axis.

Frequently, the directional properties of the radiation from an antenna are described by another form of radiation pattern, the power pattern. The power pattern gives angular dependence of the power density and is found from the θ, ϕ variation of the r-component of the Poynting vector. For z-directed sources, $H_\phi = E_\theta/\eta$ so the r-component of the Poynting vector is $(1/2)\,E_\theta H_\phi^* = |E_\theta|^2/2\eta$ and the normalized power pattern is simply the square of its field pattern magnitude $P(\theta) = |F(\theta)|^2$. The general normalized *power pattern* is

$$\boxed{P(\theta, \phi) = |F(\theta, \phi)|^2} \tag{2-119}$$

The normalized power pattern for a z-directed current element is

$$P(\theta) = \sin^2\theta \tag{2-120}$$

and for a z-directed uniform line source is

$$P(\theta) = \left\{\sin\theta\,\frac{\sin\left[(\beta L/2)\cos\theta\right]}{(\beta L/2)\cos\theta}\right\}^2 \tag{2-121}$$

Frequently, patterns are plotted in decibels. *It is important to recognize that the field (magnitude) pattern and power pattern are the same in decibels.* This follows directly from the definitions. From the field pattern in decibels,

$$|F(\theta, \phi)|_{\text{dB}} = 20\log|F(\theta, \phi)| \tag{2-122}$$

and the power pattern in decibels,

$$P(\theta, \phi)_{\text{dB}} = 10\log P(\theta, \phi) = 10\log|F(\theta, \phi)|^2 = 20\log|F(\theta, \phi)| \tag{2-123}$$

we see that

$$\boxed{P(\theta, \phi)_{\text{dB}} = |F(\theta, \phi)|_{\text{dB}}} \tag{2-124}$$

Fig. 13-6 shows the same antenna pattern plotted in the four common pattern plot formats.

2.4.6 Radiation Pattern Parameters

A typical radiation pattern shown in Fig. 2-10a is a polar plot in linear units (rather than decibels). It consists of several *lobes*. The *main beam* (or *main lobe* or *major lobe*) is the lobe containing the direction of maximum radiation. There is also usually a series of lobes smaller than the main lobe. Any lobe other than the main lobe is called a *minor lobe*. Minor lobes are composed of *side lobes* and *back lobes*. Back lobes are directly opposite the main lobe, or sometimes they are taken to be the lobes in the half-space opposite the main lobe. The term side lobe is sometimes reserved for those minor lobes near the main lobe, but is most often taken to be synonymous with minor lobe; we will use the latter convention.

The radiation from an antenna is represented mathematically through the radiation pattern function, $F(\theta, \phi)$ for field and $P(\theta, \phi)$ for power. This angular distribution of radiation is visualized through various graphical representations of the pattern, which we discuss in this section. Graphical representations also are used to introduce definitions of pattern parameters that are commonly used to quantify radiation pattern characteristics.

A three-dimensional plot as in Fig. 2-10b gives a good overall impression of the entire radiation pattern, but cannot convey accurate quantitative information. Cuts through this pattern in various planes are the most popular pattern plots. They usually include the *E*- and *H*-plane patterns; see Figs. 2-4b and 2-4c. Pattern cuts are often given for fixed ϕ values, leaving the pattern a function of θ alone; we will assume that is the case here. Typically, the side lobes are alternately positive- and negative-valued; see Fig. 5-1a. In fact, a pattern in its most general form can be complex-valued. Then we use the magnitude of the field pattern $|F(\theta)|$ or the power pattern $P(\theta)$.

A measure of how well the power is concentrated into the main lobe is the (relative) *side lobe level*, which is the ratio of the pattern value of a side lobe peak to the pattern value of the main lobe. The largest side lobe level for the whole pattern is the *maximum (relative) side lobe level,* frequently abbreviated as SLL. In decibels, it is given by

$$\text{SLL}_{dB} = 20 \log \frac{|F(\text{SLL})|}{|F(\text{max})|} \qquad (2\text{-}125)$$

where $|F(\text{max})|$ is the maximum value of the pattern magnitude and $|F(\text{SLL})|$ is the pattern value of the maximum of the highest side lobe magnitude. For a normalized pattern, $F(\text{max}) = 1$.

The width of the main beam is quantified through *half-power beamwidth* HP, which is the angular separation of the points where the main beam of the power pattern equals one-half the maximum value:

$$\text{HP} = |\theta_{\text{HP left}} - \theta_{\text{HP right}}| \qquad (2\text{-}126)$$

where $\theta_{\text{HP left}}$ and $\theta_{\text{HP right}}$ are points to the "left" and "right" of the main beam maximum for which the normalized power pattern has a value of one-half (see Fig. 2-10). On the field pattern $|F(\theta)|$, these points correspond to the value $1/\sqrt{2}$. For example, the $\sin \theta$ pattern of an ideal dipole has a value of $1/\sqrt{2}$ for θ values of $\theta_{\text{HP left}} = 135°$ and $\theta_{\text{HP right}} = 45°$. Then $\text{HP}|135° - 45°| = 90°$. This is shown in Fig. 2-4b. Note that the definition of HP is the magnitude of the difference of the half-power points and the assignment of left and right can be interchanged without changing HP. In three dimensions, the radiation pattern major lobe becomes a solid object and the half-power contour is a continuous curve. If this curve is essentially elliptical, the pattern cuts that contain the major and minor axes of the ellipse determine what the IEEE defines as the *principal half-power beamwidths.*

(*a*) Broadside   (*b*) Intermediate   (*c*) Endfire
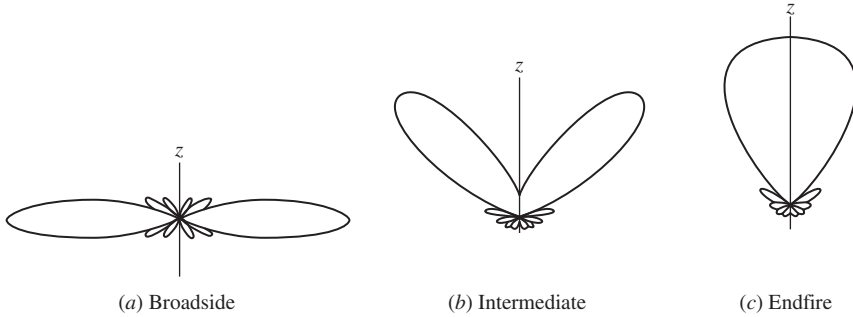
Figure 2-11 Polar plots of uniform line source patterns $|f(\theta)|$.

Antennas are often referred to by the type of pattern they produce. An *isotropic antenna*, which is hypothetical, radiates equally in all directions giving a constant radiation pattern. An *omnidirectional antenna* produces a pattern that is constant in one plane; the ideal dipole of Fig. 2-4 is an example. The pattern shape resembles a "doughnut." In contrast to the unrealizable isotropic pattern, the omnidirectional pattern is achievable, and it has many practical applications. For example, antennas in cellular telephone handsets should respond equally in all directions around the handset in the horizontal plane because of their arbitrary orientation when held by a human operator. A second application that often requires an omnidirectional pattern is a base station antenna, where uniform coverage is needed to serve users on the ground anywhere around the tower.

We often refer to a directive antenna as being broadside or endfire. A *broadside antenna* is one for which the main beam maximum is in a direction normal to the plane containing the antenna. An *endfire antenna* is one for which the main beam is in the plane containing the antenna. For a linear current on the *z*-axis, the broadside direction is $\theta = 90°$ and the endfire directions are $0°$ and $180°$. For example, an ideal dipole is a broadside antenna. For *z*-directed line sources, several patterns are possible. Fig. 2-11 illustrates a few $|f(\theta)|$ patterns. The entire pattern in three dimensions is imagined by rotating the pattern about the *z*-axis. The full pattern can then be generated from the *E*-plane patterns shown. The broadside pattern of Fig. 2-11*a* is called a *fan beam*. The full three-dimensional endfire pattern for Fig. 2-11*b* has a single lobe in the endfire direction. This single lobe is referred to as a *pencil beam*. Note that the $\sin\theta$ element factor, which must multiply these patterns to obtain the total pattern, will have a significant effect on the endfire pattern.

2.5   DIRECTIVITY AND GAIN

One very important quantitative description of an antenna is how much it concentrates energy in one direction in preference to radiation in other directions. This characteristic of an antenna is called directivity and is equal to its gain if the antenna is 100% efficient. Gain is expressed relative to an isotropic radiator or sometimes a half-wavelength dipole.

Toward the definition of directivity, let us begin by recalling that the power radiated by an antenna from (2-29) is

$$P = \iint \mathbf{S} \cdot d\mathbf{s} = \frac{1}{2}\mathrm{Re}\iint (\mathbf{E} \times \mathbf{H}^*) \cdot d\mathbf{s} \tag{2-127}$$

$$= \frac{1}{2}\mathrm{Re}\int_0^{2\pi}\int_0^{\pi}(E_\theta H_\phi^* - E_\theta H_\phi^*)r^2 \sin\theta\, d\theta\, d\phi \tag{2-128}$$
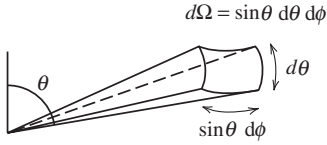
Figure 2-12 Element of solid angle $d\Omega$.

In general, there will be both θ- and ϕ-components of the radiation fields. From (2-107), we find that

$$H_\phi = \frac{E_\theta}{\eta} \quad \text{and} \quad H_\theta = -\frac{E_\phi}{\eta} \qquad (2\text{-}129)$$

Using these in (2-128) gives

$$P = \frac{1}{2\eta}\iint (|E_\theta|^2 + |E_\phi|^2)r^2\, d\Omega \qquad (2\text{-}130)$$

where $d\Omega =$ element of solid angle $= \sin\theta\, d\theta\, d\phi$, which is shown in Fig. 2-12. The integral can be evaluated over any surface enclosing the antenna; however, for simplicity a spherical surface centered on the origin is usually used. Since the magnitude variations of the radiation fields are $1/r$, we find it convenient to introduce *radiation intensity*, which is defined from

$$U(\theta, \phi) = \tfrac{1}{2}\mathrm{Re}(\mathbf{E} \times \mathbf{H^*}) \cdot r^2\hat{\mathbf{r}} = S(\theta, \phi)r^2 \qquad (2\text{-}131)$$

Radiation intensity is the power radiated in a given direction per unit solid angle and has units of watts per square radian (or steradian, sr). The advantage to using radiation intensity is that it is independent of distance r. Radiation intensity can be expressed as

$$U(\theta, \phi) = U_m|F(\theta, \phi)|^2 \qquad (2\text{-}132)$$

where U_m is the maximum radiation intensity, and $|F(\theta, \phi)|^2$ is the power pattern normalized to a maximum value of unity in the direction $(\theta_{max}, \phi_{max})$, and

$$U_m = U(\theta_{max}, \phi_{max}) \qquad (2\text{-}133)$$

The total power radiated is obtained by integrating the radiation intensity over all angles around the antenna:

$$P = \iint U(\theta, \phi)\, d\Omega = U_m\iint |F(\theta, \phi)|^2\, d\Omega \qquad (2\text{-}134)$$

An isotropic source with uniform radiation in all directions is only hypothetical but is sometimes a useful concept. The radiation intensity of an isotropic source is constant over all space, at a value of U_{ave}. Then $P = \iint U_{ave}\, d\Omega = U_{ave}\iint d\Omega = 4\pi U_{ave}$ since there are 4π sr in all space (see Prob. 2.5-1). For nonisotropic sources, the radiation intensity is not constant throughout space, but an average power per steradian can be defined as

$$U_{ave} = \frac{1}{4\pi}\iint U(\theta, \phi)\, d\Omega = \frac{P}{4\pi} \qquad (2\text{-}135)$$

The average radiation intensity U_{ave} equals the radiation intensity $U(\theta, \phi)$ that an isotropic source with the same input power P would radiate.

As an example, consider the ideal dipole again; we find from (2-74) and (2-131) that

$$U(\theta, \phi) = \frac{1}{2}\left(\frac{I\Delta z}{4\pi}\right)^2 \beta\omega\mu \sin^2\theta \tag{2-136}$$

so

$$U_m = \frac{1}{2}\left(\frac{I\Delta z}{4\pi}\right)^2 \beta\omega\mu \tag{2-137}$$

and

$$F(\theta, \phi) = \sin\theta \tag{2-138}$$

The average radiation intensity follows from the total radiated power expression (2-77) for an ideal dipole as

$$U_{\text{ave}} = \frac{P}{4\pi} = \frac{(\beta\omega\mu/12\pi)(I\Delta z)^2}{4\pi} = \frac{1}{3}\left(\frac{I\Delta z}{4\pi}\right)^2 \beta\omega\mu$$

$$= \frac{2}{3}\, U_m \quad \textit{ideal dipole} \tag{2-139}$$

Thus, $U_m = 1.5 U_{\text{ave}}$ for the ideal dipole, which means that in the direction of maximum radiation, the radiation intensity is 50% more than that which would occur from an isotropic source radiating the same total power.

Directivity.   Directivity is defined as the ratio of the radiation intensity in a certain direction to the average radiation intensity. The reference direction is usually taken to be that of the maximum radiation, giving

$$D = \frac{U_m}{U_{\text{ave}}} = \frac{U_m}{P/4\pi} \tag{2-140}$$

Unless otherwise stated, directivity will be assumed to be this maximum directivity value. If both the numerator and denominator are divided by r^2, then (2-140) becomes a ratio of power densities at distance r from the antenna:

$$D = \frac{U/r^2}{U_{\text{ave}}/r^2} = \frac{\frac{1}{2}\max[\text{Re}(\mathbf{E} \times \mathbf{H}^*) \cdot \hat{\mathbf{r}}]}{P/4\pi r^2} \tag{2-141}$$

Thus, directivity has two interpretations, as a ratio of maximum to average radiation intensities (which are dimensionless) and as a ratio of maximum to average power densities (which have units of W/m^2).

Directivity can be tied more directly to the pattern function. First, we define *beam solid angle*, Ω_A:

$$\boxed{\Omega_A = \iint_{\text{sphere}} |F(\theta, \phi)|^2 \, d\Omega} \qquad \textit{beam solid angle} \tag{2-142}$$

As illustrated in Fig. 2-13, beam solid angle of the actual radiation pattern has the same solid angle as constant radiation of a level equal to the maximum value radiation intensity over all the beam area, or

$$P = U_m \, \Omega_A \tag{2-143}$$
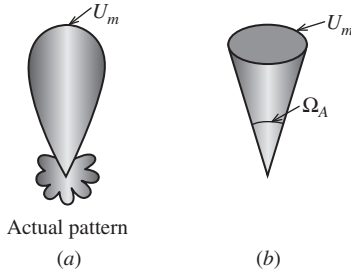
Figure 2-13 Antenna beam solid angle Ω_A. (*a*) Plot of radiation intensity $U(\theta, \phi)$ from an actual antenna. (*b*) Plot of radiation intensity with all radiation from the actual antenna concentrated into a cone of solid angle Ω_A with constant radiation intensity equal to the maximum of the actual pattern.

This also follows from (2-134) and (2-142). Finally, substituting (2-135) in (2-140) along with (2-142) gives

$$\boxed{D = \frac{4\pi}{\Omega_A}} \qquad directivity \qquad (2\text{-}144)$$

These results show that directivity is entirely determined by the pattern shape; it is independent of the details of the antenna hardware. Directivity as a function of pattern angle is expressed simply as the directivity multiplied by the power pattern:

$$D(\theta, \phi) = D|F(\theta, \phi)|^2 \qquad (2\text{-}145)$$

Since $|F(\theta, \phi)|^2$ has a maximum value of unity, the maximum value of directivity as a function of angle is D.

The concept of directivity is illustrated in Fig. 2-14. If the radiated power were distributed isotropically over all of space, the radiation intensity would have a maximum value equal to its average value as shown in Fig. 2-14*a*, that is, $U_m = U_{ave}$ or $\Omega_A = 4\pi$. Thus, the directivity of this isotropic pattern is unity. The distribution of radiation intensity $U(\theta, \phi)$ for an actual antenna is shown in Fig. 2-14*b*. It has a maximum radiation intensity in the direction $(\theta_{max}, \phi_{max})$ of $U_m = DU_{ave}$ and an average radiation intensity of $U_{ave} = P/4\pi$. There is D times as much power density in the direction $(\theta_{max}, \phi_{max})$ as there would be if the same total power were radiated by an isotropic source. Thus, by directing the radiated power P in a preferred direction (the maximum radiation direction) the radiation intensity is increased in that direction by a factor of D over what it would be if the same radiated power had been isotropically radiated.

(*a*) Radiation intensity distributed isotropically.

(*b*) Radiation intensity from an actual antenna.
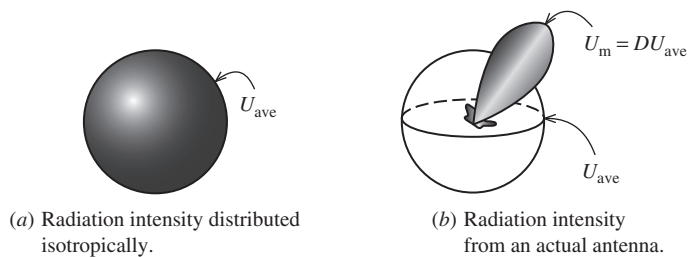
Figure 2-14 Illustration of directivity.

EXAMPLE 2-2   *Directivity of an Ideal Dipole*

The directivity of an ideal dipole can now be easily calculated using (2-139) in (2-140)

$$D = \frac{U_m}{U_{\text{ave}}} = \frac{U_m}{\frac{2}{3}U_m} = \frac{3}{2} \qquad \textit{ideal dipole} \qquad (2\text{-}146)$$

Usually directivity is calculated directly from (2-144) and the directivity calculation reduces to one of finding the beam solid angle. To illustrate, we use the ideal dipole. Substituting (2-138) in (2-142) leads to

$$\Omega_A = \int_0^{2\pi} \int_0^{\pi} |\sin\theta|^2 \sin\theta \, d\theta \, d\phi = 2\pi\frac{4}{3} = \frac{8\pi}{3} \qquad (2\text{-}147)$$

and we obtain the same value of directivity from

$$D = \frac{4\pi}{\Omega_A} = \frac{4\pi}{8\pi/3} = \frac{3}{2} \qquad (2\text{-}148)$$

Thus, the directivity of an ideal dipole is 50% greater than that of an isotropic source, which has a directivity of 1.

EXAMPLE 2-3   *Directivity of a Sector Omnidirectional Pattern*

An ideal omnidirectional antenna would have constant radiation in the horizontal plane ($\theta = 90°$) and would fall rapidly to zero outside that plane. Suppose that the pattern in the vertical plane is constant out to $\pm\frac{\pi}{6}(\pm 30°)$ from horizontal. The pattern expression is then written as

$$F(\theta) = \begin{cases} 1 & \frac{1}{3}\pi < \theta < \frac{2}{3}\pi \\ 0 & \text{elsewhere} \end{cases} \qquad (2\text{-}149)$$

The solid angle of the pattern from (2-142) is

$$\Omega_A = \int\int |F(\theta,\phi)| \, d\Omega = \int_0^{2\pi} \int_{\pi/3}^{2\pi/3} \sin\theta \, d\theta \, d\phi$$

$$= (2\pi)[-\cos\theta]_{\pi/3}^{2\pi/3} = (2\pi)(0.5 + 0.5) = 2\pi \qquad (2\text{-}150)$$

The directivity from (2-144) is

$$D = \frac{4\pi}{\Omega_A} = \frac{4\pi}{2\pi} = 2. \qquad (2\text{-}151)$$

Gain.   As noted, directivity is solely determined by the radiation pattern of an antenna. When an antenna is used in a system (say as a transmitting antenna) we are interested in how efficiently the antenna transforms available power at its input terminals into radiated power, as well as its directive properties. To quantify this, *gain* is defined as 4π times the ratio of the maximum radiation intensity to the net power accepted by the antenna from the connected transmitter, or

$$G = \frac{4\pi U_m}{P_m} \qquad \textit{gain} \qquad (2\text{-}152)$$

This is a power quantity and is sometimes referred to as power gain. This definition does not include losses due to mismatches of impedance or polarization, which are discussed in Sec. 4.2.

Gain can be expressed as a function of angle by including the power pattern as $G(\theta, \phi) = G|F(\theta, \phi)|^2$. Gain can also be given as a value in a specific direction. If no direction is specified and the gain value is not given as a function of θ and ϕ, it is assumed to be the maximum gain.

Directivity from (2-140) is $D = 4\pi U_m/P$. Comparing this with (2-152), we see that the only difference between gain and directivity is the power value used. Directivity can be viewed as the gain an antenna would have if all input power appeared as radiated power, that is, $P_{in} = P$. Gain includes the fact that real antennas do not behave in this fashion and some of the input power is lost on the antenna. The portion of input power P_{in} that does not appear as radiated power is absorbed on the antenna and nearby structures. This prompts us to define *radiation efficiency, e_r,* as

$$e_r = \frac{P}{P_{in}} \qquad \textit{radiation efficiency} \tag{2-153}$$

Radiation efficiency is bounded as

$$0 \le e_r \le 1 \tag{2-154}$$

Many antennas are highly efficient, with radiation efficiency close to unity. However, electrically small antennas are very inefficient, and radiation efficiency is much less than unity.

Using $P = e_r P_{in}$ from (2-153) in (2-140) leads to

$$\boxed{G = e_r D} \tag{2-155}$$

Thus, the gain of an antenna is equal to its purely directional characteristic of directivity multiplied by radiation efficiency. The term *absolute gain,* which is synonymous with gain, is sometimes used.

Units for Directivity and Gain. Because gain is a power ratio it can be calculated in decibels as follows:

$$G_{dB} = 10 \log G \tag{2-156}$$

Similarly for directivity:

$$D_{dB} = 10 \log D \tag{2-157}$$

For example, the directivity in decibels of an ideal dipole is

$$D_{dB} = 10 \log 1.5 = 1.76 \, dB \qquad \textit{ideal dipole directivity} \tag{2-158}$$

Frequently gain is used to describe the performance of the antenna relative to some standard reference antenna. This *relative gain* is defined as the ratio of the maximum radiation intensity from the antenna U_m to the maximum radiation intensity from a reference antenna $U_{m,ref}$ with the same input power, or

$$G_{ref} = \frac{U_m}{U_{m,ref}} \tag{2-159}$$

This is a convenient definition from a measurement standpoint. The formal definition of gain employs a lossless isotropic antenna as a reference antenna. This can be shown by noting that the lossless isotropic reference antenna has a maximum radiation intensity of $P_{in}/4\pi$ because all of its input power is radiated, and substituting this into (2-159) for $U_{m,ref}$ leads to (2-152).

It is common at frequencies below 1 GHz to quote gain values relative to that of a half-wave dipole because a half-wave dipole is the usual reference antenna used in gain measurements at these frequencies; see Sec. 13.3. The directivity of a half-wave dipole is 1.64 or 2.15 dB; see Sec. 3.2. Gain relative to a half-wave dipole carries the units of dBd and is related to gain relative to the normal implied reference of a lossless, isotropic antenna as

$$G(\text{dBd}) = G(\text{dB}) - 2.15 \qquad (2\text{-}160)$$

The unit of gain is sometimes denoted as dBi instead of just dB to emphasize that an isotropic antenna is used as the reference. As a numerical example, the following forms are equivalent gain values for an antenna with 6.1 dB gain:

$$G = 6.1\,\text{dB} = 6.1\,\text{dBi} = 3.95\,\text{dBd} \qquad (2\text{-}161)$$

As a note of caution, sometimes the unit of dB is used in antenna specifications when dBd is intended, especially in the land mobile industry below 1 GHz.

2.6   ANTENNA IMPEDANCE

The primary function of a transmitting antenna is to convert a bound wave to an unbound (i.e., radiated) wave, and vice versa for a receiving antenna. Whereas the transmission line connected to an antenna binds the wave and prevents it from radiating, the antenna itself should instead enable radio waves to leave the structure. The antenna is an interface between wave phenomena on and beyond the antenna to the connecting circuit hardware. The antenna input terminals form the interface point and the circuit parameter of impedance is used to characterize the input to the antenna. The *input impedance* of an antenna (or simply *antenna impedance)* will be affected by other antennas or objects that are nearby, but the discussion here assumes an isolated antenna.

As with conventional circuits, antenna impedance is composed of real and imaginary parts.

$$\boxed{Z_A = R_A + jX_A} \qquad \textit{input impedance} \qquad (2\text{-}162)$$

Fig. 2-15a shows the general antenna model and Fig. 2-15b shows the equivalent model for a transmitting antenna. As a consequence of reciprocity, the impedance of an antenna is identical for receiving and transmitting operation. Receiving antennas and models for them are discussed in Sec. 4.2. The input resistance, R_A, represents dissipation which occurs in two ways. Power that leaves the antenna and never returns (i.e., radiation) is one
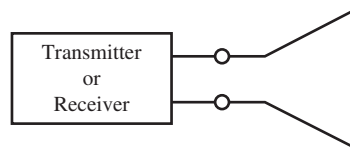
(*a*) General antenna model.
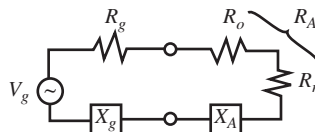
(*b*) Equivalent model for a transmitting antenna.

Figure 2-15 Antenna models.

part of the dissipation. There are also ohmic losses just as in a lumped resistor. Electrically small antennas can have significant ohmic losses but other antennas usually have ohmic losses that are small compared to their radiation dissipation. The input reactance, X_A, represents power stored in the near fields of the antenna.

First, we discuss the input resistance. The average power dissipated in an antenna is

$$P_{\text{in}} = \frac{1}{2} R_A |I_A|^2 \qquad (2\text{-}163)$$

where I_A is the current at the input terminals. Note that a factor of $\frac{1}{2}$ is present because current I_A is the peak value in the time waveform. Separating the dissipated power into radiative and ohmic losses gives

$$P_{\text{in}} = P + P_o$$
$$\frac{1}{2} R_A |I_A|^2 = \frac{1}{2} R_r |I_A|^2 + \frac{1}{2} R_o |I_A|^2 \qquad (2\text{-}164)$$

where we define the radiation resistance of an antenna referred to the input terminals as

$$R_r = \frac{2P}{|I_A|^2} \qquad (2\text{-}165)$$

It follows from (2-164) that

$$R_A = R_r + R_o \qquad (2\text{-}166)$$

where R_o is the resistance associated with ohmic losses that include the directly driven part of the antenna as well as losses in other portions of the antenna structure such as a ground plane. Ohmic resistance of an antenna is defined as

$$R_o = \frac{2P_o}{|I_A|^2} = \frac{2(P_{\text{in}} - P)}{|I_A|^2} \qquad (2\text{-}167)$$

The radiated power is found using (1-35):

$$P = \frac{1}{2} \iint\limits_{s_{\text{ff}}} (\mathbf{E} \times \mathbf{H}^*) \cdot d\mathbf{s} \qquad (2\text{-}168)$$

where s_{ff} is a surface in the far field, usually spherical. P is real-valued because the power density $\mathbf{S} = \frac{1}{2} \mathbf{E} \times \mathbf{H}^*$ is real-valued in the far field.

Radiation resistance can be defined relative to the current at any point on the antenna, but we reserve R_r for radiation resistance referred to the input terminals. Radiation resistance relative to the maximum current I_m that occurs on the antenna R_{rm} is obtained by using I_m in place of I_A in (2-165). In this section, we discuss center-fed electrically short antennas, which always have a current maximum at the input, so $R_r = R_{rm}$. We discuss this topic again in Sec. 6.1.

The power radiated from an ideal dipole of length $\Delta z \ll \lambda$ and input current $I_A = I$ is given by (2-77) which together with (2-165) gives the radiation resistance:

$$R_r = \frac{2P}{|I_A|^2} = \frac{2}{I^2} \frac{\omega\mu\beta}{12\pi} (I\Delta z)^2 = \frac{\sqrt{\mu\omega}\sqrt{\mu}\sqrt{\varepsilon}}{\sqrt{\varepsilon}\,6\pi} \beta(\Delta z)^2$$

$$= \eta \frac{\beta^2}{6\pi} (\Delta z)^2 = \eta \frac{2}{3}\pi \left(\frac{\Delta z}{\lambda}\right)^2$$

$$R_r = 80\pi^2 \left(\frac{\Delta z}{\lambda}\right)^2 \Omega \qquad \textit{ideal dipole} \qquad (2\text{-}169)$$

For ideal dipoles, R_r is very small since $\Delta z \ll \lambda$.

Except for low frequencies, the skin depth $\delta = \sqrt{2/\omega\mu\sigma}$ is much smaller than the conductor radius and then the ohmic resistance for an antenna of length L that carries an axially uniform current is directly proportional to the wire length and inversely proportional to the wire circumference:

$$R_o \approx \frac{L}{2\pi a} R_s \tag{2-170}$$

where L is the length of the wire, a the wire radius, and R_s the surface resistance:

$$R_s = \sqrt{\frac{\omega\mu}{2\sigma}} \tag{2-171}$$

The ideal dipole has a uniform current as shown in Fig. 2-16a. In reality, the current on a straight wire antenna must smoothly go to zero at the wire ends. The current distribution on a center-fed wire dipole of length $\Delta z \ll \lambda$, called a **short dipole**, is approximately triangular in shape as illustrated in Fig. 2-16b. If end loading such as with metal plates (see Fig. 3-3) is added to the short dipole, the radial current reduces to zero at the edge of the plates, giving a nearly uniform current on the vertical portion of the dipole, which permits use of the ideal dipole model. More will be said about short dipoles in Sec. 3.1.

Pattern calculations for the ideal dipole were performed in Sec. 2.3 assuming that the magnitude and phase differences of rays coming from points on the wire due to different path lengths were negligible. Since the short dipole also satisfies $\Delta z \ll \lambda$, the pattern will also be the same $\sin\theta$ radiation pattern as the ideal dipole. In addition, the ideal dipole and short dipole will have the same directivity value of 1.5 because pattern shape completely determines directivity.

The triangular current distribution of the short dipole leads to an equivalent length that is one-half that of its physical length. This is because the equivalent length is proportional to the area under the current versus distance curves shown in Fig. 2-16, which follows from the radiation integral of (2-103) with $\exp(j\beta z' \cos\theta) \approx 1$ for short dipoles. The radiated fields are, in turn, proportional to this equivalent length. Because the radiation resistance is proportional to the integral of the far-zone electric field squared and the patterns of the ideal and short dipoles are the same, the radiation resistances are proportional to the equivalent lengths squared. The area of the triangle-shape current on the short dipole is one-half that of the uniform current shape, so the radiation resistance is one fourth that of the ideal dipole. Dividing (2-169) by 4 gives

(a) Ideal dipole.   (b) Short dipole.
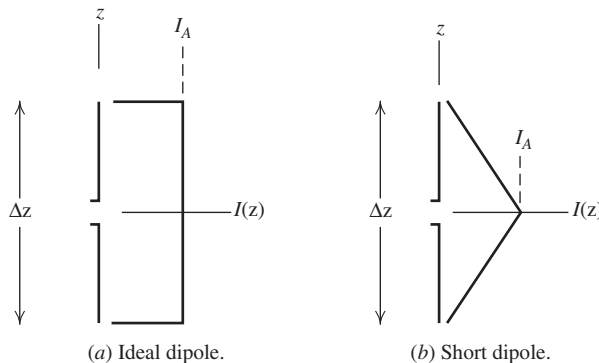
Figure 2-16 The ideal dipole and short dipole with current distributions; $\Delta z \ll \lambda$. I_A is the value of the input current at the terminals in the center of each antenna. The short dipole of (b) is that which is encountered in practice.

$$R_r = 20\pi^2 \left(\frac{\Delta z}{\lambda}\right)^2 \Omega \qquad \textit{short dipole} \qquad (2\text{-}172)$$

The ohmic resistance for the short dipole is less than that of the ideal dipole. The ohmic resistance of a short dipole is found by first determining the power dissipation from ohmic losses, which at any point along the antenna is proportional to the current squared. In fact, in general the total power dissipated is evaluated by integrating the current squared over the wire antenna, which together with (2-167) yields

$$R_o = \frac{2P_o}{|I_A|^2} = \frac{1}{|I_A|^2} \frac{R_s}{2\pi a} \int_{-L.2}^{L/2} |I(z)|^2 \, dz. \qquad (2\text{-}173)$$

It is easy to show that this reduces to (2-170) for a uniform current of length $L = \Delta z$. The short dipole triangular current of Fig. 2-16b can be written as a function of position along the wire as

$$I(z) = I_A \left(1 - \frac{2|z|}{\Delta z}\right), \qquad |z| \leq \frac{\Delta z}{2} \qquad (2\text{-}174)$$

Using this in (2-173) yields

$$R_o = \frac{\Delta z}{2\pi a} \frac{R_s}{3} \qquad \textit{short dipole} \qquad (2\text{-}175)$$

Notice that this is one-third that for an ideal dipole of the same length Δz given by (2-170). Since the radiation resistance for the short dipole is one-fourth that of an ideal dipole, the radiation resistance is decreased more relative to the ohmic resistance, and thus the efficiency is lower for the short dipole than it is for an ideal dipole of the same length.

   In contrast to radiated power that contributes to the real part of the input impedance, the reactive part of the input impedance represents power stored in the near field. This behavior is very similar to a complex load impedance in circuit theory. Antennas that are electrically small (i.e., much smaller than a wavelength) have a large input reactance, in addition to a small radiation resistance. For example, the short dipole has a capacitive reactance, whereas an electrically small loop antenna has an inductive reactance. This is an expected result from low-frequency circuit theory. The reactance of a short dipole is approximated by [H.6: *Ant. Eng. Hdbk.*, 4th ed., p. 4–5].

$$X_A = -\frac{120}{\pi \frac{\Delta z}{\lambda}} \left[\ln\left(\frac{\Delta z}{2a}\right) - 1\right] \Omega \qquad \textit{short dipole} \qquad (2\text{-}176)$$

This gives a large capacitive reactance for very short dipoles. The total input impedance of the short dipole is $R_r + R_o + jX_A$, where X_A is given above, R_r is given by (2-172), and R_o is given by (2-175).

   Antenna impedance is important to the transfer of power from a transmitter to an antenna or from an antenna to a receiver. For example, to maximize the power transferred from a receiving antenna, the receiver impedance should be a conjugate match to the antenna impedance (equal resistances, equal magnitude and opposite sign reactances). Receivers have real-valued impedance, typically 50 Ω, so it is necessary to "tune out" the antenna reactance with a matching network. There are two disadvantages to using matching networks: Ohmic losses in the network components such as tuning coils reduce

efficiency, and second, a matching network provides a match only over a narrow band of frequencies, which reduces the operational bandwidth. Impedance-matching techniques are discussed in Sec. 6.4.

2.7   RADIATION EFFICIENCY

In general, an efficiency factor is the ratio of wanted power to total power supplied. The radiation efficiency e_r of an antenna, which was introduced in Sec. 2.5 and defined in (2-153), is the ratio of radiated power (which is the wanted power) to the net power accepted by the antenna:

$$e_r = \frac{P}{P_{in}} = \frac{P}{P + P_o} \qquad (2\text{-}177)$$

where

P = power radiated
P_o = power dissipated in ohmic losses on the antenna
$P_{in} = P + P_o$ = input power = power accepted by the antenna

We can derive an expression for the radiation efficiency of an antenna totally in terms of resistances by substituting (2-164) and (2-165) into (2-177):

$$e_r = \frac{\frac{1}{2} R_r |I_A|^2}{\frac{1}{2} R_r |I_A|^2 + \frac{1}{2} R_o |I_A|^2} = \frac{R_r}{R_r + R_o} = \frac{R_r}{R_A} \qquad (2\text{-}178)$$

where (2-166) was used. Thus, the radiation resistance relative to the total input resistance gives the radiation efficiency (often referred to just as efficiency). This fundamental equation contains the obvious result that for a fixed radiation resistance as the ohmic resistance increases, the efficiency decreases. For many antennas radiation efficiency is nearly 100%, but electrically small antennas often have very low radiation efficiency. Low radiation efficiency is acceptable for receiving and for low-power transmitting applications. However, radiation efficiency must be as close to 100% as possible in high-power transmitting antennas for two reasons: the ohmic losses cost money in power consumption and the heat generated can possibly damage or destroy the antenna. The following example illustrates calculation of impedance and efficiency.

EXAMPLE 2-4   *Impedance and Radiation Efficiency of a Car Radio Whip Antenna*

The 31-in-long ($h = 0.787$ m) fender-mount whip antenna was used on cars for reception of AM and FM radio signals for many decades. Current production cars mainly use on-glass antennas, which are discussed further in Sec. 12.3. The fender-mount whip antenna in this example, however, provides a point of reference and usually has near maximum performance. For simplicity, we assume that the fender provides a good ground plane for the monopole. The whip is made of steel 1/8-in in diameter ($a = 0.15875$ cm). For an operating frequency in the AM band of 1 MHz ($\lambda = 300$ m) the electrical length is $h = 0.0026\,\lambda$, indicating that the whip is definitely an electrically small antenna. Using the short monopole model as will be developed in (3-15), the radiation resistance is

$$R_r = 40\pi^2 \left(\frac{0.787}{300}\right)^2 = 0.00271\,\Omega \qquad (2\text{-}179)$$

Using the conductivity of stainless steel (see App. B.l) in (2-171) gives

$$R_s = \sqrt{\frac{2\pi \times 10^6 \cdot 4\pi \times 10^{-7}}{2 \cdot 2 \times 10^6}} = 1.405 \times 10^{-3}\,\Omega \qquad (2\text{-}180)$$

The ohmic resistance from (2-175) is

$$R_o = \frac{h}{2\pi a}\frac{R_s}{3} = \frac{0.787}{2\pi \cdot 1.5875 \times 10^{-3}}\frac{1.40 \times 10^{-3}}{3} = 0.0370\,\Omega \qquad (2\text{-}181)$$

The radiation efficiency from (2-178), (2-179), and (2-181) is

$$e_r = \frac{R_r}{R_r + R_o} = \frac{0.00271}{0.00271 + 0.0370} = 6.82\% \qquad (2\text{-}182)$$

The low efficiency in this broadcast reception application is overcome by using a high-power transmitter operating into a tall, efficient antenna. Thus, the cost and complexity of the system are concentrated into the transmitting station, allowing for inexpensive and simple receiving antennas. The gain is calculated from (2-155) as $G = e_r D = 0.0682(1.5) = 0.1023 = -9.9$ dB. So the gain drops from 1.76 to -9.9 dB when loss is included for a net loss of 11.7 dB, which is 10 log (0.0682). The input reactance is found using the above parameter values in (2-176), giving $X_A = -37,882\,\Omega$; also see Fig. 6-6. This is a very large capacitive reactance, leading to a severe impedance mismatch. The total input impedance is

$$Z_A = R_r + R_o + jX_A = 0.00271 - j37,882\,\Omega \qquad (2\text{-}183)$$

Full numerical simulation of the wire dipole model in this example using the method of moments discussed in Chap. 14 gives values very close to all of those found above.

In addition to reduction in radiation efficiency, antenna ohmic losses have another undesirable effect. As with any resistive element in an electrical system, ohmic losses on antennas are noise sources. For receiving applications above 1 MHz when the signal is low, internal noise can swamp out a signal. Receiving antennas also pick up external noise. For frequencies around 1 MHz and below, external noise, mainly due to lightning, is significant and always present. This external noise picked up by the antenna via the radiation resistance is usually larger than the internal noise and antenna loss can be tolerated. Antenna noise is discussed further in Sec. 4.3.

The low efficiency associated with electrically small antennas can be increased greatly by using superconducting material. However, practical superconducting antennas operating at room temperature remain an immature technology in most applications. Hansen [see H.8.11] discusses superconducting antennas in some depth.

There are several other efficiency factors in antenna theory and they are discussed in Sec. 9.3. The measurement of radiation efficiency is discussed in Sec. 13.3.4.

2.8 ANTENNA POLARIZATION

A monochromatic electromagnetic wave, which varies sinusoidally with time, is characterized at an observation point by its frequency, magnitude, phase, and polarization. The first three of these are familiar parameters, but polarization is often not well understood by students and practicing engineers. The polarization of an antenna is the polarization of the wave radiated in a given direction by the antenna when transmitting. In this section, we first discuss the possible polarizations of an electromagnetic wave, and then antenna polarization will follow directly from wave polarization. A complete discussion of wave and antenna polarization is found in [H.11.1: Stutzman].

The phase front (surface of constant phase) of a wave radiated by a finite-sized radiator becomes nearly planar over small observation regions. This wave is referred to as a *plane*
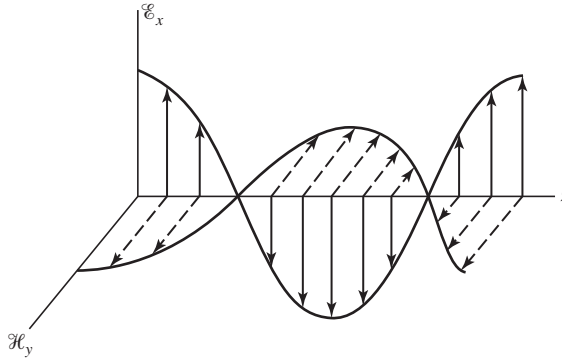
Figure 2-17 The spatial behavior of the electric (solid) and magnetic (dashed) fields of a linearly (vertical) polarized wave for a fixed instant of time. (Reproduced by permission from [H.11.1: Stutzman]. © 1993 by Artech House, Inc., Norwood, MA.)

wave and its electric and magnetic fields lie in a plane. The *polarization* of a plane wave is the figure the instantaneous electric field traces out with time at a fixed observation point. An example is the vertical, linearly polarized wave in Fig. 2-17, which shows the spatial variation of the electric field at a fixed instant of time. As time progresses, the electric field ($\mathscr{E}_x$) at a fixed point oscillates back and forth along a vertical line. For a completely polarized wave, the figure traced out is, in general, an ellipse. As indicated in Fig. 2-17, the temporal and spatial variations of the magnetic field are similar to those for the electric field, except that the magnetic field is perpendicular to the electric field. Waves can have a nonperiodic behavior, but we will not consider such randomly polarized wave components because antennas cannot generate them; noise is randomly polarized.

There are some important special cases of the polarization ellipse. If the electric field vector moves back and forth along a line, it is said to be *linearly polarized*; see Figs. 2-18*a* and 2-18*b*. An example is the electric field from an ideal dipole or any linear current. If the electric field vector remains constant in length but rotates around in a circular path, it is *circularly polarized*. Rotation at radian frequency ω is in one of two directions, referred to as the sense of rotation. If the wave is traveling toward the observer and the vector rotates clockwise, it is *left-hand polarized*. The left-hand rule applies here: With the thumb of the left hand in the direction of propagation, the fingers will curl in the direction of rotation of the instantaneous electric field $\mathscr{E}$. If it rotates counterclockwise, it is *right-hand polarized*. Left- and right-hand circularly polarized waves are shown in Figs. 2-18*c* and 2-18*d*. A helix antenna produces circularly polarized waves and the sense of rotation of the wave is the same as that of the helix windings; for example, a right-hand wound helix produces a right-hand circularly polarized wave; see Sec. 7.3. Finally, Figs. 2-18*e* and 2-18*f* show the most general cases of left-hand and right-hand sensed *elliptical polarizations*.

The time-space behavior of the important special case of circular polarization is difficult to visualize. Fig. 2-19 provides a space perspective view of a left-hand circularly polarized wave. As the vector pattern translates along the $+z$-axis, the electric field at a fixed point appears to rotate clockwise in the *xy*-plane (yielding a left-hand circularly polarized wave). This is illustrated with the time sequence of vectors in the fixed plane shown in Fig. 2-19.

A general polarization ellipse is shown in Fig. 2-20 with a reference axis system. The wave associated with this polarization ellipse is traveling in the $+z$-direction. The sense of rotation can be either left or right. The instantaneous electric field vector $\mathscr{E}$ has components $\mathscr{E}_x$ and $\mathscr{E}_y$ along the *x*- and *y*-axes. The peak values of these components are E_1 and E_2. The angle γ describes the relative values of E_1 and E_2 from

(*a*) Vertical linear polarization.

(*b*) Horizontal linear polarization.

(*c*) Left-hand circular polarization.

(*d*) Right-hand circular polarization.

(*e*) Left-hand elliptical polarization.

(*f*) Right-hand elliptical polarization.
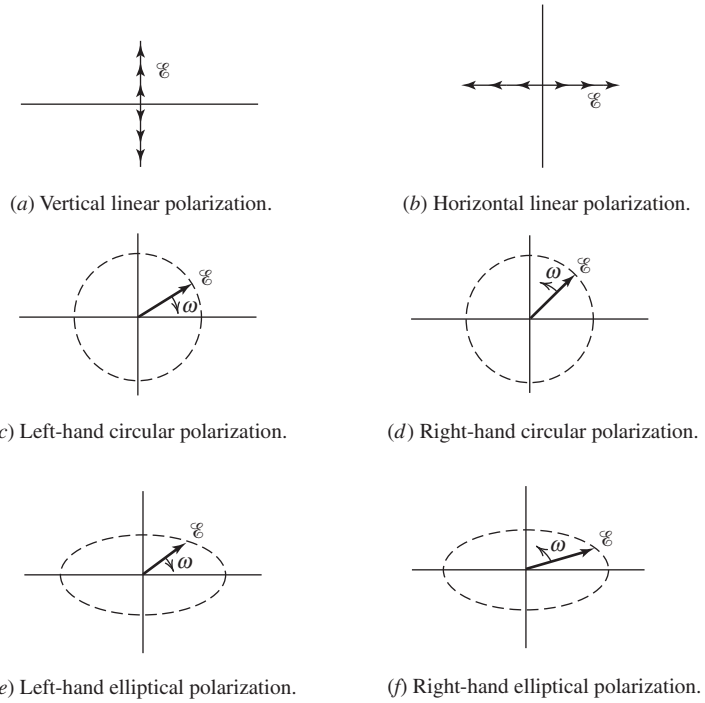
Figure 2-18 Some wave polarization states. The wave is approaching.
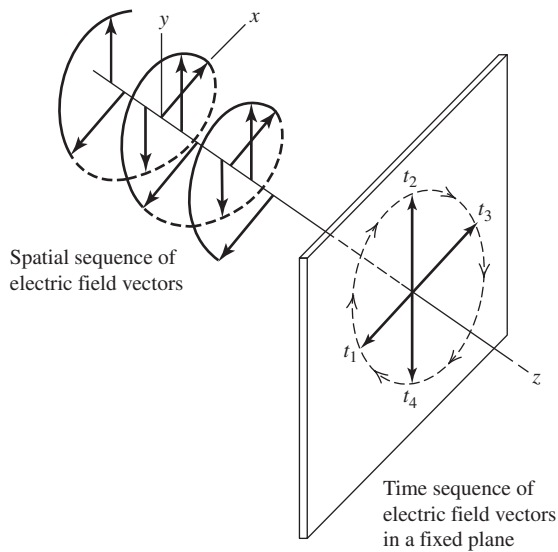
Figure 2-19 Perspective view of a left-hand circularly polarized wave shown at a fixed instant of time and the time sequence of electric field vectors as the wave passes through a fixed plane in the $+z$-direction. (Reproduced by permission from [H.11.1: Stutzman]. © 1993 by Artech House, Inc., Norwood, MA.)
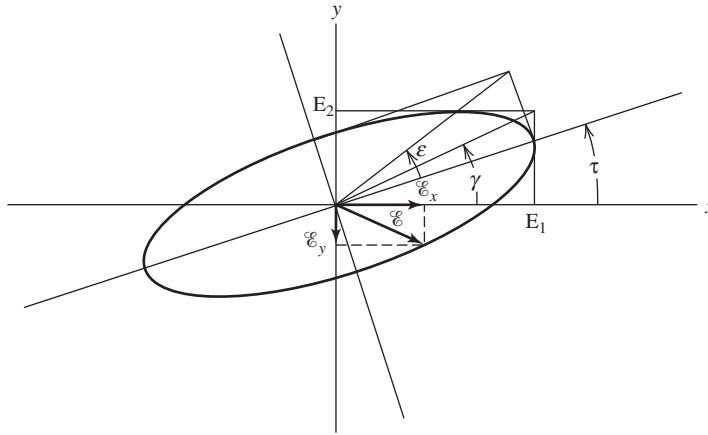
Figure 2-20 The general polarization ellipse. The wave direction is out of the page in the $+z$-direction. The tip of the instantaneous electric field vector $\mathscr{E}$ traces out the ellipse. (Reproduced by permission from [H.11.1: Stutzman]. © 1993 by Artech House, Inc., Norwood, MA.)

$$\gamma = \tan^{-1}\frac{E_2}{E_1}, \qquad 0° \leq \gamma \leq 90° \tag{2-184}$$

The tilt angle of the ellipse τ is the angle between the x-axis (horizontal) and the major axis of the ellipse. The angle ε is seen from Fig. 2-20 to be

$$\varepsilon = \cot^{-1}(-AR), \quad 1 \leq |AR| \leq \infty, \qquad -45° \leq \varepsilon \leq 45° \tag{2-185}$$

where the axial ratio of the ellipse $|AR|$ is the ratio of the major axis electric field component to that along the minor axis. A linearly polarized wave has $AR = \infty$ and (2-185) gives $\varepsilon = 0°$. The sign of AR is positive for right-hand sense and negative for left-hand sense. Axial ratio is often expressed in dB as $20 \log |AR|$.

The instantaneous electric field for the wave of Fig. 2-20 can be written as (with $z = 0$ for simplicity)

$$\mathscr{E} = \mathscr{E}_x\hat{\mathbf{x}} + \mathscr{E}_y\hat{\mathbf{y}} = E_1 \cos \omega t\, \hat{\mathbf{x}} + E_2 \cos (\omega t + \delta)\hat{\mathbf{y}} \tag{2-186}$$

where δ is the phase by which the y-component leads the x-component. This representation describes the ellipse shape as time t progresses. If the components are in-phase ($\delta = 0$), the net vector is linearly polarized. The orientation of the linear polarization depends on the relative values of E_1 and E_2, For example, if $E_1 = 0$, vertical linear polarization results; if $E_2 = 0$, horizontal linear results; if $E_1 = E_2$, the polarization is linear at 45° with respect to the axes. Linear polarization is a collapsed ellipse with infinite axial ratio. If δ is nonzero, the axial ratio is finite. When $\delta > 0$, $\mathscr{E}_y$ leads $\mathscr{E}_x$ in phase and the sense of rotation is left-hand. For $\delta < 0$, the sense is right-hand. If $E_1 = E_2$ (thus, $\gamma = 45°$) and $\delta = \pm 90°$, the polarization is circular ($+90°$ is left-hand and $-90°$ is right-hand). The axial ratio magnitude of a circularly polarized wave is unity.

The phasor form of (2-186) is

$$\mathbf{E} = E_1\hat{\mathbf{x}} + E_2 e^{j\delta}\hat{\mathbf{y}} \tag{2-187}$$

which can be written as (see Prob. 2.8-3)

$$\mathbf{E} = \sqrt{E_1^2 + E_2^2}(\cos\gamma\,\hat{\mathbf{x}} + \sin\gamma\, e^{j\delta}\hat{\mathbf{y}}) = |E|\hat{\mathbf{e}} \tag{2-188}$$

The factor $|E|$ is the field magnitude and $\hat{\mathbf{e}}$ is the complex vector representation for the field and is normalized to unity magnitude. Thus, γ and δ completely specify the polarization state of the wave. In fact, either pair of angles (ε, τ) or (γ, δ) uniquely define the polarization state of a wave. The transformations between these angles are

$$\gamma = \frac{1}{2}\cos^{-1}(\cos 2\varepsilon \ \cos 2\tau) \tag{2-189}$$

$$\delta = \tan^{-1}\left(\frac{\tan 2\varepsilon}{\sin 2\tau}\right) \tag{2-190}$$

As an example, a linearly polarized wave tilted $45°$ relative to the x-axis ($\tau = 45°$) has $E_1 = E_2 = 1/\sqrt{2}$ and $AR = \infty$. From (2-185), $\varepsilon = \cot^{-1}(\infty) = 0°$. Then (2-189) and (2-190) yield $\gamma = 45°$ and $\delta = 0°$, which are correct

The *polarization of an antenna* is the polarization of the wave radiated by the antenna when transmitting. Therefore, all of the above discussions on wave polarization apply directly to antenna polarization. The polarization of waves radiated by an antenna vary with angle around the antenna; if no angle is specified when the polarization of an antenna is stated, it is assumed to be in the direction of the beam peak. Usually the polarization characteristics of an antenna remain fairly constant over the main beam. However, radiation from side lobes can differ significantly in polarization from that of the main beam. When measuring polarization of an antenna over its pattern, the magnitude and relative phase of two orthogonal components are measured, often E_θ and E_ϕ. The principal plane patterns of a linearly polarized antenna, such as a dipole along the z-axis, are completely specified when a linearly polarized probe antenna is oriented to respond to the E_θ-component in the E- and H-planes. On the other hand, the principal plane patterns for a circularly polarized (CP) antenna are not the E- and H-planes because these planes are not uniquely defined in this case. Instead, the principal planes of a CP antenna are two orthogonal planes referenced to some stated and fixed physical direction, often denoted as the $\phi = 0°$ and $\phi = 90°$ planes.

Reciprocal antennas have identical radiation patterns on transmit and receive, and this extends to the vector nature of the radiation including polarization. Thus, a transmit antenna is polarization matched to a receive antenna if they have the same polarization states; that is, they have the same polarization ellipse axial ratio, sense, and major axis orientation. For example, a right-hand circularly polarized receiving antenna is polarization matched to a right circularly polarized wave. As a mechanical analogy, a right-hand threaded rod corresponds to a right-hand circularly polarized (RHCP) wave and a right-hand tapped hole represents a RHCP antenna. The rod and hole are matched when screwed either in or out, corresponding to reception or transmission. More will be said about the important topic of reception of polarized waves in Sec. 4.2.

Good examples of the application of antenna polarization are found in broadcast FM radio and TV. Originally FM and TV broadcast transmitting antennas in the U.S. were horizontally polarized. [H.6: *Ant. Eng. Hdbk.*, 3rd ed., p. 28-4] In the 1960s, the Federal Communications Commission (FCC) allowed FM broadcasters to transmit vertical as well as horizontal polarization (or CP), which improved reception in vehicles using a vertically polarized whip antenna. In practice, FM radio broadcasters commonly use distorted dipoles that produce elliptical polarization, which provides an increased vertical component but not one equal to the horizontal signal strength. In 1977, the FCC permitted TV broadcasters to use CP (right-hand) by allowing transmitters to maintain the maximum EIRP in horizontal polarization and add the same amount to vertical polarization. This makes the receivers less sensitive to the orientation of their linearly polarized antenna and to ghosts caused by reflected waves that arrive at the receiver with the opposite sense (i.e., left-handed).

Certain antennas are naturally circularly polarized, such as helices and spirals; see Chap. 7. Circular polarization can also be generated by feed network design. An example

is two perpendicular half-wave dipoles fed from a common transmission line such that the magnitudes are the same and with an extra quarter-wavelength of line (which gives a 90° phase shift) to one dipole as shown in Fig. 4-6*b*. This will produce circular polarization normal to the plane of the antennas with the sense determined by which leg has the delay line.

The principles and terminology introduced in this chapter will be applied to some important simple radiating systems in the next chapter.

REFERENCES

1. J. W. Arthur,' "The Fundamentals of Electromagnetic Theory Revisited," *IEEE Trans. on Ant. & Prop. Mag.*, Vol. 50, pp. 19–65, Feb. 2008.
2. J. Van Bladel, "Lorenz or Lorentz?," *IEEE Trans. on Ant. & Prop. Mag.*, Vol. 33, p. 69, April 1991.
3. S. Laybros and P. F. Combes, "On Radiating-Zone Boundaries of Short, $\lambda/2$, and λ Dipoles," *IEEE Trans. on Ant. & Prop. Mag.*, Vol. 46, pp. 53–64, Oct. 2004.

PROBLEMS

2.1-1 Use (2-6) in (2-1) to derive (2-7).

2.1-2 Use (2-13) in (2-19) together with (2-11), (2-12), and (2-20) to derive a form of (2-18) that includes loss through σ.

2.1-3 Assuming ε and μ are real and $\mathbf{M} = 0$, derive (2-28) through (2-34) using the identity (C-19).

2.1-4 Write the complex power equation for a series *RLC* network driven by a voltage generator in a form analogous to the Poynting theorem.

2.2-1 Derive (2-47) starting with (2-18).

2.2-2 Including time variation, the phase expression for a wave propagating in the *z*-direction is $\omega t - \beta z$. For a constant phase point on the wave, this expression is constant; take the time derivative to derive velocity expression in (2-53).

2.2-3 (a) Show that $\psi = Ce^{-j\beta r}/r$ satisfies (2-57) at all points except the origin. (b) By integrating (2-55) over a small volume containing the origin, substituting $\psi = Ce^{-j\beta r}/r$, and letting r approach zero, show that $C = (4\pi)^{-1}$, thus proving (2-58).

2.3-1 Show that (2-73b) follows from (2-72).

2.3-2 Derive the general electric field expression of (2-73b) valid in all regions in the following two ways. (a) Use the magnetic field expression (2-73a) in (2-63). (b) Use the vector potential expression (2-65) in (2-46).

2.3-3 For the ideal dipole of a *z*-directed current element $I\,\Delta z$ in free space and located at the origin of a spherical coordinate system: (a) Find the complex Poynting vector "expression" in the general case, where r can be in the near-field region. Use the fields of (2-73). (b) Then find the expression for the time-average power flowing out through a sphere of radius r enclosing the current element. Your answer will be that of (2-77). Why? (c) Use the results from (a) to derive (2-79).

2.3-4 Show that the electric field for the ideal dipole in (2-73b) satisfies Maxwell's equation $\nabla \cdot \mathbf{E} = 0$.

2.3-5 Prove that in the near field of an ideal dipole the ratio of the real to reactive power density magnitudes is $(\beta r)^3$ thereby verifying that the real and reactive power densities are equal for $\beta r = 1$, or $r = \lambda/2\pi$. Hint: Use (2-76) and (2-79).

2.4-1 Prove (2-95) by using (2-89) in (2-94) and retaining only $1/r$ terms; that is, using $\beta r \gg 1$.

2.4-2 Eq. (2-98) can be derived without initially assuming that the rays are parallel. Derive (2-98) by writing $R = [(\mathbf{r} - \mathbf{r}')\,(\mathbf{r} - \mathbf{r}')]^{1/2}$, expanding, factoring out an r, neglecting the smallest term, and using the first two terms of the binomial expansion.

2.4-3 By examining only the r variation in (2-73), show that the real and imaginary parts of each field component are equal for $\beta r = 1$. For the E_θ, first develop a reactive near field form, $\beta r \ll 1$, by neglecting the first term.

2.4-4 (a) Derive an expression for the wave impedance (E_θ/H_ϕ) for an ideal dipole. (b) Evaluate and plot this equation for βr from 0 to 20. Note the value for $\beta r = 10(r = 1.6\lambda)$ compared to the intrinsic impedance of free space.

2.4-5 Plot the error in E_θ magnitude for an ideal dipole in (2-73b) relative to the far-field value in (2-74a). What is the value for $r = 5\lambda$?

2.4-6 (a) Derive the reactive-radiating near field boundary criterion of $r = 0.62\sqrt{D^3/\lambda}$ for a length D. Do this by first deriving the value of θ to be 54.7° for which the fourth term of (2-86) is maximum; then find the r value of the maximum of the fourth term corresponding to a phase error of $\pi/8$ radians, (b) Find the line source length for which this boundary criterion equals that based on an ideal dipole, *i.e.,* $r = 0.16\lambda$.

2.4-7 Show that a line source of length D will have a maximum error $(R = r - D/2)$ in the distance to the far field relative to the distance from the center of the line source, r, of 10% when $r = 5D$ by using $R = 0.9r$.

2.4-8 Using the far-field boundary criterion of $r_{ff} = 2D^2/\lambda$ for a linear antenna of length D, find r_{ff} for the following antenna lengths: $D = 5\lambda$, a half-wave dipole $(D = \lambda/2)$, and a short dipole $(D = 0.01\lambda)$. Is the far-field distance you computed valid in each case; if not, why not?

2.4-9 Use the far-field distance criteria in (2-100) to make a single graph of r/λ (vertical axis) versus D/λ showing each far-field boundary criterion. Indicate the region of the graph corresponding to the far field.

2.4-10 Examine the far-field criteria for a 31-in. (0.787 m) long fender-mount car radio antenna at 1 MHz. Which criterion is the limiting condition? The graph in Prob. 2.4-9 will be a help.

2.5-1 Show that there is 4π sr in all space by integrating $d\Omega$ over a sphere.

2.5-2 A *power* pattern is given by $|\cos^n \theta|$ for $0 < \theta < \pi/2$ and is zero for $\pi/2 < \theta < \pi$. (a) Calculate the directivity for $n = 1, 2,$ and 3. (b) Find the HP values in degrees for each n. (c) Sketch the patterns for the n values on one polar plot and comment on the them. (d) Explain the directivity value for the case of $n = 0$.

2.5-3 An antenna has a far-field pattern which is independent of ϕ but which varies with θ as follows:

$$
\begin{aligned}
F &= 1 && \text{for} && 0° \le \theta \le 30° \\
F &= 0.5 && \text{for} && 60° \le \theta \le 120° \\
F &= 0.707 && \text{for} && 150° \le \theta \le 180° \\
F &= 0 && \text{for} && 30° < \theta < 60° && \text{and} && 120° < \theta < 150°
\end{aligned}
$$

Find the directivity. Also find the directivity in the direction $\theta = 90°$.

2.5-4 For a single-lobed pattern the beam solid angle is approximately given by

$$\Omega_A \approx HP_E HP_H$$

where HP_E and HP_H are the half-power beamwidths in radians of the main beam in the E and H planes. Show that

$$D \approx \frac{41,253}{HP_{E°} \cdot HP_{H°}}$$

where $HP_{E°}$ and $HP_{H°}$ are the E and H plane half-power beamwidths in degrees.

2.5-5 A horn antenna with low side lobes has half-power beamwidths of 29° in both principal planes. Use the approximate expression in Prob. 2.5-4 to compute the directivity of the horn in decibels.

2.5-6 A sector pattern has uniform radiation intensity over a specified angular region and is zero elsewhere. An example is

$$
F(\theta) = \begin{cases} 1 & \dfrac{\pi}{2} - \alpha < \theta < \dfrac{\pi}{2} + \alpha \\ 0 & \text{elsewhere} \end{cases}
$$

Derive an expression for the directivity corresponding to this pattern.

Evaluate this expression for the specific case in Example 2-3.

2.5-7 An airplane is flying parallel to the ground (in the z-direction). For a surface search radar an antenna is required that uniformly illuminates the ground over some region. The so-called cosecant pattern will do this. From the figure we see that $h = r \cos\left(\frac{\pi}{2} - \theta\right)$ or

$$r = \frac{h}{\sin \theta} = h \csc \theta.$$

This expresses how much farther the radiation must travel to reach the ground as θ is deceased. The radiation pattern

$$F(\theta) = \csc \theta,$$

will just compensate for the $1/r$ field variation with distance. If, in addition, the ϕ variation is a sector pattern of small angular extent ϕ_0, then

$$F(\theta, \phi) = \begin{cases} \csc \theta & \theta_1 < \theta < \dfrac{\pi}{2}, \, 0 < \phi < \phi_0 \\ 0 & \text{elsewhere} \end{cases}$$

Derive an expression for the directivity.
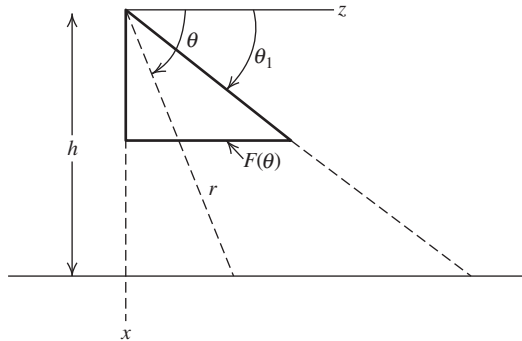
2.5-8 *Gaussian Pattern.* A circularly symmetric, narrow beam antenna pattern is frequently modeled by a Gaussian shape given by

$$F(\theta) = e^{-4 \ln\left(\sqrt{2}\right)(\theta / \text{HP})^2}$$

Derive expressions for the directivity associated with this pattern in terms of the half-power beamwidth HP in radians and in degrees. Do this by approximating $\sin \theta$ by θ in the integration for Ω_A and extending the integration limit to infinity.

2.5-9 An antenna has a directivity of 20 and a radiation efficiency of 90%. Compute the gain in dB.

2.5-10 Compute the gain of an antenna which has a radiation efficiency of 95% and the following radiation pattern:

$$F(\theta) = \begin{cases} 1 & 0 \le \theta < 20° \\ 0.707 & 20° \le \theta < 120° \\ 0 & 120° \le \theta < 180° \end{cases}$$

2.6-1 Use the ohmic resistance formula (2-173) to verify the expression for R_o for: (a) A uniform current distribution given by (2-170). (b) A triangular current distribution given by (2-175).

2.6-1 A tuned transmit antenna. (a) Using the equivalent circuit of Fig. 2-15 and assuming reactive tuning ($X_\text{L} = -X_\text{A}$), derive expressions for the power delivered to radiation, P, and to ohmic losses in R_g and R_o, P_g and P_o respectively. (b) If in addition to reactive tuning the generator and antenna resistances are matched, $R_g = R_r + R_o$, reduce the power expressions derived in (a) and determine what fraction of the power supplied is dissipated in the generator resistance.

2.7-1 A 2-m-long dipole made of 6.35-mm (0.25-in.) diameter aluminum is operated at 500 kHz. Compute its radiation efficiency, assuming: (a) the current is uniform; (b) the current is triangular.

2.7-2 A citizen's band radio at 27 MHz uses a half-wavelength long antenna that has a radiation resistance of 70 Ω. Compute the radiation efficiency if the antenna is made with 6.35-mm-diameter aluminum wire. As a rough approximation assume that the current is triangular.

2.7-3 A cordless telephone operating at 50 MHz has a 38-cm long monopole antenna made of 4-mm diameter aluminum tubing. Compute the radiation efficiency. Use (3-13) to find the monopole resistance values from the dipole values.

2.7-4 *Design of a short dipole antenna to achieve a specified efficiency.* (a) Derive an expression for the ratio of radiation resistance to ohmic resistance for a short dipole, $r = R_r/R_o$. (b) Validate the expression by calculating the efficiency for the case of Example 2-4. (c) Calculate the length of a short dipole corresponding to 90% radiation efficiency for the case No. 18 copper wire and operation at 100 MHz.

2.8-1 The instantaneous electric field components of an elliptically polarized wave are $\mathscr{E}_x = E_1 \cos(\omega t - \beta z)$ and $\mathscr{E}_y = E_2 \cos(\omega t - \beta z + \delta)$. Specify E_1, E_2, and δ for the following polarizations:

    (a) Linear with $E_1 \neq 0$ and $E_2 \neq 0$.
    (b) Right circular.
    (c) Left circular.

2.8-2 Write the frequency domain form of the total vector electric fields given in Prob. 2.8-1.

2.8-3 Start with (2-187) and prove (2-188). Use the fact that the magnitude of **E** follows from $|\mathbf{E}|^2 = \mathbf{E} \cdot \mathbf{E}^*$. Also note that γ in Fig. 2-20 is in a triangle with sides E_1 and E_2 and hypotenuse $|\mathbf{E}|$.

2.8-4 Prove that a RHCP wave normally incident on a plane perfect conductor changes to LHCP upon reflection.

2.8-5 *Feed network to generate CP.* (a) Use parallel wire transmission lines in a sketch of the wiring feeding two orthogonal dipoles, oriented along the *x*- and *y*-axes, to produce RHCP in the +*z*-direction. Clearly show the extra quarter-wavelength of transmission line. (b) What is the polarization of radiation in the −*z*-direction?

Chapter 3

Simple Radiating Systems

The previous chapter introduced the mathematical principles of antennas and illustrated them with a few elementary radiators. This chapter extends and uses the principles to treat some popular simple antenna elements and arrays. We leave the many variations and the derivational details to subsequent chapters when we revisit each radiating system. Specifically, we discuss electrically small dipoles, half-wave dipoles, monopoles, electrically small loops, and two-element arrays. Not only are these antennas fundamental to antenna practice, but they are the basic building units found in most antenna systems. In addition, complicated antenna systems are understood by examining their component parts, which are usually one or more of these canonical simple antennas. Image theory is also presented in this chapter for use in examining antennas operated in the presence of a ground plane.

3.1 ELECTRICALLY SMALL DIPOLES

An antenna whose dimensions are small compared to the wavelength at the frequency of operation is an electrically small antenna, which was first mentioned in Sec. 1.4. How much smaller than a wavelength an electrically small antenna must be depends on the application, but generally is taken to be on the order of a tenth of a wavelength in extent or less; more will be said on this topic in Sec. 11.5. Electrical size and physical size can be quite different. An antenna operating at low frequencies can be physically large but electrically small—i.e., a small fraction of a wavelength in extent. This is especially true for frequencies in the low MHz range and below. We saw in Sec. 2.7 that electrically small antennas are inherently inefficient.

However, this often is not a serious problem in receiving systems and physically small antennas offer advantages in size, weight, cost, and mobility. In this section, we revisit two electrically small antennas introduced in Secs. 2.3 and 2.7, the ideal dipole and the short dipole. We also discuss practical forms of electrically small dipoles.

The simplest practical electrically small antenna is the short dipole shown in Fig. 3-1a as a wire with a feed point in the center. The resemblance of the arms of the dipole to the feelers of an insect has been attributed to the origin of the term antenna [H.3: Schelkunoff and Friis, p. 5]. The current distribution of the short dipole is nearly triangular in shape as modeled in Figs. 2-16b and 3-1c. This is because the current distribution on thin wire antennas (diameter $\ll \lambda$) is approximately sinusoidal and also must be zero at the wire ends. Because the arms of the short dipole are a fraction of a wavelength long, only a small portion of the sine wave current appears on the arm and is therefore nearly linear; see Fig. 3-1c.

The decreasing current toward the wire ends requires that charges peel off and appear on the wire surface as shown in Fig. 3-1a. The current and charge distributions shown
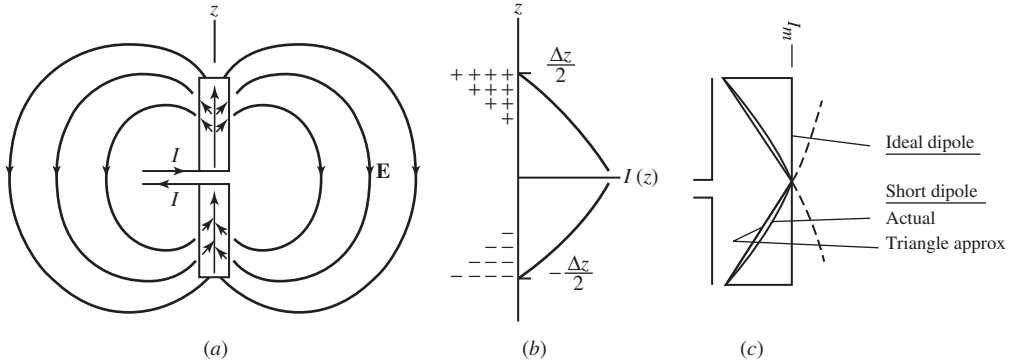
Figure 3-1 Short dipole, $\Delta z \ll \lambda$. (a) Current on the antenna and the electric fields surrounding it. (b) Current and charge distributions. (c) Short dipole current distribution and its triangular approximation compared to the ideal dipole current distribution.

in Fig. 3-1b are for an instant of time when the input current at the dipole terminals is maximum. Because the input current is changing sinusoidally with time, the current and charge distributions on the dipole do also. This charge accumulation leads to a displacement current density $j\omega\varepsilon\mathbf{E}$ in the space surrounding the dipole. The displacement current density, in turn, gives rise to an electromagnetic wave that propagates outward from the source as illustrated in Fig. 1-5. Displacement current in space couples a transmitting antenna to a receiving antenna, much as a conduction current provides coupling between lumped elements in a circuit. The radiation pattern of all forms of the electrically small dipole (with its radiating portion along the z-axis) is $\sin\theta$ as shown in Fig. 2-4.

In the ideal dipole, all charge accumulates at the ends of the antenna. In fact, the ideal dipole can be analyzed as either a uniform current or two point charges oscillating at radian frequency ω (see Prob. 3.1-1) as shown in Fig. 3-2. The charge dipole model shows that charge accumulates at the ends of the antenna, leading to a higher radiation resistance. In fact, the ideal dipole radiation resistance of (2-169) is four times that of the short dipole given by (2-172).

The input reactance of a short dipole is capacitive. This can be seen by visualizing the antenna as an open-circuited transmission line as in Fig. 1-6. When the distance from the end of the antenna to the feed point is much less than a quarter wavelength, the input impedance is capacitive, because from transmission line theory the impedance a distance s from an open-circuit termination is $-jZ_o\cos(\beta s)$. Simple transmission line theory only gives qualitative results when radiation is present. An approximate expression for the capacitive reactance of the short dipole is given in (2-176). Moment method computation techniques are used for accurate impedance evaluation; see Chap. 14. Loading coils are frequently used to tune out this capacitance.

Figure 3-2 Ideal dipole models. (a) Uniform current model. (b) Charge dipole model with $I = j\omega q$.
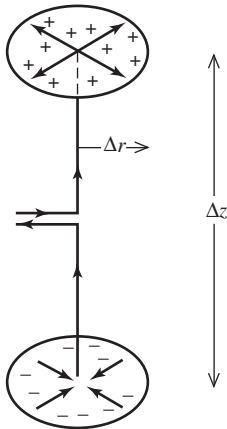
Figure 3-3 Capacitor-plate antenna. The arrows on the antenna indicate current. The charges on the plates are also shown.

The larger radiation resistance associated with the uniform current of the ideal dipole can be realized in practice by providing a mechanism for charge accumulation at the wire ends. One method of accomplishing this is to place metal plates at the ends of the wire. This is called a **capacitor-plate antenna**, or top-hat-loaded dipole antenna. Fig. 3-3 shows the construction of the antenna and the current and charges on it. If $\Delta z \ll \lambda$, the radial currents on the plates produce fields that almost cancel in the far field, since the currents are opposite-directed and the phase difference due to separation is small ($\beta \Delta z \ll 2\pi$). If, in addition, $\Delta z \ll \Delta r$, the plates provide for charge storage such that the current on the wire is constant. The capacitor-plate antenna then closely approximates the uniform current ideal dipole model. Frequently in practice, radial wires are used for the top loading in place of the solid plates.

Another small antenna used to approximate the ideal dipole is the **transmission line loaded antenna** as shown in Fig. 3-4a. The results of transmission line theory can be borrowed to determine the current distribution. The current is essentially sinusoidal along the wire with a zero at the ends. This current distribution is sketched in Fig. 3-4b for $L < \lambda/4$. If $\Delta z \ll \lambda$, the fields from the currents on the horizontal wires essentially cancel in the far field. If also $\Delta z \ll L$, the horizontal wires provide an effective place for the charge to be stored and the current on the vertical section is nearly constant, as illustrated in Fig. 3-4b. Then radiation comes from a short section over which the current is nearly constant and the antenna approximates an ideal dipole.

The monopole form of the transmission line loaded dipole shown in Fig. 3-4c is called the **inverted-L antenna** (ILA). The small ILA has a total wire length that is much less than a wavelength. The inverted-L antenna is popular for narrowband applications. The ILA is a popular antenna for garage door transmitter units (see Example 11-3). The reactance of the ILA antenna can be reduced by adding more horizontal wires. For example, an additional horizontal wire added to the left side in Fig. 3-4a will have the reactance by the paralleling of identical capacitive elements. As more horizontal wires are added, the reactance is further reduced and the structure approaches that of the capacitor plate antenna of Fig. 3-3. Evolutionary forms of the inverted-L antenna have larger bandwidth and have found widespread use in small handheld radio units where the wire length is near a quarter wavelength; these antennas will be discussed in Sec. 11.6.3.

At different portions of the frequency spectrum, electrically small antennas are used for different reasons. For instance, in the VLF region where wavelength is very large, an electrically short vertical radiator is used with a large top-hat load. The top-hat loading makes the antenna appear like the capacitor-plate antenna of Fig. 3-3. Further up the spectrum, such as in the AM broadcast band, receiving antennas are usually small

(*a*) Transmission line loaded dipole.

(*b*) Current on the transmission line loaded dipole with the wire folded out. The dashed line indicates current on the horizontal section.

(*c*) The inverted-L antenna (ILA).
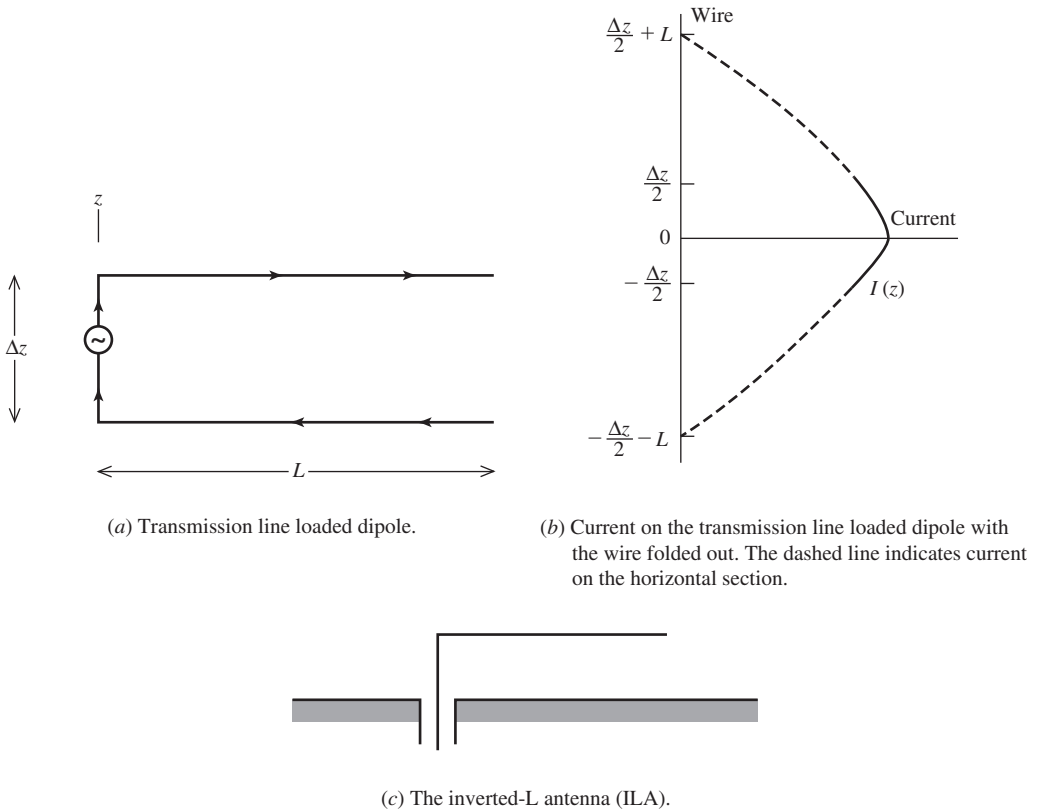
Figure 3-4 Transmission line loaded antennas.

electrically, as we saw in Sec. 2.7. AM transmitting antennas are not small, but are of resonant size as discussed in the next section. At VHF frequencies and above, electrically small antennas are only used in special situations.

One of the lowest-frequency communication systems ever implemented was the U.S. Navy's *Project ELF* for communicating with submerged submarines. Two transmitter sites, which operated in Wisconsin and Michigan from 1985 to 2004, had multiple conductors in lengths of from 22.5 to 45 km (14 to 28 mi) supported by power line poles. The ELF frequency of 76 Hz was selected to avoid the high loss of seawater, which increases rapidly with frequency. The purpose of this narrowband communication system was only to transmit a simple coded signal with a message instructing submarines to surface and use another communication system to conduct full two-way communications in air. The very long free space wavelength of 3950 km (2450 mi) at this frequency means that the 45-km conductor length corresponds to only 0.01 wavelength, leading to a very inefficient system.

3.2 HALF-WAVE DIPOLES

One of the most popular antennas, both as a modeling building block and in direct application, is the **half-wave dipole**. The half-wave dipole is realized with a straight wire fed in the center. The simple and accurate model is that of a linear current whose amplitude distribution is sinusoidal with a maximum at the center as shown in Fig. 3-5*a*. The half-wave dipole is in widespread use because of its simple physical structure and, as
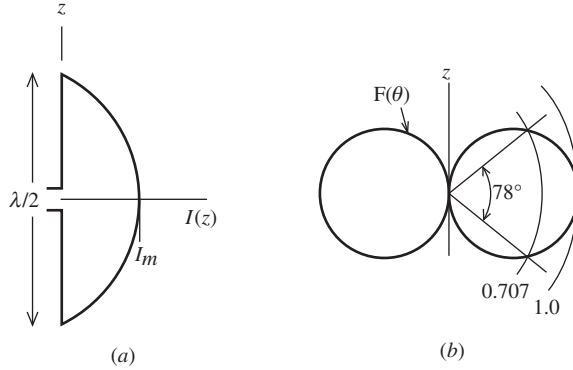
Figure 3-5 The half-wave dipole. (*a*) Current distribution $I(z)$. (*b*) Radiation pattern $F(\theta)$.

a resonant antenna, it has zero input reactance as well as a desirable input resistance. The input impedance of an infinitely thin dipole of exactly one-half wavelength is $Z_A = 73 + j42.5\,\Omega$. The input impedance of wire dipoles is discussed in detail in Sec. 6.1, but for now it is sufficient to point out that tuning of a wire half-wave dipole antenna is accomplished by trimming the wire ends until a zero input reactance is achieved, with more shortening required for thicker wire diameter; see Prob. 3.2-7 and data in Sec. 14.5. The tuned half-wave dipole then has an input impedance of about $70 + j0\,\Omega$. Thus, the half-wave dipole has the advantages of being naturally well-matched in impedance to common devices. Additionally, it has a radiation efficiency very close to 100% for common wire materials.

The evolution of the half-wave dipole antenna from an open-circuited parallel wire transmission line by bending the two conductors outward was developed in Sec. 1.3 and Fig. 1-6. Essentially the bent ends allow the internal fields of the transmission line to be exposed to space and radiate. The sinusoidal current distribution on the transmission line is assumed to also exist on the dipole. More sophisticated analytical and numerical analysis methods (see Chap. 14 for the latter) reveal that wire antennas have a current distribution very close to sinusoidal in shape and that differences in pattern calculations based on a pure sinusoidal current are negligible, especially for a half-wave wire dipole with a diameter small compared to the operating wavelength. The sinusoidal current distribution for the half-wave dipole along and centered on the z-axis as shown in Fig. 3-5a is given by

$$I(z) = I_m \sin\left[\beta\left(\frac{\lambda}{4} - |z|\right)\right], \quad |z| \le \frac{\lambda}{4} \tag{3-1}$$

where $\beta = 2\pi/\lambda$. This equation yields the required zero current condition at the ends of the dipole where $z = \pm\lambda/4$ and a maximum value I_m at the center $(z = 0)$, as shown in Fig. 3-5a. The radiated electric field pattern is found using the steps for a z-directed line source in (2-103) to (2-106), leading to

$$E_\theta = j\omega\mu\frac{2I_m}{\beta}\frac{e^{-j\beta r}}{4\pi r}\sin\theta\frac{\cos[(\pi/2)\cos\theta]}{\sin^2\theta} \tag{3-2}$$

In this expression, we identify the element factor of $g(\theta) = \sin\theta$ and the normalized pattern factor:

$$f(\theta) = \frac{\cos[(\pi/2)\cos\theta]}{\sin^2\theta} \tag{3-3}$$

Both $g(\theta)$ and $f(\theta)$ have a maximum of unity for $\theta = \pi/2$. The complete far-field pattern, normalized to unity maximum (see (2-115)) is

$$F(\theta) = g(\theta)f(\theta) = \frac{\cos[(\pi/2)\cos\theta]}{\sin\theta} \qquad \textit{half-wave dipole} \qquad (3\text{-}4)$$

The E-plane pattern is plotted in Fig. 3-5b in linear, polar form. The radiation is uniform (in the H-plane) around the dipole because the pattern function in (3-4) is not a function of ϕ; this is expected because of the symmetry of the dipole structure. The three-dimensional pattern is donut shaped similar to the short dipole. We return to the topic of dipole pattern evaluation in Sec. 6.1.

So far we have introduced three dipole antennas: the ideal, short, and half-wave dipoles. The characteristics and performance of these dipoles are listed in Table 3-2 in Sec. 3-4. The ideal and short dipoles, with uniform and triangular current distributions, respectively, have identical patterns. Both have a half-power beamwidth of $90°$ and a directivity of 1.5; see Fig. 2-4. The half-wave dipole has a narrower beamwidth of $78°$ and, thus, a higher directivity value of 1.64 or 2.15 dB, which will be derived in Sec. 6.1.

The length of the half-wave dipole is large enough electrically that radiation contributions from current elements along its extent will have significant phase variation in the far field as a function of angle around the antenna due to the difference in path length from the current elements to the far field. In the broadside direction ($\theta = \pi/2$) all path lengths are equal and there is full phasor addition and a pattern maximum. For other directions, there is partial cancellation due to phase differences between current elements that are as large as $180°$ from the dipole ends in the endfire directions ($\theta = 0, \pi$). The major influence on the endfire direction is the element factor $f(\theta) = \sin\theta$, which is zero caused by the vanishing projected area of the antenna in end view. As the dipole length is increased beyond a wavelength, the current element contributions in the far field are increasingly out of phase, leading to partial or total cancellation in the far field in the broadside direction and reinforcement in off-broadside directions, forming multiple lobes in the E-plane; see Figs. 6-3 and 6-4. This explains why practical center-fed wire antennas are less than a wavelength long.

3.3   MONOPOLES AND IMAGE THEORY

So far we have assumed that antennas have been isolated and in a free space environment. However, the environment surrounding an antenna will affect its performance, especially pattern and impedance. Environmental effects are small for a high-gain antenna that is elevated above the terrain. But effects may be large on a broad-beam antenna with objects nearby. The most common object near an antenna is a *ground plane*. The ideal form of a ground plane is the *perfect ground plane* which is planar, infinite in extent, and perfectly conducting. A real ground plane that is flat compared to a wavelength (i.e., surface deviations $\ll \lambda$) introduces small deviations from the planar assumptions. The perfectly conducting assumption is mild, and any good conductor such as aluminum or copper is very accurately modeled as a perfect conductor. The infinite extent assumption, however, can lead to errors. But, in most cases a perfect ground plane is well approximated by a solid metal plate or a planar wire grid system that is large compared to the size of the antenna. In this section, we introduce image theory as a useful tool to model an antenna operating in the presence of a perfect ground plane and then image theory is used to analyze monopole antennas. Antennas operating in the presence of imperfect ground planes are treated at appropriate points in this book: wire antennas close to finite ground planes are addressed in Sec. 6.6; the real earth as a ground plane is discussed in Sec. 6.7; and antennas operating in the presence of the human body are discussed in Sec. 11.9. Antennas in the presence of imperfect ground planes can be accurately

(*a*) Physical model.   (*b*) Equivalent model using image theory.
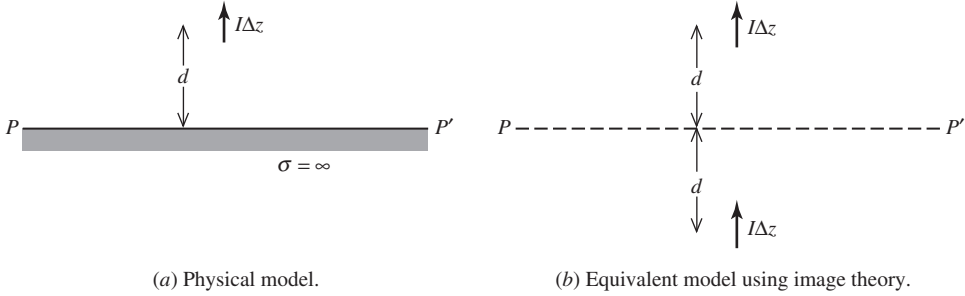
Figure 3-6 Ideal dipole above and perpendicular to a perfectly conducting ground plane.

evaluated using simulation tools based on Moment Method or Geometrical Theory of Diffraction techniques, which are discussed in Chaps. 14 and 16, respectively.

3.3.1   Image Theory

An antenna operating in the presence of a perfect ground plane produces two rays at each observation angle, a direct ray from the antenna and a second ray due to reflection from the ground plane such that Snell's law of reflection is satisfied. This is the approach used in Secs. 6.6 and 6.7 to analyze antennas above perfect and imperfect ground planes. Here, we develop the solution from first principles and it will be seen that the image antenna acts as an equivalent source for the reflected ray.

Consider first an ideal dipole near a perfect ground plane and oriented perpendicular to the ground plane as shown in Fig. 3-6*a*. Ground planes are usually horizontal, so this situation is referred to as a vertical ideal dipole above a perfect ground plane. We wish to find the fields **E** and **H** above the plane PP'. The uniqueness of the solution to a differential equation (the wave equation) plus its boundary conditions permits introduction of an equivalent system that is different below PP' but satisfies the same boundary conditions on PP' and has the same sources above PP'. Such an equivalent system, which produces the same fields above PP' as the original system, has an image source the same distance below the plane PP' and similarly directed. In this case, the image source is the virtual ideal dipole as shown in Fig. 3-6*b*.

It is a simple matter to prove that the boundary condition of zero tangential electric field along plane PP' is satisfied by this source configuration. To do this, we examine the electric field expression for an ideal dipole given by (2-73*b*). The complete expression must be used because the ground plane can be, and usually is, in the near field of the antenna. The radial component varies as $\cos\theta$ and the θ-component varies as $\sin\theta$, where θ is the angle from the axis along the direction of the current element. Let θ_1 and θ_2 be the angles from the line of the current elements to a point on the plane PP' for the primary source and its image, respectively. The radial components from the sources are then

$$E_{r1} = C\cos\theta_1 \tag{3-5}$$

$$E_{r2} = C\cos\theta_2 \tag{3-6}$$

The constant C is the same for each field component since the amplitude of the sources is the same and points on the boundary are equidistant from the current elements. From Fig. 3-7*a* we see that

$$\theta_1 + \theta_2 = 180° \tag{3-7}$$

(a) Radial components.

(b) Theta components.
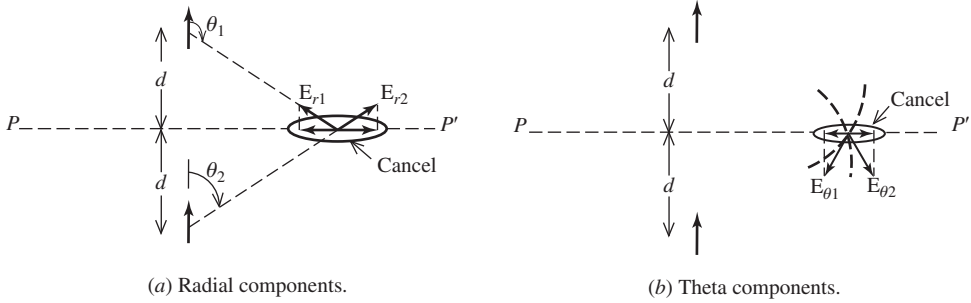
Figure 3-7 The ideal dipole and its image in a ground plane of Fig. 3-6. The source and its image acting together give zero tangential electric field intensity along the plane PP' where the original perfect ground plane was located.

so,

$$E_{r1} = C\cos(180° - \theta_2) = -C\cos\theta_2 \qquad (3\text{-}8)$$

Comparing this to (3-6), we see that

$$E_{r1} = -E_{r2} \quad along\ boundary \qquad (3\text{-}9)$$

Thus along the plane PP', the radial components are equal in magnitude and opposite in phase. E_{r2} is directed radially out from the image source since θ_2 is less than $90°$, and then $\cos\theta_2$ is positive. On the other hand, E_{r1} is radially inward toward the primary source since (3-8) is negative. Fig. 3-7a illustrates this and shows that the projections of each along PP' will cancel. A similar line of reasoning for the θ-components leads to

$$E_{\theta1} = D\sin\theta_1 = D\sin\theta_2 \qquad (3\text{-}10)$$

$$E_{\theta2} = D\sin\theta_2 \qquad (3\text{-}11)$$

where D is a constant and thus

$$E_{\theta1} = E_{\theta2} \quad along\ boundary \qquad (3\text{-}12)$$

Fig. 3-7b demonstrates that the net projection of these θ-components along plane PP' is zero.

   We have shown that the total tangential electric field intensity is zero along the image plane PP' for an ideal dipole perpendicular to the plane and its image was acting together. Therefore, since the source configuration above the plane and the boundary conditions were not altered, the system of Fig. 3-6b is equivalent to the original problem of Fig. 3-6a. The systems are equivalent in the sense that the fields above the plane PP' are identical. The above derivation can be reversed by starting with the two sources of Fig. 3-6b and then introducing a perfect ground plane with its surface along plane PP', thus arriving at Fig. 3-6a. The essential feature to remember is that *the fields above a perfect ground plane from a primary source acting in the presence of the perfect ground plane are found by summing the contributions of the primary source and its image, each acting in free space.*

   An ideal dipole oriented parallel to a perfect ground plane (i.e., horizontal) has an image that again is equidistant below the image plane, but in this case the image is oppositely directed as shown in Fig. 3-8. The equivalent model of Fig. 3-8b, which gives the same fields above plane PP' as the physical model of Fig. 3-8a, can be proved by simple sketches similar to those of Fig. 3-7.

(*a*) Physical model.   (*b*) Equivalent model using image theory.
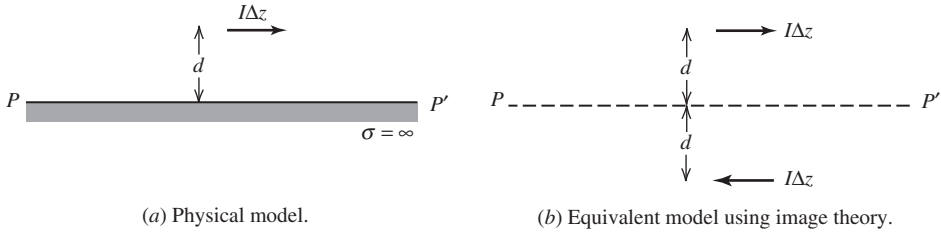
Figure 3-8 Ideal dipole above and parallel to a perfect ground plane.

The image of a current element oriented in any direction with respect to a perfect ground plane can be found by decomposing it into perpendicular and parallel components, forming the images of the components, and constructing the image from these image components. An example is shown in Fig. 3-9. The image of an arbitrary current distribution is obtained in a similar fashion. The current is decomposed into perpendicular and parallel current elements whose images are readily found. The image current distribution is then the vector sum of these image current elements.

(*a*) Physical model.   (*b*) Equivalent model using image theory.
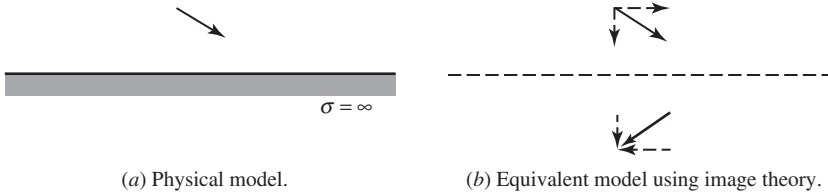
Figure 3-9 Ideal dipole above and obliquely oriented relative to a perfect ground plane.

3.3.2   Monopoles

The principles of image theory are illustrated in this section with several forms of the monopole antenna. A **monopole** is a dipole that has been divided in half at its center feed point and fed against a ground plane. Three monopoles and their images in a perfect ground plane are shown in Fig. 3-10. High-frequency monopoles are often fed from coaxial cables behind the ground plane as shown in Fig. 3-11*a*.

The currents and charges on a monopole are the same as on the upper half of its dipole counterpart, but the terminal voltage is only half that of the dipole. The voltage is half because the gap width of the input terminals is half that of the dipole, and the same electric field over half the distance gives half the voltage. The input impedance for a monopole is therefore half that of its dipole counterpart, or

$$Z_{A,\text{mono}} = \frac{V_{A,\text{mono}}}{I_{A,\text{mono}}} = \frac{\frac{1}{2}V_{A,\text{dipole}}}{I_{A,\text{dipole}}} = \frac{1}{2}Z_{A,\text{diople}} \tag{3-13}$$

This is easily demonstrated for the radiation resistance. Because the fields only extend over a hemisphere, the power radiated is only half that of a dipole with the same current. Therefore, the radiation resistance of a monopole is given by

$$R_{r,\text{mono}} = \frac{P_{\text{mono}}}{\frac{1}{2}|I_{A,\text{mono}}|^2} = \frac{\frac{1}{2}P_{\text{dipole}}}{\frac{1}{2}|I_{A,\text{dipole}}|^2} = \frac{1}{2}R_{r,\text{dipole}} \tag{3-14}$$

For example, the radiation resistance of a short monopole is from (2-172)

(*a*) Monopole antenna.

(*b*) Capacitor plate monopole.

(*c*) Transmission line monopole (inverted-L antenna).
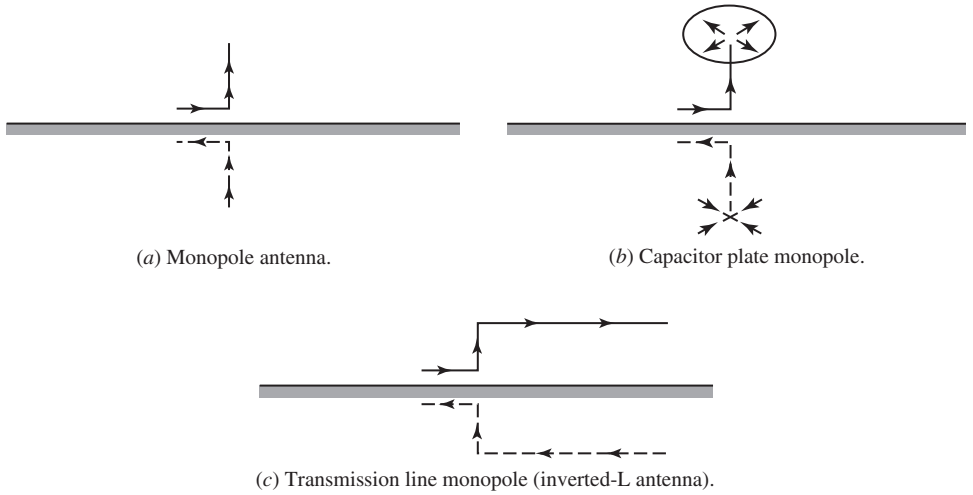
Figure 3-10 Monopole antennas over perfect ground planes with their images (dashed).

(*a*) Monopole fed against a large
solid ground plane.

(*b*) Equivalent dipole model
shown with currents.

Short monopole current

Quarter-wave
monopole current

Image currents

(*c*) Practical monopole antenna with radial wires to simulate a ground plane.
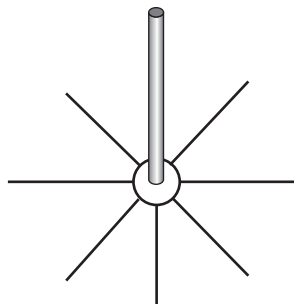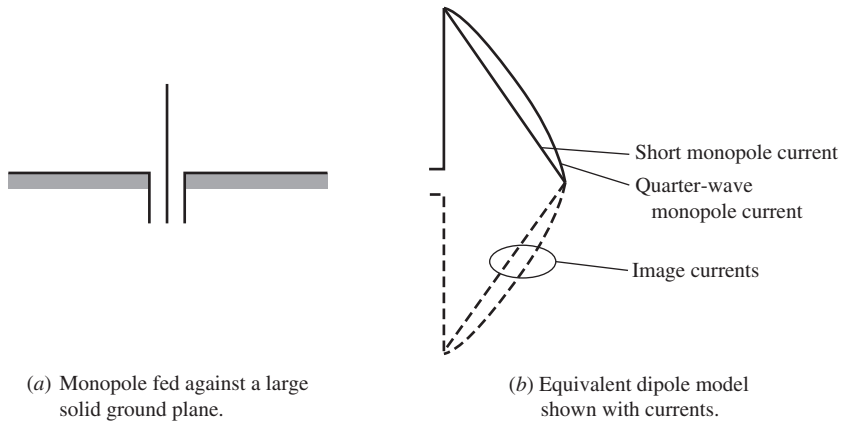
Figure 3-11 Monopole antennas fed against a ground plane with a coaxial cable.

$$R_{r,\text{mono}} = 40\pi^2 \left(\frac{h}{\lambda}\right)^2 \quad \text{for} \quad h \ll \lambda \tag{3-15}$$

where h is the length of the monopole and $\Delta z = 2h$.

The radiation pattern of a monopole above a perfect ground plane, as in Fig. 3-11, is the same as that of a dipole similarly positioned in free space since the fields above the image plane are the same. Therefore, a monopole fed against a perfect ground plane radiates one-half the total power of a similar dipole in free space because the power is distributed in the same fashion but only over half as much space. As a result, the beam solid angle of a monopole above a perfect ground plane is one-half that of a similar dipole in free space, leading to a doubling of the directivity:

$$D_{\text{mono}} = \frac{4\pi}{\Omega_{A,\text{mono}}} = \frac{4\pi}{\frac{1}{2}\Omega_{A,\text{dipole}}} = 2D_{\text{dipole}} \tag{3-16}$$

This can be shown in another way. If a dipole in free space has a maximum radiation intensity of U_m, a monopole of half the length above a perfect ground plane with the same current will have same value of U_m because the fields are the same. The total radiated power for the dipole is P, so the power radiated from the monopole is $\frac{1}{2}P$. The directivity from (2-140) for the two antennas is

$$D_{\text{mono}} = \frac{U_m}{\frac{1}{2}P/4\pi} = 2D_{\text{dipole}} \tag{3-17}$$

The directivity increase does not come from an increase in the radiation intensity (and, hence, field intensity) but rather from a decrease in average radiation intensity. This, in turn, comes about because only half the power radiated by a dipole is radiated by a monopole. The directivity of a short monopole, for example, is $2(1.5) = 3$.

The directivity of a quarter-wave monopole is twice that of a half-wave dipole in free space; that is, from Table 3-2 and (3-17)

$$D = 2(1.64) = 3.28 = 5.16\,\text{dB} \quad \lambda/4\text{-monopole} \tag{3-18}$$

The input impedance of an infinitesimally thin quarter-wave monopole from Table 3-2 and (3-13) is

$$Z_A = \frac{1}{2}(72 + j42.5) = 36 + j21.3\,\Omega \quad \lambda/4\text{-monopole} \tag{3-19}$$

In frequency bands below VHF, a quarter-wave monopole will be physically large. For example, in the standard AM radio broadcast band, a quarter-wavelength vertical monopole antenna will be 75 m (246 ft) tall at 1 MHz. Such large structures are usually not self-supporting. Guy wires are added for mechanical support with insulators spaced along the guy wires as shown in Fig. 3-12*a* to break up currents that would run onto the guy wires from the tower. Tower height can be reduced below a quarter wavelength by using the guy wires for top loading as shown in Fig. 3-12*b*. The currents on this *umbrella-loaded monopole* antenna continue from the top of the tower onto the guys, leading to a more uniform current on the tower and thus increasing the radiation resistance. However, the currents on the guys introduce a partial cancelling effect on the tower current. To provide adequate structural support for wind loading, the guy wires are spaced 120° around the tower top, as well as at one or two more levels for tall towers; see [H.6: *Ant. Eng. Hdbk.*, 4th ed., Chap. 27] for more details on tower antennas. As noted in Sec. 2.6, the radiation resistance is proportional to the area under the current versus

(a) Monopole with insulators to
reduce currents in guys.

(b) Umbrella-loaded monopole
with a top loading effect.
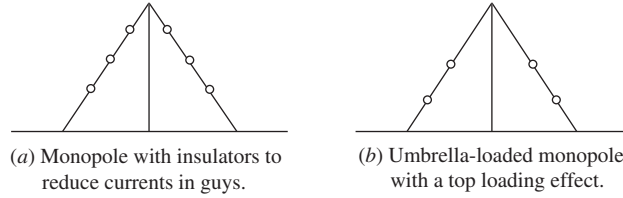
Figure 3-12 Monopoles with supporting guy wires.

distance curve. Using this fact and assuming, as usual, that the current distribution is sinu-soidal shaped, the modified radiation resistance formula for a top-loaded monopole is [1]

$$R_r = 40\pi^2 \left(\frac{h}{\lambda}\right)^2 (1 + \cos \beta h)^2 \quad \textit{top-loaded monopole} \tag{3-20}$$

For a quarter-wave monopole $\beta h = \pi/2$ and $\cos \beta h = 0$ and this expression reduces to the expected result of (3-15), which holds for a short monopole less than 0.15λ tall and with zero current on top (i.e., a linear current distribution). Measured data for umbrella-loaded monopoles are presented in [2].

3.4 SMALL LOOP ANTENNAS AND DUALITY

A closed loop current whose maximum dimension is less than about a tenth of a wave-length is called a **small loop antenna**. Again, small is interpreted as meaning, electrically small, or small compared to a wavelength. In this section, we use two methods to solve for the radiation properties of small loop antennas. First, we show that the small loop is the dual of an ideal dipole, and by observing the duality contained in Maxwell's equa-tions, we use the results previously derived for the ideal dipole to write the fields of a small loop. Next, we derive the fields of a small loop directly and show that the results are the same as those obtained using duality.

3.4.1 Duality

Frequently, an antenna problem arises for which the structure is the dual of an antenna whose solution is known. If antenna structures are duals, it is possible to write the fields for one antenna from the field expressions of the other by interchanging parameters using the principle of duality. Before examining the small loop, we discuss the general principle of duality as applied to antennas.

Dual antenna structures are similar to dual networks. For example, consider a simple network of a voltage source applied to a series connection of a resistor R and an inductor L as in Fig. 3-13a. The dual network of Fig. 3-13b is a current source $I (=) V$ applied to the parallel combination of conductance $G (=) R$ and capacitance $C (=) L$. The symbol "$(=)$" means replace the quantity on the left with the quantity on the right, much as the equal sign in a computer program statement. Since the networks are duals, the solutions are duals. In this example, the original series network can be described by the mesh equation

$$V = RI + j\omega LI \tag{3-21}$$

The dual of this mesh equation is a node equation obtained by replacing V by I, R by G, and L by C. The node equation for the dual parallel network is then

(a) Original network. (b) Dual network.
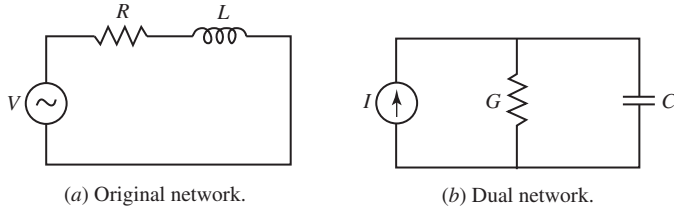
Figure 3-13 Dual networks: $I(=)V, G(=)R, C(=)L$.

$$I = GV + j\omega CV \tag{3-22}$$

Returning to the antenna problem, suppose we have an electric current source with current density $\mathbf{J}_1$ and boundary conditions on materials present $(\varepsilon_1, \mu_1, \sigma_1)$. Maxwell's equations for this system from (2-16) and (2-15) are

$$\nabla \times \mathbf{E}_1 = -j\omega\mu_1\mathbf{H}_1 \tag{3-23}$$

$$\nabla \times \mathbf{H}_1 = -j\omega\varepsilon_1' \mathbf{E}_1 + \mathbf{J}_1 \tag{3-24}$$

where $\mathbf{E}_1$ and $\mathbf{H}_1$ are the fields generated by $\mathbf{J}_1$ with materials $(\varepsilon_1, \mu_1, \sigma_1)$ present. Now suppose a fictitious magnetic current source with magnetic current density $\mathbf{M}_2$ exists with materials $(\varepsilon_2, \mu_2, \sigma_2)$ present. Maxwell's equations for this system from (2-15) and (2-21) are

$$\nabla \times \mathbf{H}_2 = j\omega\varepsilon_2' \mathbf{E}_2 \tag{3-25}$$

$$\nabla \times \mathbf{E}_2 = -j\omega\mu_2 \mathbf{H}_2 - \mathbf{M}_2 \tag{3-26}$$

where $\mathbf{E}_2$ and $\mathbf{H}_2$ are the fields arising from $\mathbf{M}_2$.

Table 3-1 Dual Radiating Systems. Radiating System #1 with Electric Currents and System #2 with Magnetic Currents Are Duals If One Can:

Replace the Following in System #2	By the Following in System #1
$\mathbf{M}_2$	$\mathbf{J}_1$
ε_2'	μ_1
μ_2	ε_1'
$\mathbf{E}_2$	$-\mathbf{H}_1$
$\mathbf{H}_2$	$\mathbf{E}_1$

The electric and magnetic systems are duals if the procedure in Table 3-1 can be performed. This is easy to demonstrate. To see if (3-25) and (3-26) are the duals of (3-23) and (3-24), we substitute the quantities in the left-hand column of Table 3-1 into (3-25) and (3-26) for the corresponding quantities of the right-hand column:

$$\nabla \times \mathbf{E}_1 = j\omega\mu_1(-\mathbf{H}_1) \tag{3-27}$$

$$\nabla \times (-\mathbf{H}_1) = -j\omega\varepsilon_1' \mathbf{E}_1 - \mathbf{J}_1 \tag{3-28}$$

Thus, the equations of the electric system, (3-23) and (3-24), are dual to the equations of the magnetic system, (3-25) and (3-26), just as (3-21) and (3-22) are dual equations. Since the equations of the systems are dual, the solutions will be also. Before illustrating this, we summarize the principle of duality:

If the sources of two systems are duals, that is,

$$\mathbf{M}_2 \ (=) \ \mathbf{J}_1 \tag{3-29}$$

and **if** the boundary conditions are also dual,[1] that is,

$$\mu_2 \ (=) \ \varepsilon_1', \quad \varepsilon_2' \ (=) \ \mu_1 \tag{3-30}$$

then the fields of system #2 can be found from the solution of system #1 by the substitutions

$$\mathbf{E}_2 \ (=) - \mathbf{H}_1, \quad \mathbf{H}_2 \ (=) \ \mathbf{E}_1 \tag{3-31}$$

in the field expressions for system #1 along with the substitutions in (3-30).

Now we use duality to find the fields of a small current loop from a knowledge of the fields of an ideal electric dipole. A current loop can be represented as a fictitious (ideal) magnetic dipole with uniform magnetic current I^m and length Δz. The sources are duals as required by (3-29) if we let

$$I^m \ (=) \ I^e \tag{3-32}$$

where I^e is the current of an ideal electric dipole of length Δz. Since no materials are present, there are no boundary conditions. The ideal electric dipole has field solutions of the form

$$\mathbf{E}_1 = E_{\theta 1}\hat{\boldsymbol{\theta}} + E_{r1}\hat{\mathbf{r}} \tag{3-33}$$

$$\mathbf{H}_1 = H_{\phi 1}\hat{\boldsymbol{\phi}} \tag{3-34}$$

The fields of the dual magnetic dipole are then found from (3-31) as

$$\mathbf{E}_2 \ (=) - \mathbf{H}_1 = -H_{\phi 1}\hat{\boldsymbol{\phi}} \tag{3-35}$$

$$\mathbf{H}_2 \ (=) \ \mathbf{E}_1 = E_{\theta 1}\hat{\boldsymbol{\theta}} + E_{r1}\hat{\mathbf{r}} \tag{3-36}$$

if we make the substitutions

$$\mu_2 = \mu \ (=) \ \varepsilon_1' = \varepsilon \quad \text{and} \quad \varepsilon_2' = \varepsilon \ (=) \ \mu_1 = \mu \tag{3-37}$$

which follow from (3-30) and the fact that in both systems the surrounding medium is a homogeneous material of μ and ε. Note that β remains the same since replacing μ by ε and ε by μ in $\omega\sqrt{\mu\varepsilon}$ yields $\omega\sqrt{\varepsilon\mu}$. Now, using and (3-32) and (3-37) in the ideal electric dipole field expressions of (2-73) together with (3-35) and (3-36) gives

$$\mathbf{E}_2 = \frac{I^m \Delta z}{4\pi} j\beta \left(1 + \frac{1}{j\beta r} \right) \frac{e^{-j\beta r}}{r} \sin\theta \ \hat{\boldsymbol{\phi}} \tag{3-38}$$

$$\mathbf{H}_2 = \frac{I^m \Delta z}{4\pi} j\omega\varepsilon \left[1 + \frac{1}{j\beta r} + \frac{1}{(j\beta r)^2} \right] \frac{e^{-j\beta r}}{r} \sin\theta \ \hat{\boldsymbol{\theta}}$$

$$+ \frac{I^m \Delta z}{2\pi} j\omega\varepsilon \left[\frac{1}{j\beta r} + \frac{1}{(j\beta r)^2} \right] \frac{e^{-j\beta r}}{r} \cos\theta \ \hat{\mathbf{r}} \tag{3-39}$$

---

[1] Note that $\varepsilon_1' = \varepsilon_1 - j(\sigma_1/\omega)$. If magnetic conductors of magnetic conductivity σ_2^m were assumed to exist in system #2, then μ_2 would become $\mu_2' = \mu_2 - j(\sigma_2^m/\omega)$ and ε_1' would be replaced by μ_2', or equivalently σ_1 replaced by σ_2^m.

(a) Small current loop and   (b) Ideal electric dipole.
      equivalent magnetic dipole.
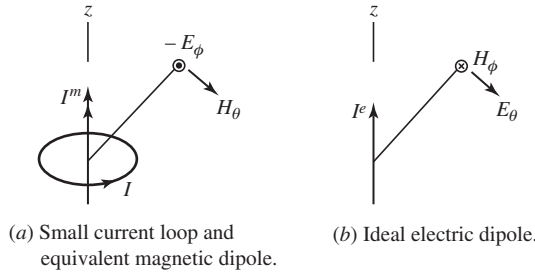
Figure 3-14 Radiation field components of ideal magnetic and electric dipoles.

These are the complete field expressions (valid in the near-field region) for a small loop of electric current. The far-field components are obtained by retaining only those terms that vary as r^{-1}, giving

$$\mathbf{E}_2 = -I^m \, \Delta z \, j\beta \frac{e^{-j\beta r}}{4\pi r} \sin\theta \, \hat{\boldsymbol{\phi}} \tag{3-40}$$

$$\mathbf{H}_2 = I^m \, \Delta z \, j\omega\varepsilon \frac{e^{-j\beta r}}{4\pi r} \sin\theta \, \hat{\boldsymbol{\theta}} \tag{3-41}$$

These radiation fields as well as those of the ideal electric dipole are shown in Fig. 3-14. Both antennas have the same radiation pattern, $\sin\theta$. The magnetic field component H_ϕ of the ideal electric dipole is easily remembered by use of the right-hand rule. Place the thumb of your right hand along the current of the dipole and pointing in the direction of current flow. Your fingers will then curl in the direction of the magnetic field. This statement is implicit in Ampere's law of (3-24). A similar relationship holds for the magnetic dipole, except the left-hand rule is used and the field obtained is the electric field component $-E_\phi$. This follows from (3-26).

3.4.2 The Small Loop Antenna

Using duality, we found the field expressions for a small loop of uniform current. However, these expressions contain the equivalent magnetic dipole current amplitude I^m. By solving the small loop problem directly, we can establish the relationship between the current I in the loop and I^m. This can be accomplished by dealing only with the far-field region.

It turns out that the radiation fields of small loops are independent of the shape of the loop and depend only on the area of the loop. Therefore, we will select a square loop as shown in Fig. 3-15a to simplify the mathematics. The current has constant amplitude I and zero phase around the loop. Each side of the square loop is a short uniform electric current segment that is modeled as an ideal dipole. The two sides parallel to the x-axis have a total vector potential that is x-directed and is given by

$$A_x = \frac{\mu I \ell}{4\pi} \left(\frac{e^{-j\beta R_1}}{R_1} - \frac{e^{-j\beta R_3}}{R_3} \right) \tag{3-42}$$

which follows from (2–65). The minus sign in the second term arises because the current in side 3 is negative x-directed. Similarly for sides 2 and 4, we find

$$A_y = \frac{\mu I \ell}{4\pi} \left(\frac{e^{-j\beta R_2}}{R_2} - \frac{e^{-j\beta R_4}}{R_4} \right) \tag{3-43}$$
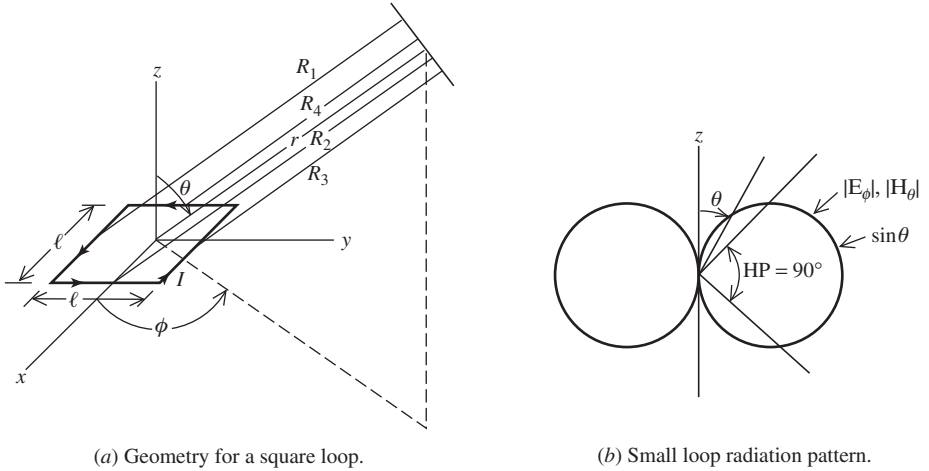
(a) Geometry for a square loop.

(b) Small loop radiation pattern.

Figure 3-15 The small loop antenna.

The far-field approximation is that the distances used for amplitude variations are nearly equal (i.e., $R_1 \approx R_2 \approx R_3 \approx R_4 \approx r$) and the phase differences are found from assuming parallel rays emanating from each side. By comparing the parallel path lengths, we find from geometrical considerations that

$$R_1 = r + \frac{\ell}{2}\sin\theta\sin\phi, \quad R_2 = r - \frac{\ell}{2}\sin\theta\cos\phi$$

$$R_3 = r - \frac{\ell}{2}\sin\theta\sin\phi, \quad R_4 = r + \frac{\ell}{2}\sin\theta\cos\phi \tag{3-44}$$

Substituting these into the exponents and r into the denominators of (3-42) and (3-43), we have

$$A_x = \frac{\mu I \ell e^{-j\beta r}}{4\pi r}\left(e^{-j\beta(\ell/2)\sin\theta\sin\phi} - e^{+j\beta(\ell/2)\sin\theta\sin\phi}\right)$$

$$A_y = \frac{\mu I \ell e^{-j\beta r}}{4\pi r}\left(e^{+j\beta(\ell/2)\sin\theta\cos\phi} - e^{-j\beta(\ell/2)\sin\theta\cos\phi}\right)$$

or

$$A_x = -2j\frac{\mu I \ell e^{-j\beta r}}{4\pi r}\sin\left(\frac{\beta\ell}{2}\sin\theta\sin\phi\right)$$

$$A_y = 2j\frac{\mu I \ell e^{-j\beta r}}{4\pi r}\sin\left(\frac{\beta\ell}{2}\sin\theta\cos\phi\right) \tag{3-45}$$

Since the loop is small compared to a wavelength, $\beta\ell = 2\pi\ell/\lambda$ is also small and the sine functions in (3-45) can be replaced by their arguments, giving

$$A_x \approx -j\frac{\mu I e^{-j\beta r}}{4\pi r}\beta \ell^2 \sin\theta \, \sin\phi$$

$$A_y \approx j\frac{\mu I e^{-j\beta r}}{4\pi r}\beta \ell^2 \sin\theta \, \cos\phi$$

(3-46)

Combining components to form the total vector potential gives

$$\mathbf{A} = A_x\hat{\mathbf{x}} + A_y\hat{\mathbf{y}} = j\beta \ell^2 \frac{\mu I e^{-j\beta r}}{4\pi r}\sin\theta \, (-\sin\phi\hat{\mathbf{x}} + \cos\phi\hat{\mathbf{y}}) \qquad (3\text{-}47)$$

The term in parentheses is the unit vector $\hat{\boldsymbol{\phi}}$ in (C-6), so

$$\mathbf{A} = j\beta S \frac{\mu I e^{-j\beta r}}{4\pi r}\sin\theta \, \hat{\boldsymbol{\phi}} \qquad (3\text{-}48)$$

where S is the area of the loop. All of $\mathbf{A}$ is transverse to the direction of propagation, so the radiation electric field from (2-104) is $-j\omega\mathbf{A}$, giving

$$\mathbf{E} = \eta\beta^2 S \frac{I e^{-j\beta r}}{4\pi r}\sin\theta \, \hat{\boldsymbol{\phi}} \qquad (3\text{-}49)$$

since $\omega\mu\beta = \omega^2\mu\sqrt{\mu s} = \sqrt{\mu/\varepsilon}\omega^2\mu\varepsilon = \eta\beta^2$. The radiation magnetic field is

$$\mathbf{H} = \frac{1}{\eta}\hat{\mathbf{r}}\times\mathbf{E} = -\beta^2 S \frac{I e^{-j\beta r}}{4\pi r}\sin\theta \, \hat{\boldsymbol{\theta}} \qquad (3\text{-}50)$$

Comparing (3-49) or (3-50) to the magnetic dipole radiation fields of (3-40) or (3-41), we find that

$$I^m \Delta z = j\omega\mu I S \qquad (3\text{-}51)$$

This completes the relationship between the small current loop and its equivalent magnetic dipole. The complete field expressions for a small loop of *magnetic moment IS* are found from (3-38) and (3-39) using (3-51). The fields depend only on the magnetic moment (current and area) and not the loop shape. And the radiation pattern for a small loop, independent of its shape, equals that of an ideal electric dipole; see Fig. 3-15b. The radiation fields from a large loop are derived in Sec. 6.8.

Another realization for a magnetic dipole in addition to the small loop is a narrow slot in a ground plane, with the fields being found from an equivalent magnetic current along the long axis of the slot. The slot antenna is discussed further in Secs. 8.10 and 16.7.

The loop antenna has been used since. Hertz first used it as a receiver in his experiments in 1886; see Fig. 1-1. It has an omnidirectional doughnut radiation pattern that is needed in many applications. The horizontal small loop (in the *xy*-plane) and vertical (*z*-directed) short dipole both have uniform radiation in the horizontal (*xy*) plane, but the loop provides horizontal polarization (E_ϕ), whereas the dipole is vertically polarized (E_θ). Next we discuss the impedance properties of the small loop and introduce the multiturn loop and ferrite core loop.

The impedance of a small loop antenna is quite different from its ideal dipole dual. Whereas the ideal dipole is capacitive, the small loop is inductive. We discuss the input resistance first. The radiation resistance is found by calculating the power radiated using the small loop radiation fields with (2-128), which yields

$$P = 10I^2(\beta^2 S)^2 \qquad (3\text{-}52)$$

The radiation resistance is then

$$R_r = \frac{2P}{I^2} = 20(\beta^2 S)^2 \approx 31{,}200 \left(\frac{S}{\lambda^2}\right)^2 \Omega \quad \textit{small loop antenna} \tag{3-53}$$

This result provides a reasonable approximation to the radiation resistance of an actual small loop antenna for a loop perimeter less than about three-tenths of a wavelength.

The radiation resistance of a loop antenna can be increased significantly by using multiple turns. The magnetic moment of an N turn loop is NIS, where S is the area of a single turn. The radiation resistance is then

$$R_r = 20(\beta^2 NS)^2 \approx 31{,}200 \left(\frac{NS}{\lambda^2}\right)^2 \Omega \quad \textit{small multiturn loop antenna} \tag{3-54}$$

Note that radiation resistance depends on the square of the number of turns, N^2. Provided the total wire length of the multitum loop is small compared to a wavelength (less than about 0.1λ), the current remains nearly constant over the wire length and previous small loop analysis applies so that the pattern is the same as a single turn loop as in Fig. 3-15.

Another popular way to enhance the radiation resistance of a loop antenna is to wind the turns around a ferrite core forming the **ferrite rod antenna** (also called a *loop-stick antenna* or *ferrite-core multiturn loop*) which is used up to VHF frequencies, where usually losses become excessive. The radiation resistance is found from (3-54) but the phase velocity is less than that of the ferrite core because it is of finite extent. Thus, the relative permeability of the bulk ferrite material is replaced by an *effective (relative) permeability*, μ_{eff}, and $\beta = \omega\sqrt{\mu\varepsilon} = \frac{2\pi}{\lambda}\sqrt{\mu_o\varepsilon_o}\sqrt{\mu_{\text{eff}}}$.

Modifying (3-54) to include this effect gives the radiation resistance for a loop-stick antenna as

$$R_r \approx 31{,}200 \left(N\mu_{\text{eff}}\frac{S}{\lambda^2}\right)^2 \Omega \quad \textit{ferrite rod antenna} \tag{3-55}$$

The effective permeability depends not only on the permeability of the core material but also on the core geometry as a function of the *relative core length* (length-to-diameter ratio), R, in the following way [3, 4]:

$$\mu_{\text{eff}} = \frac{\mu_r}{1 + D(\mu_r - 1)} \tag{3-56a}$$

where [4]

$$D = \text{demagnetization factor} = 0.37\,R^{-1.44} \tag{3-56b}$$

This equation reveals that μ_{eff} is always less than μ_r but for longer/thinner cores (larger R values) μ_{eff} approaches μ_r. This expression assumes the core is much longer than the winding on it; for other cases μ_{eff} is reduced, reaching 0.7 times that value in (3-56) for a core equal in length to the winding. Receivers for AM broadcast radio around 1 MHz are a popular application for the ferrite rod antenna with typical values in a wide range around 50 for μ_{eff} and 10 for R. Ferrites being lossy eliminates them for use in transmitters.

Small loop antennas also have ohmic resistance. For a rectangular loop of wire by ℓ_1 by ℓ_2, the ohmic resistance of the wire is given approximately by

$$R_w = \frac{2\ell_1\ell_2}{\pi(2a)^2}R_s\left\{\frac{1}{[(\ell_1/2a)^2 - 1]^{1/2}} + \frac{1}{[(\ell_2/2a)^2 - 1]^{1/2}}\right\} \tag{3-57}$$

where a is the wire radius and R_s is the surface resistance of (2-171). If and ℓ_1 and ℓ_2 are much larger than a (i.e., the wire is thin), then (3-57) reduces to

$$R_w = \frac{2(\ell_1 + \ell_2)}{2\pi a} R_s \tag{3-58}$$

This formula can be generalized to loops of arbitrary shape as follows:

$$R_w = \frac{L_m}{2\pi a} R_s \tag{3-59}$$

where L_m is the mean length of the wire loop, i.e., the perimeter.
    For a circular loop, this becomes

$$R_w = \frac{2\pi b}{2\pi a} R_s = \frac{b}{a} R_s \quad \text{circular loop} \tag{3-60}$$

where b is the mean loop radius and a is the wire radius; see (2-170).
    As mentioned previously, the small loop antenna is inherently inductive. The inductance of a small by ℓ_1 by ℓ_2 rectangular loop is given by

$$L = \frac{\mu}{\pi} \left(\ell_2 \cosh^{-1} \frac{\ell_1}{2a} + \ell_1 \cosh^{-1} \frac{\ell_2}{2a} \right) \quad \text{rectangular loop} \tag{3-61}$$

For a small circular loop of radius b, the inductance for $a \ll b$ is

$$L = \mu b \left[\ln \left(\frac{8b}{a} \right) - 2 \right] \quad \text{circular loop} \tag{3-62}$$

The loss and inductance of an N-turn loop must be multiplied by N^2. Ferrite rod antennas will also have losses in the ferrite core, but often they are smaller than coil losses. See [H.8.1: Miron] for a method and program to compute core losses.
    Radiation resistance decreases much faster with decreasing frequency for a small loop $(R_r \infty f^{-4})$ than for a short dipole $(R_r \infty f^{-2})$. Multiturn loops are used to increase radiation resistance; see (3-55). But wire losses can be reduced by decreasing the number of turns in a loop and using a ferrite core (i.e., loop-stick antenna) to maintain radiation resistance. In practice, a variable capacitor placed in parallel with the loop is used to tune out inductance.
    Small loop antennas have several applications, especially as a receiving antenna. For example, single turn small loop antennas have been popular in pagers. As already mentioned, multiturn small loops are popular in AM broadcast receivers. Small loop antennas are also used in direction finding receivers and for field strength probes.

EXAMPLE 3-1   A Small Circular Loop Antenna

A circular loop antenna has a single turn of 4-mm diameter copper wire. The loop circumference of 0.8 m is 0.08λ at the operating frequency of 30 MHz ($\lambda = 10$ m), which qualifies the antenna as electrically small. We wish to compute the input impedance and radiation efficiency. First, the radiation resistance from (3-54) is

$$R_r = 31{,}200 \left(\frac{S}{\lambda^2}\right)^2 = 31{,}200 \left(\frac{0.0509}{10^2}\right)^2 = 8.09 \times 10^{-3} \ \Omega \tag{3-63}$$

where the loop area is $S = \pi b^2 = 0.0509 \, \text{m}^2$ based on the radius $b = 0.8/\pi = 0.127 \, \text{m}$. The surface resistance found from (2-171) with the copper conductivity of $\sigma = 5.8 \times 10^7 \, \text{S/m}$ from Appendix B.1

$$R_s = \sqrt{\frac{2\pi f \mu_o}{2\sigma}} = 1.43 \times 10^{-3} \ \Omega \tag{3-64}$$

in (3-60) gives the wire loss resistance as

$$R_w = \frac{b}{a} R_s = \frac{0.127}{0.002} 1.43 \times 10^{-3} = 0.0910 \ \Omega \tag{3-65}$$

The total input resistance is then $R_A = R_r + R_w = 8.09 \times 10^{-3} + 0.091 = 0.0991 \, \Omega$ and the radiation efficiency is

$$e_r = \frac{R_r}{R_A} = \frac{8.09 \times 10^{-3}}{0.0991} = 0.082 = 8.2\% \tag{3-66}$$

The inductance from (3-62) is $L = 6.77 \times 10^{-7} \, \text{H}$ and $X_A = 2\pi L = 127 \, \Omega$; thus, the input impedance is $Z_A = R_A + jX_A = 0.0991 + j127 \, \Omega$. This compares favorably to the moment method simulation result that includes wire loss of $Z_A = 0.105 + j131 \, \Omega$. We note two things: the wire loss is much greater than the radiation resistance and the inductive reactance is much greater than the real part of the input impedance. This indicates that a small loop behaves largely as a radiating inductor.

The basic antenna elements that are used both conceptually and as physical elements in complex antenna systems have now been introduced. The characteristics of these elements are summarized in Table 3-2. All patterns are of the donut shape (omnidirectional) with uniform radiation in plane perpendicular to the dipole family of elements and uniform in the plane of the loop of the small loop antennas.

3.5   TWO-ELEMENT ARRAYS

There is no ability to control the radiation patterns of the individual element antennas we have discussed so far. However, combining the outputs of multiple antenna elements provides the possible of significantly changing the pattern, including pattern adjustment in response to time changing requirements. Such a configuration of multiple radiating elements is referred to as an **array antenna**, or simply, an **array**. In this section, we discuss the most basic form of an array, the two-element array. Larger arrays and the many applications for arrays are in Chap. 8.

Array analysis begins using elements that radiate equally in all directions, called an *isotropic point source* (or *isotropic source, isotropic radiator, isotropic element*, or *point source*). An isotropic transmitting (receiving) element transmits (receives) equally in all directions in three dimensions; see Fig. 2-14*a* for the pattern of an isotropic radiator. Although hypothetical, isotropic radiators provide a simple initial problem formulation. The full pattern with the effects of the real antennas used as elements can then be included by a simple multiplication process. Full array analysis involves summation of the phasors representing the amplitude and phase of each element and will be presented in detail in Chap. 8. In this section, we use an approach called the inspection method followed by a phasor summation approach illustrated with examples. We begin with isotropic element examples. The pattern of an array of isotropic elements is called the *array factor*. The term "factor" is included to indicate that in real arrays one must multiple the array factor by the element pattern to obtain the full array pattern. The examples use a transmit viewpoint, but by reciprocity the obtained pattern is the same on reception.

Table 3-2 Summary of Characteristics of Some Common Antenna Elements

Antenna Type	Wire Length	Current	Pattern	HP	D	D[dB]	$R_r\,[\Omega]$	Loss res. $[\Omega]$	Reactance
Isotropic[*]	—	—	isotropic	360°	1	0	—	—	—
Ideal Dipole	$L \ll \lambda$	Uniform	$\sin\theta$	90°	1.5	1.76	$80\pi^2\left(\dfrac{L}{\lambda}\right)^2$	$\dfrac{R_s}{2\pi a}L$	Capacitive
Short Dipole	$L \ll \lambda$	Triangle	$\sin\theta$	90°	1.5	1.76	$20\pi^2\left(\dfrac{L}{\lambda}\right)^2$	$\dfrac{R_s}{2\pi a}\dfrac{L}{3}$	Capacitive; see (2-176)
Half-wave Dipole	$L = 0.5\,\lambda$	Sinusoid	$\dfrac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta}$	78°	1.64	2.15	~ 70	$\dfrac{R_s}{2\pi a}\dfrac{\lambda}{4}$	Zero if trimmed to tune
Quarter-wave Monopole	$h = 0.25\,\lambda$	Sinusoid	$\dfrac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta}$	–	3.68	5.15	~ 35	$\dfrac{R_s}{\pi a}\dfrac{\lambda}{4}$	Zero if trimmed to tune
Small loop	$L_m \ll \lambda$	Uniform	$\sin\theta$	90°	1.5	1.76	$31{,}200\left(\dfrac{S}{\lambda^2}\right)$	$\dfrac{L_m}{2\pi a}R_s$	Inductive; see (3-61),(3-62)

*Hypothetical reference pattern

EXAMPLE 3-2 **Two Isotropic Point Sources with Identical Amplitude and Phase Currents, and Spaced One-Half Wavelength Apart (Fig. 3-16)**

Fig. 3-16*a* shows how the pattern of this example can be approximated by inspection. At points in the far field along the perpendicular bisector of the line joining the point sources (*x*-axis), path lengths from each point source are equal. Since the amplitudes and phases of each source are also equal, the waves arrive in phase and equal in amplitude in the far field along the *x*-axis. Thus, the total field is double that for one source. The situation is different along the axis of the array (*z*-axis). If we look to the right along the $+z$-axis, waves coming from the left source must travel one-half wavelength before reaching the source on the right. This amounts to a 180° phase lag. The waves then continue traveling to the right along the $+z$-axis and maintain this same phase relationship on out to the far field. Thus, in the far field, waves from the two sources traveling in the $+z$-direction arrive 180° out-of-phase (due to the one-half wavelength separation of the sources) and are equal in amplitude (since the

(*a*) Inspection method.

(*b*) Polar plot of the array factor
$f(\theta) = \cos[(\pi/2)\cos\theta]$.

(*c*) 3D polar pattern.

(*d*) Geometry for pattern calculation using rays.
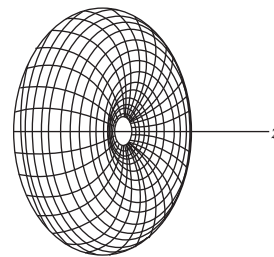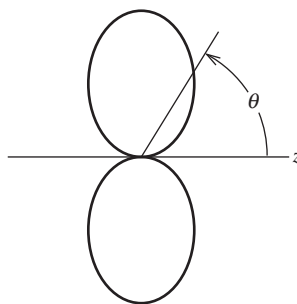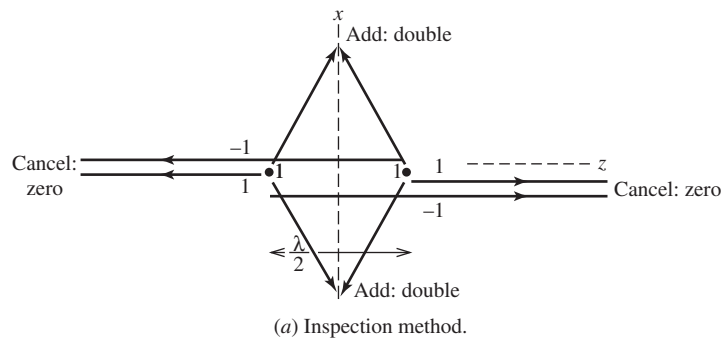
Figure 3-16 Two isotropic point sources with identical amplitude and phase currents, and spaced one-half wavelength apart (Example 3-2).

sources are). Therefore, there is a perfect cancellation and the total field is zero. The same reasoning can be used to see the effect in the $-z$-direction. The total pattern has a relative value of 2 in the $\pm x$-directions, 0 in the $\pm z$-directions, and a smooth variation in between (because the phase difference between waves from two sources changes smoothly from 0 to 180° as the observer moves from the broadside direction to the axial direction along a constant radius from the array center). This pattern is sketched in Fig. 3-16b. The pattern in three dimensions can be imagined by holding the z-axis in your fingertips and spinning the pattern shown to sweep out the total pattern. The three-dimensional pattern in Fig. 3-16c is a doughnut-type pattern similar to that for an ideal dipole.

We can also calculate the array factor exactly. If we use phases corresponding to the path length differences shown in Fig. 3-16d, the array factor is the sum of two phasors:

$$\text{AF} = 1e^{-j\beta(d/2)\cos\theta} + 1e^{j\beta(d/2)\cos\theta} = 2\cos\left(\beta\frac{d}{2}\cos\theta\right) \tag{3-67}$$

The distance between the elements is $d = \lambda/2$, so $\beta d/2 = \pi/2$ and (3-67) becomes

$$\text{AF} = 2\cos\left(\frac{\pi}{2}\cos\theta\right) \tag{3-68}$$

Normalizing the array factor for a maximum value of unity gives

$$f(\theta) = \cos\left(\frac{\pi}{2}\cos\theta\right) \tag{3-69}$$

This is maximum for $\theta = \pi/2$ since $\cos\left[(\pi/2)\cdot 0\right] = 1$ and zero for $\theta = 0$ since $\cos\left[(\pi/2)\,1\right] = 0$. This result agrees with the inspection method that leads to Fig. 3-16b.

EXAMPLE 3-3   **Two Isotropic Point Sources with Identical Amplitudes and Opposite Phases, and Spaced One-Half Wavelength Apart (Fig. 3-17)**

If we consider the array to be transmitting, the gross features of the pattern can be determined by inspection as shown in Fig. 3-17a. The path lengths from each point source to a point on the x-axis are the same. But the left source is 180° out-of-phase with respect to the right source; thus, waves arriving at points on the x-axis are 180° out-of-phase and equal in amplitude, giving a zero field. Along the z-axis (in both directions), the 180° phase difference in the currents is compensated for by the half-wavelength path difference between waves from the two sources. For example, in the $+z$-direction the waves from the left source arrive at the location of the right source, lagging the phase of waves from the right source by 360° (180° from the distance traveled and 180° from the excitation lag). This is an in-phase condition and thus the waves add in the far field, giving a relative maximum. From these few pattern values, the entire pattern can be sketched, yielding a plot similar to that of Fig. 3-17b. The three-dimensional polar plot of the pattern shown in Fig. 3-17c has the shape of a dumbbell.

We calculate the array factor exactly by phasor addition and Fig. 3-16d as

$$\text{AF} = -1e^{-j\beta(d/2)\cos\theta} + 1e^{j\beta(d/2)\cos\theta} = 2j\sin\left(\beta\frac{d}{2}\cos\theta\right) \tag{3-70}$$

Using $d = \lambda/2$ and normalizing, we have

$$f(\theta) = \sin\left(\frac{\pi}{2}\cos\theta\right) \tag{3-71}$$

Plotting this pattern, we obtain the same result as with the inspection method (see Fig. 3-17b).

(a) Inspection method.

(b) Polar plot of the array factor magnitude
$|f(\theta)| = |\sin[(\pi/2) \cos\theta]|$.

(c) 3D polar pattern.
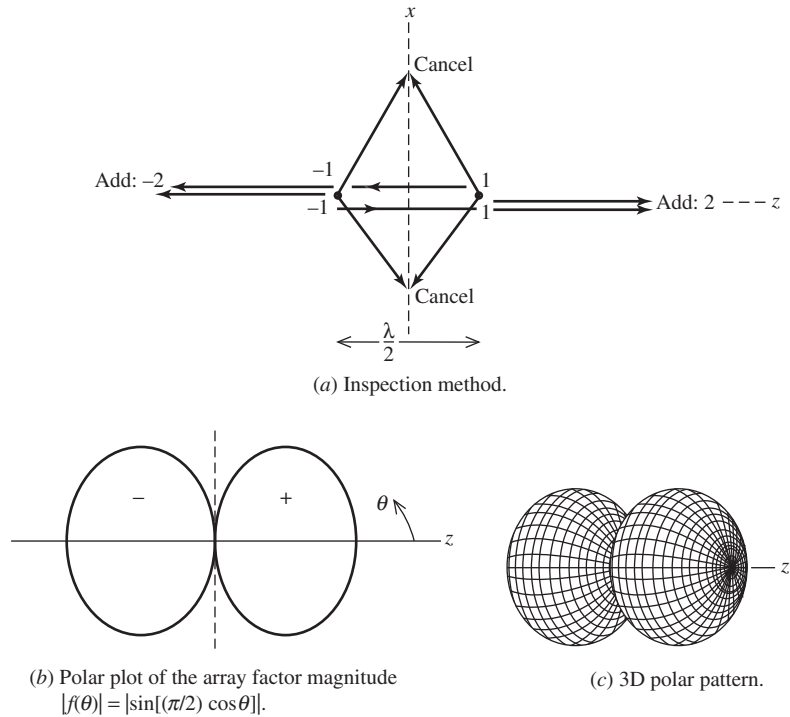
Figure 3-17 Two isotropic point sources with identical amplitudes and opposite phases, and spaced one-half wavelength apart (Example 3-3).

EXAMPLE 3-4   ***Two Isotropic Point Sources with Identical Amplitudes and 90° Out-of-Phase, and Spaced a Quarter-Wavelength Apart (Fig. 3-18)***

Waves leaving the left source of the transmitting array in Fig. 3-18 and traveling in the $+z$-direction arrive at the right source delayed by 90° due to the quarter-wavelength path. But the excitation of the right source lags the left source by 90° so waves in the $+z$-direction are in step and add in the far field. For waves leaving the right-hand source and traveling in the $-z$-direction, the phase at the location of the left source is 180° with respect to the wave from the left source (90° from the path difference and 90° from the excitation). See Fig. 3-18b. At angles between $\theta = 0°$ ($+z$-direction) and 180° ($-z$-direction), there is a smooth pattern variation from 2 (perfect addition) to 0 (perfect cancellation). This pattern is shown in Fig. 3-18c and is the so-called *cardioid pattern*. It is used frequently in the area of acoustics for microphone patterns. The response is strong in the direction of the microphone input and weak in the direction where the speakers are aimed to reduce feedback.

The phasor addition method yields the array factor expression as follows:

$$\text{AF} = 1e^{-j\beta(d/2)\cos\theta} + 1e^{-j(\pi/2)}e^{j\beta(d/2)\cos\theta}$$

$$= e^{-j(\pi/4)}\left[e^{-j[\beta(d/2)\cos\theta - \pi/4]} + e^{j[\beta(d/2)\cos\theta - \pi/4]}\right] = e^{-j(\pi/4)}2\cos\left(\frac{\beta d}{2}\cos\theta - \frac{\pi}{4}\right) \quad (3\text{-}72)$$

Substituting $d = \lambda/4$ and normalizing give

$$f(\theta) = \cos\left[\frac{\pi}{2}(\cos\theta - 1)\right] \quad (3\text{-}73)$$

(a) Array configuration.

(b) Inspection method.

(c) Polar plot of the array factor
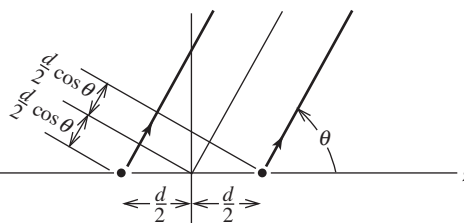$f(\theta) = \cos[(\pi/4)(\cos\theta - 1)]$.

(d) 3D polar pattern.
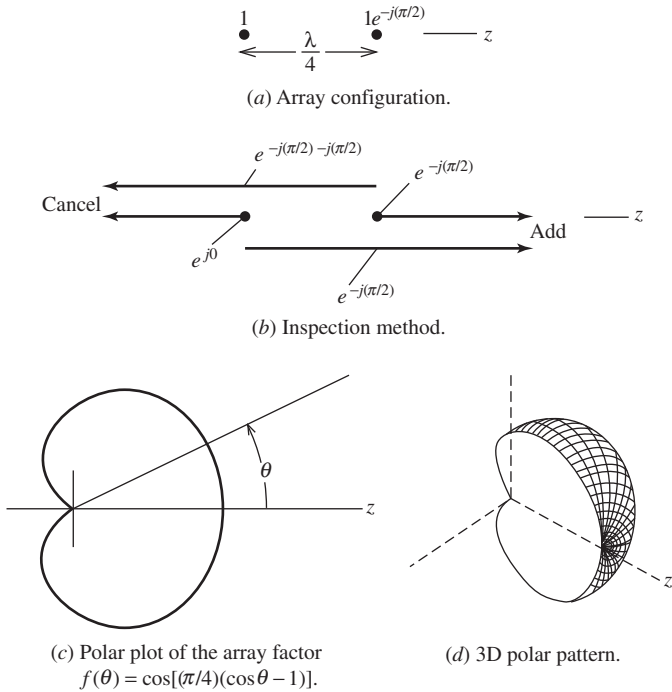
Figure 3-18 Two isotropic point sources with identical amplitudes and the right element lagging the left by 90°, and spaced a quarter-wavelength apart (Example 3-4). This pattern shape is called a cardioid pattern.

This function has a maximum value of unity for $\theta = 0°, 1/\sqrt{2}$ for $\theta = 90°$, and 0 for $\theta = 180°$. This agrees with the pattern of Fig. 3-18c obtained by inspection.

Array factor plots for arrays of two elements of equal amplitude excitations for various combinations of excitation phase, α, and element spacing, d, are shown in Fig. 3-19; the first such pattern display was in 1937 by Brown [5]. The absolute field strength has been preserved to allow comparison. Also shown is a unit circle representing the radiation from a hypothetical isotropic point source with the same input current. The reader should locate the plots for Examples 3-2 to 3-4 on the figure. Fig. 3-19 shows how phase control can significantly change the shape of an array factor, even for spacings as small as $d = \lambda/8$. The topic of phased arrays is treated in Sec. 8.9

Example 3-5 that follows serves to illustrate both how to obtain the full pattern of an array and to introduce the practical implementation of ground plane backing of a single element to produce an array effect. As indicated in Fig. 3-20, this arrangement produces what is effectively a two-element array but requires only one physical antenna. There are many practical implementations of ground-plane backed antennas; see Secs. 6.6 and 12.2.

EXAMPLE 3-5   *A Half-Wave Dipole, One-Quarter Wavelength in Front of a Ground Plane*

Fig. 3-20 shows a half-wave dipole spaced parallel to and a quarter wavelength in front of a ground plane $(s = \lambda/4)$. Image theory of Sec. 3.3 can be applied to this geometry to deal with the presence of the ground plane. Image theory is used to create the equivalent problem yielding the same fields for $z > 0$ by removing the ground plane and introducing an image dipole of the
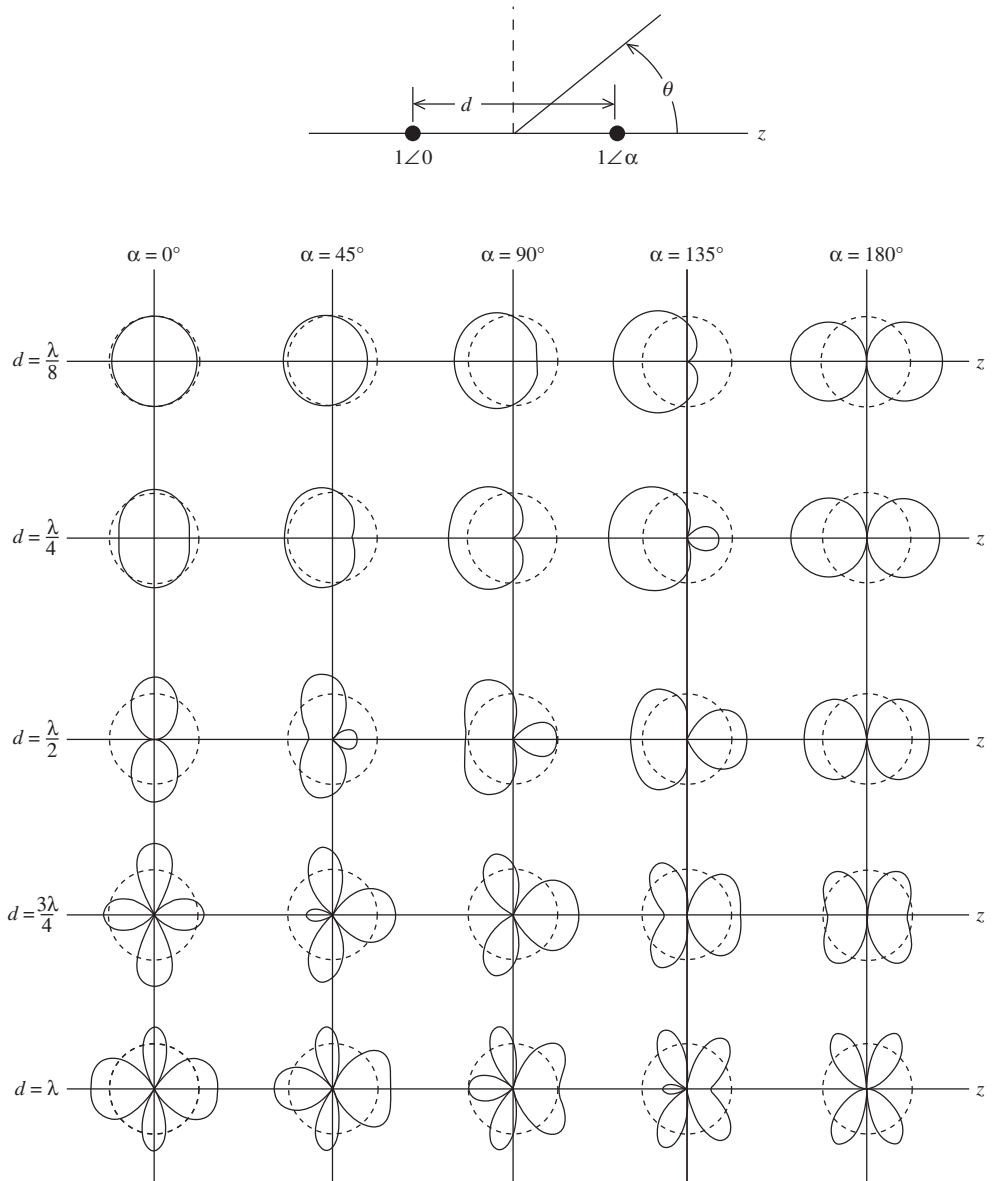
Figure 3-19 Array factor plots for arrays of two isotropic elements with equal current amplitudes, and various combinations of excitation phase, α, and element spacing, d. Also shown is a unit circle representing the radiation from a hypothetical isotropic point source with the same input current.

same length that is parallel to the source dipole and equidistant from the ground plane, thus $d = \lambda/2$. The principle in Fig. 3-8 indicates that the image dipole will have an amplitude of excitation equal to the source dipole and be $180°$ out of phase, which is essentially the array geometry of Example 3-3. So the array factor is given by (3-71) as $f(\theta) = \sin\left(\frac{\pi}{2}\cos\theta\right)$. The element is a half-wave dipole, but not oriented along the z-axis as usual, but along the x-axis. The element pattern is given by (see Example 8-7) $g_a(\theta,\phi) = \dfrac{\cos\left((\pi/2)\sin\left(\theta\right)\cos\phi\right)}{\sqrt{1 - \sin^2\theta\cos^2\phi}}$. As mentioned earlier and as will be detailed in Sec. 8.4, the complete array pattern is obtained by multiplying the array factor by the element pattern, giving

(a)

(b)   *(c)*
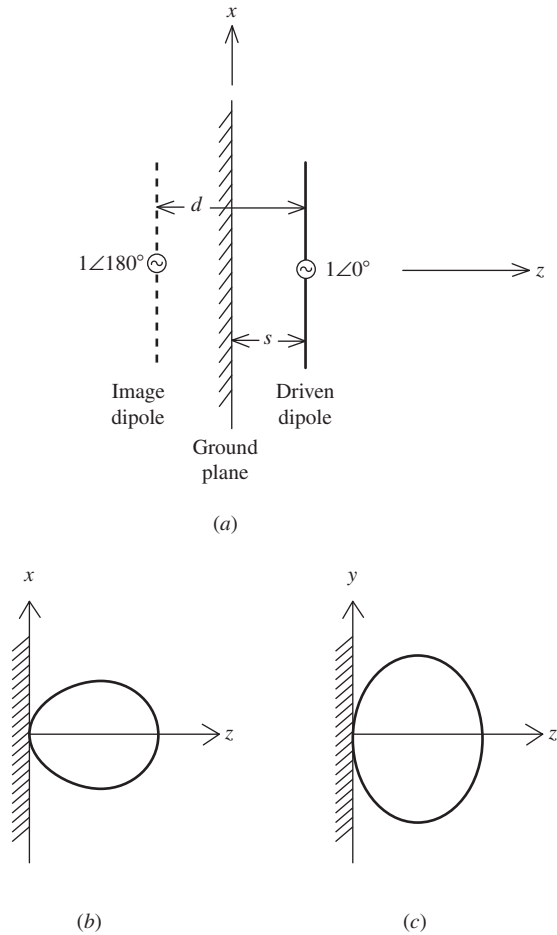
Figure 3-20 A half-wave dipole parallel to and a quarter wavelength in front of a ground plane $(s = \lambda/4)$. The image is parallel to the source, equal in amplitude, $180°$ out of phase and equispaced to the ground plane, so that $d = \lambda/2$ (Example 3-5). (*a*) The geometry. (*b*) The E-plane pattern. (*c*) The H-plane pattern.

$$F(\theta, \phi) = g_a(\theta, \phi) f(\theta) \tag{3-74a}$$

$$= \frac{\cos\left((\pi/2)\sin(\theta)\cos\phi\right)}{\sqrt{1 - \sin^2\theta \cos^2\phi}} \sin\left(\frac{\pi}{2}\cos(\theta)\right) \tag{3-74b}$$

The principal plane patterns using this function are plotted in Fig. 3-20*b* and *c*. The radiation in the back half plane $(z < 0)$ is not plotted because the fields are zero there due to the shorting effect of the ground plane. Note that the E-plane pattern is narrower than the H-plane pattern $(\text{HP}_E = 72°, \text{HP}_H = 120°)$ because the element pattern is omnidirectional in the H-plane, i.e., $g_a(\theta, \phi = 90°) = 1$. Also note that a unidirectional beam is formed by this array using only a single antenna element. This is a very desirable pattern with many applications. The directivity obtained by integrating the pattern of (3-74*b*) to find the beam solid angle, Ω_A, and then using $D = 4\pi/\Omega_A$ yields 7.5 dB. This increase in directivity over that for the isolated half-wave dipole (2.15 dB) is due to the array effect with the image and to eliminating the radiation in the back half-space. Prob. 3.5-5 revisits this configuration using simulation.

REFERENCES

1. V. Trainotti and L. Dorado, "Short Low- and Medium-Frequency Antenna Performance," *IEEE Trans. on Ant. and Prop. Mag.*, Vol. 47, pp. 66–90, Oct. 2005.
2. A. F. Gangi, S. Sensiper, and G. R. Dunn, "The Characteristics of Electrically Short, Umbrella Top-loaded Antennas," *IEEE Trans. on Ant. and Prop.,* Vol. AP-13, pp. 864–871, Nov. 1965. Also see W. Weeks, *Antenna Engineering,* McGraw-Hill, New York, 1968; pp. 44–46.
3. W. Polydoroff, *High Frequency Magnetic Materials*, John Wiley: New York, 1960.
4. R. Pettengill, H. Garland, and J. Mendl, "Receiving Antenna Design for Miniature Receivers," *IEEE Trans. on Ant. and Prop.,* Vol. AP-26, pp. 528–530, July 1977.
5. G. H. Brown, "Directional Antennas," *Proc. IRE*, Vol. 25, pp. 78–145, Jan. 1937.

PROBLEMS

3.1-1 Use the oscillating charge model for an ideal dipole as shown in Fig. 3-2*b* to derive the electric field expressions of (2-73*b*). *Hints:* The far-field scalar potential function for this problem is

$$\Phi = \frac{q}{4\pi\varepsilon_\circ}\left[\frac{e^{-j\beta(r-(\Delta z/2)\cos\theta)}}{r-(\Delta z/2)\cos\theta} - \frac{e^{-j\beta(r+(\Delta z/2)\cos\theta)}}{r+(\Delta z/2)\cos\theta}\right]$$

where the parallel ray approximation was used and the e^{jwt} time dependence was suppressed. Use $r \gg \Delta z$, $\lambda \gg \Delta z$, and $I = j\omega q$ to show that

$$\Phi = \frac{e^{-j\beta r}}{4\pi r^2}\frac{I\Delta z}{j\omega\varepsilon_\circ}(1 + j\beta r)\cos\theta$$

Then make use of (2-39).

3.1-2 The current density on an actual short dipole antenna of Fig. 3-1*b* can be written as

$$\mathbf{J} = \hat{\mathbf{z}}J_o\sin\left[\beta\left(\frac{\Delta z}{2} - |z|\right)\right]$$

Find an expression for the associated charge density.

3.1-3 Show that the capacitance of the capacitor of the capacitor-plate antenna of Fig. 2-3 is given by

$$C = \frac{\pi(\Delta r)^2\varepsilon_o}{\Delta z}$$

Assume that capacitance is entirely due to the end plates and neglect fringing.

3.1-4 (a) Using the capacitance formula in Prob. 3.1-3, calculate the capacitive reactance of a capacitor-plate dipole for which $\Delta r = 0.01\,\lambda$ and $\Delta z = 0.02\,\lambda$. (b) Calculate the radiation resistance of this antenna.

3.2-1 Sketch the current distribution on a half-wave dipole for various instants during the time cycle of the current oscillation.

3.2-2 Show that the pattern factor for half-wave dipole in (3-3) is normalized to unity at $\theta = \pi/2$.

3.2-3 Calculate and plot the radiation pattern $F(\theta)$ for a half-wave dipole in (3-4) for $0 \le \theta \le 180°$. Plot in linear, polar form as shown in Fig. 3-5*b*.

3.2-4 Show that the ohmic resistance of a half-wave dipole from (2-173) is given by

$$R_\mathrm{o} = \frac{R_s}{2\pi a}\frac{\lambda}{4}$$

3.2-5 Use the results of Prob. 3.2-4 to calculate the radiation efficiency of a half-wave dipole at 100 MHz if it is made of aluminum wire 6.35 mm (0.25 in.) in diameter. Assume the radiation resistance to be $70\,\Omega$.

3.2-6 Derive the far-field electric field expression of (3-2) for a half-wave dipole.

3.2-7 *Simulation* Use a simulation code (a method of moments code is most appropriate; see Chap. 14 and App. G.1) to compute the input impedance of half-wave dipole that is $0.48\,\lambda$ long made of $0.0003\,\lambda$ radius wire. Neglect conductor losses.

3.3-1 Show that the image theory model of Fig. 3-8*b* for an ideal dipole parallel to a perfect ground plane yields zero tangential electric field along plane PP'.

3.3-2 For a thin monopole as shown in Fig. 3-10*a* that is a quarter wavelength long:
  (a) Rough sketch the radiation pattern in polar form as a function of θ, if the monopole is along the *z*-axis.
  (b) What is the directivity?
  (c) What is the input impedance?

3.3-3 *Umbrella-loaded monopole antenna* Calculate the radiation resistance of an umbrella-loaded monopole that is $0.2\,\lambda$ tall using (3-20). Compare to values obtained from the short dipole formula (3-15) and the perfect quarter-wave monopole result of (3-19), and comment.

3.3-4 *Simulation* (a) Use a simulation code to compute the input impedance and gain of a half-wave dipole that is $0.49\,\lambda$ long made of $0.0001\,\lambda$ radius wire. Neglect conductor losses. (b) Repeat (a) for the quarter-wave monopole version of the dipole with a perfect ground plane. Compare to the results in (a) and comment.

3.4-1 Use (2–98) to derive the far-field distance expressions (3-44) for the small square loop.

3.4-2 Verify that the power radiated from a small loop is given by (3-52).

3.4-3 Show that (3-58) follows from (3-57).

3.4-4 Find the input impedance of a circular loop antenna of wire radius $a = 0.0005\,\lambda$ for loop perimeters of $0.01\,\lambda$, $0.05\,\lambda$, and $0.1\,\lambda$, and a frequency of 300 MHz. Assume the wire to be lossless. (a) Compute the radiation resistance and reactance values using (3-53) and (3-62). (b) *Simulation* Use a simulation code to evaluate radiation resistance and reactance values and compare to the values found in (a).

3.4-5 Repeat Prob. 3.4–4 for the case of a square loop antenna, comparing to radiation resistance and reactance formulas of (3-53) and (3-61).

3.4-6 Find the input impedance neglecting ohmic losses of a small square loop antenna with the same area as the loop in Example 3-1 also operating at 30 MHz and using a lossless wire of 4 mm diameter. (a) Use appropriate formulas to calculate the impedance. (b) *Simulation* Use a simulation code to find the impedance and compare to the values in part (a) and to the calculated values in Example 3-1. What are the gain and HP values?

3.4-7 A single-turn circular loop antenna with a radius 0.5 m and made of No. 8 AWG copper wire operates at 10 MHz. Calculate: (a) radiation resistance, (b) ohmic resistance, (c) radiation efficiency, and (d) input reactance. (e) *Simulation* Perform a numerical simulation including wire loss to determine the same quantities and compare them. Also, compare gain values.

3.4-8 A single-turn circular loop antenna with a radius 0.15 m and made of No. 20 AWG copper wire operates at 1 MHz. Calculate: (a) radiation resistance, (b) ohmic resistance, (c) radiation efficiency, and (d) input reactance. (e) *Simulation* Perform a numerical simulation including wire loss to determine the same quantities and compare them. Also, compare gain values.

3.4-9 An AM broadcast receiver operating at 1200 kHz uses a ferrite rod antenna with 1000 turns of No. 30 copper wire wound evenly on a rod core of ferrite with $\mu_r = 50$ that is 30 mm long and 8 mm in diameter. Find the radiation resistance, the radiation efficiency neglecting ferrite core losses, and the reactance.

3.4-10 An AM broadcast receiver operating at 1 MHz uses a ferrite rod antenna with 500 turns of No. 30 copper wire wound on a core of ferrite with $\mu_r = 38$ that is 25 mm long and with a cross-section that is 1 cm by 3 mm. Find the radiation resistance and the radiation efficiency neglecting ferrite core losses.

3.4-11 A single-turn square loop antenna that is 0.5 m on each side operates at 30 MHz. The wire is aluminum with a diameter of 2 cm. Compute: (a) radiation resistance, (b) input impedance, and (c) radiation efficiency.

3.4-12 (a) Derive an expression for the wave impedance (E_ϕ/H_θ) for a magnetic dipole using (3-38) and (3-39). (b) Evaluate and plot this equation for βr from 0 to 20. Note the value for $\beta r = 10\,(r = 1.6\,\lambda)$ compared to the intrinsic impedance of free space.

3.5-1 Write a computer program to evaluate and plot the polar diagram radiation patterns in Examples 3-2 to 3-4.

3.5-2 A two-element array consists of center-fed, short dipoles both along the z-axis and spaced 0.4λ apart center-to-center. They are driven with equal amplitude and phase voltage sources. (a) Derive the analytical pattern expression, (b) *Simulation* Use a computer code to simulate the array assuming $L = 0.08\,\lambda$ and $a = 0.001\,\lambda$ with $f = 300$ MHz. Plot the radiation pattern. Compare the half-power beam width values to those based on (a).

3.5-3 Repeat Prob. 3.5-2 for $d = 0.8\,\lambda$.

3.5-4 A two-element array consists of two resonated, center-fed, half-wave dipoles both along the z-axis and spaced (center-to-center) $1\,\lambda$ apart. They are driven with equal amplitude and phase voltage sources. (a) Derive the analytical pattern expression. (b) *Simulation* Use a computer code to simulate the array assuming $L = 0.48\,\lambda$ and $a = 0.0005\,\lambda$ with $f = 300$ MHz. Plot the radiation pattern. Compare the half-power beam width values to those based on (a).

3.5-5 *Simulation* (a) Use a simulation code to evaluate the dipole in front of a ground plane case of Example 3-5. The dipole operating at 300 MHz and its image have $L = 0.47\,\lambda$ and $a = 0.001\,\lambda$. Give the input impedance and gain values. Plot the radiation patterns in the two principal planes and compare the half-power beamwidth values to those based on the pattern of (3-74b). (b) Use numerical integration of the pattern of (3-74b) to find the directivity and compare to the simulation result.

Chapter **4**

System Applications for Antennas

4.1 INTRODUCTION

An antenna is a device that interfaces the RF circuits of the transmitter/receiver to free space. At the beginning of a system design project, the antenna specifications should be developed in concert with the full system specifications. If the system is designed without consideration for the antenna, there may not be adequate volume allocated to the antenna for it to meet performance specifications. An analogy to the system/device tradeoff is a patient visiting a doctor and asking for an antibiotic. The doctor will, of course, not immediately write a prescription, but will instead ask the patient about his/her symptoms and will perform appropriate tests to evaluate the whole-body status. Similarly, the antenna engineer must gather information on the system parameters, and should participate in system specification writing for the project at hand. The electrical, mechanical, and environment specifications for the antenna subsystem flow directly from the full system design phase. With the current evolution of antenna technology toward functionality distributed through the system, there will be increasing involvement of the antenna engineer in system design.

As mentioned in Sec. 1.2, antenna application areas are communications, sensing and imaging (passive and active), and industrial uses (e.g., control, medical, cooking). An example of a non communication wireless application is an implantable "microchip" RFID tag under the skin of a pet. In this chapter, we focus on antennas in communication systems because of the major importance of communication applications. However, nearly all material in the chapter is applicable to all of the application areas. Also included are topics in noise that are necessary for communication systems and for radiometry.

4.2 RECEIVING PROPERTIES OF ANTENNAS

A receiving antenna converts the power density arriving from a distant source to a current on the connecting transmission line. It is important to carefully account for all losses because the received signals are weak. Essential to this is proper modeling of the receiving antenna. This section addresses various receiving antenna models and discusses losses associated with the system configuration, including polarization and impedance mismatch losses.
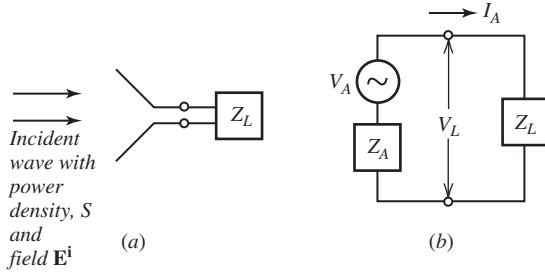
Figure 4-1 Equivalent circuit for a receiving antenna. (*a*) Receive antenna connected to a receiver with load impedance Z_L. (*b*) Equivalent circuit.

A receiving antenna collects power from an incoming plane wave.[1] There are two ways to compute the received power, one based on the incident electric field intensity and one on the incident power density. We will use both approaches because of the important relationships that arise in the derivations, and the resulting quantities are both used in practice. Beginning with the field approach, the receiving antenna converts the incident wave electric field E^i to an open-circuit voltage V_A through the *effective length h* (also called *effective height*) of the receiving antenna that is defined as

$$h = \frac{V_A}{E^i} \quad \text{effective length of an antenna} \tag{4-1}$$

So, effective length is the ratio of the induced voltage to incident electric field. V_A is the open-circuit voltage across the antenna terminals as shown in Fig. 4-1; that is, with Z_L infinite. The quantities V_A and E^i are peak values rather than rms. This effective length is referred to by the IEEE as the "effective length of a linearly polarized antenna." We generalize the effective length concept to arbitrary wave and antenna polarizations, and also include phase by defining *vector effective length* **h** through the relation:

$$V_A = \mathbf{E}^i \cdot \mathbf{h}^* \tag{4-2}$$

The complex conjugate in (4-2) will be explained in association with the polarization mismatch discussion below; if either the wave or the antenna is linearly polarized, it can be omitted.

The vector effective length expression can also be derived from basic electromagnetics by making use of reciprocity. [1] It is an intuitive relationship with the dot product representing the projection of the incident field vector, $\mathbf{E^i}$, in volts per meter onto the vector effective length **h** in meters, resulting in an output voltage V_A in volts. **h** contains the receiving antenna pattern information and is often written as $\mathbf{h}(\theta, \phi)$. When only a single value for h is quoted, the assumption is that it is for the pattern maximum direction. Maximum output voltage occurs when the wave arrives from a direction of peak receive antenna pattern response and is matched to the polarization of the receive antenna.

EXAMPLE 4-1 *Vector Effective Length of an Ideal Dipole*

As an example, consider the radiation electric field of an ideal dipole, which from (2-74*a*) is

$$\mathbf{E} = \frac{j\omega\mu\, I}{4\pi} \frac{e^{-j\beta r}}{r} \Delta z \sin\theta \,\, \hat{\boldsymbol{\theta}} \tag{4-3}$$

---

[1] The wave at the receiver behaves as a plane wave over the extent of the receiving antenna because the spherical wave from a distant source arrives with a large radius of curvature. This is sometimes called local plane wave behavior.

Since **h** contains information on the size of the antenna and the angular dependence of the radiation pattern, then we can write

$$\mathbf{E} = \frac{j\omega\mu}{4\pi} \frac{e^{-j\beta r}}{r} \mathbf{h} \tag{4-4}$$

where

$$\mathbf{h} = \Delta z \sin\theta \; \hat{\boldsymbol{\theta}} \tag{4-5}$$

Note that the dimension of **h** is length, and for the ideal dipole, the effective length is the same as the projection of the physical length viewed from angle θ. This is not, however, true in general. Polarization information for the wave and antenna are contained in $\mathbf{E^i}$ and **h**, respectively; the topics of polarization and polarization mismatch are discussed below in Secs. 4.4.4 and 4.4.5. The vector effective length of a small loop antenna is treated in Prob. 4.2-1.

In the incident power density approach, the receiving antenna is viewed as converting the incident power density (also called the time-average flux density) S into received power by the "collecting area" of the receiving antenna, called *maximum effective aperture* (or area), A_{em}, as expressed by the defining relation:

$$P_{rm} = S \; A_{em} \tag{4-6}$$

The maximum available received power P_{rm} will be realized if the antenna is directed for maximum response (i.e., the main beam peak is in the direction of the incoming wave), is polarization matched to the wave, and is impedance matched to its load. The "maximum" refers to assumption that there are no ohmic losses on the antenna.

We can now connect the two antenna-wave interaction formulations. Maximum power will be received when the load impedance is conjugate matched to the antenna impedance ($R_L = R_A$ and $X_L = -X_A$):

$$Z_L = R_L + jX_L = R_A - jX_A \quad \text{conjugate match for maximum power transfer} \tag{4-7}$$

The received power obtained by examining the circuit in Fig. 4-1:

$$P_{rm} = \frac{1}{2}|I_A|^2 R_L = \frac{1}{2} \frac{|V_A|^2}{(R_A + R_L)^2 + (X_A + X_L)^2} R_L \tag{4-8}$$

For a conjugate match this reduces to

$$P_{rm} = \frac{1}{8} \frac{|V_A|^2}{R_A} \quad \text{impedance-matched case} \tag{4-9}$$

Using the voltage for the case of a linearly polarized antenna from (4-1) for simplicity in (4-9) and equating to (4-6) gives

$$P_{rm} = \frac{1}{8} \frac{|V_A|^2}{R_A} = \frac{1}{8} \frac{|E^i h|^2}{R_A} = SA_{em} = \frac{1}{2} \frac{|E^i|^2}{\eta} A_{em} \tag{4-10}$$

where the power density expression for the incoming wave (Poynting vector magnitude) was

$$S = \frac{1}{2}|\mathbf{E} \times \mathbf{H}^*| = \frac{1}{2} \frac{|E^i|^2}{\eta} \tag{4-11}$$

which follows from (2-30) for a plane wave. Solving (4-10) for h gives the desired effective length result [H.11.1: Stutzman, p. 141]:

$$h = 2\sqrt{\frac{R_A A_{em}}{\eta}} \tag{4-12}$$

Effective length is proportional to the square root of the maximum effective area and thus has units of meters, as would be anticipated. In practice, effective length is used in association with linear antennas such as dipoles and effective aperture is used with aperture antennas such as parabolic reflectors.

EXAMPLE 4-2   *Maximum Effective Aperture of an Ideal Dipole*

Maximum effective aperture for the ideal dipole is found using (4-9) and (4-10) with (4-5):

$$A_{\text{em}} = \frac{P_{\text{rm}}}{S} = \frac{\frac{1}{8}\frac{|V_A|^2}{R_r}}{\frac{1}{2}\frac{|E^i|^2}{\eta}} = \frac{1}{4}\frac{\eta}{R_r}\frac{|E^i|^2(\Delta z)^2}{|E^i|^2} = \frac{1}{4}\frac{\eta}{R_r}(\Delta z)^2 = \frac{1}{4}\frac{\eta(\Delta z)^2}{\eta\frac{2}{3}\pi\left(\frac{\Delta z}{\lambda}\right)^2} = \frac{3}{8\pi}\lambda^2$$

$$= 0.119\,\lambda^2 \tag{4-14}$$

where the ideal dipole radiation resistance value from (2-169) was used. This result is counterintuitive because it shows that the maximum effective aperture of an ideal dipole is independent of its length Δz (as long as $\Delta z \ll \lambda$). But the dipole length cannot be reduced arbitrarily without penalty. R_r is proportional to $(\Delta z/\lambda)^2$ so that even though A_{em} remains constant as the dipole is shortened, its radiation resistance decreases rapidly and the maximum effective aperture will probably not be realized because of the difficulty in achieving the required conjugate impedance match of the receiver to the antenna.

Another quantity used to characterize a receiving antenna is *antenna factor, K*, defined as

$$K = \frac{E^i}{V_L} \qquad \textit{antenna factor} \tag{4-13}$$

Antenna factor gives the output voltage V_L across a load, usually for a standard load such as a 50 Ω, caused by the incident field intensity E^i. Antenna factor has units of m^{-1}. It includes impedance mismatches but not transmission line losses. If a high-impedance load is used, the voltage in (4-13) is an open-circuit voltage and equals the inverse of the effective length in (4-1); that is, $K = 1/h$. Antenna factor is used to determine electric field by measuring the voltage received with a probe and multiplying by the antenna factor of the probe antenna, as will be explained in Sec. 13.5. Antenna factor will also be applied electromagnetic compatibility measurements in Sec. 4.7.

4.3   ANTENNA NOISE AND RADIOMETRY

Noise corrupts all types of receiving systems and in a given direction receiving antennas receive noise just as well as they receive signals. The signal arrives in one or a few directions, whereas noise arrives from all directions as indicated in Fig. 4-2a. The peak of the receive antenna beam is directed toward the signal arrival direction and the pattern, as much as possible, is controlled to reduce response to noise. Any materials in the environment surrounding the antenna will radiate electromagnetic waves, as dictated by blackbody radiation theory. The antenna sums the arriving noise radiation, through its antenna pattern, because noise is incoherent and noise powers add, producing a noise contribution to the output along with the signal. The equivalent terminal behavior is modeled in Fig. 4-2b by considering the radiation resistance of the antenna to be a noisy resistor at temperature T_A such that the same output noise power from the antenna in the actual environment is produced. The *antenna temperature T_A* is not the actual physical temperature of the antenna but is an equivalent temperature that produces same noise power, P_{NA}, as the antenna operating in its surroundings. Antenna noise temperature is often much larger than the ambient

(a) Antenna receiving noise from directions (θ, ϕ)
producing antenna temperature T_A.

(b) Equivalent model.
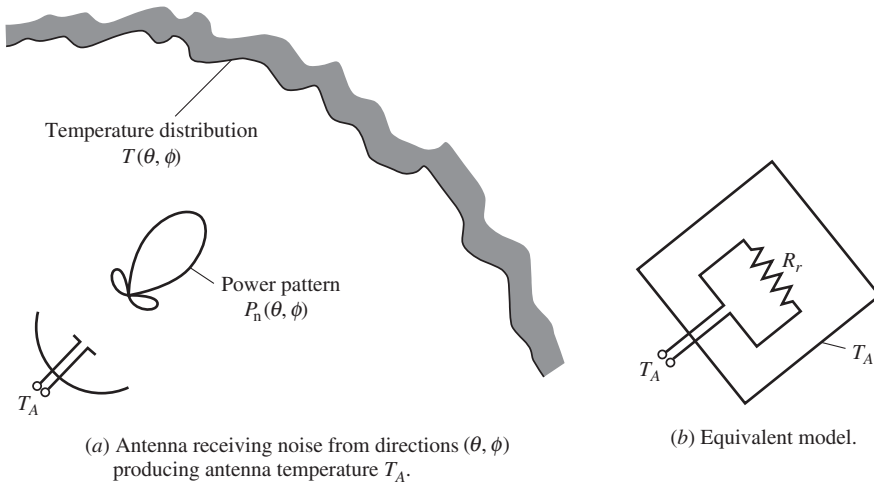
Figure 4-2 Antenna temperature.

temperature of an antenna. From noise theory, the noise power available from a noise resistor in bandwidth Δf at temperature T is

$$P_N = k\, T\, \Delta f \tag{4-15}$$

where

   P_N = available power due to noise [W]
   k = Boltzmann's constant = 1.38×10^{-23} J K^{-1}
   T = noise temperature [K]
   Δf = receiver bandwidth [Hz]

Such noise is referred to as *Nyquist noise* or *Johnson noise.* The antenna output noise power, P_{NA}, is found from this equation using the antenna noise temperature, T_A. The total system noise power is the sum of the antenna output noise and the input receiver noise power. The noise temperatures add also and the total system noise temperature is found as

$$T_{\mathrm{sys}} = T_A + T_r \tag{4-16}$$

where T_r is the receiver input noise temperature.

   There are two motivations for studying the noise performance of antennas: for noise corruption to communications and to active remote-sensing systems, and for noise calculations of passive remote-sensing systems. In most communications and active remote-sensing systems, the noise power must be sufficiently below the signal power for proper operation. This is especially true in long-distance communication systems where the arriving signals are weak. In these cases, the system is evaluated using "carrier-to-noise ratio" CNR, which is the ratio of the signal power to the system noise power:

$$\mathrm{CNR} = \frac{P_D}{P_{\mathrm{Nsys}}} \qquad \textit{carrier-to-noise ratio} \tag{4-17}$$

where

   P_D = carrier power delivered from the antenna [W]
   P_{Nsys} = system noise power = $k\, T_{\mathrm{sys}}\, \Delta f$ [W]

and T_{sys} is found from (4-16). P_D is the average modulated carrier power. The parameter closely related to CNR is signal-to-noise ratio SNR which uses the signal power after demodulation; however, the term SNR is often used where CNR is intended.

A *radiometer* is a receiving system used to measure noise in passive remote-sensing applications. It picks up noise from "hot" objects and can form an image by scanning through a noise scene, usually with a narrow-beam antenna. This is an unusual situation where an instrument tries to distinguish noise radiated by a desired object from other noise that is unwanted. An example is a satellite microwave radiometer viewing precipitation on Earth's surface from space. Water particles are lossy and produce noise just as a resistor does. The outline and intensity of a storm is imaged by detecting noise from the storm which is distinguished from the background by its different noise temperature. Radiometers are also used in radio astronomy to measure noise from celestial objects such as the sun, planets, stars, and galaxies, referred to as "radio stars." Many celestial objects have radiation over a continuous large band of frequencies and the power density is quantified in a unit that includes per Hz called a *jansky* ($= 10^{-26} \text{ Wm}^{-2} \text{ Hz}^{-1}$) in honor of Karl Jansky (1905–1950), who first measured extraterrestrial noise during an accidental radio astronomy observation in 1933.

Noise power is found by first evaluating antenna temperature. As seen in Fig. 4-2a, T_A is found by the antenna collecting noise through the scene temperature distribution $T(\theta,\phi)$ weighted by the response function of the antenna, the normalized power pattern $P_n(\theta,\phi)$; a power pattern is used because noise adds on a power basis. This is expressed mathematically by integrating over the temperature distribution:

$$T_A = \frac{1}{\Omega_A} \int_0^\pi \int_0^{2\pi} T(\theta,\phi) P_n(\theta,\phi) d\Omega \quad \textit{antenna temperature} \tag{4-18}$$

Fig. 4-2a shows an Earth terminal antenna looking at the sky but (4-18) is completely general. If the scene is of constant temperature T_o over all angles, T_o comes outside the integral in (4-18) and $T_A = T_o$, which is an expected result. In this case, the antenna is completely surrounded by noise of temperature T_o and its output antenna temperature equals T_o independent of the antenna pattern shape. Consider the case of a discrete source of small solid angular extent Ω_s with constant temperature T_s and observed with an antenna beam directed toward its center that is broad compared to the source. Then $P_n(\theta,\phi) \approx 1$ over the source and (4-18) reduces to

$$T_A = \frac{\Omega_s}{\Omega_A} T_s \quad \textit{small discrete source} \tag{4-19}$$

For a circularly symmetric source or circularly symmetric antenna beam, the solid angles Ω_S and Ω_A are proportional to the angular extent of the source squared and HP^2, respectively.

The antenna noise power P_{NA} is found from (4-15) using T_A from (4-18) once the temperature distribution $T(\theta,\phi)$ is determined. Of course, this depends on the scene, but in general $T(\theta,\phi)$ consists of two components: sky noise and ground noise. The noise temperature of ground in most situations is well approximated by 290 K, but is much less for reflecting surfaces and smooth surfaces at or near grazing incidence angles.

Unlike ground noise, sky noise is a strong function of frequency as shown in Fig. 4-3, which is the noise temperature available from an antenna, not including losses on the antenna and ground noise pickup, as a function of frequency. Sky noise is made up of atmospheric, cosmic (extraterrestrial), and manmade noise; see [2] for a review of natural noise. "Galactic noise" from our own Milky Way galaxy is especially strong in directions toward the galactic center. Below about 10 MHz external noise is very strong and usually exceeds the internal noise of a receiver, so antenna efficiency and directivity
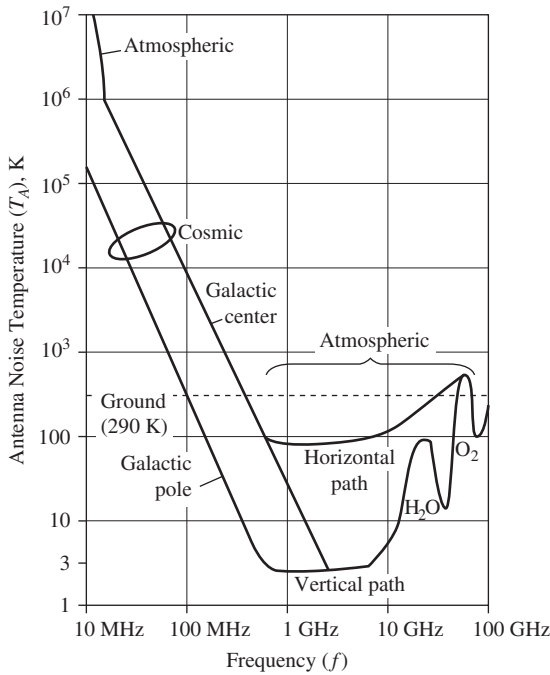
Figure 4-3 Sky noise contribution to antenna temperature as a function of operating frequency and pointing angle.

have little effect on signal-to-noise ratio. Thus, inefficient antennas are acceptable for reception, such as the ferrite loop antenna used in the AM radio band. At these frequencies, the dominant source of noise is atmospheric, arising mainly from lightning, which propagates over large distances via reflection from the ionosphere. Manmade noise produced by power lines, electric motors, and so on also can be significant, especially in urban areas. From 10 to 100 MHz, natural noise is a combination of atmospheric and cosmic noise (galactic and solar). From 100 MHz to 1 GHz, cosmic noise dominates. The band between 1 and 10 GHz is the lowest noise regime and noise increases with decreasing elevation angle due to the finite height of the atmosphere with the path length through the atmosphere becoming longer for lower elevation angles. Above 10 GHz atmospheric noise again dominates, increasing with frequency due to water vapor and hydrometeor absorption, which vary with season and location. Atmospheric gases also are noise sources and include strong, broad spectral lines, most notably the water vapor and oxygen lines at 22 and 60 GHz, respectively.

The important observation from Fig. 4-3 is the "bathtub" type dip from 1 to 10 GHz. The low noise, along with wide available signal bandwidth as well, makes this frequency region very desirable for many wireless applications. Also evident in this band is the irreducible sky background temperature of 3 K, which is theorized to be the residual effect of a "big bang" explosion that created the universe. This observation was made by Arthur Penzias and Robert Wilson in 1965 while doing radio astronomy at 4 GHz; they won the Nobel Prize for the work. An interesting side note is that searches for extraterrestrial life focus on these frequencies under the theory that extraterrestrial beings would know that humans should have receivers operating in this range because of low noise and low atmospheric loss.

Of course, the antenna pattern shape and direction strongly influence antenna temperature as seen in (4-18). The ground noise temperature contribution to antenna noise is very low for high-gain antennas pointed skyward and with low side lobes in the direction of the earth. Broad beam antennas, on the other hand, pick up a significant amount of ground noise as well as sky noise.

There are applications for radiometry for both military and civilian use. An example is detection of vehicles under tree cover, even at night, from an airborne platform by observing either high radiation from a hot engine or low radiation from the metallic body, in comparison to the soil background radiation. A second example is a millimeter-wave radiometer used in an airport for imaging passengers to detect concealed weapons.

EXAMPLE 4-3   *Direct Broadcast Satellite Reception*

Reception of high-quality television channels at home with an inexpensive, small terminal is possible as the result of many years of technology development, including new antenna designs. DirecTv (trademark of Hughes Network Services) transmits from 12.2 to 12.7 GHz with 120 W of power and an EIRP of about 55 dBW in each 24-MHz transponder that handles five compressed digital video channels. The receiver uses a 67-K noise temperature low-noise block down-converter and when combined with antenna noise gives a system noise temperature of 125 K. Here we do the noise analysis for this satellite link and will use the result in the link calculation of Example 4-4. The needed parameter values are:

$$\Delta f = \text{effective signal bandwidth} = 20\,\text{MHz}$$
$$T_{\text{sys}} = 125\,\text{K}$$
$$P_D = -116.9\,\text{dBW, from Example 4-4}$$

The noise power is evaluated using (4-15) with the system noise temperature:

$$P_{\text{Nsys}} = kT_{\text{sys}}\Delta f = 1.38 \times 10^{-23} \cdot 125 \cdot 20 \times 10^6 = 2.45 \times 10^{-14} = -134.6\,\text{dBW} \quad (4\text{-}20)$$

Thus, the carrier to noise ratio from (4-17) is

$$\text{CNR(dB)} = P_D(\text{dBW}) - P_{\text{Nsys}}(\text{dBW}) = -116.9 - (-134.9) = 17.7\,\text{dB} \quad (4\text{-}21)$$

which is a reasonable margin for proper operation.

4.4   ANTENNAS IN COMMUNICATION SYSTEMS

It is important to have an appreciation for the role played by antennas in their primary application area of communication links. The basic communication link model is shown in Fig. 4-4. In Sec. 4.2, we introduced methods for calculating the power output from a receiving antenna using maximum effective aperture. In this section, we model the complete link, including the distance of separation between the source and receiver, along with several loss mechanisms encountered in a typical link. We begin by revisiting the important parameters of antenna directivity and gain, and establishing some fundamental relationships.

4.4.1   Directivity, Gain, and Effective Aperture

For system calculations it is usually easier to work with directivity rather than its equivalent, maximum effective aperture. Directivity is directly proportional to maximum effective aperture and the formula connecting the two can be derived in a variety of ways. Here, we develop the relationship for a dipole, but the developed relationship is true for
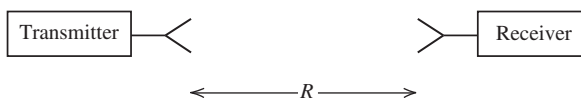
Figure 4-4 A communication link.

any antenna. A derivation for aperture antennas will be presented in Sec. 9.3. Other derivations can be found in [1] and [H.4: Collin]. Proceeding with the dipole-based derivation, the directivity of the ideal dipole, 1.5, can be written in the following manner:

$$D = \frac{3}{2} = \frac{4\pi}{\lambda^2} \frac{3}{8\pi} \lambda^2 \tag{4-22}$$

Grouping factors this way permits identification of A_{em} from (4-14), giving

$$\boxed{D = \frac{4\pi}{\lambda^2} \ A_{\text{em}}} \tag{4-23}$$

Comparing this to $D = 4\pi/\Omega_A$ from (2-144), we find

$$\boxed{\lambda^2 = A_{\text{em}} \ \Omega_A} \tag{4-24}$$

which is a general relationship that applies to any antenna. There are some interesting implications hidden in this simple formula. For a fixed wavelength (i.e., operating frequency) A_{em} and Ω_A are inversely proportional; that is, as the maximum effective aperture increases (usually as a result of increasing its physical size) the beam solid angle decreases, which means power is more concentrated in angular space. In other words, the beam solid angle decreases and as a result directivity goes up, which also follows from $D = 4\pi/\Omega_A$. Equation (4-24) also shows that for a fixed maximum effective aperture (i.e., antenna size), as wavelength decreases (frequency increases) the beam solid angle decreases, leading to increased directivity.

In practice, antennas are not completely lossless. In Sec. 2.5, we saw that the power received by an antenna is reduced by the fraction e_r (radiation efficiency) from what it would be if the antenna were lossless. Ohmic loss is included in aperture area calculations by defining *effective aperture* (or, *effective area)* as

$$A_e = e_r A_{\text{em}} \qquad \textit{effective \ aperture} \tag{4-25}$$

The expression (4-6) is generalized to include losses on the receiving antenna by replacing maximum effective aperture by effective aperture, giving the available power, P_A, as

$$P_A = S A_e \tag{4-26}$$

Again it is assumed that the peak of the receiving antenna pattern is in the direction of the incoming wave. This simple equation is very intuitive and indicates that a receiving antenna with effective aperture area A_e with units of m^2 converts incident power (flux) density in W/m^2 into power delivered to the load in W.

Gain was introduced to include losses with directivity; that is, $G = e_r D$ from (2-155). We can form a gain expression from the directivity expression by multiplying both sides of (4-23) by e_r and using (4-25) to obtain

$$G = e_r D = \frac{4\pi}{\lambda^2} e_r A_{\text{em}} = \frac{4\pi}{\lambda^2} \ A_e$$

or

$$\boxed{G = \frac{4\pi}{\lambda^2} \ A_e} \tag{4-27}$$

Ohmic losses on the antenna are included in gain, but losses associated with mismatch between the polarizations of the incident wave and receiving antenna as well as impedance mismatch between the antenna and load are accounted for separately because they are associated with how the antenna is used and are not inherent antenna characteristics.

The IEEE [H.2] defines effective area (aperture) to be in a given direction; so in general, effective aperture contains power pattern information: $A_e(\theta,\phi) = A_e|F(\theta,\phi)|^2$. Similarly, gain is often stated as a function of direction:

$$G(\theta,\phi) = \frac{4\pi}{\lambda^2}A_e(\theta,\phi) = \frac{4\pi}{\lambda^2}A_e|F(\theta,\phi)|^2 \qquad (4\text{-}28)$$

If effective aperture or gain is stated without mention of direction, it is assumed to be the maximum value. Sometimes the term *absolute gain* is used to denote maximum gain. We will show in Sec. 9.3 that effective aperture is equal to or less than the physical aperture area of the antenna A_p through *aperture efficiency* $\varepsilon_{\mathrm{ap}}$:

$$A_e = \varepsilon_{\mathrm{ap}}\, A_p \qquad (4\text{-}29)$$

where $\varepsilon_{\mathrm{ap}} \leq 1$. It is important to note that although the foregoing general relationships were developed for receiving antennas, but they also apply to transmitting antennas as ensured by reciprocity which is discussed in Sec. 13.1. Note that the effective length result in (4-12) is modified to include ohmic losses by replacing A_{em} by A_e. The foregoing relationships are used in communication system computations, which we consider next.

4.4.2   Communication Links

We are now ready to describe the power transfer in the communication link of Fig. 4-4. If the transmitting antenna were isotropic, the power density at distance R would be

$$S = \frac{U_{\mathrm{ave}}}{R^2} = \frac{P_t}{4\pi R^2} \qquad (4\text{-}30)$$

where P_t is the time-average input power accepted by the antenna, and the $1/R^2$ spherical spreading loss factor from (2-131) and (2-135) were used. In practice, the transmitting antenna with gain G_t is usually pointed in the direction of the receiver. Then the power density incident on the receiving antenna is increased from that in (4-30) by the gain, giving

$$S = \frac{G_t\, U_{\mathrm{ave}}}{R^2} = \frac{G_t\, P_t}{4\pi R^2} \qquad (4\text{-}31)$$

Using this in (4-26) gives the available received power as

$$P_r = SA_{er} = \frac{G_t\, P_t\, A_{er}}{4\pi R^2} \qquad (4\text{-}32)$$

where A_{er} is the effective aperture of the receiving antenna and we assume the receiving antenna is also pointed and polarized for maximum response. Modifying the formula to explicitly show both antenna gains by using (4-27), $A_{\mathrm{er}} = G_r\, \lambda^2/4\pi$ gives

$$\boxed{P_r = P_t \frac{G_t\, G_r\, \lambda^2}{(4\pi R)^2}} \qquad \textit{Friis transmission formula} \qquad (4\text{-}33)$$

This formula gives the available power in terms of the transmitted power, antenna gains, and wavelength. It is the most popular form of the *Friis transmission formula* and is the

basis for communications analysis. There are several built-in assumptions. The most significant assumption is a clear line-of-sight (LOS) path with no secondary wave paths caused by reflections from objects (free-space conditions). It assumes the antennas are aimed toward each other to realize the maximum gain of each. It also assumes that the transmitting and receiving antennas are matched in impedance to their connecting transmission lines, and have identical and aligned polarizations. A real communication link rarely meets all of these assumptions, but it is a simple matter to correct for the loss introduced by impedance mismatch, polarization mismatch, or antenna misalignment. Accurate inclusion of non–free space propagation conditions is more involved.

A convenient dB-form of the link equation (4-33) is obtained by taking 10 log of both sides:

$$P_r(\text{dBm}) = P_t(\text{dBm}) + G_t(\text{dB}) + G_r(\text{dB}) - 20 \log \; R(\text{km})$$
$$- 20 \log \; f(\text{MHz}) - 32.44 \tag{4-34}$$

The unit dBm is power in decibels above a milliwatt; for example, 30 dBm is 1 W. Often the powers in this equation are expressed in units of decibels above a watt, dBW. $G_t(\text{dB})$ and $G_r(\text{dB})$ are the transmit and receive antenna gains in decibels, $R(\text{km})$ is the distance between the transmitter and receiver in kilometers, and $f(\text{MHz})$ is the frequency in megahertz. Often the factors $\lambda^2/(4\pi R)^2$ in (4-33) are taken as a whole and defined to be *free space loss*. The dB form of free space loss is $L_{\text{fs}} = 20 \; R(\text{km}) - 20 \log \; f(\text{MHz}) - 32.44$, which is the last three terms in (4-34). Free space loss contains the $1/R^2$ factor associated with spherical spreading loss of a free space wave. However, it also contains λ^2 and we know free space is not frequency-dependent; thus, the term free space loss is misleading. Free space loss includes all losses in the basic link equation of (4-34) for the case of isotropic antennas (i.e., the antenna gains are both 0 dB).

Antenna misalignment is included in the Friis equation by using the gain value in the direction toward the other antenna using (4-28) for each antenna. The effect of polarization and impedance mismatch are contained in multiplicative constants; thus, the power delivered to the terminating impedance is given by

$$\boxed{P_D = pq \; P_r} \tag{4-35}$$

where
   P_D = power delivered from the antenna
   P_r = power available from the receiving antenna
   p = polarization efficiency (or polarization mismatch factor), $0 \le p \le 1$
   q = impedance mismatch factor, $0 \le q \le 1$

It is convenient to express (4-34) in dB form:

$$P_D(\text{dBm}) = 10 \log \; p + 10 \log \; q + P_r(\text{dBm}) \tag{4-36}$$

These mismatch losses will be discussed in detail following the topic of EIRP. The term *realized gain* is used to include impedance mismatch loss in antenna gain and equals qG.

4.4.3   Effective Isotropically Radiated Power (EIRP)

A frequently used concept in communication systems is that of *effective (or equivalent) isotropically radiated power,* EIRP, which is the amount of power emitted from an isotropic antenna to obtain the same power density in the direction of the antenna pattern peak with gain G_t. EIRP is simply the gain of the transmitting antenna multiplied by the net power accepted by the antenna from the connected transmitter:

(a) Directional antenna with input power P_t and gain G_t.

(b) Isotropic antenna with input power $P_t G_t$ and unity gain.
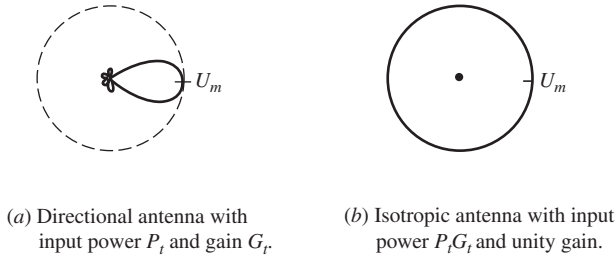
Figure 4-5 Illustration of effective isotropically radiated power, EIRP. In both (a) and (b), EIRP $= 4\pi U_m$.

$$\text{EIRP} = P_t G_t \qquad (4\text{-}37)$$

As illustrated in Fig. 4-5a, the maximum radiation intensity U_m from a transmitting antenna with input power P_t is in the direction of maximum radiation; see (2-133). The gain of the antenna from (2-152) is $G_t = 4\pi U_m / P_t$, so EIRP $= P_t(4\pi U_m / P_t) = 4\pi U_m$. The same radiation intensity U_m would be obtained from a lossless isotropic antenna (gain $G_t = 1$) with input power $P_t G_t$ as illustrated in Fig. 4-5b. In other words, to obtain the same radiation intensity produced by the directional antenna in its pattern maximum direction, an isotropic antenna would have to have an input power G_t times greater. The term *effective radiated power*, ERP is used in the broadcast industry (FM and TV) referenced to a half-wave dipole and includes losses in the transmission line from the transmitter to the antenna. However, ERP is often used in place of EIRP as defined above. For example, an FM radio station with a 100,000-W transmitter feeding a 7 dB gain antenna through a transmission line with 5 dB loss will have an ERP of $50 - 5 + 7 - 2.15 = 49.85$ dBW. EIRP is commonly used in the satellite communications industry, as illustrated in the following example.

EXAMPLE 4-4   *Direct Broadcast Satellite Reception (continued)*

This example completes the satellite TV receiver case study of Example 4-3 in which the noise power was calculated. Each 24-MHz transponder handles several compressed digital video channels and transmits 120 W of power from the satellite. The EIRP is stated to be 55 dBW. The receiving system uses a 0.46 m (18 in) diameter offset fed reflector antenna. Here, we find the received signal power using the following link parameter values:

$f = 12.45$ GHz (middle of the 12.2 to 12.7 GHz band)

$P_t(\text{dBW}) = 20.8$ dBW(120 W)

$G_t(\text{dB}) = \text{EIRP(dBW)} - P_t(\text{dBW}) = 55 - 20.8 = 34.2$ dB   using (4-37)

$R = 38,000$ km (typical slant path length)

$$G_r = \frac{4\pi}{\lambda^2} \varepsilon_{\text{ap}} A_p = \frac{4\pi}{(0.024)^2} 0.7 \left(\pi \frac{0.46}{4} \right)^2 = 2538 = 34 \text{ dB} \qquad (70\% \text{ aperture efficiency})$$

The power delivered from the antenna assuming matched conditions from (4-34) is

$$
\begin{aligned}
P_D(\text{dBW}) &= P_t(\text{dBW}) + G_t(\text{dB}) + G_r(\text{dB}) - 20 \log R(\text{km}) - 20 \log f(\text{MHz}) - 32.44 \\
&= 20.8 + 34.2 + 34 - 20 \log(38,000) - 20 \log(12450) - 32.44 \\
&= 20.8 + 34.2 + 34 - 91.6 - 81.9 - 32.4 \\
&= -116.9 \text{ dBW} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (4\text{-}38)
\end{aligned}
$$

This is a power level of only 2×10^{-12} W! Without the high gains of the antennas (68.2 dB combined) this signal would be hopelessly lost in noise.

4.4.4   Impedance Mismatch

The maximum power received occurs for the load impedance matched to the antenna impedance, which is the model of Fig. 4-1*b* for conditions in (4-9). For a general load impedance condition, (4-8) applies. The ratio of these gives the impedance mismatch factor *q*:

$$q = \frac{P_{\text{rm}}}{P_{\text{rm,matched}}} = \frac{4R_A R_L}{(R_A + R_L)^2 + (X_A + X_L)^2} \tag{4-39}$$

For the matched-impedance case of (4-7), this reduces to $q = 1$ giving no mismatch loss as expected. In the usual situation of a transmission line of characteristic impedance Z_o connected to the antenna

$$q = \frac{4R_A Z_o}{(R_A + Z_o)^2 + X_A^2} \quad (R_T = Z_o, X_T = 0) \tag{4-40}$$

In many cases it is convenient to obtain *q* by first finding the *voltage reflection coefficient, Γ,* by noting that the received power is the incident power (normalized to unity) less the relative reflected power:

$$q = 1 - |\Gamma|^2 \tag{4-41}$$

And Γ is found either from measurement or from calculation using

$$\Gamma = \frac{Z_L - Z_A^*}{Z_L + Z_A} \tag{4-42}$$

For an antenna that is conjugate matched to the load as in (4-7), the preceding two equations reduce to the expected results of $\Gamma = 0$ and $q = 1$, and there is no mismatch loss. The magnitude of the reflection coefficient can also be found from measurement of the *voltage standing wave ratio,* VSWR:

$$|\Gamma| = \frac{\text{VSWR} - 1}{\text{VSWR} + 1} \tag{4-43}$$

When the antenna is matched to the transmission line, VSWR $= 1$ and then $q = 1$. When there is a large mismatch, VSWR is large and *q* approaches zero. Measurement instruments commonly display *return loss,* RL, as a way to quantify reflections. It has units of dB and is related to reflection coefficient as

$$\text{RL} = -20 \log |\Gamma| \quad [\text{dB}] \qquad \textit{return loss} \tag{4-44}$$

For matched conditions $\Gamma = 0$ and RL $= \infty$; and infinite return loss means that all reflections of the antenna signal are "lost" in the load. A typical impedance match antenna specification is VSWR ≤ 2, corresponding to $|\Gamma| \leq 0.333$ and RL ≥ 9.5 dB. By reciprocity, the impedance mismatch results are equally applicable to transmitting and receiving cases. Reemphasizing an important point, impedance mismatch is not included in gain or effective aperture but is calculated separately because it is not inherent to the antenna but instead depends on how the antenna is used. If, for example, gain assumed a connecting 50-Ω load, it would be difficult to compute gain if other loads were used.

4.4.5 Polarization Mismatch

Just as with impedance mismatch, gain and effective aperture do not include polarization mismatch, and it must be evaluated separately. Polarization principles were introduced in Sec. 2.8 and we can use them to make antenna-wave interaction calculations; for complete details see [H.11.1: Stutzman, Chap. 6]. Stating the obvious, the polarization of a radiated wave is that of the transmitting antenna, which in general varies with pattern angle. The wave that arrives at a receiving antenna is matched to the polarization of the receiving antenna if the wave and the antenna polarizations have the following the same: axial ratios, senses (of polarization rotation), and tilt angles (i.e., the same spatial orientation of their polarization ellipse major axes). The *polarization efficiency* (or *polarization mismatch factor*), *p*, varies from 0 to 1 as the incoming wave and receiving antenna vary from completely mismatched in polarization to completely matched, respectively. A complete match ($p = 1$) exists when the wave and antenna polarization states are identical. A complete mismatch ($p = 0$) occurs when the wave and antenna are *cross-polarized.* Fig. 13-14 illustrates co- and cross-polarized situations. Examples of cross-polarized states are orthogonal linear states such as horizontal and vertical linear polarizations, and right- and left-hand circular polarizations. Polarization mismatch is a loss in the sense that less power is captured from the wave when the antenna is polarization mismatched to the wave compared to when matched. In some cases, the loss can be recovered by adjustment. For example, matching polarizations of the transmit and receive antennas on a link that have the same sense and axial ratio is simply a matter of rotating one antenna until the polarization ellipse major axes are aligned. Polarization information for the wave and antenna are contained in the wave field $\mathbf{E}^i$ and the receiving antenna vector effective length $\mathbf{h}$, respectively. The available received power is proportional to the received terminal voltage squared, which from (4-2) is $|\mathbf{E}^i \cdot \mathbf{h}^*|^2$. Normalizing by the maximum of this expression yields the fraction of power received gives *p*:

$$p = \frac{|\mathbf{E}^i \cdot \mathbf{h}^*|^2}{|\mathbf{E}^i|^2 \cdot |\mathbf{h}^*|^2} = |\hat{\mathbf{e}}^i \cdot \hat{\mathbf{h}}^*|^2 \tag{4-45}$$

where $\hat{\mathbf{e}}^i$ and $\hat{\mathbf{h}}^*$ are the complex unit vectors for the incident wave and the receiving antenna vector effective length, respectively. $\hat{\mathbf{e}}^i$ equals the polarization of the distant transmitting antenna, $\hat{\mathbf{h}}_t$, if there is a clear-line-of-sight and the intervening propagation medium does not depolarize the wave. Antenna vector effective length is defined for the transmitting situation. The conjugate in (4-45) essentially reverses the direction of travel for the receive antenna polarization reference frame and acts to place both the wave and the receiving antenna in the same receive-antenna coordinate system. If the conjugate is not used, a separate manipulation is required to account for this effect, making the calculation error prone. When using (4-45) one can write the antenna vector effective length as if the antenna were transmitting. A helpful analogy for this use of relative coordinate systems is two people shaking hands. Facing each other, they are "matched" when both use their right hands to shake even though the hands are on opposite sides. In the same fashion, a right-hand circularly polarized (RHCP) wave is matched to a RHCP receiving antenna. A classic example of confusing the senses of CP occurred with first transatlantic TV transmission on July 11, 1962, with the Telstar satellite. Receiving stations in England and France were both attempting to receive the CP signal transmitted from an Earth station in the state of Maine, in the United States. The English misunderstood the CP sense and were mismatched in polarization, but the French receiving station was properly polarized, and they received the signal first.

The wave and the antenna polarization unit vectors are written using the coordinate system of Fig. 4-6 and following the form of (2-188):

$$\hat{\mathbf{e}} = \cos\gamma_i \hat{\mathbf{x}} + \sin\gamma_i e^{j\delta_i} \hat{\mathbf{y}} \tag{4-46a}$$

$$\hat{\mathbf{h}} = \cos \gamma \hat{\mathbf{x}} + \sin \gamma e^{j\delta} \hat{\mathbf{y}} \qquad (4\text{-}46b)$$

where (γ_i, δ_i) and (γ, δ) are the polarization angles for the incoming wave and for the antenna in the arrival direction; see Fig. 2-20. The process of evaluating polarization efficiency is illustrated in the following examples.

EXAMPLE 4-5   *Reception of an LP Wave with an LP Antenna*

A linearly polarized (LP) incident wave with a tilt angle of τ_i illuminates a LP antenna at tilt angle τ as shown in Fig. 4-6a where a dipole is used to illustrate a general LP antenna. The wave arrives normal to the plane of the dipole antenna (*xy*-plane), corresponding to the usual operating situation for a receiving antenna. For the LP case, $\gamma_i = \tau_i$ and $\gamma = \tau$, which follow from (2-189) with $\varepsilon = 0$ and $\varepsilon_i = 0$. From (2-190), $\delta_i = 0$ and $\delta = 0$. Substituting these values in (4-46) permits evaluation of (4-45):

$$p = |\hat{\mathbf{e}}^i \cdot \hat{\mathbf{h}}^*| = |(\cos \tau_i \, \hat{\mathbf{x}} + \sin \tau_i \, \hat{\mathbf{y}}) \cdot (\cos \tau \, \hat{\mathbf{x}} + \sin \tau \, \hat{\mathbf{y}})|^2$$

$$= |\cos \tau_i \cos \tau + \sin \tau_i \sin \tau|^2 = \cos^2(\tau_i - \tau) = \cos^2(\Delta \tau) \qquad (4\text{-}47)$$

Thus, polarization efficiency is a function only of the relative tilt angle $\Delta \tau$ for the case when both the wave and antenna are linearly polarized. When $\Delta \tau = 0°$ the wave and antenna are aligned (e.g., $\mathbf{E}^i$ is parallel to the dipole), and the wave and antenna are co-polarized (i.e., polarization matched) and $p = 1$. When the wave and the antenna are orthogonal, $\Delta \tau = 90°$, (4-47) yields $p = 0$. Then the receiving antenna produces no output and the wave and antenna are cross-polarized. In practice, antennas are not perfectly polarized, and there will be a

Receiving antenna (LP)   Incoming wave (LP)

(*a*) LP antenna and LP wave

Receiving antenna (RHCP)   Incoming wave (RHCP)

(*b*) RHCP antenna and RHCP wave
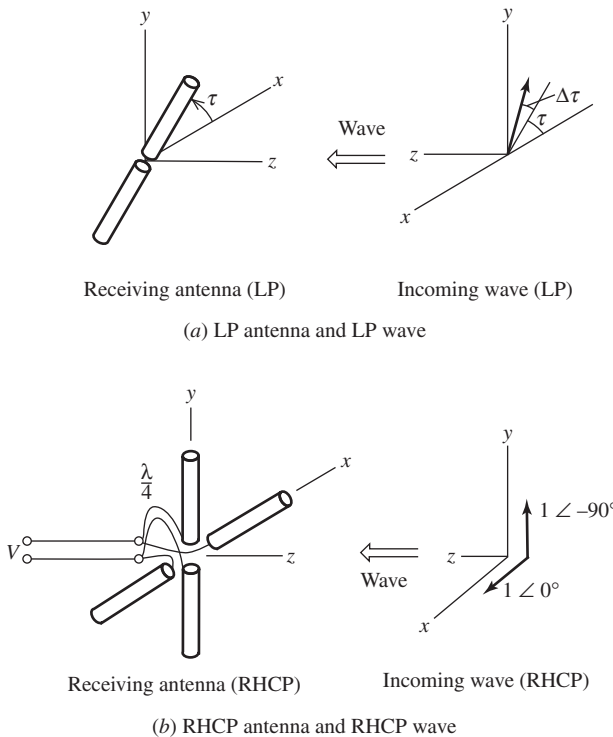
Figure 4-6 Illustration of reception of an incident wave with electric field $\mathbf{E}^i$ by a receiving antenna.

response, albeit small, to a wave state that is nominally cross-polarized to the antenna polarization state.

EXAMPLE 4-6   *Reception of CP Wave with a CP Antenna*

Calculations with circular polarization (CP) on a link reveal the value of the complex vector formation of polarization states and the role of the complex conjugate in (4-45). Consider a right-hand circularly polarized (RHCP) receiving antenna illustrated in Fig. 4-6b as crossed dipoles with a quarter-wave delay line. When operated as a transmitting antenna, this antenna has a polarization state in the $+z$-direction given by

$$\hat{\mathbf{h}} = \frac{1}{\sqrt{2}}(\hat{\mathbf{x}} - j\hat{\mathbf{y}}) \quad RHCP \tag{4-48}$$

because the magnitudes of the x- and y-components are equal and the y-component lags the x-component by 90°. This result also follows from (4-46b) with $\gamma = 45°, \delta = -90°$; see Sec. 2.8. Similarly, for a RHCP incident wave

$$\hat{\mathbf{e}}^i = \frac{1}{\sqrt{2}}(\hat{\mathbf{x}} - j\hat{\mathbf{y}}) \quad RHCP \tag{4-49}$$

The polarization efficiency from (4-45) is then

$$p = |\hat{\mathbf{e}}^i \cdot \hat{\mathbf{h}}*| = \left| \frac{1}{\sqrt{2}}(\hat{\mathbf{x}} - j\hat{\mathbf{y}}) \cdot \frac{1}{\sqrt{2}}(\hat{\mathbf{x}} - j\hat{\mathbf{y}})* \right| = 1 \tag{4-50}$$

and the wave is perfectly matched to the antenna. This result can also be explained by examining how the antenna responds to the incoming wave. The x-dipole produces a voltage of $1 \angle 180°$; the 180° is included because of the opposite reference direction of the x-axes of the wave and antenna. The y-dipole is excited by $1 \angle 90°$ and its output is delayed by 90° due to the quarter-wavelength transmission line section, producing a net $1 \angle 180°$ excitation at the connecting transmission line terminals. Combining the voltages from the two dipoles gives $2 \angle 180°$, indicating complete reinforcement of the x- and y-components. Therefore, the antenna is matched to the wave. Note that if the wave is left-hand CP, then the phase of the y-component of the wave is $+90°$ rather than $-90°$ and there is complete cancellation at the transmission line. The sign of the $\hat{\mathbf{y}}$-term in (4-49) would be positive and $p = 0$, indicating a cross-polarized situation.

EXAMPLE 4-7   *Reception of an LP wave by a CP antenna*

The LP wave of Fig. 4-6a incident on the CP antenna of Fig. 4-6b has a polarization efficiency evaluated using $\hat{\mathbf{e}}^i$ from Example 4.5 and $\hat{\mathbf{h}}$ from Example 4-6:

$$p = |\hat{\mathbf{e}}^i \cdot \hat{\mathbf{h}}^*|^2 = \left| (\cos \tau_i \, \hat{\mathbf{x}} + \sin \tau_i \, \hat{\mathbf{y}}) \cdot \frac{1}{\sqrt{2}}(\hat{\mathbf{x}} - j\hat{\mathbf{y}})^* \right|^2 = \frac{1}{2} |\cos \tau_i + j \sin \tau_i|^2 = \frac{1}{2} \tag{4-51}$$

Thus, one half of the power available from an LP wave is lost when received by a CP antenna. The same is true for a CP wave and LP antenna. In many system applications, a 3-dB loss is significant and an antenna matched to the wave must be used. On the other hand, there are operational links with one antenna linear and the other circular. For example, the wave from a

linearly polarized antenna on a spacecraft will rotate due to spacecraft motion or Faraday rotation in the ionosphere, but if a circularly polarized receive antenna is used, the incoming linearly polarized wave orientation angle will not lead to power level fluctuations. Even though a 3-dB signal loss is encountered, the received signal remains constant. If two linearly polarized antennas are used as in Example 4-5 on a satellite-to-earth link, there is the possibility of significant polarization loss because p varies from 1 to 0 as $\Delta\tau$ varies from $0°$ to $90°$, or a polarization tracking system must be used to maintain alignment.

4.5   ANTENNAS IN WIRELESS COMMUNICATION SYSTEMS

Wireless communications, or simply *wireless*, is a broad term including any electronic communication means that does not employ wires connecting the terminals. *Radio communications* refers to systems employing radios and is part of wireless communications, although the terms are often used interchangeably. In the general setting, the term wireless is intended to include systems not using radio technology such as those using infrared or ultrasonic techniques. Originally the term wireless was popular in its infancy, over a hundred years ago, followed by the term radio that lasted through most of the 20th century only to see wireless return to popularity at the end of the century (see Sec. 1.1 for a history of radio and wireless communications). The term *radio* is also used to mean a device such as an electronic receiver or transceiver.

Major application areas for wireless are: communication of voice, video, and data; position location; identification; paging; control; and medical. So-called *location-based services* expanded rapidly with the deployment of satellite constellations that provide signals for inexpensive mobile terminals to self-determine their location at no recurring cost. GPS (Global Positioning Satellite) is an example. Such services provide low-cost determination of position, but the position is only known at the terminal. A return link must be added to communicate the terminal's position to another location. Common ways to do this are to use a cellular or satellite link. Many innovative wireless applications continue to emerge in great numbers. One interesting example application in the medical profession is wireless capsule endoscopy that uses a swallowable pill with a camera for imaging the intestinal tract and telemetering the data to a body-worn receiving array.

4.5.1   Spatial Frequency Reuse and Cellular Systems

The frequency spectrum is the natural resource for wireless technology. It is finite and must be shared and reused to maximize coverage and capacity on the network. Frequencies can be reused by using different locations, angles, or polarizations at the same time on the same frequency, but the most common method for reusing frequencies is to separate same-frequency channels spatially (spatial frequency reuse), which is a technique referred to as *cellular* wireless communications.

Terminals can be fixed, portable, or mobile. Fixed terminals are stationary, whereas mobile terminals communicate on the move. Traditionally, the term portable is used for terminals that can easily be moved between operating sites. In wireless communications, portable refers to a handheld terminal used by an operator while walking and mobile implies operation at vehicular speeds. Platforms for terminals can be terrestrial, air, ship, or satellite.

Fig. 4-7 shows that the sizes of the individual cells vary greatly in serving regions spanning large-size scales, as required for the universal goal of global connectivity anywhere, anytime with anyone. [3, 4] Wireless is undergoing a convergence of services in various devices. Smart cell phones, for example, include multiple voice and data bands (cellular, GSM, DCS, PCS), position location (GPS), wireless local area network (WiFi), mobile TV, satellite communications, and so on. Of course, antennas are required for each of these physical communication links. Especially challenging is the antenna design for small handheld terminals required to support several communication bands and services.
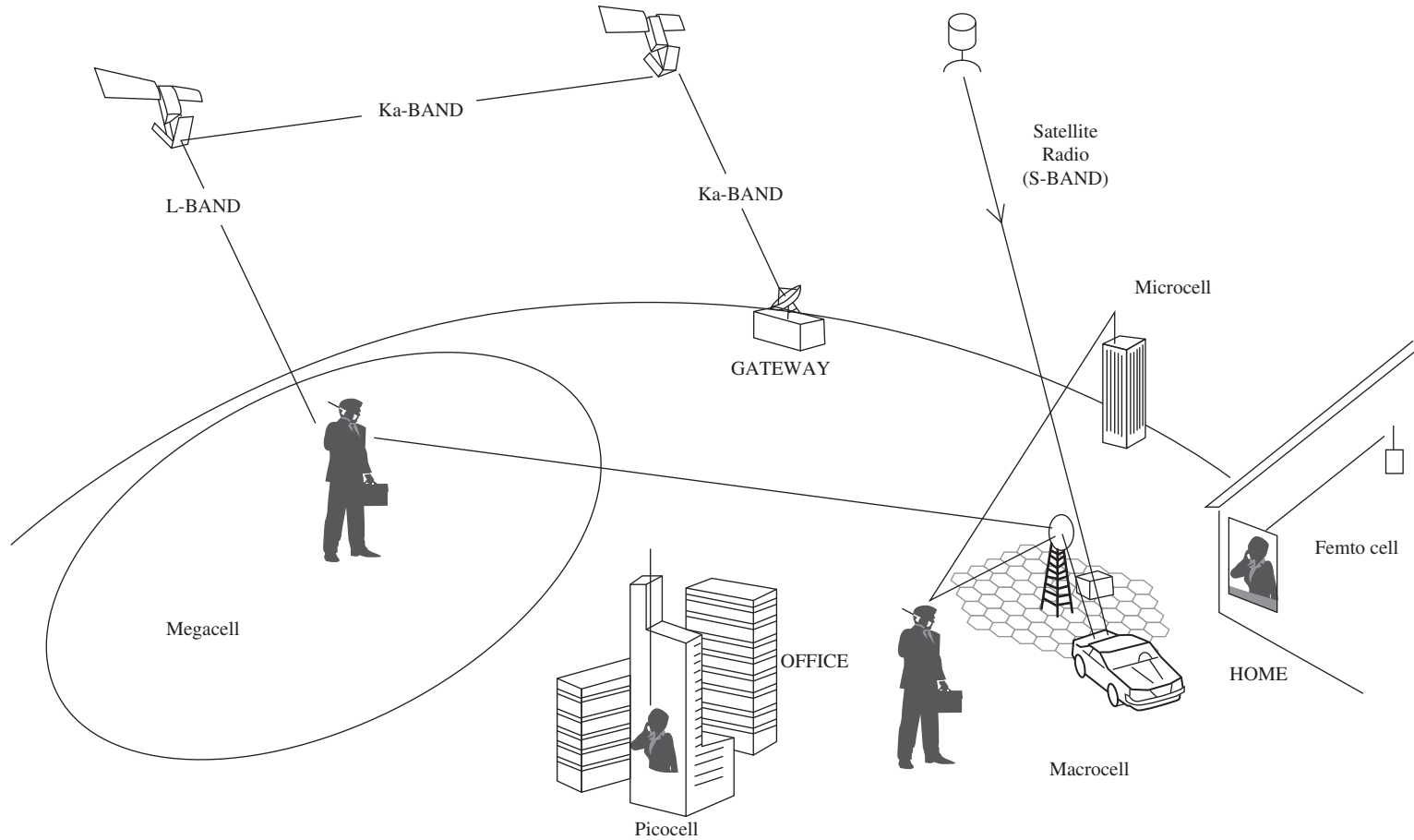
Figure 4-7 Depiction of multimode global communications with various size cells that include satellite-to-user links, inter-satellite links, feeder links to satellites, terrestrial grids, and in-building networks.

This topic and cellular base stations are discussed in detail in Chap. 12. Examples of the cell size scales listed below are depicted in Fig. 4-7.

- *Megacell*
    A megacell serves areas as large as continents using a satellite platform as the "base station." Terminals can be on planes, ships, or land (mobile, portable, or fixed).
- *Macrocell* (for WWAN, wireless wide area networks)
    Macrocellular is conventional cellular service used in rural and urban areas with base station antennas on large towers or buildings above the surrounding terrain. The coverage radius can be as much as 30 km.
- *Microcell* (for WMAN, wireless metropolitan area networks)
    Microcellular base stations have lower power and smaller cells than macrocells and are found in urban areas. Often the antennas are mounted on the side of a building or on a utility pole. The cells have a radius on the order of 1 km or less. Example services are for campuses and malls.
- *Picocell* (for WLAN, wireless local area networks)
    Buildings, small campuses, airports, train cars, and airplane interiors are examples of picocells. Wi-Fi (Wireless Fidelity) systems are probably the most prominent application for picocells.
- *Femtocell*
    Traditional cell phone service is extended to homes and other small in-building areas with a femtocell.
- *Personal Area Cell* (for WPAN, wireless personal area networks)
    Personal area cells are used primarily for communicating among computing equipment, cell phones, and so on. The range is only a few meters. Example implementing technologies are Bluetooth and UWB (Ultra Wideband).
- *Body Area Cell* (for WBAN, wireless body area networks)
    Wearable and implantable wireless devices are used in WBAN for monitoring body parameters and transmitting data to nearby terminals. The technology is emerging and many applications will evolve.

One-way links that have a single transmitter and multiple receivers are referred to as *broadcast communications.* AM and FM Broadcast radio are good examples. Broadcast involves a large, fixed, expensive transmitter and inexpensive receivers that can even be mobile. Communication links that are two-way are called *duplex* systems. One or more base stations that serve many transceivers in both directions are referred to as *point-to-multipoint communications,* such as cellular telephone. Finally, *point-to-point communications* is used to connect two sites directly, either one-way or two-way. A garage door opener is an example of one-way point-to-point communications. Other important terminology for links is *forward* and *reverse* links for base station-to-user and user-to-base station, respectively. Alternate terminology with origins in satellite communications is *downlink* and *uplink* for base station-to-user and user-to-base station, respectively.

Predicting the future in any technical field is always risky, but future wireless systems will certainly emphasize spectral efficiency through spectrum co-use and reuse. Modulation and coding improvements continue to be emphasized, but the physical side of wireless is involved as well. For example, the antenna function will continue its spread into more subsystem functions. In some of today's digital radios, it is hard to define exactly where the antenna is in the system. For example, in *software radio* the same hardware is reprogrammed or reconfigured for different functions at different times. This is accomplished substantially through software and the physical layer can be controlled by software. Multiple antennas will become more prevalent for use as either adaptive arrays or as multiple sensors for diversity or in a MIMO (multiple-input, multiple-output) configuration. In advanced configurations, a *cognitive radio* will sense its operating

environment and adjust to it, such as by changing operating frequency. New antennas will be required to support these and other innovations.

4.5.2   Propagation Effects on Communication Links

The basic communications principles in Sec. 4.4 assumed a clear LOS and no deleterious propagation effects. However, real wireless systems are a complicated interplay between the antenna and propagation effects, especially for mobile and portable applications where there is a dynamic propagation environment. Handheld portable terminals present special complications of antenna interaction with the small radio platform, the human operator, and the many nearby objects when indoors or in urban areas. The impact of antenna and propagation factors on systems is illustrated in Table 4-1, which compares typical mobile and portable wireless communication situations. The mobile case assumes a 5-dB gain rooftop-mounted antenna such as a 5/8-over 1/4-wavelength collinear antenna (see Sec. 12.3). The mobile case is representative of all early cellular communications and much of today's land mobile communications used by commercial and government vehicles. The portable unit is handheld and operates with low power and with a low-gain antenna because of its small size. In addition, portable terminals often are operated inside buildings and thus experience many more dB loss than the mobile terminal. The net loss experienced by the portable terminal compared to the mobile terminal in the example of Table 4-1 is 26 dB. Modern cellular systems make up for this reduced signal by using smaller cell sizes.

Path effects are included by expanding the link equation (4-34) to generalize the propagation loss, include mismatch losses, and add losses for other effects:

$$P_D(\text{dBm}) = P_t(\text{dBm}) + G_t(\text{dB}) + G_r(\text{dB}) - 10n \log R(\text{km}) - 20m \log f(\text{MHz})$$
$$- 32.44 + 10 \log(p) + 10 \log(q) - \sum_i L_i \tag{4-52}$$

where

$P_D(\text{dBm})$ = power delivered to the receiver in dBm
$P_t(\text{dBm})$ = power output from the transmitter in dBm
$G_t(\text{dB}), G_t(\text{dB})$ = gains in dB of the transmit and receive antennas, respectively
$R(\text{km})$ = distance between the transmitter and receiver in km
n = power on distance variation (n = 2 for free space)
$f(\text{MHz})$ = operating frequency in MHz
m = power on frequency variation (m = 2 for free space)

p, q = polarization and impedance mismatch factors, respectively
L_1, L_2 = loss in dB associated with heights of transmit and receive antennas, h_1 and h_2
L_3 = clear air absorption loss in dB
L_4 = hydrometeor (e.g. rain) absorption loss in dB
L_5 = loss for outdoor-to-indoor building loss in dB

Table 4-1   Comparison of Mobile and Personal Wireless Communication
Terminal Performance

Quantity	Mobile Communications	Portable Communications	Net Loss Using Portable Comm.
Antenna gain	5 dB	−4 dB	9 dB
Transmit power	3 W	0.6 W	7 dB
Building penetration loss	-	10 dB	10 dB
Total net loss for portable comm.	-	-	26 dB

For free space paths, $n = 2, m = 2$, and $L_i = 0$. For non–free space conditions, the powers n and m and losses L_1 and L_2 are found from models that are site-specific for known propagation environments, or, more commonly, from average models based on statistics of variations over location and time. A moving terminal and objects moving near the link cause variable fading by summing at the receiver of the multiple signals arriving via different paths. Average models are based on empirical data, analysis, or both. Here we give a brief discussion of some of the loss factors. More details are available in references listed in Secs. H.4 and H.9.1, and in [5].

A typical value for n is 4 for outdoor propagation and higher for shadowed or obstructed LOS paths. The increase from 2 for free space to 4 is extreme because of its exponential nature. For example, this increase corresponds to an additional loss of 26 dB on a 20-km path. For most models m is 2, the free space value. In the HF band and below, ionospheric effects must be considered. Clear-air absorption, L_3, is small below a few GHz, but increases as frequency increases with high attenuation in the water vapor and oxygen bands near 22 and 60 GHz, respectively. Rain absorption, L_4, is small below a few GHz, but is significant above 10 GHz; see [6] for a simple model for rain attenuation and [7] for experimental data in the 10 to 30 GHz range. Propagation losses for a signal passing from outside to inside a building can be several additional dB depending on the building material and frequency. [H.4: Siwiak] Polarization also influences propagation. Ground reflections are polarization dependent, which affects the indirect multipath components; see Sec. 6.7. Circular polarization can offer the advantage of natural multipath rejection because the reflected waves are of opposite sense to the incident wave and thus do not destructively interfere with the direct wave. [H.11.1: Stutzman, Sec. 8.2.3] Note that impedance mismatch can occur on both ends of a link.

4.5.3   Gain Estimation

An especially useful tool in antenna and communication system analysis is a technique to estimate antenna gain. Gain can be calculated and estimated in a variety of ways as will be treated in detail in Sec. 9.3. As we shall see, the technique that includes the most information about the particular antenna provides the most accurate gain value. But for now we present one very simple, yet powerful, estimation method.

Directivity depends only on the antenna pattern and thus can be calculated or estimated based only on pattern data. Directivity varies inversely with beam solid angle: $D = 4\pi/\Omega_A$ from (2-144). Beam solid angle is found by integrating the power pattern over all angles using (2-142). It is possible to measure or calculate many pattern cuts and numerically integrate to find Ω_A and then D. The simple approach we develop here is to use the principal plane pattern half-power beamwidths to estimate beam solid angle. This begins by defining *beam efficiency*, ε_M, as the ratio of the solid angle only in the main beam, Ω_M, to the solid angle of the entire pattern, Ω_A:

$$\varepsilon_M = \frac{\Omega_M}{\Omega_A} = \frac{\displaystyle\iint_{main\ beam} |F(\theta,\phi)|^2 d\Omega}{\displaystyle\iint_{4\pi} |F(\theta,\phi)|^2 d\Omega} \qquad beam\ efficiency \qquad (4\text{-}53)$$

The main beam solid angle is well approximated in many cases by the product of the half-power beamwidths in the principal plane patterns that pass through the main beam peak:

$$\Omega_m \approx \mathrm{HP_E HP_H} \qquad (4\text{-}54)$$

The directivity is then estimated by combining these results:

$$D = \frac{4\pi}{\Omega_A} = \frac{4\pi\varepsilon_M}{\Omega_M} \approx \frac{4\pi\varepsilon_M}{HP_E HP_H} = \frac{41{,}253\ \varepsilon_M}{HP_{E^\circ} HP_{H^\circ}} \tag{4-55}$$

where HP_{E° and HP_{H° are the half-power beamwidths in the E- and H-planes in degrees. If the pattern has no side lobes (i.e., no power outside the main beam), $\varepsilon_M = 1$ and

$$D \approx \frac{41{,}253}{HP_{E^\circ} HP_{H^\circ}} \qquad \textit{no side lobes} \tag{4-56}$$

This formula works well for low-directivity antennas. For example, a half-wave dipole has $HP_{E^\circ} = 78$ and $HP_{H^\circ} = 360$, and (4-56) yields $D \approx 1.47$, which is close to the correct value of 1.64.

Antennas other than simple elements (e.g., small antennas and half-wave dipoles) have lobes outside the main beam. A typical beam efficiency value for such antennas is about 0.63. If no loss is present (or is accounted for separately), $e_r = 1$ and the gain for typical antennas in practice is estimated, based on (4-55) with $\varepsilon_M = 0.63$, using:

$$D = e_r D \approx D \approx \frac{26{,}000}{HP_{E^\circ} HP_{H^\circ}} \qquad \textit{general use} \tag{4-57}$$

It must be emphasized that accompanying such a simple formula is the risk of misuse and lack of accuracy. It is used for estimates and usually only when just the beamwidth values are known.

EXAMPLE 4-8 **_A Cell Phone Reverse Link_**

Here we consider a typical cell phone reverse link (uplink) from a handset to a base station with a sector antenna. The handset transmits at 1900 MHz with 300 mW of power into a -1-dB gain, vertically-polarized omnidirectional antenna with a VSWR of 3. The antenna on a base station tower 8 km away has a slant $45°$ linear polarization, a $6°$ elevation beamwidth, and a $90°$ azimuth beamwidth. It is desired to calculate the power available from the receive antenna. First, the base station receive antenna gain is found using (4-57) to be $G_r = 16.8$ dB and is impedance matched. The impedance mismatch loss at the transmitter from (4-43) in (4-41) is $q = 0.75$ and the polarization mismatch loss from (4-47) is $p = 0.5$ Using a $1/R^4$ distance variation, which is typical for many propagation scenarios, in (4-52) gives

$$\begin{aligned}
P_D(\text{dBm}) = {} & P_t(\text{dBm}) + G_t(\text{dB}) + G_r(\text{dB}) - 40\log R(\text{km}) - 20\log f(\text{MHz}) - 32.44 \\
& + 10\log(0.5) + 10\log(0.75) \\
= {} & 24.8 - 1 + 16.8 - 36.1 - 65.6 - 32.44 - 3 - 1.25 = -97.8\ \text{dBm}
\end{aligned} \tag{4-58}$$

We can also compute the noise and CNR. From Fig. 4-3, the sky noise is very low at this frequency, and most antenna noise comes from ground pickup. Assuming an antenna noise of 200 K and a receiver noise of 250 K, the system noise is $T_{\text{sys}} = 450$ K. A channel bandwidth of 1.5 MHz then gives a noise power using (4-15) of

$$\begin{aligned}
P_{\text{Nsys}} = kT_{\text{sys}}\Delta f &= 1.38 \times 10^{-23}(450)1.5 \times 10^6 = 9.32 \times 10^{-15} = -140.3\ \text{dBW} \\
&= -110.3\ \text{dBm}
\end{aligned} \tag{4-59}$$

The carrier-to-noise ratio from (4-17) is

$$\text{CNR} = \frac{P_D}{P_{\text{Nsys}}} = P_D(\text{dBm}) - P_{\text{Nsys}}(\text{dBm}) = -97.8 - (-110.3) = 12.5\ \text{dB} \tag{4-60}$$

4.6 ANTENNAS IN RADAR SYSTEMS

We now turn our attention to radar. Suppose an airplane is the target of a radar as shown in Fig. 4-8. We assume that the transmit and receive antennas are collocated, forming a *monostatic radar*, and are pointed such that the pattern maxima are directed toward the target. The power density incident on the target is then

$$S^i = \frac{P_t}{4\pi R^2} G_t = \frac{P_t A_{et}}{\lambda^2 R^2} \tag{4-61}$$

where (4-27) and (4-31) were used. The power intercepted by the target is proportional to the incident power density, so

$$P^i = \sigma S^i \tag{4-62}$$

where the proportionality constant σ is the *radar cross section* RCS (m^2) and is the equivalent area of the target based on the target reradiating the incident power isotropically. Although the incident power P^i is not really scattered isotropically, we are only concerned about the power scattered in the direction of the receiver and can assume the target scatters isotropically. Because P^i appears to be scattered isotropically, the power density arriving at the receiver is

$$S^s = \frac{P^i}{4\pi R^2} \tag{4-63}$$

The power available at the receiver from (4-32) is

$$P_r = A_{er} S^s \tag{4-64}$$

Combining the above four equations gives

$$P_r = A_{er} \frac{\sigma S^i}{4\pi R^2} = P_t \frac{A_{er} A_{et}}{4\pi R^4} \frac{\sigma}{\lambda^2} \tag{4-65}$$

which is referred to as the *radar equation*. Using (4-27), we can rewrite this equation in a convenient form as

$$P_r = P_t \frac{\lambda^2 G_r G_t \sigma}{(4\pi)^3 R^4} \tag{4-66}$$

This is a basic form of the radar range equation adequate for continuous wave (CW) radars. If the transmitting and receiving antennas are the same antenna, as is usually the case, $G_r G_t = G^2$.

Combining (4-62) and (4-63) forms a definition of radar cross section, σ:

$$\sigma = \frac{4\pi R^2 S^S}{S^i}, \quad R >>> \lambda \tag{4-67}$$

Transmitter

Receiver

Target
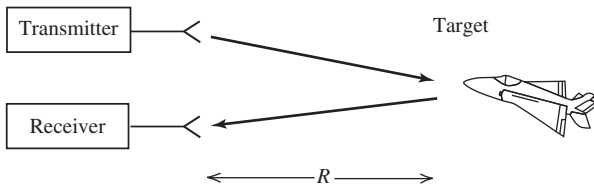
$\longleftarrow R \longrightarrow$

Figure 4-8 Radar example.

This is the ratio of 4π times the radiation intensity $R^2 S^S$ in the receiver direction, to the incident power density from the transmitter direction. The multiplication by R^2 makes σ independent of range. This is because the scattered power varies inversely with R^2 whereas the incident field is evaluated at the target.

Radar cross section for complex shaped scattering objects is a function of many variables, including incidence angle, scattering angle, frequency, polarization, and material composition. Some examples of these are in Sec. 16.16. The RCS does not bear either a direct or a simple relationship to the physical size of the radar target.

EXAMPLE 4-9   *Range of a Monostatic Radar*

Police radars are typically simple low-power CW Doppler radars that operate approximately at either 10, 24, or 35 GHz. Consider a police radar unit with the following parameters. The transmitter employs a Gunn-effect diode with a maximum output of 50 mW. The receiver employs low-noise Schottky barrier mixer diodes and requires a minimum received power of -120 dBm for proper detection. The antenna is a CP conical horn of 2.5 in. diameter with a gain of 23 dB and a HP of $12°$. An example motor vehicle has a RCS of 100 m^2. What is the maximum range of this radar for measuring the speed of the motor vehicle? From (4-66)

$$R = P_t \frac{\lambda^2 G^2 \sigma}{(4\pi)^3 P_r} = (50 \times 10^{-3}) \frac{(0.8 \times 10^{-2})^2 (2 \times 10^2)^2 (10^2)}{(4\pi)^3 (10^{-15})} = 1.59 \,\text{km}$$

In this case the maximum range for speed measurement is slightly less than one mile, which is typical for this class of radar equipment.

4.7 ANTENNAS AS UNINTENTIONAL RADIATORS

A time-varying current on a conductor will radiate. This book is mostly concerned with conductors that facilitate either transmitted RF radiation or received RF radiation. We refer to these conductors as antennas, implying they radiate and/or receive *intended* electromagnetic fields. However, there are also **un**intended electromagnetic fields that have the potential to cause *electromagnetic interference* or EMI. The study of EMI, its causes and its cures, is the field of *electromagnetic compatibility* (EMC).

Of particular interest in the 21st century are digital devices. The electromagnetic fields emitted by a digital electronic device are required by government regulation to be below a specified minimum level at a specified distance from the device. This sets emission limits. The device is also required to not be susceptible to performance degradation from exposure to unwanted electromagnetic fields. This determines susceptibility limits.

EMC has become more and more important as the result of three technological trends. First, faster processor speeds dictate that pulse rise times are shorter. This increases the energy content at increasingly higher frequencies which emit over larger distances. (Recall that displacement current is proportional to frequency). Second, solid state digital logic and signal processing circuits utilize low threshold voltages, making them susceptible to EMI. Third, physical equipment is largely built of plastic-type material, thereby decreasing the shielding inherent in the equipment itself compared to a metal case.

Electronic equipment cannot be sold unless they meet certain regulatory emission and susceptibility standards. The most cost-effective way to meet these standards is to incorporate principles in the design process that frustrate unwanted radiation and unwanted reception. *An understanding of antennas is essential to this process* so that costly post-design "fixes" are not required. Some examples of things to be avoided are shown in Fig. 4-9, which shows examples of unintentional antenna elements.

V_o

PCB With Cables

Analog Gnd.

Digital Gnd.

Power Supply

PCB With Separate
Analog & Digital Gnds.

(*a*) Unintentional dipoles.

PCB

Daughter Board

Ground Wire

Mother Board

Chassis

Mother/Daughter Board

PCB Separated From Chassis &
Connected With a Ground Wire

(*b*) Unintentional monopoles.

PCB

Signal

Ground

Clocks on PCBs

(*c*) Unintentional loop.
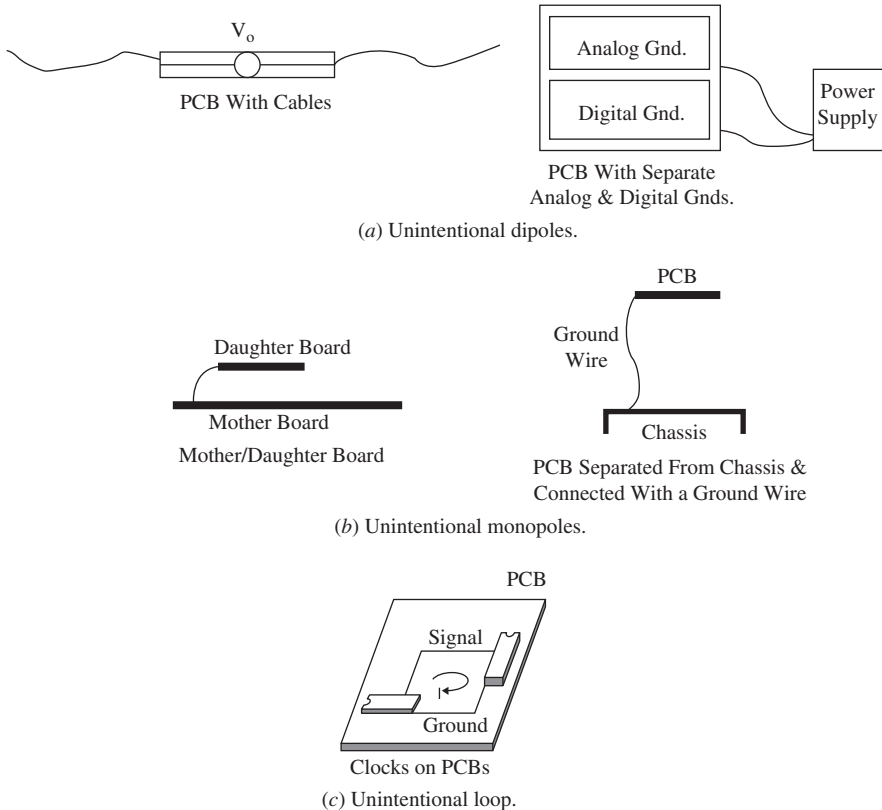
Figure 4-9 Examples of unintentional radiating elements.

An example from Paul [8] illustrates the difficulty caused by a small, unintentional current loop on a printed circuit board (PCB). Consider a 1 cm by 1 cm current loop on a PCB carrying a current of 100 mA at a frequency of 50 MHz. At the standard FCC Class B measurement distance of 3 m, the maximum value of the electric field from this current loop is 40.8 dBμV/m. The FCC Class B limit at this frequency is 40.8 dBμV/m. Thus, an electronic product containing a current loop such as this one would fail to comply with the FCC limit and could not be sold. The PCB designer with knowledge of antennas would know how to avoid having such a current loop in his design. (See Prob. 4.7-2.)

In the performance of emission measurements, a receiver such as a spectrum analyzer, a network analyzer, or an RF voltmeter is used. Its reading at a particular frequency will be in terms of microvolts usually expressed in units of dBμV (decibels above 1 microvolt). What is needed, however, is the value of the electric field, E^i, present at the measuring antenna because regulatory requirements are expressed in terms of E^i. The relationship between the antenna terminal voltage, V_L, and E^i, is expressed in (4-13) by the antenna factor, K. Antenna factor is used to quantify the emission and susceptibility of equipment to stray fields (i.e., interference). For emission testing, a probe antenna with known antenna factor receives the signal from a device under test (DUT) when transmitting at rated power output. The probe is oriented for maximum response. Calibrated antenna factor data is usually supplied with the measurement antenna because analytical determination of the antenna factor (or vector effective length) for most antennas is not easy. Typically, calibrated probe data is provided from 30 MHz to 1 GHz (see Sec. 7.9). In general, the spectrum analyzer voltage, V_r, from (4-13) expressed in dB form and including cable loss, L_c is (also see Sec. 13.5)

$$V_r(\text{dB}\mu\text{V}) = V_L(\text{dB}\mu\text{V}) - L_c(\text{dB}) = E^i(\text{dB}\mu\text{Vm}^{-1}) - K(\text{dBm}^{-1}) - L_c(\text{dB}) \quad (4\text{-}68)$$

EXAMPLE 4-10 *Antenna Factor Used in EMC Testing*

An electronic device emits an unwanted electric field at 100 MHz that produces a voltage at the terminals of an EMC measuring antenna. The strength of the electric field at the surface of the antenna is the FCC Class B limit of 43.5 dBμV/m, or $E^i = 149.62\,\mu\text{V/m}$, measured at the standard distance of 3 m from the device under test (DUT). The antenna factor of the antenna, as supplied by the antenna manufacturer, is 14 dBm^{-1}. The antenna is connected to a spectrum analyzer by 25 feet of RG-58U coaxial cable that has 4.5 dB of loss per 100 feet. What is the voltage reading of the spectrum analyzer?

The voltage at the antenna terminals is found using (4-68) as

$$\begin{aligned} V_r(\text{dB}\mu\text{V}) &= E^i(\text{dB}\mu\text{Vm}^{-1}) - K(\text{dBm}^{-1}) - L_c(\text{dB}) = 43.5 - 14 - 25 \times 4.5/100 \\ &= 28.375 \text{ dB}\mu V \end{aligned}$$

Thus, this electronic device needs to have its emission level reduced such that the voltage reading at the spectrum analyzer is reduced below 28.375 dBμV in order to satisfy the FCC requirement and have a margin of safety to allow for various uncertainties.

REFERENCES

1. J. H. Reed, Ed., *An Introduction to Ultra Wideband Communication Systems*, Prentice-Hall, Chap. 4, 2005.
2. W. Flock and E. K. Smith, "Natural Radio Noise—A Mini-Review," *IEEE Trans. on Ant. and Prop.,* Vol. AP-32, pp. 762–767, July 1984.
3. Y. Rahmat-Samii and W. L. Stutzman, Ed., "Special Issue on Wireless Communications," *IEEE Trans. on Ant. and Prop.*, Vol. 46, June 1998.
4. W. L. Stutzman and C. B. Dietrich Jr., "Moving beyond Wireless Voice Systems," *Scientific American*, Vol. 278, pp. 80–81, April 1998.
5. D. E. Kerr, *Propagation of Short Radio Waves*, Dover: New York, 1951.
6. W. L. Stutzman and J. W. LaPean, "Model Predicts Rain Attenuation in Satellite Links," *Microwaves & RF*, Vol. 31, pp. 108–115, April 1992.
7. W. L. Stutzman, et al., "Results from the Virginia Tech Propagation Experiment Using the Olympus Satellite 12, 20, and 30 GHz Beacons," *IEEE Trans. on Ant. and Prop.*,Vol. 43, pp. 54–62, Jan. 1995.
8. C. R. Paul, *Introduction to Electromagnetic Compatibility*, 2nd ed., Wiley, p. 428, 2006.

PROBLEMS

4.2-1 *Vector effective length of a small loop antenna* (a) Derive the following vector effective length expression starting with (3-49).

$$\mathbf{h} = -j\beta\mu_{\text{eff}}NS \sin\theta \,\hat{\boldsymbol{\phi}}$$

for a small loop antenna oriented in the *xy*-plane with *N* turns, effective relative permeability μ_{eff}, and area *S* of a single turn. (b) Find an expression for the ratio of the effective length of a small loop antenna to that for a short dipole antenna. (c) Evaluate the expression in (b) for the case of a single-turn loop of 0.3 m in diameter and a short dipole 0.3 m in length, both operating at 1.5 MHz. Comment on the results.

4.2-2 A ferrite rod antenna has 22 wire turns wound on a ferrite core with $\mu_{\text{eff}} = 38$ and diameter 1.17 cm. Evaluate the effective length (maximum) using the expression in Prob. 4.2-1a at 1 MHz.

4.2-3 (a) Use (4-8) and (4-10) to derive a general expression for A_{em}. (b) Show that the expression reduces to (4-12) for a conjugate match.

4.2-4 (a) Write an equation in dB form for antenna factor based on (4-13), and (b) Use the formula to calculate the antenna factor value in dB/m of a receiving antenna connected to a spectrum analyzer by a cable with 8 dB loss. The voltage measured at the spectrum analyzer (with a 50-Ohm input impedance) is 40 dBμV. The incident wave is matched to the antenna and has an intensity of 60 dBμV/m.

4.3-1 An Earth-based radiometer has an antenna beamwidth of HP = 0.079° that when centered on Mars measures an antenna temperature of 0.8 K at 9.5 GHz. Mars appears as a circular disk subtending an angle of 0.005°. Find the equivalent source temperature of Mars.

4.3-2 The main beam of an antenna with directivity of 40 dB is aimed and centered on the sun, which subtends an angle of 0.5°. Calculate the contribution to antenna temperature from the sun at 12 GHz, where the noise temperature of the sun is about 6,000 K.

4.3-3 Calculate the noise power in dBW out of an antenna at 5 GHz in a bandwidth of 1.5 MHz for the limiting conditions of elevation angles of 0° and 90°.

4.4-1 Derive an expression for the maximum effective aperture for an isotropic antenna. Compare to that for a short dipole.

4.4-2 (a) Derive an expression for antenna effective length based on (4-12) that uses directivity instead of maximum effective aperture. Note that antenna ohmic losses are included by using gain in place of directivity. (b) Evaluate the expression for an ideal dipole. (c) Evaluate the effective length of a half-wave dipole at 100 MHz. Compare to the actual physical length.

4.4-3 (a) Find the effective aperture expression in terms of λ^2 for a lossless isotropic antenna. (b) Repeat (a) for a short dipole and compare to the Example 4.2 result. (c) Repeat (a) for a half-wave dipole. (d) Evaluate the effective aperture for a half-wave dipole at 100 MHz and compare to its physical area assuming a 3-mm wire diameter.

4.4-4 Calculate the beam solid angle Ω_A for an ideal dipole in steradians (square radians) and in square degrees. Use the fact that $A_{em} = 0.119\ \lambda^2$ for an ideal dipole.

4.4-5 Suppose a transmitting antenna produces a maximum far-zone electric field in a certain direction given by $E = 90\ I\frac{e^{-j\beta r}}{r}$ where I is the peak value of the terminal current. The input resistance of the lossless antenna is $50\ \Omega$. Find the maximum effective aperture of the antenna, A_{em}. Your answer will be a number times wavelength squared.

4.4-6 A parabolic reflector antenna with a circular aperture of 3.66-m diameter has a 6.30 m² effective aperture area. Compute the gain in dB at 11.7 GHz.

4.4-7 The effective aperture of a 1.22-m diameter parabolic reflector antenna is 55% of the physical aperture area. Compute the gain in dB at 20 GHz.

4.4-8 Compute the gain in dB of 0.3-m circular diameter aperture antennas with 70% aperture efficiency at 5, 10, and 20 GHz. This problem approximates the performance of a small satellite earth terminal antenna over the range of commonly used frequencies and quantifies the frequency dependence of gain for a fixed aperture size.

4.4-9 Derive the dB form of the power transmission equation (4-34) from (4-33).

4.4-10 Write a power transfer equation similar to (4-34) but with distance R in units of miles.

4.4-11 An FM broadcast radio station has a 2-dB gain antenna system and 100 kW of transmit power. Calculate the effective isotropically radiated power in kW.

4.4-12 Derive a power transfer equation in a form involving the effective isotropically radiated power of the transmitter, the effective aperture of the receiving antenna, and free space spreading loss, $1/4\pi R^2$. Start with (4-33).

4.4-13 Calculate the received power in watts for the DBS system of Example 4-4 using (4-33).

4.4-14 A 150-MHz VHF transmitter delivers 20 W into an antenna with 10 dB gain. Compute the power in W available from a 3-dB gain receiving antenna 50 km away.

4.4-15 We wish to examine the received power on communication links with the same terminal separation distance but with different terminal configurations. Examine the ratio of the received to transmitted power, P_r/P_t, for a basic communication link to extract the frequency dependence for the following terminal antenna conditions: (a) Both have constant gain with frequency. (b) A constant-gain transmit antenna and a constant effective-aperture receive antenna. (c) Both have constant effective aperture.

4.4-16 Derive (4-39) using (4-41a).

4.4-17 A transmitting antenna is not matched to the impedance of a connecting transmission line. The radiation intensity, or equivalently the power density at a fixed distance, is reduced from the perfect impedance match case. Compute this reduction in dB for mismatch situations that produce VSWR values on the transmission line of 1.01, 1.2, 2, and 10. Also calculate the return losses.

4.4-18 Find the complex-valued unit vector for a right-hand elliptically polarized wave with an axial ratio of 2 dB and tilt angle $\tau = 45°$. Then compute the polarization efficiency for receiving antennas with the following polarizations: (a) Horizontal linear. (b) Vertical linear. (c) Right-hand circular. (d) Left-hand circular. (e) Right-hand elliptical with AR = 2 dB and $\tau = 45°$. (f) Left-hand elliptical with AR = 2 dB and $\tau = 135°$.

4.5-1 A low Earth orbit (LEO) satellite system transmits at 1.62 GHz using a 29-dB gain antenna with spot beams directed toward users on the Earth that are a maximum of 1500 km away. Find the required satellite transmit power in order for the power received by a user at the maximum distance to be at least -100 dBm if the user has a 1-dB gain antenna directed toward the satellite.

4.5-2 A cellular telephone base station transmitter at 850 MHz delivers 20 W into a 10-dB gain antenna. Compute the power in W available from a 3-dB gain mobile receiving antenna 20 km away assuming free space propagation conditions.

4.5-3 This problem compares the performance of the wireless system of Prob. 4.4-14 to wire line systems using coaxial and fiber optic cables.
  (a) Compute the coaxial cable loss in dB for the 50 km distance using an attenuation of 0.1 dB/m.
  (b) What is the net loss for the wireless system of Prob. 4.4-14; that is, find the net loss between the transmit and receive antenna input ports.
  (c) Would repeater amplifiers be necessary in the cable system?
  (d) Repeat (a) and (b) for the case of a 500-m path length.
  (e) Repeat (a) and (b) for a 500-m path at 300 MHz for a cable attenuation of 0.14 dB/m. The antenna gains are the same.
  (f) Compute the loss in dB for distances of 50 km and 500 m of a fiber optic cable with an attenuation of 1 dB/km.
  (g) Tabulate numerical results.

4.5-4 A pager receiver operating at 152.65 MHz uses a loop antenna constructed of a copper band with rectangular cross-section 3.65 mm $\times$ 0.70 mm in a single turn that is 41.6 mm long and 13 mm wide. (a) Compute the radiation efficiency. (b) Compute the effective aperture after first finding the gain. (c) Compute the power output from the antenna for an input electric field intensity of 13 μV/m.

4.5-5 Perform the calculations for the forward link of Example 4-8 also for operation at 1900 MHz. Assume the base station transmits 20 W and the handheld terminal has an antenna temperature of 290 K and a receiver input noise temperature of 1500 K. Assume the transmit antenna is impedance matched to its transmission line.

4.6-1 A monostatic radar detects a 10 m^2 target at a range of 266.4 km. It transmits 100 kW at a frequency of 3 GHz. The antenna gain is 40 dB and the received power is -100 dBm or 10^{-13}W. At what range would a stealthy target with an RCS of -30 dBsm$(10^{-3}$m$^2)$ be detected?

4.7-1 The voltage, V_L, induced at the terminals of an antenna is 0.2 V for every V/m of the incident electric field. What voltage level in dBμV at the base of the antenna would correspond to the FCC Class B limit of 43.5 dBμV/m for the incident electric field at 144 MHz?

4.7-2 Suggest two ways to reduce the radiated emission level of the 1 cm by 1 cm loop discussed in the Sec. 4.7.

Line Sources

The line source has been introduced in Sec. 2.4.4. It was called a line current. It can also be thought of as a current filament. It is not an actual antenna, but it can be used to model many antennas, and this is the motivation for studying it here. For example, consider the 4λ by 2λ aperture in Fig. 5-1, where an electric field is directed across the narrow dimension of the aperture, and there is a cosine variation across the wide dimension. What are the side lobe levels in the E-plane and in the H-plane, and what are the half-power beamwidths in both of these principal planes? The material in this chapter will enable us to easily answer these questions concerning the radiation pattern.

5.1 THE UNIFORM LINE SOURCE

We begin our discussion of line sources by considering an important special case, that of the uniform line source. A uniform line source has a current distribution with uniform amplitude and linear phase progression given by

$$
I(z') = \begin{cases} I_o e^{j\beta_o z'} & -\dfrac{L}{2} < z' < \dfrac{L}{2} \\[2mm] 0 & \text{elsewhere} \end{cases} \tag{5-1}
$$

where β_o is the phase shift per unit length along the line source. The free space wave number, β, and the guided wave number, β_o, are usually not the same, although they can be.
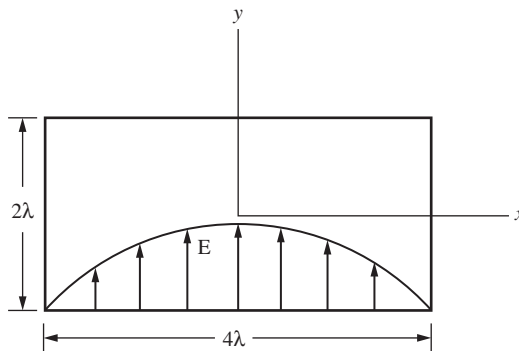
Figure 5-1 Rectangular aperture with a cosine electric field distribution.

The unnormalized pattern factor of the uniform line source is

$$f_{\mathrm{un}}(u) = \int_{-L/2}^{L/2} I(z')e^{j\beta z' \cos\theta}dz' = I_o L\frac{\sin u}{u} \tag{5-2}$$

where

$$u = (\beta \cos\theta + \beta_o)\frac{L}{2} \tag{5-3}$$

The evaluation of (5-2) is similar to that given in (2-110) for a broadside uniform line source.

It is convenient to introduce an angle θ_o (see Prob. 5.4-1) such that

$$\beta_o = -\beta \cos\theta_o \tag{5-4}$$

Then (5-3) becomes

$$u = \frac{\beta L}{2}(\cos\theta - \cos\theta_o) \tag{5-5}$$

The far-zone electric field from (2-106) and (5-2) is

$$E_\theta = \frac{j\omega\mu e^{-j\beta r}}{4\pi r}I_o L \ \sin\theta\frac{\sin u}{u} \tag{5-6}$$

The pattern factor of this uniform line source field expression is

$$f(u) = \frac{\sin u}{u} \tag{5-7}$$

The pattern factor is shown in Fig. 5-2 without using absolute values. The maximum occurs for $u = 0$ and is unity (0 dB) there. The nulls occur at multiples of π and are separated by π, except for the beamwidth between first nulls, which is 2π.
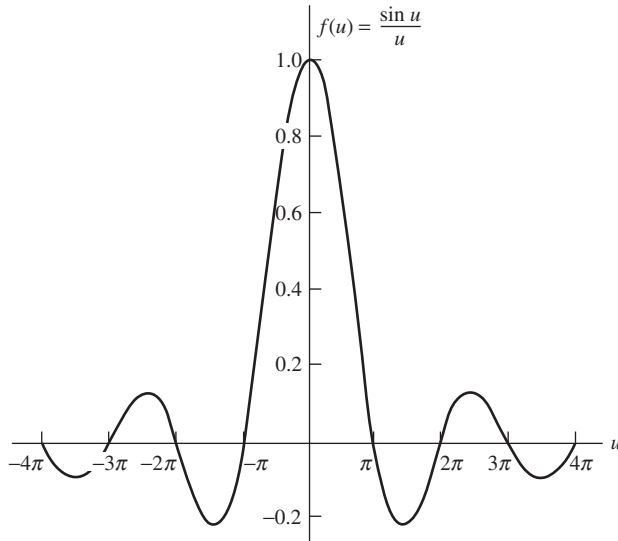
Figure 5-2 Pattern factor of the uniform line source of length L and $u = (\beta L/2) \cos\theta$.

The half-power beamwidth of the uniform line source pattern factor is found from solving

$$\frac{1}{\sqrt{2}} = \frac{\sin u_{\text{HP}}}{u_{\text{HP}}} \tag{5-8}$$

The solutions to this are $u_{\text{HP}} = \pm 1.39$. Then from (5-5)

$$\theta_{\text{HP}} = \cos^{-1}\left(\frac{2}{\beta L} u_{\text{HP}} + \cos \theta_o\right) = \cos^{-1}\left(\pm 0.443 \frac{\lambda}{L} + \cos \theta_o\right) \tag{5-9}$$

The plus sign corresponds to the half-power point on the right of the main beam maximum and the minus sign to the left half-power point. So from (2-126),

$$\begin{aligned}
\text{HP} &= \left|\theta_{\text{HP left}} - \theta_{\text{HP right}}\right| \\
&= \left|\cos^{-1}\left(-0.443\frac{\lambda}{L} + \cos \theta_o\right) - \cos^{-1}\left(0.443\frac{\lambda}{L} + \cos \theta_o\right)\right|
\end{aligned} \tag{5-10}$$

This formula is general but useful only when both half-power points appear in the visible region $(0 \le \theta \le 180°)$, which in turn requires that the arguments of the arccosines in (5-10) are between -1 and $+1$. For a broadside uniform line source, $\theta_o = 90°$ and (5-10) reduces to (see Prob. 5.1-1)

$$\text{HP} = 2 \sin^{-1}\left(0.443\frac{\lambda}{L}\right) \quad (\theta_o = 90°) \tag{5-11}$$

For long $(L \gg \lambda)$ line sources, this is approximately

$$\text{HP} \approx 0.886 \frac{\lambda}{L} \text{ rad} = 50.8 \frac{\lambda}{L} \text{ degrees} \quad (\theta_o = 90°) \tag{5-12}$$

since $\sin^{-1}(x) \approx x$ for $x \ll 1$. For an endfire uniform line source, only one half-power point appears in the visible region and then

$$\text{HP} = 2 \cos^{-1}\left(1 - 0.443\frac{\lambda}{L}\right) \quad (\theta_o = 0° \text{ or } 180°) \tag{5-13}$$

For long $(L \gg \lambda)$ line sources, this may be approximated (see Prob. 5.1-2) as

$$\text{HP} \approx 2\sqrt{0.886 \frac{\lambda}{L}} \text{ rad} \quad (\theta_o = 0° \text{ or } 180°) \tag{5-14}$$

Since (5-14) leads to wider beamwidths than does (5-12), we conclude that beamwidth increases as the pattern is scanned away from broadside (see Fig. 5-4).

The half-power beamwidth expression $\text{HP} = 0.886(\lambda/L)$ for the broadside uniform line source was developed using two approximations. The effect of the element factor $\sin \theta$ was neglected, and also it was assumed that the line source was long. With a few examples, we can see how these approximations affect the beamwidth. In Table 5-1, half-power beamwidth values for three uniform line sources are presented for various levels of approximation. The first column is the HP found from the complete pattern expression

$$F(\theta) = \sin \theta \frac{\sin[(\beta L/2) \cos \theta]}{(\beta L/2) \cos \theta} \quad (\theta_o = 90°) \tag{5-15}$$

Table 5-1 Half-Power Beamwidth Evaluation for Broadside Uniform
Line Sources

Length L	Exact Value from Complete Pattern F of (5-15)	Value from Pattern Factor $f = \dfrac{\sin u}{u}$	Value from $HP = 0.886\dfrac{\lambda}{L}$
2λ	$24.766°$	$25.591°$	$25.382°$
5λ	$10.112°$	$10.166°$	$10.153°$
10λ	$5.071°$	$5.080°$	$5.076°$

The third column is the HP obtained from only the pattern factor of (5-7). The last column is that of (5-12). Note that even for five wavelengths, all values are in very close agreement. We can also see that as the length increases, the approximations; improve.

The largest side lobe is the first one (i.e., the one closest to the main beam). The side lobe maxima locations are found by differentiating (5-7) and setting it equal to zero. This leads to

$$u_{\mathrm{SL}} = \tan u_{\mathrm{SL}} \qquad (5\text{-}16)$$

The intersections of the straight-line curve u_{SL} with the curve $\tan u_{\mathrm{SL}}$ give the side lobe maximum locations (the main beam maximum is at $u_{\mathrm{SL}_0} = 0$). The first side lobe maximum occurs for $u_{\mathrm{SL}_1} = \pm 1.43\pi$. This is not precisely midway between the pattern nulls at π and 2π. The side lobe maxima are slightly closer to the main beam than midway between their nulls. Evaluating (5-7) at the first side lobe maximum location gives 0.217 or -13.3 dB.
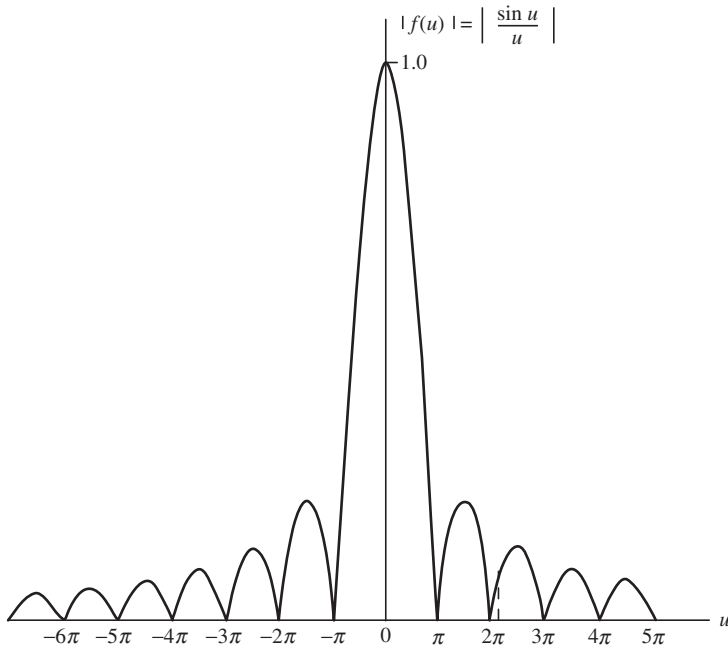
The polar plot of the pattern factor of a uniform line source can be obtained graphically from a universal pattern factor. The uniform line source universal pattern factor is shown in Fig. 5-3a. It is used for *all* source lengths L and scan angles θ_o. A typical case is shown in Fig. 5-3b. The transformation (5-5) between u and θ is illustrated graphically by the dashed lines. Pattern values for a given value of θ can be found from the universal pattern factor using this graphical transformation. The radius of the circle used in the transformation is $\beta L/2$ and its origin is at the value of u equal to $-(\beta L/2)\cos\theta_o$.

As an example, consider a three-wavelength uniform line source. The universal pattern factor is shown in Fig. 5-4a. The polar plot for the broadside case is illustrated in Fig. 5-4b. The pattern factor for a main beam maximum angle of $45°$ is polar-plotted in Fig. 5-4c. The endfire case is shown in Fig. 5-4d. Notice that the main beam (and also the side lobes) widen near endfire, as pointed out earlier. The current distributions required to produce those patterns are shown in Fig. 5-5. The amplitudes are constant in all cases, as illustrated in Fig. 5-5a. The required linear phase distributions for main beam scanning are depicted in Fig. 5-5b.

The effects of the element factor on the total pattern are shown in Fig. 5-6 for the three-wavelength uniform line source. In the broadside case of Fig. 5-6a, the element factor has a relatively minor effect. However, in the endfire case of Fig. 5-6b where the pattern factor alone produces a single endfire beam, the element factor effect on the total pattern produces a null in the endfire direction, thus bifurcating the main beam.

Next, we consider the directivity of the uniform line source. The directivity can be found easily if the element factor is assumed to have a negligible effect on the pattern. Then, we can work with the pattern factor f alone. First, the beam solid angle is from (2-142) and (5-7):

$$\Omega_A = \int_0^{2\pi}\int_0^{\pi}\left|\frac{\sin u}{u}\right|^2 \sin\theta\, d\theta\, d\phi \qquad (5\text{-}17)$$

(a) Universal pattern factor.
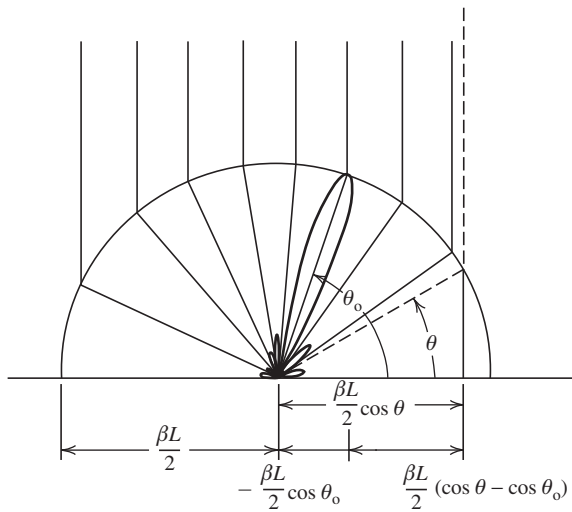
(b) Polar plot of pattern factor for $L = 4\lambda$.

Figure 5-3 Illustration of obtaining a polar plot from the universal factor of a uniform line source.

with the element factor g set to unity. If we change the θ integration variable to u as given by (5-3), then $du = -(\beta L/2) \sin\theta \, d\theta$ and (5-17) becomes

$$\Omega_A = \int_0^{2\pi} d\phi \int_{(\beta+\beta_o)L/2}^{(-\beta+\beta_o)L/2} \frac{\sin^2 u}{u^2} \frac{du}{-(\beta L/2)}$$

$$= 2\frac{\lambda}{L} \int_{(\beta_o-\beta)L/2}^{(\beta_o+\beta)L/2} \frac{\sin^2 u}{u^2} \, du \tag{5-18}$$
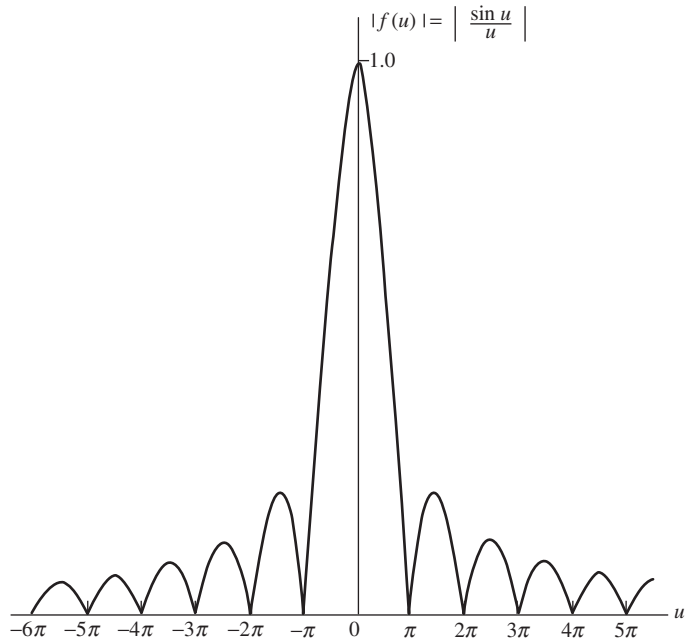
(a) Universal pattern factor.
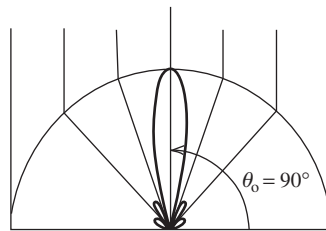
(b) Polar plot of pattern factor for $\beta_o L/2 = 0$, ($\theta_o = 90°$).
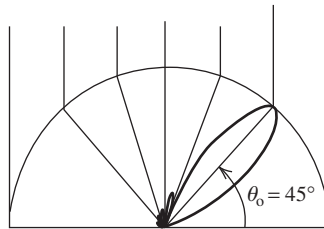
(c) Polar plot of pattern factor for $\beta_o L/2 = -2.12\pi$, ($\theta_o = 45°$).
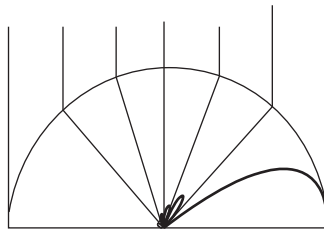
(d) Polar plot of pattern factor for $\beta_o L/2 = -\beta L/2 = -3\pi$, ($\theta_o = 0°$).

Figure 5-4 Pattern factors for a three-wavelength long ($L = 3\lambda$) uniform line source for various scan conditions.

$I(z')$

I_o

$-\dfrac{3\lambda}{2}$ $-\lambda$ $-\dfrac{\lambda}{2}$ 0 $\dfrac{\lambda}{2}$ λ $\dfrac{3\lambda}{2}$

z

(*a*) Current amplitude distribution.

Phase $[I(z)] = \beta_o z = -\beta z \cos\theta_o$.

3π

2π

π

$\beta_o = 0,\ \theta_o = 90°$

z

$-\dfrac{3\lambda}{2}$ $-\lambda$ $-\dfrac{\lambda}{2}$ $\lambda/2$ λ $\dfrac{3\lambda}{2}$

$-\pi$

-2π

$\beta_o = -\dfrac{\sqrt{2}\,\pi}{\lambda},\ \theta_o = 45°$

-3π

$\beta_o = -\dfrac{2\pi}{\lambda},\ \theta_o = 0°$

(*b*) Current phase distributions.
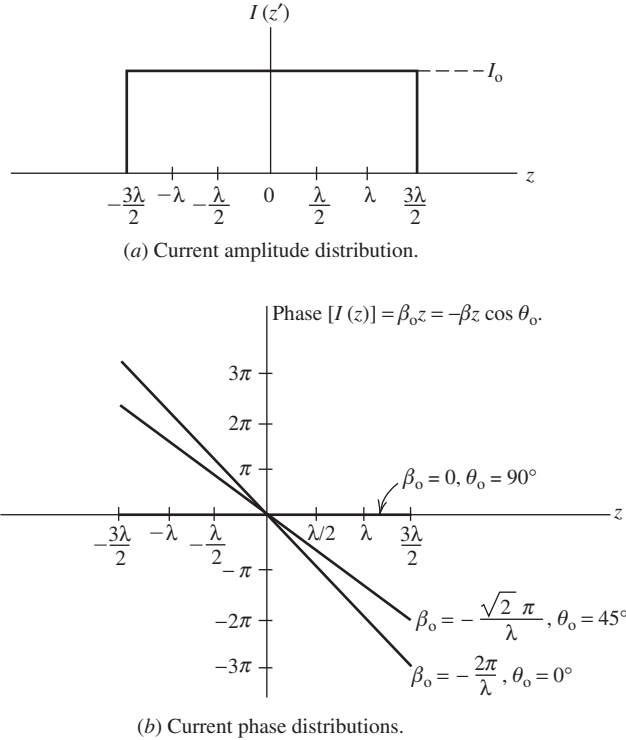
Figure 5-5 Current distributions for the three-wavelength uniform line source patterns of Fig. 5-4.

The evaluation of this expression for the general case is discussed in Prob. 5.1-7. For the broadside case ($\beta_o = 0$), the limits on the integral are $-\beta L/2$ to $\beta L/2$. If further $L \gg \lambda$, then $\beta L/2 \gg 1$ and we approximate the limits as $-\infty$ to $+\infty$, and if we use (F-12), the definite integral has a value of π. Thus, $\Omega_A \approx 2\lambda\pi/L$ and $D = 4\pi/\Omega_A$ yields

$$D_u = 2\frac{L}{\lambda} \qquad \text{(broadside, } L \gg \lambda) \qquad (5\text{-}19)$$

where the subscript u indicates a uniform line source. For the endfire case ($\beta_o = \pm\beta$), the integral limits are 0 and $\beta L/2$ that are approximated as 0 and ∞ when $L \gg \lambda$; this yields a value of $\pi/2$ for the integral. So, $\Omega_A \approx \lambda\pi/L$ and

$$D_u = 4\frac{L}{\lambda} \qquad \text{(endfire, } L \gg \lambda) \qquad (5\text{-}20)$$

The uniform line source exhibits the most directivity that can be obtained from a linear phase source of fixed length. Other current distributions will yield lower directivities as will be shown in the next section.

From the beamwidth and directivity relationships presented here for the uniform line source, we can begin to get a feel for the pattern changes as a function of source length and scan angle. First, consider the pattern factor alone. As the length increases, the beamwidth decreases and the directivity increases. The side lobe level (if the line source is long enough for the first side lobe maximum to be visible) remains constant with

Element factor
$g(\theta) = \sin\theta$

Pattern factor
$f(\theta)$
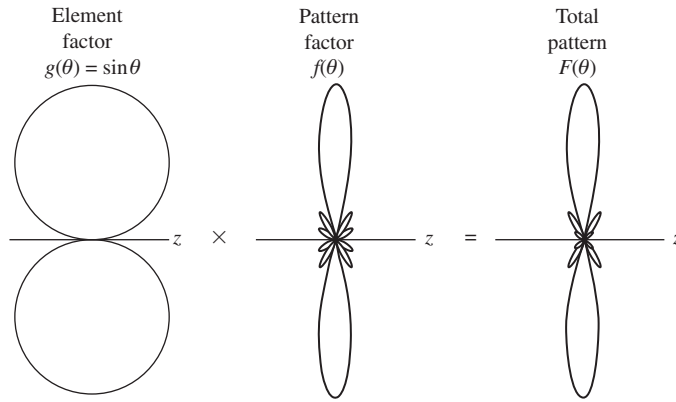
Total pattern
$F(\theta)$

(a) Broadside case ($\theta_{o} = 90°$, $\beta_{o} = 0$). Pattern factor is from Fig. 5-4b.
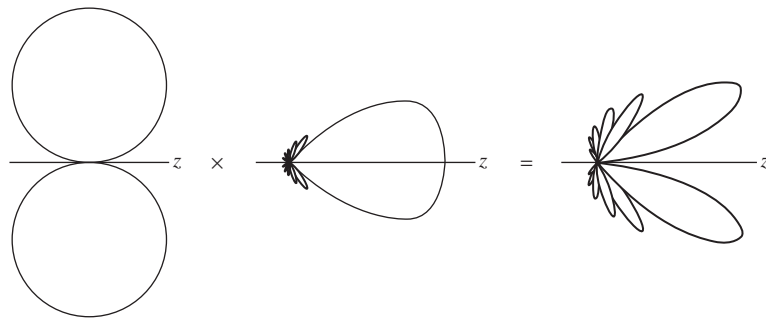
(b) Endfire case ($\theta_{o} = 0°$, $\beta_{o}L/2 = -3\pi$). Pattern factor is from Fig. 5-4d.

Figure 5-6 Total patterns for a three-wavelength uniform line source.

length variations; it is always $-13.3\,\mathrm{dB}$ for a uniform line source. For a scanned line source, the beamwidth increases as the main beam is scanned away from broadside. However, the total main beam volume (obtained by rotating the E-plane pattern about the z-axis) decreases and, consequently, Ω_A decreases, which in turn leads to an increase in directivity. The beamwidth and directivity change slowly for scan angles near broadside but change rapidly near endfire. The complete pattern must include the element factor effects. For long sources ($L \gg \lambda$), the pattern factor $f(\theta)$ has a much narrower pattern than the element factor $g(\theta) = \sin\theta$ and the total pattern obtained from $g(\theta)\,f(\theta)$ is closely approximated by $f(\theta)$. The side lobe level, beamwidth, and directivity values are then accurately determined from the pattern factor $f(\theta)$ alone, except near endfire where the element factor becomes significant since it forces the total pattern to zero in the $\theta = 0$ and $180°$ directions, as illustrated in Fig. 5-6b.

EXAMPLE 5-1   *Plane Wave Incident on a Slit*

A simple physical example of a uniform line source is a long narrow slit in a good conductor that has a uniform plane wave incident on it, as illustrated in Fig. 5-7. Phase fronts (planes of constant phase) are indicated by the parallel lines. In Fig. 5-7a, the wave is normally incident on the slit. Thus, the slit has a uniform amplitude excitation and also has uniform phase since
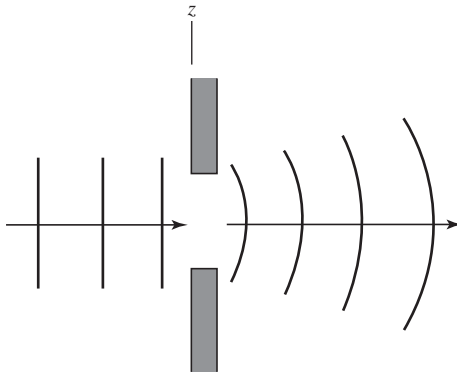
(*a*) Broadside case: $\theta_o = 90°$ and $\beta_o = -\beta \cos 90° = 0$.
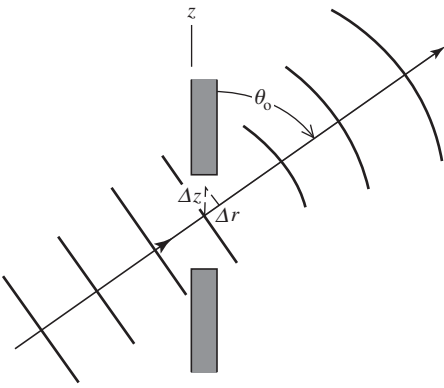
(*b*) Intermediate case: $\beta_o = -\beta \cos \theta_o$.
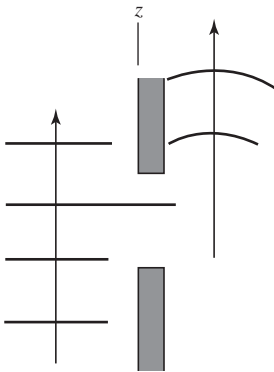
(*c*) Endfire case: $\theta_o = 0$ and $\beta_o = -\beta$.

Figure 5-7 Example of a uniform line source: an infinitely long slit of width L in a good conductor illuminated by a uniform plane wave from the left.

the phase fronts are parallel to the slit. The slit then behaves as a uniform line source with uniform phase across it. This equivalence of a field distribution to a current distribution will be discussed further in Sec. 9.1. It is obvious, however, that the maximum radiation on the right-hand side of the slit will be primarily in the direction of propagation of the incident wave coming from the left, that is, $\theta_o = 90°$. For a line source, the phase shift, along the source is related to the direction of maximum radiation θ_o by $\beta_o = -\beta \cos \theta_o$, where in this case β is the phase constant of the incident plane wave. Since $\theta_o = 90°$, $\beta_o = -\beta \cos 90° = 0$. This says that there is no phase shift along the slit. We already observed that this must be true for a plane wave normally incident on the slit.

If the direction of propagation of the plane wave makes an angle θ_o with the slit plane, there will be a phase shift along the slit due to different arrival times of the wave. In fact; this phase shift is given by $\beta_o z'$, where β_o is the phase shift per meter along the slit and we have assumed zero phase at $z' = 0$. But the phase shifts β rad/m in the direction of propagation, so the phase shift for distance Δr along the direction of propagation is $\Delta\phi = -\beta \ \Delta r$ (since the wave propagates as $e^{-j\beta r}$). The same phase shift is encountered in the distance Δz along the slit, or $\Delta\phi = \beta_o \ \Delta z$ (see Fig. 5-7b). But $\Delta r = \Delta z \ \cos \theta_o$, and since the phase shifts are equal, we have $\Delta\phi = -\beta \ \Delta z \ \cos \theta_o = \beta_o \ \Delta z$. Thus, $\beta_o = -\beta \cos \theta_o$ as given by (5-4), which was then a convenient definition. It is obvious from Fig. 5-7b that the maximum radiation from the slit or its equivalent line source will occur in the direction of propagation of the wave $\theta = \theta_o$.

In Fig. 5-7c the incident wave is traveling parallel to the slit. The phase shift per meter along the slit is obviously equal to the negative of the wave phase constant. This also follows from $\beta_o = -\beta \cos \theta_o = -\beta$ for $\theta_o = 0°$. The radiated wave on the right side is endfire in this case.

5.2   TAPERED LINE SOURCES

Many antennas that can be modeled by line sources are designed to have tapered distributions. This is because if the current amplitude decreases toward the ends of a line source, the pattern side lobes are lowered and the main beam widens. In many applications, low side lobes are necessary and a wider main beam is accepted as a consequence. This tradeoff between side lobe level and half-power beamwidth is a major consideration to the antenna engineer.

As an example, consider a current distribution with the so-called cosine taper, where

$$I(z') = \begin{cases} I_o \ \cos\left(\dfrac{\pi}{L}z'\right) e^{j\beta_o z'} & -\dfrac{L}{2} < z' < \dfrac{L}{2} \\[2ex] 0 & \text{elsewhere} \end{cases} \tag{5-21}$$

The shape of this current distribution is plotted in Fig. 5-8a. The unnormalized pattern factor is then found as follows:

$$\begin{aligned} f_{\text{un}}(\theta) &= I_o \int_{-L/2}^{L/2} \cos\left(\frac{\pi}{L}z'\right) e^{j(\beta \cos\theta + \beta_o)z'} \, dz' \\[2ex] &= \frac{I_o}{2} \int_{-L/2}^{L/2} [e^{j(\pi/L + \beta \cos\theta + \beta_o)z'} + e^{-j(\pi/L - \beta \cos\theta - \beta_o)z'}] \, dz' \\[2ex] &= \frac{I_o}{2} \left[\frac{e^{j(\pi/L + \beta \cos\theta + \beta_o)z'}}{j(\pi/L + \beta \cos\theta + \beta_o)} + \frac{e^{-j(\pi/L - \beta \cos\theta - \beta_o)z'}}{-j(\pi/L - \beta \cos\theta - \beta_o)} \right]_{-L/2}^{L/2} \end{aligned} \tag{5-22}$$

Evaluating the above expression leads to

$$f_{\text{un}}(\theta) = I_o \frac{2L}{\pi} \frac{\cos\left[(\beta \cos\theta + \beta_o)L/2\right]}{1 - [(\beta \cos\theta + \beta_o)L/\pi]^2} \tag{5-23}$$

Using $\beta_o = -\beta \ \cos \theta_o$ as in (5-4) and normalizing such that the pattern factor is unity for $\theta = \theta_o$ gives

$$f(\theta) = \frac{\cos\left[(\beta L/2(\cos\theta - \cos\theta_o)\right]}{1 - [(\beta L/\pi)(\cos\theta - \cos\theta_o)]^2} \tag{5-24}$$
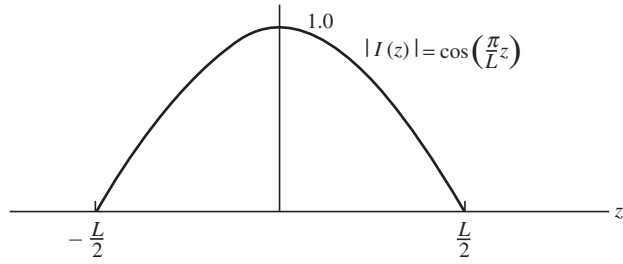
(a) Current distribution amplitude.
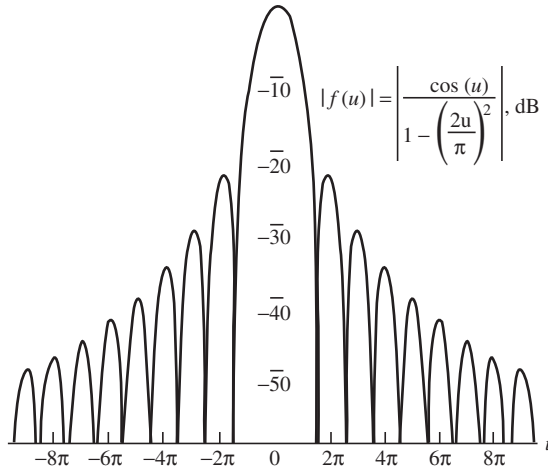
(b) Pattern Factor.

Figure 5-8 Current distribution and pattern factor for a cosine-tapered line source.

This pattern can be written in terms of u using (5-5) as

$$f(u) = \frac{\cos u}{1 - (2u/\pi)^2} \qquad (5\text{-}25)$$

This pattern is plotted in Fig. 5-8b. Compare the side lobe level to that of Fig. 5-12a for the uniform line source.

The side lobe level for the cosine-tapered line source is -23.0 dB and the beamwidth is given by

$$\text{HP} \approx 1.19 \frac{\lambda}{L} \text{ rad} = 68.2 \frac{\lambda}{L} \text{ degrees} \qquad (5\text{-}26)$$

for the broadside case. The side lobe level is 10 dB lower and the beamwidth is 38% greater than a uniform line source of the same length. Although the side lobes are reduced from those of the uniform line source, the main beam widening leads to smaller directivity than obtained from a uniform line source. The ratio D/D_u is used to compare the directivity of a tapered line source to that of a uniform line source of the same length. For the cosine taper, $D/D_u = 0.810$. The actual directivity D from (5-19) is then

$$D = 0.810 D_u = 1.620 \frac{L}{\lambda} \quad (\text{broadside, } L \gg \lambda) \qquad (5\text{-}27)$$
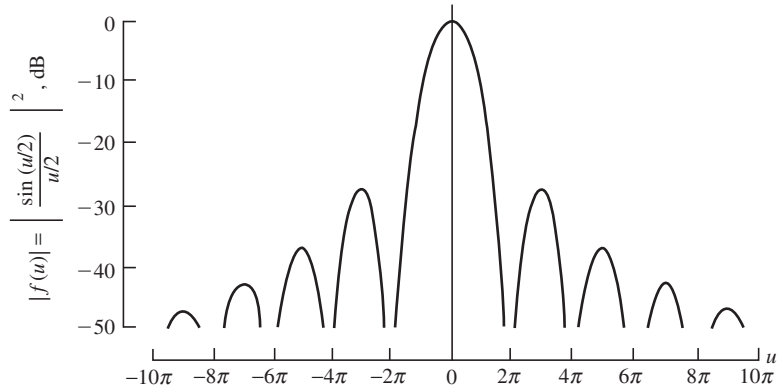
Figure 5-9 Pattern factor of a triangular tapered line source.

If the current amplitude taper is increased as in the case of a cosine-squared taper, the side lobes are reduced even more and the beamwidth is further widened. The pattern parameters of the cosine-squared case, as well as many other important cases, are summarized in Table 5-2 [1; H.6: Hansen, Vol. I, Chap. 1; H.3: Silver, p. 187].

As a further example, consider the triangular current taper given in Table 5-2a. The pattern $(\sin u/u)^2$ is the square of the uniform line source pattern. This property is apparent when the pattern of Fig. 5-9 for the triangular line source is compared to that of the uniform line source in Fig. 5-2. The first nulls of the triangular line source are twice as far out as for the uniform line source pattern. Thus, the beamwidth between first nulls is twice as large. The half-power beamwidth is 44% larger (from $0.886\lambda/L$ to $1.28\lambda/L$). Also, the side lobes of the triangular line source are twice as wide in the variable u and the side lobe level in decibels is twice as small, $-13.3\,$dB for the uniform line source and $-26.6\,$dB for the triangular line source. The directivity (from Table 5-2a) is 75% of the uniform line source value.

From Table 5-2, we can generalize and make some statements about current amplitude distributions and their influence on the far-field pattern. We assume that the current is of constant phase. *As the taper of the current amplitude from the center to the edges of a line source becomes more severe, the side lobes decrease and the beamwidth increases.* Consequently, the directivity decreases. There is then a tradeoff between the side lobe level and the beamwidth. The antenna engineer must decide on a compromise between beamwidth and side lobe level for each specific design problem.

EXAMPLE 5-2   *A Cosine Tapered Rectangular Aperture*

Consider the aperture in Fig. 5-1. Determine the side lobe level in the *E*- and *H*-planes. Also determine the half-power beamwidths (HP).

First we note that the electric field distribution is that of a cosine in the *x*-direction. In the *y*-direction, the distribution is uniform because the electric field lines are lines of uniform intensity. Consulting Table 5-2, we see that the cosine distribution in the *H*-plane gives a $-23\,$dB side lobe level and the uniform distribution in the *E*-plane gives a $-13.3\,$dB side lobe level. *The side lobe levels are independent of aperture size and depend only on the distribution of the electric field in the aperture.*

Again from Table 5-2, the HP in the *H*-plane is $1.19/4 = 0.3$ rad or $17°$. In the *E*-plane, the HP is $0.886/2 = 0.443$ rad or $25.4°$. *The half-power beamwidths depend on both the aperture size and the aperture distribution.*

Table 5-2 Characteristics of Tapered Line Source Distributions

a. Triangular taper

$$I(z) = 1 - \frac{2}{L}|z| \quad |z| \le \frac{L}{2}$$

$$f(u) = \left[\frac{\sin(u/2)}{u/2}\right]^2$$
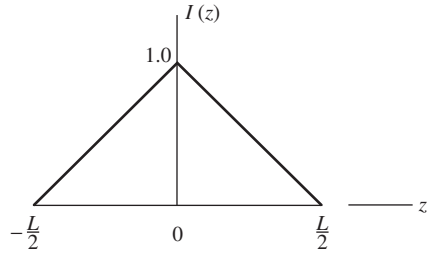
	Side Lobe Level	
HP (rad)	(dB)	D/D_u
$1.28(\lambda/L)$	-26.6	0.75

b. Cosine tapers

$$I(z) = \cos^n\left(\frac{\pi z}{L}\right) \quad |z| \le \frac{L}{2}$$

$$f(u) = \frac{\sin u}{u} \qquad n = 0$$

$$f(u) = \frac{\cos u}{1 - (2u/\pi)^2} \qquad n = 1$$

$$f(u) = \frac{1}{1 - (u/\pi)^2}\frac{\sin u}{u} \qquad n = 2$$
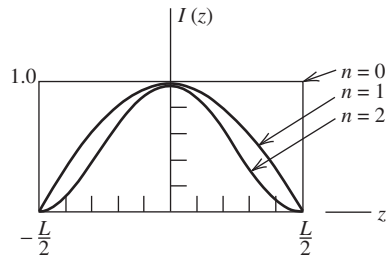
		Side Lobe Level		
n	HP (rad)	(dB)	D/D_u	Type
0	$0.886\,\dfrac{\lambda}{L}$	-13.3	1.00	Uniform line source
1	$1.19\,\dfrac{\lambda}{L}$	-23.0	0.810	Cosine taper
2	$1.44\,\dfrac{\lambda}{L}$	-31.7	0.667	Cosine-squared taper

c. Cosine on a pedestal

$$I(z) = C + (1 - C)\,\cos\frac{\pi z}{L}$$

$$f(u) = \frac{C\dfrac{\sin u}{u} + (1 - C)\dfrac{2}{\pi}\dfrac{\cos u}{1 - (2u/\pi)^2}}{C + (1 - C)\dfrac{2}{\pi}}$$
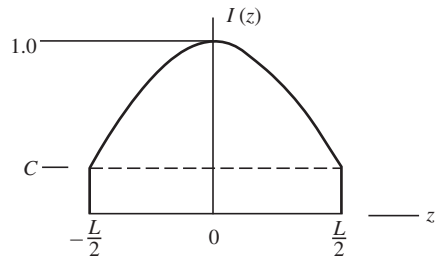
Table 5-2 Continued

		Edge Illumination		
C	$-20 \log C$ (dB)	HP (rad)	Side Lobe Level (dB)	D/D_u
0.3162	-10	$1.03 \dfrac{\lambda}{L}$	-20	0.92
0.1778	-15	$1.08 \dfrac{\lambda}{L}$	-22	0.88
0	$-\infty$	$1.19 \dfrac{\lambda}{L}$	-23	0.81

Note 1: The pattern expressions are valid for any value of $u = (\beta L/2)(\cos\theta - \cos\theta_o)$. However, the half-power beamwidth values and directivities are approximations for broadside line sources $\theta_o = 90°$ and $u = (\beta L/2)\cos\theta)$ and $L \gg \lambda$.

Note 2: The directivity for each line source is found from the ratio D/D_u as $D = \dfrac{D}{D_u} D_u = \dfrac{D}{D_u} 2\dfrac{L}{\lambda}$ for broadside line sources with $L \gg \lambda$.

Note 3: The element factor $\sin\theta$ has been neglected in the calculations leading to the values in this table. For long, broadside line sources, its effect is minimal at or near broadside.
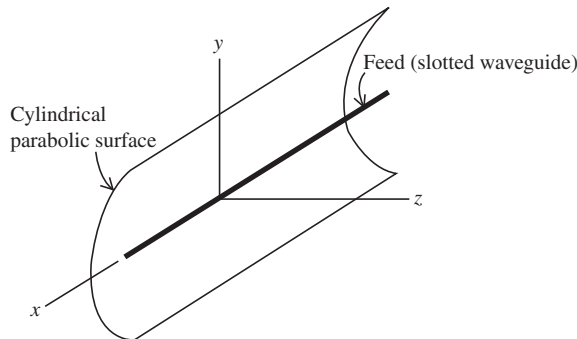
Figure 5-10 Parabolic cylindrical reflector with *x*-polarized feed along the axis of the cylinder.

EXAMPLE 5-3   *A Cylindrical Parabolic Reflector Antenna*

A cylindrical parabolic reflector antenna (see Fig. 5-10) can be modeled by line sources. Suppose the parabolic surface is 10λ across at the edges of the reflector (i.e., the aperture) in Fig. 5-10 and that the field distribution in the aperture in the *y*-direction is that of a cosine on a pedestal with -15 dB edge illumination. Then from Table 5-2c, the half-power beamwidth is

$$HP = 1.08\lambda/L = 0.108 \ \text{rad} = 6.2°$$

and the side lobe level is -22 dB. Fig. 16-21 shows a pattern calculated via a one-dimensional aperture integration for an aperture distribution that is nearly a cosine on

a pedestal (Fig. 16-22). It has $-22\,$dB side lobes and a half-power beamwidth of $6.2°$ in the H-plane. There is no E-plane data because the antenna in Sec. 16.6 is infinite in the x-direction (i.e., a two-dimensional rather than a three-dimensional problem). Suppose, however, that the antenna is also 10λ in the x-direction (E-plane), as implied by Fig. 5-10, and the aperture distribution is uniform in the E-plane. The E-plane pattern would then be modeled by a uniform line source. Aperture directivity can be calculated from line source formulas if the principal plane distributions are separable; see (9-86). Much more will be said about aperture antennas in Chap. 9, where, as in this example, linesource results will play an important role.

5.3   FOURIER TRANSFORM RELATIONS BETWEEN THE FAR-FIELD PATTERN AND THE SOURCE DISTRIBUTION

The far-field pattern and its (nonperiodic) source distribution form a Fourier transform pair. To see this, consider (2-103) and (2-106) where $I(z') = |I(z')|e^{j\beta_o z'}$ and write

$$F_{un}(\theta) = \sin\theta \int_{-L/2}^{L/2} |I(z')|e^{j(\beta\cos\theta+\beta_o)z'}\,dz' \tag{5-28}$$

or

$$f_{un}(\theta) = \frac{F_{un}(\theta)}{\sin\theta} = \int_{-L/2}^{L/2} |I(z')|e^{j(\beta\cos\theta+\beta_o)z'}\,dz' \tag{5-29}$$

where once again, the element pattern $\sin\theta$ has been absorbed into the far field of the line source. Thus, $f_{un}(\theta)$ can be viewed as the far field of a line source in which the element pattern is isotropic. Since $I(z')$ is zero for $z' > L/2$ and $z' < -L/2$, the limits on (5-29) may be extended to infinity. Thus,

$$f_{un}(\theta) = \int_{-\infty}^{\infty} |I(z')|e^{j(\beta\cos\theta+\beta_o)z'}\,dz' \tag{5-30}$$

which is recognized as one-half of a Fourier transform pair. The other half of the (antenna) pair is

$$I(z') = \frac{1}{2\pi} \int_{-\infty}^{\infty} f_{un}(\theta)e^{-jz'\beta\cos\theta}\,d(\beta\cos\theta) \tag{5-31}$$

From circuit theory, the Fourier transform (circuit) pair can be written as

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(\omega)e^{j\omega t}\,d\omega \tag{5-32}$$

and

$$g(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}\,dt \tag{5-33}$$

If we let $\cos\theta$ and $\beta z'$ correspond to t and ω, respectively, then z'/λ corresponds to frequency f. The quantity z'/λ is called spatial frequency with units of hertz per radian. For real values of θ and $|\cos\theta| \leq 1$, the field distribution associated with $f_{un}(\theta)$ represents radiated power, whereas for $|\cos\theta| > 1$, it represents reactive or stored power (e.g., see Sec. 2.4.3). The pattern $f_{un}(\theta)$, or angular spectrum, represents an angular distribution of Plane waves. For $|\cos\theta| \leq 1$, the angular spectrum is the same as the far-field pattern $f_{un}(\theta)$.

In circuit theory, a very narrow pulse (in time) has a large or wide-frequency spectrum. To pass such a pulse through a filter requires that the filter have a wide passband. Similarly, an antenna with a very narrow far-field pattern must pass a wide band of spatial frequencies. That is, the antenna must be electrically large. Thus, the antenna may be viewed as a spatial filter, a concept widely used in radiometry and radio astronomy.

Probably the greatest value in recognizing that the source and far-field function form a Fourier transform pair is that one can utilize the vast amount of information available on Fourier transform theory, particularly in circuit theory. For example, Table 5-3 shows some, common Fourier transform pairs found in antenna theory. Some of these also appear in Table 5-2 that was constructed without reliance on a knowledge of Fourier transforms. The pairs in Table 5-3 work in either direction. That is, the source distributions and far-field patterns in Table 5-3 may be interchanged, at least theoretically. In some cases, however, the resulting distributions are not practical as far-field patterns or as source distributions. And finally, although our familiarity with (5-30) may imply that the application of (5-31) is equally straightforward, it is not. Application (5-31) leads to techniques in antenna synthesis. Antenna synthesis is discussed in Chap. 10.

5.4   FAST WAVE AND SLOW WAVE DISTRIBUTIONS

Previously in this chapter we considered the phase progression on a line source and described it by $e^{j\beta_o z'}$. This exponential form satisfies the frequency domain wave equation and is therefore a mathematical description of a wave traveling along the z-axis (see Prob. 2.2-2). The wave number, β_o, is either more or less than the free space wave number, $\beta = 2\pi/\lambda$, depending on the properties of the actual antenna structure supporting the wave that the line source represents. Thus, we write

$$\beta_o = 2\pi/\lambda_g \tag{5-34}$$

where λ_g is the guided wavelength associated with a phase velocity, v, on the actual antenna. Since $v = f\lambda_g$ it follows that

$$\cos\theta_o = -\frac{\beta_o}{\beta} = -\frac{\lambda}{\lambda_g} = -\frac{c}{v} \tag{5-35}$$

Previously in this chapter we saw that the phase progression determined the angle of maximum radiation. Alternatively from (5-35) we see that the angle of maximum radiation depends upon the ratio of the velocity of light in free space to the phase velocity on the antenna being modeled. Thus, the current of a uniform line source may also be expressed as
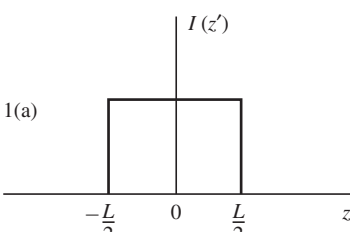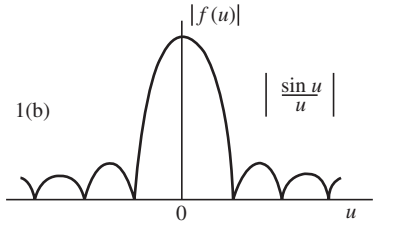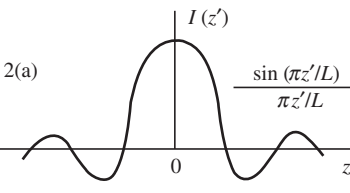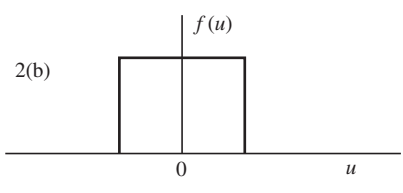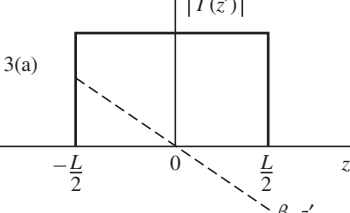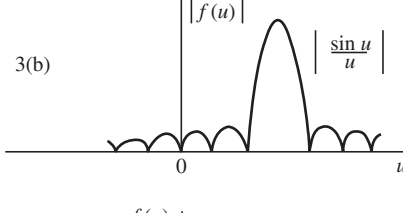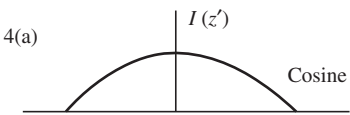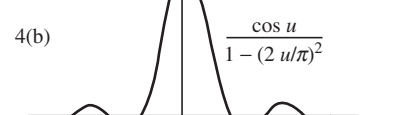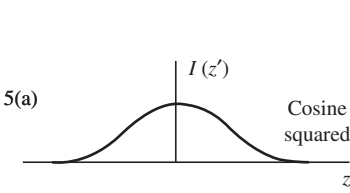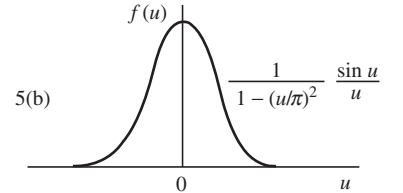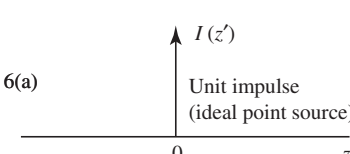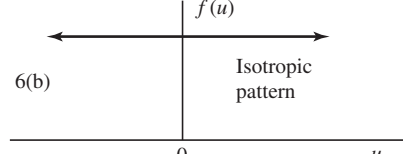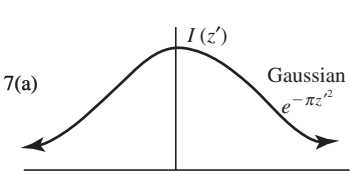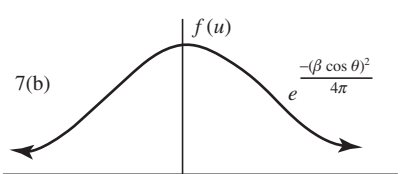
$$I(z') = I_o e^{j\beta(c/v)z'} \quad -\frac{L}{2} < z' < \frac{L}{2} \tag{5-36}$$

and we may classify the wave as either fast or slow according to

$$\text{Fast wave,} \quad c/v < 1$$
$$\text{Slow wave,} \quad c/v > 1$$

For a wave to be a fast wave, v must be greater than c. Consider a sinusoidal plane wave incident upon a large perfectly conducting sheet. At normal incidence, the crest of a wave strikes the sheet everywhere at the same instant, and c/v is zero because v is infinite. If the plane wave is traveling parallel to the sheet, $c = v$ and the wave is neither fast nor slow. To have a slow wave, v must be less than c, which is not possible in this example situation but is possible in some guided wave antennas such as the Yagi-Uda and axial mode helix.

Table 5-3 Some Common Fourier Transform Pairs

1(a) $I(z')$ [rectangular function from $-\frac{L}{2}$ to $\frac{L}{2}$]

1(b) $|f(u)|$ $\left|\dfrac{\sin u}{u}\right|$

2(a) $I(z')$ $\dfrac{\sin(\pi z'/L)}{\pi z'/L}$

2(b) $f(u)$

3(a) $|I(z')|$ $\beta_{\mathrm{o}} z'$

3(b) $|f(u)|$ $\left|\dfrac{\sin u}{u}\right|$

4(a) $I(z')$ Cosine

4(b) $f(u)$ $\dfrac{\cos u}{1 - (2\,u/\pi)^2}$

5(a) $I(z')$ Cosine squared

5(b) $f(u)$ $\dfrac{1}{1 - (u/\pi)^2}\dfrac{\sin u}{u}$

6(a) $I(z')$ Unit impulse (ideal point source)

6(b) $f(u)$ Isotropic pattern

7(a) $I(z')$ Gaussian $e^{-\pi z'^2}$

7(b) $f(u)$ $e^{\frac{-(\beta\cos\theta)^2}{4\pi}}$
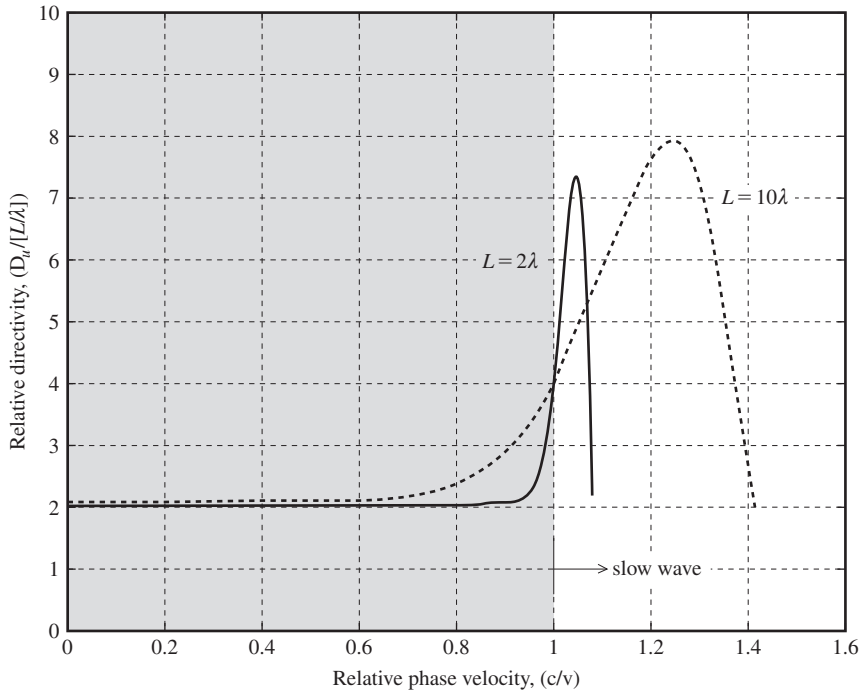
Figure 5-11 Relative directivity as a function of length and phase velocity.

Fig. 5-11 shows the relative directivity, $D_u/(L/\lambda)$, of the pattern factor for 2λ and 10λ line sources versus c/v. Note how well both curves agree with (5-19) and (5-20) even though the 2λ line source is not $\gg \lambda$. Also note that the directivities in both cases exceed the endfire $(c/v = 1)$ directivity when the phase velocity is slowed to less than c. The peak value in the curves occurs when the phase velocity is slowed such that the phase delay along the line source exceeds that of ordinary endfire by 2.94 radians (sometimes approximated by π radians for convenience). This is called the Hansen-Woodyard increased directivity condition. We will study two antennas that adopt this condition naturally, the Yagi-Uda in Chap. 6 and the helix in Chap. 7.

Note that the directivity in Fig. 5-11 does not include the effect of the $\sin\theta$ element factor in (5-6). This is of little consequence in the broadside case as Fig. 5-6a indicates. In the endfire case, it is of consequence if the antenna element factor is similar to $\sin\theta$. However, in some cases the element factor may be quite different for endfire antennas such as the Yagi-Uda and axial mode helix.

5.5   SUPERDIRECTIVE LINE SOURCES

We have seen in Table 5-2 that the maximum (ordinary) directivity from a line source occurs when the amplitude distribution is uniform. It is, however, possible to obtain greater directivity under certain conditions that result in superdirectivity. In general, linear sources with $L > \lambda$ are superdirective if the directivity is higher than that obtained using a phase distribution $e^{\pm j\beta_o z'}$ with $|\cos\theta_o| \le 1$ or $c/v \le 1$. Thus, the Hansen-Woodyard condition of the previous section is a superdirective condition because $c/v > 1$ implying that $|\cos\theta_o| > 1$.

Superdirectivity is produced by an interference process whereby the main beam is scanned into the invisible region (see Fig. 5-12), where $|u| > \pi L/\lambda$ or $|\cos\theta_o| > 1$. This causes energy to be stored in the near field, resulting in a large antenna Q. The reactive

(a) Broadside.

(b) Ordinary endfire.

(c) Hansen–Woodyard.
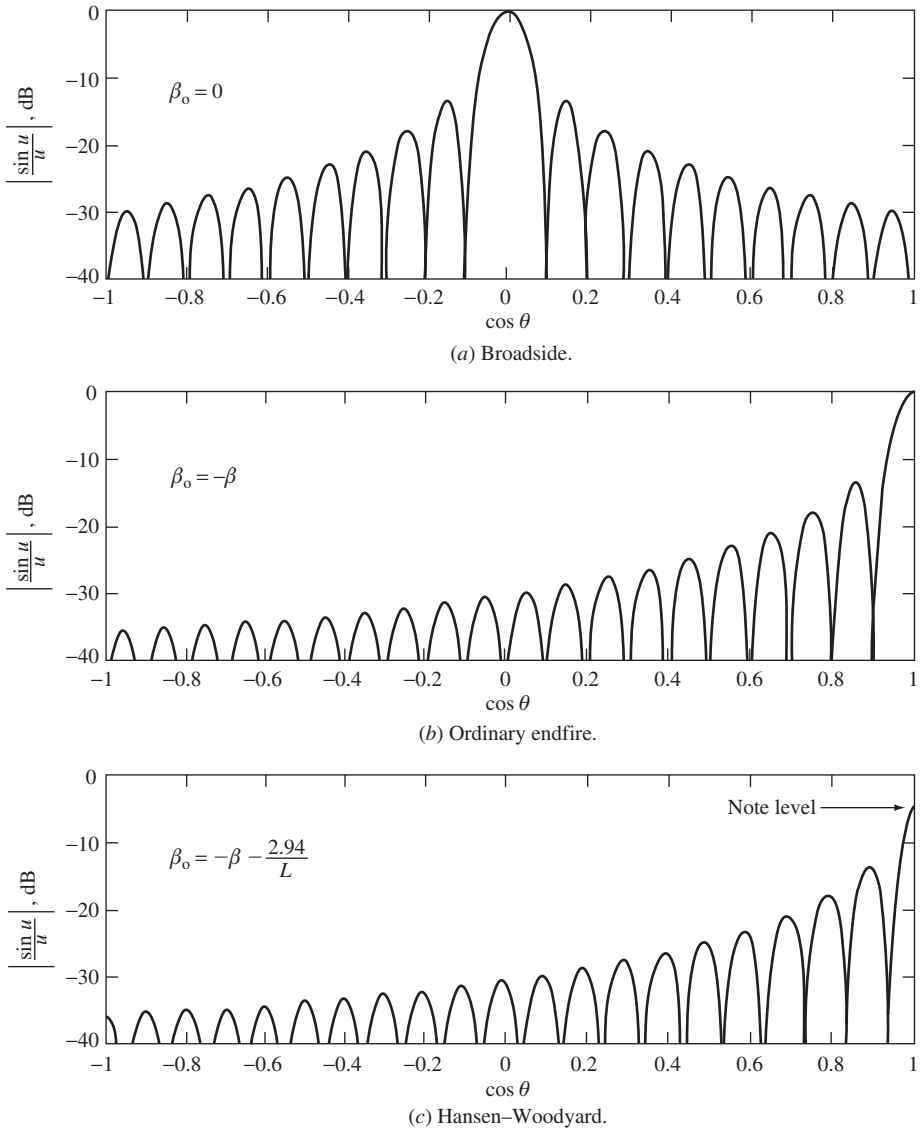
Figure 5-12 Far-field patterns for a 10λ line source: The corresponding superdirective ratios are found in Table 5-4.

power is found approximately by integrating over the invisible region and the radiated power is found, of course, by integrating the pattern factor over the visible region (where $|u| \leq \pi L/\lambda$). To quantify superdirectivity, a superdirective ratio R_{SD} may be defined as the ratio of radiated power plus reactive power to radiated power, which for a broadside line source is [2]

$$R_{SD} = \frac{\displaystyle\int_{-\infty}^{\infty} |f(u)|^2 \, du}{\displaystyle\int_{-\pi L/\lambda}^{\pi L/\lambda} |f(u)|^2 \, du} \tag{5-37}$$

For other than broadside, the limits in the denominator change (see Probs. 5.5-1 and 5.5-2).

Table 5-4 Superdirectivity for a 10λ
Line Source

Case	R_{SD}
Broadside	1.01
Ordinary endfire	2.01
Hansen-Woodyard	8.03

Since the Q may be expressed as the ratio of reactive power to radiated power,

$$Q \approx \frac{\displaystyle\int_{-\infty}^{-\pi L/\lambda} |f(u)|^2 \, du + \int_{\pi L/\lambda}^{\infty} |f(u)|^2 \, du}{\displaystyle\int_{-\pi L/A}^{\pi L/\lambda} |f(u)|^2 \, du} \tag{5-38}$$

Comparing the previous two equations gives $R_{SD} \approx 1 + Q$. If the line source is electrically small, $L \ll \lambda$, the element factor $\sin\theta$ must be included in (5-38) to determine the radiation Q [2].

To investigate superdirectivity for a uniform line source, the directivity can be written as

$$D_u \simeq R_{SD} \frac{2L}{\lambda} \tag{5-39}$$

Comparing (5-19) and (5-20) for long line sources, we conclude that $R_{SD} = 1$ for the broadside line source and $R_{SD} = 2$ for the ordinary endfire line source. For a 10λ line source, this is approximately true as Table 5-4 shows. The exact values for R_{SD} in the table were obtained using (5-37). Table 5-4 indicates that for the Hansen-Woodyard case, moderate levels of superdirectivity are achievable. The value of $R_{SD} = 2.01$ for the ordinary endfire case does not indicate a superdirective condition since R_{SD} was achieved with the linear phase distribution $e^{\pm j\beta_o z'}$ and $|\cos\theta_o| \leq 1$. In the broadside case, superdirectivity ratios even modestly greater than unity are not practical since rapid precisely controlled variations of phase are required. Even if such rapid phase variations could be achieved in practice, the resulting superdirectivity would not result in supergain because of a decrease in e_r, the radiation efficiency, due to the ohmic losses that would inevitably occur.

EXAMPLE 5-4   **_Superdirectivity of a_ 10λ _Broadside Line Source_**

It is desired to calculate R_{SD} for a broadside line source when $L = 10\lambda$:

$$f(u) = \frac{\sin u}{u} \quad \text{where} \quad u = \frac{\beta L}{2} \cos\theta \quad \text{since } \theta_o = 90°$$

The superdirective ratio is found using (5-37). The numerator of (5-37) has a value of π from (F-12). The denominator of (5-37) is evaluated using integration by parts. Let

$$dy = \frac{1}{u^2} \, du \qquad \text{and} \qquad x = \sin^2 u$$

so that

$$y = -\frac{1}{u} \quad \text{and} \quad dx = 2 \sin u \, \cos u \, du$$

Now

$$\int x \, dy = xy - \int y \, dx$$

or in this case,

$$\int_{-10\pi}^{10\pi} \frac{\sin^2 u}{u^2} \, du = -\frac{\sin^2 u}{u} \Big|_{-10\pi}^{10\pi} - \int_{-10\pi}^{10\pi} \left(\frac{-1}{u}\right) 2 \sin u \cos u \, du$$

$$= 0 + \int_{-10\pi}^{10\pi} \frac{\sin 2u}{2u} d(2u) = 2 \, \text{Si}(20\pi)$$

where the sine integral of (F-13) has been used. Thus,

$$R_{\text{SD}} = \frac{\pi}{2 \, \text{Si}(20\pi)} = \frac{3.14159}{3.10976} = 1.01$$

and the broadside entry in Table 5-4 has been confirmed.

REFERENCES

1. J. F. Ramsay, "Lambda Functions Describe Antenna/Diffraction Patterns," *Microwave J.*, Vol. 6, pp. 69–107, June 1967.
2. G. A. Thiele, P. L. Detweiler, and R. P. Penno, "On the Lower Bound of the Radiation Q for Electrically Small Antennas," *IEEE Trans. on Ant. and Prop.*, Vol. AP-51 pp. 1263–1268, June 2003.

PROBLEMS

5.1-1 Show that

$$\cos^{-1}(-x) - \cos^{-1}(x) = 2 \sin^{-1}(x)$$

thereby proving (5-11). To do this, introduce α such that $x = \sin \alpha$ and use $\cos(\alpha \pm \pi/2) = \mp \sin \alpha$

5.1-2 Prove the half-power beamwidth expression for an endfire, uniform line source. Start with (5-13) and derive (5-14). *Hint:* Let $\alpha = \cos^{-1}(1 - y)$ where $y = 0.443(\lambda/L)$, then form, $\cos^2 \alpha$, neglect y^2, expand 1 as $\cos^2 \alpha + \sin^2 \alpha$, and use $\sin \alpha \approx \alpha$.

5.1-3 Show that the far-zone electric field expression E_θ for a broadside, uniform line source approximates that of an ideal dipole for short line sources ($L \ll \lambda$).

5.1-4 Compute the half-power beamwidths (in degrees) and the directivities (in decibels) for the following uniform fine sources:

(a) Eight-wavelength broadside, uniform line source
(b) Eight-wavelength endfire, uniform line source
(c) Sixteen-wavelength broadside, uniform line source
(d) Sixteen-wavelength endfire, uniform line source

5.1-5 (a) Use the universal pattern factor for a uniform line source to obtain polar plots of a four-wavelength uniform line source for two cases: broadside and endfire ($\theta_o = 0°$).

(b) Measure the half-power beamwidths from the polar plots obtained in part (a).
(c) Calculate the half-power beamwidths in degrees using (5-12) and (5-14). The agreement between these results and those of (b) depends mainly on how accurately you constructed the polar plot.

5.1-6 Verify the half-power beamwidth values in Table 5-1 for the three levels of approximation for uniform line sources with the following lengths: (a) 2λ, (b) 5λ, and (c) 10λ.

5.1-7 *Uniform line source directivity.*

(a) Show that (5-18) leads to the expression

$$\frac{\beta L}{D_u} = \frac{\cos a - 1}{a} + \frac{\cos b - 1}{b} + \text{Si}(a) + \text{Si}(b)$$

where D_u is the directivity of a uniform line source with excitation phase shift per unit lengthy of β_o, $a = (\beta - \beta_o)L$, $b = (\beta + \beta_o)L$, and Si is the sine integral function defined in (F-13).

(b) Plot the directivity relative to that of the broadside, very long, uniform line source case, that is, $D_u/(2L/\lambda)$, for $\beta L = 10$ and $\beta L = 100$ as a function of θ_o from $90°$ to $0°$.

(c) What does the expression in part (a) reduce to for the broadside case $(\theta_o = 90°)$?

(d) As L becomes much larger than a wavelength, show that your result in part (c) gives (5-19).

(e) Use the result from part (c) and plot the directivity relative to that of a broadside, very long, uniform line source (i.e., $\lambda D_u/2L$) for βL from 1 to 10. This result shows how well the long line-source directivity approximation behaves.

5.2-1 Verify for the cosine-tapered line source pattern of (5-25) that (a) HP $= 1.19(\lambda/L)$ in the broadside case for $L \gg \lambda$, and (b) the side lobe level is -23.0 dB.

5.2-2 Construct the linear, polar plot of the pattern factor for a broadside cosine-tapered line source that is three wavelengths long. Proceed as in Fig. 5-4.

5.2-3 A 3-m long, broadside line source operating at 1 GHz has a cosine-squared tapered current distribution.

(a) Compute the half-power beamwidth in degrees.

(b) Compute the directivity in decibels.

5.2-4 Evaluate the half-power beamwidths in degrees and the directivities in decibels of 10λ long line sources with the following current distributions: (a) uniform, (b) triangular, (c) cosine, (d) cosine-squared, and (e) cosine on a -10-dB pedestal.

5.2-5 *Triangular current-tapered line source.*

(a) From the current distribution in Table 5-2a, derive the normalized pattern factor $f(u)$.

(b) Verify that the half-power beamwidth is given by $1.28(\lambda/L)$ for $L \gg \lambda$ and the side lobe level is -26.6 dB. You may do this by substitution, and also you may find some of the results from the uniform line source helpful.

5.2-6 The pattern from a triangular-tapered current distribution is the square of that of the uniform current distribution. From Fourier transform theory, how are the current distributions related?

5.2-7 Dipole antennas with lengths less than a half-wavelength have current distributions that are nearly triangular (see Fig. 2-16b).

(a) Write the complete electric field expression E_θ in the far field for a broadside line source with a triangular current distribution.

(b) Approximate the expression of part (a) for short dipoles $(L \ll \lambda)$.

(c) Compare this to the far-field expression for E_θ of an ideal dipole. Discuss.

5.2-8 Derive the pattern factor expression in Table 5-2 for a cosine-squared line source current distribution. Also verify the half-power beamwidth expression.

5.2-9 A broadside line source has a cosine on a -10-dB pedestal current distribution. It operates at 200 MHz and has a length of 20 m. Compute (a) the half-power beamwidth in degrees and (b) the directivity in decibels.

5.2-10 Derive the pattern factor expression in Table 5-2 for a cosine on a pedestal current distribution for a line source.

5.2-11 The directivity of a broadside line source can be calculated from

$$D = \frac{2}{\lambda} \frac{\left| \int_{-L/2}^{L/2} I(z)\ dz \right|^2}{\int_{-L/2}^{L/2} |I(z)|^2\ dz}$$

This is the one-dimensional analogy of (9-66). Use this formula to:
- (a) Derive $D_u = 2L/\lambda$, the directivity of a uniform line source.
- (b) Derive an expression for D/D_u of a cosine on a pedestal current distribution. Evaluate for $C = 1, 0.3162, 0.1778$, and 0.

5.4-1 Substitute (5-1) into (5-2) and set the partial derivative of the exponential term with respect to z' equal to zero to derive (5-35). This is called determining the angle of maximum radiation by the principle of stationary phase (see Sec. 16.15).

5.4-2 An array of wires is composed of wires each parallel to the y-axis with their centers on the z-axis. The wires are spaced a quarter wavelength apart. The first wire is located at the coordinate origin, the next at $z = \lambda/4$, and so on. A wave is traveling in the $+z$-direction over the wires and induces currents in them.

- (a) If the current phases, starting at the origin, are $0°$, $-80°$, $-160°$, $-240°$, $-320°$, etc., is the wave fast or slow?
- (b) If the current phases, starting at the origin, are $0°$, $-100°$, $-200°$, $-300°$, $-400°$, etc., is the wave fast or slow?

5.5-1 Verify the ordinary endfire value for R_{SD} in Table 5-4. *Note:* $u = (\beta L/2)(\cos\theta - 1)$.

5.5-2 Verify the Hansen–Woodyard value for R_{SD} in Table 5-4. *Note:* $u = 0.5 [\beta L(\cos\theta - 1) - 2.94]$.

Chapter 6

Wire Antennas

In this chapter, we discuss the important topic of wire antennas. Wire antennas are the oldest and still the most prevalent of all antenna forms. Just about every imaginable shape and configuration of wires has a useful antenna application. Wire antennas can be made from either solid wire or tubular conductors. They are relatively simple in concept, easy to construct and very inexpensive.

To obtain completely accurate solutions for wire antennas, the current on the wire must be solved for, subject to the boundary condition that the tangential electric field is zero along the wire. This approach gives rise to an integral equation, for which many approximate solutions have been reported over the last several decades [1]. These classical solutions are rather tedious and limited to a few simple wire shapes. On the other hand, modern numerical methods implemented on the digital computer are rather simple in concept and applicable to many wire antenna configurations. These numerical (moment method) techniques are discussed in Chap. 14. In this chapter, we adopt a simple approach to solving for the properties of wire antennas. This affords a conceptual understanding of how wire antennas operate, as well as yielding surprisingly accurate engineering results. For example, during the discussion of the loop antenna in Sec. 6.8 a detailed comparison of results from simple theory and the more exact numerical methods demonstrates the accuracy of simple theory.

In this chapter, we discuss several resonant wire antennas such as straight wire dipoles, vee dipoles, folded dipoles, Yagi-Uda arrays, and loops. A resonant antenna is a standing wave antenna (e.g., a dipole) with zero input reactance at resonance.

Other wire antennas that are broadband, such as traveling-wave antennas, the helix, and log-periodic, are presented in the next chapter. Methods of feeding wire antennas and their performance in the presence of an imperfect ground plane are included here. Most of the developments in this chapter utilize the principles set forth thus far. Design data and guidelines for the construction and use of wire antennas are emphasized.

6.1 DIPOLE ANTENNAS

We have discussed short dipoles in Secs. 2.3 and 3.1 and the half-wave dipole in Sec. 3.2. In this section, dipoles of arbitrary length are examined. The dipole antenna has received intensive study [1]. We will use a simple but effective approach that involves an assumed form for the current distribution. The radiation integral may then be evaluated and thus also the pattern parameters. For dipoles, we assume that the current distribution is sinusoidal. This is a good approximation verified by measurements. The current must, of course, be zero at the ends. We are, in effect, using the current distribution that is found on an open-circuited parallel wire transmission line. It is assumed that if the end of such a transmission line is bent out to form a wire antenna, the current distribution along the bent

portion is essentially unchanged. Although this is not strictly true, it is a good approximation for thin antennas, for which the conductor diameter is on the order of 0.01λ or smaller.

6.1.1 Straight Wire Dipoles

A straight dipole antenna is shown in Fig. 6-1 oriented along the z-axis. It is fed at the center from a balanced two-wire transmission line, that is, the currents on each wire are equal in magnitude and opposite in direction. The current distribution along the antenna is assumed to be sinusoidal and can be written as

$$I(z) = I_m \, \sin\left[\beta\left(\frac{L}{2} - |z|\right)\right], \quad |z| < \frac{L}{2} \tag{6-1}$$

The dipole is surrounded by free space, thus, the phase constant is that of free space, β.

It is helpful to visualize the current distribution on an antenna. Fig. 6-2 shows the current on a dipole for $L < \lambda/2$. The solid lines indicate actual currents on the antenna and the dotted lines indicate extensions of the sine wave function. As a note of caution with this visualization, the dotted portion of the current distribution does not appear on the transmission line. For this case, I_m in (6-1) is not the maximum current attained on the antenna. The maximum current on the antenna shown in Fig. 6-2 is at the input terminals where $z = 0$ and is of a value $I_m\sin(\beta L/2)$. The arrows in Fig. 6-2 show the current direction. The currents on the top and bottom halves of the antenna are in the same
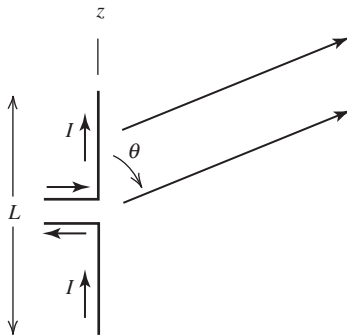
Figure 6-1 The dipole antenna.
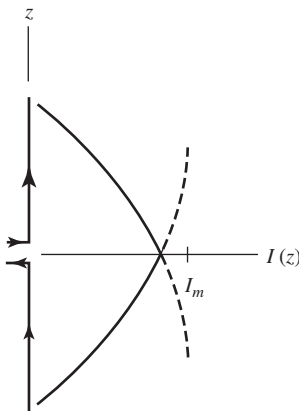
Figure 6-2 Current distribution on a dipole of length $L < \lambda/2$.
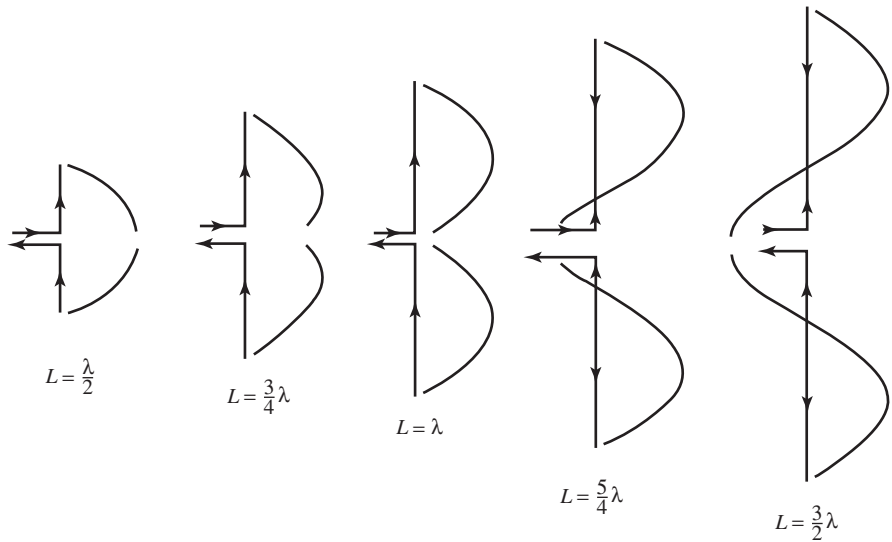
Figure 6-3 Current distributions for various center-fed dipoles. Arrows indicate relative current directions for these maximum current conditions.

direction at any instant of time, and thus the radiation effects from each half reinforce. The transmission line, however, has oppositely directed currents that have canceling radiation effects for typical close conductor spacings. (See Fig. 3-4*a*.)

In Fig. 6-3, current distributions on various dipoles are plotted together with the antennas used to generate them. The sinusoidal curves superimposed on the antennas indicate the intensity of the current on the wire—that is, the value of the curve at point z is the current value *on the wire* at the same point z. Again, the arrows indicate current directions. To construct plots such as these, begin on the z-axis at one end of the wire where the current is zero and draw a sine wave while moving toward the feed point. The current on the other half is then the mirror image. For dipoles longer than one wavelength, the currents on the antenna are not all in the same direction. Over a half-wave section, the current is in-phase and adjacent half-wave sections are of opposite phase. We would then expect to see some large canceling effects in the radiation pattern. This will be shown later to be precisely what happens. For all the current distributions presented, the plots represent the maximum excitation state. It is assumed that a sinusoidal waveform generator of radian frequency $\omega = 2\pi c/\lambda$ is connected to the input transmission line. The standing wave pattern of the current at any instant of time is obtained by multiplying (6-1) by cos ωt, which follows from (2-6).

To obtain the dipole radiation pattern, we first evaluate the radiation integral

$$f_{\text{un}} = \int_{-L/2}^{L/2} I(z') \ e^{j\beta z' \cos\theta} \ dz' \tag{6-2}$$

Substituting the current expression from (6-1) gives

$$f_{\text{un}} = \int_{-L/2}^{0} I_m \sin\left[\beta\left(\frac{L}{2} + z'\right)\right] e^{j\beta z' \cos\theta} \ dz'$$

$$+ \int_{0}^{L/2} I_m \ \sin\left[\beta\left(\frac{L}{2} - z'\right)\right] e^{j\beta z' \cos\theta} \ dz' \tag{6-3}$$

Evaluating these integrals (see Prob. 3.2-6) gives the unnormalized pattern

$$f_{un} = \frac{2I_m}{\beta} \frac{\cos[(\beta L/2) \ \cos\theta] - \cos(\beta L/2)}{\sin^2\theta} \tag{6-4}$$

Using this in (2-103) and (2-106) leads to the complete far-zone electric field

$$E_\theta = j\omega\mu \ \sin\theta \frac{e^{-j\beta r}}{4\pi r} \frac{2I_m}{\beta} \frac{\cos[(\beta L/2) \ \cos\theta] - \cos(\beta L/2)}{\sin^2\theta} \tag{6-5}$$

Noting that $\omega\mu/\beta = \eta$, we see that this expression simplifies to

$$E_\theta = j\eta \frac{e^{-j\beta r}}{2\pi r} I_m \frac{\cos[(\beta L/2) \ \cos\theta] - \cos(\beta L/2)}{\sin\theta} \tag{6-6}$$

The θ-variation of this function determines the far-field pattern. For $L = \lambda/2$, it is

$$F(\theta) = \frac{\cos[(\pi/2) \ \cos\theta]}{\sin\theta} \quad (L = \lambda/2) \tag{6-7}$$

This expression was also derived in Sec. 3.2; see (3-4). This is the normalized electric field pattern of a **half-wave dipole**. The half-power beamwidth is 78° and its pattern plot is shown in Fig. 6-4a.
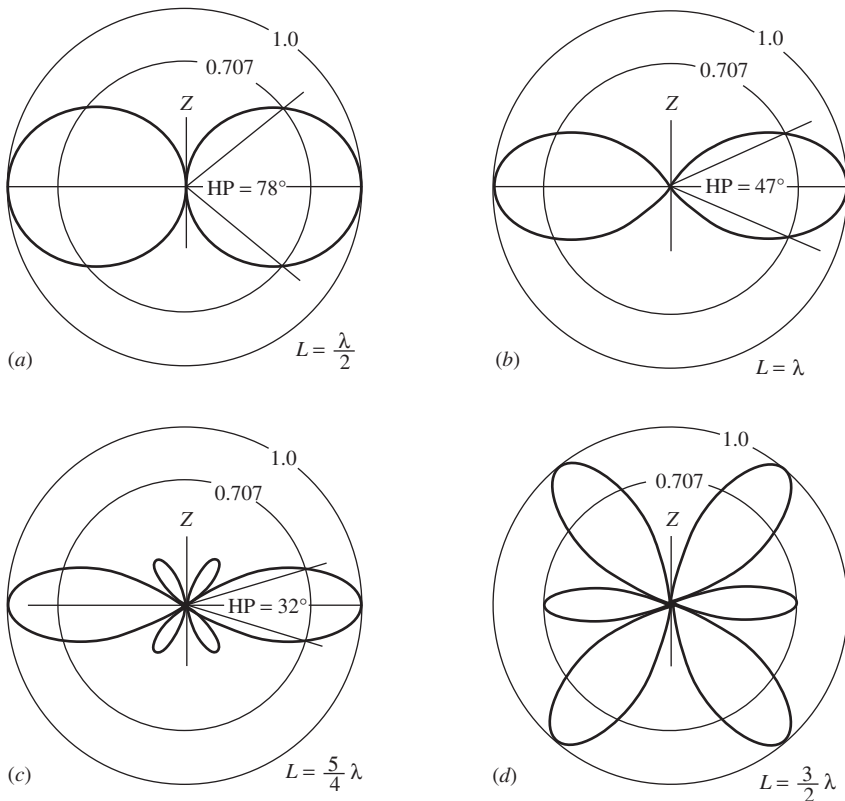
Figure 6-4 Radiation patterns of center-fed straight dipole antennas of length L.

For a center-fed dipole with $L = \lambda$, the normalized electric field pattern from (6-6) is

$$F(\theta) = \frac{\cos(\pi \cos \theta) + 1}{2 \sin \theta} \quad (L = \lambda) \tag{6-8}$$

The half-power beamwidth for this **full-wave dipole** is $47°$. Its pattern is shown in Fig. 6-4b. If $L = \frac{3}{2}\lambda$, the pattern function is

$$F(\theta) = 0.7148 \frac{\cos\left(\frac{3}{2}\pi \cos \theta\right)}{\sin \theta} \quad \left(L = \frac{3}{2}\lambda\right) \tag{6-9}$$

The factor 0.7148 is the normalization constant. As predicted earlier for dipoles of length greater than one wavelength, the pattern of the three-halves wavelength dipole shown in Fig. 6-4d has a multiple lobe structure due to the canceling effect of oppositely directed currents on the antenna. This effect is also visible in the $\frac{5}{4}$ wavelength case.

As L/λ becomes very small, the dipole pattern variation in (6-6) approaches $\sin \theta$. Thus, we see again that the pattern of a short dipole along the z-axis is $\sin \theta$. Recall that the short dipole pattern has a $90°$ half-power beamwidth; see Fig. 2-4b.

To obtain the radiation resistance, first the radiated power must be found. Substituting (6-6) into (2-130) gives

$$P = \frac{1}{2\eta} \int_0^{2\pi} \int_0^{\pi} \eta^2 \frac{I_m^2}{(2\pi r)^2} \left\{ \frac{\cos[(\beta L/2) \cos \theta] - \cos(\beta L/2)}{\sin \theta} \right\}^2 r^2 \sin \theta \, d\theta \, d\phi$$

$$= \frac{\eta}{8\pi^2} I_m^2 \int_0^{2\pi} d\phi \; 2 \int_0^{\pi/2} \frac{\{\cos[(\beta L/2) \cos \theta] - \cos(\beta L/2)\}^2}{\sin \theta} \, d\theta \tag{6-10}$$

Changing the integration variable to $\tau = \cos \theta$, $d\tau = -\sin \theta \, d\theta$, gives

$$P = \frac{\eta}{2\pi} I_m^2 \int_1^0 \frac{\{\cos[(\beta L/2)\tau] - \cos(\beta L/2)\}^2}{1 - \tau^2} (-d\tau)$$

$$= \frac{\eta}{4\pi} I_m^2 \int_0^1 \left(\frac{\{\cos[(\beta L/2)\tau] - \cos(\beta L/2)\}^2}{1 + \tau} \right.$$

$$\left. + \frac{\{\cos[(\beta L/2)\tau] - \cos(\beta L/2)\}^2}{1 - \tau} \right) d\tau \tag{6-11}$$

where in the last expression the identity

$$\frac{1}{1 - u^2} = \frac{1}{2}\left(\frac{1}{1 + u} + \frac{1}{1 - u} \right) \tag{6-12}$$

was used. Eq. (6-11) can be evaluated in terms of sine and cosine integral functions; see (F-13) and (F-14). A simpler expression for the special case of the half-wave dipole is obtainable in terms of a single cosine integral function. Thus, when $\beta L/2 = \pi/2$, (6-11) becomes

$$P = \frac{\eta}{4\pi} I_m^2 \int_0^1 \left[\frac{\cos^2(\pi\tau/2)}{1 + \tau} + \frac{\cos^2(\pi\tau/2)}{1 - \tau} \right] d\tau \tag{6-13}$$

Changing variables again as $v = 1 - \tau$ and $w = 1 + \tau$ and substituting into (6-13)

$$P = \frac{\eta}{4\pi} I_m^2 \left[\int_1^0 \frac{-\sin^2(\pi v/2)}{v} dv + \int_1^2 \frac{\sin^2(\pi w/2)}{w} dw \right] = \frac{\eta}{4\pi} I_m^2 \int_0^2 \frac{\sin^2(\pi v/2)}{v} dv$$

$$= \frac{\eta}{4\pi} I_m^2 \int_0^2 \frac{1 - \cos \pi v}{2v} \, dv \tag{6-14}$$

Changing the variable of integration to $t = \pi v$ leads to

$$P = \frac{\eta}{8\pi} I_m^2 \int_0^{2\pi} \frac{1 - \cos t}{t} \, dt = \frac{\eta}{8\pi} I_m^2 \, \text{Cin}(2\pi) = \frac{\eta}{8\pi} I_m^2 (2.44) \tag{6-15}$$

where $\text{Cin}(x)$ is related to the cosine integral function by (F-16) and is tabulated in [2]. In this case, $\text{Cin}(2\pi) = 2.44$. Using this and $\eta = 120\pi$ in (6-15) leads to the radiation resistance for a half-wave dipole as

$$R_r = \frac{2P}{I_m^2} = \frac{2(15 \ I_m^2 \ 2.44)}{I_m^2} = 73 \, \Omega \quad \left(L = \frac{\lambda}{2} \right) \tag{6-16}$$

The infinitely thin dipole antenna also has a reactive impedance component. For the half-wave dipole, the reactance is inductive, and the complete input impedance is

$$Z_A = 73 + j42.5 \, \Omega \quad \left(L = \frac{\lambda}{2} \right) \tag{6-17}$$

This can be calculated for an infinitely thin dipole by a classical procedure known as the induced emf method [H.4: Jordan & Balmain, Chap. 14]. However, the input impedance of dipoles with finite wire diameter can be calculated using the moment method of Chap. 14, where the form of the current is not assumed. The results of such a calculation for the input resistance and reactance of a small-diameter, center-fed dipole are given in Figs. 6-5 and 6-6. The resonance effects are evident in these plots. Note that the input reactance is capacitive for small lengths, as we pointed out in Sec. 3.l.

The dotted curve in Fig. 6-5 is the input resistance from (2-169) for an ideal dipole with uniform current. It does not give good results for an actual wire dipole as shown by the solid curve of Fig. 6-5. However, the triangular current approximation with $R_{ri} = 20\pi^2 (L/\lambda)^2$ from (2-172) does give a good approximation to the input radiation resistance for short dipoles as demonstrated by the dashed curve of Fig. 6-5. Some simple formulas that approximate the input resistance of wire dipoles are given in Table 6-1 [3]. For example, using the second formula for $L = \lambda/2$ gives $R_{ri} = 24.7(\pi/2)^{2.4} = 73.0 \, \Omega$, which agrees with (6-17). The values obtained from Table 6-1 also agree closely with those of Fig. 6-5.

Input resistance is related to radiation resistance. There are several ways to define radiation resistance by using different current reference points. Usually, radiation resistance is defined using the current distribution maximum I_m, whether or not it actually occurs on the antenna. We shall use the symbol R_{rm} for this definition. It is also useful to refer the radiation resistance to the input terminal point. In this case, the symbol R_{ri} is used. These definitions can be related by writing the radiated power as

$$P = \frac{1}{2} I_m^2 R_{rm} = \frac{1}{2} I_A^2 R_{ri} \tag{6-18}$$

For dipoles that are odd integer multiples of a half-wavelength long, $I_m = I_A$ and $R_{rm} = R_{ri}$. A third radiation resistance, denoted by R_r, is often used; it is the radiation
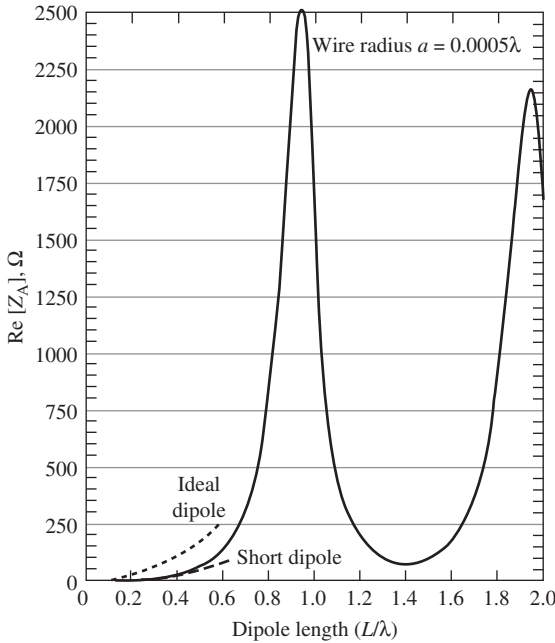
Figure 6-5 Calculated input resistance of a center-fed wire dipole of 0.0005λ radius as a function of length L (solid curve). Also shown is the input resistance $R_{ri} = 80\pi^2(L/\lambda)^2$ of an ideal dipole with a uniform current distribution (dotted curve) and the input resistance $R_{ri} = 20\pi^2(L/\lambda)^2$ of a short dipole with a triangular current distribution approximation (dashed curve).
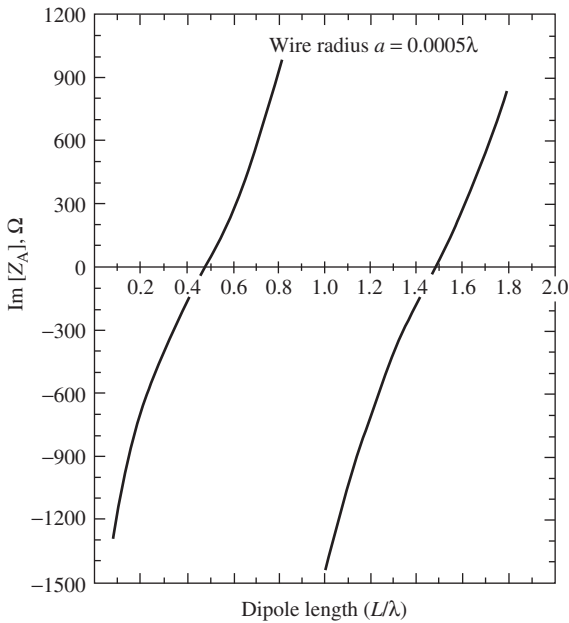
Figure 6-6 Calculated input reactance of center-fed wire dipole of radius 0.0005λ as a function of length L.

resistance relative to the maximum current that occurs on the antenna. For dipoles less than a half-wavelength long, the current maximum on the antenna always occurs at the center, and then $R_{ri} = R_r$ for center-fed dipoles; this was discussed in Sec. 2.6. In practice, we are interested in input resistance, so R_{ri} is of primary importance. It is related to R_{rm} for center-fed dipoles by setting $z = 0$ in (6-1), giving

Table 6-1 Simple Formulas for the Input
Resistance of Dipoles

Length L	Input Resistance $(R_{ri}), \Omega$
$0 < L < \dfrac{\lambda}{4}$	$20\pi^2 \left(\dfrac{L}{\lambda}\right)^2$
$\dfrac{\lambda}{4} < L < \dfrac{\lambda}{2}$	$24.7 \left(\pi\dfrac{L}{\lambda}\right)^{2.4}$
$\dfrac{\lambda}{2} < L < 0.637\lambda$	$11.14 \left(\pi\dfrac{L}{\lambda}\right)^{4.17}$

$$I_A = I_m \ \sin\frac{\beta L}{2} \tag{6-19}$$

and substituting into (6-18), which yields

$$R_{ri} = \frac{I_m^2}{I_A^2} R_{rm} = \frac{R_{rm}}{\sin^2(\beta L/2)} \tag{6-20}$$

R_{ri} is the component of input resistance due to radiation and equals the total input resistance R_A if ohmic losses are neglected, which we shall do unless otherwise indicated.

For dipole lengths, $L = \lambda, 2\lambda, 3\lambda, \ldots, \beta L/2 = \pi, 2\pi, 3\pi, \ldots$, and R_{ri} from (6-20) is infinite. For example, the one-wavelength dipole of Fig. 6-3c has a current zero at its feed point and thus an infinite input impedance. This, of course, is based on the perfect sine wave current distribution. Dipoles of finite thickness have large but finite values of input impedance for lengths near integer multiples of wavelength, as seen in Fig. 6-5. This effect arises from the deviation of the current distribution from that of (6-1) for dipole lengths near integer multiples of the wavelength: there is always a finite input current on an actual dipole. For other-length dipoles, the sinusoidal current distribution is a good approximation for thin-wire dipole antennas.

By reducing the length of the half-wave dipole slightly, the antenna can be made to resonate $(X_A = 0)$. The input impedance of the infinitely thin half-wave dipole is then about $70 + j0\,\Omega$. In Fig. 5-6, the dipole of radius 0.0005λ resonates for lengths corresponding to the intersections with the horizontal $(X_A = 0)$ axis. The first intersection is the half-wave dipole case and the resonant length is slightly less than $\lambda/2$. It turns out that as the wire thickness increases, the dipole must be shortened more to obtain resonance. Approximate length values for resonance are given in Table 6-2. For the dipole of 0.0005λ wire radius, the length-to-diameter ratio, $L/2a$, is 500 for the half-wave case. From Table 6-2, we see that about 4% shortening $(L = 0.48\lambda)$ would be required to produce resonance. This agrees closely with the resonance point from Fig. 6-6. In practice, wire antennas are constructed slightly longer than required. Then a transmitter is connected to the antenna and the standing wave ratio (or reflected power) is monitored on the feed transmission line. The ends of the antenna are trimmed until a low value of standing wave ratio is obtained. Note that as the length is reduced to obtain resonance, the input resistance also decreases. For example, for a thin dipole with $L/2a = 50$ and $L = 0.475\lambda$, the second formula of Table 6-1 gives $R_A = 64.5\,\Omega$; the reactance is, of course, zero.

Table 6-2 Wire Lengths Required to Produce a Resonant
Half-Wave Dipole for a Wire Diameter of $2a$ and Length L

Length to Diameter Ratio, $L/2a$	Percent Shortening Required	Resonant Length L	Dipole Thickness Class
5000	2	0.49λ	Very thin
50	5	0.475λ	Thin
10	9	0.455λ	Thick
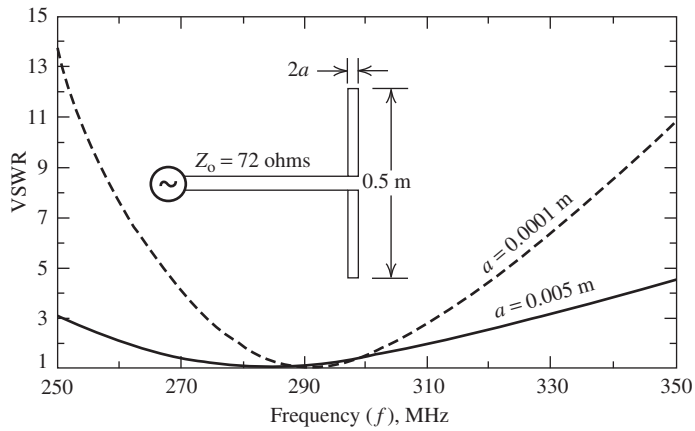
Figure 6-7 Calculated VSWR as a function of frequency for dipoles of different wire diameters.

Since dipoles are resonant-type structures, their bandwidth is low. The VSWR as a function of frequency for a half-wave dipole is shown in Fig. 6-7. In general, bandwidth is defined as "the range of frequencies within which the performance of the antenna, with respect to some characteristic, conforms to a specified standard" [4]. In this case, let the specified standard be a VSWR less than 2.0:1. From Fig. 6-7, we see that the bandwidths are $310 - 262 = 48$ MHz and $304 - 280 = 24$ MHz, respectively, for $a = 0.005$ m ($L/2a = 50$) and $a = 0.0001$ m ($L/2a = 2500$). In terms of percent relative to the design frequency (300 MHz), the bandwidths are 16% and 8%. It is an important general principle that the thicker the dipole, the wider is its bandwidth. Also, note that the minimum VSWR for the thicker dipole occurs at a lower frequency than for the thinner one. In fact, if we use the rules in Table 6-2, the resonant frequencies are calculated to be 285 and 294 MHz for wire radii of 0.005 and 0.0001 m. These values agree well with the minimum points of the curves in Fig. 6-7.

The improved bandwidth offered by the thick circularly cylindrical dipole in Fig. 6-7 can also be achieved with a flat metallic strip as Fig. 6-8 indicates. The relationship between the circularly cylindrical dipole radius and the width of the metallic strip for equivalent performance under certain conditions is $a = 0.25w$ [5]. The advantage of the flat strip dipole is primarily ease of construction.

Finally, the directivity of a half-wave dipole is found from $D = 4\pi U_m/P$. The radiated power P was evaluated in (6-15). Using the far-zone electric field of (6-6) leads to the maximum radiation intensity as
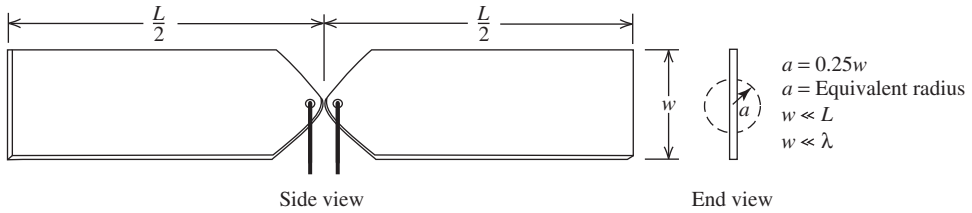
Figure 6-8 Thin metal strip dipole.

$$U_m = \frac{r^2}{2\eta}|E_\theta|^2_{\max} = \frac{1}{2\eta}\frac{\eta^2 I_m^2}{(2\pi)^2} = \frac{\eta}{8\pi^2}I_m^2 \tag{6-21}$$

So,

$$D_{(\lambda/2)\text{ dipole}} = \frac{4\pi U_m}{P} = \frac{4\pi(\eta/8\pi^2)I_m^2}{(\eta/8\pi)I_m^2(2.44)} = 1.64 = 2.15 \text{ dB} \tag{6-22}$$

This is only slightly greater than the directivity value of 1.5 for an ideal dipole with uniform current. So for very short dipoles, the directivity is 1.5 and increases to 1.64 as the length is increased to a half-wavelength. As length is increased further, directivity also increases. A full-wave dipole has a directivity of 2.41. Even more directivity is obtained for a length of about 1.25λ. As the length is increased further, the pattern begins to break up (see Fig. 6-4d) and directivity drops sharply. See Prob. 6.1-12.

6.1.2 The Vee Dipole

Wire dipole antennas that are not straight also appear in practice. One such antenna is the **vee dipole** shown in Fig. 6-9. This antenna may be visualized as an open-circuited transmission line that has been bent so that ends of length h have an included angle of γ. The angle γ for which the directivity is greatest in the direction of the bisector of γ is given by

$$\gamma = 152\left(\frac{h}{\lambda}\right)^2 - 388\left(\frac{h}{\lambda}\right) + 324, \qquad 0.5 \le \frac{h}{\lambda} < 1.5$$

$$\gamma = 11.5\left(\frac{h}{\lambda}\right)^2 - 70.5\left(\frac{h}{\lambda}\right) + 162, \qquad 1.5 \le \frac{h}{\lambda} \le 3.0 \tag{6-23}$$

where the resulting angle γ is in degrees. The corresponding directivity is

$$D = 2.95\left(\frac{h}{\lambda}\right) + 1.15 \tag{6-24}$$
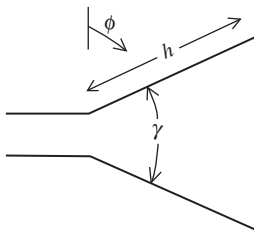
Figure 6-9 The vee dipole antenna.
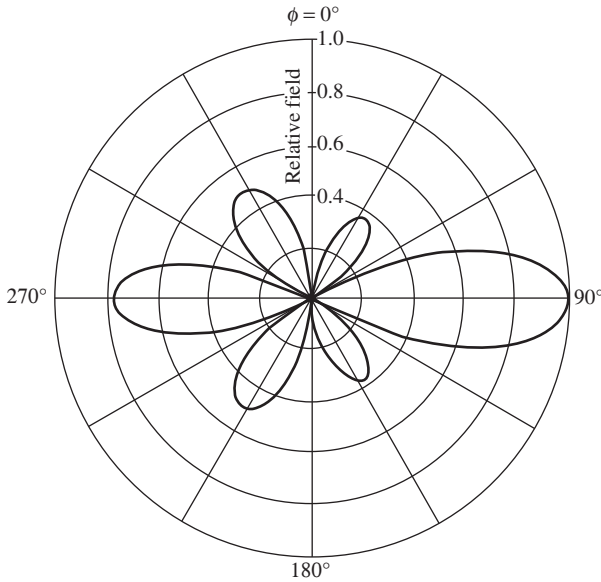
Figure 6-10 Far-field pattern of a vee dipole shown in Fig. 6-9 with arm length $h = 0.75\lambda$, $\gamma = 118.5°$, and $a = 0.0005\lambda$.

These equations were empirically derived for antennas with $0.5 \le h/\lambda \le 3.0$ using the computational methods (MoM) of Chap. 14.

The directivity of a vee dipole can be greater than that of a straight dipole. This can be seen from the pattern in Fig. 6-10 where $h = 0.75\lambda$ and λ from (6-23) is $118.5°$. Notice that the direction of maximum radiation is $\phi = 90°$ while radiation in the $\phi = 270°$ direction is about 2 dB less. Even more significant is the low level of the side lobes. For the most part, it is the reduced side lobe levels of the vee dipole that give it a greater directivity than the straight dipole version (see Fig. 6-4d). The directivity for the vee dipole of Fig. 6-10 from (6-24) is $D = 2.94(0.75) + 1.15 = 3.355 = 5.26$ dBi. The directivity of a 1.5λ long straight wire dipole is about 2.2, or 3.4 dBi.

The input impedance of a vee dipole antenna is generally less than that of a straight dipole of the same length. For example, the input impedance of the vee dipole in Fig. 6-10 is $106 + j17\,\Omega$, which is less than for the straight dipole version ($L = 1.5\lambda$) as found from Figs. 6-5 and 6-6.

6.2   FOLDED DIPOLE ANTENNAS

An extremely practical wire antenna is the **folded dipole**. It consists of two parallel dipoles connected at the ends forming a narrow wire loop, as shown in Fig. 6-11 with dimension d much smaller than L and much smaller than a wavelength. The feed point is at the center of one side. The folded dipole is essentially an unbalanced transmission line with unequal currents. Its operation is analyzed by considering the current to be composed of two modes: the transmission line mode and the antenna mode. The currents for these modes are illustrated in Fig. 6-12.

The currents in the transmission line mode have fields that tend to cancel in the far field since d is small. The input impedance Z_t for this mode is given by the equation for a transmission line with a short circuit load (see Prob. 6.2-3)

$$Z_t = jZ_o \, \tan \, \beta \frac{L}{2} \tag{6-25}$$

where Z_o is the characteristic impedance of the transmission line.
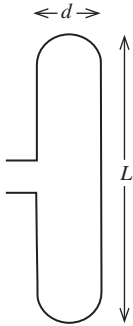
Figure 6-11 The folded dipole antenna.

(*a*) Transmission line mode.   (*b*) Antenna mode.
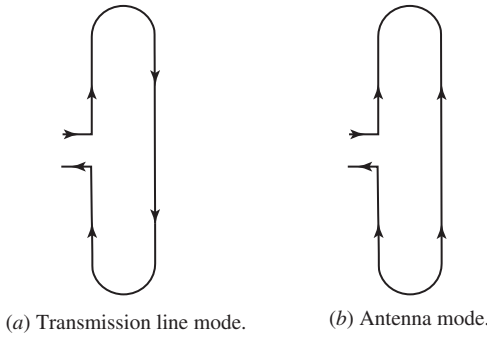
Figure 6-12 The current modes on a folded dipole antenna.

(*a*) Transmission line mode.   (*b*) Antenna mode.
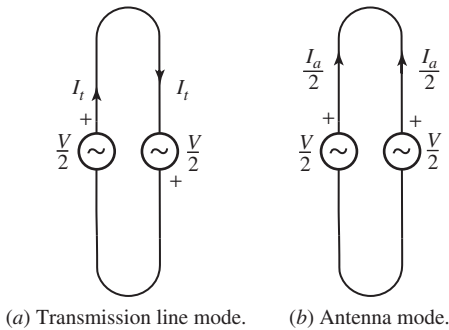
Figure 6-13 Mode excitation and current for a voltage *V* applied to the terminals of a folded dipole. Superposition of these modes gives the complete folded dipole model.

In the antenna mode, the fields from the currents in each vertical section reinforce in the far field since they are similarly directed. In this mode the charges "go around the corner" at the end, instead of being reflected back toward the input as in an ordinary dipole, which leads to a doubling of the input current for resonant lengths. The result of this is that the antenna mode has an input current that is half that of a dipole of resonant length.

Suppose a voltage *V* is applied across the input terminals of a folded dipole. The total behavior is determined by the superposition of the equivalent circuits for each mode in Fig. 6-13. Note that if the figures for each mode are superimposed and the voltages are added, the total on the left is *V* and on the right is zero, as it should be. The transmission line mode current is

$$I_t = \frac{V}{2Z_t} \tag{6-26}$$

For the antenna mode, the total current is the sum of each side, or I_α. The excitation for this current is $V/2$; thus, the antenna current is

$$I_\alpha = \frac{V}{2Z_d} \tag{6-27}$$

where to a first-order approximation Z_d is the input impedance for an ordinary dipole of the same wire size. The total current on the left is $I_t + \frac{1}{2}I_\alpha$ and the total voltage is V, so the input impedance of the folded dipole is

$$Z_A = \frac{V}{I_t + \frac{1}{2}I_a} \tag{6-28}$$

Substituting (6-26) and (6-27) in (6-28) yields

$$Z_A = \frac{4Z_t Z_d}{Z_t + 2Z_d} \tag{6-29}$$

As an example, consider the popular **half-wave folded dipole**. From (6-25) with $L = \lambda/2$, $Z_t = jZ_o \ \tan[(2\pi/\lambda)(\lambda/4)] = jZ_o \ \tan(\pi/2) = \infty$. Then (6-29) gives

$$Z_A = 4Z_d \quad \left(L = \frac{\lambda}{2}\right) \tag{6-30}$$

Thus, the half-wave folded dipole provides a four-fold increase in impedance over its dipole version. Since the half-wave dipole (at resonance) has a real input impedance, the half-wave folded dipole has also.

The current on the half-wave folded dipole is particularly easy to visualize. We will discuss this current and also rederive the impedance. If the vertical wire section on the right in Fig. 6-11 were cut directly across from the feed point and the wire folded out without disturbing the current, it would appear as shown in Fig. 6-14. The current is not zero at the ends, because they are actually connected. Perhaps a better way to view this is to fold the current back down and note that currents on the folded part are now upside down, as shown in Fig. 6-15a. The same total current (and thus the same pattern) is obtained with both the folded and the ordinary dipoles in Fig. 6-15. The difference is that the folded dipole has two closely spaced currents equal in value, whereas in the ordinary dipole they are combined on one wire. From this, it is easy to see that the input currents in the two cases are related as

$$I_f = \frac{1}{2}I_d \quad \left(L = \frac{\lambda}{2}\right) \tag{6-31}$$
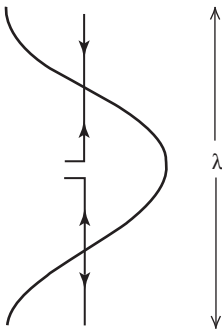
Figure 6-14 Current for the antenna mode of a half-wave folded dipole that has been folded out without disturbing the current.
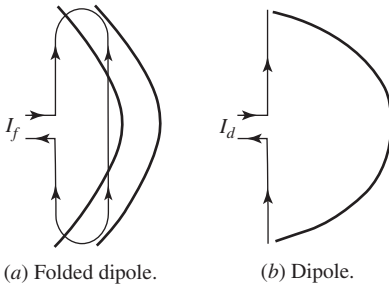
(*a*) Folded dipole.   (*b*) Dipole.   **Figure 6-15** Currents on half-wave dipoles.

The input powers are

$$P_f = \frac{1}{2} Z_A I_f^2 \quad \text{(folded dipole)} \tag{6-32}$$

and

$$P_d = \frac{1}{2} Z_d I_d^2 \quad \text{(dipole)} \tag{6-33}$$

Since the total currents are the same in the half-wave case, the radiated powers are also. Equating (6-32) and (6-33) and using (6-31) give

$$\frac{1}{2} Z_d I_d^2 = \frac{1}{2} Z_A \frac{1}{4} I_d^2$$

or

$$Z_A = 4Z_d \quad \left(L = \frac{\lambda}{2}\right) \tag{6-34}$$

This result is an independent confirmation of the result in (6-30).

The input impedance of a half-wave folded dipole (at resonance) is four times that of an ordinary dipole. A resonant half-wave dipole has about $70\,\Omega$ of input resistance, so a half-wave folded dipole then has an input impedance of

$$Z_f = 4(70) = 280\,\Omega \quad \text{(half-wave)} \tag{6-35}$$

This impedance is very close to the $300\,\Omega$ of common twin-lead transmission line.

The input resistance curve for a folded dipole of finite wire thickness are given in Fig. 6-16 as a function of folded dipole length. The solid curves were obtained from the transmission line model. The wire spacing $d = 12.5a$ is such that the characteristic impedance corresponds to a $300\,\Omega$ transmission line $[Z_o = (\eta/\pi)\ln(d/a) = 120\ln(12.5a/a) \approx 300\,\Omega]$. The folded dipole input impedance is then found from (6-25) and (6-29). As an example, consider a folded dipole of length $L = 0.8\lambda$, spacing $d = 12.5a$, and radius $a = 0.0005\lambda$. From (6-25),

$$Z_t = j300 \tan 0.8\pi = -j218\,\Omega \tag{6-36}$$

From Figs. 6-5 and 6-6,

$$Z_d = 950 + j950 \tag{6-37}$$
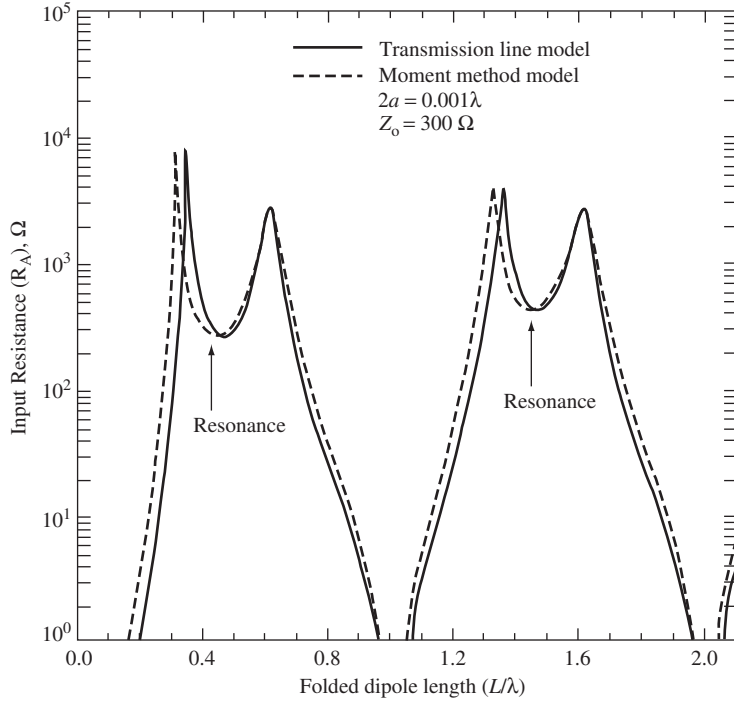
Figure 6-16 Input impedance of a folded dipole. The solid curves are calculated from the transmission line model. The dashed curves are calculated from more accurate moment methods. The wire radius a is 0.0005λ and wire spacing d is $12.5a$.

Using these in (6-29) yields

$$Z_A = 28 - j461 \ \Omega \qquad (L = 0.8\lambda) \tag{6-38}$$

This result agrees with the value shown in Fig. 6-16 for R_A.

Also shown in Fig. 6-16 as dashed curves is the input impedance calculated using the moment methods of Chap. 14. The agreement between the simple transmission line model and the simulation results is quite good. Both methods show that the real part of the input impedance is slightly less than $300\,\Omega$ at the first resonance ($L \approx 0.48\lambda$) and slightly larger than $300\,\Omega$ at the second resonance ($L \approx 1.47\lambda$). It is this characteristic of the folded dipole that makes it useful at harmonically related frequencies. Note, too, the very low value of Z_A when $L \approx \lambda, 2\lambda, \dots$. This can easily be explained from the transmission line model, since then $\tan(\beta L/2) \approx \tan \pi = 0$ and thus $Z_t = 0$ and Z_A from (6-29) is zero.

The folded dipole is used as an FM broadcast band receiving antenna, and it can be simply constructed by cutting a piece of $300\,\Omega$ twin-lead transmission line about a half-wavelength long (1.5 m at 100 MHz). The ends are soldered together such that the overall length L is slightly less than a half-wavelength at the desired frequency (usually 100 MHz). One wire is then cut in the middle and connected to the twin-lead transmission line feeding the receiver.

Occasionally, two different wire sizes are used for a folded dipole as shown in Fig. 6-17. The input impedance for the half-wave case is given by

$$Z_A = (1 + c)^2 Z_d \quad \left(L = \frac{\lambda}{2} \right) \tag{6-39}$$

(a) Side view.   (b) End view.
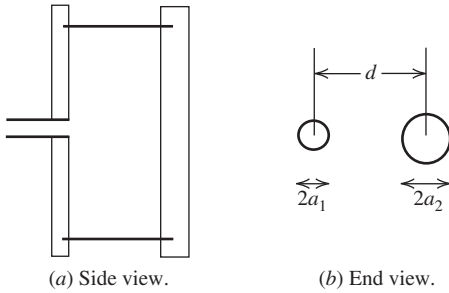
Figure 6-17 Folded dipole antenna constructed from two different size conductors.

For given values of d, a_1 and a_2, the value of c can be found [H.6: *Ant. Eng. Hdbk.*, 4th ed., pg. 4-16]. As is frequently the case, if a_1 and a_2 are much less than d, c is approximately given by

$$c \approx \frac{\ln(d/a_1)}{\ln(d/a_2)} \tag{6-40}$$

The folded dipole antenna is a very popular wire antenna. The reasons for this are its impedance properties, ease of construction, and structural rigidity. The equal-size conductor half-wave folded dipole has an input impedance very close to that of a 300-ohm twin-lead-type transmission line as seen from Fig. 6-16. Also, by changing the conductor radii, the input impedance can be changed. In addition to having desirable impedance properties, the half-wave folded dipole has a wider bandwidth than an ordinary half-wave dipole. In part for these reasons, a folded dipole is frequently used as a feed antenna for Yagi-Uda arrays and other popular antennas.

6.3   YAGI-UDA ANTENNAS

One way to obtain more gain out of dipole type antennas is to use more than one in an array configuration. Usually arrays have all elements identical and are active, requiring a direct connection to each element by a feed network. This kind of array is studied in Chap. 8, whereas the array configuration studied in this section usually has only one element directly connected to a transmitter and/or receiver while the other elements are parasitic. Such an array is referred to as a *parasitic array*. The elements that are not directly driven (called parasitic elements) receive their excitation by near-field electromagnetic coupling from the driven element. A parasitic linear array of parallel dipoles is called a **Yagi-Uda antenna**, a Yagi-Uda array, or simply "Yagi." Yagi-Uda antennas are very popular because of their simplicity and relatively high gain. In this section, the principles of operation and design data for Yagis will be presented [6].

The first research done on the Yagi-Uda antenna was performed by Shintaro Uda at Tohoku University in Sendai, Japan in 1926 and was published in Japanese in 1926 and 1927. The work was reviewed in an article written in English by Uda's professor, H. Yagi, in 1928 [7]. The Yagi antenna, as it is popularly called, was used by the British in their pre–World War II VHF radars before the availability of high-power microwave sources made possible radar at microwave frequencies. The Yagi remains a popular HF and VHF antenna today for non-radar use.

The basic unit of a Yagi consists of three elements. To understand the principles of operation for a three-element Yagi, we begin with a driven element (or "driver") and add parasites to the array. Consider a driven element that is a resonant half-wave dipole. If a parasitic element is positioned very close to it, the field incident on the parasite element is that of the field leaving the driver:

$$E_{\text{incident}} = E_{\text{driver}} \qquad (6\text{-}41)$$

The current excited on the parasite and the resulting radiated electric field, E_{parasite}, also tangent to the wire, is almost equal in amplitude and opposite in phase to the incident wave from the driver due to the small separation distance between them. This is because the electric field arriving at the parasite from the driver is tangential to it and the total electric field tangential to a perfect electric conductor is zero. Thus, the field radiated by the parasite is such that the total tangential field on the parasite is zero, or $0 = E_{\text{incident}} + E_{\text{parasite}}$. Combining this fact with (6-41) gives

$$E_{\text{parasite}} = -E_{\text{incident}} = -E_{\text{driver}} \qquad (6\text{-}42)$$

From the two-element array theory in Sec. 3.5, we know that two closely spaced equal amplitude, opposite phase elements will have an endfire pattern; for example, see Fig. 3-17 for $d = \lambda/8$ and $\alpha = 180°$. The difficult-to-simulate pattern of the two-element parasitic array for (impractical) 0.04λ spacing is shown in Fig. 6-18b. The gain of the array in Fig. 6-18a is exceedingly low because of the considerable amount of destructive inference between the fields radiated by the two elements. The array is very superdirective.

A feature of the Yagi is that different lengths of parasitic elements behave in substantially different ways. If the parasitic element is longer than resonant length, L_R, it acts like a reflecting element. If the parasitic element is shorter than resonant length, L_D, it can act as a directing element, or director, because it directs radiation along the z-axis in Fig. 6-19. These effects can be simulated for the close spacing used in Fig. 6-18 but are better done with more realistic reflector spacing, S_R, and director spacing, S_D.

Fig. 6-20 shows the E-plane pattern for a two-element array consisting of only the reflector element and driver in Fig. 6-19 with spacings of 0.2λ. The driver length, L, is 0.47λ, and L_R is 0.482λ. The radiation pattern shows the reflecting action of the longer parasitic element. Fig. 6-21 shows the E-plane pattern for a two-element array consisting of only one director element and driver, $L = 0.47\lambda$, in Fig. 6-19. The director length, L_D, is 0.42λ. The radiation pattern shows the directing action of the shorter element. The patterns in Figs. 6-20 and 6-21 suggest that further enhancement in radiation in the z-direction might be obtained by using both a reflector element and a director element.

(a) Array configuration.

(b) H-plane pattern computed from simple array theory.
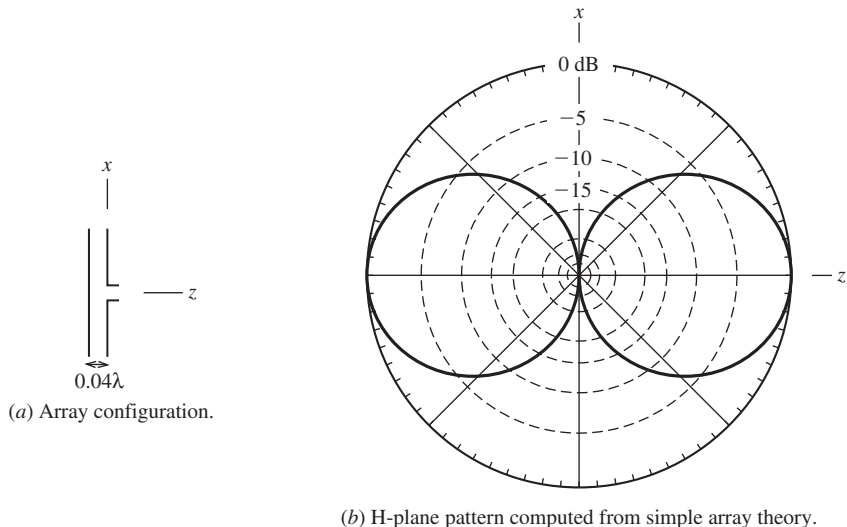
Figure 6-18 Two-element array of half-wave resonant dipoles, one a driver and the other a parasite. The currents are idealized according to (6-41).
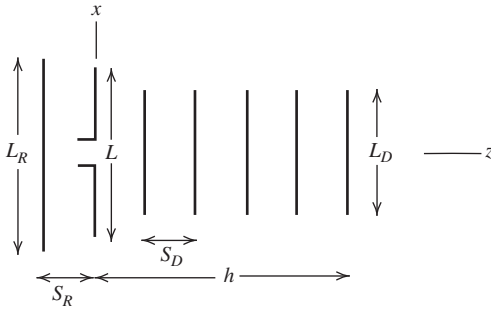
Figure 6-19 Configuration for a general Yagi-Uda antenna.

(*a*) Array configuration.

(*b*) *E*-plane pattern computed from a moment method code.
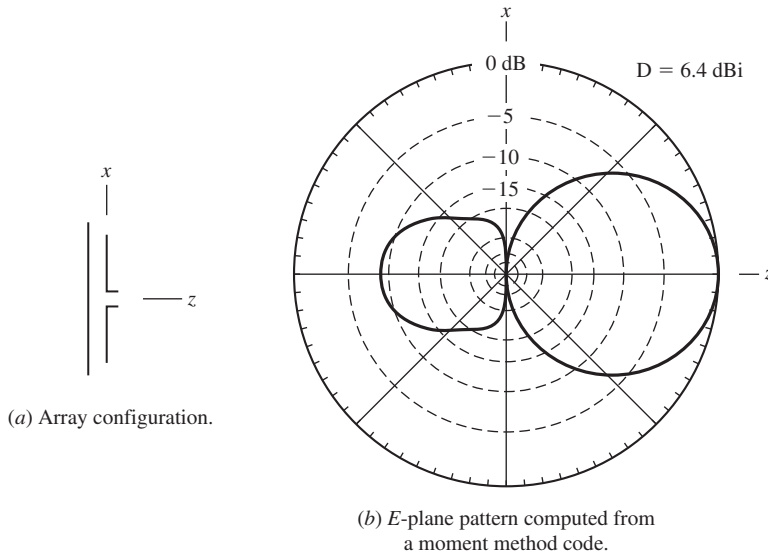
Figure 6-20 Two-element Yagi-Uda antenna consisting of a driver of length $L = 0.47\lambda$ and a reflector of length 0.482λ, spaced 0.2λ apart. The wire radius for both is 0.00425λ.

This is indeed the case. In Fig. 6-22 is the pattern of a three-element array using the parameters of Figs. 6-20 and 6-21, except that the director used in Fig. 6-22 is 0.442λ instead of 0.420λ. The 0.442λ length is near optimum for the three-element array but does not work very well in the two-element array. We will see that yet further enhancement in radiation in the z-direction can be obtained by using more than one director as in Fig. 6-19.

The maximum directivity obtainable from a three-element Yagi is about 9 dBi or 7 dBd [8]. Optimum reflector spacing S_R (for maximum directivity) is between 0.15 and 0.25 wavelengths as Fig. 6-23 shows. Note that the gain above an isolated dipole is more than 2.5 dBd, whereas if a flat plate were used, instead of a simple wirelike element, the gain would be 3 dBd. Thus, a single wire-like reflector element is almost as effective as a flat plate in enhancing the gain of a dipole.

Director-to-director spacings are typically 0.2 to 0.35 wavelength with the larger spacings being more common for long arrays and closer spacings for shorter arrays. Typically, the reflector length is 0.5λ and the driver is of resonant length when no parasitic elements are present [9]. The director lengths are typically 10–20% shorter than their resonant length, the exact length being rather sensitive to the number of directors N_D and the interdirector spacing S_D.

(a) Array configuration.

(b) E-plane pattern computed from
a moment method code.
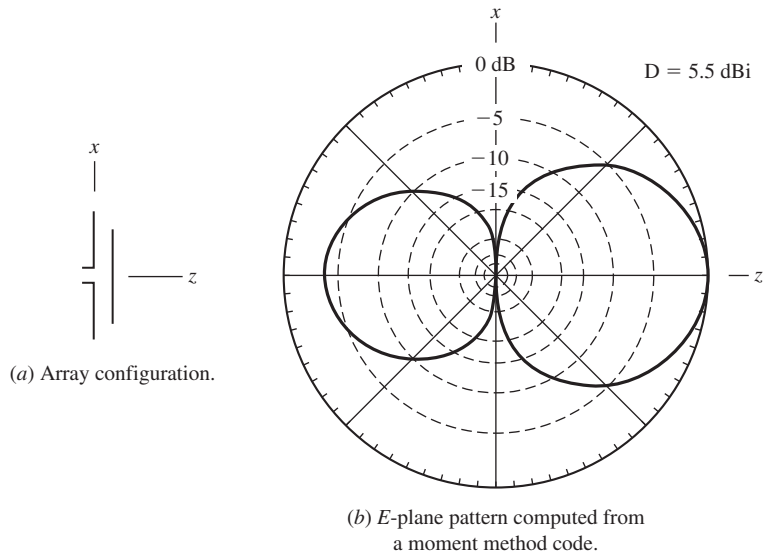
Figure 6-21 Two-element Yagi-Uda antenna consisting of a driver of length $L = 0.47\lambda$ and a director of length 0.42λ spaced 0.2λ apart. The wire radius for both is 0.00425λ.

(a) Array configuration.

(b) E-plane pattern computed from
a moment method code.
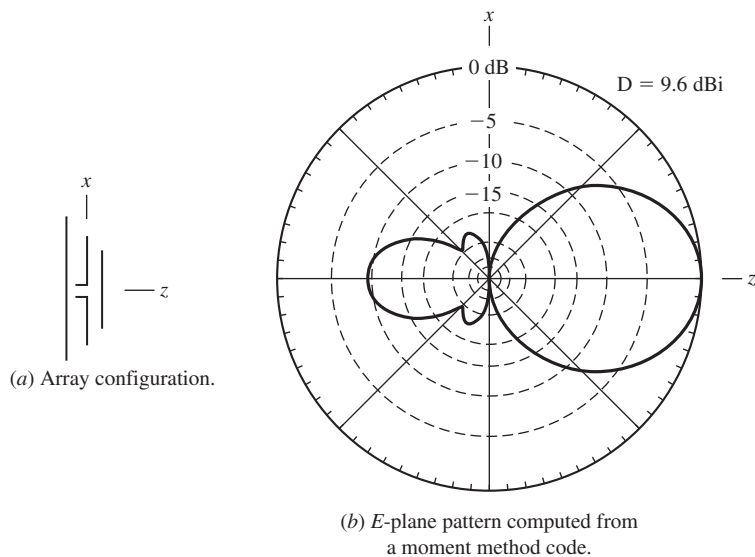
Figure 6-22 Three-element Yagi-Uda antenna consisting of a driver of length $L = 0.47\lambda$, a reflector of length 0.482λ, and a director of length 0.442λ, spaced 0.2λ apart. The wire radius for all is 0.00425λ. (See Table 6-3.)

The gain of the Yagi is related to its boom length as our study of uniform line sources in the previous chapter suggests, but for a parasitic array such as the Yagi, there is a smaller increase in gain per element as directors are added to the array (if we assume S_D is fixed) since the Yagi is not uniformly excited (see Fig. 6-27). In fact, the addition of directors up to about five or six provides a significant increase in gain expressed in dB, whereas the addition of more directors is beyond the "point of diminishing returns" as
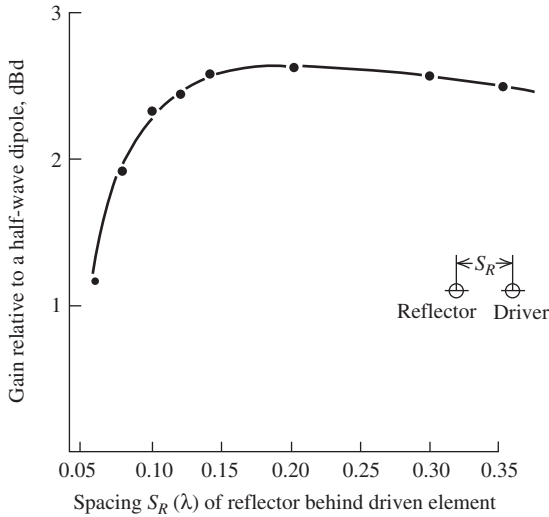
Figure 6-23 Measured gain [11] in dBd of a dipole and reflector element for different spacings S_R.
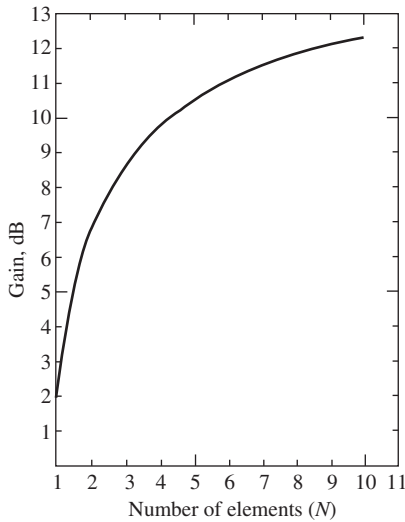
Figure 6-24 Gain of a typical Yagi-Uda antenna versus the total number of elements. The element spacings $S_R = S_D = 0.15\lambda$. The conductor diameters are 0.0025λ. (Used with permission from Green [10] © 1966 The Institute of Engineers, Australia.)

Fig. 6-24 shows. Fig. 6-24 plots the gain versus the number of elements N in the array (including one reflector and one driver) [10] for an interelement spacing for all elements of $S_R = S_D = 0.15\lambda$. Note that adding one director to increase N from 3 to 4 gives about a 1-dB gain increase, whereas adding one director to increase N from 9 to 10 yields only about an additional 0.2-dB gain.

The addition of more reflector elements results in a fractional dB increase in gain and is usually not done. The main effects of the reflector are on the driving point impedance at the feed point and on the back lobe of the array. Pattern shape, and therefore gain, are mostly controlled by the director elements. The director spacing and director length are interrelated, but the more sensitive parameter is the director length, which becomes more critical as the boom length increases.

An extensive decade-long experimental investigation by Viezbicke [11] at the National Bureau of Standards (now NIST) has produced a wealth of information on Yagi-Uda

Table 6-3 Optimized Lengths of Parasitic Dipoles for Yagi-Uda Array Antennas of Six Different Boom Lengths, L_B

$d/\lambda = 0.0085$ $S_R = 0.2\lambda$	Boom Length of Yagi-Uda Array, λ					
	0.4	**0.8**	**1.20**	**2.2**	**3.2**	**4.2**
Length of reflector, L_R/λ	0.482	0.482	0.482	0.482	0.482	0.475
D_1	0.442	0.428	0.428	0.432	0.428	0.424
D_2		0.424	0.420	0.415	0.420	0.424
D_3		0.428	0.420	0.407	0.407	0.420
D_4			0.428	0.398	0.398	0.407
D_5				0.390	0.394	0.403
D_6				0.390	0.390	0.398
D_7				0.390	0.386	0.394
D_8				0.390	0.386	0.390
D_9				0.398	0.386	0.390
D_{10}				0.407	0.386	0.390
D_{11}					0.386	0.390
D_{12}					0.386	0.390
D_{13}					0.386	0.390
D_{14}					0.386	
D_{15}					0.386	
Spacing between directors (S_D/λ)	0.20	0.20	0.25	0.20	0.20	0.308
Gain relative to half-wave dipole, dBd	7.1	9.2	10.2	12.25	13.4	14.2
Design curve (Fig. 6-25)	(A)	(C)	(C)	(B)	(C)	(D)
Front-to-back ratio, dB	8	15	19	23	22	20

Note: The left-margin label reading "Length of director, L_{D_n}/λ" applies to the director rows D_1 through D_{15}.

Source: P.P. Viezbicke, "Yagi Antenna Design," NBS Tech. Note 688, National Bureau of Standards, Washington, DC, Dec. 1968.

antenna design. An objective of the experimental investigation, conducted several wavelengths above the ground, was to determine optimum designs for a specified boom length, $L_B = S_R + h$. Boom lengths from 0.2 to 4.2λ were included in the study. Some of Viezbicke's work is summarized in Table 6-3, which can be used for design purposes. Viezbicke's work and its summaries in [12] and [13] show how to correct the free-space parasitic element lengths for both the diameter of the conductors used (see Fig. 6-25) and the diameter of a metal boom (see Fig. 6-26), if a metal boom is used. A metal boom can be used because the voltage distribution on the parasitic elements goes through a zero at the element center. Ideally, an infinitely thin metallic boom down the center of the array would not change the voltage distribution. However, metallic booms of practical size do have an effect that must be compensated for by increasing the parasitic element lengths. Alternatively, the parasitic elements can be insulated from the boom, in which case no compensation is required.

The Yagi-Uda antenna with at least several directors can be viewed as an endfire traveling wave antenna the supports a surface wave of the slow wave type (i.e., $c/v > 1$). That is, the driver-reflector pair launches a wave onto the directors that slows the wave down to a velocity, v, where $v < c$. Or, in terms of phase delay, the phase delay per unit distance along the axis of the array in the forward direction is greater than that of the ordinary endfire condition. One might expect that the additional phase delay beyond ordinary endfire required for maximum gain is that of the Hansen-Woodyard condition for a uniform line source (see Sec. 5.4). If the boom length is quite long, this is approximately true.
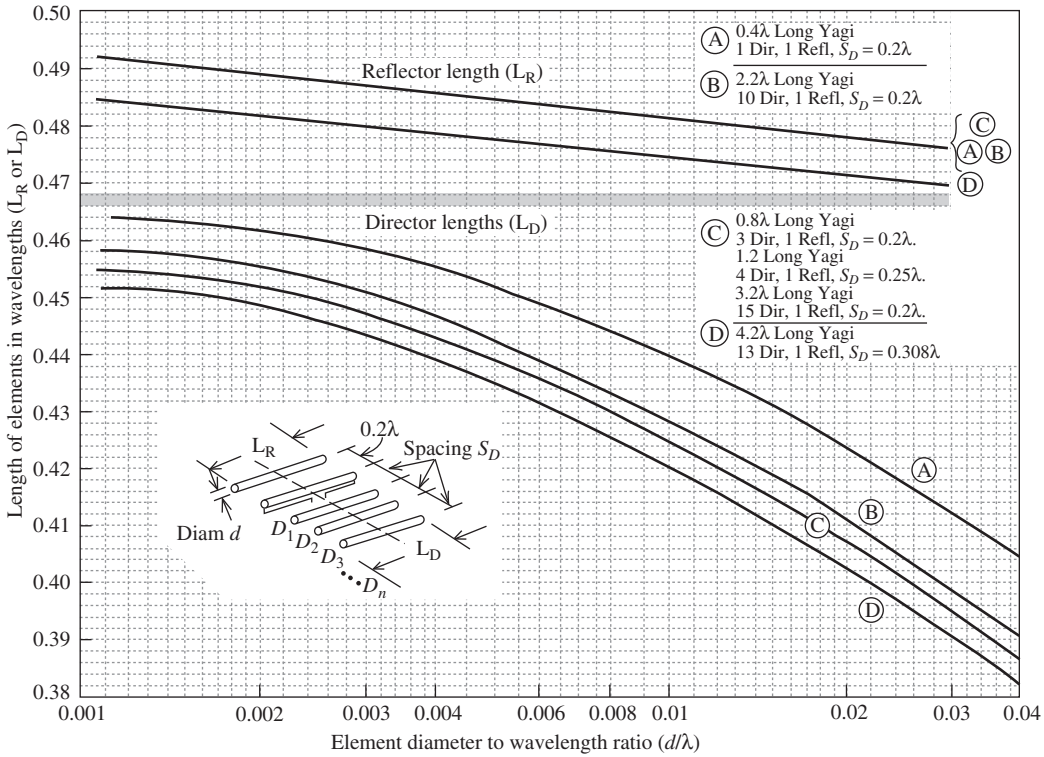
Figure 6-25 Design curves for Yagis in Table 6-3. (From [11]. P. Viezbicke, "Yagi Antenna Design," NBS Technical Note 688, U.S. Government Printing Office, Washington, DC, Dec. 1976.)
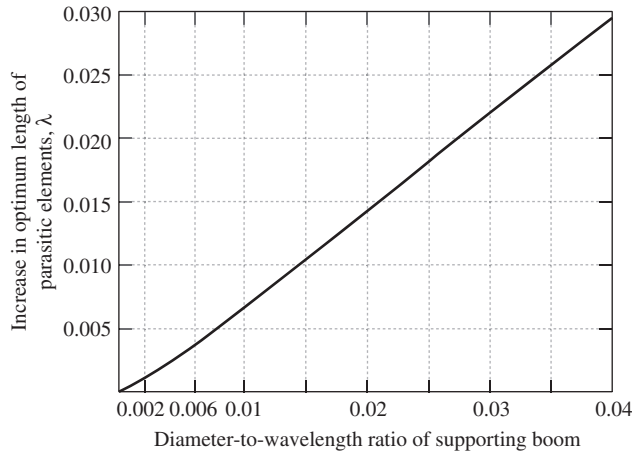
Figure 6-26 Graph showing effect of supporting metal boom on the length of Yagi parasitic elements. (From [11]. P. Viezbicke, "Yagi Antenna Design," NBS Technical Note 688, U.S. Government Printing Office, Washington, DC, Dec. 1976.)

The director currents on a well-designed Yagi are nearly equal as Fig. 6-27 indicates [6]. If the partial boom length h, as measured from the driver to the furthest director (see Fig. 6-22), is long $(h \gg \lambda)$, the Hansen-Woodyard condition requires that the phase difference between the surface wave and a freespace wave at the director furthest from the driver (the terminal director) be approximately 180°. Thus,
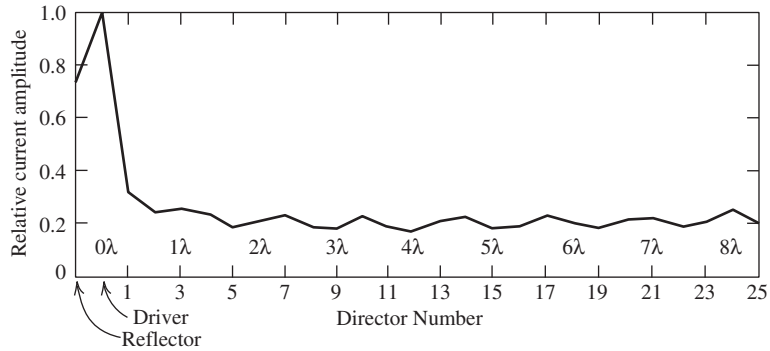
Figure 6-27 Relative current amplitudes for a 27-element Yagi array (Used with permission from Thiele [6] © 1969 IEEE.)
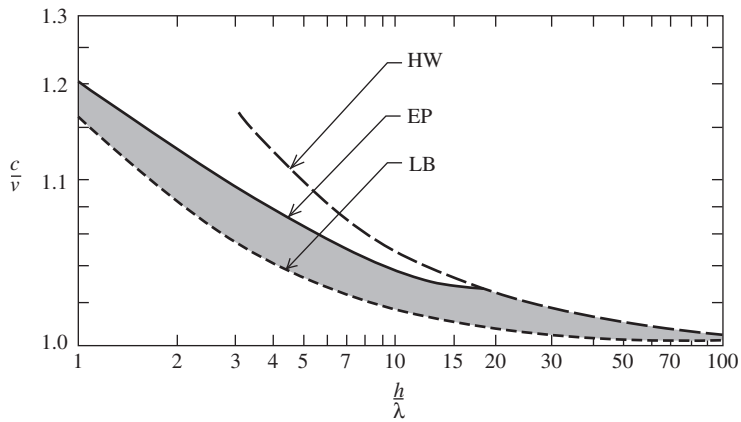
Figure 6-28 Relative phase velocity $c/v = \lambda/\lambda_g$ for maximum gain surface wave antennas as a function of antenna length h/λ. HW = Hansen–Woodyard condition (5-46); EP = Ehrenspeck and Poehler experimental values; LB = lower bound (for idealized surface wave excitation), (Used with permission from [H6: Johnson] © 1993 The McGraw-Hill Companies.)

$$h\beta_g - h\beta = \pi \tag{6-43}$$

or

$$\lambda/\lambda_g = c/v = 1 + \lambda/2h \tag{6-44}$$

where β_g is a guided phase constant, or wavenumber, along the forward axis of the array, λ_g represents the corresponding guided wavelength, and λ is the unguided (free-space) wavelength. Note that $\beta_g = -\beta \cos \theta_o \approx -\beta(c/v)$ and recall that $c/v > 1$ implies $|\cos \theta_o| > 1$. Eq. (6-44) is plotted as the upper dashed line in Fig. 6-28.

Experimental work by Ehrenspeck and Poehler [9] showed that the optimum terminal phase difference is about 60° for short Yagis, rising to about 120° for $4\lambda < h < 8\lambda$, and then approaching 180° for $h > 20\lambda$ This is the solid curve in Fig. 6-28. Various data for Yagis and other endfire structures indicate that the optimum c/v values lie on or just below the solid curve in the shaded area. Other surface wave structures, with more efficient surface wave excitation than the Yagi, can have optimum c/v values that lie in

the shaded region nearer to the dashed lower bound, LB, curve but all of these surface wave structures approach the Hansen-Woodyard condition if they are very long.

Viewing the directors as a reactive surface over which the surface wave travels, and noting from Table 6-3 that director lengths tend to be shorter for longer boom lengths, lead us to surmise that the surface wave couples less to the reactive surface for long boom lengths so that the total phase delay is not excessive and falls on the solid curve in Fig. 6-28. Examining Table 6-3 for those arrays with a director spacing $S_D = 0.2\lambda$ shows that indeed the directors tend to be shorter for longer boom lengths.

The Yagi is one of the more popular antennas used in the HF–VHF–UHF frequency range. It provides moderately high gain while offering low weight and low cost. It has a relatively narrow bandwidth (e.g., a few percent), which may be improved somewhat by using feeds other than a dipole, such as a folded dipole. The folded dipole also provides a higher input impedance than a dipole even though the driving point impedance of both are usually reduced considerably from their self-impedances by mutual coupling effects. Further, increased gain can be obtained by arraying or "stacking" Yagi antennas. Maximum gain results for a separation of almost one wavelength (see Fig. 8-16). Thus, for a given application, if a somewhat narrow bandwidth can be tolerated, the Yagi-Uda antenna can provide good gain (e. g., 9–12 dB) at low cost.

EXAMPLE 6-1   *TV Channel 12 Yagi Antenna Design*

A 12-element Yagi for TV channel 12 at 205.25 MHz is to be designed using 1-cm-diameter elements insulated from a metallic boom [21]. The boom length is to be 2.2λ. Table 6-3 indicates that 0.2λ spacing is required. The wavelength at 205.25 MHz is 1.46 m. Thus, the spacing between all elements is 29.2 cm. To obtain the element lengths, the following four steps are followed:

1. Plot the element lengths from Table 6-3 on the design curves "B" in Fig. 6-25. The design curves are for conductor diameters of 0.0085λ.
2. Since the 1-cm conductor to be used is 0.0065λ in diameter, the element lengths in the table must be increased slightly. This is accomplished by drawing a vertical line at 0.0065λ on the horizontal axis. This line intersects the two applicable design curves, which gives the compensated lengths of the reflector and first director:

$$L_R = 0.483\,\lambda, \qquad L_{D_1} = 0.4375\lambda$$

Notice the distance along the director "B" curve between the intersection of the vertical line at 0.0065λ and the location of the first director length from step 1 above. All the directors must be increased in length by an amount that is determined by this distance.
3. With a pair of dividers, measure the distance on the B director curve between the initial length and corrected length of the first director. Slide each of the other director lengths to the left by this amount to determine their compensated lengths:

$$L_{D_2} = 0.421\lambda$$
$$L_{D_3} = L_{D_{10}} = 0.414\lambda$$
$$L_{D_4} = L_{D_9} = 0.405\lambda$$
$$L_{D_5} = L_{D_6} = L_{D_7} = L_{D_8} = 0.398\lambda$$

4. The fourth step generally is to correct the lengths for the metallic boom using Fig. 6-26, if one is used. In this case, the boom is metallic, but the elements are insulated from it and no correction factor is required. (See Prob. 6.4-5.)

(a) E-plane.

(b) H-plane.
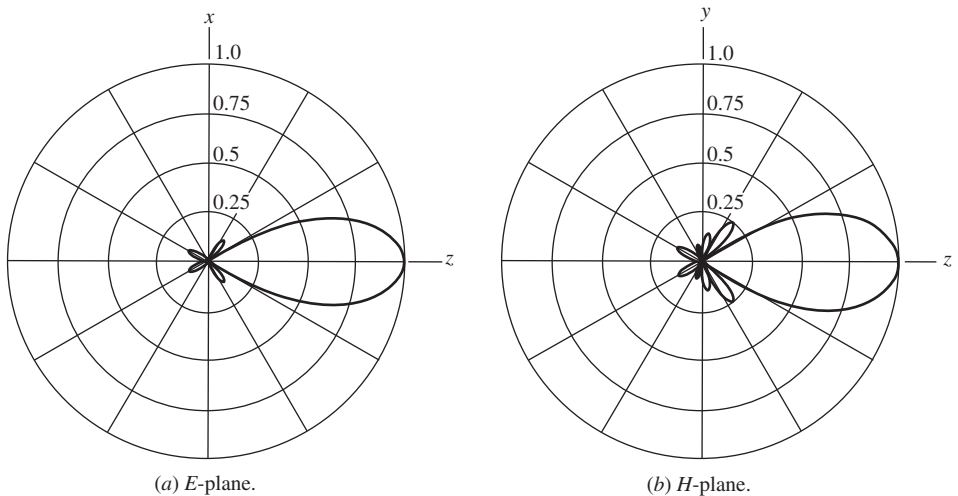
Figure 6-29 Simulated patterns for the 12-element Yagi of Example 6-1.

Calculated patterns for Example 6-1 using numerical methods (e.g., Chap. 14) are shown in Fig. 6-29 and the calculated directivity is 11.82 dBd, which agrees well with the gain value of 12.25 dBd in Table 6-3. The calculated input impedance for a dipole driver is $26.5 + j23.7\,\Omega$. The calculated front-to-back ratio is 38.5 dB, owing to the almost total absence of a back lobe in the calculated pattern.

6.4 FEEDING WIRE ANTENNAS

There are several questions that arise when connecting an antenna to a transmitter or receiver. What sort of transmission line is best to use? What is the best way to "match" impedances between the antenna, the transmission line and the transmitter or receiver? Is it necessary to transform an unbalanced transmission line to a balanced antenna? This section will provide answers to these questions.

6.4.1 Transmission Lines

Consider an antenna operating in the transmission mode as in Fig. 6-30. Both antenna systems in Fig. 6-30 employ a transmission line of characteristic impedance Z_o between generator and antenna. Wire antennas are most commonly fed with either coaxial cable or some form of parallel wire transmission line as shown in Fig. 6-31.

The most common value for Z_o of a circular coaxial line is 50 Ohms (real). This is a compromise among several factors. Z_o for maximum power handling in a coaxial line is about 30 Ohms, for maximum voltage about 60 Ohms, and for minimum conductor losses Z_o is about 77 Ohms [14]. The characteristic impedance of a coaxial cable is primarily determined by the ratio of the inner and outer diameters according to

$$Z_o = \frac{60}{\sqrt{\varepsilon_r}} \ln\left(\frac{b}{a}\right) \qquad (6\text{-}45)$$

Parallel wire transmission lines tend to have higher characteristic impedances with 75 and 300 ohms being common for "twin-lead" and 450 and 600 Ohms being common for open (i.e., mostly air between conductors, $\varepsilon_r = 1$) wire transmission line. The spacing between conductors is the primarily practical determinant for the value of Z_o as seen in
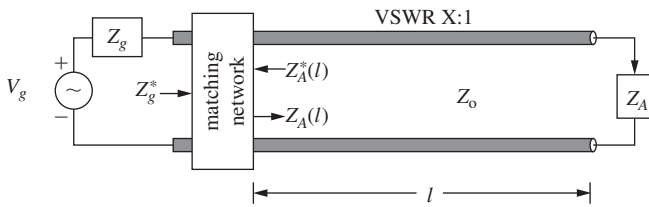
(a) Conjugate matching.
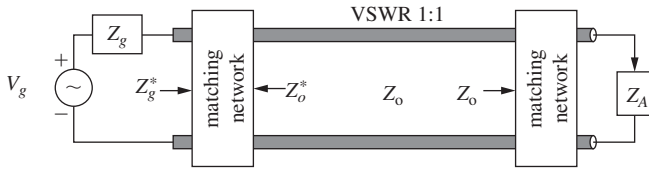
(b) Reflectionless matching.

Figure 6-30 Conjugate and reflectionless matching of a transmission line.
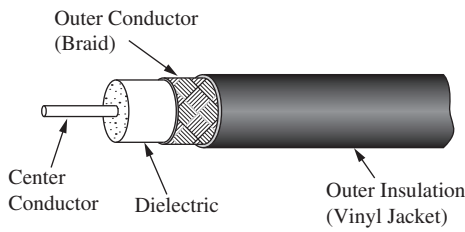
(a) Coaxial cable.
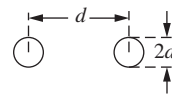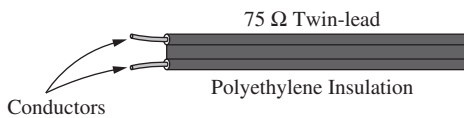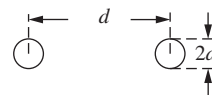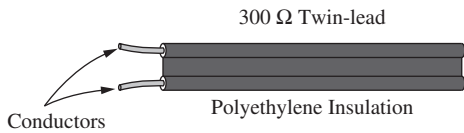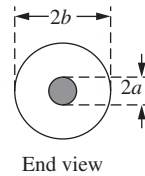
(b) Two examples of twin-lead.
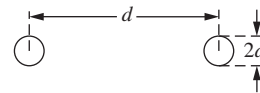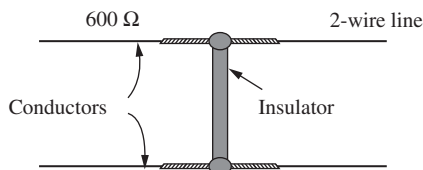
(c) Open two-wire line.

Figure 6-31 Common transmission lines for feeding wire antennas.

$$Z_o = \frac{120}{\sqrt{\varepsilon_r}} \cosh^{-1}\left(\frac{d}{2a}\right) = \frac{120}{\sqrt{\varepsilon_r}} \ln\left(\frac{d + \sqrt{d^2 - (2a)^2}}{2a}\right) \qquad (6\text{-}46)$$

The advantages of the open-wire line are high-power handling capability and low loss. At VHF and above, parallel wire lines are seldom used because the spacing between conductors becomes electrically large enough to disturb the transmission line behavior of the line. That is, radiation by the transmission line may occur. Parallel wire lines are balanced transmission lines, whereas coaxial lines are unbalanced lines, as will be explained in Sec. 6.4.3.

6.4.2   Matching Networks

Consider an antenna in the transmission mode as in the equivalent circuit of Fig. 2-15. From circuit theory we know that maximum transfer of power from the generator to the real part of the load provided by the antenna, $R_A = R_o + R_r$, will occur when the load impedance, Z_A, on the right is the conjugate match of the source impedance, Z_g, on the left. In other words, $Z_A = Z_g$ for maximum power transfer. For resonant size antennas, $R_r \gg R_o$ and thus $R_A \approx R_r$.

Fig. 6-30 is an extension of Fig. 2-15. Fig. 2-15 is a useful equivalent circuit of an antenna system that does not show the transmission line connecting the antenna to a source (transmitter or receiver). When the separation between the antenna and the source is greater than only about 0.1λ, the connection between the source and antenna must be treated by transmission line theory. Separations of many wavelengths are common.

Maximum power transfer is not the only consideration when feeding an antenna. Two other considerations are (1) achieving a minimum amount of power reflected from the load and (2) achieving a desired current distribution. In general, maximum power transfer does not guarantee zero reflection from the load.

Minimum reflection at the load can be particularly important when a high-power transmitter is being used. Reflected power results in a standing wave of voltage, VSWR, and of current, ISWR, on the transmission line. The VSWR produces a voltage, V_{max}, on the line which is greater than the transmitter output voltage, V_o. This may result in voltage breakdown on the line in the form of arcing. Further, the reflected power can cause damage to the transmitter's output final amplifier and/or cause "frequency pulling" which is a change from the desired transmission frequency. In cases where high power is not involved, a low VSWR may be a luxury and not a necessity. This is demonstrated in Table 6-4. For example, a VSWR of 2:1 leads to 89% power transmission. An 11% reduction in radiated or received power would not be noticeable on most commercial and amateur communication links.

So how do we achieve a system with a low VSWR on the transmission line, and/or maximum power transfer? Consider Fig. 6-30a, which shows a single matching network that provides for a conjugate match at a single frequency. The matching network may be placed anywhere on the line. In 6-30b, the matching network on the right provides for a reflectionless match between the antenna impedance, Z_A, and the characteristic impedance, Z_o, of the line. The matching network on the left provides for a conjugate match between the source and the transmission line. If the transmission line has very low loss and the characteristic impedance of the line is essentially pure real, the system in Fig. 6-30 can provide both a conjugate match for maximum power transfer and zero reflection from the load. This is the ideal situation. The system in Fig. 6-30a is potentially broader band than that in Fig. 6-30b.

Matching networks have disadvantages. For example, if a matching network is designed to obtain a near-perfect match, it will usually be narrow-band. If the matching network is designed to be broadband, it will not usually yield a near-perfect match at all frequencies over the band, and sometimes not at any frequencies over the band. Matching

Table 6-4 VSWR and Transmitted Power for a Mismatched Antenna

VSWR	Percent Reflected Power $=\|\Gamma\|^2 \times 100$ $=\left(\dfrac{VSWR-1}{VSWR+1}\right)^2 \times 100$	Percent Transmitted Power $=q \times 100$ $=(1-\|\Gamma\|^2) \times 100$	Maximum Voltage on Transmission line $=(1+\|\Gamma\|)V_o$	Return Loss $=-20 \log (\Gamma)$
1.0	0.0	100.0	1.00 V_o	∞ dB
1.1	0.2	99.8	1.09 V_o	26.4 dB
1.2	0.8	99.2	1.17 V_o	20.9 dB
1.5	4.0	96.0	1.20 V_o	14.0 dB
2.0	11.1	88.9	1.33 V_o	9.54 dB
3.0	25.0	75.0	1.50 V_o	6.02 dB
4.0	36.0	64.0	1.60 V_o	4.47 dB
5.0	44.4	55.6	1.67 V_o	3.52 dB
5.83	50.0	50.0	1.71 V_o	3.01 dB
10.0	66.9	33.1	1.82 V_o	1.74 dB

networks add cost, weight, and power loss to a system and can decrease the bandwidth of the antenna system. Fortunately, the advantages of using matching networks usually outweigh the disadvantages.

The radiation quality factor of an antenna, Q_A, is the approximate inverse of the antenna bandwidth provided $Q_A > 10$. That is, the bandwidth as a percent, B_p, is

$$B_p \approx \frac{1}{Q_A} \times 100\% \qquad \text{provided } Q_A > 10 \tag{6-47}$$

For the system in Fig. 6-41a, the system Q, Q_s, comprised of the antenna radiation Q, Q_A, and the matching network Q, Q_m, is

$$\frac{1}{Q_s} = \frac{1}{Q_A} + \frac{1}{Q_m} \tag{6-48}$$

For example, if Q_A and $Q_m = 10$ (each individually have a bandwidth of 10%), then the bandwidth of the system, Q_s, is 5%. The desirability for using wideband matching networks in many applications is readily apparent.

Since an antenna of efficiency e_r and a matching network of efficiency e_m are in series, the efficiency of the system, e_s, is the product of the two efficiencies. For example, if the antenna efficiency is 0.98 and the matching network efficiency is 0.92, the system efficiency is 0.89 and the power loss due to the two efficiencies is about 0.5 dB. The system Q can be decreased (and bandwidth increased) by introducing loss into the matching network. But the penalty in doing this is a decrease in system efficiency.

Low-loss broadband matching networks are a challenging design problem and are beyond the scope of this text. The most general broadband matching problem is the matching of an arbitrary impedance to another arbitrary impedance over a specified frequency range within a specified deviation from perfect match. However, the antenna designer is unlikely to be faced with such an arbitrary problem, because at least one of the impedances involved, Z_o, will be nearly resistive and not very frequency, sensitive. For serious discussion of broadband matching, the reader is referred to [H.8.2: Munk, pp. 288-305; H.6: *Ant. Eng. Hdbk.*, 4th ed., Chap. 52].
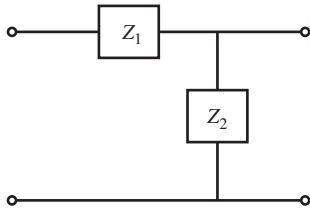
(*a*) L-section network.
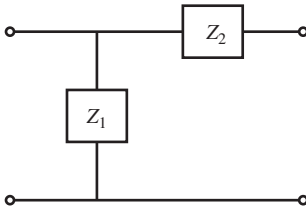
(*b*) Inverted L-section network.
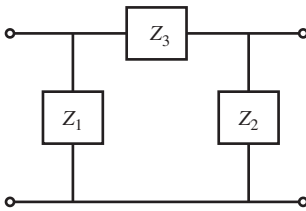
(*c*) Pi-network.

Figure 6-32 Some lumped element matching networks.

Matching networks employ different hardware at high and low frequencies. At UHF and low microwave frequencies, distributed tuning devices such as stubs are introduced to transform the real part of an impedance to that of the transmission line as well as tuning out the reactive component. At low frequencies reactive tuning is accomplished with variable capacitors and coils because the electrical dimensions of these lumped elements are small with respect to the wavelength. One way impedance levels can be changed is with air-filled coils functioning as a transformer.

Examples of using lumped elements at low frequencies are the L-network and the Pi-network shown in Fig. 6-32. The L-network uses only reactive elements to provide a conjugate match between two complex impedances, but the necessary component values may be difficult to realize in some cases. The Pi-network provides a more flexible arrangement, in part because there are three variables instead of two [H.6: *Ant. Eng. Hdbk.*, 4th ed., Chap. 52]. The Pi-network can be used to match a low impedance to a rather high one, say 50 ohms to many hundred ohms, or 50 ohms to a low impedance of just a few Ohms. The Pi-network also provides somewhat more control over the (narrow) bandwidth.

Some examples of matching networks used at higher frequencies include stubs, quarter-wave transformers, cascaded quarter-wave transformers, Chebyshev transformers, binomial transformers, combinations of transformers and stubs, and tapered lines. A discussion of these techniques can be found in [H.6: *Ant. Eng. Hdbk.*, 4th ed., Chap. 52].

There are also ways to change the input impedance of an antenna without using a matching network. For example, the input impedance of a dipole can be changed by displacing the feed point off-center. If the feed point is a distance z_f from the center of the dipole, the current at the input terminals is
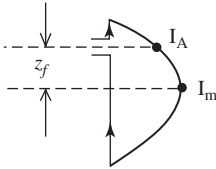
Figure 6-33 Half-wave dipole with displaced feed.
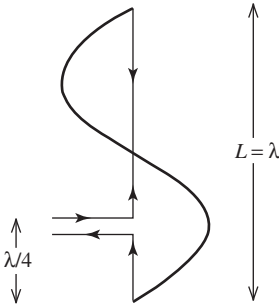
Figure 6-34 Current distribution on a full wave dipole for an off-center feed.

$$I_A = I_m \ \sin\left[\beta\left(\frac{L}{2} - |z_f|\right)\right] \qquad (6\text{-}49)$$

In the case of a half-wave dipole as shown in Fig. 6-33, $\beta L/2 = \pi/2$, and this reduces to

$$I_A = I_m \ \cos \beta z_f \qquad (6\text{-}50)$$

The input resistance (not including ohmic losses) is found from (6-50) in (6-20), giving

$$R_A = \frac{I_m^2}{I_A^2} R_{rm} = \frac{R_m}{\cos^2 \beta z_f} \quad \left(L = \frac{\lambda}{2}\right) \qquad (6\text{-}51)$$

As the feed point approaches the end of the wire, this result indicates that the input resistance increases toward infinity. In practice, the input resistance becomes very large as the feed point moves out. The pattern is essentially unchanged as the feed point shifts. For longer dipoles, the pattern and impedance differ significantly from the center-fed case as the feed point is displaced. For example, a full-wave dipole fed a quarter-wavelength from one end, as shown in Fig. 6-34, will have a current distribution that is significantly different from the center-fed full-wave dipole of Fig. 6-3c and that has a broadside null in the pattern.

The off-center feed arrangement is unsymmetrical and can lead to undesirable phase reversals in the antenna, as shown in Fig. 6-34. A symmetrical feed that increases the input resistance with increasing distance from the center point of the wire antenna is the *shunt feed*. A few forms of shunt matching are shown in Fig. 6-35. We will discuss the operation of the tee match; the remaining shunt matches behave in a similar fashion. The center section of the tee match may be viewed as being a shorted transmission line in parallel with a dipole of wide feed gap spacing. The shorted transmission line is less than a quarter-wavelength long and thus its impedance is inductive. Capacitance can be introduced to tune out this inductance by either shortening the dipole length or placing variable capacitors in the shunt legs. As the distance D is increased, the input impedance increases and peaks for a D of about half of the dipole length. As D is increased further, the impedance decreases and finally equals the folded dipole value when D equals the

(*a*) Delta match.   (*b*) Tee match.   (*c*) Gamma match.
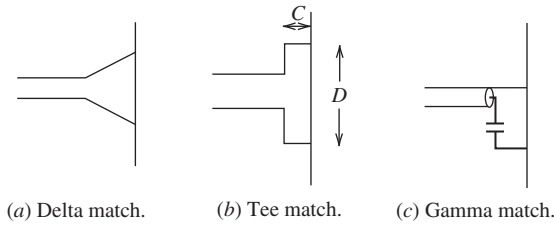
Figure 6-35 Shunt matching configurations.

dipole length. The exact impedance value depends on the distances C and D, and the ratio of the dipole wire diameter to the shunt arm wire diameter (similar to the folded dipole behavior). In practice, sliding contacts are made between the shunt arms and the dipole for impedance adjustment. Shunt matches will radiate, and do so in an undesirable fashion.

6.4.3   Baluns

We now turn our attention to a separate but related problem of balancing currents on wire antennas. Many wire antennas are symmetrical in nature, and thus, the currents should also be symmetrical (or balanced). An example of balanced and unbalanced operation of a half-wave dipole is shown in Fig. 6-36. In the balanced case, the currents on the transmission line are equal in magnitude and opposite in direction, which yields very small radiation from the transmission line for closely spaced conductors. For unbalanced operation, as illustrated in Fig. 6-36*b*, the current I_1 is greater than I_2 and there is a net current flow on the transmission line leading to uncontrolled radiation that is not in the desired direction or of the desired polarization. Also, the unbalanced current on the antenna will change the radiation pattern from the balanced case. Thus, it is clear that balanced operation is desirable.

Transmission lines are referred to as balanced or unbalanced. Parallel wire lines are inherently balanced in that if an incident wave (with balanced currents) is launched down the line, it will excite balanced currents on a symmetrical antenna. On the other hand, a coaxial transmission line is not balanced. A wave traveling down the coax may have a balanced current mode, that is, the currents on the inner conductor and the inside of the outer conductor are equal in magnitude and opposite in direction. However, when this wave reaches a symmetrical antenna, a current may flow back on the outside of the outer conductor, which unbalances the antenna and transmission line. This is illustrated in Fig. 6-37. Note that the currents on the two halves of the dipole are unbalanced. The current I_3 flowing on the outside of the coax will radiate. The currents I_1 and I_2 in the coax are shielded from the external world by the thickness of the outer conductor. They could actually be unbalanced with no resulting radiation; it is the current on the outside surface of the outer conductor that must be suppressed. To suppress this outside surface current, a *balun* (contraction for "balanced to unbalanced transformer") is used.

(*a*) Balanced currents, $I_1 = I_2$.   (*b*) Unbalanced currents, $I_1 > I_2$.
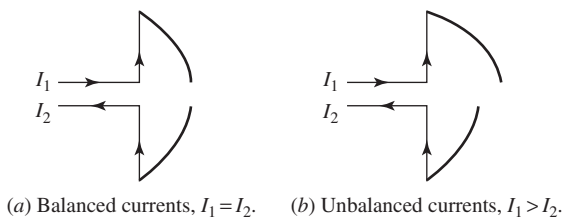
Figure 6-36 Balanced and unbalanced operation of a center-fed half-wave dipole.
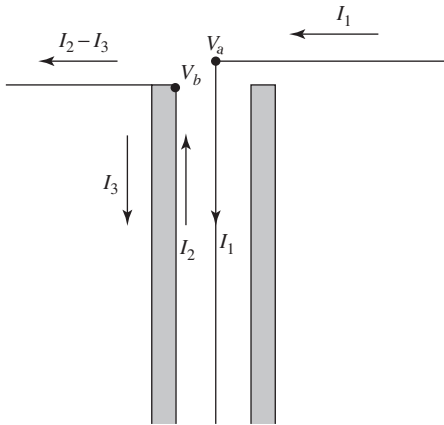
Figure 6-37 Cross section of a coaxial transmission line feeding a dipole antenna at its center.

The situation in Fig. 6-37 may be understood by examining the voltages that exist at the terminals of the antenna. These voltages are equal in magnitude but opposite in phase (i.e., $V_a = -V_b$). Both voltages act to cause a current to flow on the outside of the coaxial line. If the magnitude of the currents on the outside of the coax produced by both voltages are equal, the net current would be zero. However, since one antenna terminal is directly connected to the outer conductor, its voltage V_b produces a much stronger current than the other voltage V_a.

In other words, the current I_2, upon reaching terminal "b" in Fig. 6-37, splits with $I_2 - I_3$ going onto the dipole arm and I_3 going down the outside of the coax. What a balun does is force I_3 to zero or nearly so, such that the currents on the two arms of the dipole are the same or nearly so. Thus, a *balun transforms the balanced input impedance of the dipole to the unbalanced impedance of the coaxial line such that there is no net current on the outer conductor of the coax.*

To illustrate how a balun works, consider the *sleeve* (or *bazooka) balun* in Fig. 6-38. The sleeve and outer conductor of the coaxial line form another coaxial line of characteristic impedance Z'_o that is shorted a quarter-wavelength away from its input at the antenna terminals. The equivalent circuit for Fig. 6-37 is that of Fig. 6-39a. The equivalent circuit of Fig. 6-38 is that of Fig. 6-39b, which shows that both terminals see a very high impedance to ground. Thus, the situation in Fig. 6-39b is equivalent to the balanced condition of Fig. 6-39c wherein the currents I_1 and I_2 are equal.
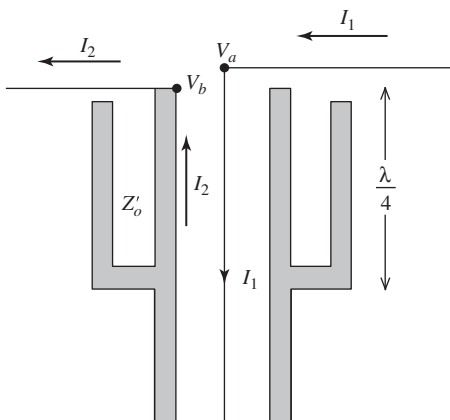
Figure 6-38 Cross section of a sleeve balun feeding a dipole at its center.

(*a*) Equivalent circuit of coax-fed dipole in Fig. 6-37.

(*b*) Equivalent circuit of sleeve balun-fed dipole in Fig. 6-38.

(*c*) Final equivalent circuit for Fig. 6-38 with quarter-wave
   transmission line removed, $I_1 = I_2$.
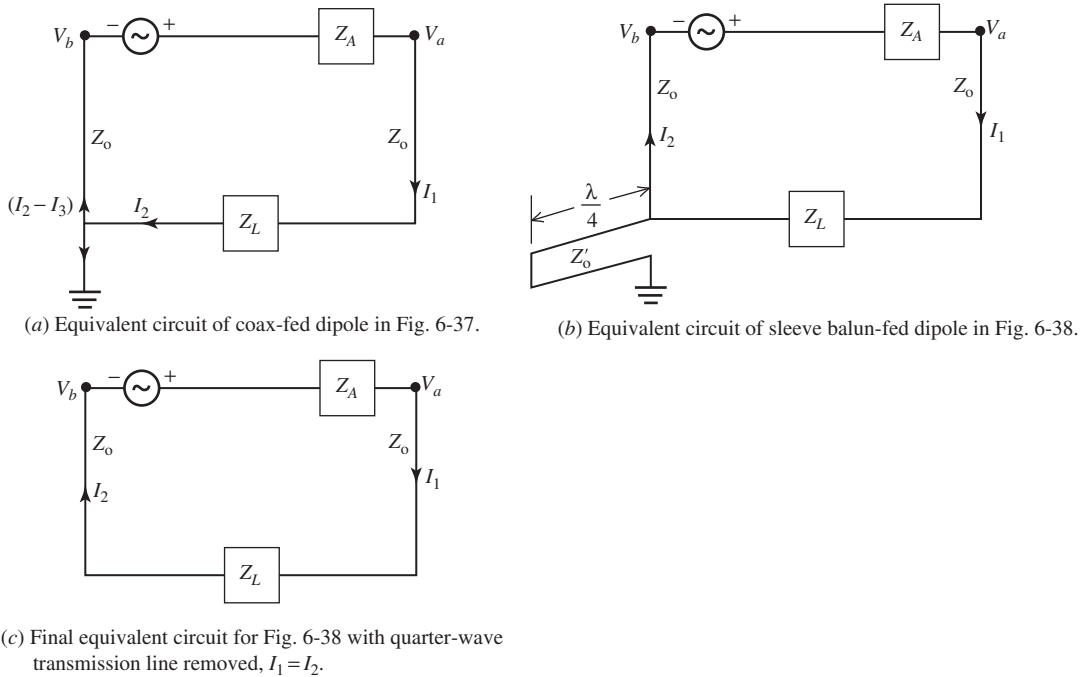
Figure 6-39 Equivalent circuits for a dipole fed from a coaxial transmission line of characteristic impedance Z_o and load impedance Z_L.

An easily constructed balun form is the *folded balun* shown in Fig. 6-40. The quarter-wavelength of coax from the *a* terminal to the outer conductor of the transmission line does not affect the antenna impedance Z_A. The extra quarter-wavelength of coax together with the outer conductor of the main transmission line forms another equivalent transmission line, which is a quarter-wavelength long and is shorted at *C*. Therefore, the short circuit at *C* is (ideally) transformed to an infinite impedance at the antenna terminals, which is in parallel with Z_A, leaving the input impedance unchanged. The quarter-wavelength line induces a cancelling current on the outside of the coaxial transmission line, so that the net current on the outside of the main coax below point *C* is zero, as shown in Fig. 6-40*a*.

The split-coax balun in Fig. 6-40*b* functions in the same way as the folded balun. The sleeve, folded and split-coax baluns are, of course, not broadband because of the quarter-wavelength involved in their construction.

Broadband baluns can be constructed by tapering a balanced transmission line to an unbalanced one over at least several wavelengths of transmission line length as indicated in Fig. 6-41. Fig. 6-41*a* shows a balanced transmission line tapering to an unbalanced microstrip line and Fig. 6-41*b* illustrates a balanced line tapering to an unbalanced coaxial line.

The baluns we have considered thus far are useful from microwave frequencies down to VHF. From VHF down to lower frequencies, it is impractical in many cases to employ these configurations and transformers are used as Fig. 6-42 indicates. Fig. 6-42*a* is an air core transformer arrangement useful at lower frequencies. Fig. 6-42*b* is bifilar wound ferrite core balun that can be used from VLF through UHF.

Impedance transformation can also be included in a balun for matching purposes. For example the "four to one" balun in Fig. 6-43 will transform an unbalanced 75-Ω impedance to one that is 300-Ω balanced. Such a balun is useful with a folded dipole.

(*a*) Folded balun.
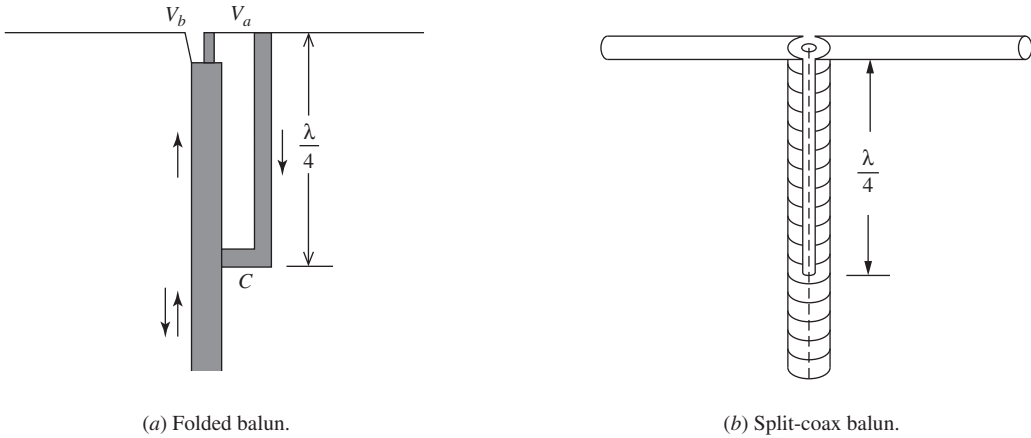
(*b*) Split-coax balun.

Figure 6-40 The folded balun and the split-coax balun equivalent.

(*a*) Tapered microstrip balun.

(*b*) Tapered coaxial balun.
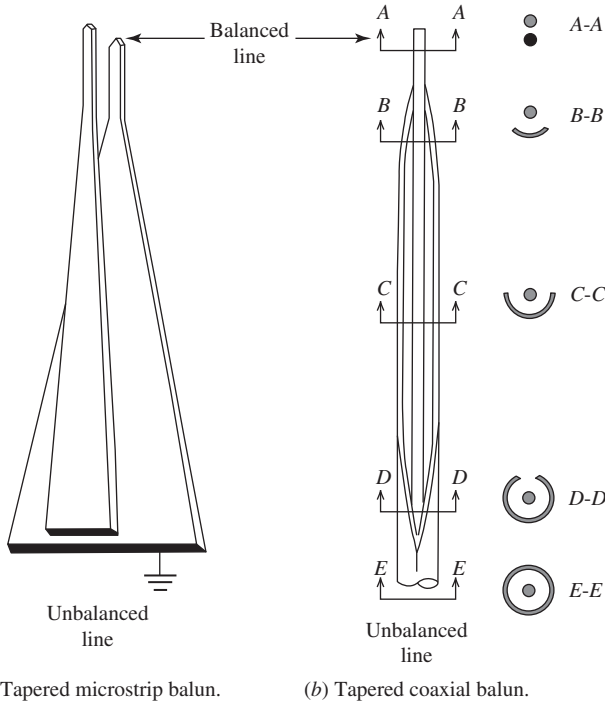
Figure 6-41 Broadband baluns.

To understand how the 4:1 balun works, consider Fig. 6-43*a*, which shows the 300-Ω load split into two 150-Ω impedances in series and still balanced. Note that the midpoint between the two terminals is at zero potential with respect to ground and therefore may be grounded as in Fig. 6-43*b*. Also in Fig. 6-43*b*, a half-wavelength of coax is used to transform the negative terminal to a positive terminal which is then connected to the upper positive terminal. This places the two 150-Ω impedances in parallel and unbalanced

(a) Air core transformer
     balun.

(b) Ferrite core bifilar
     wound wire balun.
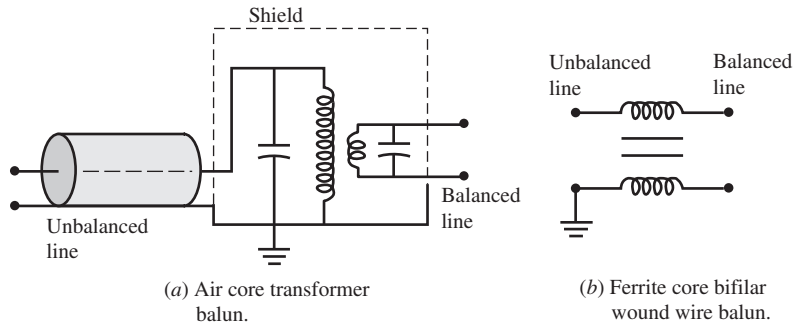
Figure 6-42 Baluns used at lower frequencies.

(a) 300 Ω load replaced with
     two 150 Ω loads in series.

(b) λ/2 line puts two 150 Ω loads in
     parallel, transforming 300 Ω
     balanced to 75 Ω unbalanced.
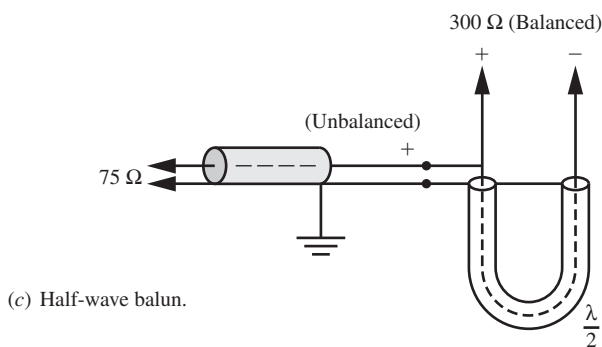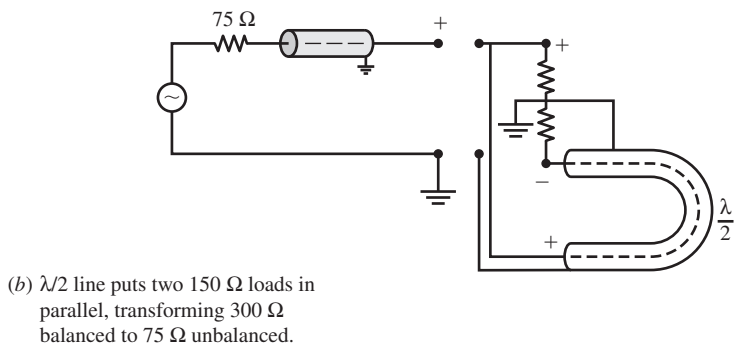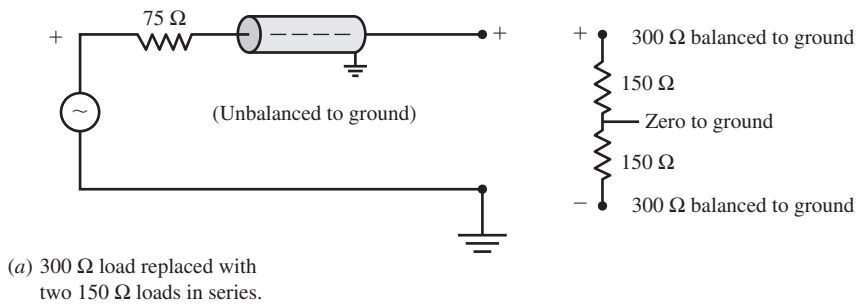
(c) Half-wave balun.

Figure 6-43 A half-wave balun that provides an impedance setup ratio of 4:1.

as seen from the coaxial line on the left. The antenna load still sees 300-Ω balanced. Thus we have a 4:1 balun as shown in Fig. 6-43c. A balun that leaves the impedance unchanged is often referred to as a "one to one" balun.

6.5 LOADED WIRE ANTENNAS

The characteristics of a wire antenna can be altered by incorporating resistance, inductance, and/or capacitance into the wire. This is called "loading" and can be accomplished either distributively or discretely with lumped R, L, and/or C elements.

6.5.1 Lumped Loaded Wire Antennas

Lumped loading is usually employed for one of several reasons: (1) to obtain multi-band operation, (2) to alter the current distribution for purposes of pattern control (e.g., phase shifter), (3) to physically shorten the antenna.

Multi-band operation can be achieved by means of a "tuned trap" which is a resonant parallel LC circuit. Fig. 6-44 achieves two-band operation at high and low frequencies f_H and f_L, respectively through the use of a trap resonant at f_H. When operation takes place at f_H, the trap has a very high impedance effectively isolating the upper section of the antenna from the lower section, thereby producing a $\lambda/4$ monopole at the higher frequency. With the trap resonant at f_H, the trap behaves as an inductor when operation takes place at f_L, physically lengthening the antenna at the lower frequency. The amount of physical lengthening at f_L depends on the L/C ratio of the trap elements [H.8.1: ARRL Antenna Book]. Without the trap, the monopole is about one-half wavelength at the higher frequency, f_H, and is very difficult to impedance match because of the current minimum at the antenna input (see Prob. 6.1-4).

In many instances it is desirable to obtain more gain from a (vertical) monopole than that provided by a quarter-wavelength monopole. To obtain more gain, a $3\lambda/4$ or $7\lambda/8$ monopole may be used as Fig. 6-45 illustrates. In practice, the $\lambda/4$ stub is usually replaced by an inductor. Both the stub and its inductor substitute provide the necessary phase shift to place the current on the upper and lower portions of the monopole in phase as indicated. Alternatively, the inductor may be roughly thought of as the wire in the stub coiled into a coil.

The purpose of the series inductance at the base of the $3\lambda/4$ monopole is to improve the VSWR. Fig. 6-46a shows the input impedance with the base inductor in place. Note
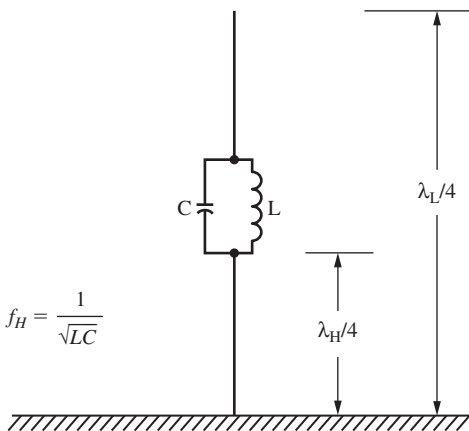
Figure 6-44 A trap monopole antenna.
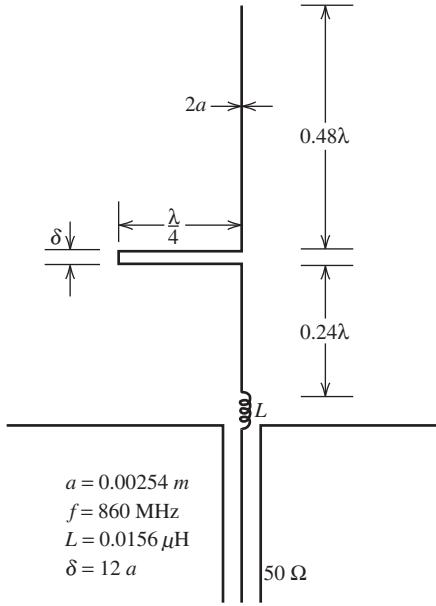
Figure 6-45 Three-quarter wavelength loaded monopole with series loading.

$2a$

0.48λ

$\dfrac{\lambda}{4}$

δ

0.24λ

L

$a = 0.00254\ m$
$f = 860\ MHz$
$L = 0.0156\ \mu H$
$\delta = 12\ a$

$50\ \Omega$

that a zero reactance occurs at 860 MHz. Fig. 6-46b shows the VSWR referenced to 50 Ohms. The VSWR curve could be made more symmetrical by slightly increasing the inductance at the base.

The far-field pattern is shown in Fig. 6-46c. Note the strong broadside radiation compared to the $3\lambda/2$ dipole case in Fig. 6-4. The gain of the $3\lambda/4$ monopole is at least 3dB greater than the $\lambda/4$ monopole, or about 8 to 9 dBi [H.3: Kraus, 3rd ed., p. 824].

Another form of common lumped loading is the top-loaded monopole shown in Fig. 6-47 (see also Fig. 3-10b). If the monopole is without top loading and $h < \lambda/5$, its radiation resistance will be half of (2-172) for the *short* dipole whose current goes to zero at the free space end, or

$$R_r = 40\pi^2 \left(\frac{h}{\lambda}\right)^2 \tag{6-52}$$

Recall that the total length of the dipole is Δz whereas the height of the monopole is h (i.e., $\Delta z = 2h$). If we can make the current on the monopole be uniform, then the radiation resistance will be four times greater since the radiation resistance of a monopole with uniform distribution will be half that of (2-169), or

$$R_r = 160\pi^2 \left(\frac{h}{\lambda}\right)^2 \tag{6-53}$$

as given in (3-15) and discussed in Sec. 3.3.1.

The uniform current distribution can be achieved, or nearly achieved, by the incorporation of the top load in Fig. 6-47. The top load contains that portion of the current that goes to zero at the end of a linear non-loaded monopole thereby rendering the vertical current portion nearly uniform. With the addition of the top load, the radiation resistance is increased by at least a factor of four, and the reactance is less capacitive because the

(a) Input impedance.

(b) Voltage standing wave ratio.

(c) Far-field pattern.
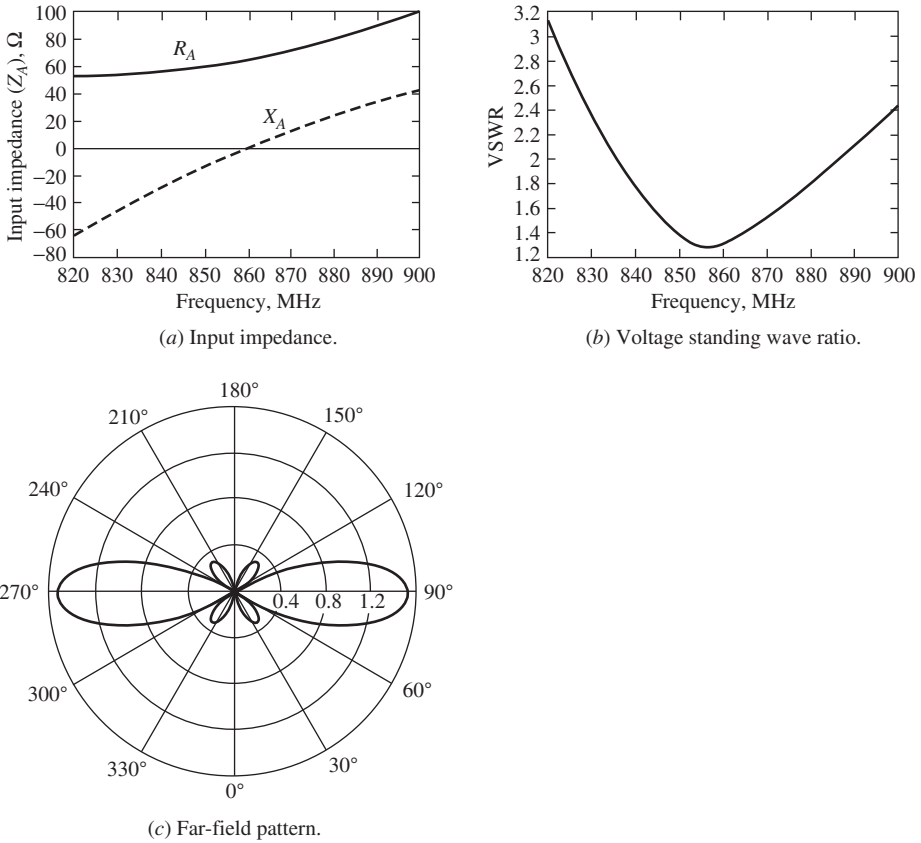
Figure 6-46 Performance of the three-quarter wavelength loaded monopole in Fig. 6-45.
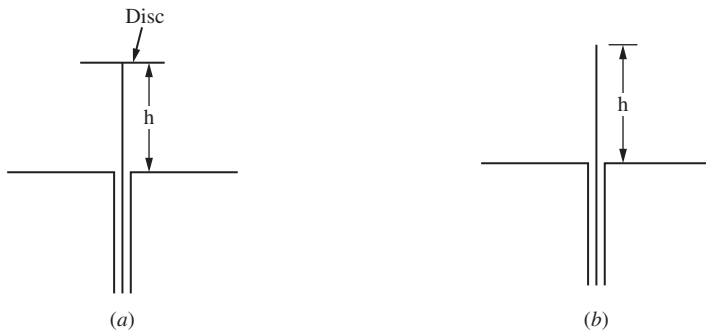
Figure 6-47 Loaded and unloaded monopoles. (a) End loaded monopole. (b) Unloaded monopole.

monopole appears to be electrically longer. That is, the longer-appearing monopole is more inductive.

Theoretically, we could use an inductor or coil of wire at the very end of the monopole to accomplish a similar result to that of the top load. For practical reasons, this is not often done.

6.5.2   **Distributively Loaded Wire Antennas and Fractals**

One form of distributive loading is resistive. The addition of resistance to the antenna conductor will lower the radiation Q but increase the bandwidth at the expense of antenna efficiency. In part because there are other more efficient ways of increasing antenna bandwidth, resistive loading of an antenna is seldom used.

A more common form of distributive loading is to use inductance to shorten the physical antenna length while preserving electrical length as we saw with lumped loading. Fractal geometries provide a way of naturally introducing inductive loading into an antenna. A fractal geometry is generated via the successive iteration of a generator shape. For example, in Fig. 6-48*a* the first iteration is the application of the generator shape (i.e., the triangle-like perturbation in the arm of the monopole). The second iteration is obtained from the first by applying the generator shape to the straight segments of the first iteration, and so on to higher-order iterations. Another example is the Sierpinski gasket monopole in Fig. 6-48*b*, which uses triangular voids as the generator shape in the original solid triangular plate.

Examination of Fig. 6-48*a* suggests distributive loading. That is, the deviation from straight wire geometry to the Koch fractal shape increases primarily the inductance between various parts of the Koch monopole. The increase in inductance allows the monopole to be shorter than the corresponding linear monopole and still be resonant.

Some fractals possess a property called *self-similarity*, wherein small regions of the geometry duplicate the overall geometry on a reduced scale. This property can lead to multi-band operation of the antenna. Such is the case with the Sierpinski gasket monopole of Fig. 6-48*b*. The overall size of the Sierpinski monopole determines the lowest frequency of operation while the individual voids help determine one or more higher frequencies of operation. The voids are a form of distributive loading.

Fractals were introduced to the antenna community in 1996 and have generated a bit of controversy. Originally fractals were thought to be broadband, but it now known that they can only be multiband. Fractals do indeed work but not necessarily better than other known antenna geometries [H.6: *Modern Antenna Hdbk*, Chap. 10; H.8.11: Hansen, pp. 74–81]. This is shown by Best in [H.6: *Modern Antenna Hdbk*, Chap. 10], where he shows that for the same wire length, the normal mode helix (see Fig. 7-12) has a lower
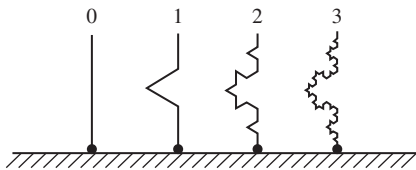
(*a*) Koch monopole, three iterations.

first iteration   fourth iteration
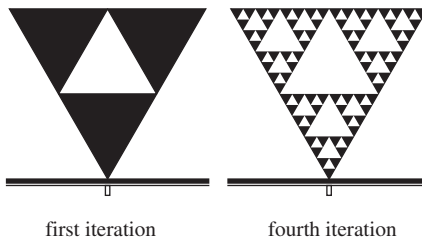
(*b*) Sierpinski gasket monopole, first and fourth iterations.

Figure 6-48 Two examples of fractal antennas.

resonant frequency than the Koch monopole because it is, in effect, a better inductor. In the case of the Sierpinski monopole, Best shows that the multiple resonances are not due to the details of the fractal geometry but can be duplicated by nonfractal slit voids in the triangular plate. In summary, fractal geometries through distributive loading can provide benefits such as reduced antenna size and possibly multi-band operation, but probably not in an optimum way [16].

Another form of distributive loading is dielectric loading. If an antenna such as a resonant quarter-wave monopole is embedded in an infinite dielectric, its resonant length is shortened by $1/\sqrt{\varepsilon_r}$. If the medium is not infinite, and the outer medium is air, the shortening will be less. Further, the larger ε_r, the more the radiation resistance is reduced with implications for impedance matching complexities and reduced matching circuit efficiency. And finally, bandwidth is also reduced as ε_r is increased. Lumped loading techniques are usually preferred over dielectric loading [H.9.1: Fujimoto, Chap. 3].

6.6   GROUND PLANE BACKED WIRE ANTENNAS

Earlier in the chapter we saw that the impedance bandwidth of a wire antenna could be improved by making a wire thicker. In this section, three ways of improving the gain of a wire antenna using flat metal plates are presented. The first way employs a single flat plate, and the next two each employ two flat plates. In the first two antennas, the flat plates function as a reflector to provide unidirectional radiation. In the third case, which also provides unidirectional radiation, one of the plates functions as a reflector and the other provides for a cavity-like resonance. These kinds of reflector antennas are simple to construct and cost effective. They sometimes are referred to as aperiodic reflectors. There is, of course, another major kind of reflector that uses curved surfaces, the most common one being the focused parabolic reflector. The parabolic reflector is treated in detail in Chap. 9.

6.6.1   The Flat Plate Reflector

The simplest reflector system is a dipole parallel to a large flat plate and positioned a distance, s, from it as shown in Fig. 6-49. When the flat plate is infinite, the method of images (Sec. 2.3) can be used to replace the dipole and infinite flat plate by two parallel dipoles a distance $2s$ apart. The image element will differ in phase by $180°$ from the original dipole and thus the pair will be a two-element endfire array giving us
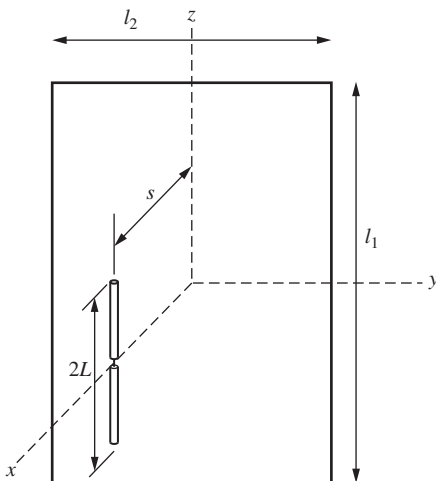
Figure 6-49 Dipole parallel to a flat rectangular plate.
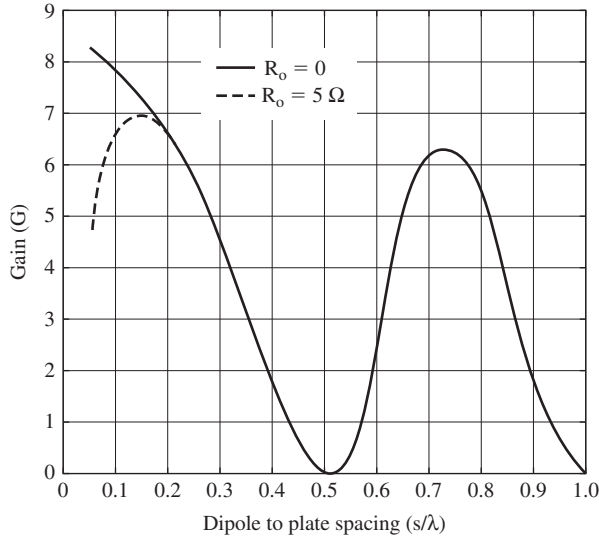
Figure 6-50 Calculated gain of a half-wavelength dipole vs. distance from a flat plate reflector. (Reproduced by permission from [H.3: Wolff], Norwood, MA: Artech House, Inc. 1988 © by Artech House, Inc.)

the radiation pattern in the half-space $x > 0$. The pattern in the H-plane is given by (3-71) and is illustrated in Fig. 3-20. The pattern in the E-plane is (3-71) multiplied by (3-3) and results in the narrower pattern in Fig. 3-20.

The two-element array model in Sec. 3.5 is an ordinary endfire array when $s = \lambda/4$. For $s < 0.25\lambda$ some superdirectivity will occur because the phase delay from image to dipole exceeds that of ordinary endfire. For $s < 0.25\lambda$ the directivity can exceed 2.15 dBi + 3.0 dB = 5.15 dBi as in Fig. 6-50 for the lossless $R_o = 0$ case [H.3: Wolff, p. 297]. When ohmic losses, R_o, are included, the gain peaks at about $s = 0.15\lambda$ but is still above 5.15 dBi. Fig. 6-50 is an example of superdirectivity not being the same as supergain because of ohmic losses (see Sec. 5.5).

The proximity of the dipole to the plate affects not just the gain but the input impedance of the dipole as well. This is because there is a coupling (see Sec. 8.7) between the dipole and the plate (or alternatively between the dipole and its image). The input resistance of the dipole as a function of the distance s is shown in Fig. 6-51. The deviation in input resistance from the 73 Ω value given in (6-17) is due to the interaction of the dipole with the flat plate.

When the plate is not infinite, its finite size will affect the pattern, affect the gain some, and have a very minor effect on the input impedance. These effects are not just due to the size of the plate but diffraction (see Sec. 16.7) from the edges of the plate. The half-power (HP) beamwidth is noticeably affected by the diffraction. Consider a half-wavelength dipole a quarter-wavelength in front of a flat plate. Then for a plate 1λ by 1λ, HP in the H-plane is 93°, for a 2λ by 2λ plate HP is 113°, and for a 3λ by 3λ HP is 111° [H.3: Milligan, 2nd. ed., p. 227]. By contrast, the half-power beamwidth for the infinite ground plane is 120° (see Ex. 3-5 and Prob. 3.5-5).

The size of the plate and its diffraction determine the front-to-back ratio. For example, for a plate 1λ by 1λ, the front-to-back ratio is about 16 dB, for 2λ by 2λ about 22 dB, and for 3λ by 3λ about 28 dB [H.3: Milligan, 2nd ed., p. 227]. For most applications, a 1.5λ by 1.5λ plate is adequate. The plate does not have to be solid, but instead can be replaced with a grid of wires or rods in the same plane as the incident electric field vector.

Fig. 6-52 presents a nomograph for determining the transmission through a grid of wires. The nomograph applies to the component of the electric field tangential to the
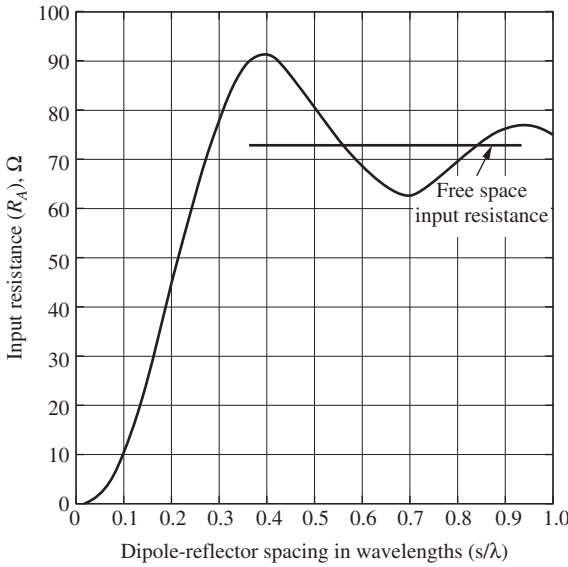
Figure 6-51 Calculated input resistance of a resonant dipole ($L = 0.4781\lambda$, $a = 0.001\lambda$) vs. distance from a flat plate reflector. (After [H.3: Kraus].)

wires in the grid. The wires are of diameter $2a$ with a spacing s between centers of the wires. To use the nomograph, a straight construction line is drawn from a location on the left corresponding to the spacing in wavelengths to a second point on the right corresponding to the ratio of the spacing to the wire radius. Where the construction line crosses the line in the middle labeled "Transmitted Power (dB)" determines the transmitted power. The formula used to construct the nomograph may be found in [17]. The accuracy of the nomograph was found to be better than ± 1dB when compared with transmitted power experimental data. The nomograph also applies to a grid of perpendicular wires by separately treating the other orthogonal set of parallel wires forming the mesh.

As an example, Fig. 6-52 predicts that for a wire spacing of 0.15λ, and a spacing to wire radius ratio of 12.5, the transmitted power is -13 ± 1 dB, corresponding to a reflected power between approximately 94% and 96% of the incident power. This represents a small loss in reflected power of between about 0.3 and 0.2 dB, respectively, a negligible uncertainty for most applications. Thus, the nomograph provides reasonable guidance for wire grid reflector design. The use of a wire grid reduces the weight and wind loading of the antenna while providing an efficient reflecting surface.

6.6.2 Corner Reflector Antennas

Another practical antenna that produces a gain of 10 to 12 dB over a half-wave dipole is the corner reflector antenna invented by J. D. Kraus in 1938 [4]. His first experimental model was a $90°$ corner reflector. Although other corner angles can be used, the $90°$ corner illustrated in Fig. 6-53 is the most practical and the one that will be discussed here. The corner reflector is used as a gain standard at UHF frequencies.

The corner reflector antenna can be easily analyzed using the method of images and array theory. Consider Fig. 6-54 that shows the source and its three images. The array factor contribution from the feed element 1 and image from 4 from (3-67) is $2 \cos[\beta s \cos(\phi)]$. The contribution from images 2 and 3 will be the same but rotated $90°$ and of opposite phase, or $-2 \cos[\beta s \cos (90° - \phi)]$. Thus, if we assume that the conducting reflecting sheets are infinite in extent, the array factor in the xy-plane (H-plane) valid in the region $-45° \leq \phi \leq 45°$ is

$$\text{AF}(\theta = 90°, \ \phi) = 2 \ \cos(\beta s \cos\phi) - 2 \ \cos(\beta s \sin\phi) \tag{6-54}$$
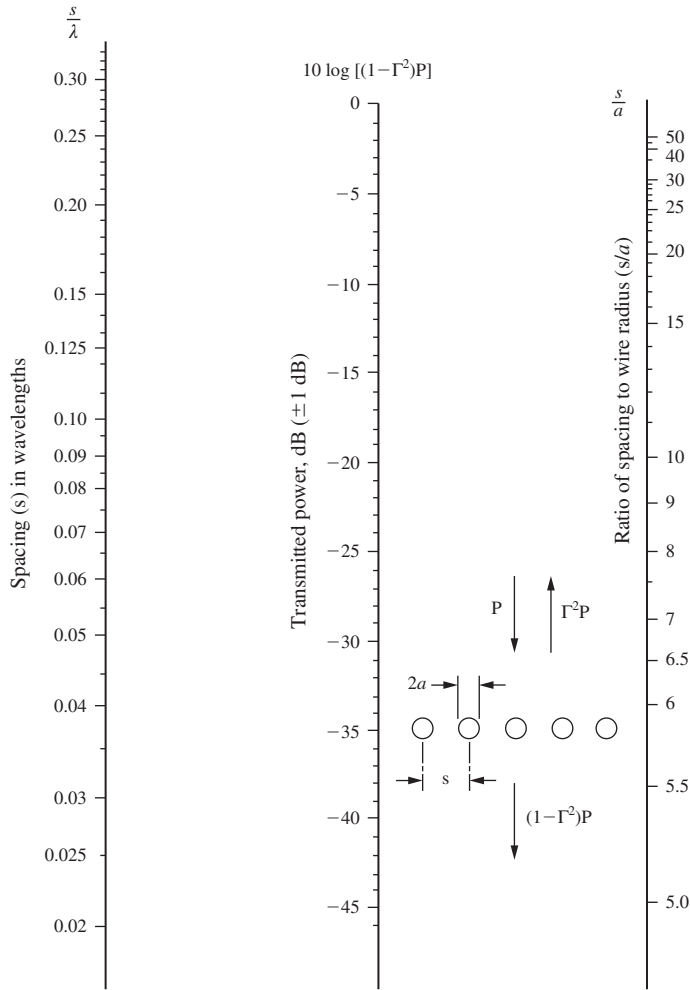
Figure 6-52 Power transmission through a grid of parallel wires. Γ is the voltage reflection coefficient of the wire-grid. (Used with permission from Mumford [17] © 1961 IEEE.)

It follows that in the *xz*-plane or *E*-plane, the array factor may be constructed by a somewhat similar reasoning process:

$$AF(\theta, \phi = 0°) = \{-2 + 2 \ \cos[\beta s \ \cos(90° - \theta)]\}g(\theta) \qquad (6\text{-}55)$$

where the element factor $g(\theta)$ is usually that of a half-wave dipole.

The pattern shape, gain, and feed point impedance will all be a function of the feed-to-corner spacing s (see Fig. 6-55a). For the 90° corner reflector, the pattern will have no minor lobes within $-45° \leq \phi \leq 45°$ and good directivity if $0.25\lambda \leq s \leq 0.75\lambda$. The directivity will be greatest at $s = 0.5\lambda$ [35] when the conducting plates are of infinite extent, but the input impedance of a dipole feed will be high (i.e., around 125 Ω). Reducing the spacing to 0.35λ will in theory produce a 70-Ω input impedance with a negligible decrease in gain. Often, a bow tie (see Fig. 7-34) is used for the feed because it has superior impedance bandwidth properties compared to an ordinary linear dipole.

Making the conducting plates of finite extent is, of course, necessary for a practical design. It can be shown by ray tracing that a length value of $L = 2s$ is a reasonable minimum length so that the main beam is not degraded by the finite extent of the
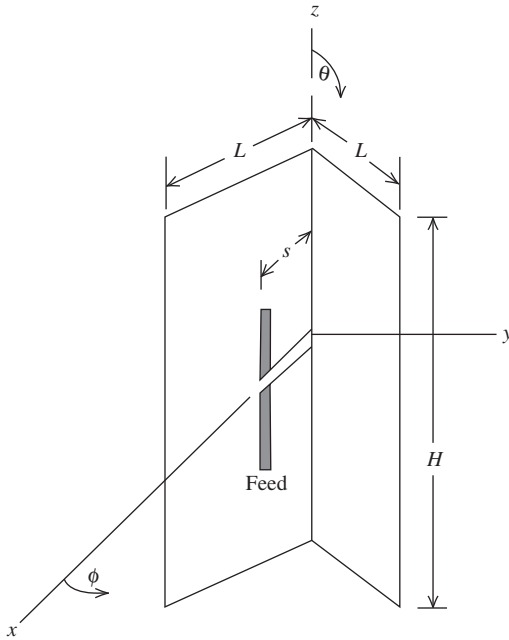
Figure 6-53 Right-angle corner reflector with metal plates in the $\phi = \pm 45°$ planes.
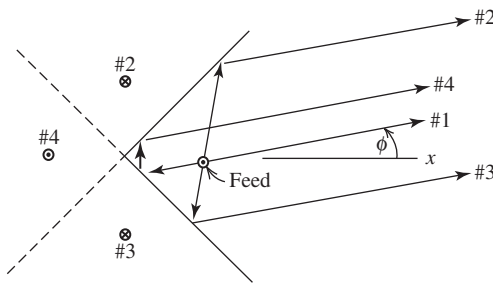
Figure 6-54 Right-angle corner reflector with images shown and how they account for reflections.

conducting plates. The dimension H is usually chosen to be from 1.2 to 1.5 times the length of the feed so as to minimize the direct radiation by the dipole feed into the back region. The finite extent of the plates will result in a pattern broader than that predicted for infinite plates as in, for example, Fig. 6-55. The effect of the finite plate size on the feed driving point impedance is usually negligible. Additional discussion of the corner reflector can be found in [H.3: Miligan, 2nd ed., pp. 237–241].

6.6.3   Backfire Antennas

The *short backfire antenna (SBFA)* is an efficient antenna that is simple and compact. The SBFA consists of two flat parallel reflectors of different sizes, usually spaced about 0.5λ apart as shown in Fig. 6-56. The dipole is usually positioned midway between the reflectors. The arrangement is sometimes called a leaky cavity resonator, but its radiation mechanism is not completely understood, in part because two or more radiation mechanisms are involved. The primary radiation direction from the SBFA is normal to the smaller reflector.

(*a*) H-plane.   (*b*) E-plane.
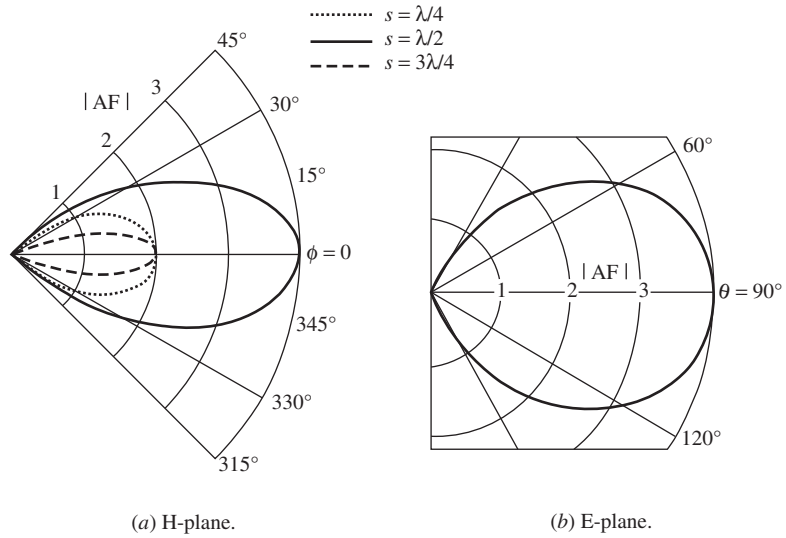
Figure 6-55 Principal plane patterns, |AF|, for a right angle corner reflector composed of two (semi-infinite) half-planes and a $\lambda/2$ dipole feed. (Only the $3\lambda/4$ spacing is used in the *E*-plane.)
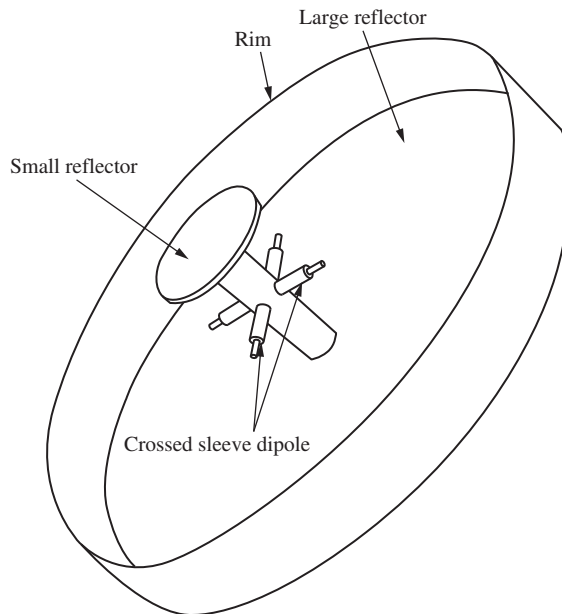
Figure 6-56 A short backfire antenna with crossed dipoles between the small and large reflectors. (Reproduced by permission from [H.8.12: Kumar and Hristov] Norwood, MA: Artech House, Inc. 1989 © by Artech House, Inc.)

The SBFA is of interest because of its several advantages: simple to design and construct; lightweight and compact; high gain for its size; low side lobes and back lobes; dual polarization and circular polarization capability; easy to impedance match; and high-power capability [H.8.11: Kumar, Chap. 5]. SBFAs are good candidates for use in antenna arrays and for applications where low weight is a strong consideration, such as on satellites and missiles.
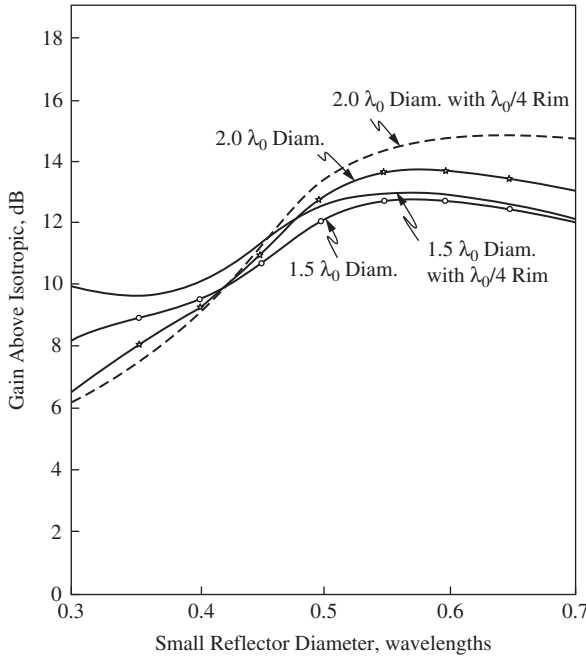
Figure 6-57 Measured gain at 1500 MHz vs. small reflector diameter for two sizes of large reflector with and without rim [18].

Dod [18] performed an experimental investigation of the SBFA shown in Fig. 6-56. His investigation shows the effect of the small disc size on half-power beamwidth, HP, and side lobe level, SLL, when the large reflector is 2.0λ in diameter. As the size of the small reflector is increased, HP is reduced and SLL increases. This behavior may suggest that the small reflector is acting like a director element in a Yagi because the small reflector diameter is less than its resonant diameter of about 0.58λ. This behavior is similar to the Yagi directors in Sec. 6.3 being shorter than their resonant length. Dod also experimented with the large reflector 1.5λ in diameter. In this case the same trends were found, but HP was smaller and SLL higher than for the larger diameter large reflector.

Fig. 6-57 shows the gain dependency on the size of the small reflector [18]. The curves tend to peak near the resonant size of the disc. The gain for the 2.0λ large reflector gives a higher gain than the 1.5λ large reflector when the small reflector diameter is near resonant size. The H-plane HP is about $37°$ and the E-plane HP about $34°$ when the large and small reflectors are 2.0λ and 0.5λ, respectively. If a rim 0.5λ high had been used, the gains would have been higher than in Fig. 6-57.

Fig. 6-58 shows the VSWR for the SBFA when the large reflector is 2.0λ in diameter [18]. Using VSWR $= 2:1$, the impedance bandwidth for this design is only about 3.3%. The low VSWR at 1.5 GHz was largely obtained by adjusting the length of the sleeves (see Sec. 7.5) on the dipoles and adjusting the length of the dipoles themselves. The impedance bandwidth can be improved at some expense to the gain by adjusting the dipole location relative to the large reflector [19, 20].

The curve in Fig. 6-58 is indicative of resonance behavior that may be interpreted either as a cavity resonance, or possibly a Yagi type resonance if one views the SBFA as behaving like a three-element Yagi with a single director and a super reflector. The parameters in Fig. 6-58 are typical ones for the SBFA except for the rim height which often is larger. Fig. 6-59 shows a typical SBFA pattern. The principal plane patterns are usually very similar. Side lobes are typically -20 dB or lower and back lobes about -30 dB. The pattern bandwidth is normally much larger than the impedance bandwidth for a typical design.
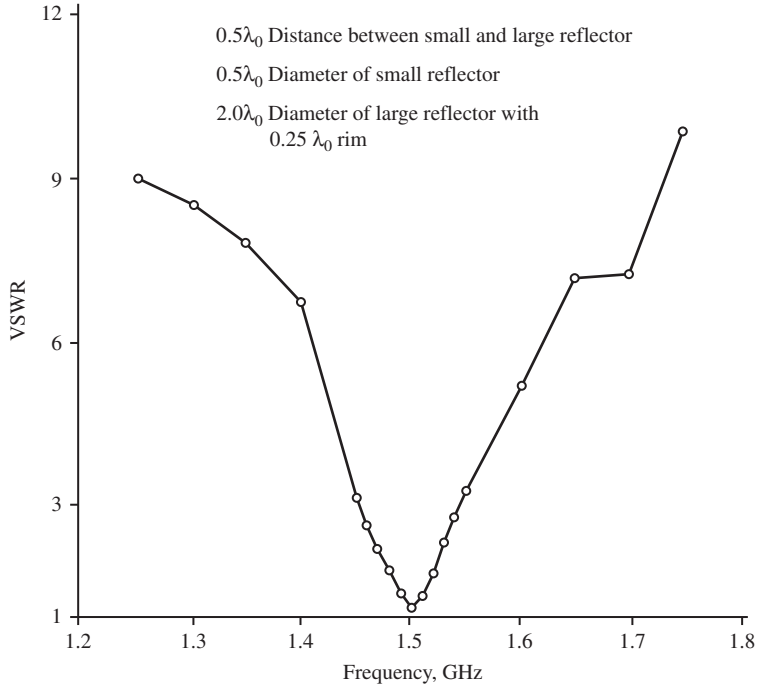
Figure 6-58 Measured VSWR vs. frequency [18].
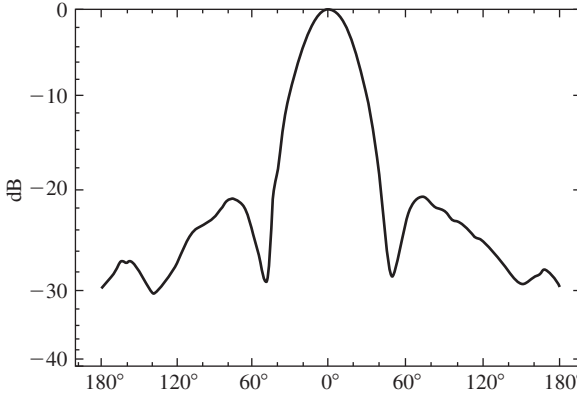
Figure 6-59 Principal plane pattern for a typical SBFA.

The rim on the large reflector in Fig. 6-56 has a strong effect on the side and back lobes as well as the gain. It has been found that a rim height on the order of 0.5 to 0.6λ can add as much as 5 dB to the gain [22] compared to no rim at all. Define an antenna aperture for the SBFA as being a plane area the size of the large reflector and resting on its rim. Assuming the circular aperture in Fig. 6-56 is the effective aperture, and then using (4-27) with $A_e = A_p$, where A_p is the physical aperture, the gain of the antenna would be

$$G = \frac{4\pi}{\lambda^2}A_p = \frac{4\pi}{\lambda^2}(\pi \cdot 1.0\lambda^2) = 39.48 = 16\,\text{dBi} \qquad (6\text{-}56)$$

This is consistent with measured gains reported in [21]. The assumption that the physical aperture is the effective aperture implies a uniform aperture distribution and represents a difficult to achieve upper bound on the gain. (Recall that the maximum directivity of the broadside line source in Chap. 5 occurred when the line source had a uniform distribution.) However, near field measurements have indicated that the aperture has a cosinusoidal distribution and that this distribution extends beyond the rim creating a virtual aperture larger than the physical aperture [22]. This helps explain the unexpectedly high values of gain. The extension of the aperture fields beyond the edge of the rim can be explained by diffraction theory in Chap. 16. Diffraction theory can also explain why the best rim height is 0.5 to 0.6λ.

The SBFA is easier and more economical to build than a parabolic dish and can produce somewhat better gains for the same diameter and less depth. Further, the SBFA produces as much gain as a Yagi 10 times longer in the boom length [23]. Thus, it is not surprising that a SBFA was used for communications in the Apollo spacecraft program, where a compact, lightweight, high-gain antenna was required on the command module.

6.7   WIRE ANTENNAS ABOVE AN IMPERFECT GROUND PLANE

The operation of low-frequency (roughly VHF and below) antennas is affected significantly by the presence of typical environmental surroundings, such as the Earth, buildings, and so forth. In Sec. 2.3, we discussed the principles for analyzing antennas above a perfect ground plane. A perfect ground plane in its ideal form is an infinite, plane, perfect conductor. It is well approximated in practice by a planar good conductor that is large relative to the antenna extent. Image theory from Sec. 3.3 reveals that an antenna above a perfect ground plane, or an approximation of it, has an equivalent form that is an array. Array theory can then be used to obtain the radiation pattern above the ground plane.

In this section, we consider ground planes that are not well approximated by a perfect ground plane. Since low-frequency antennas are most affected by their surroundings and low-frequency antennas are usually wire antennas, the illustrations will be for wire antennas above a ground plane. The general principles can, however, be applied to many antenna types.

A ground plane can take many forms, such as radial wires around a monopole, the roof of a car, or the real Earth. In many situations, the Earth is well approximated as being infinite and planar, but it is a poor conductor. Good conductors have conductivities on the order of 10^7 S/m. Earth conductivity varies greatly, but is typically 10^{-3} to 12×10^{-3} S/m with rich soil at the high end and rocky or sandy soil at the low end. With these low conductivities, electric fields generated by a nearby antenna penetrate into the Earth and excite currents that, in turn, give rise to $\sigma|E|^2$ ohmic losses. This loss appears as an increase in the input ohmic resistance and thus lowers the radiation efficiency of the antenna.

6.7.1   Pattern Effects of a Real Earth Ground Plane

The pattern of an antenna over a real Earth is different from the pattern when the antenna is operated over a perfect ground plane. Approximate patterns can be obtained by using image theory. The same principles discussed in Sec. 3.3.4 for images in perfect ground planes apply, except that the strength of the image in a real ground will be reduced from that of the perfect ground plane case (equal amplitude and equal phase for vertical elements, and opposite phase for horizontal elements). The strength of the image can be approximated by weighting it with the plane wave reflection coefficient for the appropriate polarization of the field arriving at the ground plane. To illustrate, consider a short vertical dipole a distance h above a ground plane, shown in Fig. 6-60 together with its image. There is a direct and a reflected ray arriving in the far field. As can be seen, the reflected ray appears to be coming from the image antenna. The primary source and its image form an array. The electric field above the ground plane for this example, using (2-74a), is
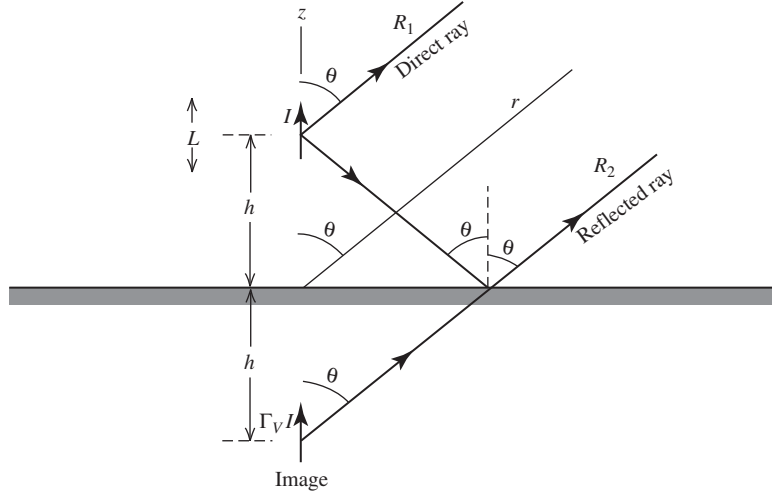
Figure 6-60 A short vertical dipole of current L above a real Earth ground plane, together with its image of current $\Gamma_V I$.

$$E_\theta = j\omega\mu \ \sin\theta\left(IL\frac{e^{-j\beta R_1}}{4\pi R_1} + \Gamma_V IL\frac{e^{-j\beta R_2}}{4\pi R_2}\right) \tag{6-57}$$

where L is the length of the short dipole and $\Gamma_V I$ is the current for the image dipole. Γ_V is the plane wave reflection coefficient for a planar Earth and vertical incident polarization, when E is in the plane of incidence defined by a normal to the Earth and the ray from the source to the normal at the Earth. Using parallel rays for far-field calculations gives the far-field distance expressions

$$R_1 = r - h \ \cos\theta \qquad \text{and} \qquad R_2 = r + h \ \cos\theta \tag{6-58}$$

Then (6-57) reduces to

$$E_\theta = j\omega\mu\frac{IL}{4\pi}\frac{e^{-j\beta r}}{r} \ \sin\theta(e^{j\beta h \cos\theta} + \Gamma_V e^{-j\beta h \cos\theta}) \tag{6-59}$$

where $R_1 \approx R_2 \approx r$ was used in the denominator. This expression is valid above the ground plane. It contains an element pattern $\sin\theta$ and an array factor, in the brackets, for a two-element array with elements spaced $2h$ apart.

Similarly for a horizontally oriented short dipole as shown in Fig. 6-61, we have (in the xz-plane)

$$E_\theta = j\omega\mu\frac{IL}{4\pi}\frac{e^{-j\beta r}}{r} \ \cos\theta(e^{j\beta h \cos\theta} + \Gamma_V e^{-j\beta h \cos\theta}) \tag{6-60}$$

where the minus sign appears because the image current is in the opposite direction. This expression is valid only in the xz-plane. Γ_V is used because E is in the plane of incidence. The field in the yz-plane is given by

$$E_\phi = j\omega\mu\frac{IL}{4\pi}\frac{e^{-j\beta r}}{r} \ (e^{j\beta h \cos\theta} + \Gamma_H e^{-j\beta h \cos\theta}) \tag{6-61}$$
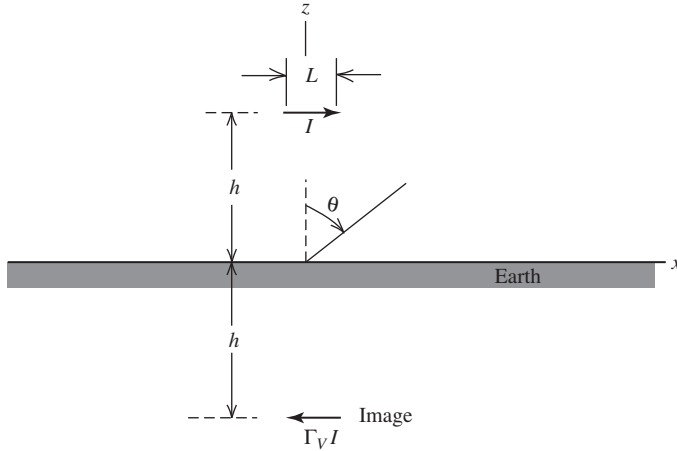
Figure 6-61 A short horizontal dipole of current I above a real Earth ground plane together with its image of current $\Gamma_V I$ for the xz-plane. The image current in the yz-plane is $\Gamma_H I$.

The element pattern is unity because a dipole has an omnidirectional pattern in the plane normal to the dipole axis. The horizontal reflection coefficient Γ_H is used in this case because the electric field is perpendicular to the plane of incidence.

There is no minus sign in the second term of (6-61) because of the definition of Γ_H, which is [H.4: Jordan, Chap. 16].

$$\Gamma_H = \frac{\cos\theta - \sqrt{\varepsilon_r' - \sin^2\theta}}{\cos\theta + \sqrt{\varepsilon_r' - \sin^2\theta}} \tag{6-62}$$

This is the plane wave reflection coefficient for an incident electric field perpendicular to the plane of polarization (i.e., the plane formed by the surface normal and the direction of propagation). Further, for Γ_V we have [H.4: Jordan, Chap. 16].

$$\Gamma_V = \frac{\varepsilon_r' \cos\theta - \sqrt{\varepsilon_r' - \sin^2\theta}}{\varepsilon_r' \cos\theta + \sqrt{\varepsilon_r' - \sin^2\theta}} \tag{6-63}$$

This is the plane wave reflection coefficient for an incident electric field parallel to the plane of polarization. ε_r' is the relative complex effective dielectric constant (see Sec. 2.1) for the ground and is given by

$$\varepsilon_r' = \frac{\varepsilon'}{\varepsilon_o} = \varepsilon_r - j\frac{\sigma}{\omega\varepsilon_o} \tag{6-64}$$

ε_r and σ are the relative dielectric constant and conductivity of the ground plane. The Earth has an average value of $\varepsilon_r = 15$. Ground conductivities across the United States vary from 10^{-3} to 12×10^{-3} S/m [H.4: Jordan, Chap. 16].

It is convenient to express the imaginary part of ε_r' as

$$\frac{\sigma}{\omega\varepsilon_o} = 18 \times 10^3 \frac{\sigma}{f_{\text{MHz}}} \tag{6-65}$$
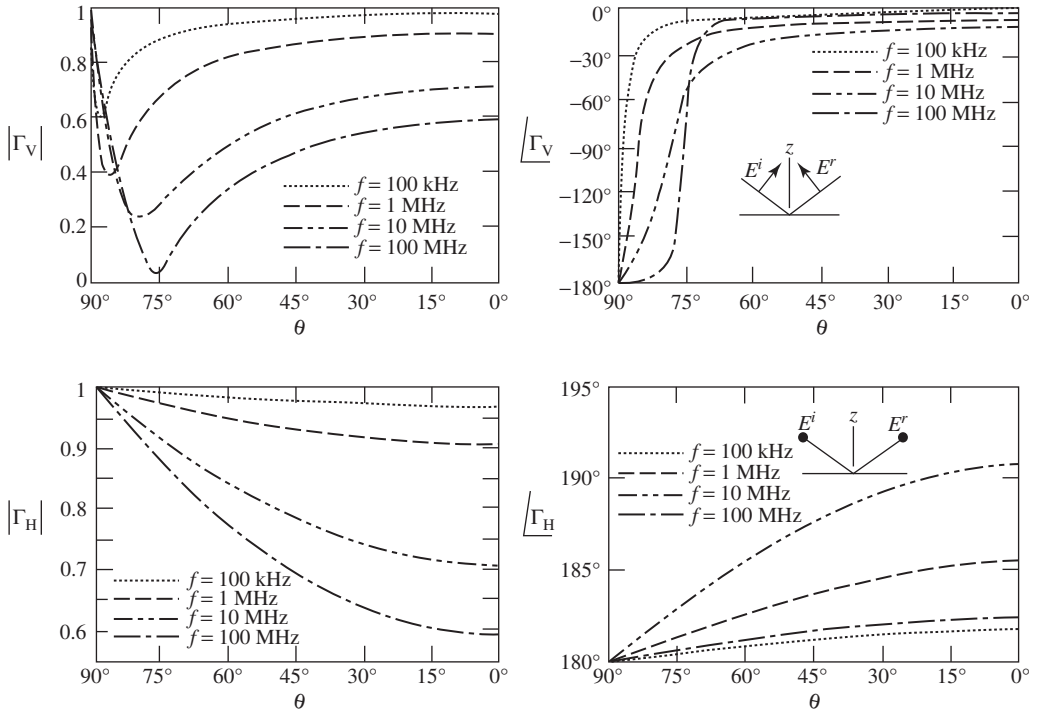
Figure 6-62 Magnitude and phase of Γ_V and Γ_H at four frequencies. Note that the horizon is at $\theta = 90°$. $\sigma = 12 \times 10^{-3}\,\mathrm{S/m}$ and $\varepsilon_r = 15$.

At low frequencies (e.g., 1 MHz and below), the imaginary part or loss-producing part of the complex permittivity dominates. At high frequencies (e.g., 100 MHz and above), the real part dominates.

The reflection coefficients are shown in Fig. 6-62 for a typical ground conductivity $\sigma = 12 \times 10^{-3}$. There is a great deal of information in these curves. First, we note that Γ_H is close to -1 at low elevation angles ($\theta \sim 90°$) for all frequencies and ground conductivities. The situation is much different for Γ_V where both the magnitude and phase usually vary rapidly at low elevation angles due to the presence of the pseudo-Brewster angle when $\angle \Gamma_V = -90°$. This behavior does not occur when either $\sigma \to \infty$ or $\omega \to 0$ since $|\Gamma_V| \to 1$ as can be shown from (6-63) and (6-64).

The use of plane wave reflection coefficients to obtain the image strength is only an approximation since antennas near a ground plane do not form plane waves incident on the ground plane. In addition to the radiation we have described above, there is a surface wave that propagates along the ground plane surface. For HF and VHF frequencies, the surface wave attenuates very rapidly. For grazing angles (θ near $90°$), $\Gamma_V \approx -1$ and vertical antennas close to a real Earth have zero radiation for $\theta = 90°$; see (6-59). In this case, the surface wave accounts for all propagation, as in daylight standard broadcast AM. The effect of neglecting the surface wave, and using the procedure given above, has been found not to be critical for vertical antennas [24]. For horizontal antennas, the antenna should be at least 0.2λ above the Earth for the plane wave reflection coefficient method to be valid [25].

The elevation pattern for a short vertical dipole at the surface of various ground planes is shown in Fig. 6-63. When the ground plane is perfect ($\sigma = \infty$), the pattern above the ground plane is the same as that of a short dipole in free space, $\sin\theta$. Thus, in the perfect ground plane case, radiation is maximum along the ground plane, whereas for a real Earth ground plane, the radiation maximum is tilted up away from the ground
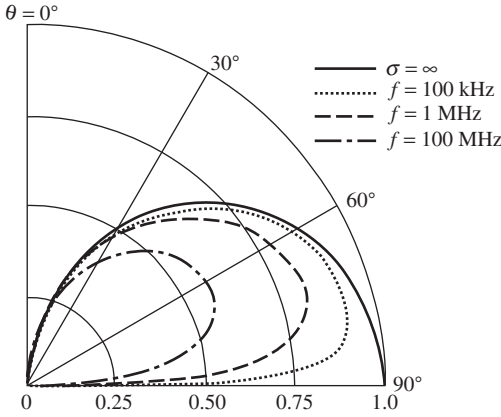
Figure 6-63 Elevation plane patterns of a vertical short dipole at the surface of a flat Earth with $\varepsilon_r = 15$ and $\sigma = 1.2 \times 10^{-2}$ S/m for three frequencies compared to the perfect ground plane ($\sigma = \infty$) case.

plane and is reduced in intensity, for the same input power, due to reduced efficiency. This is a general trend. The effect of a lossy Earth on vertical antennas is to tilt the radiation pattern upward. A good radial ground system (to be discussed in Sec. 6.7.2) makes the pattern behave more nearly like that for a perfect ground plane, that is, increase the low angle radiation (along the ground plane). Low angle radiation is particularly important for long-distance communication links that rely on ionospheric reflection (skip).

A short vertical dipole that is $\lambda/4$ above the ground plane forms a $\lambda/2$ spaced array with its image. For the perfect ground plane, $\varepsilon_r' = \infty$ and Γ_V from (6-63) is +1. The array is then a $\lambda/2$ spaced, equally excited, in-phase collinear array. The pattern for this is plotted in Fig. 6-64a ($\sigma = \infty$). For a real Earth ground plane, $\Gamma_V \approx -1$ at grazing angles (θ near 90°). The array contributions thus cancel, giving a null along the ground plane as shown in Fig. 6-64a. As the height h is increased to $\lambda/2$, the equivalent array of Fig. 6-60 has a λ spacing and multiple lobes appear in addition to the effects described for $h = \lambda/4$. The elevation patterns for $h = \lambda/2$ are plotted in Fig. 6-64b.

For a horizontal short dipole as shown in Fig. 6-61, the radiation is not the same for all planes through the z-axis, as for vertical antennas. In the yz-plane (perpendicular to the axis of the dipole), the radiation electric field is given by (6-61). The reflection coefficient Γ_H is exactly -1 for a perfect ground plane and approximately -1 for real Earth ground planes at all angles θ if the frequency is low. The element pattern is isotropic since the elements are seen in end view in the H-plane (yz-plane). Thus, the array factor completely determines the pattern. The low elevation pattern effects of finite conductivity are much less pronounced for horizontal antennas than for vertical antennas.

The field expressions of (6-59) to (6-61) for short dipoles above a ground plane can be used for other antenna types by using the appropriate element pattern. In particular, sin θ in (6-59) and cos θ in (6-61) are replaced by the free space pattern of the antenna considered.

6.7.2   Ground Plane Construction

The previous section illustrates the effect of an imperfect Earth ground plane on the performance of an antenna in proximity to the air-Earth interface. These effects can be overcome by using a highly conducting ground plane. An excellent ground plane is one constructed by using a metallic sheet that is much larger than the antenna extent. For example, consider the monopoles on finite size perfectly conducting ground planes in Fig. 16-27. For such ground planes the only departure from an ideal pattern, such as that

(a) Short dipole $\lambda/4$ above the ground, $h = \lambda/4$.

(b) Short dipole $\lambda/2$ above the ground, $h = \lambda/2$.
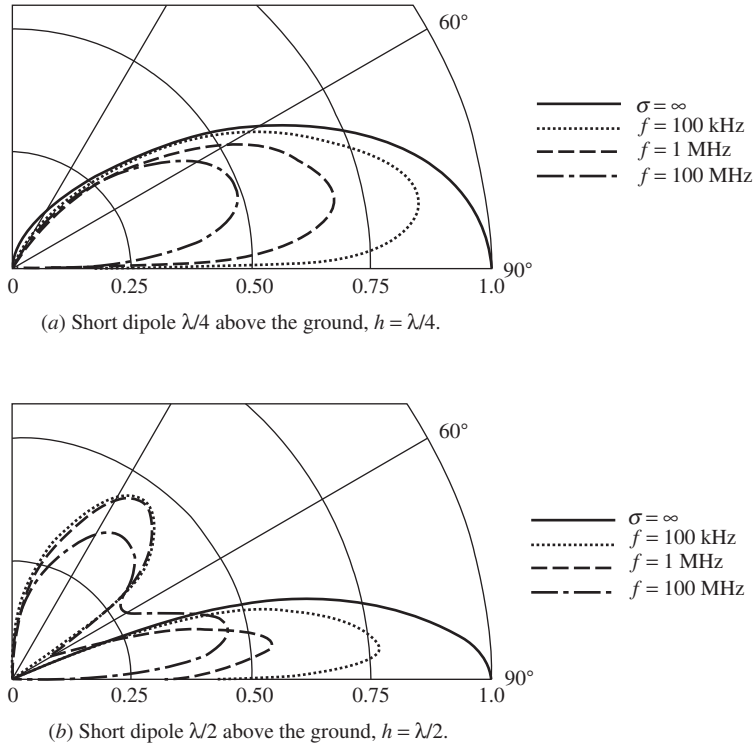
Figure 6-64 Elevation plane patterns of a vertical short dipole a distance h above a flat Earth (Fig. 6-60) with $\varepsilon_r = 15$ and $\sigma = 1.2 \times 10^{-2}$ S/m for three frequencies compared to the perfect ground plane $(\sigma = \infty)$ case.

in Fig. 6-63 for $\sigma = \infty$, is caused by the finite size of the ground plane. The pattern for $0 < \theta < 90°$ in Fig. 16-28 is an example of finite ground plane size on the radiation pattern when compared to Fig. 6-63. The edges of the ground plane can have an effect on the antenna input impedance as well. This effect is illustrated in Fig. 16-41.

A solid metallic ground plane is impractical at low frequencies because of the minimum required radial dimension of a quarter-wavelength. In this section techniques are discussed for increasing the apparent conductivity of the real Earth ground to minimize the effects of the Earth upon the radiation pattern and the power lost in the Earth.

Consider a vertical monopole antenna with its base at ground level. (See Sec. 3.3 for a discussion of monopoles over a perfect ground plane.) Currents flowing up the antenna leave the antenna and form displacement currents in air. Upon entering the Earth, conduction currents are formed that converge toward the base of the antenna. Losses in an Earth ground can be reduced by providing a highly conductive return path. This is commonly achieved with a *radial ground system*. The size of the wires used is not critical and is determined by the mechanical strength required. Number 8 AWG wire is typical. They need not be buried, but it is usually convenient to do so. However, they should not be buried too deep in order to minimize the extent of Earth through which the fields must pass. Sometimes, the radial wires are linked together at the base of the monopole by a ring-shaped ground strap. Occasionally, one or more stakes are driven into the ground near the base of the monopole.

The ohmic resistance of the radial system and Earth ground adds in series to the ohmic resistance of the monopole structure to determine the total ohmic resistance of the input impedance. The ohmic losses in the ground plane, with or without radials, can be

separated into *H*-field and *E*-field losses, calculated individually, and then simply added together [27]. The *H*-field loss is associated with the tangential magnetic field on the ground surface whereas the *E*-field loss is due to the vertical electric field immediately above the ground plane. The *H*-field loss is usually much greater than the *E*-field loss. The ohmic losses in the ground plane radial wires are negligible compared to the two ohmic losses in the Earth itself [27].

Especially for high-power transmitting antennas, it is important to have a well-designed radial system to achieve high efficiency. On the other hand, for a simple monopole structure in free space, as few as three equally spaced radial wires form an adequate radial system. Often these radials slant downward so that the maximum radiation is in the horizontal plane. (Fig. 16-28 is a monopole on a flat ground plane whose maximum radiation is at $\theta < 90°$ due to the finiteness of the ground plane.) For a monopole on the lossy Earth, an additional benefit of a good radial system is maximum radiation being near the horizon instead of being elevated as in Fig. 6-63.

The most sophisticated ground system, such as used with a standard broadcast AM-transmitting antenna, is 120 radial wires spaced equally 3° apart, around the tower out to a distance of about a quarter-wavelength from the tower. In general, the length of the radials is roughly equal to the height of the monopole antenna. The value of the total ohmic resistance of a ground system with 120 radials for typical soils is plotted in Fig. 6-65 for a few frequencies as a function of radial length [26]. Note that at 3 MHz a ground system with 120 radials that are about a quarter-wavelength long (25 m) gives a ground system resistance of about 1.6 Ω. Since the surface resistance of the Earth varies as the square root of frequency [see (2-171)], the ground system resistance will be constant for lower frequencies if the length of the radials is increased in proportion to the square root of wavelength. For frequencies above 3 MHz, the curve for radial length in Fig. 6-65 is only slightly to the right of the 3-MHz curve. This is because after the radials reach a
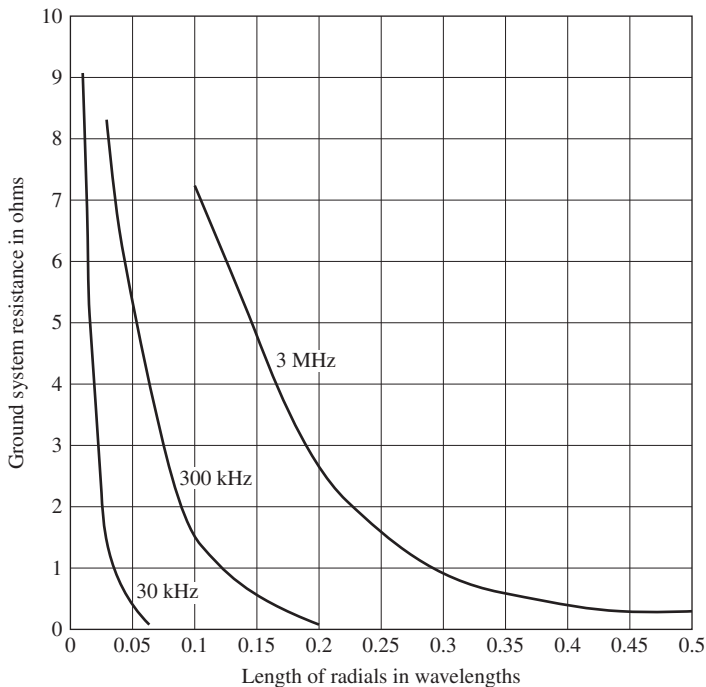
Figure 6-65 Typical resistance of radial ground systems using 120 radials in average soil. (Data from [26].)

length of about a quarter-wavelength, most of the large current densities occur within the region of the radials and further length increase is of no major consequence.

The construction principles for a radial wire ground system on top of or in the Earth can be summarized rather simply. The function of a radial system is to prevent the electromagnetic fields from the antenna from penetrating into the ground and exciting currents that, in turn, lead to $\sigma|E|^2$ ohmic loss. As can be seen from the discussion, if 120 quarter-wavelength-long radials are employed, the ohmic resistance introduced by the ground system will be at most a few ohms, and usually well under an ohm. In most applications, it is impractical to install as many as 120 radials. Generally speaking, 50 radials about a quarter-wavelength-long will reduce Earth losses to a few ohms. When only a few radials are used, the added resistance of the ground can be several ohms. Also if the radial lengths (almost independent of the number used) are reduced below a tenth of a wavelength, the ground system resistance will increase significantly.[1] The radial wires can be laid on top of the ground or buried slightly (but never deeply buried). Wire selection is largely determined by mechanical considerations. As the number of radials is increased, the less current each one will have to carry and thus the smaller the wire diameter required. At the base of the antenna, the radials should be connected together and to one or more ground stakes.

6.8 LARGE LOOP ANTENNAS

The topic of this section is electrically large loop antennas which have a loop perimeter L that is about a tenth of a wavelength or more. Before discussing large loops, we review small loops. Electrically small loop antennas, with a perimeter much less than a wavelength, have a doughnut-shaped pattern that is omnidirectional in the plane of the loop and polarized in the plane of the loop (the xy-plane of Fig. 3-15). The pattern of a small dipole along the z-axis is also omnidirectional in the xy-plane, but has a radiated electric field that is polarized perpendicular to the xy-plane; see Fig. 2-4. We found in Sec. 3.4 that the pattern and radiation resistance of electrically small loop antennas are insensitive to loop shape and depend only on the loop area. The doughnut pattern with a maximum in the plane of the small loop and with zero radiation along the axis normal to the loop is a consequence of the current amplitude and phase being constant around the loop. Electrically small loops antennas have low radiation resistance and high inductive reactance, making them unsuitable for transmitting applications.

Large loops have a current amplitude and phase that vary with position around the loop, causing the impedance and pattern to depend on loop size. There is a pattern maximum perpendicular to the plane of the loop if its perimeter is about a wavelength or more. Large loop antennas are used in *direction finding (DF)* systems that make use of the sharp null in the horizontal plane (along the x-axis) when the loop is oriented with the y-axis vertical in Fig. 6-66. This allows accurate angle-of-arrival determination in the horizontal plane, which is the most common application. The performance of a fixed physical size large loop antenna will change with varying frequency, which is characteristic of a resonant antenna. Large loop antennas usually have either a circular or square shape. At low frequencies where physical sizes are large, loops are often square because they are more easily constructed by using crossed, diagonal wooden support members to maintain the wire shape square. Self-supporting circular loops are common at UHF frequencies and above. Both circular and square loops are usually operated near the first resonance point, which occurs for a perimeter of slightly greater than one wavelength. Storer [28] provides analytical expressions for the current distribution and impedance of

---

[1]More details and references for ground system design are available in [H.3: Weeks, Sec. 2.6].
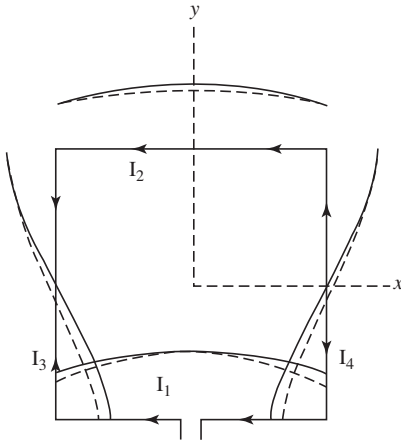
Figure 6-66 The one-wavelength square loop antenna. Each side is of length $\lambda/4$. The solid curve is the sinusoidal current distribution of (6-66). The dashed curve is the current magnitude obtained from more exact numerical methods.

circular loops, accompanied by extensive tables and plots. Also available in the literature for circular loop antennas are general far-field expressions [29] and approximate formulas for the directivity and radiation resistance [30]. A good summary of circular loops is found in [H.3: Burberry, Chap. 4]. The uniform current circular loop is addressed in Prob. 6.8-1. A specially configured square loop antenna, called the *Alford loop antenna*, with a perimeter of about 1λ achieves close to a uniform current and thus produces nearly omnidirectional, horizontally-polarized radiation in the horizontal plane (with the loop in the horizontal plane), which is hard to achieve with other antennas in a compact form factor and with reasonable impedance and efficiency. Large circular and square loops with the same perimeter have similar performance; see Prob. 6.8-5.

First, we analyze the **one-wavelength square loop antenna** using the same techniques that we used for other resonant wire antennas by assuming a sinusoidal current distribution. The assumption of sinusoidal current distribution has been verified to a good approximation through experiment [31]. Next, accurate numerical method results are compared to the approximate analysis results. Finally, computed impedance values are given for the square loop as a function of perimeter size.

The one-wavelength square loop antenna has the assumed sinusoidal current distribution shown in Fig. 6-66 (solid curve) that is continuous around the loop, starting with a maximum at the feed point at the center of the side parallel to the x-axis. Each side has a quarter-wavelength of current that is expressed as

$$\mathbf{I}_1 = \mathbf{I}_2 = -\hat{\mathbf{x}}I_o \, \cos(\beta x'), \qquad |x'| \le \frac{\lambda}{8} \tag{6-66}$$

$$\mathbf{I}_4 = -\mathbf{I}_3 = -\hat{\mathbf{y}}I_o \, \sin(\beta y'), \qquad |y'| \le \frac{\lambda}{8}$$

The solution for the radiation properties proceeds in the usual manner. First, the vector potential from (2-101) is

$$\mathbf{A} = \mu \frac{e^{-j\beta r}}{4\pi r} \int_{\text{loop}} \mathbf{I} e^{j\beta \hat{\mathbf{r}} \cdot \mathbf{r}'} \, dl \tag{6-67}$$

To find the phase delay function, the expressions for vectors from the origin to arbitrary positions on each side must be written. They are

$$\mathbf{r}_1' = x'\hat{\mathbf{x}} - \frac{\lambda}{8}\hat{\mathbf{y}} \qquad \mathbf{r}_2' = x'\hat{\mathbf{x}} + \frac{\lambda}{8}\hat{\mathbf{y}}$$

$$\mathbf{r}_3' = -\frac{\lambda}{8}\hat{\mathbf{x}} + y'\hat{\mathbf{y}} \qquad \mathbf{r}_4' = \frac{\lambda}{8}\hat{\mathbf{x}} + y'\hat{\mathbf{y}}$$

$$(6\text{-}68)$$

where the numbered subscripts indicate the corresponding loop side. Using the expansion of $\hat{\mathbf{r}}$ from (C-4) and (6-68) in (6-67) with the loop integral broken into integrals over each side gives

$$
\mathbf{A} = \mu\frac{e^{-j\beta r}}{4\pi r}I_o\left[-\hat{\mathbf{x}}\int_{-\lambda/8}^{\lambda/8}\cos(\beta x')e^{j\beta x'\,\sin\theta\,\cos\phi}\left(e^{-j(\pi/4)\sin\theta\,\sin\phi} + e^{j(\pi/4)\sin\theta\,\sin\phi}\right)\,dx'\right.
$$

$$
\left. +\hat{\mathbf{y}}\int_{-\lambda/8}^{\lambda/8}\sin(\beta y')e^{j\beta y'\,\sin\theta\,\sin\phi}\left(-e^{-j(\pi/4)\sin\theta\,\cos\phi} + e^{j(\pi/4)\sin\theta\,\cos\phi}\right)\,dy'\right]
$$

$$
= \mu\frac{e^{-j\beta r}}{4\pi r}I_o\left[-\hat{\mathbf{x}}2\,\cos\left(\frac{\pi}{4}\sin\theta\,\sin\phi\right)\int_{-\lambda/8}^{\lambda/8}\cos(\beta x')e^{j\beta x'\,\sin\theta\,\cos\phi}\,dx'\right.
$$

$$
\left. +\hat{\mathbf{y}}2j\,\sin\left(\frac{\pi}{4}\sin\theta\,\cos\phi\right)\int_{-\lambda/8}^{\lambda/8}\sin(\beta y')e^{j\beta x'\,\sin\theta\,\sin\phi}\,dy'\right]
$$

$$(6\text{-}69)$$

The first factors in each of the above two terms in brackets are the array factors for the pairs of sides 1, 2 and 3, 4, respectively. Evaluation of the integrals and subsequent simplification (see Prob. 6.8-2) lead to

$$
\mathbf{A} = \mu\frac{e^{-j\beta r}}{4\pi r}\frac{2\sqrt{2}I_o}{\beta}\left\{\hat{\mathbf{x}}\frac{\cos[(\pi/4)\,\cos\Omega]}{\sin^2\gamma}\left[\cos\gamma\,\sin\left(\frac{\pi}{4}\cos\gamma\right) - \cos\left(\frac{\pi}{4}\,\cos\gamma\right)\right]\right.
$$

$$
\left. -\hat{\mathbf{y}}\frac{\sin[(\pi/4)\,\cos\gamma]}{\sin^2\Omega}\left[\cos\Omega\,\cos\left(\frac{\pi}{4}\cos\Omega\right) - \sin\left(\frac{\pi}{4}\,\cos\Omega\right)\right]\right\}
$$

$$(6\text{-}70)$$

where

$$\cos\gamma = \sin\theta\,\cos\phi \qquad\text{and}\qquad \cos\Omega = \sin\theta\,\sin\phi \qquad (6\text{-}71)$$

The angles γ and Ω have a geometrical interpretation; they are the spherical polar angles (similar to θ) for the x- and y-axes; see (C-4).

The far-zone electric field components are

$$
E_\theta = -j\omega A_\theta = -j\omega\mathbf{A}\cdot\hat{\boldsymbol{\theta}} = -j\omega(A_x\hat{\mathbf{x}}\cdot\hat{\boldsymbol{\theta}} + A_y\hat{\mathbf{y}}\cdot\hat{\boldsymbol{\theta}})
$$

$$
= -j\omega(A_x\,\cos\theta\,\cos\phi + A_y\,\cos\theta\,\sin\phi) \qquad (6\text{-}72a)
$$

$$
E_\phi = -j\omega\mathbf{A}\cdot\hat{\boldsymbol{\phi}} = -j\omega(-A_x\,\sin\phi + A_y\,\cos\phi) \qquad (6\text{-}72b)
$$

Substituting A_x and A_y from (6-70) gives

$$E_\theta = \frac{jI_o\eta e^{-j\beta r}}{\sqrt{2}\pi r}\ \cos\theta \Biggl\{ \frac{\sin\phi\ \sin[(\pi/4)\ \sin\theta\ \cos\phi]}{1-\sin^2\theta\ \sin^2\phi}$$

$$\cdot \left[\sin\theta\ \sin\phi\ \cos\left(\frac{\pi}{4}\sin\theta\ \sin\phi\right) - \sin\left(\frac{\pi}{4}\sin\theta\ \sin\phi\right) \right]$$

$$- \frac{\cos\phi\ \cos[(\pi/4)\ \sin\theta\ \sin\phi]}{1-\sin^2\theta\ \cos^2\phi}$$

$$\cdot \left[\sin\theta\ \cos\phi\ \sin\left(\frac{\pi}{4}\sin\theta\ \cos\phi\right) - \cos\left(\frac{\pi}{4}\ \sin\theta\ \cos\phi\right) \right] \Biggr\} \qquad (6\text{-}73a)$$

$$E_\phi = \frac{jI_o\eta e^{-j\beta r}}{\sqrt{2}\pi r}\ \Biggl\{ \frac{\cos\phi\ \sin[(\pi/4)\ \sin\theta\ \cos\phi]}{1-\sin^2\theta\ \sin^2\phi}$$

$$\cdot \left[\sin\theta\ \sin\phi\ \cos\left(\frac{\pi}{4}\sin\theta\ \sin\phi\right) - \sin\left(\frac{\pi}{4}\sin\theta\ \sin\phi\right) \right]$$

$$+ \frac{\sin\phi\ \cos[(\pi/4)\ \sin\theta\ \sin\phi]}{1-\sin^2\theta\ \cos^2\phi}$$

$$\cdot \left[\sin\theta\ \cos\phi\ \sin\left(\frac{\pi}{4}\sin\theta\ \cos\phi\right) - \cos\left(\frac{\pi}{4}\ \sin\theta\ \cos\phi\right) \right] \Biggr\} \qquad (6\text{-}73b)$$

These expressions are rather involved but were derived in a straightforward fashion using the principles set forth in Sec. 2.4.

The far-field expressions simplify somewhat in the principal planes. In the *xy*-plane, which is the plane of the loop (an *E*-plane), $\theta = 90°$ and then (6-73) reduces to

$$E_\theta\left(\theta = \frac{\pi}{2}\right) = 0 \qquad (6\text{-}74a)$$

$$E_\phi\left(\theta = \frac{\pi}{2}\right) = \frac{jI_o\eta e^{-j\beta r}}{\sqrt{2}\pi r}\frac{\pi}{4}\Biggl\{ \frac{\sin[(\pi/4)\ \cos\phi]}{(\pi/4)\ \cos\phi}\left[\sin\phi\ \cos\left(\frac{\pi}{4}\sin\phi\right) - \sin\left(\frac{\pi}{4}\sin\phi\right)\right]$$

$$+ \frac{\cos[(\pi/4)\ \sin\phi]}{(\pi/4)\ \sin\phi}\left[\cos\phi\ \sin\left(\frac{\pi}{4}\cos\phi\right) - \cos\left(\frac{\pi}{4}\ \cos\phi\right)\right] \Biggr\} \qquad (6\text{-}74b)$$

The E_ϕ expression is plotted in Fig. 6-67a (solid curve) in normalized form. Along the *x*-axis ($\phi = 0°$ and $180°$), $E_\phi = 0$. This is true because the sides 3 and 4 alone each have patterns that are zero in the broadside direction since the current distributions on these sides are odd about their midpoints. Along the *y*-axis, (6-74b) reduces to

(*a*) The *xy*-plane (the plane of the loop and an *E*-plane) pattern plot of E_ϕ. In this plane, HP = 99°.

(*b*) The *xz*-plane (an *E*-plane) normalized pattern plot of E_θ. In this plane, HP = 85°. The patterns from the two methods coincide in this plane.

(*c*) The *yz*-plane (the *H*-plane) pattern plot of E_ϕ.
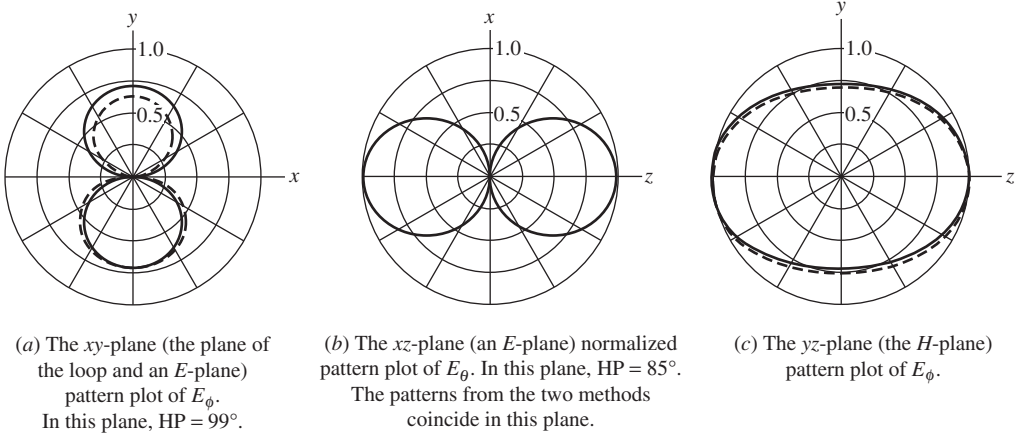
Figure 6-67 Principal plane patterns for one-wavelength square loop antenna. The solid curves are the patterns based on a sinusoidal current distribution of Fig. 6-66. The dashed curves are the patterns arising from the current distribution obtained by the more exact numerical methods.

$$E_\phi\left(\theta = \frac{\pi}{2},\ \phi = \frac{\pi}{2}\right) = \frac{jI_o\eta e^{-j\beta r}}{\sqrt{2}\pi r}\frac{1}{\sqrt{2}} \tag{6-75}$$

In the *xz*-plane, which is an *E*-plane, (6-73) yields

$$E_\phi(\phi = 0) = 0 \tag{6-76a}$$

$$E_\theta(\phi = 0) = \frac{jI_o\eta e^{-j\beta r}}{\sqrt{2}\pi r}\frac{\sin\theta\ \sin[(\pi/4)\ \sin\theta] - \cos[(\pi/4)\ \sin\theta]}{\cos\theta} \tag{6-76b}$$

The normalized form of this E_θ expression is plotted in Fig. 6-67*b* (solid curve). It can be shown that (6-76*b*) goes to zero for $\theta = 90°$, as it should by (6-74*a*).

In the *yz*-plane, which is the *H*-plane, (6-73*a*) reduces to

$$E_\theta\left(\phi = \frac{\pi}{2}\right) = 0 \tag{6-77a}$$

$$E_\phi\left(\phi = \frac{\pi}{2}\right) = -\frac{jI_o\eta e^{-j\beta r}}{\sqrt{2}\pi r}\cos\left(\frac{\pi}{4}\sin\theta\right) \tag{6-77b}$$

Fig. 6-67*c* (solid curve) gives the plot of the normalized form of this E_ϕ expression. The $\cos[(\pi/4)\ \sin\theta]$ pattern is the array factor for two point sources at the midpoints of sides 1 and 2. Note that in the *z*-direction, (6-76) and (6-77) give the same result (for $\theta = 0°$): an electric field parallel to the *x*-axis given by

$$E_x = -\frac{jI_o\eta e^{-j\beta r}}{\sqrt{2}\pi r} \tag{6-78}$$

which is a factor of $\sqrt{2}$ greater than E_x in the *y*-direction given in (6-75). This can be seen from the pattern in Fig. 6-67*c* where the value is 0.707 in the *y*-direction wand is unity in the *z*-direction. Fig. 6-67*a* also shows this result because the *xy*-plane pattern is shown relative to the peak of the entire 3D pattern, which is in the *z*-direction.

From the patterns in Fig. 6-67, we can make some general conclusions about the radiation properties of the one-wavelength square loop antenna. Radiation is maximum normal to the plane of the loop (along the z-axis) and in that direction is polarized parallel to the loop side containing the feed (x-direction). In the plane of the loop, there is a null in the direction parallel to the side containing the feed point (x-direction) and a lobe peak perpendicular to the null direction (y-direction). These results are quite different from the small loop, which has a null perpendicular to the plane of the loop, and maximum radiation in the plane of the loop where the pattern is omnidirectional.

We now use the square loop to examine the accuracy of the sinusoidal current distribution assumption by comparing to moment method simulation results, which solve for the actual current; see Chap. 14 and App. G. The current distribution obtained by simulation for the square loop of wire radius of 0.001λ, shown in Fig. 6-66 as the dashed curve, agrees closely with the assumed sinusoidal current (solid curve). The effects of the slight differences between the currents on the radiation patterns are revealed in Fig. 6-67. The patterns calculated from the pattern expressions developed in this section (solid curves) agree closely with those obtained through simulations (dashed curves). Simulations are capable of including the asymmetry introduced by the feed, which causes the simulated current to be asymmetric and, as a consequence, the pattern to be asymmetric in the xy-plane. The half-power beamwidths in this plane are 99° and 103° based on simple theory and simulation, respectively. The calculated and simulated patterns in the xz-plane are in full agreement and have a beamwidth of 85°; for reference, a half-wave dipole has a beamwidth of 78°. This detailed comparison of numerical to approximate analytical results demonstrates that simple theory using an assumed current, as employed thus far in this book, provides good engineering results.

The impedance of the square loop antenna calculated using moment method numerical techniques is plotted as a function of perimeter size for a wire radius of 0.001λ in Fig. 6-68. As would be expected for a loop antenna, the reactance is inductive below a perimeter of about 0.4λ after which it becomes capacitive and alternates between the two for larger perimeters. This behavior is similar to the impedance for a dipole in Fig. 6-6 except the dipole begins as capacitive for short lengths. Also note in comparing the loop to the dipole that for increasing loop perimeter anti-resonance (large reactance) is encountered first, followed by resonance (zero reactance), whereas the dipole has the reverse sequence. Input impedance values calculated based on the constant current assumption used in Sec. 3.4.2 lose accuracy as the perimeter becomes about 0.1λ or greater, indicating that the uniform current assumption is no longer valid and large loop theory is required; Prob. 6.8-3 addresses this in detail. Storer [28] found the same result for circular loops. Fig. 6-68 indicates the general conclusion that large loop antennas exhibit reasonable input impedance values for perimeters greater than about 0.7λ.

For a one-wavelength perimeter, the input impedance is $102 - j142\,\Omega$ and the gain[2] is 3.1 dB. Resonance occurs at about $L = 1.1\lambda$ where the impedance is $133\,\Omega$ (real), as can be seen in Fig. 6-68. The gain of the resonated one-wavelength square loop ($L = 1.1\lambda$) with $a = 0.001\lambda$ is 3.41dB, which is 1.26 dB higher than that of a resonated half-wave dipole ($3.41 - 2.15$ dB). The peak gain for the large square loop is 4.01 dB at $L = 1.4\lambda$, but the impedance is $657 + j669\,\Omega$. The best design approach for most applications is to size the loop for resonance at the middle of the operating frequency band to facilitate impedance matching over as broad of bandwidth as possible. Simulations for circular loops using the same wire radius of 0.001λ give similar results; see Prob. 6.8-5. For a one-wavelength circumference circular loop, the impedance is $121 - j95\,\Omega$ and the gain is 3.45 dB. Resonance occurs for a circumference of 1.06λ where the impedance is $142\,\Omega$ (real) with a gain of 3.67 dB, which are close to the square loop values. The peak gain of a

---

[2]Although lossless wire is assumed in the simulations, the loss is very low and gain is assumed to equal directivity as with any electrically large antenna.

(*a*) Input resistance.   (*b*) Input reactance.
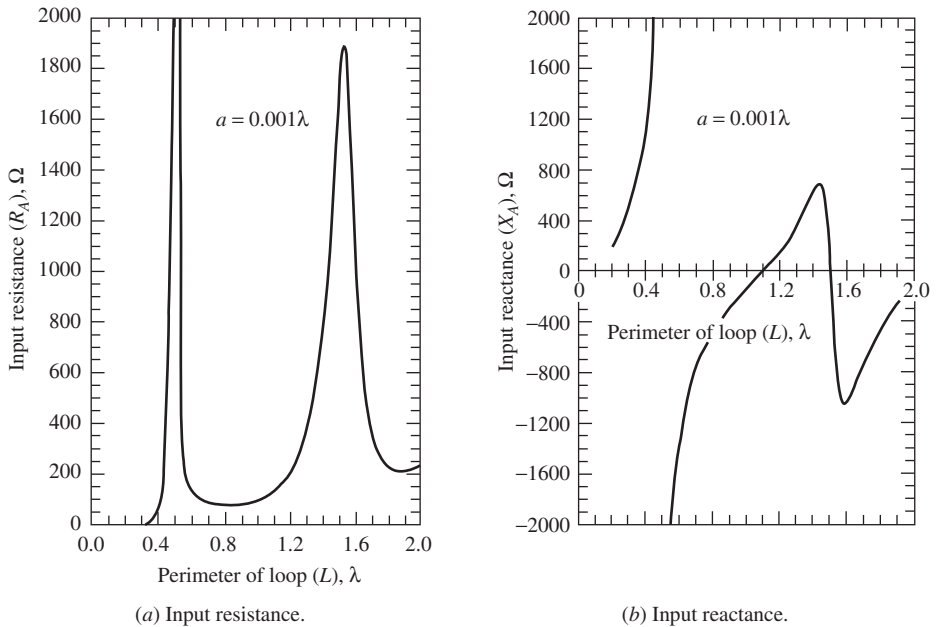
Figure 6-68 Input impedance of a square loop antenna as a function of the loop perimeter in wavelengths. The loop is fed in the center of one side and has a wire radius of $a = 0.001\,\lambda$. Numerical calculation methods were used.

large circular loop is 4.46 dB at a circumference of 1.42λ. We can conclude that the circular loop has a few tenths of a dB more gain than a similar-sized square loop. If thicker wire is used, the gain increases. Also, resonance occurs at a larger perimeter value and the input resistance is larger for thicker wire.

REFERENCES

1. R. W. P. King, "The Linear Antenna—Eighty Years of Progress," *Proc. IEEE*, Vol. 55, pp. 2–16, Jan. 1967.
2. M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, NBS Applied Math. Series 55, U.S. Gov. Printing Office, Washington, D.C., 1964, Chap. 5.
3. R. F. Schwartz, "Input Impedance of a Dipole or Monopole," *Microwave J.*, Vol. 15, p. 22, Dec. 1972.
4. "IEEE Standard Defintions of Terms for Antennas," *IEEE Standard* 145–1993, 1993.
5. C. M. Butler, "The Equivalent Radius of a Narrow Conducting Strip," *IEEE Trans. Ant. & Prop.*, Vol. 30, pp. 755–758, July 1982.
6. G. A. Thiele, "Analysis of Yagi-Uda Type Antennas," *IEEE Trans. on Ant. & Prop.*, Vol. AP-17, pp. 24–31, Jan. 1969.
7. H. Yagi, "Beam Transmission of Ultra-Short Waves," *Proc. IRE*, Vol 16, p. 715, 1928.
8. S. Uda and Y. Mushiake, *Yagi–Uda Antenna*, Saski Printing and Publishing Company, Ltd., Sendai, Japan, 1954.
9. H. W. Ehrenspéck and H. Poehler, "A New Method for Obtaining Maximum Gain for Yagi Antennas," *IRE Trans. on Ant. & Prop.*, Vol. AP-7, pp. 379–386, Oct. 1959.
10. H. E. Green, "Design Data for Short and Medium Length Yagi–Uda Arrays," *Institution of Engineers* (*Australia*), *Elec. Eng. Trans.*, pp. 1–8, March 1966.
11. P. Viezbicke, *"Yagi-Antenna Design," NBS Technical Note 688*, U.S. Government Printing Office, Washington, D.C., Dec. 1976.
12. J. L. Lawson, *Yagi Antenna Design*, American Radio Relay League, Newington, CT, 1986.
13. J. H. Reisert, "How to Design Yagi Antennas," *Ham Radio Magazine*, pp. 22–31, Aug. 1977.
14. U. Inan and A. Inan, *Engineering Electromagnetics*, Addison-Wesley, p. 758, 1999.

15. G. A. Thiele, E. P. Ekleman, Jr., and L. W. Henderson, "On the Accuracy of the Transmission Line Model for Folded Dipole," *IEEE Trans. on Ant. & Prop.*, Vol. AP-28, pp. 700–703, Sept. 1980.
16. J. M. Gonzalez-Arbesu, S. Blanch, and J. Romeu, "Are Space-Filling Curves Efficient Small Antennas?" *IEEE Ant. & Wireless Prop. Letters*, Vol. 2, pp. 147–150, 2003.
17. W. W. Mumford, "Some Technical Aspects of Microwave Radiation Hazards," *Proc. IRE*, Vol. 49, pp. 427–447, Feb. 1961.
18. L. R. Dod, "Experimental Measurements of the Short-Backfire Antenna," NASA Tech. Report No. X-525-66-480, Oct. 1966.
19. G. S. Kirov, "Design of Short Backfire Antennas," *IEEE Ant. & Prop. Mag.*, Vol. 51, pp. 110–120, Dec. 2009.
20. S. Ohmori, et al., "An Improvement in Electrical Characteristics of a Short, Backfire Antenna," *IEEE Trans. on Ant. & Prop.*, Vol. 31, pp. 644–646, July 1983.
21. H. W. Ehrenspeck, "A New Class of Medium-Size High-Efficiency Reflector Antennas,"*IEEE Trans. on Ant. & Prop.*, Vol. AP-22, pp. 329–332, March 1974.
22. K. M. Chen, D. P. Nyquist, and J. L. Lin, "Radiation Fields of the Short Backfire Antenna," *IEEE Trans. on Ant. & Prop.*, Vol. AP-16, pp. 596–597, Sept. 1968.
23. H. W. Ehrenspeck, "The Short Backfire Antenna," *Proc. IEEE*, Vol 53, pp. 1138–1140, Aug. 1965.
24. E. K. Miller et al., "Analysis of Wire Antennas in the Presence of a Conducting Half-Space. Part I. The Vertical Antenna in Free Space," *Canadian J Phys.,* Vol. 50, pp. 879–888, 1972.
25. E. K. Miller et al., "Analysis of Wire Antennas in the Presence of a Conducting Half-Space. Part II. The Horizontal Antenna in Free Space," *Canadian J. Phys.,* Vol. 50, pp. 2614–2627, 1972.
26. B. W. Griffith, *Radio-Electronic Transmission Fundamentals*, McGraw-Hill, New York, 1962, Chap. 43.
27. L. A. Dorado and V. Trainoti, "Accurate Evaluation of Magnetic- and Electric Field Losses in Ground Systems," *IEEE Trans. on Ant. & Prop. Mag.*, Vol. 49, pp. 58–70, Dec. 2007.
28. J. E. Storer, "Impedance of Thin-Wire Loop Antennas," *AIEE Trans.*, Part 1, Vol. 75, pp. 606–619, Nov. 1956.
29. D. H. Werner, "An Exact Integration Procedure for Vector Potentials of Thin Circular Loop Antennas," *IEEE Trans. on Ant. & Prop.*, Vol. 44, pp. 157–165, Feb. 1996.
30. J. D. Mahony, "Approximations to the Radiation Resistance and Directivity of Circular-Loop Antennas," *IEEE Trans. on Ant. & Prop. Mag.*, Vol. 36, pp. 52–55, Aug. 1994.
31. Phillis Kennedy, "Loop Antenna Measurements," *IRE Trans. on Ant. & Prop.*, Vol. 4, pp. 610–618, Oct. 1956.

PROBLEMS

6.1-1 Use the integral from (F-11) in (6-3) to prove (6-4).

6.1-2 Starting with (6-6), show that for $L \ll \lambda$, the radiation pattern of a dipole reduces to that of a short dipole, $\sin\theta$.

6.1-3 (a) The outputs from two collinear, closely spaced, half-wave dipoles are added together as indicated by a summing device in the figure below. The transmission lines from the antennas to the summer are of equal length. Write the pattern $F_a(\theta)$ of this antenna system using array techniques.
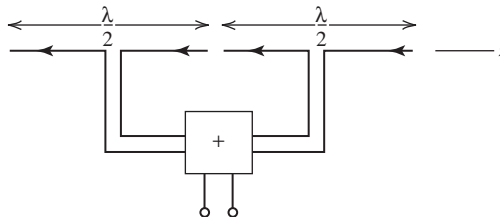
(b) Now consider a center-fed, full-wave dipole that is along the z-axis. Write its pattern expression $F_b(\theta)$.

(c) Now draw the current distributions $I_a(z)$ and $I_b(z)$ for both antennas. From these current distributions, can you make a statement about the patterns from the two antennas? Return to the pattern expressions and prove your statement mathematically.

6.1-4 The center-fed, full-wave dipole is rarely used because it has a current minimum at the feed point. If it is instead fed as shown below, sketch the current distribution. Also, rough sketch how you think the pattern should look, and explain how you obtained it.
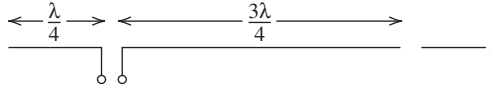
6.1-5 (a) The array of Prob. 6.1-3a is parallel fed, in-phase array. Show how the parallel wire transmission lines are connected to perform the summing function. Also put current arrows on each wire.

(b) Consider an array similar to that of part (a) except now the array elements (half-wave dipoles) are fed $180°$ out-of-phase. Show how the transmission lines are arranged to accomplish this subtraction function. Again, show the currents on each wire.

6.1-6 Use the results of the cosine-tapered current distribution in Sec. 5.2 to derive the pattern of a half-wave dipole in (6-7).

6.1-7 Verify that the normalization constant in (6-9) is 0.7148 for the pattern of a $3\lambda/2$ dipole. What are the angles θ_o in degrees for maximum radiation?

6.1-8 A resonant half-wave dipole is to be made for receiving TV Channel 7 of frequency 177 MHz. If $\frac{1}{2}$-in.-diameter tubular aluminum is used, how long (in centimeters) should the antenna be?

6.1-9 Use (6-1) and (6-4) to calculate and plot the current distribution and far-field pattern for dipoles of length 2.0 and 2.5λ. Compare with Figs. 6-3 and 6-4.

6.1-10 (a) Show that the power radiated by a center-fed dipole of arbitrary length L with a sinusoidal current is

$$ P = \frac{\eta I_m^2}{4\pi} \left\{ 0.5772 + \ln(\beta L) - \text{Ci}(\beta L) + \frac{1}{2} \ \sin(\beta L)[\text{Si}(2\beta L) - 2 \ \text{Si}(\beta L)] \right. $$

$$ \left. + \frac{1}{2}\cos(\beta L)\left[0.5772 + \ln\left(\frac{\beta L}{2}\right) + \text{Ci}(2\beta L) - 2 \ \text{Ci}(\beta L)\right] \right\} $$

(b) Derive an expression for the directivity and then plot directivity as a function of dipole length for L from 0 to 3λ.

6.1-11 Use the length reduction procedure for half-wave resonance in Table 6-2 to calculate the resonant frequencies of the two dipoles in Fig. 6-7.

6.1-12 Design an optimum directivity vee dipole to have a directivity of 6 dB.

6.1-13 To show that the vee dipole results of (6-23) and (6-24) give roughly the correct results for a full-wave straight wire dipole, use $D = 2.41$ and determine γ.

6.1-14 A formula for the shortening factor of a nominal half-wave dipole is

$$ L_{\text{res}} = 0.5\left(1 - \frac{0.2257}{\ln\left(2\frac{L}{2a}\right) - 1}\right) $$ where L_{res} is the resonant length of a half-wave dipole of radius

"a" and $L = \lambda/2$. Calculate values for the resonant length of the dipoles in Table 6-2 and compare with Table 6-2. Optional: use a moment method code to find the resonant lengths and compare with Table 6-2 and the above formula.

6.2-1 (a) It is desired to have a simple formula for the length of a thin-wire half-wave folded dipole antenna. Show that it is $L(\text{cm}) = 14{,}250/f(\text{MHz})$.

(b) Determine the length in centimeters of half-wave folded dipoles for practical application as receiving antennas for each VHF TV channel and the FM broadcast band (100 MHz). Tabulate results.

6.2-2 Calculate the input impedance of a folded dipole of length $2a = L = 0.4\lambda$, wire size 0.001λ, and wire spacing $d = 12.5a$ using the transmission line model. Compare your results to values from Fig. 6-16.

6.2-3 The equation for the impedance on a transmission line a distance l from a load, Z_L, is given by

$$Z(l) = Z_o \frac{Z_L \cos \beta l + j Z_o \sin \beta l}{Z_o \cos \beta l + j Z_L \sin \beta l}$$

where Z_o is the characteristic impedance of the transmission line. Show that this reduces to (6-25) when the load is a short circuit at $l = \frac{L}{2}$.

6.3-1 Use array theory from Sec. 3.5 to analyze the array of Fig. 6-18a.

(a) Plot the E-plane pattern and compare to Fig. 6-18b
(b) Plot the H-plane pattern.
(c) Use (6-7) for the dipole (element) pattern.

6.3-2 Numerical methods reveal that the currents on the elements of the two-element parasitic array of Fig. 6-18 are nearly sinusoidal and the current amplitudes and phases at each element center are $1.0 \angle -88°$ for the driver and $0.994 \angle 81.1°$ for the parasite. Use simple array theory to obtain and plot the H-plane pattern in linear, polar form.

6.3-3 Phasor diagrams are often helpful in obtaining a rough idea about how arrays perform. To illustrate, use phasor diagrams to obtain the relative far-zone field values in the endfire directions of the two-element parasitic array of Prob. 6.3-2 (i.e., find the front-to-back ratio). To do this, find the total phasor at each element location including the spatial phase delay due to the element separation. Assume the amplitudes of each element are unity and the phases are $-88°$ for the driver and $81.1°$ for the parasite.

6.3-4 Design a 10.2-dBd gain Yagi for operation at 50.1 MHz. The parasitic elements are insulated from the metal boom. The diameter of the elements is 0.0021λ.

6.3-5 Design a 14.2-dBd gain Yagi for operation at 432 MHz. The parasitic elements are mounted through a metal boom making electrical contact with it. The diameter of the elements is 0.00343λ. The boom diameter is 0.0275λ.

6.3-6 *Construction project—a 10¢ Yagi.* This project is designed to demonstrate how a high gain antenna can be built for under 10¢! Locate a channel on your (or a cooperating friend's) TV receiver that has marginal reception, such as a snowy picture when a modest antenna (like rabbit ears) is used. The construction phase proceeds as follows. Find a large rigid piece of corrugated cardboard and trim it so that it is several centimeters longer than the total array length and about 5 cm narrower than the director length. Now locate several thick coat hangers. Straighten them as much as possible and cut them to the lengths required for the reflector and directors. The feed element is a folded dipole constructed from a piece of twin-lead transmission line. Cut it to a length that is a little longer than the driver dimension. Strip the ends and solder the two wires at each end together such that the overall length is equal to the driver dimension. Next cut one wire of the driver at the center of the folded dipole and solder the ends to a long piece of twin-lead that serves as a transmission line for the antenna. Lay out all element positions on the cardboard by marking appropriately. Tape the folded dipole onto the cardboard at the driver location. The coat hanger parasitic elements are positioned by merely inserting them into the corrugations in the cardboard. Now connect the transmission line to the receiver. Rotate the antenna and note the effect on the reception. Large performance differences should be observed. Note that it may be necessary to elevate the antenna by placing it in the attic, for example. With this construction, it is very easy to change the element spacings by placing the coat hanger elements into difference corrugations. Very little difference will be observed for small distance changes. Normally, the best performance is achieved for horizontal polarization, that is, elements parallel to the ground.

6.3-7 *Construction project—a slightly more expensive Yagi.* A fairly rugged Yagi antenna can be constructed using the following technique. Select a TV channel with marginal reception and design a Yagi for that frequency. The materials required for this project are a 1×2 in. board of length slightly greater than the overall length of the array and a few meters of aluminum wire (usually No. 8 AWG). Trim and straighten wires for the reflector and directors. Drill holes in the wooden mast at the appropriate positions for the reflector and the directors. The holes should be just slightly greater than the wire diameter. Be sure all holes are along a straight line. The driver is a folded dipole oriented such that the plane of the dipole is perpendicular to the line of the array. Drill one hole in the mast about 2 cm above the array line. At the same

distance below the array line, drill in from each side of the mast about 0.5 cm. Cut a piece of wire more than twice the length of the driver. Push it through the top hole and center it. Bend the wire at the required length at each end and fold it back to the mast. Now carefully trim away any excess wire such that the wire ends can just be forced into the shallow holes and still form a symmetric folded dipole. Now wrap the bared ends of a twin-lead transmission line to the ends of the folded dipole close to the mast (at the feed point). Be sure to get a good mechanical contact. Also leave a tab of polyethylene where you stripped the twin lead. Small wire brads can be wedged between the wire ends at the feed point and the mast, and at the same time pinch the twin-lead connection between the antenna wire and the brad. Solder the feed point connections. Tack the polyethylene tab to the bottom of the mast to provide strain relief. Insert the remaining elements into their holes, center them, and nail brads into the hole alongside the wires to secure their positions. The construction is now complete and you can connect the transmission line to the receiver and test the reception. Try several antenna locations and orientations.

6.3-8 A two-element Yagi has a current on the driven element of $1 \angle 164°$ and a current on the parasitic element of $0.5 \angle 238°$. The spacing between the elements is 0.2λ. Does the parasitic element act like a director or reflector element? Use a phasor diagram to show why.

6.3-9 A two-element Yagi with 0.2λ element spacing has a current on the driven element of $1 \angle 254°$ and a current on the parasitic element of $0.6 \angle -32°$. Does the parasitic element act like a director or reflector element? Why?

6.4-1 A *receiving* antenna with a real impedance R_L attached to its terminals has the equivalent circuit shown. Prove that maximum power transfer to the load for a fixed real antenna impedance R_A occurs for $R_L = R_A$.
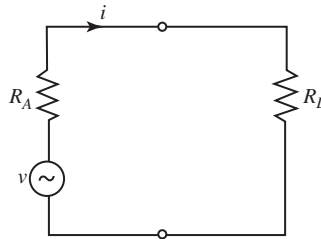
6.4-2 A transmitter with a real impedance of R_t is connected to a lossless transmission line of real characteristic impedance R_o and then to an antenna of real input impedance R_o.
  (a) Derive an expression for the transmit efficiency, that is, power delivered to antenna/total power dissipated. Neglect any mismatch effects.
  (b) Find the percent efficiency for $R_t = R_o$, $R_t = 0.5R_o$, and $R_t = 0.1R_o$.

6.4-3 Using the expression for $Z(l)$ in Prob. 6.2-3 and referring to Fig. 6-30a, what is the impedance $Z_A(l)$ if the antenna impedance is $40 - j20$ ohms and $l = 42.3\lambda$?

6.5-1 A monopole is embedded in a large dielectric medium with relative permittivity of 3.4. Determine the length in free-space wavelengths of the quarter-wave monopole if the wire radius is classified as thin.

6.6-1 A wire-grid is to pass through not more than 2.5% of the incident power. The spacing between wires in the grid is to be 0.12λ. What wire radius is required?

6.6-2 A wire-grid is to have a reflection loss of 0.5 dB. The spacing to wire radius ratio is 20. What must the spacing be in wavelengths?

6.6-3 A wire-grid is to have a reflection loss of 1 dB. What percentage of the incident power will be reflected?

6.6-4 A wire-grid is to be used to provide a measure of shielding. Not more than 1% of the incident power may be transmitted through a grid with 0.1λ spacing. What wire radius is required?

6.6-5 (a) Calculate and plot the magnitude of the array factor in (6-54) and verify that the maximum value of $|AF(\theta = 90°, \phi = 0°)|$ obtained when $s = 0.5\lambda$.
(b) Reason that this must be the case, using Fig. 6-54.

6.7-1 Derive an expression for the directivity of an ideal (infinitesimal) dipole as a function of its height h above a perfect ground plane. The dipole is oriented perpendicular to the ground plane. Make use of the results in Secs. 3.3 and 8.5.

6.7-2 A short dipole is a quarter-wavelength above a perfect ground plane. Use simple array theory for the dipole and its image to obtain polar plot sketches of the E- and H-plane patterns when the dipole is oriented (a) vertically and (b) horizontally.

6.7-3 Repeat Prob. 6.7-2 for a short dipole a half-wavelength above a perfect ground plane.

6.7-4 A horizontal short dipole is a quarter-wavelength above a planar real Earth and is operating at 1 MHz. The conductivity of the Earth is $\sigma = 12 \times 10^{-3}$ S/m and the relative dielectric constant is $\varepsilon_r = 15$. For this frequency, σ, and ε_r, we can approximate $|\Gamma_H|$ by 0.9 and the phase of Γ_H by $-190°$ for all θ.
   (a) Calculate and plot the H-plane elevation pattern in polar form in the upper half-space.
   (b) Compare the pattern with that of the short dipole over a perfectly conducting ground plane (i.e., the results of Prob. 6.7-2b.

6.7-5 A quarter-wave resonant monopole is to be used as a transmitting antenna at 1 MHz. A radial system of 120 radials is to be used. If 97% efficiency is to be achieved, how long must the radial wires be? Neglect any tower ohmic resistance.

6.8-1 *The uniform circular loop antenna.* A circular loop in the xy-plane with its center at the origin and a radius b carries a uniform amplitude, uniform phase current given by

$$\mathbf{I} = I_o \hat{\boldsymbol{\phi}}'$$

   (a) Due to symmetry, the pattern will not be a function of ϕ and $\mathbf{A}$ will have only a ϕ-component. Using these facts, show that

$$\mathbf{A} = \hat{\boldsymbol{\phi}}\, \mu \frac{e^{-j\beta r}}{4\pi r} I_o b \int_0^{2\pi} \cos\phi'\, e^{j\beta b \sin\theta \cos\phi'}\, d\phi'$$

   in the far field. Use symmetry to reason that $\phi = 0$ can be assumed and only a ϕ-component exists.
   (b) Find an expression for E_ϕ. *Hint:* Use (F-7).
   (c) Show that this result reduces to that for a small loop antenna in (3-49).
   Hint: $J_1(x) \approx x/2$ for $x \gg 1$.

6.8-2 Show that (6-69) yields (6-70). To perform the integrations, decompose the functions $\cos(\beta x')$ and $\sin(\beta y')$ into sums of exponential functions using (E-6) and (E-7).

6.8-3 *Simulation* This problem applies small loop theory to moderately large square loops to investigate where small loop theory becomes inaccurate. Calculate the input impedance (radiation resistance and input reactance) based small loop theory for perimeter values 0.01, 0.05, 0.1, 0.15, and 0.2λ and compare to values using a simulation code. Use a wire radius of 0.001λ and a frequency of 300 MHz. Comment on the results.

6.8-4 *Simulation* Use a simulation code to evaluate a square loop antenna with a 1.1λ perimeter and wire radius of 0.001λ.
   (a) Compute the input impedance and compare to Fig. 6-68.
   (b) Plot the radiation patterns and comment closeness to Fig. 6-67 patterns.
   (c) Compute the gain.

6.8-5 *Simulation* This problem compares circular and square large loop antennas using a simulation code.
   (a) For the geometry of Fig. 6-66 use a simulation code to evaluate gain and input impedance of a square loop antenna for $L = 1, 1.5$ and $2\ \lambda$ and a wire radius of 0.001λ. Compare the impedance values to those in Fig. 6-68.
   (b) Repeat (a) for circular loops with circumferences equal to the perimeter values of the square loops in (a). Use the same wire radius.
   (c) Tabulate performance values for both antennas. Comment on similarities.

6.8-6 In a communication link a half-wavelength folded dipole perpendicular to the z-axis is used to transmit and a square, one-wavelength loop as in Fig. 6-66 is used to receive. Sketch a perspective view of the antennas in a common coordinate system for maximum power transfer; include the orientations of the antennas and their feed points.

6.8-7 *Design of a* TV *antenna.* Design a square loop antenna to receive over-the-air TV Channel 37. It is to be constructed of 2-mm diameter wire. List the values of input impedance and gain. *Hint*: Use a resonant loop and a simulation code.

6.8-8 *Project.* Research the Alford loop antenna in the literature and prepare a report that includes background information with a discussion of applications, operating principles, calculated performance using a simulation code for an example practical realization, and comments on the results.

Chapter 7

Broadband Antennas

7.1 INTRODUCTION

Many system applications require an antenna to operate over a large range of frequencies. An antenna with wide bandwidth is referred to as a **broadband antenna**, also often called a *wideband antenna*. Bandwidth is defined both as percentage of the center frequency and as a ratio. First, let f_U and f_L be the upper and lower frequencies, respectively, for which satisfactory performance is obtained. The antenna operating band spans the frequencies from f_L to f_U and the center frequency is $f_C = (f_U + f_L)/2$, which is the arithmetic mean of the limit frequencies. The *antenna bandwidth* is defined as BW $= f_U - f_L$ and carries units of frequency. Antenna bandwidth is, however, usually expressed in relative terms rather than in absolute frequency, either as a percent or a ratio. *Fractional bandwith*, B, is defined as BW/f_C and is often expressed as *percent bandwidth,* which is defined as

$$B_p = B \times 100 = \frac{\text{BW}}{f_C} \times 100 = 2\frac{f_U - f_L}{f_U + f_L} \times 100 \quad [\%] \tag{7-1}$$

Ratio bandwidth is defined as a ratio of the upper to lower frequencies:

$$B_r = \frac{f_U}{f_L} \tag{7-2}$$

Percentage bandwidth and ratio bandwidth are illustrated in Fig. 7-1.

The conversion between percent bandwidth to ratio bandwidth is found using:

$$B_r = \frac{200 + B_p}{200 - B_p} \qquad B_p = 2\frac{B_r - 1}{B_r + 1} \times 100 \quad [\%] \tag{7-3}$$

For examples, 50% bandwidth corresponds to $B_r = 1.67$; 67% to $B_r = 2$ (an "octave"); 100% to $B_r = 3$; and 164% to $B_r = 10$ (a "decade"). Other center frequencies that are used in the definition of percent bandwidth are design frequency, resonant frequency, lowest VSWR frequency, and geometric mean frequency of $f_C = \sqrt{f_L f_U}$. Geometric mean center frequency is popular when frequency is considered on a logarithmic scale, including when fundamental limits on antenna size are being examined (see Sec. 11.5). B_p values using arithmetic and geometric mean differ by only a small amount until B_r reaches 1.5, where the difference is *2 %* (see Prob. 7.1-1). Bandwidth values are quoted without noting whether percent or ratio is being used. Percent bandwidth is stated with the unit % and ratio bandwidth is often expressed a ratio format of $B_r : 1$. The bandwidth of narrowband antennas is usually expressed in percent bandwidth, whereas broadband antennas are usually quoted in ratio bandwidth.
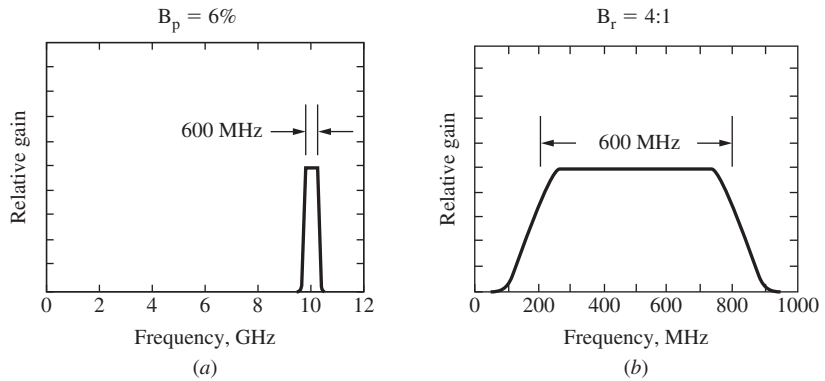
Figure 7-1 Relative gain vs. frequency. (*a*) Percentage bandwidth example. (*b*) Ratio bandwidth example.

Antenna performance is evaluated compared to desired performance values of impedance, pattern, gain, or polarization. For example, suppose the antenna gain specification is 12 dB (minimum) and the selected antenna has 14-dB gain at the center frequency. Thus, in this case the operating band is determined by the frequencies where the gain falls by 2 dB. Electrically small antennas are almost always limited by adequate impedance match. Impedance bandwidth is often stated such that for all frequencies from f_L to f_U the VSWR is less than or equal to 2. Resonant antennas are also limited by impedance match to the connecting transmission line. For example, the half-wave dipoles in Fig. 6-7 have bandwidths of 8% and 16% for VSWR ≤ 2. The performance quantity that is responsible for limiting bandwidth varies with the antenna. Circularly polarized antennas are often limited by the specification on axial ratio. With no qualifier on bandwidth, it is assumed that all performance quantities meet the stated specifications over the operating band.

Most antennas have a single continuous operating band spanning f_L to f_U. However, antennas with multiple, separated operating bands (called a *multiband antenna*) are becoming more common, especially in wireless communications. Each operating band of a multiband antenna has an associated bandwidth. System aspects are also important in the antenna selection process. When an antenna must handle all frequencies in the operating band simultaneously, the bandwidth is called *instantaneous bandwidth*. Systems that employ impulses in time for sensing (such as radar) or communication often require wide instantaneous bandwidth. Impulse ultra-wideband systems require wide instantaneous bandwidth antennas, as will be discussed in Sec. 7.10. Many applications do not require that the antenna use all of its bandwidth at one time. The most common such situation is an antenna with acceptable radiation properties over a wide bandwidth but with an input impedance that is acceptable only over a narrow band. This is solved by tuning (adjusting) an external matching network, or sometimes by adjusting internal portions of the antenna, for proper impedance match over a sub-band of frequencies within the whole bandwidth, called the *tunable bandwidth*.

The term "broadband" in communications is a measure of bandwidth that depends on the application, but we are specific in defining a broadband antenna as follows: *A broadband antenna is one with performance quantities of interest that do not change significantly over about an octave of bandwidth* $(B_r = 2)$ *or more.* Electrically small and resonant antennas rarely qualify as broadband, whereas aperture antennas, to be discussed in Chap. 9, are capable of broadband performance. This chapter focuses on antennas that are specifically selected because of their broadband properties.

As we will see in this chapter, broadband antennas usually require structures that do not have abrupt changes in physical dimensions but instead utilize materials with smooth

geometries and emphasize angles. Smooth physical structures tend to produce patterns and input impedances that also change smoothly with frequency. This simple concept guides the design of broadband antennas.

7.2   TRAVELING-WAVE WIRE ANTENNAS

The wire antennas we have discussed thus far have been resonant structures. The wave traveling outward from the feed point to the end of the wire is reflected, setting up a standing-wave-type current distribution. This can be seen by examining the expression for the current in (6-1) for the top half of the dipole that can be written as

$$I_m \sin\left[\beta\left(\frac{L}{2} - z\right)\right] = \frac{I_m}{2j} e^{j(\beta L/2)}(e^{-j\beta z} - e^{j\beta z}) \tag{7-4}$$

The first term in brackets can be taken to represent an outward traveling wave and the second term a reflected wave. The minus sign is the current reflection coefficient at an open circuit.

If the reflected wave is not strongly present on an antenna, it is referred to as a **traveling-wave antenna**. A traveling-wave antenna acts as a guiding structure for traveling waves, whereas a resonant antenna supports standing waves. Traveling waves can be created by using matched loads at the ends to prevent reflections. Also, very long antennas may dissipate most of the power, leading to small reflected waves by virtue of the fact that very little power is incident on the ends. In this section, several wire forms of traveling-wave antennas will be discussed. Some of the antennas in this section are essentially the traveling-wave counterparts of resonant wire antennas presented in Chap. 6. They tend to be broadband with bandwidths of as much as $2:1$.

The simplest traveling-wave wire antenna is a straight wire carrying a pure traveling wave, referred to as the **traveling-wave long wire antenna**. A long wire is one that is greater than one-half wavelength long. The traveling-wave long wire is shown in Fig. 7-1 with a matched load R_L to prevent reflections from the wire end. Exact analysis of this structure, as well as others to be presented in this section, is formidable. We shall make several simplifying assumptions that permit pattern calculations that do not differ greatly from real performance. First, the ground plane effects will be ignored and we will assume that the antenna operates in free space. A traveling-wave long wire operated in the presence of an imperfect ground plane is called a **Beverage antenna**, or **wave antenna**. The ground plane may be accounted for in certain cases by using the techniques of the previous chapter. Second, the details of the feed are assumed to be unimportant. In Fig. 7-2, the long wire is shown being fed from a coaxial transmission line as one practical method. The vertical section of length d is assumed not to radiate, which is approximately true for $d \ll L$. Finally, we assume that the radiative and ohmic losses along the wire are small. When attenuation is neglected, the current amplitude is constant and the phase velocity is that of free space [1]. We can then write

$$I_t(z) = I_m e^{-j\beta z} \tag{7-5}$$

which represents an unattenuated traveling wave propagating in the $+z$-direction with the phase constant β of free space.

The current of (7-5) is that of a uniform line source with a linear phase constant of $\beta_o = -\beta$. From (5-4), $\beta_o = -\beta\cos\theta_o$, so the pattern factor maximum radiation angle (not including element factor effects) is $\theta_o = 0°$, which implies an endfire pattern. The complete radiation pattern from (5-6) is

$$F(\theta) = K \, \sin\theta \, \frac{\sin[(\beta L/2)(1 - \cos\theta)]}{(\beta L/2)(1 - \cos\theta)} \tag{7-6}$$
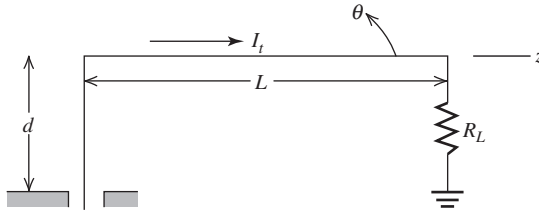
Figure 7-2 Traveling-wave long wire antenna.
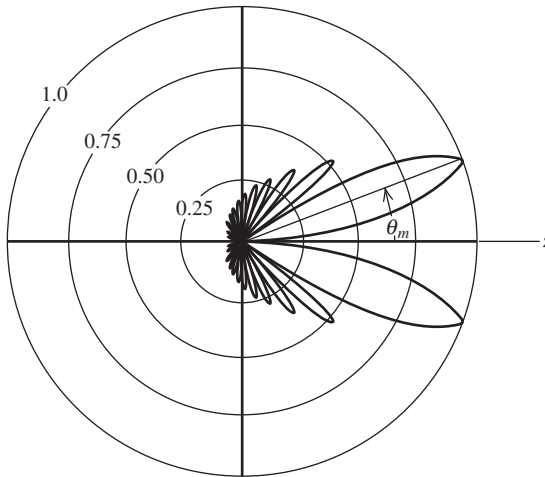
Figure 7-3 Pattern of a traveling-wave long wire antenna. $L = 6\lambda$ and $\theta_m = 20°$.

where K is a normalization constant that depends on the length L. The polar pattern for $L = 6\lambda$ is shown in Fig. 7-3. The length $L = n\lambda$ results in n forward lobes in the angular range $0 < \theta_m < 90°$. In this example, $n = 6$. The element factor $\sin\theta$ forces a null in the endfire direction. Hence, instead of having a single endfire lobe (which the pattern factor produces), the "main beam" is a rotationally symmetric cone about the z-axis. The maximum radiation angle in this case is $\theta_m(L = 6\lambda) = 20.1°$. In general, it is a function of L. Solving (7-6) for θ_m as a function L produces the plot of Fig. 7-4. An approximate expression for the angle of maximum radiation is [H.3: Wolff]

$$\theta_m = \cos^{-1}\left(1 - \frac{0.371}{L/\lambda}\right) \tag{7-7}$$

The beam direction values from Fig. 7-3 or (7-7) for a traveling-wave long wire of length L may be used to calculate an approximate beam direction for a standing-wave straight wire antenna (i.e., dipole). For example, θ_m for $L = 3\lambda/2$ from Fig. 7-3 is $40°$ and θ_m for the dipole of Fig. 6-4d is $42.6°$. As L increases, the traveling-wave and standing-wave antenna main beam maximum angles approach each other [H.6: *Ant. Eng. Hdbk.,* 1st ed., Chap. 6]. The standing-wave wire antenna is distinguished from its traveling-wave counterpart by the presence of a second major lobe in the reverse direction; see Fig. 6-4d. This can be seen by noting that the traveling-wave current of (7-5) corresponds to the first term of the standing-wave current of (7-4). The second term of (7-4), which is the reflected wave, produces a pattern similar in shape but oppositely directed. Thus, a traveling-wave antenna has a beam with a maximum in the $\theta = \theta_m$ direction and a
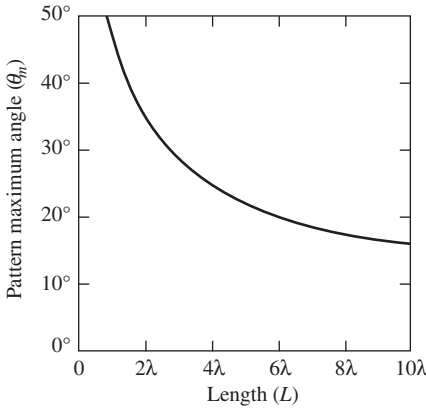
Figure 7-4 Pattern maximum angle for a traveling-wave long wire antenna of length L operating in free space. See (7-7).

standing-wave antenna of the same length has an additional beam in the $\theta = 180° - \theta_m$ direction.

The input impedance of a traveling-wave antenna is always predominantly real. This can be understood by recalling that the impedance of a pure traveling wave on a low-loss transmission line is equal to the (real) characteristic impedance of the transmission line. Antennas that support traveling waves operate in a similar manner. The radiation resistance of a traveling-wave long wire antenna is 200 to 300 Ω (see Prob. 7.2-5). The termination resistance should equal the value of the radiation resistance.

The resonant vee dipole discussed in Sec. 6.1.2 can be made into a traveling-wave antenna by terminating the wire ends with matched loads. The **traveling-wave vee antenna** is shown in Fig. 7-5. The pattern due to each arm separately is expressed by (7-5), an example of which is shown in Fig. 7-3. From Fig. 7-5, it is seen that when $\alpha \approx \theta_m$, the beam maxima from each arm of the vee will line up in the forward direction. A more accurate analysis of vee (see Prob. 7.2-8) includes the spatial separation effects of the arms. Pattern calculations as a function of α reveal that a good vee pattern is obtained when

$$\alpha \approx 0.8\ \theta_m \qquad (7\text{-}8)$$

where θ_m is found from Fig. 7-4 or (7-7). For $L = 6\lambda$, $\theta_m = 20°$ from Fig. 7-4 or (7-8) yields $\alpha \approx 16°$; the pattern for a vee with this geometry is shown in Fig. 7-5. The large side lobes
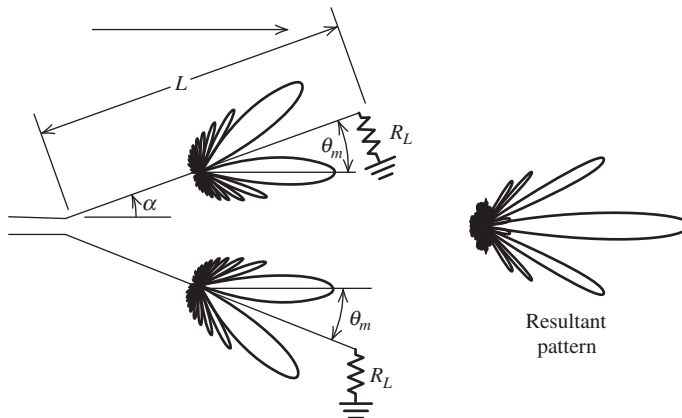
Figure 7-5 The traveling-wave vee antenna. In this case, $L = 6\lambda$ and $\alpha = 0.8\theta_m = 16°$.
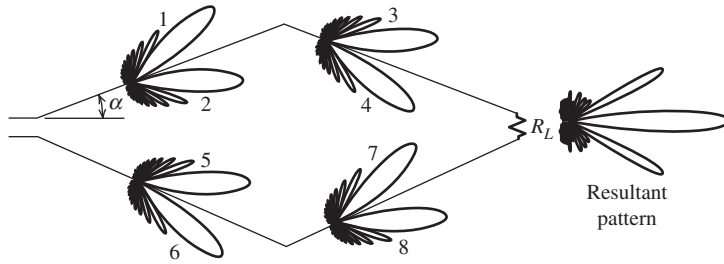
Figure 7-6 The rhombic antenna. Each side is of length L. Component beams 2 and 3, and 5 and 8 line up to form the main beam of the resultant pattern. In this case, $L = 6\lambda$ and $\alpha = 16°$.

arise from portions of the beams from each half of the vee that do not line up along axis. The pattern of the vee out of the plane of the vee is rather complicated due to the merging of the conical beams for each half of the vee. The traveling-wave vee provides relatively high gain for a bent wire structure.

By extending the ideas of a traveling-wave vee antenna, we obtain a **rhombic antenna** as shown in Fig. 7-6. The operation of this antenna is visualized most easily by viewing it as a transmission line that has been spread apart and consequently the characteristic impedance is increased. The load resistor R_L is of such a value as to match the transmission line. The antenna carries outward traveling waves that are absorbed in the matched load. Since the separation between the lines is large relative to a wavelength, the structure will radiate. If designed properly, a directive pattern with a single beam in the z-direction can be obtained.

A rhombic antenna operating in free space can be modeled as two traveling-wave vee antennas put together. If we choose $\alpha = 0.8\,\theta_m$ as for the vee, the beams of the rhombic in Fig. 7-6 numbered 2, 3, 5, and 8 will be aligned. Again, θ_m follows from Fig. 7-4. Due to the spatial separation of the two vees, the rhombic pattern will not be the same as that of a single vee. (See Prob. 7.2-9.)

The effects of a rhombic operating above a real Earth ground can be included by the techniques of Sec. 6.7.1. For a rhombic that is oriented horizontally, the reflection coefficient Γ_H is approximately -1 and the real Earth may be modeled as a perfect conductor; Fig. 6-62 illustrates that this assumption has a minor effect for horizontal antennas at lower frequencies. The array factor of a rhombic a distance h above a perfect ground plane produces a null along the ground plane. There are several designs for rhombics above a ground plane in the literature [2–5]. One such design is for the alignment of the major lobe at a specific elevation angle. Then the rhombus angle α and the elevation angle of the main beam are equal, and the height above ground is given by

$$h = \frac{\lambda}{4\sin\alpha} \tag{7-9}$$

and the length of each leg is

$$L = \frac{0.371\lambda}{\sin^2\alpha} \tag{7-10}$$

For example, if $\alpha = 14.4°$, then $L = 6\lambda$ and $h = 1\lambda$. Rhombic impedances are typically on the order of 600 to 800 Ω.

The efficiency of the rhombic antenna is decreased significantly because of the matched termination. The power that is not radiated is absorbed in the load R_L. However, this loss of power is essentially that which would have appeared in a large back lobe as a result of reflected current if the matched load were not present. The traveling wave feature not only improves the pattern but also produces wider impedance bandwidth. Well-designed
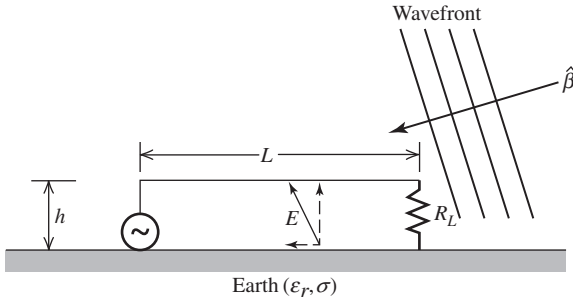
Figure 7-7 Beverage receiving antenna on an imperfect ground plane.
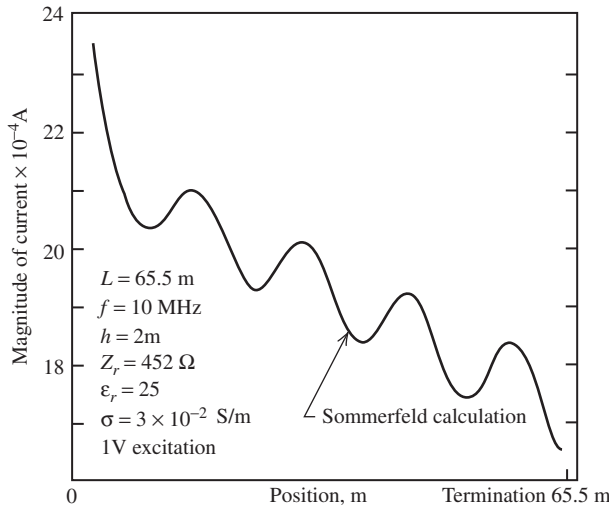
Figure 7-8 Current distribution on a Beverage antenna, $L = 2.18\lambda$, $h = 0.067\lambda$.

traveling-wave antennas have input impedances that have little reactance since there is little or no reflected power.

The traveling-wave antennas discussed above have been examined in free space. One traveling-wave antenna that requires the imperfectly conducting properties of real Earth is the Beverage antenna shown in Fig. 7-7. The height h is a small fraction of a wavelength, and the length L is usually between 2 and 10 wavelengths. The incoming vertically polarized plane wave in Fig. 7-7 produces a horizontal component of the electric field that is not totally shorted out by the imperfectly conducting Earth. It is this horizontal component that is responsible for inducing a current on the antenna conductor of length L. Alternatively, the Beverage antenna and its image in the lossy Earth may be viewed as an unbalanced transmission line. As was noted in the folded dipole discussion in Sec. 6.2, unbalanced transmission lines can radiate.

Fig. 7-8 shows the current on a Beverage antenna of length 2.18λ. The curve was calculated by the rigorous Sommerfeld theory for antennas in proximity to the imperfectly conducting Earth [1,2]. Because there are large dissipative losses in the Earth as well as radiative losses, the current decays from the feed end to the load end. This portion of the current distribution can be accounted for by a modification of (7-5):

$$I_t(z) = I_m e^{-\alpha z} e^{-j\beta z} \tag{7-11}$$

The current also shows a standing wave superimposed on the major portion of the current distribution. This is evidence of a small reflected wave from the load end. The reflected
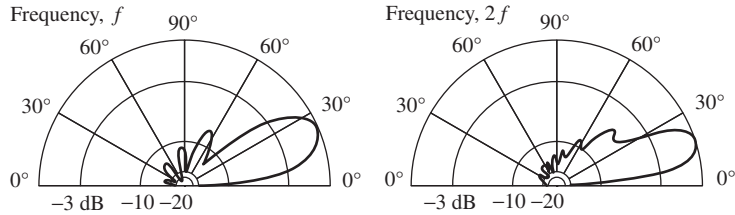
Figure 7-9 Typical elevation patterns for a Beverage antenna.

wave also appears to diminish in strength as it moves toward the feed end. The reflected wave can be represented by a relationship similar to (7-11). The approximate theory presented in Sec. 6.7 cannot be used to obtain the current distribution in Fig. 7-8 because the Beverage interacts too strongly with the Earth due to its close proximity to the ground and its long horizontal length. Instead, the rigorous but complicated Sommerfeld theory [1] must be used.

Fig. 7-9 shows some typical elevation plane patterns for a Beverage antenna used by the military as a tactical field antenna. The antenna is particularly well-suited for this purpose because it does not need to be elevated much above the ground. Note several things about the patterns. The higher-frequency pattern has a narrower beamwidth in both the elevation and azimuthal planes, and has a lower angle of radiation as expected. Further, the back lobe radiation is low, particularly in the higher frequency case where the antenna is electrically twice as long and there is more opportunity for dissipative loss.

Often, the feed end of the Beverage is higher than the load end to achieve as low a radiation angle relative to the ground (take-off angle) as possible. The Beverage antenna is usually used in the LF and lower HF portions of the frequency spectrum. It is believed to have been first used on Long Island in the early days of radiotelephone for trans-atlantic communication between the United States and London at 50 and 60 kHz. The Beverage was the first antenna to use the traveling-wave principle.

7.3 HELICAL ANTENNAS

If a conductor is wound into a helical shape and fed properly, it is referred to as a **helical antenna**, or simply as a **helix**. The typical geometry for a helix is shown in Fig. 7-10. If
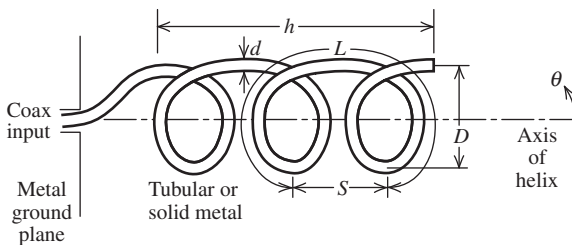
Figure 7-10 Geometry and dimensions of a helical antenna. This is a left-hand wound helix.

one turn of the helix is uncoiled, the relationships among the various helix parameters are revealed, as shown in Fig. 7-11. The symbols used to describe the helix are defined as follows:

D = diameter of helix (between centers of coil material)
C = circumference of helix = πD
S = spacing between turns = $C \tan \alpha$
α = pitch angle = $\tan^{-1} \frac{S}{C}$
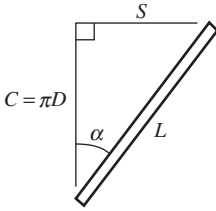
Figure 7-11 One uncoiled turn of a helix.

L = length of one turn = $\sqrt{C^2 + S^2}$
N = number of turns
L_w = length of helix coil = NL
h = height = axial length = NS
d = diameter of helix conductor

Note that when $S = 0$ $(\alpha = 0°)$, the helix reduces to a loop antenna, and when $D = 0$ $(\alpha = 90°)$, it reduces to a linear antenna.

The helix can be operated in two modes: the normal mode and the axial mode. The normal mode yields radiation that is most intense normal to the axis of the helix. This occurs when the helix diameter is small compared to a wavelength. The axial mode provides a radiation maximum along the axis of the helix. When the helix circumference is on the order of a wavelength, the axial mode will result. We introduce both the normal mode and axial mode helices in this section. See [H.3: Kraus, 3rd ed., H.4: Volakis, 4th ed., Chap. 12; H.8.3: Nakano] for more details.

7.3.1   Normal Mode Helix Antennas

In the normal mode of operation, the radiated field is maximum in a direction normal to the helix axis and for certain geometries, in theory, will emit circularly polarized waves. For normal mode operation, the dimensions of the helix must be small compared to a wavelength, that is, $D \ll \lambda$ and usually $L \ll \lambda$ as well. The normal mode helix is electrically small and thus its efficiency is low.

Since the helix is small, the current is assumed to be constant in magnitude and phase over its length. The far-field pattern is independent of the number of turns and may be obtained by examining one turn. One turn can be approximated as a small loop and ideal dipole as shown in Fig. 7-12. The far-zone electric field of the ideal dipole from (2-74a) is

$$\mathbf{E}_D = j\omega\mu \, IS \, \frac{e^{-j\beta r}}{4\pi r} \sin\theta \, \hat{\boldsymbol{\theta}} \tag{7-12}$$

where S, the spacing between helical turns, is the length of the ideal dipole in Fig. 7-12. The far-zone electric field of the small loop from (3-49) is
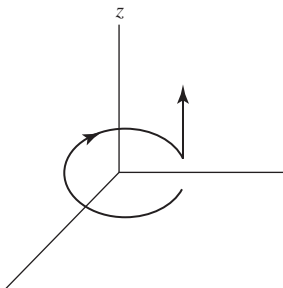
Figure 7-12 One turn of a normal mode helix approximated as a small loop and an ideal dipole.

(a) Geometry. (b) Radiation pattern of both $|E_\theta|$ and $|E_\phi|$.
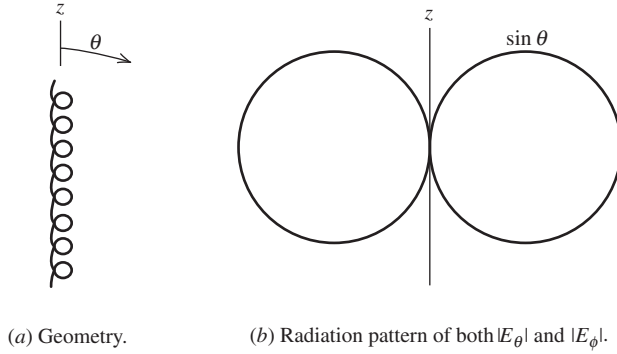
Figure 7-13 The normal mode helix and its radiation pattern.

$$\mathbf{E}_L = \eta\beta^2 \frac{\pi}{4} D^2 I \frac{e^{-j\beta r}}{4\pi r} \sin\theta \, \hat{\boldsymbol{\phi}} \qquad (7\text{-}13)$$

where $\pi D^2/4$ is the area of the loop. The total radiation field for one turn, as modeled in Fig. 7-12, is the vector sum of the fields in (7-12) and (7-13). Note that both components have a $\sin\theta$ pattern (see Fig. 7-13) and they are $90°$ out-of-phase. The ratio of the electric field components is

$$\frac{|E_\theta|}{|E_\phi|} = \frac{4\omega\mu S}{\sqrt{(\mu/\varepsilon)}\omega\sqrt{\mu\varepsilon}(2\pi/\lambda)\pi D^2} = \frac{2S\lambda}{\pi^2 D^2} = \frac{2\frac{S}{\lambda}}{\left(\frac{C}{\lambda}\right)^2} \qquad (7\text{-}14)$$

This equals the axial ratio of the polarization ellipse when greater than unity and the inverse of axial ratio when less than unity; see Sec. 2.8. Limiting axial ratio values are 0 (with $S = 0$) corresponding to a small loop with horizontal polarization and ∞ (with $D = 0$) corresponding to a short dipole with vertical polarization.

Since the (perpendicular) linear components are $90°$ out-of-phase, circular polarization is obtained if the axial ratio is unity. This occurs for

$$C = \pi D = \sqrt{2S\lambda} \qquad (7\text{-}15)$$

which was found by setting (7-14) equal to unity. This circular polarization is obtained in all directions, except of course where the pattern is zero (along the axis of the helix). From Fig. 7-11, it is seen that

$$L \sin\alpha = S \quad \text{or} \quad \alpha = \sin^{-1}\frac{S}{L} \qquad (7\text{-}16)$$

and

$$C^2 + S^2 = L^2 \qquad (7\text{-}17)$$

For circular polarization in the normal mode, the circumference of the helix given by (7-15) used in (7-17) gives

$$S_{\text{CP}}^2 + 2S_{\text{CP}}\lambda - L^2 = 0 \qquad (7\text{-}18)$$

This is a quadratic equation that may be solved for S as

$$S_{\text{CP}} = \frac{-2\lambda \pm \sqrt{4\lambda^2 + 4L^2}}{2} = \lambda\left[-1 \pm \sqrt{1 + \left(\frac{L}{\lambda}\right)^2}\right] \qquad (7\text{-}19)$$

Choosing the plus sign to keep the physical length S positive and substituting into (7-16) yields the pitch angle required for circular polarization:

$$\alpha_{CP} = \sin^{-1}\left[\frac{-1 + \sqrt{1 + (L/\lambda)^2}}{L/\lambda}\right] \qquad (7\text{-}20)$$

Usually, the normal mode helix is oriented vertically and operated such that the ratio in (7-14) is greater than unity, leading to predominantly vertically polarized radiation. This antenna is very popular in small transceivers such as handheld personal radios; see Sec. 11.6.1. For these applications, the wire length L_w is about a quarter-wave-length and the antenna is operated as a monopole fed against a ground plane. In this context, it is often referred to as a **normal mode helix antenna (NMHA)**, or **resonant (quarter-wave) stub helix antenna**. The pattern is, of course, nearly omnidirectional. The advantages of the stub helix over a conventional straight-wire monopole of the same height is that the helix acts as an inductor, tending to cancel the capacitance inherent in electrically short antennas. The current along the wire of the helix is approximately sinusoidal. The radiation resistance of a resonant stub helix above a perfect ground plane for heights under $\lambda/8$ is [H.6: *Ant. Eng. Hdbk.*, 3rd ed., Chap. 13]

$$R_r \approx 640\left(\frac{h}{\lambda}\right)^2 \, \Omega \quad \text{resonant, stub helix} \qquad (7\text{-}21)$$

The counterpart short monopole from (3-15) has $R_r = 395(h/\lambda)^2 \, \Omega$. Since electrically short antennas suffer from low radiation resistance, the higher radiation resistance of the stub helix is another improvement over the conventional short (straight wire) monopole.

The normal mode helix antenna is not a broadband antenna but is similar in structure to the axial mode helix antenna, which is broadband and is discussed in the next section.

EXAMPLE 7-1   *A Stub Helix Antenna for Handheld Cellular Radios*

Consider a stub helix operating in the cellular telephone band at 883 MHz ($\lambda = 34$ cm). The four-turn helix is 2.25 in. (5.7 cm) high, or $h = 0.168\lambda$, and 0.2 in. (0.5 cm) in diameter, giving a circumference of $C = 0.046\lambda$. The turn spacing is $S = h/N = 5.7$ cm/4 = 1.43 cm = 0.042λ. The helix length is $L_w = [(NC)^2 + h^2]^{1/2} = 0.25\lambda$, which is of resonant length.

The axial ratio from (7-14) is

$$|AR| = \frac{2\left(\dfrac{S}{\lambda}\right)}{\left(\dfrac{C}{\lambda}\right)^2} = \frac{2(0.042)}{(0.046)^2} = 38$$

This confirms the nearly vertical linear nature of the polarization. The radiation resistance from (7-21) is $R_r = 640(0.168)^2 = 18 \, \Omega$. A straight monopole of the same height would have $R_r = 395(0.168)^2 = 11 \, \Omega$.

7.3.2 Axial Mode Helix Antennas

Axial mode helices are used when a moderate gain of up to about 15 dB and circular polarization are required. The relatively small cross section of the helix makes it popular at UHF frequencies, where it is widely used for satellite communications. In this section, we consider the axial mode, monofilar (single winding) helix antenna in detail with emphasis on design considerations. In the axial mode, the helix radiates as an endfire antenna with a single main beam along the axis of the helix (the $+z$-direction in Fig. 7-10). The radiation is close to circularly polarized near the axial direction. Further, the main beam narrows as turns are added to the helix. The axial mode occurs when the helix circumference C is on the order of one wavelength. Helices with a few turns perform well over the frequency range corresponding to

$$\frac{3}{4}\lambda \leq C \leq \frac{4}{3}\lambda \tag{7-22}$$

This gives a bandwidth ratio of

$$B_r = \frac{f_U}{f_L} = \frac{c/\lambda_U}{c/\lambda_L} = \frac{4/3}{3/4} = \frac{16}{9} = 1.78 \tag{7-23}$$

This is close to the conventional definition of a 2:1 bandwidth ratio for a wideband antenna. For long helices, the upper operating frequency is lower than $4\lambda/3$, reducing the bandwidth below 1.78.

Kraus [H.3:Kraus, 1st ed., Chap. 4] performed the pioneering work on the axial mode helix and provided a simple explanation of its operation as well as empirical formulas for pattern, gain, polarization, and impedance. Subsequent experiments [3, 6] produced more accurate models for helix antenna performance. The development that follows is based on these works.

An approximate model for the axial mode helix that offers a simple explanation for its operation assumes that the helix carries a pure traveling wave that travels outward from the feed. The electric field associated with this traveling wave rotates in a circle, producing radiation that is nearly circularly polarized off the end of the helix.

In contrast to the normal mode helix, which has a current that is nearly uniform in phase over the helix winding, the phase of the axial mode helix current shifts continuously along the helix, which is characteristic of a traveling wave. Since the circumference is close to one wavelength, the current at opposite points on a turn are about 180° out-of-phase. This cancels the current direction reversal introduced by the half-turn. Thus, the radiation from opposite points on the helix is nearly in-phase, leading to reinforcement along the axis in the far field. This radiation mechanism closely parallels that of the one-wavelength loop discussed in Sec. 6.8.

Both early measurements [H.3: Kraus] and theoretical analysis [3] showed that the phase velocity of the traveling wave closely follows that of a Hansen-Woodyard endfire antenna. Traditional modeling of the helix for pattern evaluation adopts an array approach with each turn being an element of the array as in [H.3: Stutzman and Thiele, 2nd ed, Sec. 6.2]. Here we follow a slightly easier path of a line source model. We assume the phase is that of Hansen-Woodyard endfire and the amplitude is to be uniform, which is reasonable for pattern calculation purposes. Specifically, the turn-to-turn phase shift from the Hansen-Woodyard endfire array case from (8-38b) is

$$\alpha_h = -\left(\beta S + \frac{\pi}{N}\right) \tag{7-24}$$

where the turn spacing S is equivalent to the element spacing d. The phase shift per unit length, which is needed for the line source model, is

$$\beta_o = \frac{\alpha_h}{S} = -\left(\beta + \frac{\pi}{NS}\right) = -\beta - \frac{\pi}{h} \tag{7-25}$$

where (7-24) and $h = NS$ were used. The pattern factor of the line source from (5-7) is $\sin u / u$ where from (5-3):

$$u = (\beta \, \cos \theta + \beta_o)\frac{h}{2} = \frac{1}{2}[\beta h(\cos \theta - 1) - \pi] \tag{7-26}$$

As mentioned, the element factor is $\cos \theta$. So the complete line source pattern is the product of the element factor and the pattern factor, as in (5-6), is

$$F(\theta) = \frac{\pi}{2}\cos \theta \frac{\sin\left\{\frac{1}{2}\left[\beta h(\cos \theta - 1) - \pi\right]\right\}}{\frac{1}{2}[\beta h \, (\cos \theta - 1) - \pi]} \tag{7-27}$$

where $\phi/2$ is needed for normalization to unity at $\theta = 0°$. This expression applies to E_θ and E_ϕ.

A remarkable feature of the axial mode helix is that the phase velocity automatically adjusts to maintain the increased directivity condition over a wide range of helix parameter values: $5° < \alpha < 20°$ and $\frac{3}{4}\lambda < C < \frac{4}{3}\lambda$. This result was found empirically. Just as the directors on a Yagi-Uda antenna can be viewed as a slow wave structure, the helix can be viewed as a slow wave structure supporting a circularly polarized traveling wave. The significant difference is that while the Yagi-Uda is narrowband the helix is broadband because of the self-adjusting phase.

The electrical performance of the axial mode helix is, of course, influenced by the several geometric parameters. A pitch angle of $\alpha = 13°$ is popular because the helix is broadband at that value, but pitch angles from 11 to 14° are good as well [4]. Increasing the number of turns increases gain and reduces the axial ratio. Although helix antennas can be accurately simulated by numerical computational codes, most notably the method of moments, it is helpful to have simple formulas for approximating key performance parameters and to use as design aids. Empirical formulas based on extensive measured data were originally developed by Kraus [H.3: Kraus, 1st ed.]. Subsequent investigations have improved the accuracy. The formulas to be presented are to be used under the condition of the circumference being bounded as in (7-22) as well as $12° < \alpha < 15°$ and $N > 5$.

The half-power beamwidth in degrees is given by

$$\text{HP} = \frac{65°}{\frac{C}{\lambda}\sqrt{N\frac{S}{\lambda}}} \tag{7-28}$$

Originally the numerator constant used by Kraus was 52°, but this value has been found to be too low. The formula applies to all cut planes through the helix axis because of the high degree of symmetry in the structure. The gain approximation follows directly from the HP approximation using (4-57):

$$G = \frac{26,000}{\text{HP}^2} = 6.2\left(\frac{C}{\lambda}\right)^2 N\frac{S}{\lambda} \tag{7-29}$$

This formula is presented as a gain formula, but does not include wire losses (usually small) and matching network losses (can be 1 dB or more) that need to be accounted for separately. It does include gain reductions due to pattern imperfections such as null filling

and high side lobes that often occur in practice. Originally the leading constant value used by Kraus was 15, but this has been shown to be too high [5]. Extensive experimental investigation has shown the numerator constant in this formula to vary from 4.2 to 7.7 for the range of parameters noted [6]. Based on (7-29), gain varies with frequency as f^3, although experiments indicate this to be strictly true only for N about 10. Variation with N from (7-29) is linear, so that doubling the helix length adds 3 dB to the gain. However, measurements show this to be high [6]. The circumference variation of $(C/\lambda)^2$ is limited to the range in (7-22), and experiments have shown that gain peaks around $C = 1.1\lambda$ [6].

The classical empirical formula for axial ratio [H.3: Kraus, 1st ed.] is:

$$|\text{AR}| = \frac{2N + 1}{2N} \tag{7-30}$$

indicating that circular polarization improves ($|\text{AR}|$ lowers) as the number of turns increases. For a 10-turn helix, (7-30) yields $|\text{AR}| = 21/20 = 1.05 = 0.4\,\text{dB}$. This is an overly optimistic value that is not achieved in practice. Experimental data indicate that axial ratio is less than $1.2(1.5\,\text{dB})$ for $0.8 < C/\lambda < 1.2$. [6] Axial ratio can be improved by tapering the last few turns of the helix. The sense of the radiated wave is determined by the sense of the windings of the helix as shown in Fig. 7-14. Thus, a left- (right-) hand wound helix is left- (right-) hand sensed polarized. It is important to remember that the sense of polarization is the same when using the antenna for transmit or receive.

The classical formula for input resistance is

$$R_A = 140\frac{C}{\lambda}\,\Omega \tag{7-31}$$

The input impedance is real-valued due to the nearly pure traveling-wave behavior of a properly designed helix. This simple formula must be regarded as only an approximation since the input impedance of actual helix antennas is affected by details of the feed. However, the input impedance does remain nearly resistive over a wide bandwidth.

An axial mode helix performs well and is represented approximately by the foregoing empirical relations when the increased directivity condition is satisfied. This occurs over the bandwidth of (7-22) for helices of a few turns. However, long helices have a reduced upper operating frequency; for example, f_U corresponds to about $C \approx \lambda$ for $N = 50$ [30].

If only a pure traveling wave exists on the helix, the ground plane would have little effect. However, other modes are present, including a wave reflected from the end of the helix that returns to the feed region. This makes the ground plane geometry important. An approximate guideline is that the ground plane should be at least $3\lambda/4$ in diameter. Ground structures such as cups or cones are often used in place of a larger ground plane [7]. The inner conductor of the coaxial connector is attached to the helix and the outer conductor is connected to the ground structure as indicated in Fig. 7-14. The conductor diameter is frequently selected to provide a rigid, self-supporting structure. Helix winding conductor diameter d usually is between 0.005λ and 0.05λ. A 50-Ω input impedance can be achieved by including an impedance transformer in the feed or by adjusting the location of the wire from the coax to the helix winding.

There are many variations of the helix antenna. For example, tapering the helix by gradually reducing the diameter near the end of the helix improves impedance, pattern,

(a) Left-hand sensed helix.   (b) Right-hand sensed helix.
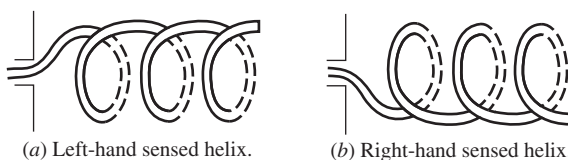
Figure 7-14 Left- and right-hand wound helices. For the axial mode helix, the sense of the windings determines the sense of polarization of the antenna.

and polarization [8]. A compact volume can be achieved at the sacrifice of bandwidth with a helix that conforms to a spherical surface rather than a cylinder [9]. A recent modification is the *stub-loaded helix antenna* that uses internal stubs to reduce the volume with only small reduction in bandwidth [10].

EXAMPLE 7-2   *A 10-Turn Axial Mode Helix Antenna*

The helix antenna is rather easy to construct and will perform approximately as predicted by the simple theory presented in this section, as will be demonstrated in this example. Calculations are compared to experimental results for a 10-turn helix constructed for a center frequency of 8 GHz ($\lambda = 3.75$ cm), where the helix was designed to have a circumference of $C = 0.92\lambda = 3.45$ cm. The helix was built with a pitch angle of $\alpha = 13°$. The spacing between turns is then $S = C \tan \alpha = 3.45 \tan 13° = 0.796$ cm. The measured radiation patterns for the two principal planes are shown in Figs. 7-15a and 7-15b [3]. These patterns are nearly alike and compare well to the pattern computed from simple theory with (7-27), which is plotted in Fig. 7-15c. The beamwidth of the measured patterns is about 44°. For comparison, the computed pattern of Fig. 7-15c has HP = 39°, and the approximate empirical formula of (7-28) gives a beamwidth of

$$\text{HP} = \frac{65°}{\frac{C}{\lambda}\sqrt{N\frac{S}{\lambda}}} = \frac{65°}{0.92\sqrt{10(0.212)}} = 48° \tag{7-32}$$

whereas (7-27) yields

$$\text{HP} = 47° \tag{7-33}$$

(*a*) Measured E_θ pattern at 8 GHz [10].   (*b*) Measured E_ϕ pattern at 8 GHz [10].

(*c*) Pattern computed from (7-27).
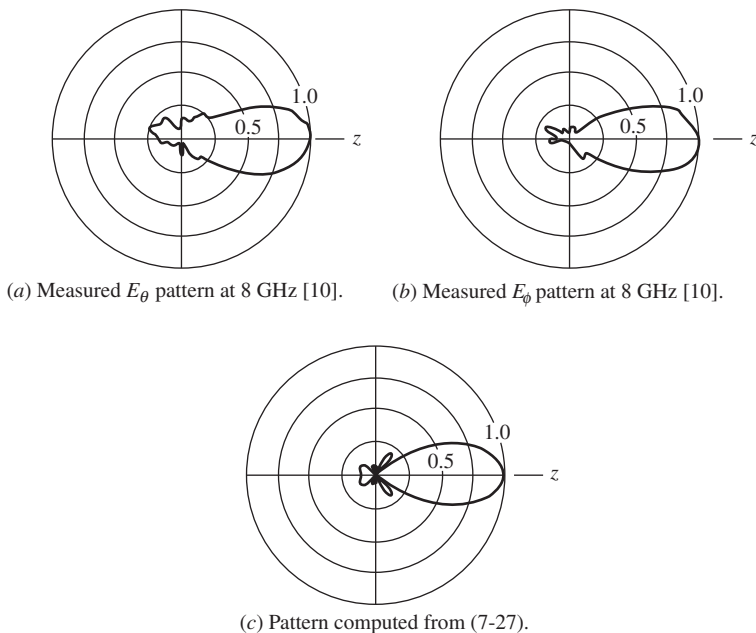
Figure 7-15 Radiation patterns of a 10-turn axial mode helix with $C = 0.92 \lambda$ and $\alpha = 13°$ (Example 7-2).

The gain predicted by (7-29) is

$$G = 6.2\left(\frac{C}{\lambda}\right)^2 N\frac{S}{\lambda} = 6.2(0.92)^2 10(0.212) = 11.1 = 10.5 \text{ dB} \tag{7-34}$$

7.4   BICONICAL ANTENNAS

Broadbanding of dipole and monopole antennas was tried early in the history of antennas. Oliver Lodge built a biconical antenna in 1897, and Marconi constructed a vertical monocone antenna (see Fig. 1-3) in 1901; see [H.8.1: Bevensee] for a survey of conical antenna development.

The bandwidth of a simple dipole antenna can be increased by using thicker wire as indicated in Fig. 6-7. This concept can be extended to further increase bandwidth if the conductors are flared to form a biconical structure. Then the fixed wire diameter is replaced by a smoothly varying diameter and a fixed angle (of the conical surfaces). In this section, the idealized biconical antenna is considered first, followed by two practical forms—the finite biconical antenna and the discone.

7.4.1   The Infinite Biconical Antenna

If the conducting halves of an antenna are two infinite conical conducting surfaces end-to-end, but with a finite gap at the feed point, the **infinite biconical antenna** of Fig. 7-16 results. Since the structure is infinite, it can be analyzed as a transmission line. With a time-varying voltage applied across the gap, currents will flow radially out from the gap along the surface of the conductors. These currents, in turn, create an encirculating magnetic field H_ϕ. If we assume a TEM transmission line mode (all fields transverse to direction of propagation), the electric field will be perpendicular to the magnetic field and be θ-directed. When the potential on the top cone is positive and the bottom cone is negative, the electric field lines extend from the top to the bottom cone as indicated in Fig. 7-16.
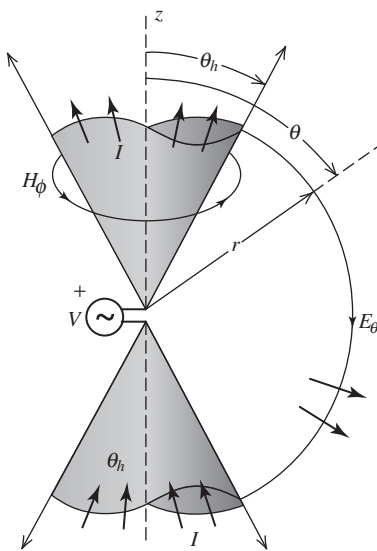
Figure 7-16 Infinite biconical antenna. The field components and current are shown.

In the region between the cones, $\mathbf{J} = 0$, $\mathbf{H} = H_\phi \hat{\boldsymbol{\phi}}$, and $\mathbf{E} = E_\theta \hat{\boldsymbol{\theta}}$. Then Ampere's law, $\nabla \times \mathbf{H} = j\omega\varepsilon\mathbf{E} + \mathbf{J}$, reduces to

$$\frac{1}{r \sin\theta} \frac{\partial}{\partial\theta} (\sin\theta \, H_\phi) = j\omega\varepsilon E_r = 0 \tag{7-35}$$

for the r-component and

$$-\frac{1}{r} \frac{\partial}{\partial r} (rH_\phi) = j\omega\varepsilon E_\theta \tag{7-36}$$

for the θ-component. From (7-35), we see that $\partial/\partial\theta \, (\sin\theta \, H_\phi) = 0$ so

$$H_\phi \propto \frac{1}{\sin\theta} \tag{7-37}$$

Since the structure acts as a guide for spherical waves, we can write (7-37) as

$$H_\phi = H_o \frac{e^{-j\beta r}}{4\pi r} \frac{1}{\sin\theta} \tag{7-38}$$

Then, substituting this into (7-36), we obtain

$$E_\theta = \frac{-1}{j\omega\varepsilon} \frac{1}{r} \frac{H_o}{4\pi} \frac{1}{\sin\theta} \frac{\partial}{\partial r} (e^{-j\beta r}) = \frac{\beta H_o}{\omega\varepsilon} \frac{1}{r} \frac{e^{-j\beta r}}{4\pi} \frac{1}{\sin\theta} = \eta H_o \frac{e^{-j\beta r}}{4\pi r} \frac{1}{\sin\theta} \tag{7-39}$$

This equation is simply $E_\theta = \eta H_\phi$, which confirms our statement that the wave is TEM. The field components vary as $1/\sin\theta$, so the radiation pattern is

$$F(\theta) = \frac{\sin\theta_h}{\sin\theta}, \qquad \theta_h < \theta < \pi - \theta_h \tag{7-40}$$

which is normalized to unity at its maxima on the conductor surfaces. This pattern is plotted in Fig. 7-17.

In order to determine the input impedance, we first find the terminal voltage and current. Referring to Fig. 7-16, we see the voltage is found by integrating along a constant radius r and it is
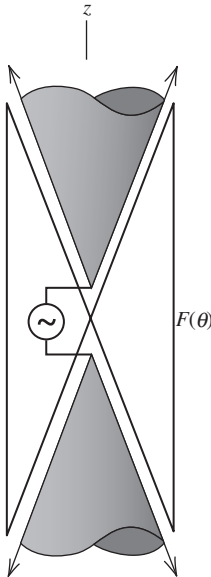
Figure 7-17 Radiation pattern of an infinite biconical antenna.

$$V(r) = \int_{\theta_h}^{\pi-\theta_h} E_\theta r \ d\theta \tag{7-41}$$

This can be performed for any r since the cones are equipotential surfaces. Substituting (7-39) into the above equation yields

$$V(r) = \frac{\eta H_o}{4\pi} e^{-j\beta r} \int_{\theta_h}^{\pi-\theta_h} \frac{d\theta}{\sin\theta} = \frac{\eta H_o}{4\pi} e^{-j\beta r} \left[\ln\left|\tan\ \frac{\theta}{2}\right|\right]_{\theta_h}^{\pi-\theta_h}$$

$$= \frac{\eta H_o}{2\pi} e^{-j\beta r} \ \ln\left(\cot\frac{\theta_h}{2}\right) \tag{7-42}$$

The boundary condition on H_ϕ at the conductor surface is $J_s = H_\phi$. The total current on one cone is found by integrating the current density J_s around the cone as shown in Fig. 7-16, so

$$I(r) = \int_0^{2\pi} H_\phi r \ \sin\theta \ d\phi = 2\pi r H_\phi \ \sin\theta \tag{7-43}$$

Substituting (7-38) in the above gives

$$I(r) = \frac{H_o}{2} e^{-j\beta r} \tag{7-44}$$

The characteristic impedance at any point r, from (7-42) and (7-44), is

$$Z_o = \frac{V(r)}{I(r)} = \frac{\eta}{\pi} \ln\left(\cot\ \frac{\theta_h}{2}\right) \tag{7-45}$$

Since this is not a function of r, it must be also the impedance at the input$(r = 0)$. Using $\eta \approx 120\pi$ in (7-45) gives the input impedance

$$Z_A = Z_o = 120 \ \ln\left(\cot\ \frac{\theta_h}{2}\right) \ \Omega \tag{7-46}$$

For θ_h less than $20°$,

$$Z_A = Z_o \approx 120 \ \ln\left(\frac{2}{\theta_h}\right) \ \Omega \tag{7-47}$$

where θ_h is in radians. The input impedance is real because there is only a pure traveling wave. Since the structure is infinite, there are no discontinuities present to cause reflections setting up standing waves, which would show up as a reactive component in the impedance (except at a few resonance points). If $\theta_h = 1°$, $Z_A = 568 + j0 \ \Omega$. If $\theta_h = 50°$, $Z_A = 91 + j0 \ \Omega$.

If one cone is flared all the way out to form a perfect ground plane, a single infinite cone above a ground plane results. This monopole version of the infinite bicone then has an input impedance which is half that of the infinite bicone.

7.4.2 Finite Biconical Antennas

A practical biconical antenna is made by ending the two cones of the infinite bicone forming the **finite biconical antenna** shown in Fig. 7-18. Inside an imaginary sphere of radius h just enclosing the antenna, TEM waves exist together with higher-order modes
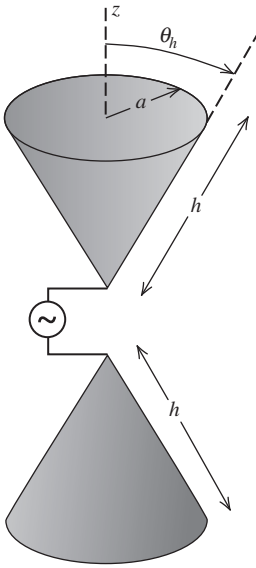
Figure 7-18 Finite biconical antenna.

created at the ends of the cones. These higher-order modes are the major contributors to the antenna reactance. The ends of the cones cause reflections that set up standing waves that lead to a complex input impedance.

The reactive part of the input impedance can be held to a minimum over a progressively wider bandwidth by increasing the angle θ_h in Fig. 7-18. At the same time, the real part of the input impedance becomes less sensitive to changing frequency (or changing h in Fig. 7-18). This is illustrated by measured data in Fig. 7-19 for a conical monopole where the antenna impedance is plotted versus the height of the monopole L_h. These data clearly show that one can achieve the 2:1 impedance bandwidth necessary for one part of our definition of a broadband antenna. This is our first example of an antenna that can be more dependent on an angle in its geometry description than on its length. Frequency-independent antennas, considered later in this chapter, exploit this property. Another property that we will observe in many broadband and frequency-independent antennas is that some important dimension must be at least $\lambda/4$. Careful examination of Fig. 7-19 reveals that the impedance bandwidth starts when the height of the conical monopole is about $\lambda/4$ and extends upward beyond $\lambda/2$. The pattern of a conical monopole or finite biconical for small cone angles is very similar to that of an ordinary monopole or dipole of the same length.

A much simpler alternative to the finite biconical antenna is the common "**bow-tie**" antenna (shown later in Fig. 7-34). It offers less weight and costs less to build, but will have a somewhat more sensitive input impedance to changing frequency than the finite biconical. The bow-tie is commonly constructed by printing the two halves on a thick substrate material, one on top and the other on the bottom. The monopole version of the bow-tie, called a **triangular antenna**, has impedance characteristics similar to those for the conical monopole of Fig. 7-19, but with somewhat smaller variations with antenna length [11]. A unidirectional bow-tie with 37% bandwidth is made by mounting it parallel to and $3/8\lambda$ in front of a ground plane [12].

7.4.3 Discone Antennas

If one cone of the finite biconical antenna is replaced with a disk-shaped ground plane, the structure becomes a disk-cone, or **discone**, antenna (see Fig. 7-20). The discone antenna was developed by Kandoian [13] in 1945, followed several years later by
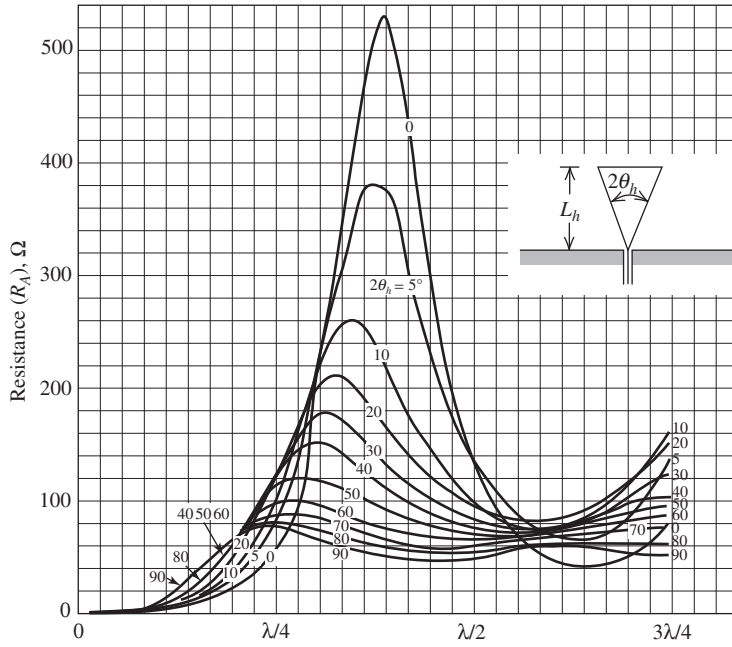
(a) Input resistance.
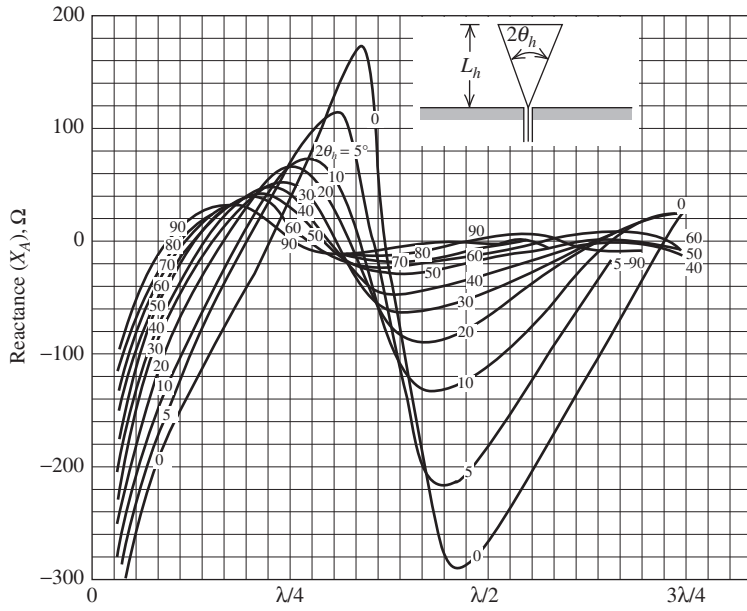
(b) Input reactance.

Figure 7-19 Measured input impedance of a conical monopole with flare angle $2\theta_h \leq 90°$ versus monopole height L_h. (Used with permission from [H.6: Jasik] © 1961 The McGraw-Hill Companies.)
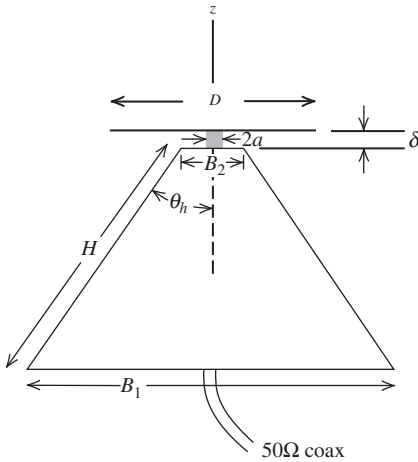
Figure 7-20 Discone antenna. Typical dimensions are $H \sim 0.7\lambda$, $B_1 \sim 0.6\lambda$, $D \sim 0.4\lambda$, and $\delta \ll D$.

experimental design studies [14, 15]. It is used (like a vertical dipole) for vertical polarization and nearly uniform azimuth coverage (i.e., an omnidirectional pattern). The discone offers satisfactory operation over a wide frequency range (several octaves) while maintaining acceptable pattern and impedance properties.

The discone is constructed as shown in Fig. 7-20. The outer conductor of the coaxial transmission line is connected to the cone and the inner conductor is attached to the disk ground plane. The cone and disk can be either solid metal or radial wires. Ideally, the pattern between the ground plane and the cone is that of the infinite bicone. This omnidirectional pattern is well suited to broadcast applications.

The discone antenna can be designed for broadband impedance performance (typically $50\,\Omega$), while maintaining acceptable pattern behavior with frequency [13]. Typical center frequency dimensions are $H = 0.7\lambda$, $B = 0.6\lambda$, $D = 0.4\lambda$, $\theta_h = 25°$, and $\delta \ll D$. For example, the discone with the patterns of Fig. 7-21 has a center frequency of 1 GHz ($\lambda = 30\,\text{cm}$). So at 1 GHz, $H = 21.3$ cm $= 0.71\lambda$, $B_1 = 19.3$ cm $= 0.64\lambda$, and $\theta_h \approx \sin^{-1}[(B/2)/H] = 27°$. Nail [14] has given optimum design formulas of $D = 0.7B_1$ and $\delta = 0.3B_2$, independent of H and θ_h.

The pattern performance over a 3:1 bandwidth is revealed in Fig. 7-21. At low frequencies, the structure is small relative to a wavelength, and the pattern is not too different from that of a short dipole (see Fig. 7-21a). As frequency increases, the electrical size of the ground plane increases and the pattern is confined more to the lower half-space (see Fig. 7-21b). For further increases in frequency, the antenna behavior approaches that of an infinite structure. For example, at 1500 MHz, the pattern (of Fig. 7-21c) is very close to that of the monopole version of the infinite biconical antenna in Fig. 7-16. Measurements with several values of disk parameter D and spacing between the cone δ showed that the patterns are insensitive to these parameters [15].

Nail's optimum design formulas [12] are for $B_2 \approx \lambda_U/75$ at the highest operating frequency of the antenna and $\delta \ll D$. For larger values of B_2 and $\delta \approx 0.5B_2$, it has been found that Nail's equations need to be altered when a type N-connector is used between the skirt and the disk [14]. In this case, $\delta = 0.5B_2$, $2a = 0.33B_2$, $D = 0.75B_1$, $L = 1.15\lambda_L$ based on experimental measurements. It has been reported in [16] that a VSWR below 1.5:1 over an octave bandwidth is easily achievable and $45° < 2\theta_h < 75°$ yields the best results. This is evident from Fig. 7-22 [H.6; Rudge Vol. 2, pg. 735] which shows VSWR for various discone angles. The best low-frequency performance along with a nearly "flat" VSWR curve at higher frequencies occurs for approximately $45° < 2\theta_h < 75°$.
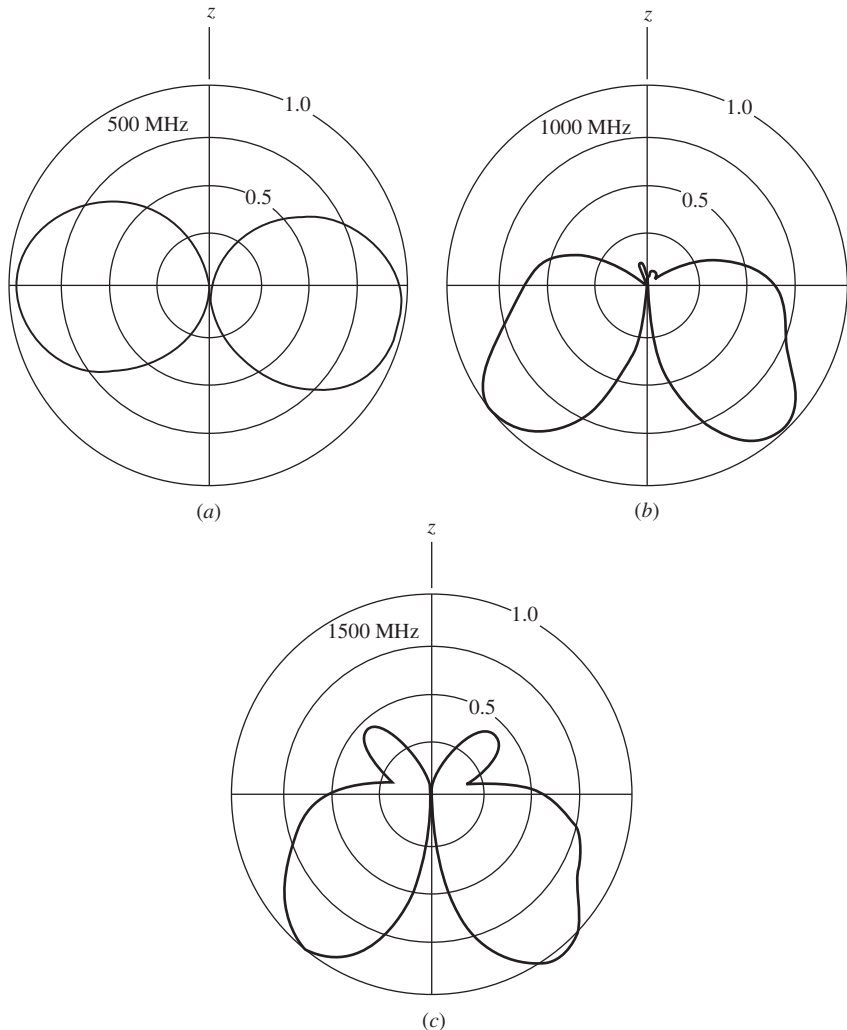
Figure 7-21 Measured patterns of a discone antenna for $H = 21.3$ cm, $B_1 = 19.3$ cm, and $\theta_h = 25°$.

Discone, biconical, and bow-tie antennas are often constructed of wires rather than solid metal to reduce weight and cost, and several commercial products are available. Fig. 14-40 illustrates a wire-formed monocone. Simulation studies [17] have shown that if more than eight wires are used to form the monocone, the performance is similar to the solid monocone counterpart antenna; also see [18].

7.5   SLEEVE ANTENNAS

In Sec. 6.1, we saw that the dipole antenna is very frequency-sensitive and its bandwidth is much less than the octave bandwidth provided by the antennas studied previously in this chapter. However, the addition of a sleeve to a dipole or monopole can increase the bandwidth to more than an octave. In this section, we will briefly examine a few forms of the **sleeve antenna**, which incorporates a tubular conductor sleeve around an internal radiating element. Emphasis will be placed on practical configurations.
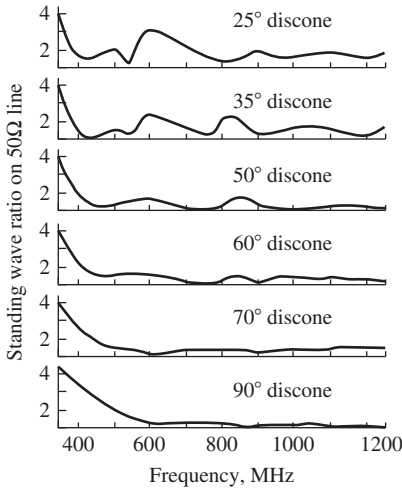
Figure 7-22 VSWR curves for discones of various cone angles, $2\theta_h$. (From [H.6: Rudge] © 1983 IET (London). Used with permission.)

7.5.1 Sleeve Monopoles

A **sleeve monopole** configuration is shown in Fig. 7-23a fed from a coaxial transmission line. The sleeve exterior acts as a radiating element and the interior of the sleeve acts as the outer conductor of the feed coaxial transmission line. In principle, the length of the sleeve can be any portion of the total length of the monopole from zero (no sleeve) to where the sleeve constitutes the entire radiating portion of the antenna. However, in practice, the sleeve is usually about $\frac{1}{3}$ to $\frac{1}{2}$ the height of the monopole. The reason for this is apparent from Figs. 7-23b and 7-23c, which suggest that the current at the virtual feed point changes only slightly as the overall monopole height varies from $\lambda/4$ to $\lambda/2$. Thus, the impedance remains somewhat constant over at least an octave. As for an ordinary monopole with no sleeve, the antenna dimensions affect the impedance more than the pattern.

Consider Fig. 7-23d. The first sleeve monopole resonance occurs at a frequency where the monopole length $\ell + L$ is approximately $\lambda/4$. Design proceeds by locating this first resonance near the lower end of the frequency band, thereby fixing the total physical length $\ell + L$. The remaining design variable is ℓ/L. It has been found experimentally that

(a) Geometry. (b) Current distribution. (Sleeve not shown) (c) Current distribution. (Sleeve not shown) (d) Geometry.
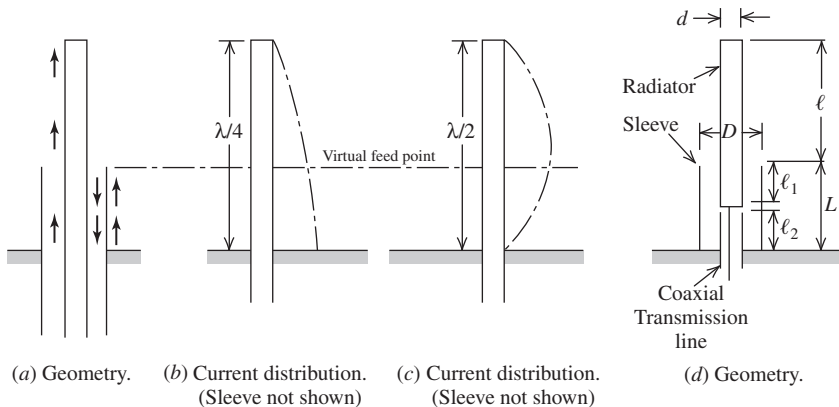
Figure 7-23 Sleeve monopole configurations: Arrows in (a) indicate polarity when $\ell + L \le \lambda/2$. Different current distributions on the center conductor are shown in (b) and (c).

Table 7-1   Specifications for Optimum Pattern Design
of a Sleeve Monopole

Pattern bandwidth	4:1
$\ell + L$	$\lambda/4$ at low end of band
ℓ/L	2.25
D/d	3.0
VSWR	less than 8:1

a value of $\ell/L = 2.25$ yields optimum (nearly constant with frequency) radiation patterns over a 4:1 band [19]. The value of ℓ/L has little effect for $\ell + L \leq \lambda/2$ since the current on the outside of the sleeve will have approximately the same phase as that on the top portion of the monopole itself, as suggested by the arrows of Fig. 7-23a. However, for longer electrical lengths the ratio ℓ/L becomes very important and has a marked effect on the radiation pattern, since the current on the outside of the sleeve will not necessarily be in-phase with that on the top portion of the monopole. Some typical specifications for optimum performance are given in Table 7-1. In some applications, the VSWR may be too high, requiring a matching network.

7.5.2   Sleeve Dipoles

The sleeve monopole has a counterpart in the **sleeve dipole** antenna that is shown in Fig. 7-24a. An approximate impedance analysis of the sleeve dipole can be carried out according to Figs. 7-24b and 7-24c where the doubly driven structure of Fig. 7-24a is replaced by the pair of asymmetrically driven structures in Fig. 7-24b. The change in diameters on the longer arms is neglected, resulting in Fig. 7-24c. The current at the input to the sleeve (virtual feed) $I_A(z_f)$ is then approximately the sum of the currents at the point $z = z_f$ from the two configurations in Fig. 7-24c. For the left half of Fig. 7-24c, the current at the virtual feed in that asymmetrical structure is $I_{as}(z_f)$. The current at the same point due to the excitation in the lower half of the sleeve dipole (i.e., the right half of Fig. 7-24c) is identical to the current at the point $-z_f$ when the excitation is at the point z_f, since the two structures are physically equivalent. Thus,

$$I_A(z_f) \approx I_{as}(z_f) + I_{as}(-z_f) \tag{7-48}$$
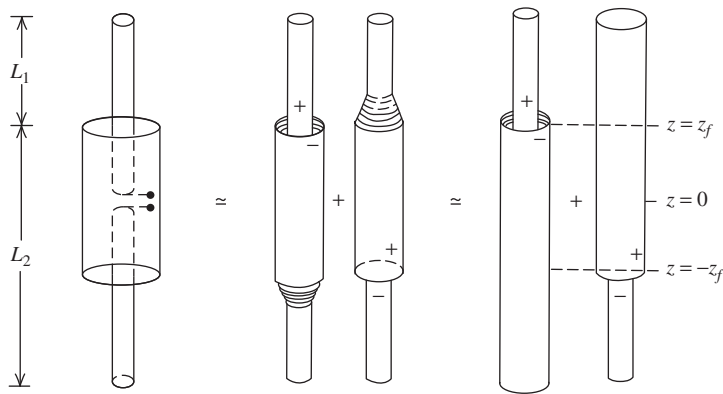
(a) Sleeve dipole.   (b) Intermediate equivalent.   (c) Final equivalent.

Figure 7-24   Sleeve dipole configuration and approximate equivalents (Used with permission from [H.3: Weeks] © 1968 The McGraw-Hill Companies.)

The input admittance to the sleeve is then

$$Y_A \approx \frac{I_{\mathrm{as}}(z_f) + I_{\mathrm{as}}(-z_f)}{V_A} \approx Y_{\mathrm{as}}\left[1 + \frac{I_{\mathrm{as}}(z_f)}{I_{\mathrm{as}}(-z_f)}\right] \tag{7-49}$$

where

$$Y_{\mathrm{as}} \approx \frac{2}{(Z_1 + Z_2)} \tag{7-50}$$

and where Z_1 is the impedance of a symmetrical antenna of half-length L_1 and Z_2 is the impedance of a symmetrical antenna with half-length L_2 [H.3: Weeks, Sec. 4.4.4]. Eq. (7-50) is useful for estimating the impedance of asymmetrical dipoles such as that in Prob. 6.l-4.

The sleeve dipole of Fig. 7-24a can be approximated with an **open-sleeve dipole** in which the tubular sleeve is replaced by two conductors close to either side of the driven element as shown in Fig. 7-25. The length of the parasites (simulated sleeve) is approximately one-half that of the center-fed dipole. The open-sleeve dipole, which we will describe from an experimental viewpoint, is operated in front of a flat reflector, or ground plane [20, 21]. The results are also applicable to sleeve dipoles without a flat reflector present.

The antenna was designed for the 225- to 400-MHz frequency band. The dipole to reflector spacing S_d was chosen to be 0.29λ at 400 MHz to avoid the deterioration of the radiation pattern that occurs for larger spacings. All the dimensions required for the design of the open-sleeve dipole are given in Table 7-2. These design values yield low VSWR over a wide bandwidth. This is illustrated in Fig. 7-26 by a comparison of the VSWR characteristics of a conventional (unsleeved) dipole and an open-sleeve dipole with a diameter D of 2.9 cm. Although these results do not represent exhaustive design
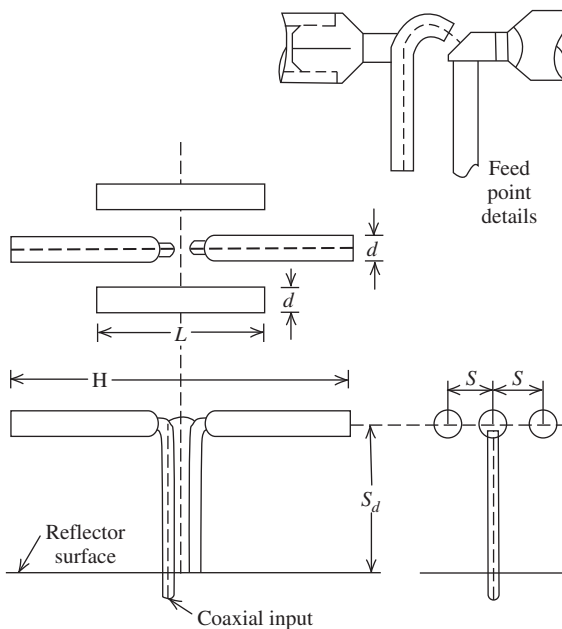
Figure 7-25 The open-sleeve dipole antenna with a flat reflector shown in front, top, and side views.

Table 7-2 Electrical Dimensions of an Open-Sleeve Dipole with a
Reflector for Lowest VSWR

Parameter (see Fig. 7-25)	Electrical Dimension at Lowest Frequency (225 MHz)	Electrical Dimension at Highest Frequency (400 MHz)
D	$0.26\,\lambda$	$0.047\,\lambda$
H	$0.385\,\lambda$	$0.684\,\lambda$
L	$0.216\,\lambda$	$0.385\,\lambda$
S	$0.0381\,\lambda$	$0.0677\,\lambda$
S_d	$0.163\,\lambda$	$0.29\,\lambda$
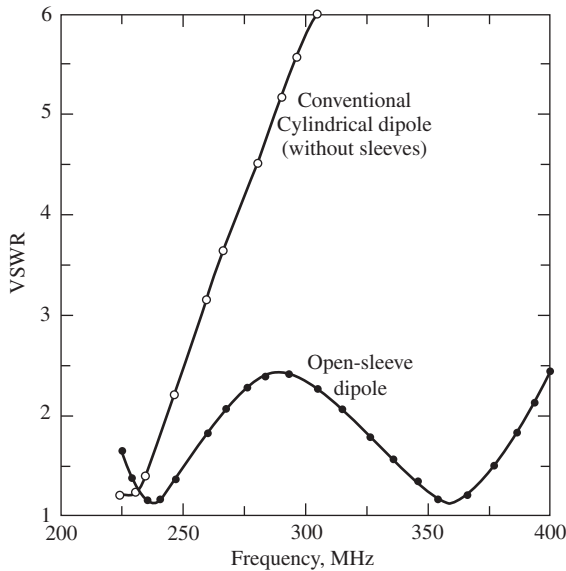
Figure 7-26 Comparison between the VSWR response of a conventional (unsleeved) cylindrical dipole and an open-sleeve dipole both with a diameter D of 2.9 cm [20].

data for the open-sleeve dipole, they do serve as a starting point in the design of open-sleeve dipoles with or without a reflector present.

7.6   PRINCIPLES OF FREQUENCY-INDEPENDENT ANTENNAS

Frequency independent antennas had their beginning in research conducted in the late 1950s at the University of Illinois and led to configurations that will be covered in this section. An early nonmilitary application of frequency-independent antennas was the log-periodic dipole TV antenna that was used on houses around the world starting in the 1960s.

Aperture antennas, to be discussed in the next chapter, are capable of bandwidths of 2:1 or more, but the main beam narrows as frequency is increased. Often, it is desirable to have the pattern of an antenna remain constant over a very wide range of frequencies. An antenna with a bandwidth of about 10:1 or more is referred to as a **frequency-independent** antenna. The purest form of a frequency-independent antenna has constant pattern, impedance, polarization, and phase center with frequency. Few antennas meet all these criteria. The axial mode helix has constant impedance and phase center location over a bandwidth of about 2:1, but the main beam narrows with increasing frequency. The

principles of frequency independence are discussed in this section and antennas capable of 10:1 bandwidth are introduced in the next two sections.

The biconical antenna represents the emergence of frequency-independent behavior. In Sec. 7.4, we found that the input impedance and pattern of the infinite biconical antenna were independent of frequency. This is precisely the behavior we desire. The feature of the biconical structure that is responsible for frequency independence is the emphasis on angles and the complete avoidance of finite lengths. This is verified by the observation that when the bicone is truncated to form the finite biconical antenna of Fig. 7-18, its bandwidth is limited. Of course, in general, if no finite lengths are present on an antenna, the structure would have to be infinite. Rumsey [H.8.3: Rumsey] noted that, in practice, frequency-independent antennas are designed to minimize finite lengths and maximize angular dependence. The concept of angle emphasis has been exploited to produce a family of frequency-independent antennas.

There is another property in addition to angle emphasis that leads to frequency-independent behavior, that of self-complementarity. Consider a metal antenna with input impedance Z_{metal}. A dual structure can be formed by replacing the metal with air and replacing air with metal. The resulting *complementary* antenna has input impedance Z_{air}. Complementary antennas are similar to a positive and negative in photography. An example is a ribbon dipole and its complement, the slot antenna, shown in Fig. 7-27.

Babinet's principle can be used to find the impedance of complementary antennas. Babinet's principle for optics states that a source of light behind complementary thin conducting sheets produces lit regions on the source-free side that when superposed give a completely lit region, just as would exist without the sheets present. Extending this to electromagnetics [19] leads to the following important relationship for the input impedances of complementary antennas [H.8.3: Mushiake, p. 16]:

$$Z_{metal}\, Z_{air} = \frac{\eta^2}{4} = \frac{(376.7)^2}{4} = 35{,}475.7\,\Omega \qquad (7\text{-}51)$$

This assumes that no dielectric or magnetic materials are present; if so, the proper η must be used in place of the free-space value. If the dipole of Fig. 7-27a is resonated by reducing its length slightly below a half-wavelength, its impedance is $Z_{metal} = Z_{dipole} = 70\,\Omega$. Then from (7-51), the impedance for the slot antenna of Fig. 7-27b, Z_{air}, is

$$Z_{slot} = \frac{\eta^2}{4Z_{dipole}} = \frac{35{,}475.7}{4(70)} = 506.8\,\Omega \quad \textit{half-wave slot antenna} \qquad (7\text{-}52)$$
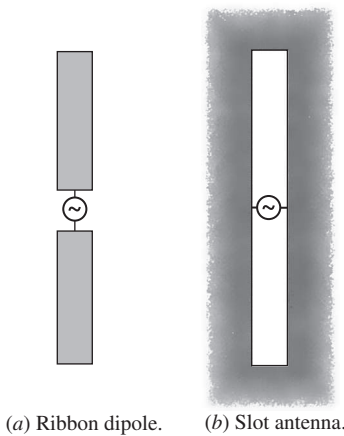
(*a*) Ribbon dipole.   (*b*) Slot antenna.   **Figure 7-27** Complementary dipole and slot antennas.

Prob. 7.6-2 addresses the slot that is complementary to the ideal dipole.

The product of the impedances of two complementary antennas is the constant $\eta^2/4$. If the antenna is its own complement, frequency-independent impedance behavior is achieved. This is the *self-complementary* property, in which the antenna and its complement are identical. A self-complementary structure can be made to exactly overlay its complement through translation and/or rotation. The value of impedance follows directly from (7-52), as noted by [H.8.3: Mushiake]

$$Z_{\text{metal}} = Z_{\text{air}} = \frac{\eta}{2} = 188.5\,\Omega \quad \textit{self-complementary antenna} \tag{7-53}$$

The frequency-independent impedance of (7-53) is the second design principle for frequency-independent antennas; that is, self-complementary antennas tend to be frequency-independent. It turns out, however, that many antennas that are not self-complementary still have small impedance variations with frequency.

Our study of the antennas in this chapter has led to a number of characteristics that are likely to produce broadband behavior. Before moving on to antenna types that yield extremely wide bandwidth in the following sections, we summarize these properties. The characteristics that yield broadband behavior are:

1. **Emphasis on angles rather than lengths.** Examples are the helix in Sec. 7.3 and the spiral in Sec. 7.7, which both avoid fixed physical length elements and produce wide bandwidth.
2. **Self-complementary structures.** The equiangular spiral of Fig. 7-29 is an example.
3. **Thick metal—"fatter is better."** Increasing the wire diameter of even resonant antennas such as a dipole widens its bandwidth; see Fig. 6-7. The biconical antenna is the ultimate fat dipole and has wide bandwidth; the biconical antenna also emphasizes angles. The bow-tie antenna of Fig. 7-34 is another example.

Ideally, frequency-independent antennas should display all three of these properties. It is found in practice that successful wide bandwidth designs emphasize these properties, but in many cases strict adherence is not required. For example, we will see in Sec. 7.8 that some log-periodic antennas deviate from the self-complementary principle and still have wide bandwidth. The usual penalty for doing this is that the impedance will vary with frequency and not be constant as predicted with (7-53). This may not be a serious problem in many applications.

A distinguishing feature of frequency-independent antennas is their *self-scaling* behavior. Most radiation takes place from that portion of the frequency-independent antenna where its width is a half-wavelength or the circumference is one wavelength—the so-called *active region*. Radiation is maximum perpendicular to the plane of the structure and can be explained in a fashion similar to the one-wavelength loop discussed Sec. 6.8. As frequency decreases, the active region moves to a larger portion of the antenna, where the width is a half-wavelength. The characteristics of angle emphasis and using thick metal yield structures that provide regions where the current can adjust as the frequency changes.

Frequency-independent antennas can be divided into two types: spiral antennas and log-periodic antennas. Spirals are discussed in the next section and log-periodics are treated in the following section.

7.7 SPIRAL ANTENNAS

Spiral antennas and their variations are usually constructed to be either exactly or nearly self-complementary. This yields extremely wide bandwidths of up to 40:1. Historically, the equiangular spiral was invented first, so we begin our discussions with it [22], [H.6: *Ant. Eng. Hdbk.*, 4th ed., Chap. 13].

7.7.1   Equiangular Spiral Antennas

The equiangular spiral curve shown in Fig. 7-28 is represented by the generating equation

$$r = r_0 e^{a\phi} \qquad (7\text{-}54)$$

where r_0 is the radius for $\phi = 0$ and a is a constant controlling the flare rate of the spiral. The spiral of Fig. 7-28 is right-handed. Left-handed spirals can be generated using negative values of a, or by simply turning over the spiral of Fig. 7-28. The equiangular spiral curve is used to create the antenna of Fig. 7-29, which is referred to as the **planar equiangular spiral antenna**. The four edges of the metallic region each have an equation for their curves of the form in (7-54). In particular, edge 1 is exactly that of Fig. 7-28, so $r_1 = r_0 e^{a\phi}$. Edge 2 has the same spiral curve but rotated through the angle δ, so $r_2 = r_0 e^{a(\phi - \delta)}$. The other half of the antenna has edges that make the structure symmetric; that is, rotating one spiral arm by one-half turn brings it into congruence with the other arm. So, $r_3 = r_0 e^{a(\phi - \pi)}$ and $r_4 = r_0 e^{a(\phi - \pi - \delta)}$. The structure of Fig. 7-29 is self-complementary, so $\delta = \pi/2$. It does not have to be constructed this way, but pattern symmetry is best for the self-complementary case.

The impedance, pattern, and polarization of the planar equiangular spiral antenna remain nearly constant over a wide range of frequencies. The feed point at the center, the overall radius, and the flare rate affect the performance. The flare rate a is more conveniently represented through *expansion ratio* ε, which is the increase factor of the radius for one turn of the spiral:
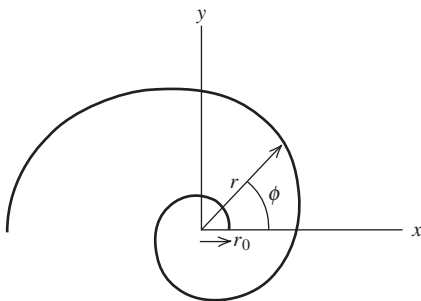
Figure 7-28 Equiangular spiral curve with $r = r_0 e^{a\phi}$ and $r_0 = 0.311$ cm and $a = 0.221$.
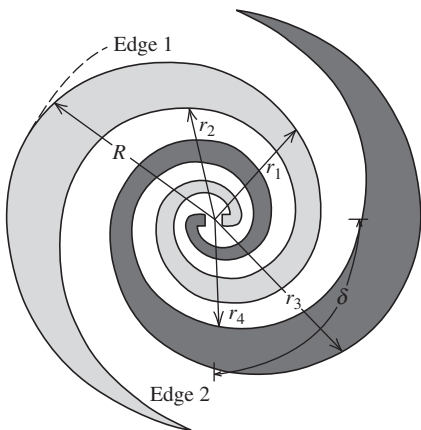
Figure 7-29 Planar equiangular spiral antenna for the self-complementary case with $\delta = 90°$.

$$\varepsilon = \frac{r(\phi + 2\pi)}{r(\phi)} = \frac{r_o e^{a(\phi+2\pi)}}{r_0 e^{a\phi}} = e^{a2\pi} \qquad (7\text{-}55)$$

A typical value for ε is 4, and then from (7-55), $a = 0.221$. The frequency at the upper end of the operating band f_U is determined by the feed structure. The minimum radius r_0 is about a quarter-wavelength at f_U for an expansion ratio ε of 4 [H.8.3: Rumsey]. A nearly equivalent criterion is a circumference in the feed region of $2\pi r_0 = \lambda_U = c/f_U$. Of course, the spiral terminates at this point and is connected to the feed transmission line. The low-frequency limit is set by overall radius R, which is roughly a quarter-wavelength at f_L. Alternatively, the circumference of a circle just enclosing the spiral can be used to set the low-frequency limit through $C = 2\pi R = \lambda_L$.

Spirals with one-half to three turns have been found experimentally to be relatively insensitive to the parameters a and δ [22]. One-and-one-half turns is about optimum. For example, again consider a one-and-one-half turn spiral with $a = 0.221$ as shown in Fig. 7-29. Each edge curve is of the form in (7-55), so the maximum radius is $R = r(\phi = 3\pi) = r_0 e^{0.221(3\pi)} = 8.03 r_0$. This equals $\lambda_L/4$, where λ_L is the wavelength at the lower band edge frequency. At the feed point, $r = r(\phi = 0) = r_0 e^0 = r_0$, and this equals $\lambda_U/4$ where λ_U is the wavelength at the upper band edge. The bandwidth is then $f_U/f_L = \lambda_L/\lambda_U = \lambda_L/4 / \lambda_U/4 = 8.03$. This 8:1 bandwidth is typical; however, bandwidths of 40:1 can be obtained.

Based on (7-53), the self-complementary equiangular spiral should have an input impedance value of $188.5 + j0$ Ω. In practice, the measured impedance values tend to be lower than this (about $120 + j0$ Ω), due to the finite thickness of the metallization and the presence of the coaxial feed line that is wound along one arm toward the feed at the center [22]. A feed of this type is referred to as an *infinite balun*. The balance function arises because any currents that are excited on the outside of the coax travel out from the feed point at the center, acting essentially like the currents on the arm and radiating upon reaching the active region. To maintain symmetry, a dummy coax is often attached to the second arm.

The radiation pattern of the self-complementary planar equiangular spiral antenna is bidirectional with two wide beams broadside to the plane of the spiral. The field pattern is approximately $\cos\theta$ when the z-axis is normal to the plane of the spiral. The half-power beamwidth is, thus, approximately $90°$. The polarization of the radiation is close to circular over wide angles, out to as far as $70°$ from broadside. The sense of the polarization is determined by the sense of the flare of the spiral. For example, the spiral of Fig. 7-29 radiates in the right-hand sense for directions out of the page and left-hand sense for directions into the page.

7.7.2 Archimedean Spiral Antennas

Another form of the planar spiral is the **Archimedean spiral antenna** shown in Fig. 7-30. This antenna, as are many spiral antennas, is easily constructed using printed circuit techniques. The equations of the two spirals in Fig. 7-30 are

$$r = r_0 \phi \qquad (7\text{-}56)$$

$$r = r_0(\phi - \pi) \qquad (7\text{-}57)$$

The Archimedean spiral is linearly proportional to the polar angle rather than an exponential for the equiangular spiral, and thus flares much more slowly.

The simple geometry of the Archimedean spiral antenna affords an opportunity to explain an important operating principle in frequency-independent antennas. This is the "band" description of radiation that is characterized by an active region responsible for radiation. Between the feed point of a frequency-independent antenna and the active region, currents exist in a transmission line mode and fields arising from them cancel in the far field. The active region occurs on that portion of the antenna that is one wavelength in
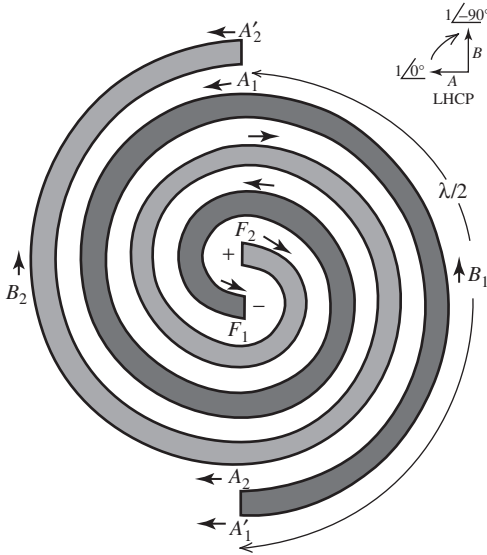
Figure 7-30 The Archimedean spiral antenna. The outside circumference in this case is one wavelength and thus is the active region. The inset is a vector diagram for the radiated electric fields, showing that the outward radiation is left-hand circularly polarized.

circumference for curved structures or has half-wavelength-long elements in antennas with straight wires or edges. Beyond the active region, currents are small, having lost power to radiation in the active region, and the antenna effectively behaves as if it is infinite in extent. Of course, the active region moves around the antenna with frequency. Since the geometry of a spiral is smooth, as frequency is reduced and the active region shifts to locations farther out on the spiral, the electrical performance remains unchanged. Hence, self-scaling occurs and frequency-independent behavior results.

We now give a physical explanation of how spiral antennas operate using Fig. 7-30. The arms are fed $180°$ out-of-phase at points F_1 and F_2. This is represented with oppositely directed current arrows. The current is inward for arm no. $1(-)$ and outward for arm no. $2(+)$. The lengths of the arms out to A, F_1A_1 and F_2A_2, are equal, so the phase shifts from the feed to A are identical, preserving the current directions as shown in Fig. 7-30. The active region where the circumference is one wavelength contains points labeled with an A or B. It can be assumed that the current magnitudes over this region are nearly the same. The phase, however, shifts as the traveling waves progress along the arms. Since the circumference is electrically large in the active region, phase must be accounted for. The phase shifts $180°$ between A_1 and A_1' and between A_2 and A_2', because of the $\lambda/2$ differential path length. Adjacent points on different arms (A_1, A_2' and A_2, A_1') are now in phase because the $180°$ phase shift counters the direction reversal introduced by the half-turn. In addition, the points opposite these pairs are in phase; that is, A_1, A_2' are in phase with A_2, A_1'. This in-phase condition leads to reinforcement of electric fields in the broadside direction, giving a radiation maximum. Interior to the active region, the electrical distance along different arms to adjacent points is not electrically large, preserving the antiphase condition due to the excitation. This is a transmission line mode and radiation is low. Often, resistive loads are added to the ends of the spiral to prevent reflection of the remaining traveling waves.

The final aspect that requires explanation is the circular polarization property. In the active region, points that are one-quarter turn around the spiral are $90°$ out of phase. For example, the phase at point B_1 lags that at point A_1 by $90°$. In addition, the currents are orthogonal in space. The current magnitudes are also nearly equal. Thus, all conditions are satisfied for circular polarized radiation: The radiated fields (created by the currents) are orthogonal, equal in magnitude, and $90°$ out of phase. As indicated by the vector diagram insert in Fig. 7-30, the wave is left-hand circular polarized. The left-hand sense results from the left-hand winding of the spiral. This is for radiation out of the page. Viewed from the other side of the page, the spiral is right-hand wound and thus produces RHCP.

Based on the above discussion, it is apparent that the spiral produces a broad main beam perpendicular to the plane of the spiral. Most applications require a unidirectional beam. This is created by backing the spiral with a ground plane. The most common construction approach is to use a metallic cavity behind the spiral, forming a **cavity-backed Archimedean spiral antenna**. This introduces a fixed physical length (the distance to the ground plane), thereby altering the true frequency-independent behavior. This is corrected in most commercial units by loading the cavity with absorbing material to reduce resonance effects; this, however, introduces loss. Typical performance parameter values for the cavity-backed Archimedean spiral are HP = 75°, |AR| = 1 dB, G = 5 dB over a 10:1 bandwidth or more. The input impedance is about 120 Ω, and is nearly real. The performance of the equiangular spiral is similar to that for the Archimedean spiral.

Very wideband antennas such as spirals are balanced structures. They are normally connected to a coaxial cable, which is an unbalanced structure. Therefore, a balun must be included with the spiral feed; see Sec. 6.4.3 for discussions of balun principles. One such wideband balun is the *tapered-coax wideband balun* that is formed by gradually cutting away the outer conductor of the coaxial cable, leaving a parallel wire line to attach to the spiral arms [23] as shown in Fig. 6-27*b*.

The pattern of the cavity-backed Archimedean spiral can be modeled by the following empirically derived function [24]:

$$F(\theta) = \cos^{5.8}(0.53 \ \theta) \tag{7-58}$$

Evaluating this pattern for the half-power points leads to a beamwidth of HP = 74°.

Fig. 7-31*a* shows an experimental model that has a diameter of 5.4 cm. The low frequency cutoff for this spiral antenna occurs when the circumference is one wavelength leading to:

$$c/\lambda = c/\pi D = 30/5.4 \, \pi = 1.77 \, \text{GHz} \tag{7-59}$$

The measured patterns are shown in Fig. 7-31*b* for three frequencies: 2.5 (slightly above cutoff), 5, and 10 GHz. Also plotted is the pattern model of (7-58). Note that the pattern remains nearly constant over a 4:1 bandwidth, characteristic of frequency-independent antennas. Representative measured antenna patterns for a cavity-backed spiral are shown in Fig. 13-13.

7.7.3   Conical Equiangular Spiral Antennas

Nonplanar forms of spiral antennas are used to produce a single main beam, thereby avoiding a backing cavity. For example, the planar equiangular spiral antenna conformed to a conical surface forms the **conical equiangular spiral antenna** shown in Fig. 7-32. The equation for a conical equiangular spiral curve is

$$r = e^{(a \sin \theta_h)\phi} \tag{7-60}$$

The planar spiral is a special case of this with $\theta_h = 90°$. The equations for the edges of one spiral of metal are that of (7-60) for r_1 and $r_2 = e^{(a \sin \theta_h)(\phi - \delta)}$, and $\delta = \pi/2$ for the self-complementary case. The other spiral arm is produced by a 180° rotation. The edges of the arms maintain a constant angle α with a radial line for any cone half-angle θ_h [H.8.3: Rumsey].

$$a = \cot \alpha \tag{7-61}$$

The conical equiangular spiral antenna has a single main beam that is directed off the cone tip in the $-z$-direction. A self-complementary shape yields the best radiation

(a) The spiral (reduced in size).

(b) Radiation patterns: measured (dashed curve) and computed (solid curve) using $\cos^{5.8}(0.53\theta)$.
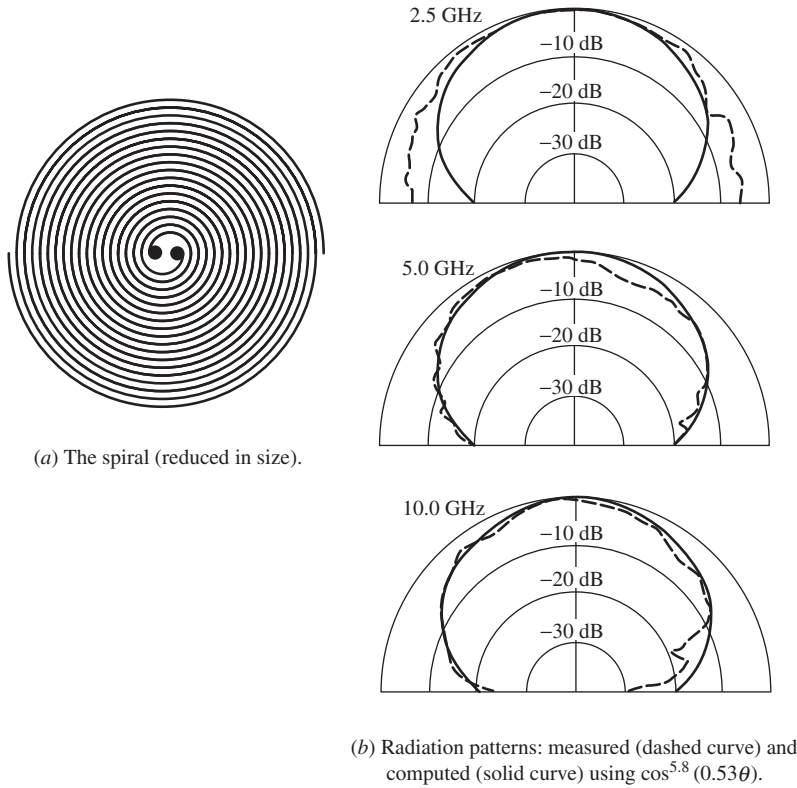
Figure 7-31 A 4:1 bandwidth cavity-backed Archimedean spiral antenna.

patterns. Typical patterns for $\theta_h \leq 15°$ and α about $70°$ have a broad main beam with a maximum in the $\theta = 180°$ direction and a half-power beam width of about $80°$. Since the structure is rotationally symmetric, the pattern is also nearly rotationally symmetric. The polarization of the radiated field is very close to circular in all directions, with the sense determined by the sense of the spiral winding. However, the polarization ellipticity increases with the off-axis angle.

The impedance [H.8.3: Rumsey] can be approximated by the relation

$$Z_o \approx 300 - 1.5\delta(\text{degrees}) \ \Omega \tag{7-62}$$

where $\delta(\text{degrees})$ is the angle δ of Fig. 7-32 in degrees. For the self-complementary case, δ is $90°$ and from (7-61) $Z_o \approx 165 \ \Omega$, which is close to the 188.5-Ω theoretical value of (7-53). The impedance is not affected significantly by θ_h or α.

The design of the conical equiangular spiral antenna is rather simple and proceeds as follows [25]. The upper frequency f_U of the operating band occurs when the truncated apex diameter is a quarter-wavelength; that is, $\lambda_U/4$. The lower band edge frequency f_L is determined by the base diameter B and occurs for $B = 3\lambda_L/8$. θ_h is usually less than $15°$ and α about $70°$. For $\theta_h = 10°$ and $\alpha = 73°$, the front-to-back ratio is 15 dB and the polarization axial ratio on the axis is below 3 dB.

7.7.4   Related Configurations

Spiral antennas can be operated in other than the fundamental odd mode described above, where the feed point terminals are excited in antiphase. Higher-order modes are possible and have an active region where the circumference is an odd multiple (3, 5, . . .) of a
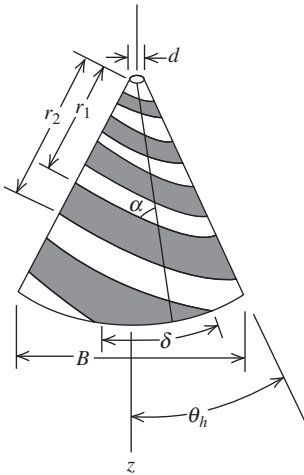
Figure 7-32 The conical equiangular spiral antenna.

Figure 7-33 The sinuous antenna.

wavelength. Even modes can be created by feeding the arms in-phase, leading to a null broadside to the spiral. This pattern is useful in direction finding (DF) [H.11.4: Lipsky]

Archimedean spirals with square arms provide a size reduction of 22%, which is based on the increase in perimeter from circular with πD compared to $4D$ for square. Also, spiral arms can be zigzagged [26], coiled, or loaded [H.9.1: Waterhouse, Chap. 4] to extend bandwidth. Additional material on spiral antennas can be found in the following references: [H.8.3: Corzine and Mosko], [H.8.3: Nakano], [H.8.6: *Ant. Eng. Hdbk*. 4th, ed., Chap. 13].

A broadband antenna that is related to the spiral is the **sinuous antenna** [25]. As can be seen from Fig. 7-33, the sinuous antenna is more complicated than the spiral antenna. However, it offers more flexible polarization uses. Two opposite arm pairs produce orthogonal linear polarizations. These pairs can be used separately for polarization diversity or for transmit/receive operation. Or, the two-arm-pair out-puts can be combined to produce simultaneous LHCP and RHCP. The operating principles are very similar to those for the planar spirals.

7.8 LOG-PERIODIC ANTENNAS

The frequency independent spiral antennas of the previous section illustrate the principle that emphasis on angles will lead to a broadband antenna. Although spiral antennas are
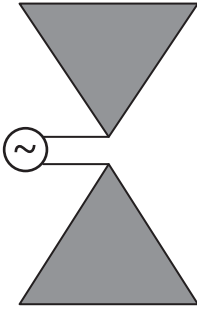
Figure 7-34 The bow-tie antenna.

not complex structures, construction would be simplified if simple geometries, involving circular or straight edges, could be utilized. Antennas of this type are discussed in this section. To see how the ideas develop, first consider the **bow-tie antenna** (also called the **bifin antenna**) of Fig. 7-34. It is the planar version of the finite biconical antenna (see Fig. 7-18). It has a bidirectional pattern with broad main beams perpendicular to the plane of the antenna. It is also linearly polarized. The bow-tie antenna is used as a receiving antenna for UHF TV channels, frequently with a wire grid ground plane behind it to reduce the back lobe. Since currents are abruptly terminated at the ends of the fins, the antenna has limited bandwidth. As we shall see shortly, by modifying the simple bow-tie antenna as shown in Fig. 7-35, the currents will then die off more rapidly with distance from the feed point. The introduction of periodically positioned teeth distinguishes this antenna as one of a broad class of log-periodic antennas. A **log-periodic antenna** is an antenna having a structural geometry such that its impedance and radiation characteristics repeat periodically as the logarithm of frequency. In practice, the variations over the frequency band of operation are minor, and log-periodic antennas are usually considered to be frequency-independent antennas.

Most of the work on frequency-independent antennas took place at the University of Illinois in the late 1950s and the 1960s [28]. A series of antennas were developed through many experiments. (For an excellent historical discussion of this evolution, see [H.3: Weeks, Sec. 7.2].) Several geometries were examined, and those that produced broadband behavior led to the determination of the properties necessary for wide bandwidth.

Frequency-independent spiral antennas were discussed in the previous section. In this section, we outline the development of the log-periodic antenna family. The metamorphosis of the log-periodic produced the log-periodic dipole antenna, which is made up of only straight wire segments.

One of the first log-periodic antennas was the **log-periodic toothed planar antenna** shown in Fig. 7-35. It is similar to the bow-tie antenna except for the teeth. The teeth act to disturb the currents that would flow if the antenna were of bow-tie-type construction. Currents flow out along the teeth and, except at the frequency limits, are not significant at the ends of the antenna. The rather unusual shape of this antenna is explained by examining the planar equiangular spiral antenna. Along a radial line from the center of the spiral, the positions of the far (or near) edges of a conductor from (7-54) are

$$r_n = r(\phi + n2\pi) = r_0 e^{a(\phi + n2\pi)} \tag{7-63}$$

The ratio of the $n + 1$th position to the nth position is

$$\frac{r_{n+1}}{r_n} = \frac{r_0 e^{a(\phi + (n+1)2\pi)}}{r_0 e^{a(\phi + n2\pi)}} = e^{a2\pi} = \varepsilon \tag{7-64}$$

which is the expansion ratio of (7-55). This is a constant, and thus the distances (or period) of the edges are of constant ratio for the planar spiral. For the structure of
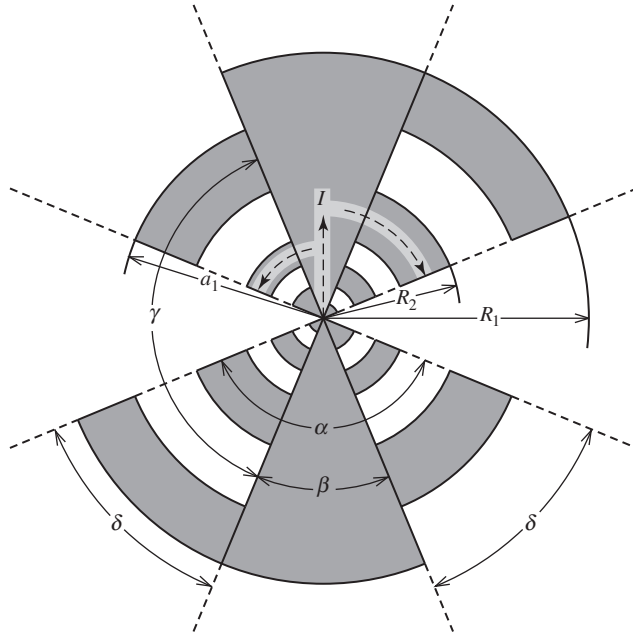
Figure 7-35 Log-periodic toothed planar antenna (self-complementary). Midband currents are shown on top half.

Fig. 7-35, the ratio of edge distances is also a constant and is given by the following scale factor:

$$\tau = \frac{R_{n+1}}{R_n} < 1 \tag{7-65}$$

The slot width is expressed by

$$\sigma = \frac{a_n}{R_n} < 1 \tag{7-66}$$

These relations are true for any n. The parameter τ gives the period of the structure. We would thus expect periodic pattern and impedance behavior with the same period. In other words, if frequencies f_{n+1} and f_n from adjacent periods lead to identical performance, then

$$\frac{f_n}{f_{n+1}} = \tau, \quad f_n < f_{n+1} \tag{7-67}$$

Forming $f_{n+1} = f_n/\tau$ from this equation and taking the logarithm of both sides, we have

$$\log f_{n+1} = \log f_n + \log(1/\tau) \tag{7-68}$$

Thus, the performance is periodic in a logarithmic fashion, hence the name log-periodic antenna. All log-periodic antennas have this property.

If the teeth sizes of the log-periodic toothed planar antenna are adjusted properly, the structure can be made self-complementary. From Fig. 7-35, we see that in general (whether self-complementary or not).

$$\gamma + \beta = 180° \quad \text{and} \quad \beta + 2\delta = \alpha \tag{7-69}$$

If the structure is self-complementary (as shown),

$$\alpha = \gamma \quad \text{and} \quad \beta = \delta \tag{7-70}$$

Substituting into (7-70) into (7-69) yields $\alpha + \beta = 180°$ and $\beta + 2\beta = \alpha$. Solving these two equations gives

$$\alpha = 135° \quad \text{and} \quad \beta = 45° \tag{7-71}$$

for a self-complementary log-periodic toothed planar antenna. As we saw in the previous section, an antenna that is self-complementary tends to be broadband and has an input impedance of 188.5 Ω.

If the widths of the teeth and gaps are equalized, $\sigma = a_n/R_n = R_{n+1}/a_n$. Using (7-65) and solving for σ give

$$\sigma = \sqrt{\tau} \tag{7-72}$$

This relationship and the self-complementary feature are popular in practice.

The properties of the log-periodic toothed planar antenna depend on τ. It has been found experimentally that the half-power beamwidth increases with increasing values of τ [H.8.3: Rumsey], increasing from about 30° at $\tau = 0.2$ to about 75° at $\tau = 0.9$. The pattern has two lobes with maxima in each normal direction to the plane of the antenna. The radiation is linearly polarized parallel to the teeth edges. This is perpendicular to what it would be if there were no teeth ($\delta = 0$), in which case the antenna would be a bow-tie. The fact that transverse current flow dominates over radial current flow is significant. Most of the current appears on teeth that are about a quarter-wavelength long (the active region). This, we have seen, is key to achieving wide bandwidths. The frequency limits of operation are set by the frequencies where the largest and smallest teeth are a quarter-wavelength long.

The log-periodic toothed planar antenna should have a performance (impedance and pattern) that repeats periodically with frequency with period τ given by (7-65). The self-complementary version of the antenna, although not producing frequency-independent operation, does lead to performance that does not vary greatly for frequencies between periods—that is, for $f_n < f < f_{n+1}$. In fact, measurements have produced nearly identical patterns over a 10:1 bandwidth [H.8.3: Rumsey].

The **log-periodic toothed wedge antenna** of Fig. 7-36 is a unidirectional pattern form of its planar version in Fig. 7-35, in which the included angle ψ is 180°. A single broad
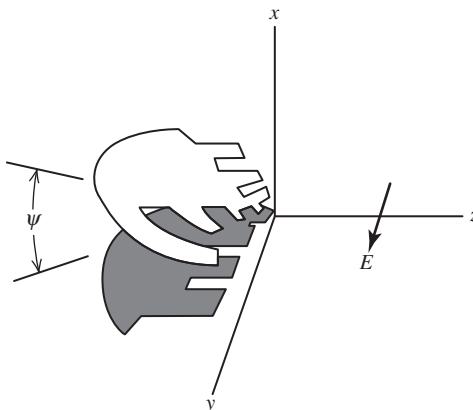
Figure 7-36 Log-periodic toothed wedge antenna.

main beam exists in the $+z$-direction. The patterns are nearly frequency-independent for $30° < \psi < 60°$. The polarization is linear and y-directed for an on-axis radiation, as indicated in Fig. 7-36. There is a small cross-polarized component (x-directed) arising from the radial current mode, as found in a biconical antenna. Typically, this cross-polarized component is 18 dB down from the copolarized (y-directed) component on-axis, indicating a strong excitation of the transverse current mode associated with frequency-independent behavior. The bandwidth of the wedge version is similar to the sheet version, but the input impedance is reduced for decreasing ψ. For the planar case ($\psi = 180°$), the self-complementary antenna, which should have an impedance of 188.5 Ω, has an impedance of about 165 Ω, whereas the wedge form with $\psi = 30°$ has a 70-Ω impedance. As ψ is decreased, the impedance variation over a period of the structure (frequency ratio of τ) increases. For example, a 3:1 variation occurs for $\psi = 60°$ relative to the geometric mean [H.8.3: Rumsey].

From a construction standpoint, it would be desirable if the toothed antennas could be made with straight edges. This simplification of the structure turns out to be of little consequence in the performance of the antenna. This is another major step in the development of the log-periodic antenna. As an example, if the tooth edges of the log-periodic toothed planar antenna in Fig. 7-35 are replaced by straight edges, the **log-periodic toothed trapezoid antenna** of Fig. 7-37 results. The performance of this antenna is similar to its curved edge version in Fig. 7-35. A **log-periodic toothed trapezoid wedge antenna** can be formed by bending the planar version into a wedge, creating an antenna similar to that of Fig. 7-36. In fact, the patterns of the two wedge forms (curved edge and trapezoid) are similar, but the trapezoid version has better impedance performance with only about a 1.6:1 variation over a period for $\psi = 60°$ [H.8.3: Rumsey].

The solid metal (or sheet) antennas we have described are practical for short wave-lengths, but for low frequencies the required structures can become rather impractical. It turns out that the sheet antennas can be replaced by a wire version in which thin wires are shaped to follow the edges of the sheet antenna. An example of this major structural simplification is that of Fig. 7-38a, which is the wire version of Fig. 7-37. This **log-periodic trapezoid wire antenna** can also be bent at the apex to form a wedge that produces a unidirectional pattern. The **log-periodic trapezoid wedge wire antenna** has a performance similar to its sheet version. Measurements for a wedge angle $\psi = 45°$ have yielded E- and H-plane half-power beamwidths of $66°$, a gain of 9.2 dB, and a front-to-back ratio of 12.3 dB. The average input impedance has been measured as 110 Ω with a VSWR of 1.45 over a 10 : 1 band [29]. As with other wedge log-periodics, the main beam maximum is straight off the apex and the radiation is linearly polarized.

Other even simpler log-periodic wire antennas exist in both planar and wedge shapes. The **log-periodic zig-zag wire antenna** of Fig. 7-38b is an example.

The final phase in this metamorphosis of log-periodic antennas is the use of only parallel wire segments. This is the **log-periodic dipole array** of Fig.7-39 [30, 31]. The log-periodic dipole array (LPDA) is a series-fed array of parallel wire dipoles of
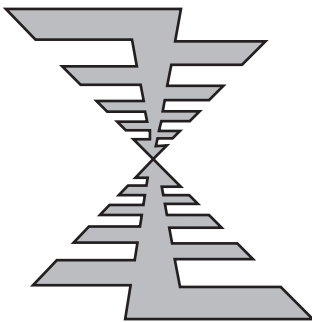
Figure 7-37 Log-periodic toothed trapezoid antenna.
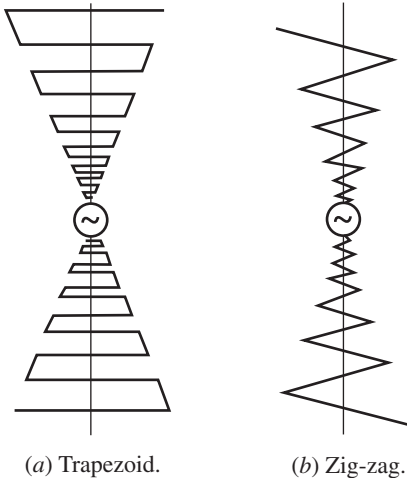
(*a*) Trapezoid.   (*b*) Zig-zag.   **Figure 7-38** Log-periodic wire antenna configurations.

successively increasing lengths outward from the feed point at the apex. Note that the interconnecting feed lines cross over between adjacent elements. This can be explained by noting that the LPDA of Fig. 7-39 resembles the toothed trapezoid of Fig. 7-37 when folded on itself, making a wedge with zero included angle. The two center fins of metal then form a parallel transmission line with the teeth coming out from them on alternate sides of the fins. This alternate arm geometry occurs for all wedge log-periodic antennas.

A particularly successful method of constructing an LPDA is shown in Fig. 7-40. A coaxial transmission line is run through the inside of one of the feed conductors. The outer conductor of the coax is attached to that conductor and the inner conductor of the coax is connected to the other conductor of the LPDA transmission line.

As shown in Fig. 7-38, a wedge of enclosed angle α bounds the dipole lengths. The scale factor τ for the LPDA is

$$\tau = \frac{R_{n+1}}{R_n} < 1 \tag{7-73}$$

Right triangles of enclosed angle $\alpha/2$ reveal that

$$\tan \frac{\alpha}{2} = \frac{L_n/2}{R_n} = \frac{L_{n+1}/2}{R_{n+1}} \tag{7-74}$$
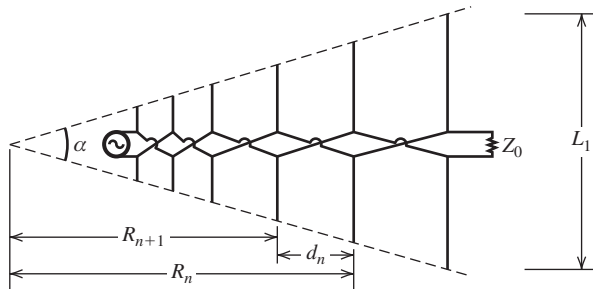
Figure 7-39 Log-periodic dipole array geometry.

Feed point
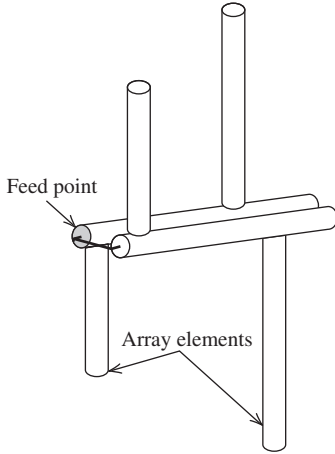
Array elements

Figure 7-40 Construction details of the log-periodic dipole array.

Thus,

$$\frac{L_1}{R_1} = \cdots = \frac{L_n}{R_n} = \frac{L_{n+1}}{R_{n+1}} = \cdots = \frac{L_N}{R_N} \tag{7-75}$$

Using this result in (7-73) gives

$$\tau = \frac{R_{n+1}}{R_n} = \frac{L_{n+1}}{L_n} \tag{7-76}$$

Thus, the ratio of successive element positions equals the ratio of successive dipole lengths.

The spacing factor for the LPDA is defined as

$$\sigma = \frac{d_n}{2L_n} \tag{7-77}$$

where the element spacings as shown in Fig. 7-39 are given by

$$d_n = R_n - R_{n+1} \tag{7-78}$$

But $R_{n+1} = \tau R_n$, so

$$d_n = R_n - \tau R_n = (1 - \tau)R_n \tag{7-79}$$

From (7-74), $R_n = L_n/2 \tan (\alpha/2)$. Using this in (7-79) yields

$$d_n = (1 - \tau)\frac{L_n}{2 \tan (\alpha/2)} \tag{7-80}$$

Substituting this in (7-77) gives

$$\sigma = \frac{d_n}{2L_n} = \frac{1 - \tau}{4 \tan (\alpha/2)} \tag{7-81}$$

or

$$\alpha = 2 \tan^{-1}\left(\frac{1 - \tau}{4\sigma}\right) \tag{7-82}$$

Combining (7-81) with (7-76), we note that all dimensions are scaled by

$$\tau = \frac{R_{n+1}}{R_n} = \frac{L_{n+1}}{L_n} = \frac{d_{n+1}}{d_n} \tag{7-83}$$

As we have seen with other log-periodic antennas, there is also an active region for the LPDA, where the few dipoles near the one that is a half-wavelength long support much more current than do the other radiating elements. It is convenient to view the LPDA operation as being similar to that of a Yagi–Uda antenna. The longer dipole behind the most active dipole (with largest current) behaves as a reflector and the adjacent shorter dipole in front acts as a director. The radiation is then off of the apex. The wedge enclosing the antenna forms an arrow pointing in the direction of the main beam maximum.

As the operating frequency changes, the active region shifts to a different portion of the antenna. The frequency limits of the operational band are roughly determined by the frequencies at which the longest and shortest dipoles are half-wave resonant—that is,

$$L_1 \approx \frac{\lambda_L}{2} \quad \text{and} \quad L_N \approx \frac{\lambda_U}{2} \tag{7-84}$$

where λ_L and λ_U are the wavelengths corresponding to the lower and upper frequency limits. Since the active region is not confined completely to one dipole, often dipoles are added to each end of the array to ensure adequate performance over the band. The number of additional dipoles required is a function of τ and σ [32], [H.8.3: Smith]. But for noncritical applications, (7-84) is sufficient.

The pattern, gain, and impedance of an LPDA depend on the design parameters τ and σ. Since the LPDA is a very popular broadband antenna featuring simple construction, low cost, and light weight, we will give the design details and illustrate them by examples. Gain contours are plotted in Fig. 7-41 as a function of τ and σ [31]. Note that high gain requires a large value of τ, which means a very slow expansion, that is, a LPDA of large overall length. Gain is only slightly affected by the dipole thickness. It increases about 0.2 dB for
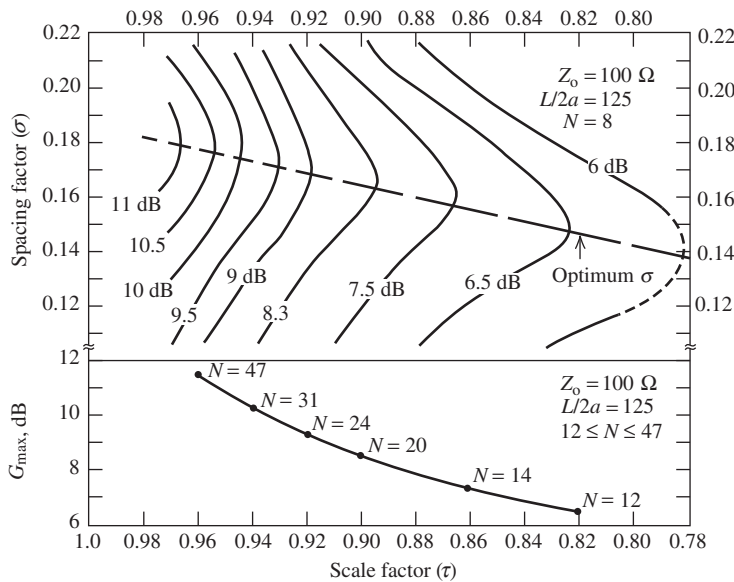
Figure 7-41 Gain of a log-periodic dipole array. (Contours, at top, adapted from Carrel [31]. Maximum gain curve, at bottom, derived from data in [32].)

a doubling of the thickness [H.8.3: Rumsey]. Gain is also affected by the feeder impedance [32] and tends to decrease as the feeder impedance is increased above $100\,\Omega$.

Fig. 7-41 (top portion) shows the variation in gain of an LPDA with σ and τ. The curves are a modification of those originally presented by Carrel [31] that have been found to have a gain that is erroneously high [32, 33, 34]. In [H.3: Balanis], Carrel's curves are reduced uniformly by 1 dB and in [35] uniformly by 1.5 dB. Based on data in [32, 34], it appears that Carrel's original curves were more in error (for gain) for lower values of τ than for higher values. Thus, the 11-dB and 6-dB contours in Fig. 7-41 are 1 dB and 2 dB lower, respectively, than those in Carrel.

The bottom portion of Fig. 7-41 shows a gain curve that is derived from data in [32] where N, the number of dipoles, varies from 12 to 47 (unlike Carrel's modified contours above for which $N = 8$). Notice that the value of $G_{\max}$ is greater than the value of the gain contour at the optimum σ line in the top portion of Fig. 7-41. The $G_{\max}$ vs. τ curve probably represents an upper bound on the LPDA gain that can be achieved in practice for feeder impedances of $100\,\Omega$ or greater.

Further details on the design and calculations for the LPDA are available in the literature [32–35]. Also, the LPDA can be constructed in a size-reduced form or by using printed circuit techniques [36–39].

EXAMPLE 7-3   *Optimum Design of a 54- to 216-MHz Log-Periodic Dipole Antenna*

An antenna that operates over the entire VHF-TV and FM broadcast bands, which span the 54- to 216-MHz frequency range for a 4:1 bandwidth, is desired. Suppose the design gain is chosen to be 6.5 dB. The corresponding values of τ and σ for optimum design from Fig. 7-40 are

$$\tau = 0.822 \quad \text{and} \quad \sigma = 0.149 \qquad (7\text{-}85)$$

Then from (7-82), we have

$$\alpha = 2 \tan^{-1} \left[\frac{1 - 0.822}{4(0.149)} \right] = 33.3° \qquad (7\text{-}86)$$

The length of the longest dipole is determined first. At the lowest frequency of operation (54 MHz), the dipole length from (7-84) should be near a half-wavelength, so

$$L_1 = 0.5\lambda_L = 0.5(5.55) = 2.78 \text{ m} \qquad (7\text{-}87)$$

The shortest dipole length should be on the order of $L_U = 0.5\lambda_U = 0.694$ m at 216 MHz. The LPDA element lengths are computed until a length on the order of 0.694 m is reached. To be specific, element lengths are found from L_1 using $L_{n+1} = \tau L_n$. For example,

$$L_2 = \tau L_1 = (0.822)(2.78) = 2.29 \text{ m}$$

and

$$L_3 = \tau L_2 = (0.822)(2.29) = 1.88 \text{ m}$$

Completing this process leads to

$$\begin{aligned} &L_1 = 2.78 \text{ m}, \quad L_2 = 2.29 \text{ m}, \quad L_3 = 2.78 \text{ m}, \quad L_4 = 1.54 \text{ m} \\ &L_5 = 1.27 \text{ m}, \quad L_6 = 1.04 \text{ m}, \quad L_7 = 0.858 \text{ m}, \quad L_8 = 0.705 \text{ m} \\ &L_9 = 0.579 \text{ m} \end{aligned} \qquad (7\text{-}88)$$

The array was terminated with nine elements since $L_9 = 0.579$ m is less than the 0.694 m length for the highest operating frequency. Elements could be added to either end to improve performance at the band edges.

The element spacings for this example are found from (7-81) as

$$d_n = 2\sigma L_n = 2(0.149)L_n = 0.298L_n \qquad (7\text{-}89)$$

Using the element lengths of (7-88) gives

$$
\begin{aligned}
d_1 &= 0.828 \text{ m}, \quad d_2 = 0.682 \text{ m}, \quad d_3 = 0.560 \text{ m}, \quad d_4 = 0.459 \text{ m} \\
d_5 &= 0.378 \text{ m}, \quad d_6 = 0.310 \text{ m}, \quad d_7 = 0.256 \text{ m}, \quad d_8 = 0.210 \text{ m}
\end{aligned}
\qquad (7\text{-}90)
$$

These dipole lengths and spacings completely specify the LPDA, as shown in Fig. 7-39. The total length of the array is the sum of the spacings in (7-90), which gives a 3.683 m. The outline of the antenna fits into an angular sector of angle $\alpha = 33.3°$.

EXAMPLE 7-4   *Characteristics of a 200- to 600-MHz LPDA*

In this example, we examine the gain, pattern, impedance, and current distribution of a LPDA as a function of frequency. Suppose it is to be constructed for operation over the 200- to 600-MHz band. For optimum performance and a design goal of 9 dB gain, we see from Fig. 7-41 that $\tau = 0.917$ and $\sigma = 0.169$. The lowest frequency of operation (200 MHz) has a wavelength of $\lambda_L = 1.5$ m, so the first element has a length of $L_1 = \lambda_L/2 = 0.75$ m. The length of the shortest element should be on the order of a half-wavelength at 600 MHz, and $\lambda_U/2 = 0.500 \text{ m}/2 = 0.250$ m. Using the design techniques illustrated in the previous example and four extra elements at the narrow end gives the 18-element LPDA shown in Fig. 7-42. (The antenna geometry details are left as a problem.)

The LPDA of Fig. 7-42 was modeled using the computer techniques of Sec. 14.10.2. The resulting dipole terminal currents at the band edges and one midway frequency are shown in Fig. 7-42. Also shown are the voltage amplitude distributions on the transmission line. These currents and voltages illustrate the active region behavior we have mentioned several times. For example, at 200 MHz there are three dipoles with strong currents on them and a total of five with significant currents. This is also true for other frequencies in the operating band, with the active region shifted to some other portion of the antenna as seen in Fig. 7-42. At the high-frequency limit, element 14 is about a half-wavelength long and the extra four elements provide support for the active region at 600 MHz.

The gain, pattern, and impedance behavior as a function of frequency are shown in Fig. 7-43. At 150 MHz, the gain is considerably less than the 9-dB design value due to the large back lobe. Also, the input impedance has a substantial imaginary part. This inferior performance is, of course, caused by insufficient antenna length required for proper support of the active region at that frequency. At the lower band edge of 200 MHz, however, the pattern has little back radiation, the gain is approaching the design goal, and the input impedance has a small imaginary part. Similarly, at 650 MHz the performance is only slightly inferior to that at the upper band edge of 600 MHz because of the added elements. At intermediate frequencies between the band edges, the gain, pattern, and impedance remain reasonably constant, indicating frequency-independent behavior. Fig. 7-43c and 7-43d are typical of intermediate frequencies. The fact that the calculated gain exceeds the 9 dB value indicated by the contour is due to a combination of at least three factors: (1) a different feeder impedance (below 100 ohms) is used, (2) the ratio of $L/2a$ being different than 125, (3) the estimated gain reduction applied here to Carrel's original contours being, perhaps, slightly excessive. Each of these three factors could conceivably account for 0.1–0.2 dB.

The termination used in this example is purely resistive and equal to the characteristic impedance of the transmission line, but reactive terminations can also be used. The use of a reactive termination can lead to unwanted resonances on the LPDA caused by energy being trapped between the termination and the stop region on the termination side of the active region. These high Q resonances can be eliminated by using a termination that is at least slightly resistive or by using a relatively high value for the LPDA transmission line impedance
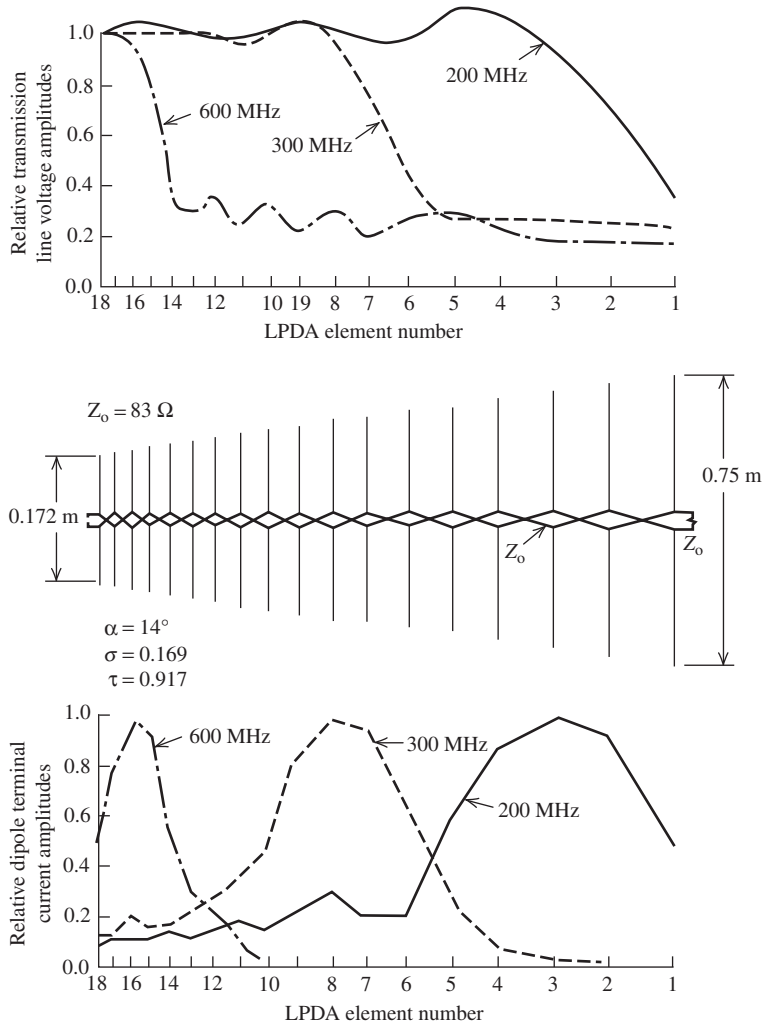
Figure 7-42 An optimum log-periodic dipole antenna for operation in the 200- to 600-MHz band (Example 7-3). (Top) Voltage distribution on the 83-Ω transmission line. (Middle) The geometry. (Bottom) Relative dipole terminal current amplitudes.

(e.g., $Z_o \sim 150 \ \Omega$) since this will cause the dipole elements to more heavily load the line. This makes the active region more efficient [39] with the result that there is relatively little energy left to propagate past the active region and cause a strong resonance effect on the radiation pattern.

7.9 WIDEBAND EMC ANTENNAS

Electromagnetic compatibility (EMC) testing divides into two major categories: emission testing (i.e., unwanted transmitted radiation) and susceptibility or immunity testing (i.e., undesirable received radiation). The antenna requirements for each are not identical. For emission testing, the EMC antenna should be broadband, have a broad radiation pattern (i.e., low directivity), have a low (<2 : 1) VSWR, and be capable of receiving electric
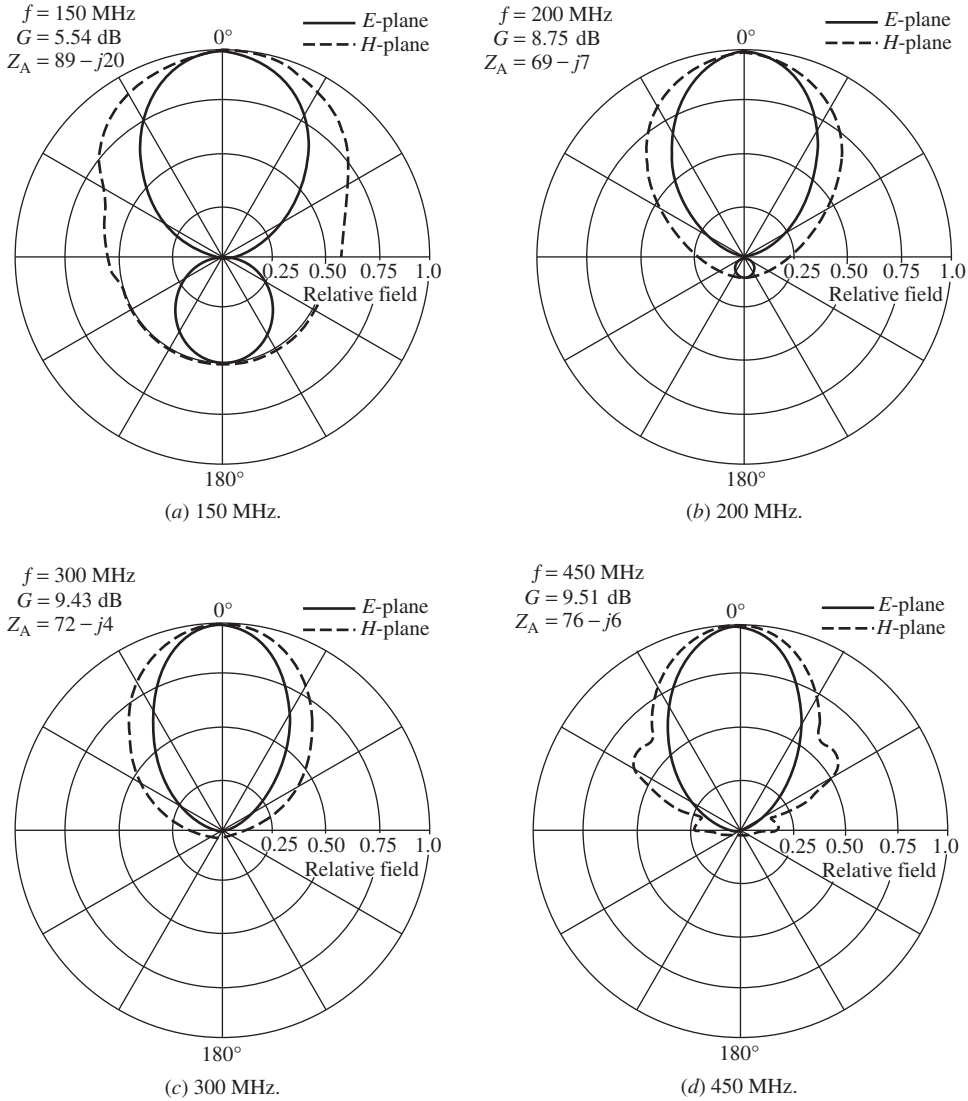
Figure 7-43 Radiation patterns at several frequencies for the log-periodic dipole antenna of Example 7-3. The gain and impedance values are also given. Calculations were done by the rigorous moment method formulation in Sec. 14.10.2.

and magnetic fields of any polarization. For susceptibility testing, the antenna should also be capable of handling high power and have a directive pattern.

EMC standards tend to originate from the European-based CISPR (Comite International Speciale des Perturbations Radio-Electroniques) and when adopted by governments become national standards. Government agencies like the FCC in the United States sometimes adopt modifications of these standards to suit their own national needs. For example, CISPR requires that radiated emission measurements from Class B devices (non-commercial devices) be conducted at a distance of 10 m from the DUT (device under test), while the FCC standard specifies 3 m as the measuring distance. Note that at 3 m it is likely that neither the DUT nor the EMC antenna will be in the far field of each

f = 600 MHz
G = 9.37 dB
$Z_A = 78 - j11$

——— E-plane
– – – H-plane

0°

0.25 0.50 0.75 1.0
Relative field

180°

(e) 600 MHz.

f = 650 MHz
G = 8.98 dB
$Z_A = 71 - j27$

——— E-plane
– – – · H-plane

0°

0.25 0.50 0.75 1.0
Relative field

180°

(f) 650 MHz.
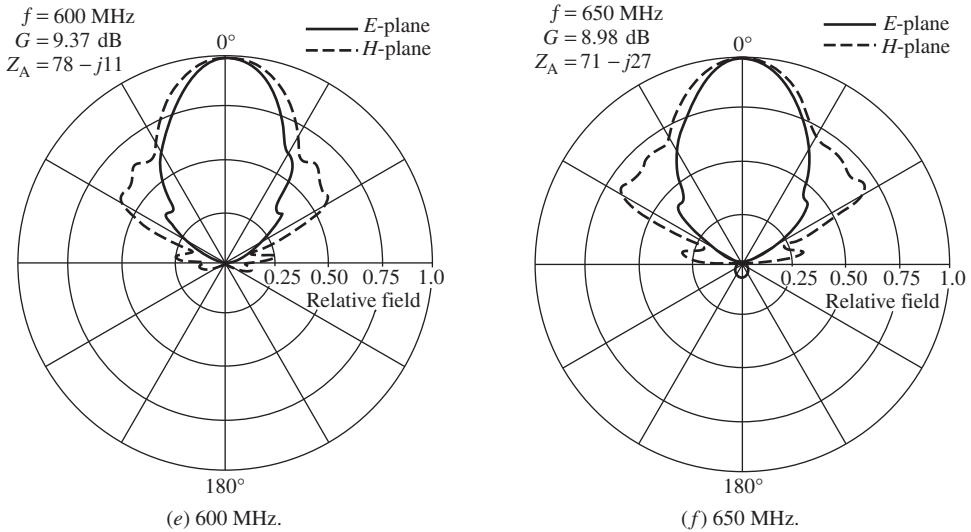
Figure 7-43 (Continued)

Table 7-3 CISPR Requirements for a Broadband Antenna

Frequency	30–1000 MHz
Polarization	Substantially linearly polarized
Cross-pol	20 dB below aligned polarization
Polar pattern	Direct and reflected ray response within 1 dB (i.e., broad antenna pattern)
VSWR	Less than 2:1

other. This has obvious negative implications for the measurements that are partly overcome by the use of a standard measuring distance.

CISPR Class B measurement standards officially require a set of tuned dipoles to cover the emission testing frequency range but allow the use of broadband antennas under most conditions. Not using tuned dipoles, however, requires the use of a calibrated antenna factor. The antenna factor was discussed in Sec. 4.7. The CISPR requirements for a broadband antenna include those in Table 7-3. The CISPR 2:1 VSWR requirement is difficult to achieve at the lowest frequencies without the use of an attenuator inserted at the antenna terminals.

Emission testing is performed from 30 MHz to at least 1 GHz. The finite biconical (Sec. 7.4.2) and the log-periodic (Sec. 7.8) are commonly used to cover the frequency range in two steps with the biconical covering $30-300$ MHz and the log-periodic covering $300-1000$ MHz. These two antennas have been combined into an antenna commercially marketed as a Bilog antenna. There are several possible configurations for the biconical portion of the Bilog, one of which is shown in Fig. 7-44. A typical antenna factor for the Bilog is shown in Fig. 7-45. Ideally the antenna factor should have low magnitude, no sharp resonances, and minimum sensitivity to external factors. For the antenna in Fig. 7-44, the VSWR and gain both deteriorate markedly from 60 MHz down to 30 MHz. But this is not a problem for EMC measurements, because the antenna factor is known.
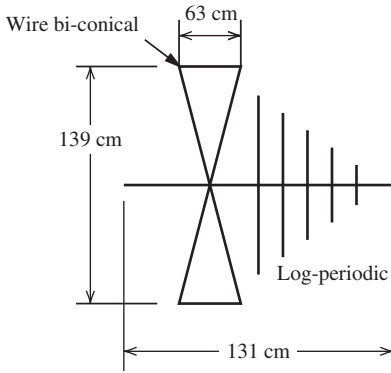
Figure 7-44 Bilog antenna. (Courtesy of TESEQ, Inc. Used with permission.)
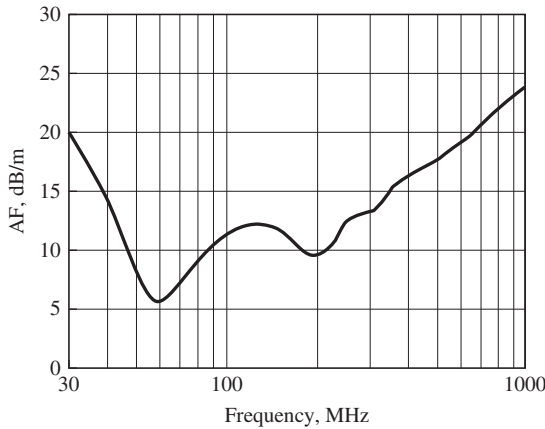
Figure 7-45 Antenna factor for the Bilog antenna in Fig. 7-44. (Courtesy of TESEQ, Inc. Used with permission.).

7.10 ULTRA-WIDEBAND ANTENNAS

Ultra-wideband (UWB) systems have very wide bandwidths with applications in radar and communications. A UWB communication system has a bandwidth much larger than necessary to support the information being carried [38]. UWB communications is well suited for wireless personal area networks (see Sec. 4.5.1) for short-range distribution of video, data, music, and so on. Non-communication applications include imaging systems (ground-penetrating radar, through-wall radar, surveillance, and medical imaging) and vehicular radar. UWB systems are of two types: multicarrier (MC-UWB) and impulse (I-UWB). MC-UWB divides the spectrum into many sub-bands, each with a carrier, and the antennas can be analyzed using conventional frequency domain methods. I-UWB uses impulses that spread the signal over many octaves of bandwidth, requiring the use of time-domain techniques and an antenna with sufficient instantaneous bandwidth. An illustration of the spreading in the frequency domain due to a short duration time pulse is the Fourier transform of a perfect impulse as shown in Table 5-3, sixth entry.

The recent expansion of commercial UWB systems in the United States is due to the FCC allocating frequency bands for UWB, including 3.1 to 10.6 GHz. The bandwidth (the band edges are defined in this case as the frequencies where the EIRP is -10 dB down from the center frequency value) must occupy 500 MHz or 20% of its center frequency, whichever is greater. No license is required but emissions must not exceed an EIRP in 1 MHz of -41.3 dBm. This low-power level limits the application to short-range systems but reduces the potential of interfering with other systems sharing the same frequencies. Conventional communication systems, of course, have less bandwidth and transmit much greater power levels.

Military systems also use UWB technology in communications and radar, as well as a special application called high-power microwave (HPM). HPM uses a directed energy source employing a very narrow pulse at very high power levels (i.e., megawatts). The antennas used, called *impulse radiating antennas* (IRA), are capable of radiating waveforms with a flat frequency spectrum from 35 MHz to 3.5 GHz, which is a 100 : 1 frequency ratio [39]. Dissipative materials are sometimes used in the IRA to broaden bandwidth.

Impulse-UWB systems should have a pulse signal that undergoes small distortion in the time domain. This requires the ideal UWB antenna to produce radiated fields that have constant magnitude with frequency and phase that varies linearly with frequency. This is accomplished with an antenna that produces an output that is the time derivative of the input [38, p. 13]. Such a non-dispersive system will produce an undistorted replica of the source pulse waveform at the receiver output and do so at angles extending over the half-power beamwidth of the pattern main beam (see. Fig. 15-31).

The *phase center* plays an import role in I-UWB antennas. It is the effective center of a sphere in the far field over which the phase is constant. To radiate without dispersion, an antenna must have a phase center which remains constant with frequency and have constant impedance and pattern over the bandwidth [H.8.3: Schantz, p. 42]. Many antennas do not have a unique phase center. Broadband antennas often have a phase center that moves in location with frequency.

There are three types of UWB antennas [H.6: Balanis, 4th ed., Chap. 19]:

Type I. Variations of dipole/monopole antennas
  Broadband dipoles/monopoles and biconical-family antennas can achieve 3:1 bandwidth.
Type II. Antennas with absorptive loading or with curved geometries
  Diffractions are controlled through absorption or curved edges. Absorptive antennas are suited for impulse UWB, but are lossy. Example curved-geometry antennas are Vivaldi antennas (see Fig. 8-35*b*) and horns with rolled edges to achieve gentle, continuously curved geometries.
Type III. Frequency-independent antennas
  These antennas have an active region that maintains its shape and electrical size with frequency change over the operating band. Examples are the LPDA, spiral, and sinuous antennas discussed earlier in this chapter. However, the phase center of these antennas moves with the active region, introducing dispersion and making the antennas inappropriate for I-UWB.

Fig. 7-46 shows the evolution of compact UWB antennas suitable for time-domain applications [40, 41]. The *spherical antenna* of Fig. 7-46*a* and its variations have been used for decades for broadband applications. It emphasizes curved geometry of Type II UWB antennas. The antenna height is one-quarter wavelength at the lowest operating frequency. The VSWR for a sphere of 2.5 cm in height on a 2.8-cm diameter circular ground plane is less that 2:1 from 3 to over 12 GHz. The phase is linear with frequency and there is nearly omnidirectional, good impulse response over this range of frequencies. Flattening the sphere into a disk to reduce antenna volume produces the *circular disk antenna* of Fig. 7-46*b* that has electrical performance which is only slightly degraded from the spherical antenna. Further size reduction is achieved by halving the disk to form the *half-circular disk antenna* of Fig. 7-46*c*. The performance is again nearly the same as its predecessors except the low-frequency end of the band is increased from 3 to 4.2 GHz for VSWR = 2. The most compact UWB antenna to date is the *compact UWB antenna* of Fig. 7-46*d*. The measured VSWR for a 3.2-cm diameter enclosing hemisphere on a 30-cm ground plane was about 2:1 from 2 to 18 GHz. [41] Time-domain performance is good as well. Details on other UWB antennas are given in the following references: [40, Chap. 4], [41, 42, 45], [H.8.3: Volakis, Sec. 3.2.8] and [H.8.3: Schantz, Chap. 6].

(*a*) Spherical antenna.

(*b*) Circular disk
antenna
(side view).

(*c*) Half circular
disk antenna
(side view).

(*d*) Compact UWB
antenna.
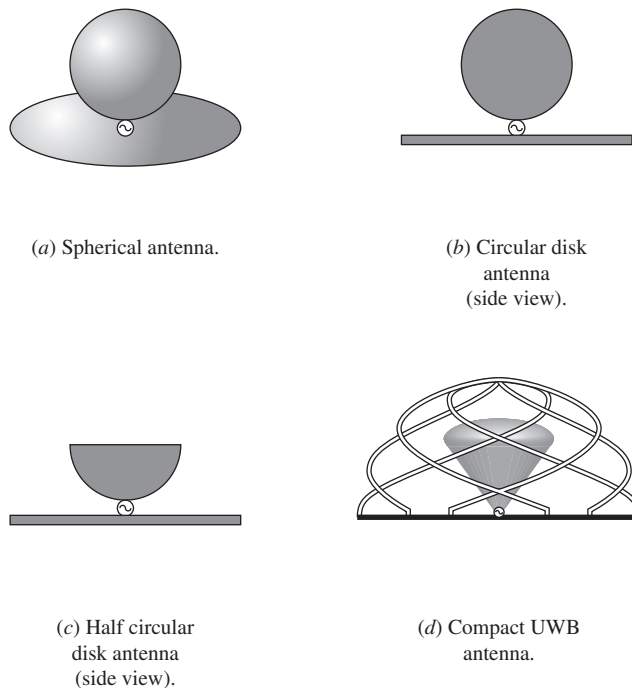
Figure 7-46 Evolution of antennas suitable for UWB applications.

REFERENCES

1. E. K. Miller et al., "On Increasing the Radiation Efficiency of Beverage-Type Antennas," Lawrence Livermore National Lab, UCRL-52300, July 1977.
2. G. J. Burke et al., "Computer Modeling of Antennas near the Ground," Lawrence Livermore National Lab, UCID-18626, May 1980.
3. T. S. Maclean and R. Kouyoumjian, "The Bandwidth of Helical Antennas," *IEEE Trans. on Ant. and Prop.*, Vol. AP-7, pp. S379–S386, Dec. 1959.
4. D. Emerson, "The Gain of an Axial-Mode Helix Antenna," *ARRL Antenna Compendium*, Vol. 4, pp. 64–68, 1996.
5. K. F. Lee and P. F. Wong, "Directivities of Helical Antennas Radiating in the Axial Mode," *IEE Proc.*, Part H, Vol. 131, pp. 121–122, April 1984.
6. H. E. King and J. L. Wong, "Characteristics of 1 to 8 Wavelength Uniform Helical Antennas," *IEEE Trans. on Ant. and Prop.*, Vol. AP-28, pp. 291–296, March 1980.
7. K. R. Carver, "The Helicone: A Circularly Polarized Antenna with Low Side Lobe Level," *Proc. IEEE*, Vol. 55, p. 559, April 1967.
8. R. G. Vaughan and J. Bach Anderson, "Polarization Properties of the Axial Mode Helix Antenna," *IEEE Trans. on Ant. and Prop.*, Vol. AP-33, pp. 10–20, Jan. 1985.
9. A Safaai-Jazi and J. C. Cardoso, "Radiation Characteristics of a Spherical Helical Antenna," *IEE Proc.—Microwave, Antennas, & Propagation*, Vol. 143, pp. 7–12, Feb. 1996.
10. R. M. Barts and W. Stutzman, "Stub Loaded Helix Antenna," U.S. Patent No. 5,986,621, Nov. 16, 1999.
11. G. H. Brown and O. M. Woodward, "Experimentally Determined Radiation Characteristics of Conical and Triangular Antennas," *RCA Review*, Vol. 13, pp. 425–452, Dec. 1952.
12. M. C. Bailey, "Broad-Band Half-Wave Dipole," *IEEE Trans. on Ant. and Prop.*, Vol. AP-32, pp. 410–412, April 1984.
13. A. G. Kandoian, "Three New Antenna Types and Their Applications," *Proc. IRE*, Vol. 34, pp. 70W–75W, Feb. 1946.
14. J. J. Nail, "Designing Discone Antennas," *Electronics*, Vol. 26, pp. 167–169, Aug. 1953.

15. T. H. Crowley and W. Marsh, "Discone Type Antennas," Ohio State Research Foundation Tech. Rept. 510–13, June 30, 1954.

16. T. Rappaport, "Discone Design Using Simple *N*-Connector Feed," *IEEE on Ant. and Prop. Newsletter*, Feb. 1988.

17. C. E. Smith, C. M. Butler, and K. Umashankar, "Characteristics of a Wire Biconical Antenna," *Microwave J.*, Vol. 22, pp. 37–40, Sept. 1979.

18. O. Givati and A. Fourie, "Analysis of Skeletal Wire Conical Antennas," *IEEE Trans. on Ant. and Prop.*, Vol. 44, pp. 844–858, June 1996.

19. A. J. Poggio and P. E. Mayes, "Pattern Bandwidth Optimization of the Sleeve Monopole Antenna," *IEEE Trans. on Ant. and Prop.*, Vol. AP-14, pp. 643–645, Sept. 1966.

20. H. E. King and J. L. Wong, "An Experimental Study of a Balun-Fed, Open-Sleeve Dipole in Front of a Metallic Reflector," Aerospace Corp. Rept. No. TR-0172 (2158)-2, Feb. 1972.

21. J. L. Wong and H. E. King, "An Experimental Study of a Balun-Fed Open Sleeve Dipole in Front of a Metallic Reflector," *IEEE Trans. on Ant. and Prop.*, Vol. AP-20, March 1972.

22. J. D. Dyson, "The Equiangular Spiral Antenna," *IRE Trans. on Ant. and Prop.*, Vol. AP-7, pp. 181–187, April 1959.

23. J. W. Duncan and V. P. Minerava, "100-1 Bandwidth Balun Transformer," *Proc. IRE*, Vol. 44, pp. 31–35, Jan. 1960.

24. D. G. Shively and W. L. Stutzman, "Wideband Arrays with Variable Element Sizes," *IEE Proc.*, Part H, Vol. 137, pp. 238–240, Aug. 1990.

25. J. D. Dyson, "The Unidirectional Equiangular Spiral Antenna," *IRE Trans. on Ant. and Prop.*, Vol. AP-7, pp. 329–334, Oct. 1959.

26. E. Roland and W. F. Patterson, "A Slow-Wave Flat Spiral Antenna," *17th USAF Antenna Symposium* (U. Ill. Allerton Park), Nov. 1967.

27. T. -T. Chu and H. G. Oltman, "The Sinuous Antenna," *Microwave Syst. News and Communication Technology*, Vol. 18, pp. 40–48, June 1988.

28. R. H. DuHammel and D. E. Isbell, "Broadband Logarithmically Periodic Antenna Structures," *IRE Intern. Conv. Record*, pp. 119–128, 1957.

29. R. S. Elliott, "A View of Frequency Independent Antennas," *Microwave J.*, Vol. 5, pp. 61–68, Dec. 1962.

30. D. E. Isbell, "Log Periodic Dipole Arrays," *IRE Trans. on Ant. and Prop.*, Vol. AP-8, pp. 260–267, May 1960.

31. R. Carrel, "The Design of Log-Periodic Dipole Antennas," *IRE Interna. Convention Rec.*, Part 1, pp. 61–75, 1961.

32. G. DeVito and G. B. Stracca, "Comments on the Design of Log-Periodic Dipole Antenna," *IEEE Trans. on Ant. and Prop.*, Vol. AP-21, pp. 303–308, May 1973.

33. G. DeVito and G. B. Stracca, "Further Comments on the Design of Log-Periodic Dipole Antennas," *IEEE Trans. on Ant. and Prop.*, Vol. AP-22, pp. 714–718, Sept. 1974.

34. P. C. Butson and G. T. Thompson, "A Note on the Calculation of the Gain of Log-Periodic Dipole Antennas," *IEEE Trans. on Ant. and Prop.*, Vol. AP-14, pp. 105–106, Jan. 1976.

35. J. K. Breakall and R. A. R. Solis, "A New Design Method for Low Side Lobe Level Log-Periodic Dipole Antennas," *Appl. Computational Electromagnetics Soc. J.*, Vol. 11, pp. 9–15, Nov. 1996.

36. D. F. DiFonzo, "Reduced Size Log Periodic Antennas," *Microwave J.*, Vol. 7, pp. 37–42, Dec. 1964.

37. S. C. Kuo, "Size-Reduced Log-Periodic Dipole Array Antenna," *Microwave J.*, Vol. 15, pp. 27–33, Dec. 1972.

38. C. K. Campbell et al., "Design of a Stripline Log-Periodic Dipole Antenna," *IEEE Trans. on Ant. and Prop.*, Vol. AP-25, pp. 718–721, Sept. 1977.

39. C. C. Bantin and K. G. Balmain, "Study of Compressed Log-Periodic Dipole Antennas," *IEEE Trans. on Ant. and Prop.*, Vol. AP-18, pp. 195–203, March 1970.

40. J. H. Reed, *An Introduction to Ultrawideband Communication Systems*, Prentice-Hall, 2005.

41. W. D. Prather et al., "Recent Developments in Ultra-Wideband Sources and Antennas," in *Ultra-Wideband Short Pulse Electromagnetics 5*, Kluwer Academic / Plenum Publishers, pp. 393–397, New York, 2002.

42. T. Yang, S-Y Suh, R. Nealy, W. A. Davis, and W. Stutzman, "Compact Antennas for UWB Applications," *IEEE Aerospace & Elec. Sys. Mag.*, Vol. 19, pp. 16–20, May 2004.

43. W. A. Davis, T. Yang, and W. Stutzman, "The Design of Ultra-Wideband Antennas," *URSI General Assembly Meeting* (Chicago), Aug. 2008.
44. D. Ghosh et al., "Transmission and Reception by Ultra-Wideband Antenas," *IEEE Ant. and Prop. Mag.*, Vol. 48, pp. 67–98, Oct. 2006.
45. T. Yang, W. A. Davis, and W. L. Stutzman, "Fundamental-Limit Perspectives on Ultrawideband Antennas," *Radio Science*, Vol. 44, Feb. 2009.

PROBLEMS

7.1-1 Plot percent bandwidth using both arithmetic and geometric mean as the center frequency as a function of B_r from 1 to 3.

7.1-2 Derive the relationships in (7-3).

7.2-1 Verify that the maximum of the radiation from a traveling-wave long wire antenna that is 6λ long occurs at an angle of $20.1°$ from the wire.

7.2-2 Compare the approximate beam maximum angle formula of (7-7) for a traveling-wave long wire with the values of Fig. 7-4 for $L/\lambda = 1, 3, 6, 10$.

7.2-3 Show that the power radiated from a traveling-wave long wire antenna is

$$P = 30I_m^2 \left[2.108 + \ln\left(\frac{L}{\lambda}\right) - \text{Ci}(2\beta L) + \frac{\sin(2\beta L)}{2\beta L} \right]$$

Use (5-8), (7-6) and (2-130).

7.2-4 Use the radiated power expression from Prob. 7.2-3 for a traveling-wave long wire to:
  (a) Derive the directivity expression

$$D = \frac{1.69 \cot^2 \left[\frac{1}{2} \cos^{-1}\left(1 - \frac{0.371}{L/\lambda} \right) \right]}{2.108 + \ln\left(\frac{L}{\lambda}\right) - \text{Ci}(2\beta L) + \frac{\sin(2\beta L)}{2\beta L}}$$

  (b) Evaluate the directivity for $L/\lambda = 2, 5, 10, 20$. $\text{Ci}(2\beta L)$ is approximately zero for these values of L.

7.2-5 Use the radiated power expression from Prob. 7.2-3 for a traveling-wave long wire to:
  (a) Find an expression for the radiation resistance.
  (b) Evaluate the radiation resistance for $L/\lambda = 2, 5, 10, 20$. $\text{Ci}(2\beta L)$ is approximately zero for these values of L.

7.2-6 Plot the linear, polar plot of a traveling-wave long wire antenna that is eight wavelengths long.

7.2-7 To be completely general, the traveling-wave long wire antenna has a current distribution given by

$$I_t(z) = I_m e^{-az} e^{-j\beta_o z}$$

where a is the attenuation coefficient representing radiation and ohmic losses. β_o is the phase constant and is related to the velocity factor $p = v/c$ as $\beta_o = \beta/p$.
  (a) Derive the pattern function

$$F(\theta) = K \sin\theta \, \frac{\sinh\left[\frac{aL}{2} + j\frac{\beta L}{2}\left(\frac{1}{p} - \cos\theta \right) \right]}{\frac{aL}{2} + j\frac{\beta L}{2}\left(\frac{1}{p} - \cos\theta \right)}$$

  (b) Show that this reduces to (7-6) for $a = 0$ and $p = 1$.
  (c) Plot the polar pattern for $a = 0$ and $L = 6\lambda$, for $p = 1.0, 0.75, 0.5$

7.2-8 *Travel-wave vee antenna.*

(a) Place the zero-phase reference point at the vertex of the vee antenna of Fig. 7-5, and derive the radiation pattern as

$$F_V(\theta) = K_V[F_1(\theta) - F_2(\theta)]$$

where

$$F_1(\theta) = e^{j(\beta L/2)[-1+\cos(\theta-\alpha)]} \sin(\theta - \alpha) \frac{\sin[(\beta L/2)(1 - \cos(\theta-\alpha))]}{(\beta L/2)(1 - \cos(\theta - \alpha))}$$

and $F_2(\theta)$ is the same as $F_1(\theta)$ except that $-\alpha$ is replaced by α. This pattern expression is valid only in the plane of the vee.

(b) Plot the polar pattern in Fig. 7-5 for $L = 6\lambda$ and $\alpha = 16°$.

7.2-9 *Rhombic antenna.*

(a) Show that the pattern of the rhombic in Fig. 7-6 is

$$F_R(\theta) = K_R\{F_1(\theta) - F_2(\theta) + e^{-j\beta L}[F_3 - F_4]\}$$

where $F_3 = e^{j\beta L \cos(\theta-\alpha)}F_2$ and $F_4 = e^{j\beta L \cos(\theta+\alpha)}F_1$. F_1 and F_2 are given in Prob. 7.2-8. This expression is valid only in the plane of the rhombic.

(b) Plot the polar pattern in Fig. 7-6 for $L = 6\lambda$ and $\alpha = 16°$

7.2-10 A rhombic antenna above ground is to be designed for a main beam maximum at an elevation angle of $20°$. Determine the rhombic configuration required.

7.3-1 Compare the radiation resistances of the resonant stub helix to a short monopole for the height values of 0.01, 0.05, 0.08, and 0.1λ.

7.3-2 Find the radiation resistance of a six-turn resonant stub helix that is 2 cm high and operates at 850 MHz.

7.3-3 An unfurlable helix was built with an overall length of 78.7 cm, a diameter of 4.84 cm, and a pitch angle of $11.7°$. The center frequency of operation is 1.7 GHz. Calculate the number of turns, the gain in decibels, the half-power beamwidth in degrees, and the axial ratio for the helix.

7.3-4 It is desired to achieve a right-hand circularly polarized wave at 475 MHz having a half-power beamwidth of $39°$. One of the easiest ways to do this is with a helix antenna. It is to be built with a pitch angle of $12.5°$, and the circumference of one turn is to be one wavelength at the center frequency of operation.

(a) Calculate the number of turns needed.

(b) What is the directivity in decibels?

(c) What is the axial ratio of the on-axis fields?

(d) Over what range of frequencies will these parameters remain relatively constant?

(e) Find the input impedance at the design frequency and at the ends of the band.

(f) Evaluate HP at the band ends.

7.3-5 A commercially available axial mode helix antenna has six turns made of 0.95-cm aluminum tubing supported by fiberglass insulators attached to a 3.8-cm aluminum shaft. The band of operation is 300 to 520 MHz. The mechanical characteristics are as follows: length of helix, 118 cm; diameter of helix (center to center), 23.2 cm; and ground screen diameter, 89 cm.

(a) Determine the pitch angle α.

(b) Compute the gain in decibels at edges of the frequency band of operation.

7.3-6 A 12-turn axial mode helix has a circumference of 0.197 m, a pitch angle of $8.53°$, and operates at 1525 MHz. Calculate and plot the radiation pattern in linear-polar form.

7.3-7 A helix antenna has five turns and a pitch angle of $12°$. It is operated such that its circumference is one wavelength. (a) Use simple array theory techniques to derive and accurately sketch the radiation pattern. (b) Calculate the half-power beamwidth, not based on results from (a).

7.3-8 One turn of an axial mode helix radiates similarly to a one-wavelength loop antenna. Explain why, then, that the helix antenna radiates circular polarization and the loop radiates linear polarization.

7.4-1 Calculate the input impedance for infinite biconical antennas of the following cone half-angles: $0.1°$, $1°$, $10°$, $20°$, $50°$.

7.4-2 Show that the radiated power of the infinite biconical antenna is

$$P = 4\pi\sqrt{\mu/\varepsilon}H_o^2\ln[\cot(\theta_h/2)]$$

and that the directivity is

$$D = \frac{1}{\sin^2\theta\,\ln[\cot(\theta_h/2)]}$$

7.5-1 *Construction project.* Select a frequency for which you have laboratory equipment to measure impedance (probably in the VHF or UHF range). Construct both an optimum open-sleeve dipole and its ordinary dipole version. (Alternatively, monopoles may be constructed.) Measure the input impedance of both antennas over a 2:1 frequency range about the center frequency. (Alternatively, measure the VSWR.)

7.5-2 Use (7-50) and Figs. 6-5 and 6-6 to estimate the input impedance of the asymmetrically fed dipole in Prob. 6.1-4.

7.6-1 Find the input impedance of a slot that is complementary to a half-wavelength long ribbon dipole with impedance $73 + j42.5\ \Omega$.

7.6-2 The far-field components for a short slot with uniform electric field and of length $\Delta z \ll \lambda$ along the z-axis are

$$E_\phi = -\frac{V_o\Delta z}{2\pi}j\beta\frac{e^{-j\beta r}}{r}\sin\theta$$

$$H_\theta = \frac{V_o\Delta z}{2\pi\eta}j\beta\frac{e^{-j\beta r}}{r}\sin\theta$$

where V_o is the excitation voltage across the center of the slot.
  (a) Find the input radiation resistance.
  (b) Verify that (7-51) is satisfied using the appropriate complementary antenna.

7.6-3 Frequency-independent antennas have constant HP with frequency. Explain this in terms of the formula $HP = K\lambda/L$.

7.7-1 What is the sense of CP radiated by the equiangular spiral antenna shown in Table 1-4?

7.7-2 Design an equiangular spiral antenna for operation over the band 450 to 900 MHz

7.7-3 *Construction project.* Construct the equiangular spiral antenna of the previous problem using aluminum foil glued to cardboard. Test its performance with a receiver (perhaps a television).

7.8-1 Design a self-complementary log-periodic toothed planar antenna for operation from 400 MHz to 2 GHz with a half-power beamwidth of 70°.

7.8-2 A log-periodic dipole array is to be designed to cover the frequency range 84 to 200 MHz and have 7.5-dB gain. Give the required element lengths and spacings for optimal design.

7.8-3 Evaluate the dipole lengths and spacings for the LPDA of Example 7-3.

7.8-4 Design an optimum LPDA to operate from 470 MHz to 890 MHz with 9-dB gain. Add one extra element to each end over that required by (7-84).

7.8-5 What is the physical length of the LPDA in Example 7-3?

Chapter 8

Array Antennas

8.1 INTRODUCTION

The topic of array antennas was introduced in Sec. 3.5 with several two-element arrays as examples of simple radiating systems. Fig. 3-19 summarizes the range of pattern control that is possible by changing the spacing and phasing of two-element arrays illustrated using elements with the same excitation amplitude and isotropic patterns. Arrays are popular because of the ability to shape the pattern through spacing and excitation adjustment along with the unique capability of scanning the pattern in angular space by dynamically adjusting the excitation phases electronically. Such an array is referred to as a **phased array**. The traditional method of scanning a directive beam is to mechanically steer an aperture antenna such as a reflector antenna. Thus, the problem of mechanically slewing of one large antenna is traded for the feed network complexities associated with properly exciting many small antennas in an array. In addition to avoiding the need for a mechanical structure required to point a large antenna, a phased array can also scan a beam at electronic speeds and can even have multiple simultaneous main beams. Although a feed network is required to excite the array elements, today's low-cost electronics, computing, and digital signal processing hardware and software make electronic scanning affordable. Arrays have the additional advantage of supporting many geometries, including non-planar surfaces to conform to existing contours such as the skin of an aircraft. There are many applications for phased arrays, most notably in radar.

The first widely used array antenna was the Yagi-Uda array invented in 1926 (see Sec. 6.3). The Yagi was followed by other fixed-phase arrays that were mechanically steered. Phased arrays emerged during World War II using mechanical phase shifters. With the invention of the ferrite phase shifter in the 1950s, full electronic scanning became possible. Both array element advances and solid-state phase shifters allowed for more affordable, compact phased arrays beginning in the 1960s. The recent innovation in arrays is the active phased array in which each element has transmit/receive electronics including phase adjustment.

Both arrays and aperture antennas are capable of high gain. Arrays can be analyzed using simple mathematics by summing phasors, but reflector antenna analysis requires integration over the aperture, as we will find out in the next chapter. In this chapter, we continue the development of arrays, beginning with linear arrays of multiple isotropic elements, uniform excitations, and equal spacings. The pattern of individual elements is then included. Techniques for evaluating the directivity of linear arrays are presented, followed by addressing nonuniform excitation of linear arrays and multidimensional arrays. Implementation issues of mutual coupling and feeding techniques are treated. Finally, more detail is presented on the types of elements used in practical arrays along with the special topic of wide-bandwidth phased arrays.
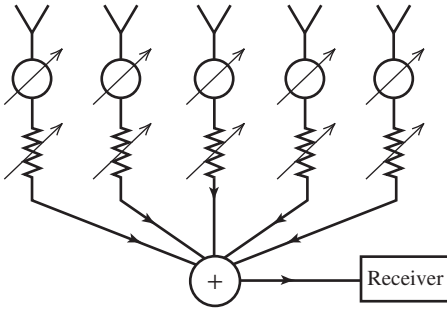
Figure 8-1 A typical linear array. The symbols $\oslash$ and $\lightning$ indicate variable phase shifters and attenuators. The output currents are summed before entering the receiver.

8.2   THE ARRAY FACTOR FOR LINEAR ARRAYS

The fundamental configuration for elements in an array is the linear array shown in Fig. 8-1. Linear arrays are used widely in practice and their operating principles can be used to understand more complex array geometries. The array of Fig. 8-1 has identical elements and is operated as a receiving antenna. The pattern characteristics of an array can be explained for operation as a transmitter or receiver, whichever is more convenient, since antennas usually satisfy the conditions for reciprocity. The output of each element can be controlled in amplitude and phase as indicated by the attenuators and phase shifters in Fig. 8-1. Amplitude and phase control provide for custom shaping of the radiation pattern and for scanning of the pattern in space.

The basic array antenna model consists of two parts, the pattern of one of the elements by itself, the element pattern, and the pattern of the array with the actual elements replaced by isotropic point sources, the array factor. The total pattern of the array is then the product of the element pattern and array factor; this will be discussed in detail in Sec. 8.4. We treat the array factor first.

The array factor corresponding to the linear array of Fig. 8-1 is found by replacing each element by an isotropic radiator, but retaining the element locations and excitations as shown in Fig. 8-2. The array is receiving a plane wave arriving at an angle θ from the line of the elements, and the planes of equal phase (i.e., wavefronts) are shown. Rays perpendicular to the wavefronts indicate the direction of travel of the wave. With the reference wavefront taken to be of zero phase, the distance to the nth element has a corresponding phase delay (found by multiplying by β) of ξ_n. That is, each element is excited with phase ξ_n, due to the spatial phase delay effect of the incoming plane wave. The amplitudes of excitation are constant, taken to be unity, because a plane wave has uniform amplitude. The resulting excitations of $1e^{j\xi_0}$, $1e^{j\xi_1}$, ... are shown for each element in Fig. 8-2. The elements themselves do not weight the outputs since they are isotropic radiators that respond equally to all incoming wave directions.

The array factor for the array of Fig. 8-1 is found from the array of Fig. 8-2 that has isotropic radiators for array elements in place of the actual elements. The array factor for this receiving array is then the sum of the isotropic radiator receiving antenna responses $\{e^{j\xi_0}, e^{j\xi_1}, \ldots\}$ weighted by the amplitude and phase shift represented by complex currents $\{I_0, I_1, \ldots\}$ introduced in the transmission path connected to each element. The array factor of the array shown in Fig. 8-2 is thus

$$\text{AF} = I_0 e^{j\xi_0} + I_1 e^{j\xi_1} + I_2 e^{j\xi_2} + \cdots \tag{8-1}$$

where $\xi_0, \xi_1, \ldots$ are the phases of an incoming plane wave at the element locations designated $0, 1, \ldots$ For convenience, these phases are usually relative to the coordinate origin; that is, the phase of the wave arriving at the nth element leads the phase of the wave arriving at the origin by ξ_n.
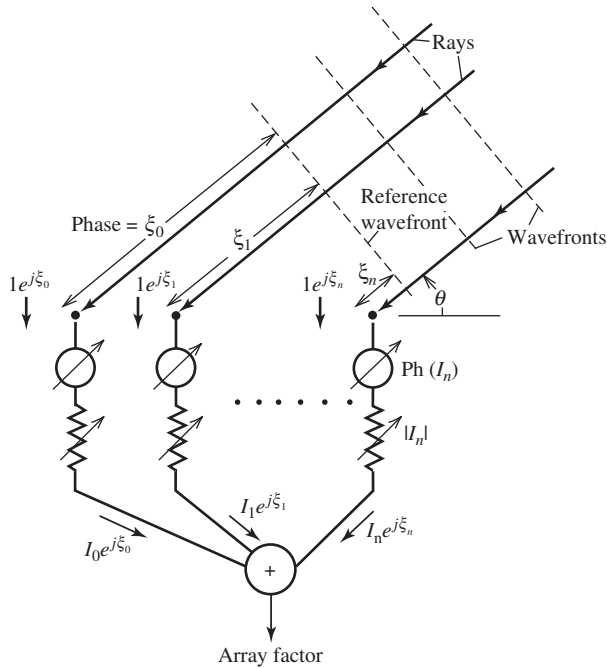
Figure 8-2 Equivalent configuration of the array in Fig. 8-1 for determining the array factor. The elements of the array are replaced by isotropic point sources.

In Sec. 3.5, we presented examples that show how to sum the phasors in (8-1) for two-element arrays of isotropic elements. For such simple cases, the major features of the radiation pattern can be found by an inspection method where the phasors for each source element are traced to the far field in principal directions and summed. We further illustrate the inspection method with the following example.

EXAMPLE 8-1 *Two Identical Isotropic Point Sources Spaced One Wavelength Apart (see Fig. 8-3)*

Two isotropic sources spaced one wavelength apart transmit signals of the same magnitude and phase as shown in Fig. 8-3a. Fig. 8-3b shows some of the transmitted far-field rays. Because of the in-phase condition, the fields arriving in the far field add perfectly (i.e., double) in the $\pm x$-directions (broadside directions) because of the equal path length to the far field. In other words, since the rays started out in phase and experience the same propagation phase shift, they arrive in phase. Off the ends of the array in the endfire directions, the phase of the field from the more distant element to the same far-field distance lags that from the close element by $360°$ due to the extra one wavelength of travel. Thus, the far-field phasors sum in phase in the $\pm z$-directions. So, the phase difference between the two far-field rays varies from $0°$ at broadside ($\theta = 90°$) to $360°$ at endfire ($\theta = 0°, 180°$). Therefore, there is a direction where the phase difference is $180°$, corresponding to a half-wavelength path length difference, and the phasors cancel. The path length difference as a function of observation angle for the two-element array is $d \cos \theta$ (see Fig. 3-16d). For this example where $d = \lambda$, the angle for $\lambda/2$-path difference can be found using $\lambda \cos \theta = \pm \lambda/2$, giving $\cos \theta = \pm 1/2$ which has solutions $\theta = 60°, 120°$ for null locations. With these null locations and knowing that maxima occur in multiples of $90°$, the radiation pattern can be sketched as in Fig. 8-3c. The pattern magnitude is plotted, but it should be noted that the endfire beams at $0°$ and $180°$ are negative-valued due to

(a) Array configuration.

(b) Inspection method.

(c) Polar plot of array factor magnitude $|f(\theta)| = |\cos(\pi \cos\theta)|$.

(d) Three-dimensional plot.
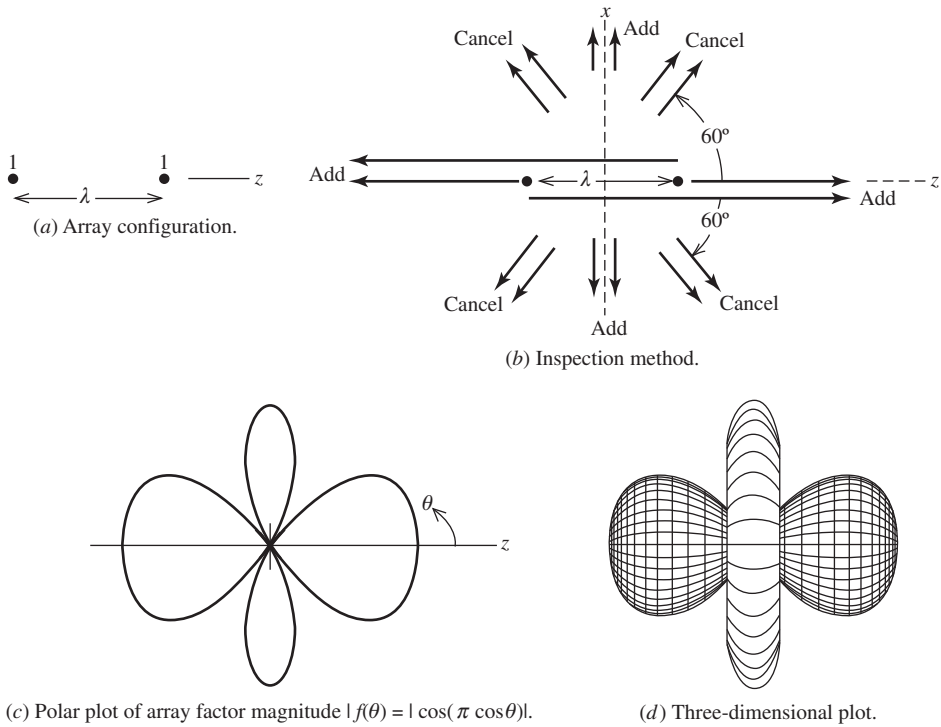
Figure 8-3 Two isotropic point sources with identical amplitude and phase currents, and spaced one wavelength apart (Example 8-1).

the $180°$ phase jump when passing through the null. The exact pattern expression is found from the two-element array factor expression in (3-67) with $d = \lambda$, which when normalized is

$$f(\theta) = \cos(\pi \cos \theta) \qquad (8\text{-}2)$$

The three-dimensional plot of this expression is shown Fig. 8-3d. This example illustrates the general result that *multiple* major lobes will appear in the pattern for an element spacing of one wavelength (or greater), which can also be seen in Fig. 3-19.

We are now ready to develop more general array factor formulas. The expression in (8-1) is general and can be applied to any geometry. First, we consider linear arrays with elements equally spaced along the z-axis as shown in Fig. 8-4. The individual isotropic elements respond equally in all directions to an incoming plane wave. But when the outputs from each element are added in a receiver (weighted by complex currents I_n), the total array response depends on θ, which is the angle with respect to the line of elements (the z-axis). The phase of the arriving wave is set to zero at the origin for convenience, thus $\xi_o = 0$. The incoming waves arriving at Element 1 travel a distance of $d \cos \theta$ less than the wave arriving at the origin. The corresponding phase of waves at Element 1 relative to the origin is $\xi_1 = \beta d \cos \theta$, the spatial phase delay. In fact, across the array the phase at each element leads its immediate nearest neighbor on the left by the same amount of $\beta d \cos \theta$. Using this result in (8-1) gives

$$\text{AF} = I_0 + I_1 e^{j\beta d \cos \theta} + I_2 e^{j\beta 2d \cos \theta} + \cdots = \sum_{n=0}^{N-1} I_n e^{j\beta n d \cos \theta} \qquad (8\text{-}3)$$
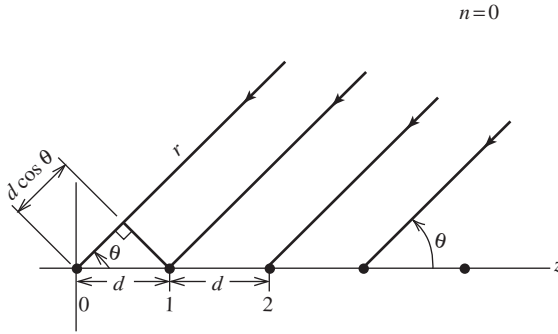
Figure 8-4 Equally spaced linear array of isotropic point sources.

Now consider the array to be transmitting. If the current has a linear phase progression (i.e., relative phase between adjacent elements is the same), we can separate the phase explicitly as

$$I_n = A_n e^{jn\alpha} \tag{8-4}$$

where the $n + 1$th element leads the nth element in phase by α. Then (8-3) becomes

$$AF = \sum_{n=0}^{N-1} A_n e^{jn(\beta d \cos\theta + \alpha)} \tag{8-5}$$

Define

$$\psi = \beta d \cos\theta + \alpha \tag{8-6}$$

Then

$$AF = \sum_{n=0}^{N-1} A_n e^{jn\psi} \tag{8-7}$$

This array factor is a function of ψ and is a Fourier series. This form is convenient for calculations, but we usually want polar plots in terms of the angle θ.

The nonlinear transformation from ψ to θ given by (8-6) can be accomplished graphically. Studying this graphical procedure adds understanding to the interaction of the several array parameters.

As an example, consider two elements spaced one-half wavelength apart and with identical currents as in Example 3-2. We found the normalized array factor in (3-69) to be $f(\theta) = \cos[(\pi/2)\cos\theta]$. In this case, ψ from (8-6) is

$$\psi = \beta d \cos\theta + \alpha = \pi \cos\theta \tag{8-8}$$

since $d = \lambda/2$ and $\alpha = 0$. Now f is expressed in terms of ψ as

$$f(\psi) = \cos\frac{\psi}{2} \tag{8-9}$$

This is a rather simple function to plot. To obtain a plot of $|f|$ as a function of θ, first plot $|f(\psi)|$ from (8-9) as shown in Fig. 8-5. Then draw a circle of radius $\psi = \pi$ below it as shown, since (8-8) is a polar equation of a circle. For an arbitrary value of ψ, say, ψ_1, drop a line straight down until it intersects the circle. The values of $\theta = \theta_1$ and $|f| = f_1$ corresponding to $\psi = \psi_1$ are indicated in the figure. Locating several points taken in this fashion produces the desired pattern sketch. Note that as θ ranges for 0 to π, ψ goes from
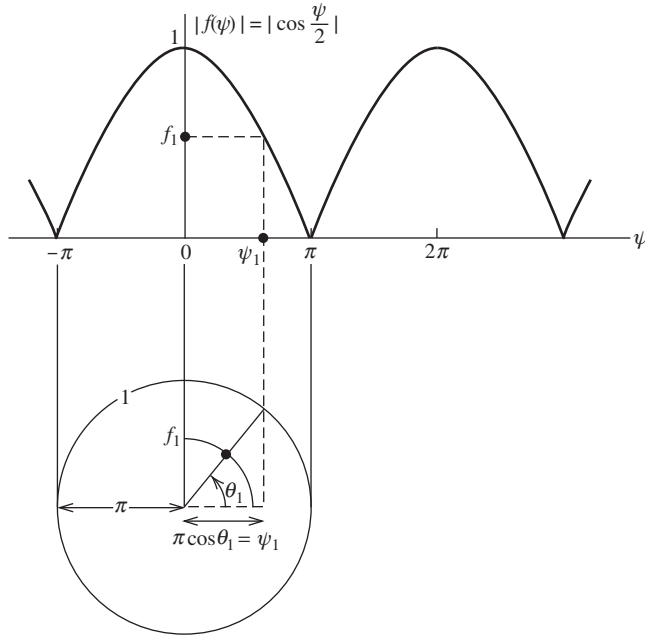
Figure 8-5 Procedure for obtaining the polar plot of the array factor of two elements spaced one-half wavelength apart with identical currents.
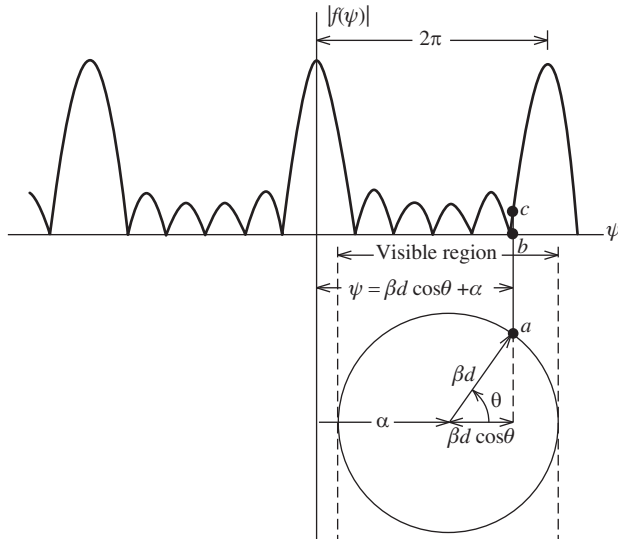
Figure 8-6 Construction technique for finding the array factor as a function of polar angle θ.

π to $-\pi$ in this case. The resulting polar plot is shown in Fig. 8-7b. It is the same as the result obtained in Fig. 3-16.

Before proceeding with more specific examples, we consider a general array factor and how a polar pattern is obtained from it. The magnitude of a typical array factor is plotted as a function of ψ in Fig. 8-6. Below it a circle is constructed with a radius equal to βd and its center located at $\psi = \alpha$. The angle θ is shown. It is very easy to use this plot. For a
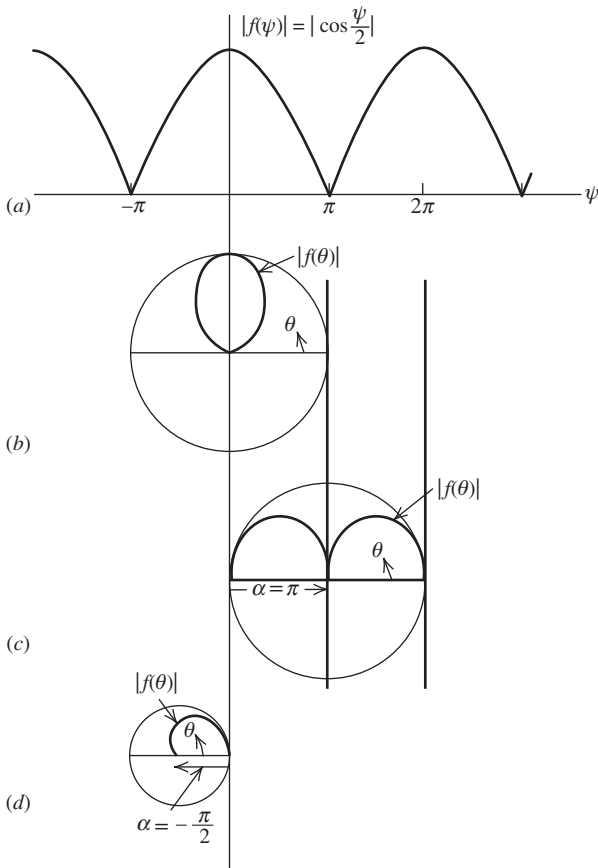
Figure 8-7 Array factors for two-element arrays with equal amplitude currents. (*a*) Universal array factor. (*b*) Polar plot for $d = \lambda/2, \beta d = \pi, \alpha = 0$ (Example 3-2). (*c*) Polar plot of $d = \lambda/2, \beta d = \pi, \alpha = \pi$ (Example 3-3). (*d*) Polar plot $d = \lambda/4$, $\beta d = \pi/2, \alpha = -\pi/2$ (Example 3-4).

given value of θ, locate the intersection of a radial line from the origin of the circle and the perimeter, point *a*. The corresponding value of ψ, at point *b*, is on a vertical line from *a*. The array factor value corresponding to these values of ψ and θ is then point *c*, also on the vertical line from *a*. Notice that the distance from the $\psi = 0$ axis to a point, say, at *a*, can be written as $\psi = \alpha + \beta d \cos \theta$, which is (8-6).

To illustrate the procedure further, we will find the polar plots of the array factors for some two-element arrays with uniform current amplitudes that were discussed earlier. The array factor as a function of ψ, from (8-7) with $N = 2$, is

$$\text{AF} = 1 + e^{j\psi} = e^{j(\psi/2)}(e^{-j(\psi/2)} + e^{j(\psi/2)}) = 2e^{j(\psi/2)} \cos \frac{\psi}{2} \qquad (8\text{-}10)$$

where $A_0 = A_1 = 1$. Taking the magnitude eliminates the exponential factor and normalization removes the factor of 2, giving

$$|f(\psi)| = \left| \cos \frac{\psi}{2} \right| \qquad (8\text{-}11)$$

which also follows from (8-9). The array factor $|f(\psi)|$ is the same for all two-element arrays with the same current amplitudes and is plotted in Fig. 8-7*a*. Of course, ψ changes with element spacing and phasing. For example, if the spacing is a half-wavelength and the phases of each element are zero ($\alpha = 0$), the pattern is obtained as shown in Fig. 8-5 with the resulting pattern plotted in Fig. 8-7*b*. This is Example 3-2 discussed earlier.

For Example 3-3, $d = \lambda/2$ and $\alpha = \pi$. The resulting polar plot of the array factor using the procedures of Fig. 8-6 is shown in Fig. 8-7c. The array factor for Example 3-4 with $d = \lambda/4$ and $\alpha = -\pi/2$ is shown in Fig. 8-7d.

By examining the general array factor expression in (8-7), some general properties can be derived that aid in the construction of pattern plots. First, the array factor is periodic in the variable ψ with period 2π. This is easily shown as follows:

$$\text{AF}(\psi + 2\pi) = \sum A_n e^{jn(\psi + 2\pi)} = \sum A_n e^{jn\psi} e^{jn2\pi} = \sum A_n e^{jn\psi} = \text{AF}(\psi) \qquad (8\text{-}12)$$

The array factor of a linear array along the z-axis is a function of θ but not of ϕ (the element pattern may be though). In other words, the array factor is a pattern that has rotational symmetry about the line of the array. Therefore, its complete structure is determined by its values for

$$0 < \theta < \pi \qquad (8\text{-}13)$$

This is called the *visible region*. This corresponds to $-1 < \cos\theta < 1$ or $-\beta d < \beta d$ $\cos\theta < \beta d$ or

$$\alpha - \beta d < \psi < \alpha + \beta d \qquad (8\text{-}14)$$

Hence, the visible region in terms of θ and ψ is given by (8-13) and (8-14), respectively. The element spacing of the array in terms of a wavelength, d/λ, determines the size of the circle in Fig. 8-6 and thus how much of the array factor appears in the visible region. The visible region in the variable ψ is of length $2\beta d$, as seen from (8-14). This is the diameter of the circle in Fig. 8-6. Suppose that exactly one period appears in the visible region. Since the period is 2π, we have $2\pi = 2\beta d = 2(2\pi/\lambda)d$ or $d/\lambda = \frac{1}{2}$. Thus, *exactly one period of the array factor appears in the visible region when the element spacing is one-half wavelength.* Less than one period is visible if $2\beta d < 2\pi$, which corresponds to $d/\lambda < \frac{1}{2}$ — that is, for spacings less than one-half wavelength. For spacings greater than one-half wavelength, more than one period will be visible. For one-wavelength spacings, two periods will be visible. For spacings larger than a half-wavelength, there may be more than one major lobe in the visible region, depending on the element phasings. Additional major lobes that rise to an intensity equal to that of the main lobe are called *grating lobes*. In the one-wavelength spaced, two-element array factor of Fig. 8-3c, there are grating lobes at $\theta = 0$ and $180°$, in addition to the desired lobe in the $\theta = 90°$ direction. In most situations, it is undesirable to have grating lobes. As a result, most arrays are designed so the element spacings are less than one wavelength.

8.3   UNIFORMLY EXCITED, EQUALLY SPACED LINEAR ARRAYS

An array is usually comprised of identical elements positioned in a regular geometrical arrangement. In fact, this is the definition adopted by the IEEE. However, arrays are encountered in practice with unequal interelement spacings. Usually, a modifier (e.g., equally or unequally spaced) is included to be completely clear about the array geometry. The examples presented in this chapter are for equally spaced arrays, and unequally spaced arrays are treated in Sec. 8.8.

8.3.1   The Array Factor Expression

A very important special case of equally spaced linear arrays is that of the uniformly excited array. This is an array whose element current amplitudes are identical, or

$$A_0 = A_1 = A_2 = \cdots \qquad (8\text{-}15)$$

In this section, we consider only element phasings of a linear form that is accounted for by interelement phase shift α. The array factor from (8-7) is then

$$\text{AF} = A_0 \sum_{n=0}^{N-1} e^{jn\psi} = A_0(1 + e^{j\psi} + \cdots + e^{j(N-1)\psi}) \tag{8-16}$$

Only a few short steps are required to sum this geometric series. First, multiply (8-16) by $e^{j\psi}$ to obtain

$$\text{AF}\ e^{j\psi} = A_0(e^{j\psi} + e^{j2\psi} + \cdots + e^{jN\psi}) \tag{8-17}$$

Subtracting this from (8-16) gives

$$\text{AF}(1 - e^{j\psi}) = A_0(1 - e^{jN\psi})$$

or

$$\text{AF} = \frac{1 - e^{jN\psi}}{1 - e^{j\psi}} A_0 \tag{8-18}$$

This is rewritten in a more convenient form as follows:

$$\text{AF} = A_0 \frac{e^{jN\psi} - 1}{e^{j\psi} - 1} = A_0\ \frac{e^{jN\psi/2}}{e^{j\psi/2}} \frac{e^{jN\psi/2} - e^{-jN\psi/2}}{e^{j\psi/2} - e^{-j\psi/2}}$$

$$\tag{8-19}$$

$$= A_0 e^{j(N-1)\psi/2}\ \frac{\sin(N\psi/2)}{\sin(\psi/2)}$$

The phase factor $e^{j(N-1)\psi/2}$ is not important unless the array output signal is further combined with the output from another antenna. In fact, if the array were centered about the origin, the phase factor would not be present since it represents the phase shift of the array phase center relative to the origin. Neglecting the phase factor in (8-19) gives

$$\text{AF} = A_0 \frac{\sin(N\psi/2)}{\sin(\psi/2)} \tag{8-20}$$

This expression is maximum for $\psi = 0$ and the maximum value from (8-16) is

$$\text{AF}(\psi = 0) = A_0(1 + 1 + \cdots + 1) = A_0 N \tag{8-21}$$

Dividing this into (8-20) gives the normalized array factor

$$\boxed{f(\psi) = \frac{\sin(N\psi/2)}{N\sin(\psi/2)}} \quad \textit{UE, ESLA} \tag{8-22}$$

This is the normalized array factor for an N element, uniformly excited, equally spaced linear array (UE, ESLA) that is centered about the coordinate origin. This function is similar to a $(\sin u)/u$ function, with the major difference that the side lobes do not die off without limit for increasing argument. In fact, the function (8-22) is periodic in 2π, which is true in general, as we showed in (8-12).

(a)

(b)
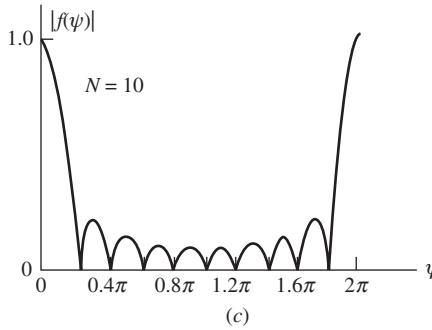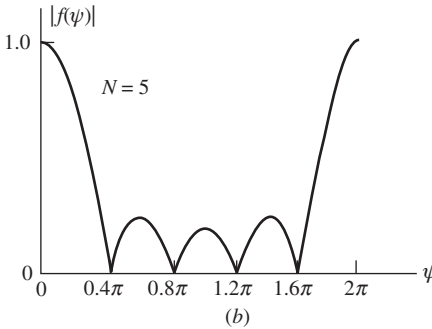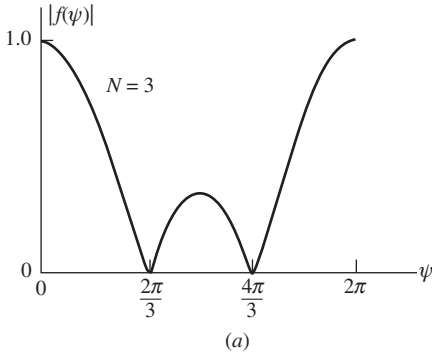
Figure 8-8 Array factor for an equally spaced, uniformly excited linear array for a few array numbers. (*a*) Three elements. (*b*) Five elements. (*c*) Ten elements.

A number of trends can be seen by examining array factor plots for various values of N as shown in Fig. 8-8:

1. As N increases, the main lobe narrows.
2. As N increases, there are more side lobes in one period of $f(\psi)$. In fact, the number of full lobes (one main lobe and the side lobes) in one period of $f(\psi)$ equals $N - 1$. Thus, there will be $N - 2$ side lobes are one main lobe in each period.
3. The minor lobes are of width $2\pi/N$ in the variable ψ and the major lobes (main and grating) are twice this width.
4. The side lobe peaks decrease with increasing N. A measure of the side lobe peaks is the *side lobe level* that we have defined as

$$\text{SLL} = \frac{|\text{maximum value of largest side lobe}|}{|\text{maximum value of main lobe}|} \tag{8-23}$$

and it is often expressed in decibels. The side lobe level of the array factor for $N = 5$ is -12 dB and it is -13 dB for $N = 20$. SLL approaches the value of a uniform line source, -13.3 dB, as N is increased.

5. $|f(\psi)|$ is symmetric about π. It is left as an exercise to show this.

The radiation (field) polar plots in the variable θ can be obtained from $f(\psi)$ as discussed in Sec. 8.2. For example, consider the two-element case. Then (8-22) becomes

$$f(\psi) = \frac{\sin\psi}{2\sin(\psi/2)} \tag{8-24}$$

This is a universal pattern function for all equal amplitude two-element arrays and is plotted in Fig 8-7a. Note that by the techniques used in Sec. 8.2, we found that the array factor for a two-element array was $\cos(\psi/2)$; see (8-9). It can be shown that this is identical to (8-24).

EXAMPLE 8-2   *Four-Element Linear Array (see Fig. 8-9)*

The universal array factor for a four-element, uniformly excited, equally spaced array is plotted in Fig. 8-9b. Let us find the array factor plot for the special case of half-wavelength spacing and 90° interelement phasing (i.e., $\alpha = \pi/2$). The array excitations are shown in
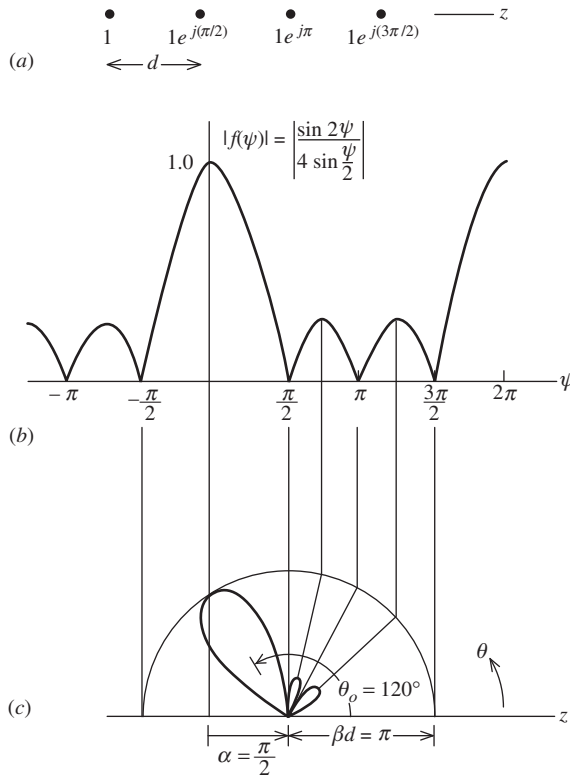
Figure 8-9 Array factor for a four-element, uniformly excited, equally spaced phased array (Example 8-2). (*a*) The array excitations. (*b*) Universal pattern for $N = 4$. (*c*) Polar plot for $d = \lambda/2$ and $\alpha = \pi/2$.

Fig. 8-9a. The pattern plot can be sketched quickly by locating prominent features such as maxima and zeros. Then vertical lines are dropped down from these points to the circle below. From the intersection points with the circle, straight lines are drawn in to the center of the circle. The perimeter of the circle has a pattern value of unity and the center a value of 0. For linear polar plots such as this one, the magnitude of the pattern factor is linearly proportional to the distance from the origin. For example, if the circle radius is 4 cm and the pattern value to be plotted is 0.25, the pattern point is 1 cm from the origin along a radial line at the appropriate angle θ. After locating the relative maxima and the zeros, a smooth curve is drawn, joining these points. The polar plot calculated using (8-22) is shown in Fig. 8-9c. Note that a polar plot can be made larger or smaller by expanding or contracting the construction circle.

8.3.2 Main Beam Scanning and Beamwidth

A maximum of an array factor occurs for $\psi = 0$. Let θ_o be the corresponding value of θ for which the array factor is maximum. Then from (8-6), we have $0 = \beta d \cos \theta_0 + \alpha$, or

$$\alpha = -\beta d \cos \theta_o \tag{8-25}$$

This is the element-to-element phase shift in the excitation currents required to produce an array factor main beam maximum in a direction θ_o relative to the line along which the array elements are disposed. Thus, if we want an array factor maximum in the $\theta = \theta_o$ direction, the required element currents from (8-4) with (8-25) are

$$I_n = e^{jn\alpha} = e^{-jn\beta d \cos \theta_o} \tag{8-26}$$

for a uniformly excited, equally spaced linear array. For the broadside case ($\theta_o = 90°$), $\alpha = 0$. For the endfire case ($\theta_o = 0°$ or $180°$), $\alpha = -\beta d$ or βd. In the example illustrated in Fig. 8-9, $\alpha = \pi/2$ and $d = \lambda/2$ so $\theta_o = \cos^{-1}(-\alpha/\beta d) = \cos^{-1}\left(-\frac{1}{2}\right) = 120°$.

A helpful way to remember the influence of the phase is that the beam pointing direction is toward the lagging element. In the example, each element lags its right neighbor, so the beam steers to the left of the array broadside direction as shown in Fig. 8-9c. This main beam scanning by phase control feature can be explicitly incorporated into ψ by substituting (8-25) into (8-6), giving

$$\psi = \beta d(\cos \theta - \cos \theta_o) \tag{8-27}$$

Scanning is discussed further in Sec. 8.9.

A measure of the width of the main beam of a uniformly excited, equally spaced linear array is given by the *beamwidth between first nulls*, BWFN, which is illustrated in Fig. 2-10a for a general pattern. The main beam nulls are where the array factor (8-22) first goes to zero in a plane containing the linear array. The zeros of the numerator of (8-22) occur for $N\psi_{\text{FN}}/2 = \pm n\pi$. When the denominator also goes to zero ($\frac{1}{2}\psi_{\text{FN}} = \pm n\pi$), the pattern factor is unity, corresponding to the main beam ($n = 0$) and grating lobes. The first nulls associated with the main beam occur for $N\psi_{\text{FN}}/2 = \pm\pi$. For a broadside array $\alpha = 0°$ and $\psi = \beta d \cos \theta$, so the angles θ for the first nulls are found from

$$\pm\pi = \frac{N}{2}\frac{2\pi}{\lambda}d \cos \theta_{\text{FN}} \tag{8-28}$$

or

$$\theta_{\text{FN}} = \cos^{-1}\left(\pm\frac{\lambda}{Nd}\right) \tag{8-29}$$

The BWFN is then

$$\text{BWFN} = |\theta_{\text{FN left}} - \theta_{\text{FN right}}| \tag{8-30}$$

$$= \left| \cos^{-1}\left(-\frac{\lambda}{Nd}\right) - \cos^{-1}\left(+\frac{\lambda}{Nd}\right) \right| \tag{8-31}$$

For long arrays (length $L = Nd \gg \lambda$), we can approximate (8-31) as follows:

$$\text{BWFN} \approx \left| \frac{\pi}{2} + \frac{\lambda}{Nd} - \left(\frac{\pi}{2} - \frac{\lambda}{Nd}\right) \right| = \frac{2\lambda}{Nd} \qquad near\ broadside \tag{8-32}$$

For an endfire array (see Fig. 2-11c), the beamwidth between first nulls is twice that from the main beam maximum to the first null. For long arrays it is approximately

$$\text{BWFN} \approx 2\sqrt{\frac{2\lambda}{Nd}} \qquad endfire \tag{8-33}$$

Half-power beamwidth (HP) is a more popular measure of the main beam size than BWFN. Both depend on the array length Nd and main beam pointing angle θ_o. For a long ($Nd \gg \lambda$) uniformly excited linear array, the HP is approximately [1]

$$\text{HP} \approx 0.886 \frac{\lambda}{Nd} \ \csc\theta_o \quad near\ broadside \tag{8-34}$$

and

$$\text{HP} \approx 2\sqrt{0.886 \frac{\lambda}{Nd}} \qquad endfire \tag{8-35}$$

Comparing the formulas for HP and BWFN, we can see that HP is roughly one-half of the corresponding BWFN value for long, uniformly excited linear arrays.

8.3.3   The Ordinary Endfire Array

In many applications, antennas are required to produce a single pencil beam. The array factor for a broadside array produces a fan beam (such as in Fig. 8-3), although the proper selection of array elements may yield a total pattern that has a single pencil beam. Another way to achieve a single pencil beam is by the proper design of an endfire array. We have said that an endfire condition results when $\theta_o = 0°$ or $180°$, which corresponds to $\alpha = -\beta d$ or $+\beta d$. Such arrays for which $\alpha = \pm\beta d$ are referred to as *ordinary endfire* arrays. If the spacing d is a half-wavelength, there will be two identical endfire lobes (see Fig. 3-17). There are several ways to eliminate one of these lobes, thus leaving a single pencil beam. Already mentioned was the use of a backing ground plane in Example 3-5. For an array, the back lobe can be reduced by decreasing the spacing below a half-wavelength. The visible region is $2\beta d$ wide in the variable ψ, and to eliminate the unwanted major lobe (grating lobe), we should reduce the visible region (and thus the spacing d) below the half-wavelength spacing value of 2π. Since the grating lobe half-width (maximum to null) is $2\pi/N$, so that $2\beta d \leq 2\pi - \pi/N$. Dividing this by $2\beta d$ gives the condition on spacing, which is given below along with the associated condition on phasing for ordinary endfire operation:

$$d < \frac{\lambda}{2}\left(1 - \frac{1}{2N}\right) \tag{8-36a}$$

$$\alpha = \pm\beta d \qquad \begin{array}{c} \textit{Ordinary endfire conditions} \end{array} \tag{8-36b}$$

An array satisfying these conditions produces a single endfire beam and no grating lobe with a peak in the direction $\theta = 0°$ for $\alpha = -\beta d$ and in the direction $\theta = 180°$ for $\alpha = \beta d$, which follows from (8-25). The example below, like many to follow in this chapter, uses five elements because there are enough lobes to fully illustrate general pattern trends.

EXAMPLE 8-3   *Five-Element Ordinary Endfire Linear Array (see Fig. 8-10)*

From (8-36a) for a five-element ordinary endfire array, we must have $d \leq (\lambda/2)(1 - 1/10) = 0.45\lambda$. If we select $d = 0.45\lambda$ and the $\theta_o = 180°$ endfire direction, the required element-to-element phase shift from (8-36b) is $\alpha = \beta d = (2\pi/\lambda)(0.45\lambda) = 0.9\pi$. The pattern construction
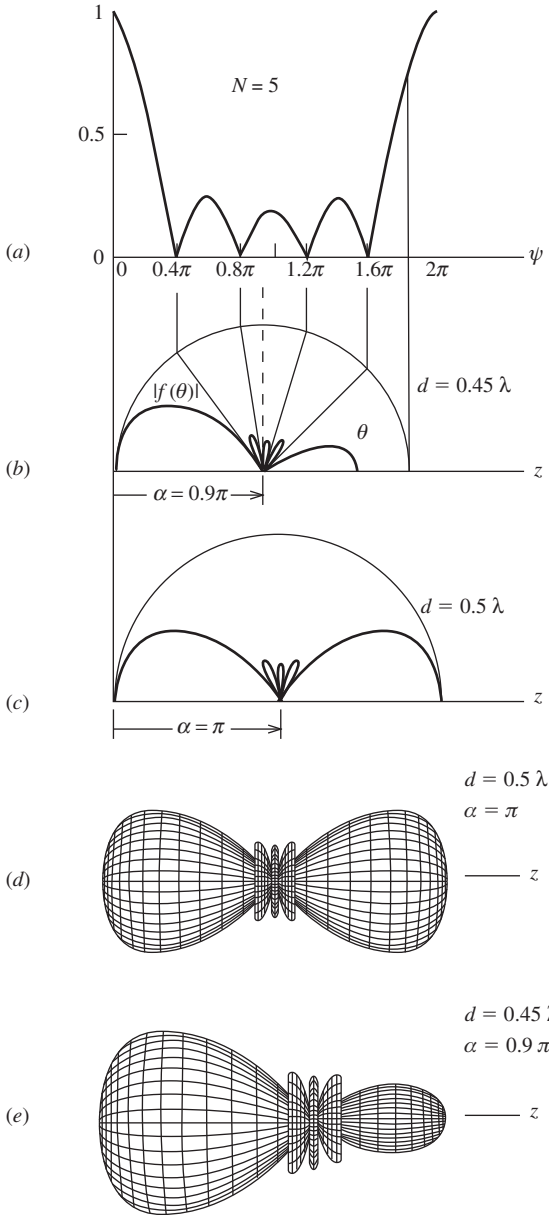
Figure 8-10 Five-element uniformly excited, equally spaced linear array (Example 8-3). (a) Universal pattern plot. (b) Polar plot for ordinary endfire case with $d = 0.45\lambda$ and $\alpha = 0.9\pi$. (c) and (d) Plots for endfire case with $d = 0.5\lambda$. (e) 3D plot for case of (c).

process is shown in Figs. 8-10*a* and *b*, and the three-dimensional pattern plot is shown in Fig. 8-10*e*. Note there is a single main lobe. For comparison, if a spacing of a half-wavelength had been used, two major lobes would appear as shown in Figs. 8-10*c* and *d*.

8.3.4 The Hansen-Woodyard Endfire Array

In the ordinary endfire case, the interelement excitation phase, $\alpha = \pm\beta d$, exactly equals the spatial phase delay of waves in the endfire direction. If the phase delay is increased beyond that for ordinary endfire by an amount δ, the *excess phase delay*, the interelement phase shift is

$$\alpha = \pm(\beta d + \delta) \tag{8-37}$$

with the $+$ and $-$ signs corresponding to beam peak directions of $\theta_o = 180°$ and $0°$, respectively. This excitation moves the maximum of the universal array factor out of the visible region ($0° \leq \theta \leq 180°$) as shown by example in Fig. 8-11, leading to a narrowing of the main beam. A popular choice for the excess phase delay is $\delta = \pi/N$. Because the phase delay is increased, the construction circle moves off the origin, which can lead to the appearance of grating lobes or high side lobes. To avoid this, the circle radius is decreased more than in the ordinary endfire case, corresponding to an element spacing d that is smaller than a half-wavelength by twice that used in the ordinary endfire case. These conditions on spacing and phasing are referred to as the *Hansen-Woodyard endfire conditions* for endfire operation and are given by:

$$d < \frac{\lambda}{2}\left(1 - \frac{1}{N}\right) \tag{8-38a}$$

Hansen-Woodyard endfire conditions

$$\alpha = \pm\left(\beta d + \frac{\pi}{N}\right) \tag{8-38b}$$
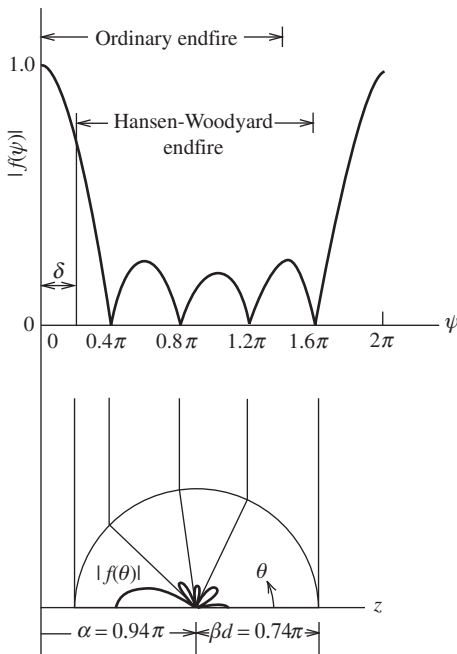
Figure 8-11 Single endfire beam for a five-element Hansen–Woodyard increased directivity array with $\alpha = 0.94\pi$ and $d = 0.37\lambda$ (Example 8-4).

where again the $+$ and $-$ signs are used with main beam peak directions of $\theta_o = 180°$ and $0°$, respectively. The original derivation by Hansen and Woodyard [2] assumed a continuous current distribution, but the result holds for an array of many closely spaced elements. The value that yields maximum directivity is $\delta_{\max} = 2.92/N$, which is close to the Hansen-Woodyard value of π/N used in (8-39b). Array directivity is discussed further in Sec. 8.5.

The pattern of an endfire array with a non-zero δ value has the pattern shape of (8-22) but is not normalized to unity maximum due to the peak of the universal array factor being invisible. Using (8-37) in (8-6) gives the modified expression for ψ of $\beta d(\cos\theta - 1) - \delta$. Solving this for the pattern maximum at $\theta_o = 0°$ or $180°$ gives $\psi_o = -\delta$ or δ, and the pattern from (8-22) is $\sin(N\delta/2)/N\sin(\delta/2)$. Using this to renormalize the general array factor to include non-zero δ gives

$$f(\psi) = \frac{N\sin\left(\dfrac{\delta}{2}\right)\sin\left(\dfrac{N\psi}{2}\right)}{\sin\left(\dfrac{N\delta}{2}\right)N\sin\left(\dfrac{\psi}{2}\right)} \tag{8-39}$$

The consequence of increased directivity for Hansen-Woodyard type arrays is higher side lobes, as illustrated in the following example.

EXAMPLE 8-4   *Five-Element Hansen-Woodyard Endfire Linear Array (see Fig. 8-11)*

From (8-38a) for a five-element Hansen-Woodyard endfire array, we must have $d \leq (\lambda/2)$ $(1 - 1/5) = 0.4\lambda$. Choosing $d = 0.37\lambda$ and the $\theta_o = 180°$ endfire direction, the required element-to-element phase shift from (8-38b) is $\alpha = \beta d + \pi/N = 0.74\pi + 0.2\pi = 0.94\pi$. The pattern shown in Fig. 8-11 has a single main beam that is narrower than for the ordinary endfire case in Example 8-3 (also using five elements), but the side lobes are higher. Nevertheless, the array exhibits increased directivity, as we will examine in more detail in Sec. 8.5.

It is helpful to summarize the endfire array design process. First, select the number of elements, using enough to achieve the required directivity. Next, choose an element spacing that is less than a half-wavelength; for example, satisfying (8-36a) or (8-38a) for ordinary or Handsen-Woodyard endfire cases. Finally, evaluate the interelement phase shift α using (8-36b) or (8-38b) for ordinary or Handsen-Woodyard endfire cases.

8.4   THE COMPLETE ARRAY PATTERN AND PATTERN MULTIPLICATION

Our treatment of arrays so far has assumed that all elements in the array are isotropic point sources. Actual arrays, of course, employ real elements that do not have an isotropic pattern. In the remainder of this chapter, we include element effects, starting in this section with the principle of pattern multiplication which gives the complete array pattern as the product of the element pattern and the array factor.

When the elements of an array are placed along a line and the currents in each element also flow in the direction of that line, the array is said to be *collinear*. As a simple example of a collinear array, suppose we have N short dipoles as shown in Fig. 8-12. The elements are equally spaced a distance d apart and have currents $I_0, I_1, I_2, \ldots, I_{N-1}$. The total current is the sum of the z-directed short dipole currents and thus is z-directed,
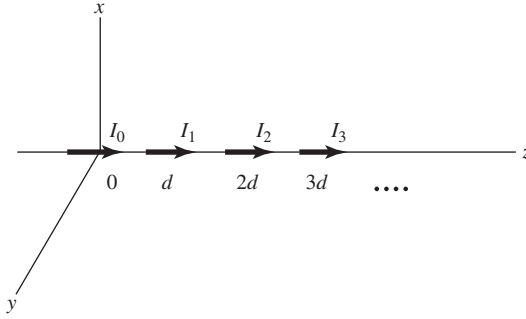
Figure 8-12 A collinear array of short dipoles.

as is the vector potential. The vector potential integral in (2-103) reduces to a sum over the element currents (modeled as ideal dipoles) as[1]

$$
\begin{aligned}
A_z &= \mu \frac{e^{-j\beta r}}{4\pi r} \Delta z \Big[I_0 + I_1 e^{j\beta d \cos\theta} + I_2 e^{j\beta 2d \cos\theta} + \cdots + I_{N-1} e^{j\beta(N-1)d \cos\theta} \Big] \\
&= \mu \frac{e^{-j\beta r}}{4\pi r} \Delta z \sum_{n=0}^{N-1} I_n e^{j\beta nd \cos\theta}
\end{aligned}
\tag{8-40}
$$

in the far field. Then from (2-106),

$$
E_\theta = j\omega\mu \frac{e^{-j\beta r}}{4\pi r} \Delta z \sin\theta \sum_{n=0}^{N-1} I_n e^{j\beta nd \cos\theta}
\tag{8-41}
$$

From this expression, we can identify $\sin\theta$ as the pattern of a single element by itself, called the *element pattern*. The remaining factor

$$
\mathrm{AF} = \sum_{n=0}^{N-1} I_n e^{j\beta nd \cos\theta}
\tag{8-42}
$$

is the array factor of (8-3). The array factor is a sum of fields from isotropic point sources located at the center of each array element and is found from the element currents (amplitudes and phases) and their locations. On the other hand, the element pattern is that factor of the radiation pattern determined by the individual properties of an element based on its current distribution and orientation in space. We shall see that this factoring process holds in general if the elements have the same pattern and are similarly oriented.

We now consider a slightly more complicated case. Suppose for the sake of explanation, we have N identical general, line-source elements forming a collinear array along

---

[1] This result could also be obtained by writing the z-directed current density as

$$
J_z = \delta(x')\ \delta(y')[I_0\ \delta(z') + I_1\ \delta(z'-d) + I_2\ \delta(z'-2d) + \cdots I_{N-1}\ \delta(z'-(N-1)\ d)]\ \Delta z
$$

and substituting this into (2-102), giving

$$
A_z = \mu \frac{e^{-j\beta r}}{4\pi r} \Delta z \int_{-\infty}^{\infty} [I_0\ \delta(z') + I_1\ \delta(z'-d) + \cdots] e^{j\beta z' \cos\theta} dz'
$$

from which (8-40) follows.

the z-axis. The nth element is centered at $z = z_n$ and has a current distribution $i_n(z')$. We are now relaxing the equal spacing constraint. The total current along the z-axis is

$$I(z') = \sum_{n=0}^{N-1} i_n(z') \tag{8-43}$$

The vector potential is then

$$A_z = \mu \frac{e^{-j\beta r}}{4\pi r} \int_{-\infty}^{\infty} \sum_{n=0}^{N-1} i_n(z') e^{j\beta z' \cos\theta} dz' \tag{8-44}$$

The far-field electric field from this and (2-106) is

$$E_\theta = j\omega\mu \frac{e^{-j\beta r}}{4\pi r} \sum_{n=0}^{N-1} E_n(\theta) \tag{8-45}$$

where

$$E_n(\theta) = \sin\theta \int_{-\infty}^{\infty} i_n(z') e^{j\beta z' \cos\theta} dz' \tag{8-46}$$

is the pattern of the nth element.

If the array possesses no symmetry, (8-45) cannot be simplified. But if the array elements are similar, a great deal of simplification is possible. By *similar* we mean that the currents of each antenna element are in the same direction, of the same length, and have the same distribution (although there may be different current amplitudes and phases for each element). Then the patterns of (8-46) will be similar; that is, they will have the same spatial variation but may have different amplitudes and phases. In the example at hand, the currents are all z-directed. Now assume that each element is of length ℓ, has a normalized current distribution over its length of $i(z')$, and an input current of I_n. Then

$$i_n(z') = \mathrm{I}_n i(z' - z_n) \tag{8-47}$$

where z_n is the position of the nth element center along the z-axis. Substituting this into (8-46) gives

$$E_n(\theta) = \sin\theta\ I_n \int_{z_n-\ell/2}^{z_n+\ell/2} i(\xi - z_n) e^{j\beta\xi \cos\theta} d\xi \tag{8-48}$$

where ξ replaced z'. Let $\tau = \xi - z_n$; then (8-48) becomes

$$E_n(\theta) = \sin\theta\ I_n \int_{-\ell/2}^{\ell/2} i(\tau) e^{j\beta(\tau+z_n)\cos\theta} d\tau$$

$$= \sin\theta \left[\int_{-\ell/2}^{\ell/2} i(\tau) e^{j\beta\tau \cos\theta} d\tau \right] I_n e^{j\beta z_n \cos\theta} \tag{8-49}$$

To maintain consistent notation, we replace τ by z', yielding

$$E_n(\theta) = \sin\theta \left[\int_{-\ell/2}^{\ell/2} i(z') e^{j\beta z' \cos\theta} dz' \right] I_n e^{j\beta z_n \cos\theta} \tag{8-50}$$

The pattern for each element of an array of similar elements given by (8-50) is a product of the pattern of the current distribution, and the amplitude and phase of excitation I_n, and the last factor represents the spatial phase due to the displacement from the origin. Substituting (8-50) into (8-45) gives

$$E_\theta = j\omega\mu \frac{e^{-j\beta\tau}}{4\pi r} \left[\sin\theta \int_{-\ell/2}^{\ell/2} i(z') e^{j\beta z' \cos\theta} \, dz' \right] \sum_{n=0}^{N-1} I_n e^{j\beta z_n \cos\theta} \tag{8-51}$$

The factor

$$\sin\theta \int_{-\ell/2}^{\ell/2} i(z') e^{j\beta z' \cos\theta} \, dz' \tag{8-52}$$

when normalized is the element pattern $g_a(\theta)$ of any element in the array of similar elements. The sum

$$\text{AF} = \sum_{n=0}^{N-1} I_n e^{j\beta z_n \cos\theta} \tag{8-53}$$

is the unnormalized array factor.

In going from (8-45) to (8-51), it was necessary to assume that the elements of the array were similar. When this is true, the electric field can be written as a product of an element pattern, as in (8-52), and an array factor, as in (8-53). Note that the array factor is the pattern of a linear array of N point sources located at positions $\{z_n\}$ on the z-axis. If the elements are equally spaced, $z_n = nd$ and (8-53) reduces to (8-42). If further, they are uniformly excited, the array reduces to (8-20). This result is not restricted to collinear elements but can be applied to any array of similar elements. This is discussed next.

The process of factoring the pattern of an array into an element pattern and an array factor is referred to as the *principle of pattern multiplication*. It is stated as follows: The electric field pattern of an array consisting of similar elements is the product of the pattern of one of the elements (the element pattern) and the pattern of an array of isotropic point sources with the same locations, relative amplitudes, and phases as the original array (the array factor).

In Sec. 2.4, we wrote the normalized electric field pattern of a single antenna as a product of a normalized element factor g and a normalized pattern factor f. For array antennas, we expand this concept and call the pattern of one element in the array an element pattern g_a. It, in turn, is composed of an element factor that is the pattern of an infinitesimal piece of current on the array element (i.e., an ideal dipole) and a pattern factor that is the pattern due to its current distribution as in (8-52). The complete (normalized) pattern of an array antenna is

$$F(\theta, \phi) = g_a(\theta, \phi) f(\theta, \phi) \tag{8-54}$$

where $g_a(\theta, \phi)$ is the normalized pattern of a single element antenna of the array (the *array element pattern*, or *element pattern*) and $f(\theta, \phi)$ is the familiar normalized array factor.

EXAMPLE 8-5   *Two Collinear, Half-Wavelength-Spaced Short Dipoles (see Fig. 8-13)*

To illustrate pattern multiplication, consider two collinear short dipoles spaced a half-wavelength apart and equally excited. The element pattern is $\sin\theta$ for an element along the z-axis and the array factor was found in (3-69) to be $\cos[(\pi/2)\cos\theta]$. The total pattern is then $\sin\theta \cos[(\pi/2)\cos\theta]$. The patterns are illustrated in Fig. 8-13. The 3D pattern, which has been
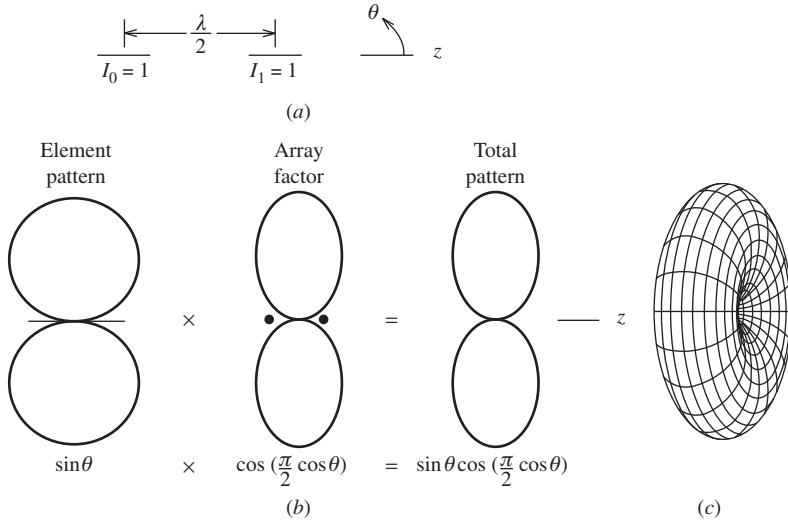
Figure 8-13 Array of two half-wavelength spaced, equal amplitude, equal phase, collinear short dipoles (Example 8-5). (*a*) The array. (*b*) The pattern. (*c*) The 3D pattern (rotated).

rotated slightly to see inside, illustrates the omnidirectional shape. The half-power beamwidth of the array factor is 60° and is 51° for the complete pattern. So, the complete pattern is essentially a slightly narrower version of the array factor.

Collinear arrays are in widespread use in base stations for land mobile communications. Half-wave dipoles spaced more than a half-wavelength apart are popular elements. The array axis is oriented vertically, producing an omnidirectional pattern in the horizontal plane as required for point-to-multipoint communications. Lengthening the array by adding elements narrows the beamwidth in the elevation plane, increasing the directivity and extending the usable range to a mobile unit. Base station antennas are treated further in Sec. 12.2.

Arrays that have parallel elements, as illustrated in Fig. 8-14, have more complicated pattern expressions because the axis of symmetry of the array (the z-axis) is no longer aligned with the axis of symmetry of the elements (the x-axis), as for a collinear array. So the pattern will be a function of both θ and ϕ rather than just θ. The element pattern for the line source element parallel to the x-axis is found from an expression analogous to (8-52):

$$\sin\gamma \int_{-\ell/2}^{\ell/2} i(x')e^{j\beta x' \cos\,\gamma}dx' \tag{8-55}$$
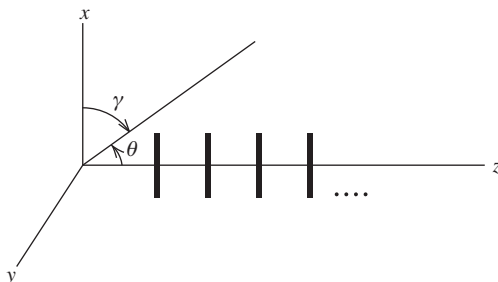
Figure 8-14 A linear array of parallel line sources.

where γ is the spherical polar angle from the x-axis with $0 \leq \gamma \leq 180°$. γ is expressed in terms of θ and ϕ through the relations [H.3: Kraus, 3rd ed., Ch. 16]:

$$\cos\gamma = \sin\theta\cos\phi \quad \text{and} \quad \sin\gamma = \sqrt{1 - \sin^2\theta\cos^2\phi} \tag{8-56}$$

The array factor of (8-53) also applies to this case of parallel elements because the elements are still along the z-axis.

EXAMPLE 8-6   *Two Parallel, Half-Wavelength Spaced Short Dipoles (see Fig. 8-15)*

The pattern of the short dipole elements parallel the x-axis is $\sin\gamma$ analogous to its $\sin\theta$ pattern when parallel to the z-axis. Using (8-56), the element pattern expressed in θ and ϕ is $\sin\gamma = \sqrt{1 - \sin^2\theta\cos^2\phi}$. The array factor is the same as for the previous example because the element locations are the same. The complete pattern is the product of the element pattern and array factor as shown in Figs. 8-15b and c. The fact that the axes of symmetry for the element and the array are orthogonal leads to different principal plane patterns and to the complicated 3D pattern of Fig. 8-15d, which has the axes tilted to show the broad null in the z-direction and the narrow null in the x-direction. While such a pattern has few applications, it is presented here to illustrate how an element pattern can have a major influence on the array pattern and for additional practice in visualizing array patterns.

EXAMPLE 8-7   *Five-Element Endfire Array of Parallel Half-Wave Dipoles*

To illustrate parallel element arrays further, suppose the element antennas of Fig. 8-14 are half-wave dipoles. Also suppose there are five elements arranged and excited for ordinary endfire as in Example 8-3. The complete pattern is the product of the single half-wave dipole element pattern and the array factor found from five isotropic sources. The element pattern for a half-wave dipole element along the x-axis is

$$g_a(\gamma) = \frac{\cos[(\pi/2)\cos\gamma]}{\sin\gamma} \tag{8-57}$$

which is (3-4) with θ replaced by γ. Using (8-56), this becomes

$$g_a(\theta, \phi) = \frac{\cos[(\pi/2)\sin\theta\cos\phi]}{\sqrt{1 - \sin^2\theta\cos^2\phi}} \tag{8-58}$$

The array factor is (8-22) with $N = 5$, or

$$f(\psi) = \frac{\sin\left(\frac{5}{2}\psi\right)}{5\sin\left(\frac{1}{2}\psi\right)} \tag{8-59}$$

For this example, $\alpha = 0.9\pi$ and $d = 0.45\lambda$ so $\psi = \beta d\cos\theta + \alpha = 0.9\pi\cos\theta + 0.9\pi$, and (8-59) is

$$f(\theta) = \frac{\sin(2.25\pi\cos\theta + 2.25\pi)}{5\sin(0.45\pi\cos\theta + 0.45\pi)} \tag{8-60}$$

The total pattern of the array in terms of θ and ϕ is then the product of (8-58) and (8-60):

(a)

(b)

Element pattern   Array factor   Total xz-plane pattern

$\sin\gamma$ $\cos(\frac{\pi}{2}\cos\theta)$ $\sin\gamma\cos(\frac{\pi}{2}\cos\theta)$

(c)

Element pattern   Array factor   Total yz-plane pattern

(d)
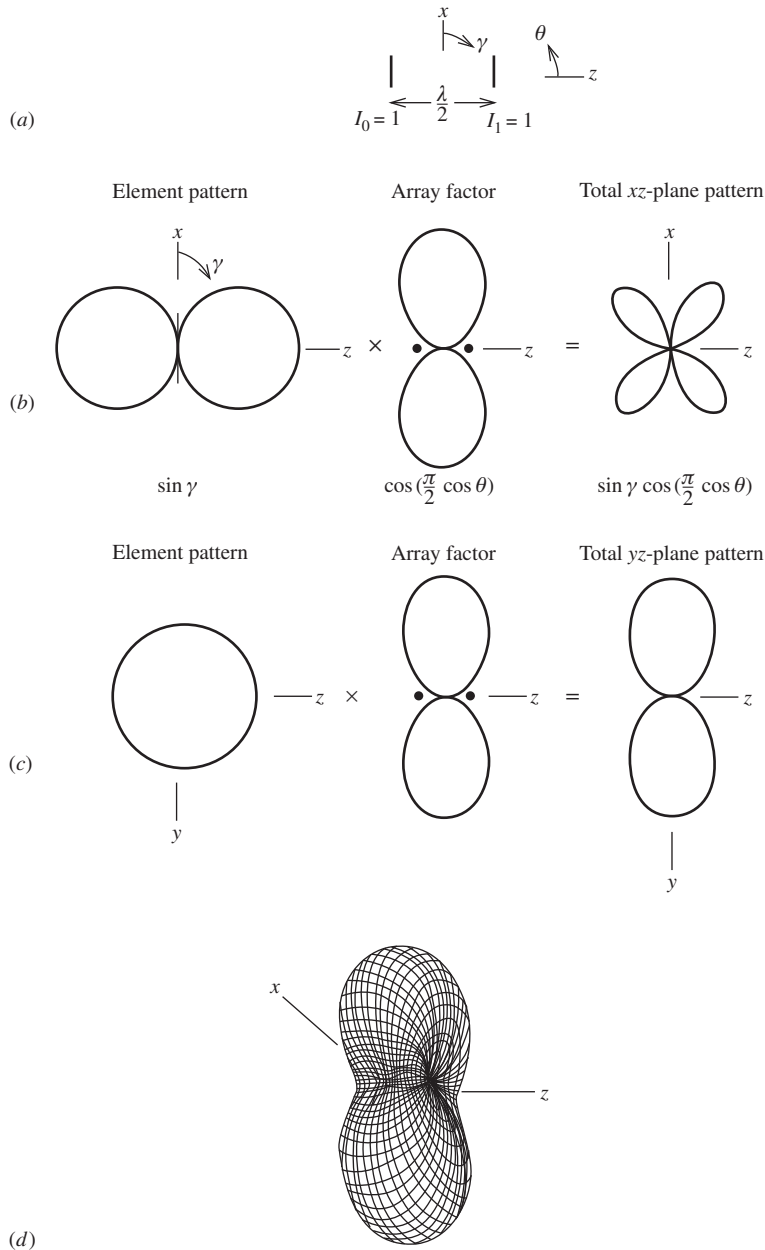
Figure 8-15 Array of two half-wavelength spaced, equal amplitude, equal phase parallel short dipoles (Example 8-6). (a) The array. (b) The xz-plane pattern. (c) The yz-plane pattern. (d) The three-dimensional pattern.

$$F(\theta,\phi) = \frac{\cos[(\pi/2)\sin\theta\cos\phi]}{\sqrt{1-\sin^2\theta\cos^2\phi}}\ \frac{\sin(2.25\pi\cos\theta+2.25\pi)}{5\sin(0.45\pi\cos\theta+0.45\pi)} \qquad (8\text{-}61)$$

The polar plot of this pattern is easily obtained by multiplying the plot in Fig. 3-5b, where the axis of symmetry is now the x-axis instead of the z-axis, times the polar plot of Fig. 8-10b. This is a polar plot similar to the array factor plot except that the endfire lobes are slightly narrower, and there is a pattern zero in the $\gamma = 0°$ direction caused by the element pattern.

8.5   DIRECTIVITY OF UNIFORMLY EXCITED, EQUALLY SPACED LINEAR ARRAYS

The directivity of an antenna is solely determined by its radiation pattern in all space surrounding the antenna. The techniques of the previous section for determining the pattern of an array have prepared us to develop, in this section, exact and approximate formulas for the directivity of uniformly excited, equally spaced linear arrays (UE, ESLA). As expected, array gain equals array directivity multiplied by the array radiation efficiency. To derive directivity expressions, we use $D = 4\pi/\Omega_A$, first finding the beam solid angle as

$$\Omega_A = \iint |F(\theta,\phi)|^2 d\Omega = \iint |g_a(\theta,\phi)|^2 |f(\theta)|^2 \, d\Omega \tag{8-62}$$

where $g_a(\theta,\phi)$ and $f(\theta)$ are the normalized element pattern and linear array factor and $d\Omega = \sin\theta \, d\theta \, d\phi$.

We begin by assuming the elements are equally spaced, uniformly excited, and isotropic. This assumption leads to approximate results for situations where the element pattern is much broader than the array factor and the main beams of both are aligned. The appropriate array factor from (8-22) is

$$|f|^2 = \left| \frac{\sin(N\psi/2)}{N\sin(\psi/2)} \right|^2 \tag{8-63}$$

$$= \frac{1}{N} + \frac{2}{N^2} \sum_{m=1}^{N-1} (N-m) \, \cos m\psi \tag{8-64}$$

where (8-64) is another form for (8-63). This identity can be shown to be true for $N = 2$ since from (8-64), $|f(\psi)|^2 = \frac{1}{2} + \frac{1}{2}\cos\psi = \cos^2(\psi/2)$ as in (8-9). With the simple expression in (8-64), it is easier to perform the integration in (8-62) in terms of the variable ψ. Using $g_a(\theta,\phi) = 1$, $\psi = \beta d \cos\theta + \alpha$, and $\sin\theta \, d\theta = -(1/\beta d)d\psi$ in (8-62) gives

$$\Omega_A = \int_0^{2\pi} d\phi \int_0^\pi |f(\theta)|^2 \sin\theta \, d\theta = 2\pi \int_{\beta d+\alpha}^{-\beta d+\alpha} |f(\psi)|^2 \left(-\frac{1}{\beta d}\right) d\psi$$

$$= \frac{2\pi}{\beta d} \int_{-\beta d+\alpha}^{\beta d+\alpha} |f(\psi)|^2 \, d\psi \tag{8-65}$$

Substituting (8-64) in the above yields

$$\Omega_A = \frac{2\pi}{\beta d} \left[\frac{1}{N} \int_{-\beta d+\alpha}^{\beta d+\alpha} d\psi + \frac{2}{N^2} \sum_{m=1}^{N-1} (N-m) \int_{-\beta d+\alpha}^{\beta d+a} \cos m\psi \, d\psi \right]$$

$$= \frac{2\pi}{\beta d} \left[\frac{1}{N} \psi \Big|_{-\beta d+\alpha}^{\beta d+\alpha} + \frac{2}{N^2} \sum_{m=1}^{N-1} (N-m) \frac{\sin m\psi}{m} \Big|_{-\beta d+\alpha}^{\beta d+\alpha} \right]$$

$$= \frac{2\pi}{\beta d} \left[\frac{1}{N} (2\beta d) + \frac{2}{N^2} \sum_{m=1}^{N-1} \frac{N-m}{m} [\sin m(\beta d+\alpha) - \sin m(-\beta d+\alpha)] \right]$$

$$= \frac{4\pi}{N} + \frac{4\pi}{N^2} \sum_{m=1}^{N-1} \frac{N-m}{m\beta d} 2 \cos m\alpha \sin m\beta d \tag{8-66}$$

where (D-6) was used in the last step.

The directivity is now found from its basic definition of $D = 4\pi/\Omega_A$ with (8-66) as [H.8.2: Ma. p. 23].

$$D = \frac{\left| \dfrac{\sin(N\,\delta/2)}{N\sin(\delta/2)} \right|^2}{\dfrac{1}{N} + \dfrac{2}{N^2}\displaystyle\sum_{m=1}^{N-1}\dfrac{N-m}{m\beta d}\sin m\beta d \cos m\alpha} \qquad UE,\ ESLA\ of\ istropic\ elements \qquad (8\text{-}67)$$

The numerator of this expression is the pattern normalization factor from (8-39), which is required to normalize the pattern to unity peak in the visible region in cases where an additional phase of δ beyond that for ordinary endfire is used; see (8-37). The formula applies for any number of elements N, any spacing d, and any interelement phasing α, but is restricted to isotropic elements. Fig. 8-16 plots this formula for various numbers of elements as a function of element spacing for the broadside case with α and δ zero. Tai [1] presented a family of such plots. All of the directivity curves fall off steeply as the spacing approaches one wavelength. This effect is due to the emergence of grating lobes in the visible region. See Fig. 8-3 as an example where full grating lobes are visible for one-wavelength spacing. Eq. (8-67) reduces to a simple result for elements spaced a multiple of a half-wavelength apart:

$$D = N \qquad for\ isotropic\ elements\ and\ d = n\frac{\lambda}{2} \qquad (8\text{-}68)$$

These values can be located on the graph in Fig. 8-16 for the broadside case, but (8-68) applies for any scan angle.

It is obvious from Fig. 8-16 that maximum directivity is achieved with a broadside UE, ESLA for a spacing of just under a wavelength, with the peak directivity becoming closer to a wavelength for larger numbers of elements. Arrays are often designed to have an element spacing near a wavelength to keep the number of elements as low as possible to limit the cost and complexity of the antenna and associated feed network. A simple formula for the directivity of a broadside array of uniformly excited isotropic elements follows directly from the uniform line source case in (5-19):

$$D \approx 2\frac{L}{\lambda} = 2\frac{Nd}{\lambda} \qquad broadside,\ UELA \qquad (8\text{-}69)$$

where we used the array length of $L = Nd$. This straight line approximation, shown in Fig. 8-16 for the case of 10 elements, is an excellent fit to the exact result up to almost
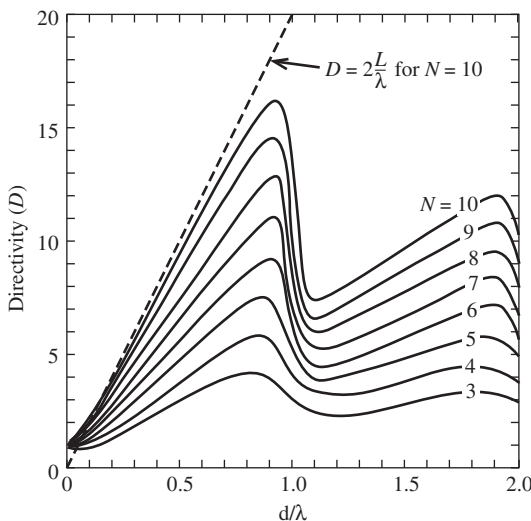
Figure 8-16 Directivity as a function of element spacing for a broadside array of isotropic elements for several element numbers N computed using (8-67).
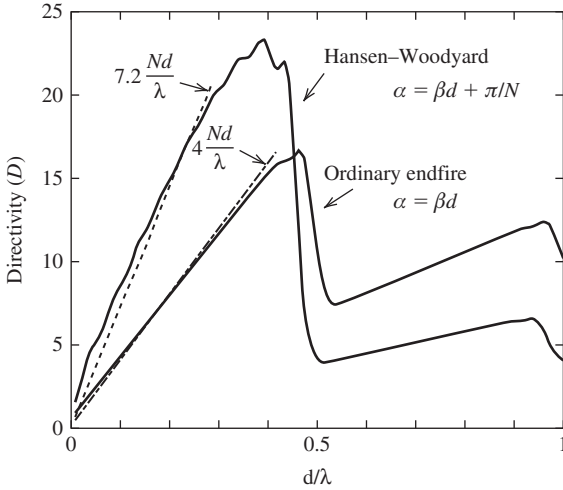
Figure 8-17 Comparison of directivities for two 10-element equally spaced, uniformly excited endfire arrays: ordinary endfire (dotted curve) and Hansen–Woodyard endfire (solid curve) calculated using (8-67). Also shown are straight line approximations using (8-70) and (8-71).

one-wavelength spacing. Note that (8-69) is exact for $d = \lambda/2$ because it equals N, as it should be based on (8-68).

The directivity of endfire arrays has trends similar to broadside arrays as illustrated by the graph in Fig. 8-17 which shows directivity versus element spacing for the case of 10 isotropic elements. Plotted are curves for both ordinary endfire and Hansen-Woodyard cases with phases given by (8-36b) and (8-38b), respectively. The results, again computed using (8-67), show that directivity increases with increasing element spacing until $d = \lambda/2$ is approached when a grating lobe begins to appear in the endfire direction opposite the main beam. Line-source theory results in (5-20) lead to the approximation

$$D \approx 4\frac{L}{\lambda} = 4\frac{Nd}{\lambda} \qquad \textit{ordinary endfire, UELA} \qquad (8\text{-}70)$$

This straight-line approximation shown in Fig. 8-17 for 10 elements is an excellent fit.

There also is a straight-line approximate formula for the Hansen-Woodyard endfire case:

$$D \approx 7.2\frac{L}{\lambda} = 7.2\frac{Nd}{\lambda} \qquad \textit{Hansen-Woodyard endfire, UELA} \qquad (8\text{-}71)$$

Constant values other than 7.2 can be found in the literature. But because the Hansen-Woodyard directivity curve does not follow a straight line as closely as does the ordinary endfire directivity curve, any linear fit would not be as accurate as the ordinary endfire straight line approximation in (8-70), which is within a few percent of the exact result. It is obvious from the Fig. 8-17 that more directivity is achieved with the Hansen-Woodyard array than with an ordinary endfire array of the same number of elements. In fact, the increase based on the ratio of (8-71) to (8-70) is $7.2/4 = 1.82 = 2.6$ dB. As an example, consider five-element endfire arrays similar to those in Examples 8-3 and 8-4. Exact directivity values for the ordinary and Hansen-Woodyard endfire cases from (8-67) are 5.88 and 9.41, respectively, while approximations (8-70) and (8-71) give 6.0 and 10.8.

The higher directivity associated with the Hansen-Woodyard array is due to the increased progressive phase (i.e., $\delta > 0$), which causes the peak of the main beam peak to move into the invisible region and narrows the visible main beam, reducing the beam solid angle and increasing directivity. It should be noted that the Hansen-Woodyard endfire array provides the highest directivity in most cases but not all. For spacings of isotropic elements much below a half-wavelength, the directivity can be increased somewhat by increasing the phase shift per element over that called for with Hansen-Woodyard operation. [2]
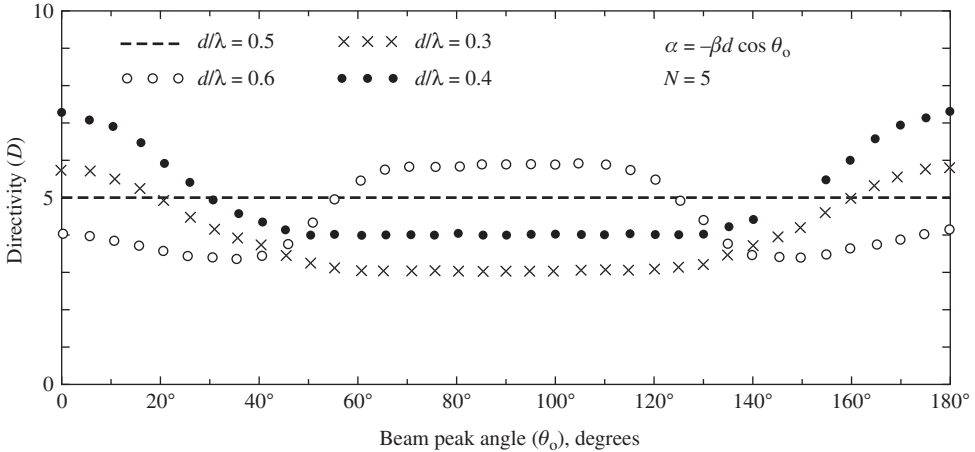
Figure 8-18 Variation of directivity with scan angle for five-element uniformly excited arrays of various element spacings. The elements are isotropic. Values were calculated using (8-67).

If the uniform excitation constraint is relaxed, it turns out that only a slight increase in directivity is possible, unless spacing is under a half-wavelength is permitted. [1] Non-uniformly excited arrays are discussed more fully in the next section.

Broadside and endfire represent the extreme beam scan directions. Fig. 8-18 plots directivity values at all scan angles for a uniform amplitude, five-element linear array of isotropic elements for various spacings. The $d = 0.5\,\lambda$ curve is a straight line equal to 5, which comes from (8-68), and demonstrates that directivity is independent of scan angle for $d = n\lambda/2$. The beamwidths calculated using (8-34) and (8-35) for broadside and endfire are $20.3°$ and $68.2°$, respectively. While the beamwidth is much narrower for the broadside case, the pattern is an omnidirectional doughnut pattern in three dimensions (see Fig. 3-16 for a similar pattern). At endfire, there are two opposing broad, unidi-rectional beams as shown in Fig. 8-10d. It turns out that the beam solid angles, Ω_A, for the two vastly different patterns are the same, giving the same directivity. For the four cases shown in Fig. 8-18, the largest directivity in the broadside direction of $\theta_o = 90°$ is for the widest spacing of $d = 0.6\,\lambda$. As seen from Fig. 8-16, directivity continues to rise for spacings up to almost $d = 0.9\,\lambda$ in the $N = 5$ case. Fig. 8-18 shows that directivity remains constant over a wide range of scan angles around broadside; this will be explained in Sec. 8.9. The greatest directivity in the endfire scan cases ($\theta_o = 0°$, $180°$) occurs for the largest spacing that satisfies the single main beam criterion of (8-36a), which is $d \le 0.45\,\lambda$ as noted in Example 8-3. For the four cases considered, the largest spacing satisfying this single endfire beam condition is 0.4λ and, thus, has the greatest directivity at endfire as can be noted from Fig. 8-18.

There are only a few expressions available similar to (8-67) for the directivity of arrays with real elements (i.e., not isotropic) and directivity is usually found by integration using (2-142) in (2-144). The often-used formula that holds for uniformly excited linear arrays with elements that can be isotropic, collinear short dipoles, or parallel short dipoles is

$$D = \frac{\left|\dfrac{\sin(N\delta/2)}{N\sin(\delta/2)}\right|^2}{\dfrac{a_0}{N} + \dfrac{2}{N^2}\displaystyle\sum_{m=1}^{N-1}\dfrac{N-m}{m\beta d}\left(a_1 \sin m\beta d + a_2 \cos m\beta d\right)\cos m\alpha} \tag{8-72}$$

where $a_0, a_1,$ and a_2 are given in Table 8-1 for the element types. [3, 4] This formula reduces to (8-67) for isotropic elements. Directivity-versus-spacing plots based on this

Table 8-1 Parameters for Use in Computing the Directivity of Uniform Current Amplitude, Equally Spaced Linear Arrays; see (8-67)

Element Type	$|g_a(\theta,\phi)|^2$	a_0	a_1	a_2
Isotropic	1	1	1	0
Collinear short dipoles (Fig. 8-12)	$\sin^2\theta$	$\dfrac{2}{3}$	$\dfrac{2}{(m\beta d)^2}$	$\dfrac{-2}{m\beta d}$
Parallel short dipoles (Fig. 8-14)	$1 - \sin^2\theta\cos^2\phi \;=\sin^2\gamma$	$\dfrac{2}{3}$	$1 - \dfrac{1}{(m\beta d)^2}$	$\dfrac{1}{m\beta d}$

formula are similar to Fig. 8-16 but the directivity values will, of course, be somewhat larger; see [1] for such plots. As an example, (8-72) yields $D = 12.2$ for an array of 10 collinear short dipoles spaced 0.6λ apart, whereas the same array but with isotropic elements has $D = 11.9$ from (8-67) or Fig. 8-16. In general, the directivity of long arrays $(L \gg \lambda)$ is primarily controlled by the array factor if the element pattern has low directivity and its major lobe is aligned with that of the array factor. In such cases, formulas based on isotropic elements like (8-69) to (8-71) can be used. Returning to the 10-element array of collinear short dipoles spaced 0.6λ apart, (8-69) gives $D = 12$, which is close to the exact value of $D = 12.2$.

The optimum (i.e., maximum directivity) array in most cases has uniform element amplitudes and linear phase shifts. For spacings under $\lambda/2$, the Hansen-Woodyard endfire array is close to optimum. Above $\lambda/2$ spacings, uniform amplitude, in-phase excitations are close to maximum with the spacing for peak directivity just under a wavelength and closer to a wavelength for larger numbers of elements. More is said on this topic in the next section.

There are several ways to estimate directivity for the purpose of quick approximate calculation. An approximate directivity formula for a UE, ESLA based on half-power beamwidth can be derived by substituting (5-19) into (5-12), giving

$$D \approx \frac{1.77}{\text{HP}}\text{rad} = \frac{101.5°}{\text{HP}_{\text{deg}}} \tag{8-73}$$

For example, an ES, UELA of 10, half-wave spaced isotropic elements has $\text{HP}_{\text{deg}} = 20.3°$ and (8-73) gives $D = 5$, which happens to be exactly correct from (8-68). For the example of 10 collinear short dipoles spaced 0.6λ apart, the beamwidth is $8.44°$ and then (8-73) gives 12.0 which is close to the correct value of 12.2. The formula also can be used with any omnidirectional antenna, but the following formula is more appropriate [5]:

$$D \approx \frac{101.5°}{\text{HP}_{\text{deg}} - 0.0027\,\text{HP}_{\text{deg}}^2} \tag{8-74}$$

For example, a half-wave dipole which has a beamwidth of $78°$, so (8-74) yields 1.65, which is close to the correct value of 1.64. The example of 10 collinear short dipoles spaced 0.6λ apart with a beamwidth of $8.44°$ has $D = 12.3$ from (8-74), and the correct value is 12.2.

The following simple formula can be found in the literature:

$$D \approx D_e D_i \tag{8-75}$$

where $D =$ directivity of the array, $D_e =$ directivity of one element in the array, and $D_i =$ directivity of the array with isotropic elements (i.e., directivity of the array factor). Application of this approximation is primarily to broadside, single main beam pattern
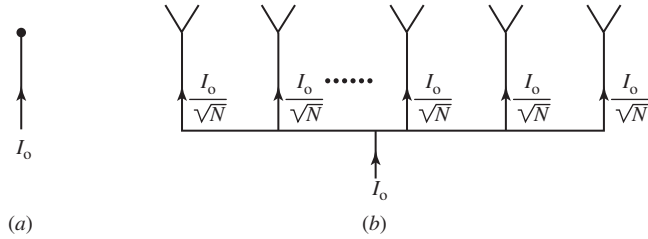
Figure 8-19 Array directivity is the ratio of the maximum radiation intensity of the array compared to that for an isotropic element with the same input power. (*a*) Reference isotropic antenna. (*b*) Array with the same total input power divided equally among the elements.

antennas (that is, grating lobes would not be accounted for). This formula is often presented without restriction, but, in fact, it must be used with caution, especially for linear antennas. [6] An alternate form of this formula that also can be found in the literature is $D \approx D_e N$. This form assumes the directivity of the array of isotropic elements equals the number of elements, $D_i = N$, which is only true for spacings equal to multiples of a half-wavelength (and for uniform current amplitudes), thus ignoring the effect spacing has on pattern and directivity. As an example, (8-75) applied to 10 collinear short dipoles spaced a half wavelength apart gives $D = D_e N = 1.5 \cdot 10 = 15$, which is way above the correct answer of 10.3. The directivity by (8-75) for 10 collinear half-wave dipoles spaced 0.75λ apart is $D = D_e D_i = 1.64 \cdot 14.5 = 23.8$, which again is much higher than the correct result obtained by integration of $D = 15.2$. On the other hand, (8-73) and (8-74) applied to this array, which has a pattern with $HP = 6.8°$, give good results with directivity values of 14.9 and $D = 15.2$, respectively.

It is important to note that array directivity represents the increase in the radiation intensity in the direction of maximum radiation over that of a single element. Consider a single isotropic element and an array of N equally excited isotopic elements as shown in Fig. 8-19. The input power to the array is assumed to divide equally among the array elements, so the element powers are $1/N$ of the input power and the element currents are $1/\sqrt{N}$ of the input current. The radiation intensity U_o for the isotropic element is proportional to its input power, which in turn is proportional to the input current squared I_o^2. The maximum radiation intensity U_{max} of the array in Fig. 8-19*b* is a factor of D greater than that for a single isotropic element with the same input power.

8.6 NONUNIFORMLY EXCITED, EQUALLY SPACED LINEAR ARRAYS

We have seen that the main beam of an endfire array can be narrowed by changing the phase from that which is required for the ordinary endfire case. We can also shape the beam and control the level of the side lobes by adjusting the current amplitudes in an array. General synthesis procedures for achieving a specified pattern are presented in Chap. 10. In this section, a few simple techniques for controlling side lobe levels and beamwidth are introduced. Several examples are given that reveal the relationship between the array current distribution and the radiation pattern. The directivity for arrays with nonuniform excitation are also examined.

The array factor of (8-7) can be written as a polynomial in terms of $Z = e^{j\psi}$ as follows:

$$\text{AF} = \sum_{n=0}^{N-1} A_n e^{jn\psi} = \sum_{n=0}^{N-1} A_n Z^n \qquad (8\text{-}76)$$

where the current amplitudes A_n are real and can be different for each n. It is a simple matter to investigate element current distributions using a computer to perform the array

factor summation. We present the results of several such calculations. The influence of the element current amplitudes will become apparent because we use the same five-element, broadside linear array with a half-wavelength element spacing throughout this section.

The pattern of a uniform array with all current amplitudes equal is plotted in linear, polar form in Fig. 8-20a, and the element currents are shown in Fig. 8-21a. If the element

$D = 5$
$HP = 20.8°$
$SLL = -12$ dB

(a) Uniform.

$D = 4.26$
$HP = 26.0°$
$SLL = -19.1$ dB

(b) Triangular.

$D = 3.66$
$HP = 30.3°$
$SLL = -\infty$ dB

(c) Binomial.

$D = 4.68$
$HP = 23.7°$
$SLL = -20$ dB

(d) Dolph-Chebyshev, −20 dB SLL.

$D = 4.23$
$HP = 26.4°$
$SLL = -30$ dB

(e) Dolph-Chebyshev, −30 dB SLL.
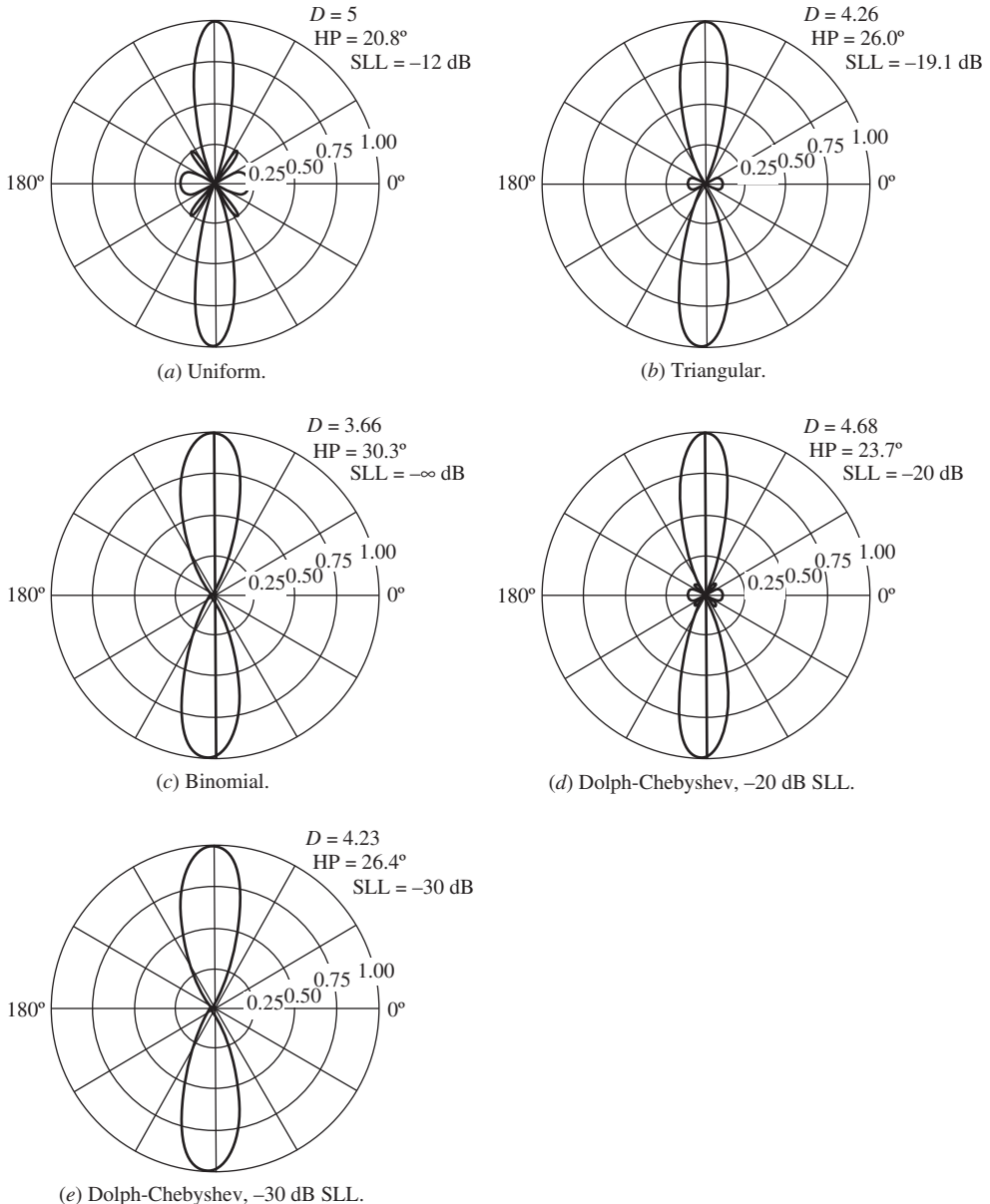
Figure 8-20 Patterns of several uniform phase $(\theta_o = 90°)$, equally spaced $(d = \lambda/2)$ five-element linear arrays with various amplitude distributions. The currents are plotted in Fig. 8-21. (a) Uniform currents, 1:1:1:1:1. (b) Triangular current amplitude distribution, 1:2:3:2:1. (c) Binomial current amplitude distribution, 1:4:6:4:1. (d) Dolph–Chebyshev current amplitude distribution, 1:1.61:1.94:1.61:1, for a side lobe level of −20 dB. See Example 8-5. (e) Dolph–Chebyshev current amplitude distribution, 1:2.41:3.14:2.41:1, with a side lobe level of −30 dB.

(a) Uniform.

(b) Triangular.

(c) Binomial.

(d) Dolph-Chebyshev, –20 dB SLL.

(e) Dolph-Chebyshev, –30 dB SLL.
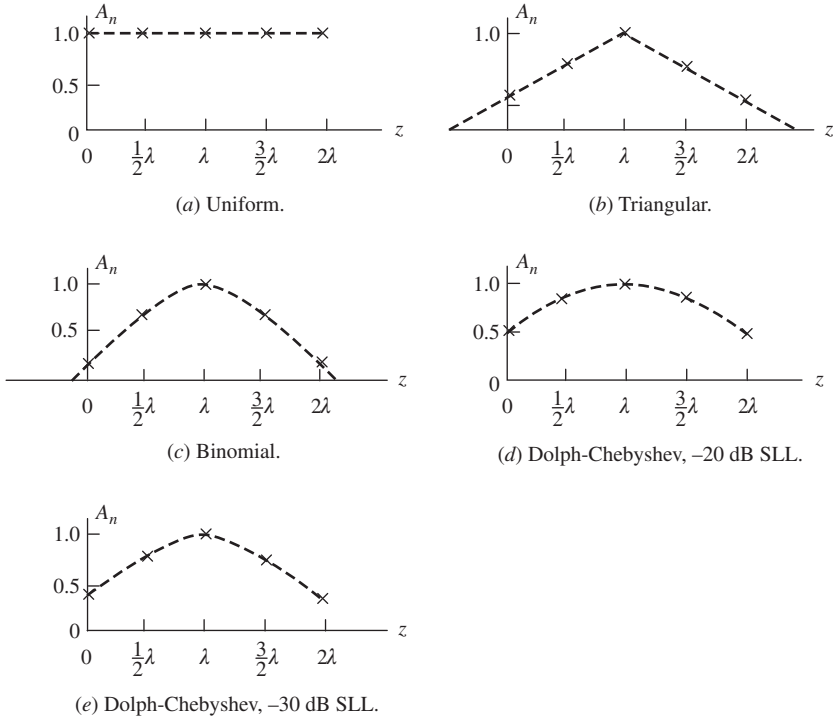
Figure 8-21 Current distributions corresponding to the patterns of Fig. 8-20. The current phases are zero $(\alpha = 0)$. Currents are normalized to unity at the array center.

current amplitudes form a triangle as shown in Fig. 8-21b, the radiation pattern of Fig. 8-20b results. Notice that the side lobes are considerably smaller than those of the uniformly illuminated array, but at the expense of increased beamwidth. This increased beamwidth $(20.8–26.0°)$ is responsible for reduced directivity (from 5 to 4.26).

The side lobe reduction introduced by the triangular amplitude taper suggests that perhaps an amplitude distribution exists such that all side lobes are completely eliminated. Indeed, this is possible if the ratios of the currents are equal to the coefficients of the binomial series. To see how this comes about, first consider a two-element array with equal amplitudes and spacing d. The array factor from (8-76) is $\text{AF} = 1 + e^{j\psi}$ and can be written in terms of $z = e^{j\psi}$ as

$$\text{AF} = 1 + Z \tag{8-77}$$

If the spacing for this broadside array is less than, or at most equal to, a half-wavelength, the array factor will have no side lobes (see Fig. 3-16). Now consider an array formed by taking the product of two array factors of this type:

$$\text{AF} = (1 + Z)(1 + Z) = 1 + 2Z + Z^2 \tag{8-78}$$

This corresponds to a three-element array with the current amplitudes in the ratio $1 : 2 : 1$. Since this array is simply the square of one that had no side lobes, the three-element array also has no side lobes. This process can also be viewed as arraying of two of the two-element arrays such that the centers of each subarray are spaced d apart. This leads to a coincidence of two elements in the middle of the total array, thus giving a current of 2. The total array factor is the product of the "element pattern," which is a two-element

subarray pattern, and the array factor that is again a two-element, equal amplitude array. Thus, the total array factor is the square of one subarray pattern. Continuing this process for an N element array, we obtain

$$AF = (1 + Z)^{N-1} \tag{8-79}$$

which is a binomial series; see (F-4). For $N = 5$,

$$AF = (1 + Z)^4 = 1 + 4Z + 6Z^2 + 4Z^3 + Z^4 \tag{8-80}$$

Therefore, the ratios of the current amplitudes are $1 : 4 : 6 : 4 : 1$. This current distribution is shown in Fig. 8-21c and the resulting pattern is shown in Fig. 8-20c. This pattern is broader than either the uniform or triangular distribution cases and has a lower directivity, but it has no side lobes.

From these five-element array examples, a trend is apparent: *As the current amplitude is tapered more from the center to the edge of an array, the side lobes decrease and the beamwidth of the main beam increases*. Many applications where interference/jamming concerns are important require low side lobes, but at the cost of reduced directivity due to the increased beamwidth. The tradeoff between beamwidth and side lobe level is a principle encountered frequently in antenna design that applies to both arrays and continuous distributions, and to slices containing the normal to two-dimensional apertures as well. An example is line sources with uniform and cosine tapered currents shown in Figs. 5-5a and 5-8a. The peak side lobe levels of the tapered distribution are 9.7 dB lower $(23 - 13.3)$ than the uniform case, but with a beamwidth that is a factor of 1.34 wider $(1.19/0.886)$, based on the data in Table 5-2.

The beamwidth/side lobe level tradeoff can be optimized. In other words, it is possible to determine the element current amplitudes such that the beamwidth is minimum for a specified side lobe level, or conversely to specify the beamwidth and obtain the lowest possible side lobe level. This array is referred to as a Dolph–Chebyshev array and it provides a pattern with all side lobes of the same level. The Dolph–Chebyshev array synthesis procedure is explained in detail in Sec. 10.4.1. For a five-element array with an element spacing of a half-wavelength and a specified side lobe level of -20 dB, the Dolph–Chebyshev current distribution is plotted in Fig. 8-21d and the corresponding pattern is shown in Fig. 8-20d. If the side lobe level for the Dolph–Chebyshev array is specified to be -30 dB, the distribution is that of Fig. 8-21e and the corresponding pattern is shown in Fig. 8-20e. Note that the main beam is slightly broader than in the previous case where the side lobe level was 10 dB higher.

The discussion of nonuniformly excited arrays thus far was for amplitudes that are tapered toward the ends of the linear array. If the amplitude distribution becomes larger at the ends of the array (called an inverse taper), we expect the opposite effect; that is, the side lobe level increases and the beamwidth decreases. Suppose, for example, that we invert the triangular distribution such that the amplitudes are $3 : 2 : 1 : 2 : 3$. The resulting pattern shown in Fig. 8-22 demonstrates the expected decrease in beamwidth and increase in side lobe level. Although the directivity for the inverse triangular tapered current is greater than that for the triangular taper of Fib. 8-20b, it is still not as large as that produced by the uniform distribution.

The directivity values were given for each of the examples in this section. We close this section by developing the directivity expression. With little additional complexity, we expand the treatment to include unequal element spacings as well as nonuniform excitation. The element positions along the z-axis are z_n and the element current amplitudes are A_n. If the element phases are linear with distance, then $\alpha_n = -\beta z_n \cos \theta_o$, where θ_o is the angle of the pattern maximum; the applications of this type of phasing are discussed in Sec. 8.9.1. The array factor of (8-53) is then appropriate and when normalized is

(a) The array factor.

(b) The current distribution, $3:2:1:2:3$.
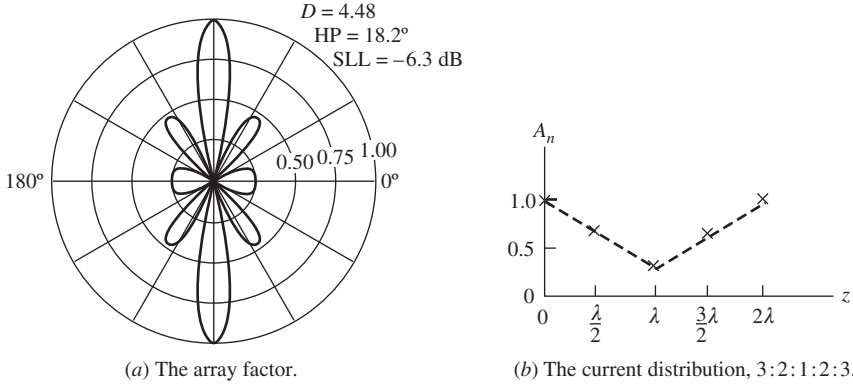
Figure 8-22 The inverse triangular tapered, five-element linear array with $d = \lambda/2$ and $\theta_o = 90°$.

$$f(\theta) = \frac{\sum\limits_{n=0}^{N-1} A_n e^{j\alpha_n} e^{j\beta z_n \cos\theta}}{\sum\limits_{n=0}^{N-1} A_n} \tag{8-81}$$

The appropriate beam solid angle expression is

$$\Omega_A = 2\pi \int_0^\pi |f(\theta)|^2 \sin\theta \; d\theta$$

$$= \frac{2\pi}{\left(\sum\limits_{k=0}^{N-1} A_k\right)^2} \sum\limits_{m=0}^{N-1} \sum\limits_{p=0}^{N-1} A_m A_p e^{j(\alpha_m - \alpha_p)} \int_0^\pi e^{j\beta(z_m - z_p)\cos\theta} \sin\theta \; d\theta \tag{8-82}$$

Evaluating the integral in the above expression and applying the result to $D = 4\pi/\Omega_A$ yields

$$D = \frac{\left(\sum\limits_{k=0}^{N-1} A_k\right)^2}{\sum\limits_{m=0}^{N-1} \sum\limits_{p=0}^{N-1} A_m A_p e^{j(\alpha_m - \alpha_p)} \dfrac{\sin[\beta(z_m - z_p)]}{\beta(z_m - z_p)}} \tag{8-83}$$

where $\alpha_n = -\beta z_n \cos\theta_o$ and the elements can have any positions z_n and current amplitudes A_n. This general result simplifies for a broadside, equally spaced array to

$$D = \frac{\left(\sum\limits_{k=0}^{N-1} A_k\right)^2}{\sum\limits_{m=0}^{N-1} \sum\limits_{p=0}^{N-1} A_m A_p \dfrac{\sin[(m-p)\beta d]}{(m-p)\beta d}} \qquad \alpha_n = 0, \quad z_n = nd \tag{8-84}$$

As another special cases, when the spacings are a multiple of a half-wavelength, (8-83) reduces to

$$D = \frac{\left(\sum\limits_{n=0}^{N-1} A_n\right)^2}{\sum\limits_{n=0}^{N-1} (A_n)^2} \quad d = \frac{\lambda}{2}, \lambda, \ldots \tag{8-85}$$

Note that this is independent of scan angle θ_o, as indicated in Fig. 8-18 for $d = \lambda/2$. Also, if the amplitudes are uniform, (8-85) yields $D = N$ as given by (8-68). For a further example, consider the triangular excitation with the pattern of Fig. 8-20b. The directivity value from (8-85) is $[2(1)+ 2(2) + 3]^2/[2(1)^2 + 2(2)^2 + (3)^2] = 4.26$. Eq. (8-85) is a very instructive formula because it shows that directivity is a measure of the coherent radiation from the linear array. The numerator is proportional to the square of the *total* coherent field, whereas the denominator is proportional to the *sum* of the squares of the field from each of the elements.

There is no closed-form expression for directivity that includes element pattern effects in an array with weighted excitations that is analogous to (8-72), which is for uniformly illuminated linear arrays of simple element types. Instead directivity is found by integrating the pattern to find Ω_A and using $D = 4\pi/\Omega_A$. This is a relatively easy task with a math applications computer package.

Directivity can be increased without limit through close element spacings with extreme amplitude and phase changes across the array. As mentioned in Sec. 5.5, there are penalties with using superdirective antennas and they are not practical except for modest increases in directivity over that achieved with uniform excitation. For equally spaced, collinear arrays of half-wave dipoles that can have nonuniform amplitude excitations, the maximum directivity is only slightly larger than for uniform excitation up to a spacing of $0.7\,\lambda$ and is nearly identical to it for greater spacings. Hansen [H.8.11: Hansen] provides more details on superdirective arrays and superdirectivity in general.

8.7   MUTUAL COUPLING IN ARRAYS

Arrays so far have been treated as having non interacting elements and as being perfectly matched in impedance to the feed network. This allows the following further assumptions: (i) The element terminal currents are proportional to their incident signals, (ii) The relative current distributions across each element are identical (although they can be of different levels proportional to the excitations), and (iii) Pattern multiplication is valid. As might be expected, in a real array the elements interact with each other which alters the currents (and thus impedances). This interaction, called *mutual coupling*, changes the current magnitude, phase, and distribution on each element from their free-space values. As a consequence, the total array pattern is altered from the no-coupling case. In this section, we discuss the effects of mutual coupling on impedance and pattern, followed by methods to evaluate array element impedance and the array pattern in the presence of mutual coupling.

First, we present names and definitions for several terms used in the field of array antennas mainly following the classic work by Hannan [8], but other names appear in the literature for the same quantities. A *passive array* uses a single generator[2] followed by power dividers and phase adjustment devices. An *active array* has a separate generator at each element, which is the most common array model. The input impedance of an element that has been removed from the array and is isolated from all objects is called *isolated element impedance* or simply *antenna impedance*. Of course, the practical case is a fully active array with all elements excited to produce the desired beam shape, scan angle, and so on. The impedance of an element in its array environment with all elements

---

[2] Arrays are usually described for the transmit case, but the results apply to receiving arrays as well.

active is referred to as the *active impedance*, *driving impedance* or simply *input impedance*.[3]
The active impedance for the elements will vary with position in the array and the excitations,
including phasing for scan. Active impedance is difficult to measure because of the
requirement for the array to be fully active, although simulations can be made for these
conditions. In practice, element impedance behavior is characterized by exciting the
one element and passively terminating all others, typically with nominal generator
impedance or a reference load such as $50\,\Omega$. This is often called *passive impedance*. In
most situations, passive impedance closely approximates active impedance because the
coupling environment is identical and the terminating impedance conditions are nearly
identical. Thus, it is more common to use the term active impedance in place of passive
impedance and we will adopt this approach. To distinguish cases of fully active arrays,
we use the term "active impedance in a fully active array."

8.7.1   Impedance Effects of Mutual Coupling

The three mechanisms responsible for mutual coupling are illustrated in Fig. 8-23*a*. First
is direct space coupling between array elements. Second, indirect coupling can occur by
scattering from nearby objects such as a support tower. Third, the feed network that
interconnects elements in the array provides a path for coupling. In many practical arrays,
feed network coupling can be minimized through proper impedance matching at each
element. This permits each element in the array to be modeled with independent gen-
erators as in Fig. 8-23*b*, where the *m*th element has an applied generator voltage and
terminal impedance given by V_m^g and Z_m^g. The voltage and current at the element term-
inals, V_m and I_m, include all coupling effects. An array of N elements is then treated as an
N port network using conventional circuit analysis, giving
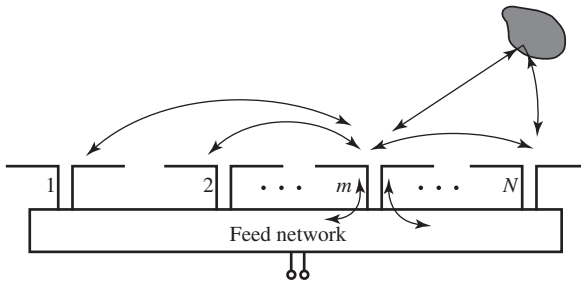
(*a*) Mechanisms for coupling between elements of an array.
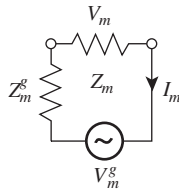
(*b*) Model for *m*th element in an array.

Figure 8-23 Mutual coupling in
a fully excited array antenna.

---

[3] Some authors have used the term *scan impedance* in place of active impedance.

$$V_1 = Z_{11}I_1 + Z_{12}I_2 + \cdots + Z_{1N}I_N$$

$$V_2 = Z_{12}I_1 + Z_{22}I_2 + \cdots + Z_{2N}I_N$$

$$\vdots$$

$$V_N = Z_{1N}I_1 + Z_{2N}I_2 + \cdots + Z_{NN}I_N$$

(8-86)

where V_n and I_n are the impressed current and voltage in the nth element, and Z_{nn} is the *self-impedance* of the nth element when all other elements are open-circuited. The *mutual impedance* Z_{mn} ($= Z_{nm}$ by reciprocity) between the two terminal pairs of elements m and n is the open circuit voltage produced at the first terminal pair divided by the current supplied to the second when all other terminals are open-circuited; that is,

$$Z_{mn} = \frac{V_m}{I_n}\bigg|_{I_i = 0} \quad \text{for all } i \text{ except } i = n$$

(8-87)

For example, $Z_{12} = Z_{21} = V_2/I_1$ with port 2 open circuited.

Mutual impedance is, in general, difficult to compute or measure. However, wide availability of commercial simulation codes based on moment methods discussed in Chap. 14 have made the evaluation of mutual impedance values for wire antennas relatively easy. Before showing simulation results, we discuss how mutual impedance between two antennas is measured, and the results are easily generalized for the determination of mutual impedance between any two elements of an arbitrary array, Z_{mn}. An antenna when isolated in free space with voltage V_1 and current I_1 has an input impedance of

$$Z_{11} = \frac{V_1}{I_1} \quad \text{single isolated element}$$

(8-88)

If a second antenna is brought into proximity with the first, then radiation from the first antenna induces currents on the second, which in turn radiates and also influences the current on the first antenna. The second antenna can either be excited or unexcited (parasitic), but in any case it has terminal current I_2. Then the total voltage at the first antenna is

$$V_1 = Z_{11}I_1 + Z_{12}I_2$$

(8-89)

Similarly, the voltage at the terminals of the second antenna is expressed by

$$V_2 = Z_{21}I_1 + Z_{22}I_2$$

(8-90)

Note that (8-86) is a generalization of (8-89) and (8-90).

Now suppose the second antenna has a load impedance Z_2^g across its terminals ($V_2^g = 0$) such that $V_2 = -Z_2^g I_2$. We can write (8-90) as

$$-Z_2^g I_2 = Z_{21}I_1 + Z_{22}I_2$$

(8-91)

Solving for I_2 and using $Z_{21} = Z_{12}$, we obtain

$$I_2 = \frac{-Z_{21}I_1}{Z_{22} + Z_2^g} = \frac{-Z_{12}I_1}{Z_{22} + Z_2^g}$$

(8-92)

Substituting this into (8-89) and dividing by I_1, we find that

$$\frac{V_1}{I_1} = Z_1 = Z_{11} - \frac{(Z_{12})^2}{Z_{22} + Z_2^g}$$

(8-93)
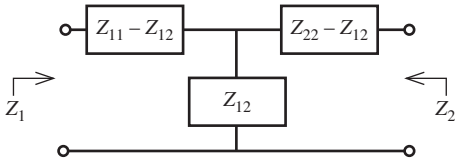
Figure 8-24 Network representation of the coupling between two antennas.

This expresses the input impedance in terms of the two self-impedances (Z_{11} and Z_{22}), the mutual impedance Z_{12}, and the load Z_2^g at the unexcited terminals of antenna 2.

The above discussion suggests the equivalent circuit of Fig. 8-24 for the coupling between two resonant antennas. For a single isolated antenna (i.e., antenna 2 very far away), $Z_{12} = 0$, and (8-93) gives the input impedance equal to the self-impedance, $Z_1 = Z_{11}$. If antenna 2 is open-circuited, then $Z_2^g = \infty$ and (8-93) gives $Z_1 = Z_{oc} = Z_{11}$. Open circuiting implies that the current all along antenna 2 is reduced to zero. This occurs for antennas such as half-wave dipoles, where resonant behavior is eliminated by open circuiting. In other antennas (such as full-wave dipoles), even with an open circuit there will be current induced on the antenna. In this case, the second antenna should be removed.

The general procedure for determining mutual impedance from open-circuit and short-circuit measurements involves the following steps. [H.6: Hansen, Vol. II, pp. 157–160]:

1. Open circuit (or remove) antenna 2. Measure $Z_{oc} = Z_{11}$ at the terminals to antenna 1. For identical antennas, $Z_{22} = Z_{11}$.
2. Short circuit antenna 2. Measure Z_{sc} at the terminals to antenna 1.
3. Compute Z_{12}, using

$$Z_{12} = \sqrt{Z_{oc}(Z_{oc} - Z_{sc})} \tag{8-94}$$

This follows from (8-93) with $Z_2^g = 0$.

The proper sign must be chosen with (8-94). This is aided by examining variations with spacing in the limit of small spacing and maintaining continuity through zero crossings [9]. Good agreement between simulations are experimental results based on this method is reported in [9].

Mutual impedance is computed by placing a generator on the input of one antenna, finding the voltage appearing at the input terminals of the second antenna, and then taking the ratio of the voltage to the current. Results of simulations using a moment method code based on principles in Chap. 14 for two free-space resonated half-wave dipoles that are oriented parallel to each other as a function of separation distance is plotted in Fig. 8-25. Similar results are presented in Fig. 8-26 for dipoles that are collinear. The mutual impedance values decrease with increasing separation distance in both dipole orientation cases. The general trends for mutual coupling in arrays are as follows [H.6: Rudge et al., Vol. II, Sec. 10.3]:

a. The magnitude of mutual impedance decreases with spacing distance d, in many cases decaying as $1/d^2$ [H.8.2: Hansen, "Phased Array Antennas," 2nd ed., p. 225].
b. The far field pattern is an indicator of coupling between elements, although the coupling mechanism is a near-field, not far-field, effect. Coupling is proportional to the element pattern level in the array plane (or surface). And, elements with a narrow pattern will have lower coupling than a elements with a broad beam.
c. Elements with polarizations (i.e., electric field orientations) that are parallel couple more than when collinear.
d. Larger elements have smaller coupling.
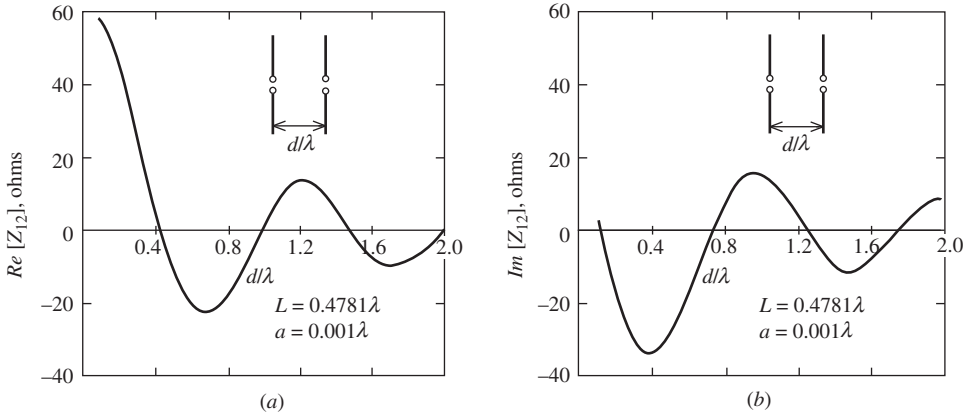
Figure 8-25 The mutual impedance between two resonant parallel dipoles as a function of their spacing relative to a wavelength. (*a*) The real part. (*b*) The imaginary part.
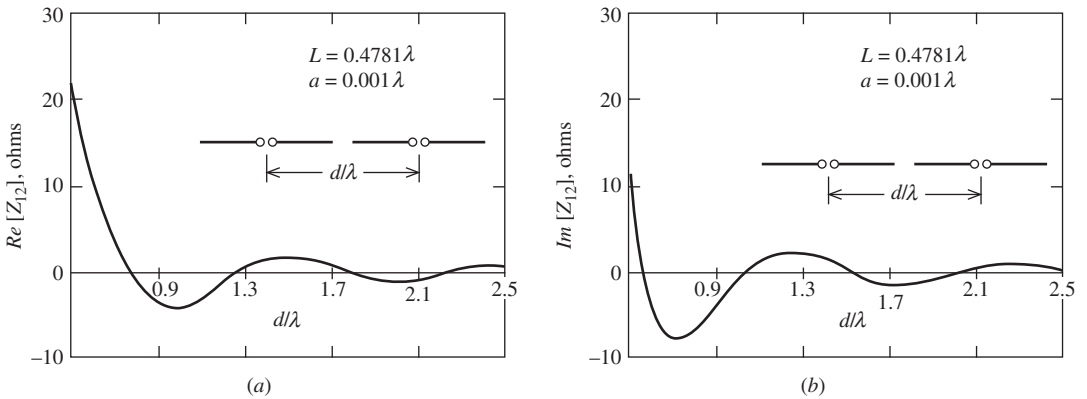
Figure 8-26 The mutual impedance between two resonant collinear parallel dipoles as a function of their spacing relative to a wavelength. (*a*) The real part. (*b*) The imaginary part.

Figs. 8-25 and 8-26 demonstrate these trends. Trend (a) is clear, and trends (b) and (c) are evident when comparing the parallel and collinear dipole results.

A second way to quantify coupled arrays in addition to the mutual impedance approach is to use the scattering parameter method that is commonly used in radio-frequency circuit analysis. For an N-element array, the scattering approach consists of incident voltage wave with an $N \times 1$ column matrix $[a]$ and reflected voltage wave $N \times 1$ column matrix $[b]$ that are related by an $N \times N$ square *scattering matrix* $[S]$ through the matrix equation $[b] = [S][a]$. The voltage wave matrix and scattering matrix entries (called *scattering parameters* or *scattering coefficients*) are complex valued, and carry magnitude and phase information. The scattering parameter S_{mn} is the ratio of the output voltage wave amplitude at port m (b_m) due to the input voltage wave amplitude at port n (a_n) with all other ports not excited. The scattering parameters are popular because they relate directly to measurements, and network analyzers output the values. The scattering matrix diagonal entries are the reflection coefficients at each port, so

$$S_{nm} = \Gamma_m \tag{8-95}$$

Just as with the impedance matrix, the scattering matrix is symmetric for a reciprocal device and $S_{nm} = S_{mn}$. The off-diagonal entries represent the coupling between ports. So the coupling between ports m and n in dB, C_{mn}, is

$$C_{mn} = 20 \log |S_{mn}| \qquad [\text{dB}] \qquad (8\text{-}96)$$

The scattering matrix entries can be found from the impedance matrix entries, Z_{mn}. For element one of a two-element array with all connecting transmission lines of characteristic impedance Z_o:

$$S_{11} = \frac{(Z_{11} - Z_o)(Z_{22} + Z_o) - Z_{12}^2}{(Z_{11} + Z_o)(Z_{22} + Z_o) - Z_{12}^2} \qquad (8\text{-}97)$$

$$S_{12} = \frac{2Z_o Z_{12}}{(Z_{11} + Z_o)(Z_{22} + Z_o) - Z_{12}^2} \qquad (8\text{-}98)$$

We can perform some simple reality checks on these equations. For no mutual coupling ($Z_{12} = 0$), (8-97) gives $S_{11} = \Gamma_1 = (Z_{11} - Z_o)/(Z_{11} + Z_o)$, which is the conventional one-port reflection coefficient formula. Also, for no mutual coupling (8-98) yields $S_{12} = 0$, as should be the case.

To illustrate how mutual coupling values are used to calculate coupling, consider the two-element array of parallel half-wave dipoles with a spacing of 0.6λ. Fig. 8-25 gives $Z_{12} = -23 - j17 \, \Omega$ and using a matched condition with $Z_{11} = Z_o = 70 \, \Omega$, (8-98) in (8-96) yields a coupling level of $C_{11} = -13.7 \, \text{dB}$. As we indicated earlier, coupling levels vary widely as dipole spacing and orientation are changed. Microstrip elements on a common substrate have coupling levels that decrease monotonically from -15 to $-40 \, \text{dB}$ as element spacing is increased from about 0.5 to 1.5λ in the H-plane and from -20 to $-30 \, \text{dB}$ in the E-plane. [H.8.4: Bancroft, p. 175] Prob. 8.7-4 addresses mutual impedance calculation using simulations.

The input impedance of an element in the array can now be computed using the mutual impedance values. For the mth element, the input impedance is found using (8-86) as

$$Z_m = \frac{V_m}{I_m} = Z_{m1} \frac{I_1}{I_m} + Z_{m2} \frac{I_2}{I_m} + \cdots + Z_{mN} \frac{I_N}{I_m} \qquad (8\text{-}99)$$

This is the input impedance in the presence of all elements; i.e., the *active impedance* in a fully excited array. All coupling effects are included via the mutual impedances found from numerical simulations or from measurements using (8-94). The formula clearly shows the dependence of the input impedance on not only the mutual impedances, but also on the terminal currents of each element. Of practical significance is that input impedance is affected by the phase of the terminal currents as it is adjusted to phase-scan the pattern. Thus, the input impedance of all elements in an array will change with scan angle.

8.7.2   Array Pattern Evaluation Including Mutual Coupling

In addition to the element impedances, mutual coupling influences the radiation properties of an array as will be discussed in this section. Complete analysis using simulations or measuring the parameters of each embedded element antenna can be used for full characterization of the array, but this is tedious and rarely done. Instead, approximate techniques for including mutual coupling effects are employed, as we will present. Arrays can be modeled using all mutual impedances explicitly, or by absorbing coupling effects into the currents (isolated-element pattern approach) or into the element pattern (active-element

pattern approach), which is the approach we take here. [H.6: Hansen, Vol. II, Chap. 3; H.8.2: Mailloux, 2nd ed., Sec. 2.1].

In the isolated-element pattern approach, all coupling effects in the total array pattern are accounted for in the excitations:

$$F_{\text{un}}(\theta, \phi) = g_i(\theta, \phi) \sum_{m=1}^{N} I_m e^{j\xi_m} \tag{8-100}$$

where ξ_m is the total phase contribution (usually referenced to the center of the array) due to spatial phase delay. It is the classical array pattern approach of Sec. 8.4 consisting of the product of an *isolated-element pattern* $g_i(\theta, \phi)$ and an array factor. Without coupling effects, the currents $\{I_m\}$ vary in proportion to the excitation voltages. Coupling effects are included using the simple circuit model of the mth element in Fig. 8-23b:

$$I_m = \frac{V_m^g}{Z_m^g + Z_m} \tag{8-101}$$

This is called "free excitation" because the element terminal voltage will change with scan angle. Instead of a constant-voltage feed, it is a constant-incident (available) power feed as is the case in Fig. 8-19. For an infinite array of identical elements in a uniform grid where all elements have same coupling, each element "sees" the same environment and the active impedances are identical, so that all $\{Z_m\}$ equal Z_A. Then the currents are proportional to the voltages across the element terminals:

$$I_m = \frac{V_m^g}{Z^g + Z_A} \propto V_m \tag{8-102}$$

This situation applies to large, equally spaced arrays. It must be pointed out that the common implementation of (8-100) for finite arrays uses the approximation of (8-102), thereby ignoring terminal current variations due to mutual coupling and only including generator voltage variations. It is difficult to obtain accurate current information so that (8-100) can be evaluated, and the active-element pattern method is usually employed; this method is discussed next.

In the active-element pattern approach, all coupling effects are accounted for through the active element. The *active-element pattern* $g_{\text{ae}}^n(\theta, \phi)$ is obtained by exciting only the nth element and loading all other elements with the generator impedance Z^g. The active-element pattern arises from direct radiation from the nth element combined with fields reradiated from the other elements, which in turn receive their power through spatial coupling from the element n. The array pattern in this formulation is

$$F_{\text{un}}(\theta, \phi) = \sum_{n=1}^{N} g_{\text{ae}}^n(\theta, \phi) I_n e^{j\xi_n} \tag{8-103}$$

Here, the currents $\{I_n\}$ are proportional to excitation voltages $\{V_n\}$ as in (8-102). All mutual coupling effects are incorporated into the active-element patterns $g_{\text{ae}}^n(\theta, \phi)$, which depend on the element characteristics and the array geometry. To represent the possibility of gain variations, the active-element pattern levels are relative to a reference element near the center of the array.

It would be tedious to measure active-element patterns for each element in an array. Fortunately, this is usually not necessary. For a large array of identical elements in an equally spaced array, each element sees the same environment of nearest neighbors, except for the edge elements. The appropriate approximation is to factor (8-103) using an *average active-element pattern,* $g_{\text{ae}}(\theta, \phi)$, which is the normalized pattern for a typical central element in the array:

$$F_{\text{un}}(\theta, \phi) = g_{\text{ae}}(\theta, \phi) \sum_{n=1}^{N} I_n e^{j\xi_n} \qquad (8\text{-}104)$$

The advantage of this approach is that the summation in this equation is the array factor based on simple theory without mutual coupling. All coupling effects are contained in the average active-element pattern, which is found by a single pattern measurement of a central element in a large array.

The normalized form of (8-104) gives the important approximate result:

$$\boxed{F(\theta, \phi) = g_{\text{ae}}(\theta, \phi) f(\theta, \phi)} \qquad (8\text{-}105)$$

where

$g_{\text{ae}}(\theta, \phi)$ = average active-element pattern
$f(\theta, \phi)$ = array factor
$F(\theta, \phi)$ = array pattern

This is of the same form as the pattern multiplication formula of (8-54), but mutual coupling effects are included. This approximation is widely used in practice. The average active-element pattern also plays an important role in array construction. The pattern of a single active element surrounded by several passively terminated elements is measured and evaluated before the fully active large array is built and tested. If the active-element pattern is not well formed, the full array will not be either. It has been shown that the active input impedance can be obtained from the active-element pattern data. [10]

The isolated-element pattern will differ from the active-element patterns of the elements in the array. But, for a sufficiently large array, the average-element pattern is representative of the majority of the active-element patterns in the array. How large an array has to be before (8-105) can be applied is an important question that is addressed through guidelines and examples. As array size becomes large (ideally, an infinite number of elements), the active-element patterns for the individual elements converge toward being identical and being equal to the average-active element pattern. Elements near the edge of a finite array have patterns that deviate from the pattern of central elements, which should be close to the average-element pattern. This is illustrated by the patterns in Fig. 8-27 that shows the active element patterns for each element in an array of eight microstrip patch antennas spaced 0.57λ apart. The polar-dB patterns shown for selected elements were measured for that element excited and all others match loaded. The patterns for elements 2 through 7 (only pattern 5 is shown) are nearly identical and are symmetric. The edge element patterns (1 and 8) are distorted due to the asymmetric array environment and due to the finite ground plane edge effects. Note that most pattern distortion is on the side away from the array. Similar results have been reported for dipoles in front of a ground plane. [H.3: Milligan, Sec. 12.5; H.8.2: Haupt, Sec. 6.4] The topic of a representative element pattern in a finite array is revisited in Sec. 8.9.1.
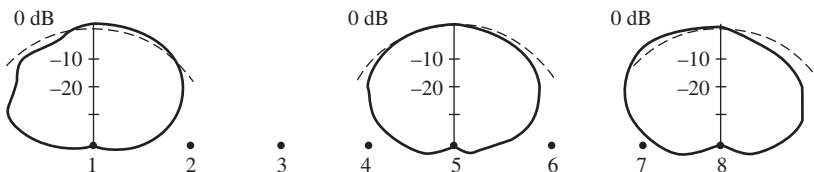
Figure 8-27 Measured active-element patterns for three elements, an interior element (5) and the end elements (1 and 8), of a linear array of eight microstrip patch elements spaced 0.57λ apart.

The full array pattern is, of course, affected by mutual coupling within the array. Arrays are modeled with voltage generator excitations based on a desired set of terminal currents. In the absence of mutual coupling, the terminal currents are proportional to the excitation voltages, (8-102) applies, and the desired pattern is realized. With coupling present the pattern is distorted, but it often turns out that the terminal currents can deviate rather far from the desired excitations and the pattern is not greatly affected. A linear array of 12 half-wave spaced, parallel, half-wave dipoles with excitation voltages designed to steer the beam 45° off broadside has a pattern in Fig. 14-27 (dashed curve) that does not differ much from the ideal no-coupling excitation case (solid curve). The beam is steered correctly as well. Although the element currents and impedances are not close to the desired values (see Table 14-1), the integrating effect of the pattern tends to smooth out the variations; also see [11]. In addition to pattern, the gain and polarization of the total array pattern are affected by mutual coupling. Guidelines similar to those used for impedance in Sec. 8.7.1 also apply to radiation from arrays.

Techniques are available to compensate for mutual coupling effects. Mutual coupling does not change the current distribution on elements in the array but does change the complex terminal currents. Compensation schemes essentially anticipate how the terminal currents will be altered by mutual coupling and change the excitation voltages to produce the desired currents. [H.3: Milligan, Sec. 3.11; 12] There are also physical means to reduce mutual coupling directly such as inserting conducting baffles (i.e., fences) between elements. [H.8.3: Haupt, Sec. 6.8].

8.8   MULTIDIMENSIONAL ARRAYS

Linear arrays have a number of limitations. For instance, they can be phase-scanned in only a plane containing the line of the elements' centers. The beamwidth in a plane perpendicular to the line of element centers is determined by the element beamwidth in that plane. This usually limits the realizable gain. Thus, multidimensional arrays are used for applications requiring a pencil beam, high gain, or main beam scanning in any direction. With advances in fabrication and integrated feed electronics, the costs of large multidimensional arrays are affordable in many situations. Multidimensional arrays are classified by three characteristics: The geometric shape of the surface on which the element centers are located, the perimeter of the array, and the grid geometry of the element centers. The surface on which elements are placed can be linear, circular, planar, etc. The perimeter of planar arrays is usually circular, rectangular, or square in shape. Fig. 8-28 illustrates a planar array with a rectangular perimeter. The array grid (or lattice) can have equal or unequal row and column spacings. A planar array with equal element spacings of d_x and d_y in the principal planes such as in Fig. 8-28 is referred to as having a rectangular grid. If $d_x = d_y$, the grid is said to be square. A triangular grid is also widely used. When the array conforms to a complicated surface such as the fuselage of an aircraft, the array is
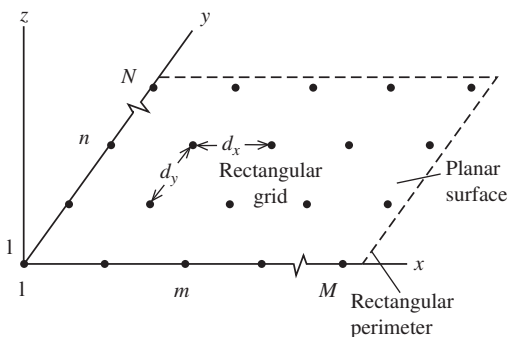
Figure 8-28 Geometry of a planar array.

said to be conformal. In this section, we present techniques for analyzing arrays of arbitrary geometry as well as a few important special case geometries.

The pattern multiplication principles developed in Sec. 8.4 for linear arrays apply to arrays of any geometry as long as the elements are similar. That is, if the elements are identical and oriented in the same direction, the total array pattern is factorable as in (8-105), which includes mutual coupling effects. This is the usual situation and permits us to confine attention to the array factor $f(\theta, \phi)$ when studying multidimensional arrays. In this section, we develop the array factor for an arbitrary geometry.

The elements for an arbitrary three-dimensional array are located with position vectors from the origin to the *mn*th element:

$$r'_{mn} = x'_{mn}\hat{\mathbf{x}} + y'_{mn}\hat{\mathbf{y}} + z'_{mn}\hat{\mathbf{z}} \tag{8-106}$$

The array factor is then

$$\text{AF}(\theta, \phi) = \sum_{n=1}^{N}\sum_{m=1}^{M} I_{mn} e^{j(\beta\hat{\mathbf{r}}\cdot\mathbf{r}'_{mn}+\alpha_{mn})} \tag{8-107}$$

which, when normalized, is $f(\theta, \phi)$. This equation is general but is directly applicable to the common situation of an array on a surface. The double summation is useful in geometries that employ "rows" and "columns." The phase term α_{mn} is that portion of the excitation current phase used to scan the main beam and is shown explicitly. A common geometry for phased arrays is planar. The array factor for a planar array in the *xy*-plane, as in Fig. 8-28, follows from (8-107) as

$$\text{AF}(\theta, \phi) = \sum_{n=1}^{N}\sum_{m=1}^{M} I_{mn} e^{j\alpha_{mn}} e^{j\xi_{mn}} \tag{8-108}$$

where
$$\xi_{mn} = \beta\hat{\mathbf{r}}\cdot\hat{\mathbf{r}}'_{mn} = \beta[x'_{mn}\sin\theta\cos\phi + y'_{mn}\sin\theta\sin\phi]$$
$$\alpha_{mn} = -\beta[x'_{mn}\sin\theta_o\cos\phi_o + y'_{mn}\sin\theta_o\sin\phi_o]$$
$$\theta_o, \phi_o = \text{main beam pointing direction}$$

This formulation is a generalization of that for a linear array. Note that the *z*-axis is normal to the plane of the array, whereas in our treatment of linear arrays the *z*-axis is along the array. If all rows parallel to the *x*-axis, have the same current distribution, and if all columns have identical current distributions, then the current is separable (e.g., $I_{mn} = I_{xm}I_{yn}$) and (8-108) reduces to

$$\text{AF}(\theta, \phi) = \sum_{m=1}^{M} I_{xm}e^{j\xi_{xm}} \cdot \sum_{n=1}^{N} I_{yn}e^{j\xi_{yn}} \tag{8-109}$$

where the phase of the current for beam steering is not shown explicitly and

$$\xi_{xm} = \beta x'_m \sin\theta\cos\phi \qquad \text{and} \qquad \xi_{yn} = \beta y'_n \sin\theta\sin\phi$$

This is a product of two linear array factors associated with the row and column current distributions. The patterns in the principal planes ($\phi = 0°, 90°$ and called the cardinal planes) are those of the corresponding linear arrays (row, column). Planar arrays normally have separable current distributions, so linear array analysis can be applied directly to find the principal plane patterns.

For planes off the principal planes, called the intercardinal planes, the pattern is the product of the row and column linear array patterns if the current distribution is separable. In the 45°-plane, the side lobes will be very low because they are a product of the side lobe levels in the principal planes; Fig. 9-7 illustrates this point using a continuous planar

aperture distribution. A technique for finding the pattern of a planar array in any cut-plane is the *projection method*, or method of collapsed distributions, for finding the array factor. The element locations are projected onto the cut-plane along with the current weighting for the elements along the projection line. The projection line contains the elements in the equivalent linear array. The weightings for coincident projections are added together. The current distribution in the original array need not be separable and the grid need not be regular. For an example of the projection method, a 4×4 planar array of elements spaced 0.5λ apart and uniformly excited has an equivalent array in the $45°$-plane of seven elements spaced 0.354λ apart with the following weightings: $1:2:3:4:3:2:1$. Prob. 8.8-5 shows that pattern of the projected linear array is identical to the planar array pattern. See [H.3: Elliott, Sec. 6.10] for more examples. The method can be used for a variety of geometries, such as finding the equivalent linear array for a circular array.

Planar arrays are used to create a highly directive beam, often with scan capability as discussed in the next section. Here we present directivity calculation methods. First consider a cell of an array, which is the area surrounding an element. The maximum possible directivity for one cell based on (4-23) and using the physical area of a cell as the maximum effective aperture, A_{em} is

$$D_{\text{MaxCell}} = \frac{4\pi}{\lambda^2} A_{\text{Cell}} \qquad (8\text{-}110)$$

This formula assumes radiation into the half-space on one side of the array face. It provides an upper bound on the directivity of one element embedded in an array. For a half-wavelength spaced square-grid planar array, the cell area is $A_{\text{Cell}} = d_x d_y = (\lambda/2)(\lambda/2) = \lambda^2/4$ and (8-110) yields $D_{\text{MaxCell}} = \pi$. Finite arrays of dipoles in a $\lambda/2 \times \lambda/2$ square grid and backed by a ground plane have a maximum element gain (see Sec. 8.9.1) somewhat above or below the maximum cell area directivity from (8-110) of 3.14, depending on the number of elements and ground plane distance. [13]. The directivity of a full array is found from a similar formula using the physical area of the full array, A_p:

$$D_{\text{MaxArray}} = \frac{4\pi}{\lambda^2} A_p = \pi D_x D_y \qquad (8\text{-}111)$$

which is also given in (9-67) for aperture antennas. The second part of this formula comes from using $A_p = L_x L_y$ and introducing the directivities of uniform line sources in the x and y directions of a planar array in the xy-plane from (5-19): $D_x = 2L_x/\lambda$ and $D_y = 2L_y/\lambda$.

This formula also can be found based on (8-75) as $D_{\text{MaxArray}} = \text{Max}\{D_e D_i\} = D_{\text{MaxCell}} N = (4\pi/\lambda^2) A_{\text{Cell}} N = (4\pi/\lambda^2) A_p$ where N is the total number of elements in the array. Forman [14] gives the range of spacing values in a square grid array for which (8-111) equals the correct directivity for a few element types, but the ranges are small and (8-111) should used only as a guideline. In general, the directivity of planar arrays increases with element spacing similar to that for linear arrays until grating lobes begin to appear and directivity decreases. King and Wong [15] note this effect, and that directivity becomes asymptotic to $N D_e$, where D_e is the directivity for actual element, for spacing values around 1λ depending on the element pattern. Formulas are available in the literature [14] for calculating the directivity of planar arrays with simple element types, although they tend to be complicated and have restrictions on their application. But such formulas are useful for checking computer programs based on pattern integration to find directivity. Forman found that (8-111) is close to the actual array directivity for an element pattern of $\cos^{0.5} \theta$; the following example demonstrates this.

EXAMPLE 8-8   *Directivity of Planar Arrays*

In this example, we examine the exact and approximate methods for calculating the directivity of a square-grid (5×5) planar array in the xy-plane. The 25 elements have the pattern

$g_a(\theta) = \cos^{0.5}\theta$ (for $\theta < 90°$ and zero otherwise). We consider two approximate directivity formulas. The first, which was mentioned in association with (8-75) for linear arrays, depends only on the element type and not the spacing:

$$D_{\text{ApraxArray}} = ND_e = 25 \cdot 4 = 100 = 20\,\text{dB} \tag{8-112}$$

where the element directivity was obtained by pattern integration. The second approximation is that of (8-111):

$$D_{\text{MaxArray}} = \frac{4\pi}{\lambda^2}A_p = \frac{4\pi}{\lambda^2}(N_x d_x)(N_y d_y) = 4\pi N\frac{d_x}{\lambda}\frac{d_y}{\lambda} \tag{8-113}$$

First consider spacings of $d = d_x = d_y = 0.5\,\lambda$. The maximum directivity from (8-113) is 19 dB. Evaluating the directivity by pattern integration gives 18.3 dB for isotropic elements and 18.98 dB for square-root cosine element patterns. Increasing the spacing to $d = 0.8\,\lambda$ increases the directivity from (8-113) to 23 dB and the directivity by numerical integration to 21.6 and 22.4 dB for isotropic and square-root cosine element patterns, respectively. From these calculations we see that (8-112) is not reliable, and (8-113) and its general form (8-111) provide a good upper bound directivity prediction for the square-root cosine element pattern array.

To produce radiation in only one hemisphere, arrays are often operated with a conducting ground plane backing the elements. The same principle applies to the single element case illustrated in Fig. 3-20. For example, a linear array backed by a ground plane can be analyzed by removing the ground plane and introducing images of the elements, forming a planar array; Probs. 8.8-2 and -3 are arrays of this type. The patterns of planar arrays can be found using sequential linear array analysis if sufficient symmetry exists. One row is considered as an "element" and its array pattern is multiplied by that of a column, considered as a linear array; Prob. 8.8-4 is an example. Grids other than rectangular are used. The triangular grid can offer similar pattern performance but requiring fewer elements; Prob. 8.8-6 compares a square grid to a triangular one of the same array area.

There are many applications for non-planar arrays such as a *conformal array* that conforms to an aircraft surface, to the superstructure of a ship, or to the skin of missile. The techniques used in this section can be used to find the radiation pattern for conformal arrays.

Directivity usually must be found by numerical integration. More examples and details on analysis techniques are available in [H.8.2: Haupt, Chap. 5] and [H.8.2: Visser, Chap. 10]. An interesting array configuration is one that uses linearly polarized elements to produce circularly polarized radiation. The planar array is made of 2×2 subarrays of four linearly polarized elements with angular orientations of $0°$, $90°$, $180°$, and $270°$, and phased in the same way. This array is preferred over one using circularly polarized elements because circularly polarized elements require more complicated feed hardware, leading to more loss, weight, and cost.

The largest antennas found in practice are arrays and are used to achieve high directivity or narrow beamwidth. Arrays are preferred to a single large continuous aperture antenna, such as a reflector, because of its large physical size and corresponding weight, complexity, and cost. Radio astronomy applications have several very large arrays in use or in the planning stages made up of thousands of elements and some have element separations of many kilometers. There are also many radars employing arrays, as will be discussed in the next section.

8.9 PHASED ARRAYS AND ARRAY FEEDING TECHNIQUES

It is often required that the antenna main beam pointing direction be varied with time, which is referred to as scanning. Scanning can be accomplished mechanically or electronically. Mechanical scan is achieved by slewing the entire antenna, which can be any

type of antenna but most often is a reflector antenna or a fixed-phase array. However, mechanical scan requires a positioning system that can be large and costly, and mechanical scan is often too slow. The alternative is to use a **phased array**, which is an array whose phase (or time delay) at each element is controlled to steer the pattern in angular space. Phased arrays offer the advantage of electronic-speed, inertia-less scanning with the ability to track multiple targets (users) simultaneously. In general, the magnitude and phases of elements can be controlled for low side lobes or to shape the main beam. Phased arrays find applications in radar, sensing, and communications. In radar, targets can be tracked to obtain their angular coordinates for surveillance. In communications, the array radiation pattern can be adjusted for varying traffic conditions such as mobile users changing location. In this section, we review electronic scanning principles along with practical constraints. This is followed by the topics of feed networks and associated enabling technologies, and operational phased arrays—past, present, and future.

8.9.1   Scan Principles

The simple formulas developed in Sec. 8.3.2 for calculating the linear phase taper across an array required to steer the pattern main beam to a desired angle holds for all linear arrays and is easily extended to other array geometries. The element phases are adjusted to form a phase front that is planar and oriented to steer the main beam in a direction normal to the formed planar wavefront.

The array factor for a linear array in (8-53) can be generalized to include element phases with a linear portion $\{\alpha_n\}$ and a nonlinear portion $\{\delta_n\}$, and to include arbitrarily positioned elements:

$$\text{AF}(\theta) = \sum_{n=0}^{N-1} I_n e^{j\xi_n} = \sum_{n=0}^{N-1} A_n e^{j(\alpha_n + \delta_n)} e^{j\xi_n} \qquad (8\text{-}114)$$

The spatial phase for the nth element located on the z-axis at z_n is

$$\xi_n = \beta z_n \, \cos\theta \qquad (8\text{-}115)$$

The current for the nth element has magnitude A_n and phase $\alpha_n + \delta_n$. The phases $\{\alpha_n\}$ vary linearly with position in the array and determine the main beam maximum direction from

$$\alpha_n = -\beta z_n \cos\theta_o \qquad (8\text{-}116)$$

and are referred to as *linear phase*, or *uniform progressive phase*. For an equally spaced array, $z_n = nd$ and $\alpha_n = n\alpha$ where $\alpha = -\beta d \cos\theta_o$ as in (8-25). Nonlinear phases $\{\delta_n\}$ can be used for pattern shaping (see Chap. 10).

As the pattern of an array is scanned off broadside, the main beam widens. This effect is called *beam broadening*. We illustrate this for a linear array of five isotropic elements spaced 0.4λ apart. Fig. 8-29 shows a series of patterns for increasing off-broadside scan angles. Notice the increase in the beamwidth of the main beam with scan off broadside. The full pattern for this array is obtained by rotating the pattern about the z-axis. Two examples of three-dimensional patterns are shown in Figs. 8-29*b* and 8-29*f*. As the main beam is scanned away from broadside, the increase in beam solid angle of the main beam is just about compensated for by the reduced solid angle of the total pattern (formed by rotation of the pattern about the array axis). Thus, directivity of the array factor remains relatively constant, with scan for spacings less than a half-wavelength and for scan angles not close to endfire; see Fig. 8-18. For spacings slightly greater than a half-wavelength, a grating lobe begins to appear for scan angles near endfire and the directivity decreases; again, refer to Fig. 8-18. Since isotropic elements were assumed, these remarks apply to

(a) $\theta_o = 90°$

(b) $\theta_o = 90°$

(c) $\theta_o = 75°$

(d) $\theta_o = 30°$ (bifurcated pattern)

(e) Endfire ($\theta_o = 0°$)

(f) Endfire ($\theta_o = 0°$)
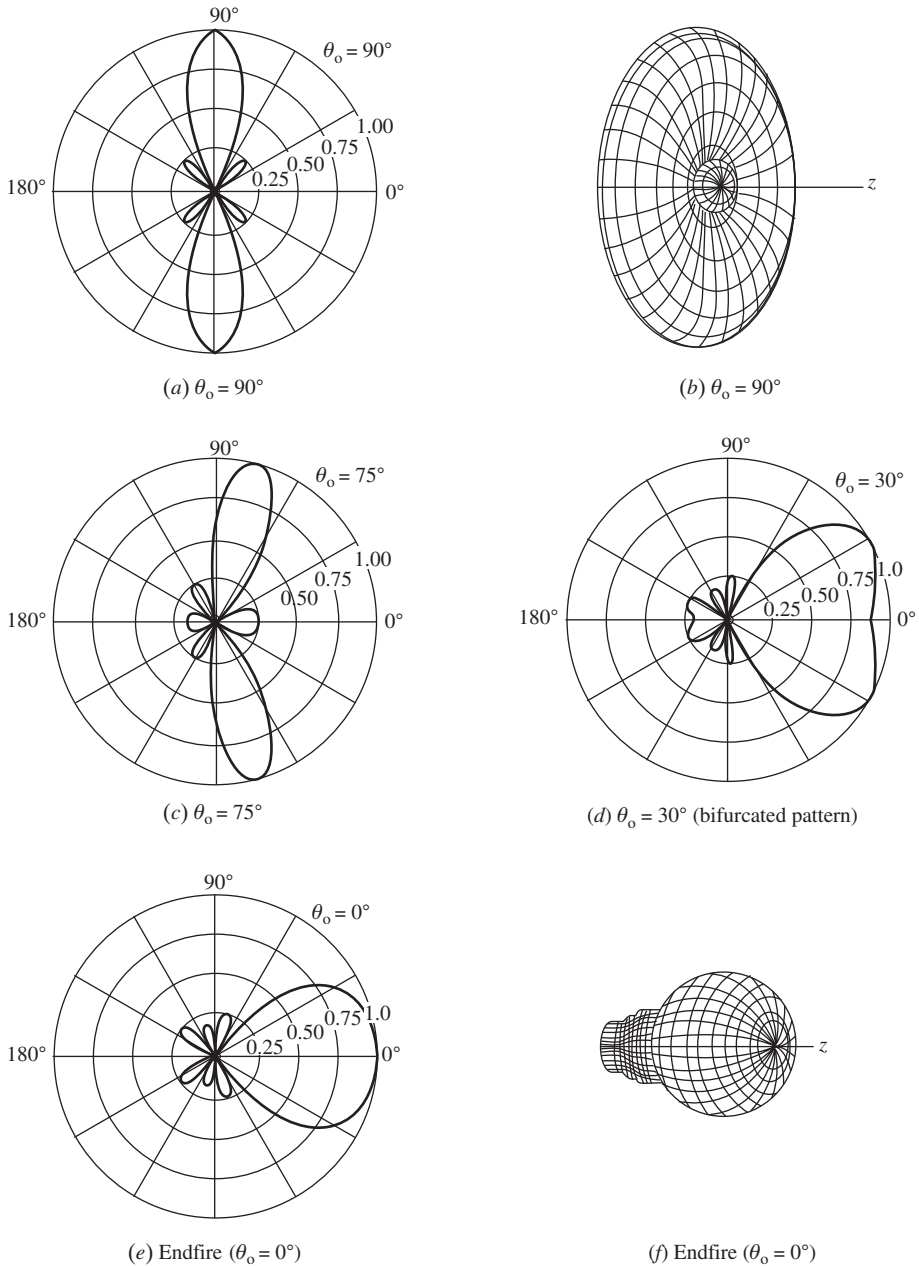
Figure 8-29 Example of phase-scanned patterns for a five-element linear array along the z-axis with isotropic elements equally spaced at $d = 0.4\lambda$ and with uniform current magnitudes for various main beam pointing angles θ_o.

array factors. When the element pattern effects are included for the case of a directive, broadside element pattern, directivity will decrease with scan angle.

As we noted in Sec. 8.2, for half-wavelength spacings there is exactly one period of the array factor in the visible region and no grating lobe will be visible, except for endfire operation that produces two endfire beams. For spacings larger than a half-wavelength, part or all of a grating lobe may be visible depending on scan angle. For one wavelength

spacing or more, there will be visible grating lobes. When spacings of several wave-lengths are used, many grating lobes are visible (see Fig. 8-3 for a simple example) and the array is called an **interferometer**. Each major lobe has a narrow beamwidth but there are many of them. Large element spacings, however, permit electrically large elements with relatively narrow beamwidth patterns that act to decrease the size of the grating lobes. In normal applications, grating lobes limit the performance of phased arrays and are to be avoided. Grating lobe peaks will be not appear in the visible region if element spacings are restricted as follows:

$$d < \frac{\lambda}{1 + |\cos \theta_o|} \quad \text{to avoid grating lobes} \tag{8-117}$$

where θ_o is the main beam pointing angle with respect to the line of the array corresponding to the largest scan angle off broadside. This relation is derived by solving (8-27) for the first grating lobe at $\psi = 2\pi$, where $\theta = 0°$. For broadside operation, $\theta_o = 90°$ and (8-117) gives $d < \lambda$. If scanning to endfire ($\theta_o = 0, 180°$) is desired, then $d < \lambda/2$. This result is based on an omnidirectional element pattern in the plane of scan. Larger spacings are permitted for directive element patterns since they diminish the effect of grating lobes.

The principles required of scanning multidimensional arrays were covered in the previous section. The expression for determining the phases of the *mn*th elements located at positions $(x'_{mn}, y'_{mn}, z'_{mn})$ needed to steer the beam peak of the array factor to the angle (θ_o, ϕ_o) is given in (8-108). For planar arrays, the orientation of the array that provides the most convenient mathematical formulation is for element centers located in the *xy*-plane. This makes the direction broadside to the array the *z*-direction, and then the angle θ is the angle off broadside. In contrast, for linear arrays we placed elements along the *z*-axis, making θ the angle relative to endfire, not broadside. In the remainder of this chapter, the *z*-axis is normal to the array, and *scan angle* refers to the angle the beam peak is scanned away from broadside.

The complete pattern of a phased array is found by multiplying the array factor by the element pattern as in (8-54). The important subtlety in pattern multiplication is that while the array factor scans with phase change the element pattern remains fixed. *As a phased array is steered, the peak of the total array pattern follows the element pattern shape*. Fig. 8-30 shows a linear array operated for a broadside beam and for scan to 30° off broadside. The array is four 0.7-λ spaced elements that have uniform current magnitudes and phases as required to steer the array factor. The element pattern (dashed curve), $\cos^2 \theta$, is unaffected by the scan angle, which has several effects on the total array pattern (solid curve). First, the peak of the main beam is reduced by the level of the element pattern at the angle of interest. In this case, the scanned beam peak is reduced to 0.77, for a 2.3-dB scan loss. Second, the beam peak direction does not go to the desired angle of 30° off broadside. Instead, the actual scan angle is a few degrees less than desired. The final effect on the scanned pattern is the altering of the whole pattern from the array factor shape. But this distortion can actually be an improvement. In the example in Fig. 8-30*b*, the side lobe at $\theta = -50°$ is a remnant of a full grating lobe the array factor has at $-68°$. So the element pattern acts to greatly reduce grating lobes in scanned patterns. These are general trends in array scanning. Mutual coupling effects can be included for scanned patterns using (8-105), where the mutual coupling effects are included in the average active-element pattern. The array grid of planar arrays controls the grating lobes. The scan performance of a rectangular-grid array behaves like the projected linear array in the scan plane.

The pattern and the element impedances of phased arrays vary with scan angle because the mutual coupling changes. The input impedance influence is evident in (8-99) because as the phases of the currents $\{I_n\}$ are changed to scan the beam the summation of mutual impedances weighted by the currents changes. The most serious effect is *scan blindness,* which is manifested by a dramatic reduction in radiated power due to high reflection
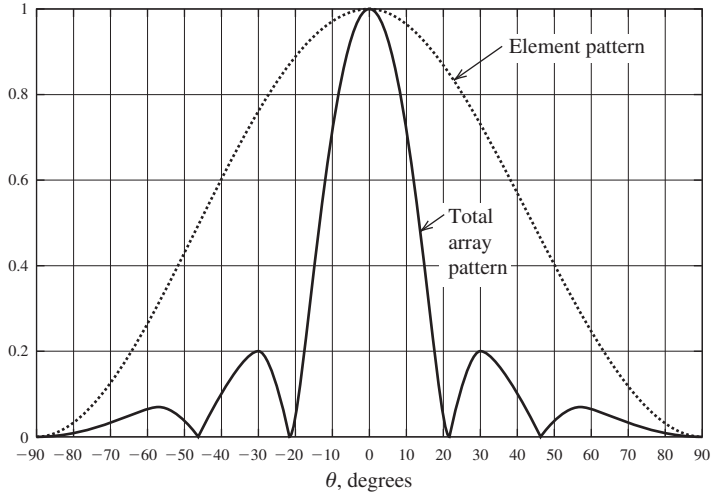
(a) Brodside case.
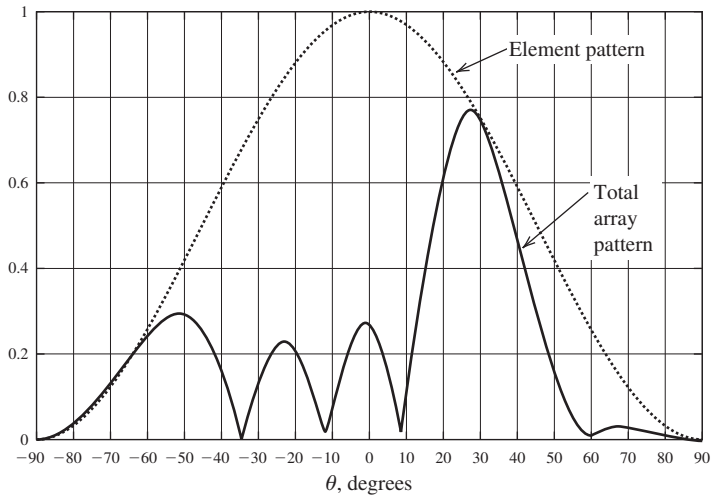
(b) Phased to scan to 30°.

Figure 8-30 The total array pattern (solid curve) for a four-element linear array that has $0.7\,\lambda$ spacings, uniform excitation magnitudes, and excitation phases for (a) broadside operation and (b) scanned 30° off broadside. The element pattern (dashed curve) is $\cos^2\theta$. Note that the element pattern remains stationary during scan.

when the array is scanned to a blindness angle. In the transmitting case, when the feed network is configured to steer the beam to a blind scan angle there is power reflected back into the transmitter, which can cause damage. If $\Gamma_m(\theta_o, \phi_o)$ is the *active reflection coefficient* of the mth element in a fully excited phased array, the power delivered to that element is

$$P_m(\theta_o, \phi_o) = P_{inc}\left[1 - |\Gamma_m(\theta_o, \phi_o)|^2\right] \tag{8-118}$$

where (4-41) was used and P_{inc} is the incident power. From this we can find the gain of the mth element operating in presence of the other elements [16], called an *element-gain pattern*:

$$G_{\text{EG}}^{m}(\theta_o, \phi_o) = G_i(\theta_o, \phi_o)\left[1 - |\Gamma_m(\theta_o, \phi_o)|^2\right] \tag{8-119}$$

where $G_i(\theta_o, \phi_o)$ is the gain pattern of an isolated element, that is, $G_i \ |F(\theta, \phi)|^2$. It is a realized gain because impedance mismatch effects due to mutual coupling are included, although feed network impedance matches are not. It is proportional to the active element pattern, so (8-119) is essentially an active element-gain pattern. In an infinite array, all element patterns are identical. In a large array with small edge effects (to be discussed shortly), the element-gain patterns of all elements except those near the edge are nearly identical and an *average element-gain pattern* (developed similarly to the pattern expression of (8-105)) can be used in (8-119) to give:

$$G_{\text{EG}}(\theta_o, \phi_o) = G_i(\theta_o, \phi_o)\left[1 - |\Gamma(\theta_o, \phi_o)|^2\right] \tag{8-120}$$

where $\Gamma(\theta_o, \phi_o)$ is the reflection coefficient of a central element in the array. Usually an array is nearly matched at broadside, so $\Gamma(0,0) = 0$, and impedance mismatch increases with scan angle.

In an infinite array, a complete mismatch occurs at the blindness angle, if it exists, when $|\Gamma(\theta_o, \phi_o)| = 1$. Then (8-120) shows that the average element-gain pattern goes to zero. In finite arrays, the blindness effect does not yield complete mismatch, but there can still be significant mismatch in large arrays. Scan blindness is associated with surface wave–like phenomena and usually occurs for spacing values less than a wavelength and at a scan angle less than that for which grating lobes occur in (8-117), and thus is often the factor limiting the scan range of a phased array. Scan blindness varies with the scan plane. In planar arrays of dipoles mismatches are larger for scan in the *E*-plane than in the *H*-plane. This is illustrated in Fig. 14-28, which plots the input impedance of a central element in a planar array of dipoles as a function of scan angle. The real part of the input impedance approaches zero in the *E*-plane, but not in the *H*-plane. In the *E*-plane, the dipole creates an electric field component perpendicular to the ground plane that can set up a surface wave, whereas in the *H*-plane the electric field from the dipole traveling along the array face is parallel to the ground plane and is shorted out. For arrays printed on a dielectric substrate, the blindness angle moves toward broadside as dielectric thickness is increased. [13] If the feed network resides on a common dielectric substrate, an additional blindness effect, called "feed blindness," can occur which has behavior opposite that of surface-wave blindness. The feed blindness scan angle moves toward endfire rather than broadside as dielectric loading increases. [17] The best way to avoid scan blindness effects is to use small element spacing, typically under a half-wavelength. Blindness effects can also be reduced by using subarrays, which are clusters of neighboring elements fed with the same excitation. [18] But subarraying introduces additional grating lobes.

Reducing the mutual coupling will help reduce blindness effects. Adding conducting fences between elements helps, as mentioned in Sec. 8.7.2. Placing elements in cavities has been used to reduce surface waves and improve blindness. [19] For more details on scan blindness and its mitigation see [H.6: Hansen, Vol. II]. Pozar [13] recommends using array designs that limit the maximum scan angle to about $10°$ less than the scan blindness angle.

Techniques are available for analyzing infinite arrays that take advantage of the unending periodicity. [H.6: Hansen, Vol. II, Chap. 3; Haupt: H.8.2, Sec. 6.5] An infinite array with a regular grid presents the same operating environment to each element. So the active element patterns, reflection coefficient, and input impedance for each element are identical. The pattern of the full infinite array is in theory an impulse function (giving infinite directivity) which traces out the element pattern as scan angle is varied. Infinite array analysis is useful in determining element performance of the central elements of a finite array of the same type. As with linear arrays discussed in Sec. 8.7.2, the patterns of elements near the edge of an array differ from that for centrally located elements.

As array size increases, the array pattern becomes more accurately predicted from the pattern multiplication formula of (8-105) using the average active-element pattern and array factor. The array size for which this formula is acceptable depends on the element type, array geometry, and other factors. An experimental study with microstrip elements found that a 7×7 array is sufficiently large. [20] For a planar array of half-wave spaced printed dipoles, an array size of 19×19 was found to be large enough for the active reflection coefficient to behave similar to being in an infinite array. [13] This is consistent with the general guideline of 10λ per side. [H.8.2: Mailloux, 2nd ed., p. 71].

Array size is usually driven by the requirements on resolution (width of the main beam) or gain (directivity reduced by efficiencies), which is also affected by element spacing. It is desirable to have wide element spacing (approaching a wavelength), which reduces the number of elements and the associated costs for elements and the feed network. However, the maximum scan angle limits how far apart elements can be in order to avoid grating lobes as given by (8-117) and to avoid scan blindness.

8.9.2   Feed Networks and Array Technology

As mentioned in Sec. 8.1, phased arrays offer the unique advantages of electronic-speed beam scanning without mechanical motion. Additional unique features are the capability of tracking multiple targets (or users) that are in different directions, dynamic pattern control, and graceful degradation as elements fail. In addition, beam nulls can be formed to cancel jamming signals (or interference). Design for low radar cross section (see Sec. 4.6) is possible by redirecting an enemy jamming signal away from the direction back toward the threat. The many attractive features of arrays come with the penalty that each element of an array must have a transmission path to the receiver (or transmitter). In addition, a phased array requires hardware and software to control the phase of the elements. The usual choice that is faced when deciding on a high-gain antenna is between a reflector with a simple feed system but with a large steering mechanism and a phased array with low-weight electronic scanning but a large feed network.

The possible array feed network architectures are shown in Fig. 8-31. The parallel and series feeds follow directly from circuit concepts. The *parallel feed* shown in Fig. 8-31a is the network form that we have assumed thus far and is the most common. It is also called a corporate feed that looks like the organizational diagram of a corporation. It has the desirable characteristic of equal line length to each element, making the network symmetric that helps with balancing mutual coupling effects. Most important, the signal phase and amplitude to each element is naturally identical, as often used in practice. The equal line lengths also make the network frequency independent, and thus wide bandwidth. The *series feed* of Fig. 8-31b offers advantages of simple construction a compact structure. A waveguide slot array is a popular series-fed array; example slots are shown in Fig. 8-35b. A disadvantage of the series feed is that the common feed line attenuates the signal to successive elements, causing loss and complicating the design. Series-fed arrays offer the possibility of *frequency scanning* by changing frequency which changes the electrical line length between elements, and thus the phase. One realization of a series feed is the "serpentine" waveguide feed which is a waveguide bent into a serpentine similar to the shape in Fig. 8-31b with radiating elements, usually dipoles, with probes tapped into the waveguide. Phase shifters can be inserted into the connecting series transmission lines or in the individual element feed lines. A series feed introduces dispersion that distorts short pulses and limits bandwidth.

The preceding feeds are sometimes referred to as "constrained" feeds because of the physical constraint of continuous connectivity from the feed to the elements. In contrast, the *space feed* uses a primary antenna illuminating a separate unit (sometimes called an "active lens") consisting of pickup elements connected to a secondary radiating array face, as illustrated in Fig. 8-31c for the transmitting case. Typically only used for large phased arrays, the space feed array takes advantage of the free space environment

(*a*) Parallel, or corporate, feed.

(*b*) Series feed.

(*c*) Space feed.

(*d*) Parallel-series feed.

Primary
transmitting
antenna

Pickup
antenna

Secondary
transmitting
antenna
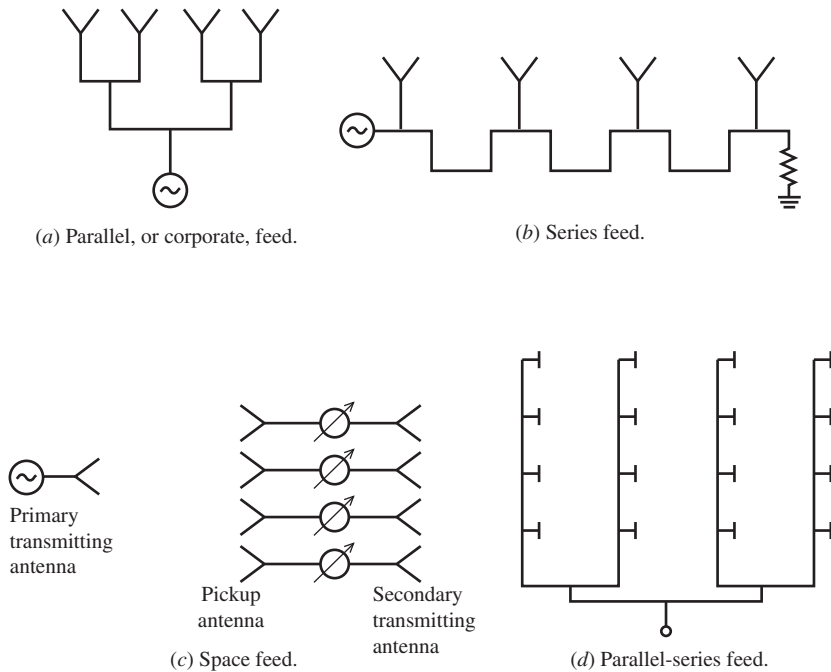
Figure 8-31 Types of array feed networks.

between the primary feed and the pickup array, leading to size, weight, and cost reduction compared to a fully connected feed network. The amplitude distribution of the array is determined by the primary feed pattern and the pickup array element placements. The phase introduced by the different path lengths from the primary feed to each pickup element is compensated for by the phase shifters on the radiating elements, which also are used for beam scanning. Just as with a reflector antenna, there is spillover loss associated with the primary feed power missing the pickup array. It is common to taper the amplitude distribution across the secondary radiating array to reduce side lobes by "thinning" the array by not connecting some elements. The elements are not removed but are dummy loaded to present a regular impedance environment to the active elements. This is called *density tapering*. The primary feed is often configured with a central feed antenna, usually a horn, surrounded by two pairs of horns, one pair in each plane, for producing difference patterns with a null on axis for direction finding in the vertical and horizontal planes. A related feed method is the *reflectarray*, which has a primary feed illuminating an array which receives, processes, and transmits a formed beam back in the direction of the feed, appearing to reflect the signal. [H.6: *Ant. Eng. Hdbk.*, 4th ed., Chap. 35] Scanning is possible by including phase shift functionality into the array.

The three feed architectures were illustrated as linear arrays, but the concept is easily extended to multidimensional arrays. The corporate feed network will become very large for arrays with many elements. The space feed is an alternative, but the *hybrid feed* of Fig. 8-31*d* is more commonly used. It is shown as a parallel-series hybrid with the rows fed in parallel and the columns series fed. By inserting phase shifters in the legs of the parallel feed the pattern can be steered in the azimuth plane (left-right as shown). The parallel-parallel hybrid feed facilitates the use of subarrays that have the same amplitude and/or phase in each element of the subarray.

Physical construction of the feed network and its associated components behind the array face is in one of two forms, brick or tile. [H.8.2: Mailloux, 2nd ed., p. 46] In *brick* construction, the complete feed hardware modules for one element (or a few elements),

often including the element itself and all in a monolithic or hybrid integrated circuit, is placed perpendicular to the array face. A brick can be as small as a single module connected to a single element. *Tile* construction consists of several parallel layers with each layer containing the same components performing the same function. For example, there could be separate layers for the elements, phase shifters, and amplifiers. Tiles are more compact than bricks but may have thermal management issues.

Beam scanning is implemented using one of the following methods. The term "phase shifter" often includes any of these means for changing the phase. For more details see [H.8.2: Mailloux, 2nd ed., Sec. 1.3], [H.6: Balanis, Ed., Chap. 25].

> *Phase shift* — The use of phase shifter devices to adjust element phases for beam scanning is the most popular method. Phase can be changed by altering the permeability using a ferrite phase shifter or by changing permittivity (dielectric constant) such as with a ferroelectric phase shifter. Phase shifters are narrow band.
>
> *Time delay* — Phase shift in time-delay devices is accomplished by adjusting line length, usually by switching in various sections of transmission line.
>
> *Hybrid* (usually phase-time delay) — Time-delay lines are generally larger than conventional phase shifters, so in applications where time delay is required, time delay and phase shift are combined. This will reduce the spreading of a pulse when the beam is scanned compared to a phase-only feed. [21].
>
> *Frequency scanning* — Phase is changed by changing frequency, which changes the electrical length of the interconnecting lines and is used with series feeds. Frequency scanned arrays are found in practice, mostly in radars rather than communications where frequency is usually required to be fixed. An example of a series-fed array is a waveguide with slots milled in one wall that act as radiating elements.
>
> *Beam switching* — Beam scanning could be accomplished by switching between separate antennas pointed in different directions, but the system would be large and complex. Instead, a *beamforming network* (*BFN*) is used to form multiple beams. A BFN that forms N beams has N input ports, one for each beam. By adding a $1 \times N$ switch to the input, beams can be selected by controlling the switch.
>
> *Digital beamforming* — Beamforming can be achieved in the digital domain by sampling the RF signal at the element level and routing to a digital processor unit for complex weighting (i.e., amplitude and phase control) and summing to form the beam. Multiple simultaneous beams and adaptive features such as interference rejection are possible as well.

Most phase-shifter technologies can be realized in digital as well as analog (allowing adjustment to any phase value) form. An M-bit digital phase shifter provides phase increments in multiples of $360/2^M$ degrees. For example, a 3-bit phase shifter has phase states of $0°$, $45°$, $90°$, $135°$, $180°$, $225°$, $270°$, and $315°$. These settings are the cumulative phases of the three bits in series in all possible on/off states where the "off" states are $0°$ and the "on" states are $45°$, $90°$, and $180°$. Arrays in practice typically use from 3 to 7 bits of phase shift. Quantization effects result from using the nearest digital phase state rather than the exact phase, with the main issue being an increase in side lobes. [H.6: Rudge, Vol. 2, Sec. 9.2.4] Similar high-side lobe problems develop when using subarrays with the same phase supplied to all elements in the subarray. [H.8.2: Mailloux, 2nd ed., Sec. 7.3] Also, digital phase shifters will place beams at discrete locations rather than directly toward a user as is possible with analog phase steering. Hybrid use of both digital and analog phase shifters may provide an optimum feed network. [H.6: Balanis, Ed., Sec. 25.5.3]

Delay line phase shifting can be realized by electromechanical adjustment of a "line stretcher," variable length transmission line. However, most time delay phase shifting is done by switching in (or bypassing) fixed-length sections of line. The phase shift in time

delay devices is proportional to frequency just as in a transmission line, leading to wide bandwidth. The frequency proportionality follows from (8-116), where the phase shift for the nth element of a linear array on the z-axis is $\alpha_n = -(2\pi/c)fnd\cos\theta_o$. A delay line with this phase will give a beam peak in the direction θ_o for all frequencies. Time delay is preferred in pulse radar to preserve the pulse shape as the scan angle is changed.

So far we have modeled arrays using unspecified phase shifters and will now discuss how they are implemented. [H.8.2: Mailloux, 2nd ed., Sec. 1.3] Although intrinsically analog, ferrite phase shifters are usually configured as digital phase shifters. Ferrite phase shifters operate by changing the permeability of ferrite material that the wave propagates through. They are often constructed in waveguide and can handle high power levels needed in phased array radars. Due to their large size and weight, they are not often used in space applications and large airborne systems. Diode phase shifters can also be analog or digital. Analog diode phase shifters are realized with Schottky barrier or varactor diodes. Digital diode phase shifters operate by using diodes (PIN, Gallium-Arsenide-based devices, and others) to switch delay lines. Also, MEMS (micro-electromechanical systems) switches are used to switch in fixed sections of line length to give the proper number of bits of phase. In addition to the foregoing through-line devices, phase shifting is also realized in devices configured for reflection from a shorted load. Ferroelectric phase shifters operate by changing the material permittivity via an applied DC voltage. When fully developed they will find wide acceptance because of their desirable characteristics of low weight and low power draw due to being voltage actuated, unlike MEMS-based units. Photonic technology is also employed in array feed networks. [H.8.2: Hansen, 2nd ed., Sec. 6.4] Using open optical beams or fiber optic cables to distribute signals reduces the size and weight of the network compared to conventional metallic feeds. Immunity to RF interference is an added advantage. However, the required converters between optical and RF frequencies are lossy and non-reciprocal.

Beamforming networks are constructed using conventional network techniques or using quasi-optical lenses. [H.8.2: Hansen, 2nd ed., Chap. 10] Conventional constrained networks can be as simple as a series of power dividers that include phase shifters if scanning is required. For example, in Fig. 8-31a two-way power dividers are located at the tee junctions, or one four-way power divider could be used. A power divider based network, however, becomes excessively large for more than a moderate number of elements in the array. One of the most popular conventional-network BFNs is the *Butler matrix*, which was introduced in 1961. [22] Compact Butler matrix units are available commercially and find application in smart antennas to be discussed in Sec. 12.4. The Butler matrix essentially performs a fast Fourier transform, radiating a set of beams that are orthogonal so that the signals into one port only appear in the beam associated with that port. A Butler matrix connects a 2^n element array to n input ports that are inputs to the beams. Fig. 8-32 shows a functional diagram of a four-beam Butler matrix. Operation is illustrated for Beam 2 using the highlighted portion of Fig. 8-32. The signal passes through two quadrature hybrid power dividers, denoted "H," each with two input ports on the bottom and two output ports on the top that have equal power outputs that are in-phase (90° out of phase) from the same-side (opposite-side) input ports. Fixed phase shifters of $-45°$ are also used. The resulting outputs of the matrix have equal amplitude and a progressive 45° interelement phase shift. The patterns are shown for half-wavelength element spacing and a $\sin\theta$ element pattern (array along the z-axis). In the isotropic element case, the direction of one beam peak is in the null direction for the other beams, which is the orthogonality property. This is approximately true for the array of Fig. 8-32. The beam peaks are 14.5° and 48.6° off broadside for the isotropic element case and 13.5° and 41° for the $\sin\theta$ element pattern. The beam crossover points at angles 0° and 30° off broadside are 3.7 dB below the beam peaks, which is the same for the $\sin\theta$ element case, except the crossover level is lower at 30° off broadside. Higher-order Butler matrix BFNs behave similarly: An 8×8 matrix for eight beams with a
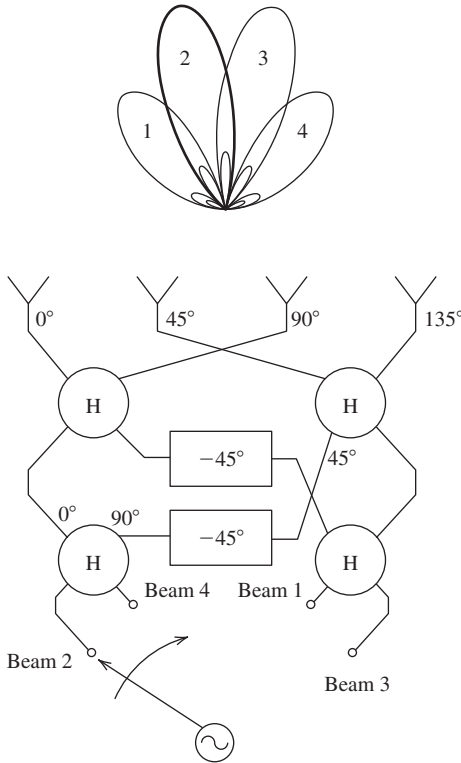
Figure 8-32 Functional diagram of a four-beam Butler matrix beamformer. "H" represents a quadrature hybrid power divider; the input ports are on the bottom, and the two output ports on the top have equal power outputs that are in-phase (90° out of phase) with the same-side (opposite-side) input port. The patterns shown are for half-wave spaced elements with $\sin\theta$ patterns (array on the z-axis). Beam 2 is active and its pattern is shown as bold.

crossover level of -3.9 dB. [H.8.3: Milligan, 2nd ed., p. 595] A BFN can be used in a *multiple beam array*, taking advantage of the orthogonal beam property where simultaneous beams can be processed separately for multiple user (target) tracking or beams can be combined and weighted for special beam shaping.

A second BFN is the *Blass matrix*, which uses transmission lines of the proper lengths to produce time-delay phasing and directional couplers to produce multiple beams. Although not as compact as the Butler matrix, it is inherently wideband. [H.8.2: Mailloux, 2nd ed., Sec. 8.1.3] Quasi-optical beamformers find use in applications calling for wide bandwidth, where the Butler matrix falls short; however, they are generally bulky and heavy. Here we mention two prominent examples, and a more detailed treatment is found in [H.8.2: Hansen, 2nd ed., Sec. 10.2.2]. The *Luneberg lens* is a spherical dielectric with relative permittivity that decreases from 2 at the center to unity at the surface. This yields the unique property that an incident signal (implemented with a small horn antenna) on the surface produces an output beam directly opposite. Multiple feeds produce multiple radiated beams. Functioning in similar fashion to the Luneberg lens is the *Rotman lens*, constructed with parallel plate waveguide, stripline, or microstrip media. Finally, we mention *parasitic beamforming*, in which only a few elements in an array are active while others are passive but have controlled loads. [H.9.1: Thiel and Smith] For example, an active monopole with parasitic monopoles on each side can be beam-steered 180° by switching between two parasitic monopole states of first having the left one shorted and the right one open loaded, and second reversing these load conditions.

The most flexible array is a *fully distributed array* (also called an *active aperture*) in which all RF functions (phase shift, up conversion, amplification in the transmit chain, and low noise amplifier and down conversion in the receive chain) are present for each element. This architecture is used in *digital beamforming*, which is the application of

digital technology to array antenna technology.[4] A digital beamforming array consists of array elements followed by transceivers that down convert and digitize signals that then stream into a digital signal processor unit. This unit supplies amplitude and phase weights to each element, which permits beam shaping, beam steering, multiple beamforming, and adaptive beamforming (which is used in smart antennas, discussed in Sec. 12.4). These features are possible because the digital signal processor unit has all array element output information available, in contrast to a conventional analog beamformer that outputs a weighted sum of the element signals.

Feed network construction usually employs the same technology as the element uses. For example, microstrip elements are fed by microstrip feed lines, and horn or open-ended waveguide elements are connected to waveguide phase shifters. For small arrays, coaxial cables are often sufficient for the feed. Modern phased arrays make use of transmit-receive (T/R) modules that provide full duplex (simultaneous transmit and receive functions). The critical components in T/R modules are a low-noise amplifier on receive and a power amplifier in the transmit path, both usually constructed in MMIC (monolithic microwave integrated circuits). [H.8.2: Fourikis, Sec. 4.6] In digital beam-formers, analog-to-digital converters are also included.

8.9.3   Operational Array Examples and the Future of Arrays

Radars made early use of arrays. The SCR-270 radar was a fixed-phase array in an 8×4 element configuration (horizontal $\times$ vertical) that was mechanically rotated in azimuth and operated at about 110 MHz. One of SCR-270 radars detected the planes approaching Hawaii in December 1941, but there was no military response, resulting in the attack on Pearl Harbor. Prompted by military preparedness concerns, phased array development programs started in the 1930s, resulting in the first phased array in 1937. Since that time there have been many technology developments that expanded capability, improved performance, and lowered cost, including RF electronics (transmit/receive active devices), feed network components (power dividers, phase shifters, etc.), signal processing units, and antenna elements (discussed in the next section). Further history of phased arrays can be found in [H.8.2: Visser, Chap. 2].

Now we consider some examples of operational arrays; more examples are found in [H.8.2: Fourikis, Chap. 1]. Arrays are used for land, sea, air, and space applications. Here we present two arrays in some detail. The first is a small flat-plate array, the AN PPS-18 radar, and the second is the large airborne AWACS phased array. The PPS-18 is mounted with the array face vertical and rotated mechanically to locate moving personnel by Doppler radar techniques. Azimuth angle is read off the mechanical positioner, and the range in steps of 25 m is determined out to 3 km from radar return delay time. The antenna consists of 256 bow-tie elements built up from 2×2 subarrays fed in parallel. Fig. 8-33 shows the top of the substrate along the other halves of two elements on the bottom of the substrate. The operating band is 9 to 9.5 GHz ($f_c = 9.25$ GHz and $\lambda_c = 3.24$ cm) so the element spacing of 2 cm in both dimensions is $d = 0.61 \lambda_c$. The array area is $A_p = 32 \cdot 32 = 1,024$ cm^2. Using the stated beamwidth in both principal planes of 6° in (4-57) gives $G_{app} = 26,000/36 = 28.6$ dB, which compares to the specified gain of 27 dB. From $G = \varepsilon_{ap}(4\pi/\lambda^2)A_p$ in (9-78), the aperture efficiency is found to be $\varepsilon_{ap} = 41\%$.

The Airborne Warning and Control System (AWACS) is an impressive example of an airborne phased array that has been in operation since 1977. The antenna is inside a rotating radome mounted on top of a U.S. Air Force E-3 (Boeing 707) airplane, shown in Fig. 8-34a. The array is the waveguide slot array with over 4,000 slots shown in

---

[4] Sometimes feed networks that use digital phase shifters are also referred to as digital beamformers.
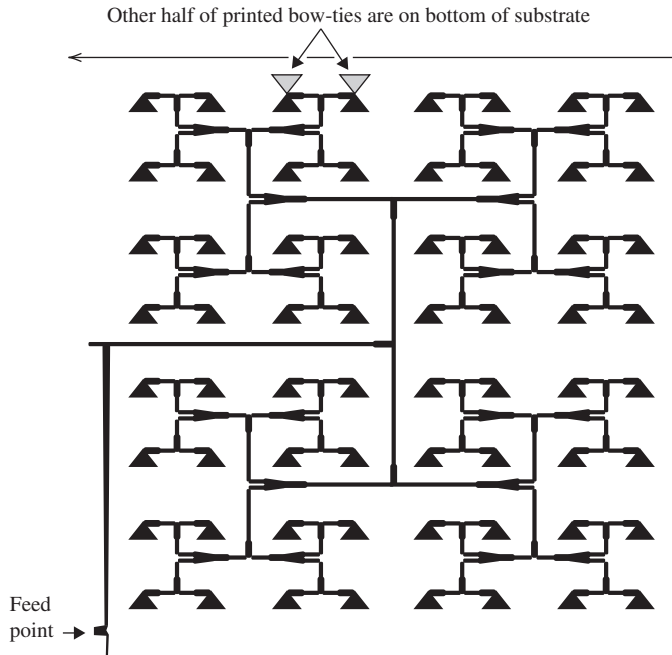
Figure 8-33 The upper right quadrant of the AN/PPS-18 Doppler radar flat-plate array. The full array has 256 bow-tie elements. The other half of the bow-tie elements (shown as dotted for two) are on the bottom of the substrate.

Fig. 8-34*b*; slot arrays are discussed in the next section. It is a hybrid parallel-series array of the type in Fig. 8-31*d*. The waveguides with series-fed slots go out in both horizontal directions from the center feed. In the vertical direction, the 28 waveguides are fed in parallel, each with a phase shifter for scanning in elevation. The entire array rotates 10 revolutions per minute for azimuth coverage. Side lobes below -40 dB are needed to distinguish low-flying aircraft targets from clutter as far as 400 km away.

In the field of antennas, the future is probably most promising for arrays. Array technology is evolving toward integration of transmit/receive modules and associated controllers, which, when affordable, enable many applications. Traditionally an antenna has been treated as a stand-alone device, but antennas are evolving into subsystems. This is especially true for digital arrays where, in fact, it is hard to even identify where the antenna is because many of its functions are embedded in software and digital processing. Arrays offer the opportunity for combining several separate functions in a single antenna using a *multifunctional antenna*. For example, the many antennas protruding from a vehicle will be replaced by one multifunctional array conforming to the surface of the vehicle. A complete multifunctional antenna performs tasks of communications over several bands and waveforms, as well as radar, sensing, and countermeasure tasks. Antennas of the future will be fully distributed digital beamforming arrays that are wideband covering more than 10 : 1 bandwidth and will support multiple polarizations. To make this possible, the enabling technologies must continue to develop by lowering the size and cost of components. There is a need for fabricating feed hardware, including several radiating elements, into a single transceiver-antenna element unit affordably. Also needed are commercial products of ferroelectric phase shifters and of MEMS switches for reconfigurable arrays. Photonic feed networks may find wider use also. As arrays develop in this fashion, reflector antennas and complex analog arrays will give way to digital beamforming arrays.

(*a*) The Boeing 707 showing the radome containing the antenna.
    (Courtesy of The Boeing Company. Used with permission.)

(*b*) The waveguide slot phased array antenna.
    (Courtesy of Northrop Grumman Corporation. Used with permission.)

Figure 8-34 The AWACS surveillance system.

8.10   ELEMENTS FOR ARRAYS

Array elements are selected based on the application for the array. An array can be a *fixed-phase array*, also called a *flat-plate array*, or a phased array. Fixed-phase arrays do not have a means to adjust the phase to the elements and thus have a pattern that is fixed in space. They are often used as replacements for reflector antennas and offer the advantage of a thinner, more appealing appearance. Flat-plate arrays are used in many parts of the world for reception of television programming via satellite or terrestrial signals. Some military radars use flat-plate arrays that are mechanically steered in the azimuth plane (i.e., parallel to the ground), such as in Fig. 8-33. Fixed-phase arrays also permit the use of narrow beam elements of moderate gain such as helix, Yagi-Uda, and horn antennas. These elements, of course, are not appropriate for use in a flat-plate array because they are long, making the array a deep structure. Wide element spacing is possible because grating lobe effects are minimized by the broadside-only operation and the narrow element pattern. Widely spaced element arrays allow the use of fewer elements for lower weight and cost. They are often used in mechanical-tracking satellite Earth terminals. As material costs decrease, flat-plate arrays will find more applications.

Phased array elements must support the desired angular range of electronic scan. Most importantly the element pattern needs to be wide enough to avoid significant gain loss with scan angle; the scan loss effect is illustrated by example in Fig. 8-30. Secondly, the element spacing should be small enough to avoid grating lobes as guided by the limit in (8-117). Third, impedance match will vary with scan angle, and the array element and array geometry should be selected to avoid scan blindness, which was discussed in Sec. 8.9.1. These considerations generally lead to the guideline that increasing scan angle requires closer element spacing. This, in turn, requires the elements to be smaller in size in order to fit into a decreasing cell area for the grid. For wide bandwidth applications, to be discussed in the next section, we will see that element spacing, and thus element size, must be reduced as bandwidth is increased. Therefore, there is considerable emphasis on using a compact element, at least in the dimensions of the array grid (x- and y-directions for the planar array of Fig. 8-28). In the remaining dimension (z-direction for our planar array example), the element can be large if the application permits. The depth of an element may not be important for fixed antennas on land and sea platforms, where size and weight are less important than for air and space vehicles. The requirement for a low profile (i.e., shallow depth) will further limit the choices of element type. Before discussing specific element types, we review general goals for element performance; more details on phased array specifications and evaluation are found in [H.6: Lo and Lee, Eds., Chap. 18].

For scanning in any plane, an element pattern that is rotationally symmetric (i.e., balanced) is desired. This prompts the concept of an ideal element pattern. The directivity expression in (8-110) indicates that one cell of a planar array in the xy-plane has directivity proportional to its area, A_{Cell}. Directivity as a function of angle varies as the projection of the cell area $\cos\theta$ viewed from angle θ off broadside as reasoned by Hannan [8]. This is similar to the element pattern for a linear current segment along the z-axis, which is $\sin\theta$ from (2-116) and is the projection factor for the length of the segment viewed from angle θ. The corresponding field pattern for the cell area is $\cos^{1/2}\theta$, which is called the *ideal element pattern* as proposed by Wheeler [23]. The average active element-gain pattern including impedance mismatch created by mutual coupling is given by (8-120) as $G_i(\theta,\phi)[1 - |\Gamma(\theta,\phi)|^2]$, which is the gain variation with angle of an element with all other elements terminated in matched loads. The angular variation of the isolated element-gain pattern (normalized pattern) is approximated as $\cos\theta$, which is the square of the ideal element pattern to make it a power pattern. The maximum directivity of a cell in the array in (8-110) extends to gain as a function of angle by including the angular variation and impedance mismatch just mentioned, giving the *ideal element-gain pattern*:

$$G_{\text{EG, ideal}}(\theta,\phi) = \frac{4\pi}{\lambda^2} A_{\text{Cell}}\left[1 - |\Gamma(\theta,\phi)|^2\right]\cos\theta \qquad (8\text{-}121)$$

If the array is infinitely large, the element patterns will be identical and often well approximated by (8-121). In finite arrays, of course, element patterns vary with location in the array. But for sufficiently large finite arrays, the full array gain is often approximated by multiplying by the number of elements, as discussed in Sec. 8.8. An approximation for the gain of the full array as a function of scan angle with all elements excited adding in-phase in direction (θ,ϕ) is to assume that the equal power supplied to all N elements adds, giving $G_{\text{Array}}(\theta,\phi) = N\, G_{\text{EG, ideal}}(\theta,\phi)$. This does not include array pattern effects, just gain variation with scan angle. It follows from the discussion in Sec. 8.8 for directivity approximations and is discussed further in [H.6: Hansen, Vol. II, p. 303].

A perfect element would be impedance matched for all angles $(1 - |\Gamma(\theta,\phi)|^2 = 1)$. The peak gain at $\theta = 0$ in (8-121) of a planar array of half-spaced elements $(d = d_x = d_y = \lambda/2)$ would be $G_{\text{MaxEle}} = 4\pi(\lambda/2)^2/\lambda^2 = \pi = 5\,\text{dB}$. Thus, a perfect element has a

Table 8-2 Elements for Array Antennas

Element	Examples	Advantages; Disadvantages
Dipole-based elements, backed by a ground plane		
Dipole	Fig. 3-20	Wide beam, simple structure;
	Fig. 8-35*a*	narrow band
Bow-tie	Fig. 7-32	Wide beam, simple structure, broadband
Microstrip elements	Sec. 11.2	Wide beam, low-profile; narrow band in basic form
Slot elements	Fig. 8-35*b*	High-power applications;
	Fig. 8-34*b*	many are narrow band
Aperture elements		
Open-ended waveguide	Fig. 9-9	Compact, moderate bandwidth; bulky structure
Horn antenna	Fig. 9-11	Moderate bandwidth, fewer array elements; limited scan due to wide spacing
Tapered slot antennas	Fig. 8-35*c* and *d*	Wide beam, broadband; not low-profile
Low-profile, broadband elements	Fig. 8-38 Sec. 8-11	Broadband, low-profile

unidirectional $\cos^{1/2}\theta$ pattern and 5 dB gain.[5] As explained in Sec. 8.9.1, a wide beam pattern such as $\cos^{1/2}\theta$ is appropriate for wide-scanning phased arrays. Scanning such a pattern 60° off broadside gives a gain loss of ($\cos^{1/2}45° = 0.841$) of only 1.5 dB. When designing large arrays that are capable of wide scan angles off broadside, it is recommended to select elements with wide beams and to perform tests with a small array before proceeding to full-scale hardware. [H.8.2: Milligan, 2nd ed., p. 128] In general, the desired element pattern depends on the array application. For fixed-beam arrays, a narrow-beam element pattern together with wider element spacing will lead to fewer elements in the array and lower cost.

Table 8-2 provides a list of commonly used array elements, along with references to places in the book where example realizations of the element are found. All of the major element categories have or will be discussed. The same is true for the subcategories, except for a few. Also noted in the table for each element are the features that are advantages or limitations. All the elements are suitable for fixed-phase arrays, but some have limited use in phased arrays, as noted in the table. In the remainder of this section, we comment on the listed elements. Elements for unidirectional beam arrays rather than bidirectional beams are assumed.

Dipoles and their variations to improve bandwidth are very popular array elements because of their wide beamwidth and easily matched input impedance, as well as their simple structure. They are used both for fixed-phase, broadside beam operation and for phased arrays capable of wide-angle scan off broadside. Dipoles, including drooping dipoles, are usually backed by a ground plane to produce a unidirectional pattern. Properties of dipoles in free space are given in Table 3-2 and in Sec. 6.1, where it was noted that about 16% bandwidth can be obtained with a wire dipole. A half-wave dipole backed by a ground plane a quarter-wavelength away produces a unidirectional as shown in Fig. 3-20. An integrated balun and mechanical support attached to the ground plane are shown in Fig. 6-40*b*. A printed form of the dipole with balun is shown in Fig. 8-35*a*.

---

[5] Note that the directivity of the $\cos^{1/2}\theta$ pattern from integration gives 6 dB because the classic formula of (8-121) is approximate.

(*a*) Printed dipole antenna.

(*b*) Waveguide slot elements.

(*c*) Linearly tapered slot antenna.

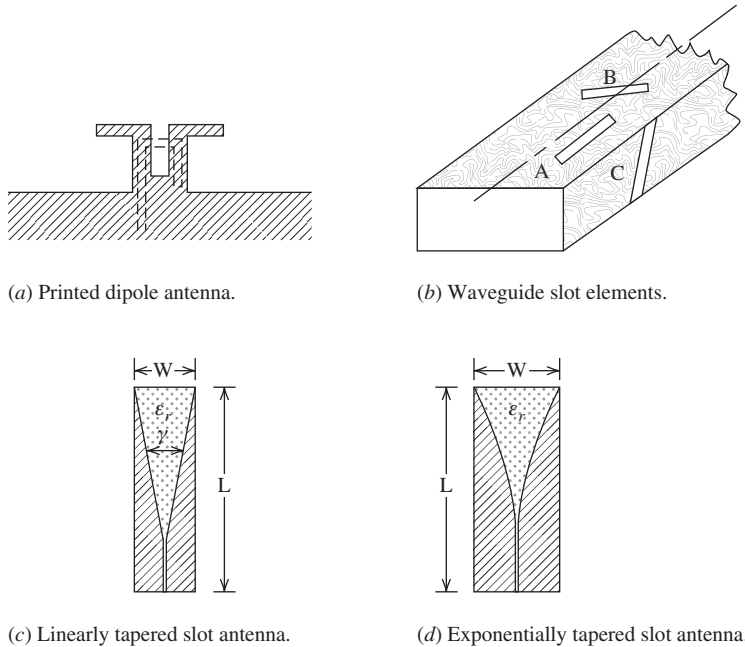(*d*) Exponentially tapered slot antenna.

Figure 8-35 Selected elements for array antennas.

The dipole is etched on to a printed circuit board together with a capacitively coupled loop, microstrip feed line, and is capable of 40% bandwidth. [H.8.2: Mailloux, 2nd ed., p. 236] Often the printed dipole is an integral part of a transceiver module. By lowering the ends of the dipole toward the ground plane, a *drooping dipole* is formed that can have 48% bandwidth. [H.8.2: Haupt, p. 271] Bandwidth can be increased to as high as 68% [24] by expanding the printed dipole into a bow-tie antenna; see Fig. 7-32. A bow-tie antenna array complete with feed network can be fabricated onto a single substrate as illustrated in Fig. 8-33. Monopole elements based on counterpart dipoles are used for fixed-phase endfire arrays and for phased arrays that scan off endfire.

Perhaps the best choice for a low-profile element is the microstrip antenna; see Sec. 11.2. It is ideally suited for printed circuit construction and has a fractional wavelength thickness. Its main disadvantage is that in its basic form, it has a bandwidth of only a few percent. Microstrip arrays are discussed in Sec. 11.3.

Another low-profile element is a slot antenna, which is a narrow slot in a ground plane as shown in Fig. 16-23. The slot is a complementary structure to the strip dipole of Fig. 6-8, where the air is replaced by metal and the metal is replaced by air to form the slot antenna. This complementary feature makes the strip dipole and slot antenna electromagnetic duals (see Sec. 3.4.1), so that the radiated electric and magnetic fields behave the same with their roles reversed. That is, any plane containing the long axis is an *E*-plane for the strip dipole but is the *H*-plane for the slot antenna. So, in Fig. 16-23, the *xz*-plane is the *E*-plane, whereas if the slot were replaced by metal forming a strip dipole, the *yz*-plane would be an *E*-plane. The patterns of a resonant half-wavelength slot in a finite ground plane are shown in Fig. 16-26. Practical single slot antennas are backed by a cavity to give a unidirectional pattern. Slot arrays are usually constructed with rectangular waveguides forming a *waveguide slot array antenna* as shown in Fig. 8-35*b*, often fabricated by milling slots in the waveguide. The array is series fed, giving a compact feed network. The slots are always designed to be resonant slots, but the array itself can be a traveling-wave (or non-resonant) or standing-wave (or resonant) array, which are terminated in a dummy load or a short, respectively. Traveling-wave arrays can be frequency scanned, whereas standing-wave arrays are used for broadside beams. Slots

are placed in the waveguide walls to interrupt the currents flowing in the walls, creating an electric potential across the narrow dimension of the slot causing it to radiate. Thus, the narrow wall of the slot determines the polarization; this was noted previously by it being the *E*-plane. The location, orientation, and length of the slots all enter into the design of a slot array, which can be complicated; see [H.6: *Ant. Eng. Hdbk.*, 4th ed., Chap. 9] for design procedures. The slots can be longitudinal (i.e., parallel to the waveguide axis as shown in A of Fig. 8-35*b*) in the waveguide broadwall, inclined (rotated) in the broadwall (B of Fig. 8-35*b*), or inclined in the narrow wall (C of Fig. 8-35*b*)—also called an edge slot. The amplitude of excitation of each slot in the array is determined by the offset from the center axis for longitudinal slots and the amount of angular rotation for the inclined slots. Excitation phase is determined by which side of the center axis the longitudinal slots are placed and the direction of rotation for inclined slots, as well as the slot length. Waveguide slot arrays have good efficiency and are an excellent choice for high-power applications, but usually are narrow band (less that 5%). An example is the AWACS array of Fig. 8-35*b*, which is a traveling-wave array with 160 narrow wall slots in the longest row.

Aperture elements such as horn antennas (see Sec. 9.4) are used only for limited scan arrays with large element spacings that can accommodate the large aperture size of a horn. An exception is an array of open-ended waveguides (see Sec. 9.2.2), which have fractional wavelength dimensions. The open-ended waveguide radiator of Example 9-3 is 0.7λ by 0.3λ. Being waveguide based, the element has moderate bandwidth (40% or more), but the physical structure can be large and heavy. Many types of directive elements can be used in limited scan arrays such as Yagi-Uda, helix, and others.

The *tapered slot antenna* (TSA),[6] also sometimes referred to as a *notch antenna* (or *flared notch antenna*), produces a useful endfire pattern over a bandwidth of 10 : 1 or more. It has moderate gain (7 to 10 dB) and balanced *E*- and *H*-plane beamwidths, making it a good array element candidate. [H.8.4: Lee and Chen, Chap. 9] There are several forms the TSA can have, but the *linearly tapered slot antenna* (LTSA) of Fig. 8-35*c* and the *exponentially tapered slot antenna* (ETSA) of Fig. 8-35*d* are two popular ones. They are constructed by etching a slot into a metal film, usually on a low-ε_r substrate. As a recent antenna innovation, the full understanding of its operation is incomplete, but basic principles are clear as we now discuss. The length *L* varies from about more than one wavelength up to many wavelengths over the operating band. The low frequency limit is determined by the slot width, *W*, which must be greater than $\lambda_L/2$. The feed region dictates the high-frequency limit. The TSA's traveling-wave behavior is responsible for its broad bandwidth. For operation near mid-band, the wave is bound in a transmission mode in the small-gap slot region near the feed, and when the wave reaches a region where the slot has widened to around a half-wavelength the power is converted to a radiation mode. TSAs are often fed using a microstrip-to-slotline transition at the slot line input. The LTSA is similar in geometry to the two-dimensional TEM (transverse electromagnetic) horn that is a planar version of a waveguide horn like the one in Fig. 9-11*b*. [H.8.2: Hansen, 2nd ed., p. 149] However, TSAs function more like a frequency independent antenna instead of an aperture antenna.

The LTSA has sides that expand linearly. With a flare angle γ in the range of 15 to 20°, it has nearly equal *E*- and *H*-plane beamwidths around 30°, even though the antenna is larger in the *E*-plane (the plane of the paper is the *E*-plane in Fig. 8-35*c* and *d*). [25] The ETSA has sides conforming to an exponentially flaring curve and is also called a *Vivaldi antenna*, a name coined in 1979 [26].

When TSAs are used as elements in an array there are unexpected benefits associated with the mutual coupling, which is explored further in the next section. The element performs best with a smaller size than when used as a single antenna. [27] The length can

---

[6] The term slot in this context does not imply there is radiation through a slot as in Fig. 16-23.

be reduced to under a wavelength and the width to under a half wavelength. This result allows closely spaced elements, which makes wide bandwidth possible. An experimental study of a dual-linearly polarized array (formed with orthogonal TSA element pairs) in a 9×8 grid demonstrated performance over a 5.9 : 1 bandwidth. The element spacing (in both planes of the planar array) varied from about $0.1 \lambda_L$ to $0.6 \lambda_U$ over the band. [28] Simulations with the array showed that scanning out to $45°$ in all scan planes is achievable. [29] The high-frequency performance of TSA arrays is limited by narrow-band resonances introduced by the mutual coupling, but they can be reduced by adding plated-through vias. [28]

Additional elements are introduced in the next section that are low-profile as well as being suitable for broadband arrays.

8.11   WIDEBAND PHASED ARRAYS

In Sec. 7.1 we defined a broadband antenna as one with a bandwidth of about 2:1 or greater. Broadband behavior is more difficult to achieve with an array than with a single element. In this section, we address arrays that operate over a significant bandwidth, typically around 2 : 1 or more, and refer to such an array as a **wideband array**. The distinction between the terms broadband and wideband are minimal, but it is common to use wideband with arrays with large bandwidth, so we adopt that convention. In this section, we first discuss fundamental principles for wideband array operation and then present types of wideband arrays that are of low-profile geometry. Multiband antennas that have multiple narrow operating bands spanning a wide frequency range are not included in this section, but instead we address arrays that operate over a continuous wide bandwidth. In some cases the array needs to support instantaneous bandwidth (see Sec. 7.1) over the band, such as with an UWB application discussed in Sec. 7.10.

A conventional wideband array with a periodic grid requires elements that span the same wide bandwidth. As discussed in Sec. 8.9.1, the element spacing d is limited by the appearance of grating lobes and other effects such as blind spots with (8-117) being the guideline. The spacing limit must be satisfied over the operating band and will give the largest physical distance value at the lowest operating frequency, f_L. Of course, the maximum physical extent of the element, D, in a row or column of the array must not exceed the spacing, so $D < d$. As an example, for a $50°$ scan angle off broadside ($\Delta\theta = 50°$, $\theta_o = 40°$), (8-117) gives $d < 0.57 \lambda$ and then the maximum extent of the element is $D = 0.57 \lambda$. It turns out that this physical size limitation is usually more difficult to satisfy than finding an element with sufficient bandwidth. The array band-width limitation can be quantified using the graph in Fig. 8-36 which compares popular broadband antennas used as elements in phased arrays. [30] The maximum lateral extent (diameter) of the Archimedean and equiangular spirals is about $D = 0.5 \lambda_L$ and for the sinuous antenna is about $D = 0.4 \lambda_L$, based on principles in Sec. 7.7. The Foursquare antenna of Fig. 8-38, to be discussed, has a maximum lateral extent of $D = 0.35 \lambda_L$. The minimum element spacing is obtained by equating to the element extent, $d = D$, which corresponds to zero separation between elements edges. As frequency is increased from the low end of the operating band where $\lambda = \lambda_L$, the spacing normalized by wavelength increases as given by $d/\lambda = (d/\lambda_L)(f/f_L)$, which is used to plot the straight lines in Fig. 8-36. For example, at $f/f_L = 2$ the spiral antennas have a minimum spacing of $d/\lambda = (d/\lambda_L)(f/f_L) = 0.5 \cdot 2 = 1$. The maximum scan angle off broadside $\Delta\theta$ found from the grating lobe appearance limit of (8-117) is shown on the right ordinate of Fig. 8-36. For the preceding scan limit example, it is seen that $\Delta\theta = 50°$ on the right ordinate corresponds to $d/\lambda = 0.57$ on the left ordinate. The plot is used to find candidate elements for an array of a given bandwidth. We illustrate for an array with a 50% bandwidth, which is a ratio bandwidth of $B_r = f_U/f_L = 1.67$. Drawing a vertical line from 1.67 on the ordinate intersects the three curves at 0.58, 0.67, and 0.84 for the element types. The corresponding maximum scan angles read from the right ordinate are $46°$, $30°$, and $11°$. That is, the
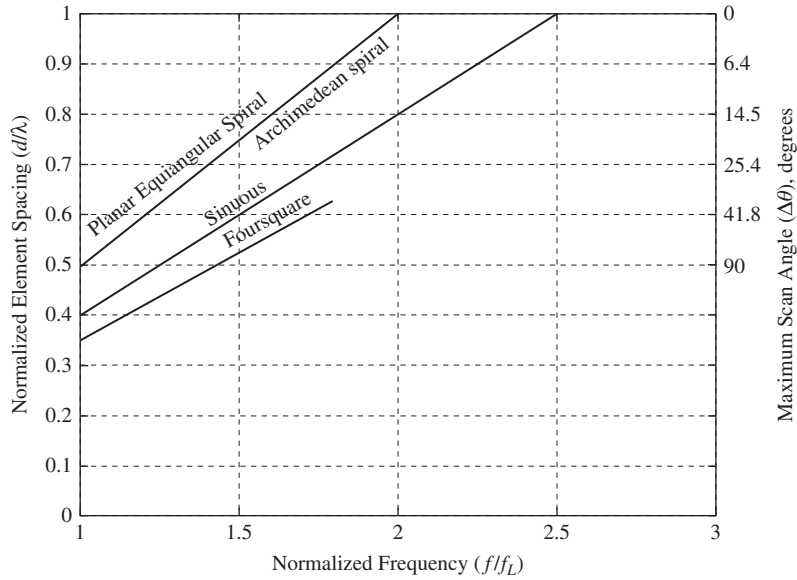
Figure 8-36 Minimum element spacing normalized to a wavelength versus normalized frequency for various broadband element types based on the physical extent of the elements. The right ordinate shows the maximum scan angle. (From [3] © 2000. Reprinted with permission from *Microwave Journal*.)

Foursquare array can scan to $46°$ at the upper frequency while the spiral arrays can only scan to $11°$ without a grating lobe appearing. Thus, we see that as the bandwidth requirement is increased for an array, the more compact the element must be in addition to covering the array bandwidth. For most applications, an extremely broadband element is unnecessary. Having a compact element is usually the most important factor.

For broadside operation there is no bandwidth limitation due to array size or feed network architecture. Of course, in all wideband arrays the element and feed network components (phase shifters, power dividers, etc.) must be wideband. However, when a phased array is scanned off broadside the main beam will steer to the intended scan angle of θ_o only at the center frequency f_c. This effect is illustrated for simplicity with a linear array along the z-axis. The phase shifters are selected to produce an array factor peak at angle θ_o at f_c, and then the phase terms in (8-5) are zero, or $0 = \beta_c d \cos \theta_o + \alpha$. When frequency is changed to f, the peak shifts to angle θ, or $0 = \beta d \cos \theta + \alpha$. Solving these two equations gives

$$\theta = \cos^{-1}\left(\frac{f_c}{f}\cos \theta_o\right) \tag{8-122}$$

This equation gives the angle for the peak of the beam of a linear array along the z-axis at frequency f with the phase shifters set to give a beam peak at angle θ_o at frequency f_c. For $f = f_c$, (8-122) gives $\theta = \theta_o$ as expected. As frequency deviates more from the center frequency, the beam peak direction correspondingly deviates ("squints") more from the desired pointing direction θ_o. The beam peak angle is closer to (farther from) broadside compared to the center frequency direction for frequencies above (below) the center frequency. Frequency dependent scan can be avoided by using time delay, as discussed in Sec. 8.9.2. See [H.8.2: Mallioux, 2nd ed., p. 31] for more details on bandwidth effects on phased arrays.

Recent innovations in arrays have led to array architectures that avoid the compact element size requirement of conventional uniform grid arrays, including reconfigurable arrays, aperiodic-grid arrays, variable element size arrays, and tightly coupled arrays. A *reconfigurable array* is altered dynamically to change its frequency band and, thus, does not fit the topic of this section. Reconfigurable arrays, discussed in Sec. 12.5, are usually implemented by activating switches internal to the elements to change their frequency range, often with MEMS switches. An *aperiodic-grid array* (i.e., an unequally spaced array) avoids the appearance of full grating lobes of a conventional periodic grid array because grating lobe formation requires a periodic geometry. Aperiodic arrays are more difficult to design because the mutual coupling environment is not repeatable across the array, leading to varying element impedances and patterns.

A special type of aperiodic-grid array employing different size elements is called an *interlaced-element array*, with one form being the *wideband array with variable element sizes* (*WAVES array*). [31] An example of a basic unit used to make large WAVES arrays shown in Fig. 8-37a consists of eight elements with two sizes of spiral antennas to cover two octaves (4 : 1) of bandwidth. The small spirals have diameter D_S and the large spirals have diameter D_L which are twice as large as the small spirals ($D_L = 2D_S$). The large elements operate over the full band of two octaves of f_1 to $f_2(= 2f_1)$ and f_2 to $f_4(= 2f_2)$. The small elements are operational only over the upper octave from f_2 to f_4. The patterns of the large spiral are shown in Fig. 7-29. At the lowest operating frequency f_1, the element spacings in the principal directions are a half-wavelength, so $S = S_x = S_y = \lambda_1/2$. At the top of the first octave $f_2 = 2f_1$ the electrical spacing is $S = \lambda_2$, which will lead to grating lobes. At f_2 the small elements become operational, lowering the spacing between the active elements to $\lambda_2/2$. Increasing frequency further to $f_4 = 2f_2$, the spacing once again becomes one wavelength, λ_4. The patterns for the array were computed using simple array theory and were measured for an experimental model operating from 2.5 to 8 GHz. [31] Representative patterns are shown in Fig. 8-37b and c for the low end of the band at 2.5 GHz with the four large elements active where the spacing is a half wavelength, and for the top end of the band at 10 GHz with all eight elements active where the spacing is one wavelength. Note that the pattern for one-wavelength spacing is acceptable because the element pattern of the spiral (see Fig. 7-29) diminishes the grating lobe effect, but smaller spacings would be needed in low-side lobe or wide-angle scanning arrays. The WAVES concept has been extended to at least three octaves of bandwidth (8 : 1) by distorting the large elements to make room for the smaller elements [32] or by using multiple layers of radiating elements, one layer for each octave. [33]

A recent innovation in wideband phased arrays is the *tightly coupled array*. While conventional array design seeks to reduce mutual coupling, it turns out that there are advantages to tightly packed elements that have high coupling. The unexpected benefit is that the bandwidth of an embedded element antenna is larger than when isolated. As with any array, the factor limiting bandwidth is the ground plane. But the second surprise benefit for tightly coupled arrays is that the elements can be placed close to the ground plane, often under $\lambda/10$ at the lowest operating frequency of the array. A related concept is that of a *connected array*, where the elements are conductively connected; examples are connected dipoles joined at the ends [34] and tapered slot elements joined on the sides [35]. The types of tightly coupled arrays are the Foursquare array, the current sheet array, the fragmented array, and the long slot array.

Perhaps the first successful tightly coupled array was developed at Virginia Tech in the 1990s. [36] Related geometries for isolated elements have evolved, such as the Fourpoint antenna shown in Fig. 12-7a which offers dual polarization and balanced patterns over 2.7 : 1 bandwidth (see Fig. 12-7b), covering common wireless bands in one compact antenna, and is an excellent choice for base stations. [37] The Foursquare element has four squares excited with a balanced feed for each opposing pair of squares, giving simultaneous dual linear polarizations. A 2×2 illustrative array is shown in Fig. 8-38. [38] The inactive squares can be thought of as sleeves for the active squares, thus
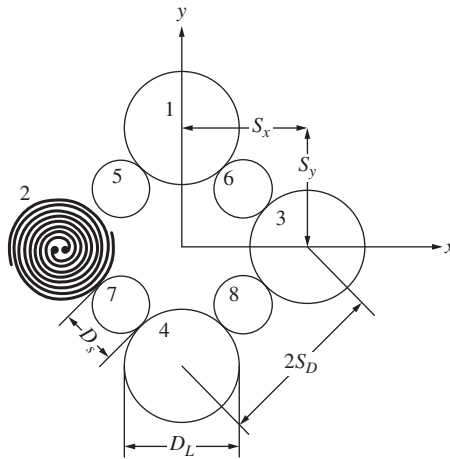
(*a*) Geometry for an eight-element, two-octave
concept demonstration array.
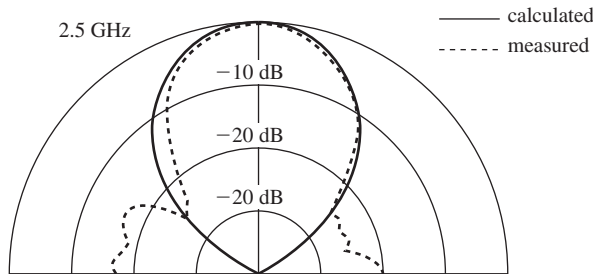
(*b*) Pattern at 2.5 GHz with four large elements active.
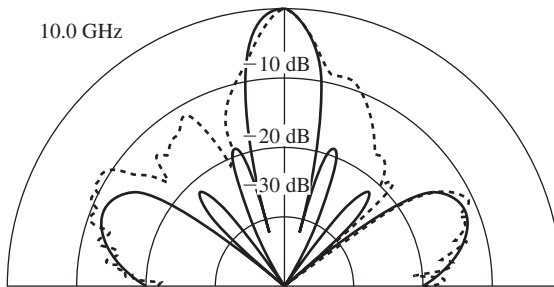
(*c*) Pattern at 10 GHz with eight large elements active.

Figure 8-37 An eight-element example wideband array with variable element sizes (WAVES array). (From [31] © 1990. Reprinted with permission from IET.)

widening bandwidth of an element as explained in Sec. 7.5. In an array, the neighboring elements support current associated with radiation from an element through coupling across the narrow gap between elements. [38] The original Foursquare arrays had a bandwidth of about 2:1. [38] The elements were about a quarter wavelength on each side and a tenth wavelength above the ground plane at the low end of the band. Subsequent research has shown 3.5:1 bandwidth is possible. [39]

The *current sheet array* (*CSA*), whose name comes from Wheeler's current sheet [23], is an array of electrically small collinear dipoles in each principal plane and backed by a ground plane less than a tenth of a wavelength away at the low end of the band. The
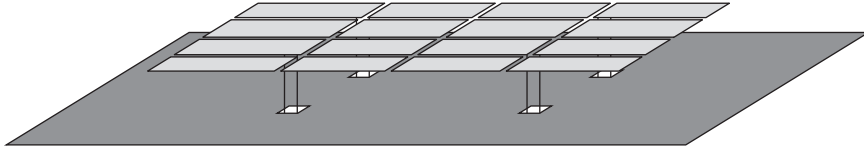
Figure 8-38 The Foursquare antenna used in a 2×2 tightly coupled array. [38]

largest element spacing is a half-wavelength at the high end of the band, just like the Foursquare array. The element ends are joined with capacitors to increase coupling. Experiments at Harris Corp. demonstrated a 9:1 bandwidth. [40] Additional details are available in [H.8.2: Munk, Chap. 6] and [H.6: Balanis, Ed., Sec. 12.2].

The *fragmented array* is designed using a genetic algorithm routine to form pixilated features. Bandwidths close to 10:1 are possible with simple ground planes. [H.6: Balanis, Ed., Sec. 12.4] The *long slot array* is a low-profile array of about $\lambda_L/8$ thickness and 4:1 bandwidth with a ground plane. [41] Array bandwidth is usually limited by the presence of a ground plane. Ground plane options include good conductors (i.e., metals), absorbing materials, or artificial materials, or a combination of these. For conducting ground planes, the conventional requirement of quarter-wavelength spacing away from the radiating elements limits the bandwidth to about 25% with conventional arrays. We have seen that tightly coupled arrays with conducting ground planes are capable of much wider bandwidths, a decade or so. Even greater bandwidths are possible by including absorbing materials such as resistive sheets or ferrite materials, but the associated loss reduces the gain. *Artificial materials* are artificial in the sense that they are not naturally occurring but are engineered using common materials in textured structures to create special electromagnetic properties. Such materials have origins in *frequency selective surfaces* that are constructed to pass certain frequency bands and reflect others for use in reflector antenna and radome applications. An example is an *artificial magnetic conductor* (*AMC*), which is the dual of a PEC. An electric antenna (such as a dipole) can be spaced arbitrarily close to a AMC ground plane. Similarly, a magnetic antenna (such as a loop) can be spaced arbitrarily close to a PEC ground plane. Close ground plane spacing enables antennas that are physically much thinner at VHF frequencies (and below) than conventional quarter-wave ground plane spaced antennas. See [H.6: *Ant. Eng. Hdbk.*, 4th ed., Chap. 34] and [H.6: Balanis, Ed., Chap. 15] for more details.

In closing, we mention that there are wideband arrays that cover multiple frequency bands by using different types of elements for each band. In this *shared-aperture array* approach, the elements of one frequency band are interleaved with elements of different frequency bands. Such an array is often used to support multiple functions (communications, radar, sensing, position location, etc.) and thus is called *multifunctional wideband array*.

REFERENCES

1. C. T. Tai, "The Optimum Directivity of Uniformly Spaced Broadside Arrays of Dipoles,"*IEEE Trans. on Ant. & Prop.*, Vol. AP-12, pp. 447–454, July 1964.
2. R. C. Hansen, "Hansen-Woodyard Arrays with Few Elements," *Microwave and Optical Technology Letters*, Vol. 5, pp. 44–46, Jan. 1992.
3. H. Bach, "Directivity Diagrams for Uniform Linear Arrays," *Microwave J.*, Vol. 15, pp. 41–44, Dec. 1972.
4. H. Bach, "Directivity of Basic Linear Arrays," *IEEE Trans. on Ant. & Prop.,* Vol. AP-18, pp. 107–110, Jan. 1970.
5. N. A. McDonald, "Approximate Relationship between Directivity and Beamwidth for Broadside Collinear Arrays," *IEEE Trans. on Ant. & Prop.,* Vol. AP-26, pp. 340–341, Mar. 1978.

6. W. L. Stutzman, "Estimating Directivity and Gain of Antennas," *IEEE Trans. on Ant. & Prop. Mag.*, Vol. 40, pp. 7–11, Aug. 1998.

7. J. Rashed-Mohassel, "Optimum Directivity of a Uniformly Spaced Broadside Array of Collinear Half-Wave Dipoles," *Microwave and Optical Technology Let.*, Vol. 7, pp. 193–196, Mar. 1994.

8. Peter Hannan, "The Element-Gain Paradox for a Phased-Array Antenna," *IEEE Trans. on Ant. & Prop.,* Vol. AP-12, pp. 423–433, July 1964.

9. A. Parfitt, D. Griffin, and P. Cole, "Mutual Coupling between Metal Strip Antennas on Finite Size, Electrically Thick Dielectric Substrates," *IEEE Trans. on Ant. & Prop.*, Vol. 41, pp. 108–115, Jan. 1993.

10. D. Pozar, "A Relation between the Active Input Impedance and the Active Element Pattern of a Phased Array," *IEEE Trans. on Ant. & Prop.*, Vol. 51, pp. 2486–2489, Sept. 2003.

11. D. Kelley and W. Stutzman, "Array Antenna Pattern Modeling Methods That Include Mutual Coupling Effects," *IEEE Trans. on Ant. & Prop.*, Vol. 41, pp. 1625–1632, Dec. 1993.

12. K. Takamizawa and W. Stutzman, "Analysis and Synthesis of Microstrip Antennas Including Mutual Coupling," Virginia Tech Antenna Group Report 88-2, Aug. 1988.

13. D. Pozar, "Analysis of Finite Phased Arrays of Printed Dipoles," *IEEE Trans. on Ant. & Prop.*, Vol. AP-33, pp. 1045–1053, Oct. 1985.

14. B. J. Forman, "Directivity Characteristics of Scannable Planar Arrays," *IEEE Trans. on Ant. & Prop.,* Vol. AP-20, pp. 245–252, May 1972.

15. H. King and J. L. Wong, "Directivity of a Uniformly Excited $N \times N$ Array of Directive Elements,"*IEEE Trans. on Ant. & Prop.*, Vol. AP-23, pp. 401–404, May 1975.

16. D. Pozar, "The Active Element Pattern," *IEEE Trans. on Ant. & Prop.*, Vol. 42, pp. 1176–1178, Aug. 1994.

17. J-P. Bayard, D. Schaubert, and M. Cooley, "E-Plane Scan Performance of Infinite Arrays of Dipoles Printed on Protruding Dielectric Substrates: Coplanar Feed Line and E-Plane Metallic Wall Effects," *IEEE Trans. on Ant. & Prop.*, Vol. 41, pp. 837–841, June 1993.

18. D. Pozar, "Scanning Characteristics of Infinite Arrays of Printed Antenna Subarrays," *IEEE Trans. on Ant. & Prop.*, Vol. 40, pp. 666–674, June 1992.

19. F. Zavosh and J. Aberle, "Infinite Phased Array of Cavity-Backed Patches," *IEEE Trans. on Ant. & Prop.*, Vol. 42, pp. 390–398, Mar. 1994.

20. M. Deshpande and M. C. Bailey, "Analysis of Finite Phased Arrays of Circular Microstrip Patches," *IEEE Trans. on Ant. & Prop.*, Vol. 37, pp. 1355–1360, Nov. 1989.

21. E. Brookner, "Phased-array Radars," *Scientific American*, pp. 94–102, Feb. 1985.

22. J. Butler and R. Lowe, "Beam-Forming Matrix Simplifies Design of Electronically Scanned Antennas," *Elec. Design*, Vol. 9, pp. 170-173, Apr. 12, 1961.

23. H. A. Wheeler, "Simple Relations Derived from a Phased-Array Antenna Made of an Infinite Current Sheet," *IEEE Trans. on Ant. & Prop.*, Vol. 40, pp. 506-514, July 1965.

24. G. Zheng et al., "A Broadband Printed Bow-Tie Antenna with a Simplified Balanced Feed," *Microwave and Optical Tech. Let.*, Vol. 47, pp. 534-536, Dec. 20, 2005.

25. P. Kool, T. Yeo, and M. Leong, "Parametric Studies of the Linearly Tapered Slot Antenna (LTSA)," *Microwave and Optical Tech. Let.*, Vol. 4, pp. 200–207, Apr., 1991.

26. P. Gibson, "The Vivaldi Aerial," *Proc. 9th European Microwave Conf.*, pp. 101–105, 1979.

27. J. Shin and D. Schaubert, "A Parameter Study of Stripline-Fed Vivaldi Notch-Antenna Arrays," *IEEE Trans. on Ant. & Prop.*, Vol. 47, pp. 879–886, May 1999.

28. H. Holter, T-H Chio, and D. Schaubert, "Experimental Results of 144-Element Dual-Polarized Endfire Tapered-Slot Phased Array," *IEEE Trans. on Ant. & Prop.*, Vol. 48, pp. 1707–1718, Nov. 2000.

29. T-H Chio and D. Schaubert, "Parameter Study and Design of Wide-Band Widescan Dual-Polarized Tapered-Slot Antenna Arrays," *IEEE Trans. on Ant. & Prop.*, Vol. 48, pp. 879–886, June 2000.

30. W. Stutzman and C. Buxton, "Radiating Elements for Wideband Phased Arrays," *Microwave J.*, Vol. 43, pp. 130-141, Feb. 2000.

31. S. Shively and W. Stutzman, "Wideband Arrays With Variable Element Sizes," *IEE Proc., Pt. H.*, Vol. 137, pp. 238–240, Aug. 1990.

32. E. Caswell, "Design and Analysis of Star Spiral with Applications to Wideband Arrays with Variable Element Sizes," Ph.D. Dissertation, Virginia Tech, 2001.

33. W. Stutzman, "Wide Bandwidth Antenna Array Design," *IEEE Southeastcon Proceedings* (Raleigh, NC), pp. 92–96, April 1985.

34. R. C. Hansen, "Linear Connected Arrays," *IEEE Ant. and Wireless Prop. Letters*, Vol. 3, pp. 154–156, 2004.

35. J. Lee, S. Livingston, and R. Koenig, "A Low-Profile Wide-Band (5:1) Dual-Pol Array," *IEEE Ant. and Wireless Prop. Let.*, Vol. 2, pp. 46–49, 2003.

36. J. R. Nealy, "Foursquare Antenna Radiating Element," U. S. Patent 5,926,137, July 20, 1999.

37. S-Y. Suh, W. Stutzman, W. Davis, A. Waltho, K. Skeba, and J. Schiffer, "A Novel Low-profile, Dual-polarization, Multi-band Base-station Antenna Element—The Fourpoint Antenna," *IEEE Vehicular Technology Conference Proc.* (Los Angeles, CA), Vol. 1, pp. 225–229, Sept. 2004.

38. C. Buxton, W. Stutzman, and J. Nealy, "Implementation of the Foursquare Antenna in Broadband Arrays," URSI National Radio Science Meeting (Orlando), July 1999.

39. Terry Vogler, "Analysis of the Radiation Mechanisms in and Design of Tightly-Coupled Antenna Arrays," Ph.D. Dissertation, Virginia Tech, 2010.

40. M. Jones and J. Rawnick, "A New Approach to Broadband Array Design Using Tightly Coupled Elements," *IEEE Military Comm. Conf.*, Oct. 2007.

41. J. J. Lee, S. Livingstone, R. Koenig, D. Nagata, and L. Lai, "Compact Light Weight UHF Arrays Using Long Slot Apertures," *IEEE Trans. on Ant. & Prop.*, Vol. 54, pp. 2009–2015, July 2006.

PROBLEMS

8.2-1 Consider an array of two elements spaced one wavelength apart with currents that are equal in amplitude and 180° out-of-phase.

(a) Use the inspection method to rough sketch the polar plot of the array factor.

(b) Derive the exact array factor as a function of θ if the elements are on the z-axis.

(c) For what angles of θ is this array factor maximum?

(d) What is the expression for the normalized array factor $|f(\theta)|$?

(e) Show that (8-9) reduces to your answer in (d).

8.2-2 Use the techniques of Fig. 8-6 to obtain a polar plot of the array factor of the array given in Prob. 8.2-1.

8.2-3 Use the techniques of Fig. 8-6 to obtain a polar plot of the array factor of a two-element, one wavelength spaced array with equal amplitude and equal phase currents (Example 8-1).

8.2-4 Usually, the interelement spacing of an array is about one-half wavelength. Spacings much greater than this produce major lobes in undesired directions. To illustrate this point, use the techniques of Fig. 8-6 to sketch the array factor for a two-element array with equal amplitude, in-phase elements in polar form for the following spacings: (a) $d = 3\lambda/4$ and (b) $d = 2\lambda$. Examples 3-2 and 8-1 and this problem show the effects of spacing on an array of fixed excitation.

8.2-5 (a) Using the array factor for a two-element broadside array ($\alpha = 0$) with equal current amplitude point source elements, show that the directivity expression is

$$D = \frac{2}{1 + (\sin \beta d)/\beta d}$$

Hint: Change from variable θ to $\psi = \beta d \cos \theta$.

(b) Plot this expression as a function of d from zero to two wavelengths.

8.3-1 Prove that the array factor magnitude $|f(\psi)|$ for a uniformly excited, equally spaced linear array is symmetric about $\psi = \pi$.

8.3-2 Show that the array factor expressions (8-9) and (8-24) for a two-element uniformly excited array are identical.

8.3-3 Drive (8-33)

8.3-4 The expression for the half-power beamwidth of the array factor for a broadside, uniformly excited, equally spaced, linear array can be approximated as

$$\text{HP} \approx k \frac{\lambda}{Nd}$$

for $Nd \gg \lambda$. Determine k for $N = 10$ and 20, and compare to (8-34)

8.3-5 In this problem, the effect of phasings and spacings on a simple array are illustrated. Consider an equally spaced five-element array with uniform current amplitudes. Sketch the array factors for:

(a) $d = \lambda/2$, broadside case ($\theta_o = 90°$)
(b) $d = \lambda$, broadside case
(c) $d = 2\lambda$, broadside case
(d) $d = \lambda/2$, $\theta_o = 45°$
(e) $d = \lambda/2$, $\theta_o = 0°$, endfire case

The five plots can be obtained from one universal pattern plot as discussed in Sec. 8.3. For the last two cases, determine the interelement phase shift α required to steer the main beam as specified.

8.3-6 Repeat Prob. 8.3-5 using a computer code.

8.3-7 Design a five-element uniformly excited, equally spaced linear array for:
(a) Main beam maximum at broadside
(b) Main beam maximum at 45° from broadside ($\theta_o = 45°$)

In each case, select the element spacing and linear phasing such that the beamwidth is as small as possible and also so that no part of a grating lobe appears in the visible region. Sketch the polar plots of the patterns.

8.3-8 Design and plot the array factor for an ordinary endfire, five-element, uniformly excited linear array with spacings $d = 0.35\lambda$. Use $\theta_o = 180°$ and find α.

8.3-9 Design a linear array of four isotropic elements for a single ordinary endfire beam using the maximum spacing. Plot the polar pattern.

8.3-10 Design a linear array of five isotropic elements for Hansen–Woodyard increased directivity with $d = 0.35\lambda$. Plot the polar pattern. (Note the differences in the results from Example 8-4.)

8.3-11 Design a linear array of 10 isotropic elements for Hansen–Woodyard increased directivity with $d = 0.4\lambda$. Plot the polar pattern.

8.3-12 Show that the array factor for a uniformly excited, equally spaced linear array approaches the pattern factor of a uniform line source (i.e., neglect the element factor) in the limit of small array element spacings.

8.3-13 Derive the pattern normalization factor used in (8-39). Consider both endfire directions (i.e., $\theta_o = 0$ and $180°$).

8.4-1 Two collinear half-wave dipoles are spaced a half-wavelength apart (but not quite touching) and have the same amplitude and phase terminal currents. Calculate and plot the element pattern, array factor, and array pattern.

8.4-2 Repeat Prob. 8.4-1 for one-wavelength spacing.

8.4-3 Two parallel half-wave dipoles are spaced a one wavelength apart and have the same amplitude and phase terminal currents. Calculate and plot the full array pattern in the two principal planes.

8.4-4 A linear array of three quarter-wavelength-long, vertical monopoles is operated against an infinite, perfectly conducting ground plane. Let the element feeds be along the z-axis, the ground plane in the yz-plane, and the monopoles in the x-direction.
(a) Design the array as a Hansen–Woodyard increased directivity endfire array; that is, determine the element spacings and phasings (choose $d = 0.3\lambda$).
(b) Use the universal array factor plot for three uniformly excited elements to obtain a polar plot of the array factor for this problem.
(c) Write the expression for the complete pattern.
(d) Plot the complete far-field patterns in the xz-plane and yz-plane.

8.4-5 Three collinear half-wave dipoles are spaced $0.9\,\lambda$ apart and have the same amplitude and phase terminal currents. (a) Use the graphical techniques of Sec. 8.2 to sketch array factor. Then sketch the element and array patterns. (b) Use the computer to calculate and plot the array pattern.

8.4-6 An array of four small loop antennas oriented with the loops in the yz-plane are spaced 0.4λ apart along the z-axis, have uniform amplitudes, and are phased for ordinary endfire. Plot the full array pattern in the two principal planes.

8.4-7 Suppose a truck uses a Citizens Band radio to communicate at 27 MHz. The antenna system is two quarter-wave monopoles parallel to the x-axis (assumed to operate above a perfect ground plane) mounted on mirrors 2.78 m apart along the z-axis and fed with equal amplitude and phase. Use simple array theory to obtain sketches of the array patterns in the three principal planes.

8.4-8 *Design problem*. Design a broadside linear array of four parallel half-wave dipoles for as narrow a beamwidth as possible and with no level outside the main beam above $-8\,\mathrm{dB}$ relative to the main beam peak. The excitations are uniform. (a) Determine the required element spacing. (b) Sketch the polar pattern in the H-plane.

8.4-9 Plot the principal plane full array patterns for the endfire array of Prob. 8.3-8 with parallel half-wave dipoles elements. Find the half-power beamwidths.

8.5-1 Calculate the directivities in decibels for the following broadside arrays of point sources:
   (a) $N = 2$, $d = \lambda/2$.
   (b) $N = 10$, $d = \lambda/2$.
   (c) $N = 15$, $d = \lambda$.

8.5-2 Evaluate (8-67) for $d = 3\lambda/8$ and $N = 10$ for:
   (a) Broadside, and compare the result to that of (8-69),
   (b) Ordinary endfire, and compare the result to that of (8-70).

8.5-3 Evaluate (8-67) and plot D versus d/λ for $N = 10$ and ordinary endfire operation. Your results should look like the curve in Fig. 8-17.

8.5-4 The approximate directivity formula of (8-69) for long, broadside linear arrays of isotropic elements can be checked in the following two ways using $\mathrm{HP} \approx 0.886\ \lambda/L$ from (8-34):
   (a) Use $D = 4\pi/\Omega_A$ to derive an equation of the form $D = cL/\lambda$ by approximating Ω_A as $2\pi\mathrm{HP}$; find c.
   (b) Begin with (8-74) to derive an equation of the form $D = cL/\lambda$; find c.

8.5-5 Develop a computer code to evaluate directivity using (8-67).
   (a) Plot the D versus d/λ curve for a broadside array with eight isotropic elements; compare to the curve in Fig. 8-16.
   (b) Plot the D versus d/λ curve for a broadside array of eight collinear short dipole elements.

8.5-6 Show that $D = N$ for an ordinary endfire linear array of uniformly excited isotropic elements when $d = \lambda/4$.

8.5-7 Evaluate the directivity of a broadside array of 12 collinear short dipole elements spaced 0.7λ apart in the following ways:
   (a) Use (8-67).
   (b) Use (8-69).
Determine the value of HP and use in the following to calculate directivity:
   (c) Eq. (8-73).
   (d) Eq. (8-74).

8.5-8 Calculate the directivity for the array of Example 8-5. How much more directivity is there compared to that of a single short dipole, in dB?

8.5-9 Calculate and plot the directivity of an array of 10 collinear short dipoles for spacings of 0.1 to $1.0\,\lambda$. Also, plot the curves for the approximate formulas (8-69) and (8-75).

8.5-10 A five-element array of five parallel short dipoles spaced $0.3\,\lambda$ apart is designed for Hansen–Woodyard endfire operation. Evaluate the directivity and compare to the directivities for the same array with isotropic elements and a similar ordinary endfire array of isotropic elements.

8.5-11 *Design problem*. Base station communication antennas are often constructed using a collinear array of half-wave dipoles oriented vertically to produce an omnidirectional pattern in the horizontal plane. The objective in this problem is to maximize gain by selecting the proper element spacing. Assuming uniform amplitude and phase excitation, design a maximum-directivity four-element array for the middle of the cellular telephone band (824 to 894 MHz). Show solution details. Give the values of spacing d in wavelengths and directivity D in decibels at midband and at the band edges. Plot the vertical-plane pattern in polar-linear form. Sketch the array, showing its physical length along, with a feed network.

8.5-12 The approximate array directivity formula $D \approx D_e D_i$ of (8-75) is attractive because of its simplicity, but can yield inaccurate results in many cases as this problem shows. Evaluate the directivity of the following arrays of short dipoles using the exact formula of (8-72).

(a) $N = 4$, broadside, collinear short dipoles, $d = \lambda/2$.
(b) $N = 4$, endfire, parallel short dipoles, $d = \lambda/2$.
(c) $N = 3$, broadside, collinear short dipoles, $d = \lambda/2$.
(d) $N = 3$, endfire, parallel short dipoles, $d = \lambda/2$.

Then compare results to (8-75) for cases to which (8-75) is applicable.

8.5-13 *Array directivity computation.*
  (a) Write a computer program to evaluate the directivity of an array by direct integration. Validate the code by comparing to values obtained using (8-72) for an array of four collinear dipoles that are spaced a half-wavelength apart.
  (b) Use the direct-integration directivity code to evaluate the directivity of arrays of collinear half-wave dipoles of 2, 3 and 4 elements spaced a half wavelength apart and uniformly excited. Compare to directivity of the same arrays with isotropic elements.

8.6-1 Calculate and plot the polar pattern for the five-element array with the triangular distribution in Fig. 8-21*b*. Find the half-power beamwidth in degrees and side lobe level in dB.

8.6-2 Calculate and plot the polar pattern for a five-element array with a triangular distribution similar to Fig. 8-21*b*, but with collinear half-wave dipole elements. Compare values of half-power beamwidth in degrees, side lobe level in dB, and directivity to the isotropic element case.

8.6-3 Prove that (8-83) follows from (8-82).

8.6-4 Prove that (8-85) follows from (8-83).

8.6-5 Verify the directivity values given in Figs. 8-20*c* through 8-20*e*, and in Fig. 8-22.

8.6-6 *Binomial array.* For a linear array of N isotropic elements spaced a half-wavelength apart and that have binomial current weightings:
  (a) Derive the normalized array factor expression in terms of θ.
  (b) Derive an expression for the directivity. The result will only be a function of N.
  (c) Evaluate the directivity expression derived in (b) for five elements.

8.7-1 Two antennas have the following self- and mutual impedances:

$$Z_{11} = 70 \angle 0°, \qquad Z_{22} = 100 \angle 45°, \qquad Z_{12} = 60 \angle -10°$$

  (a) Find the input impedance to antenna 1, if antenna 2 is short-circuited.
  (b) Find the voltage induced at the open-circuited terminals of antenna 2 when the voltage applied to antenna 1 is $1 \angle 0°$ V.

8.7-2 Derive (8-94), making use of Fig. 8-24.

8.7-3 For the parallel dipoles of Fig. 8-25 and $d = 0.6\,\lambda$, calculate the coupling level between the dipoles in dB.

8.7-4 *Mutual impedance calculation.* (a) Mutual impedance is calculated using simulation data in (8-87) for the case of antenna 1 excited and antenna 2 open circuited to find $Z_{12} = V_2/I_1$. Evaluate Z_{12} for parallel half-wave dipoles of the type in Fig. 8-25 and spaced $0.2\,\lambda$ apart. (b) Mutual impedance can also be calculated using the following formula:

$$Z_{12} = \frac{Z_{1,\text{in}} - Z_{1,\text{out}}}{2}$$

where $Z_{1,\text{in}}$ and $Z_{1,\text{out}}$ are the input impedances for antenna 1 for the two cases of two sources on the two dipoles that are in phase and then out of phase. Derive this formula. (c) Repeat the mutual impedance calculation in (a) but using the formula in (b) and appropriate simulations.

8.8-1 A planar array of four isotropic elements is arranged in the *xy*-plane with the following positions and currents: $(\lambda/4, \lambda/4), +1; (-\lambda/4, \lambda/4), +1, (-\lambda/4, -\lambda/4), -1; (\lambda/4, -\lambda/4), -1$. Use simple array modeling techniques to obtain sketches of the *xz*- and *yz*-plane patterns.

8.8-2 A four-element linear array of parallel, in-phase, half-wave dipoles is located $\lambda/4$ in front of a large planar reflector in the *xy*-plane. Assume the reflector to be a perfect ground plane. If the dipoles are parallel to the *x*-axis and spaced $\lambda/2$ apart, plot the complete pattern in the *xy*- and *yz* planes.

8.8-3 A two-element array of vertical short dipoles is operated a quarter-wavelength above a perfect ground plane as shown. The elements are a half-wavelength apart and are excited with

equal amplitude and opposite phase. Obtain polar plots for the radiation pattern of this radiating system in the *xz*- and *yz*-planes.
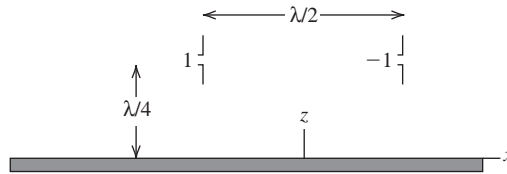
8.8-4 A two-dimensional, uniformly excited array of isotropic elements as shown below is to be analyzed. Use the principle of pattern multiplication with a pair of elements oriented vertically as an "array" and the four elements in a row as an "element." Plot the patterns in the *xz*-, *yz*- and *xy*-planes.
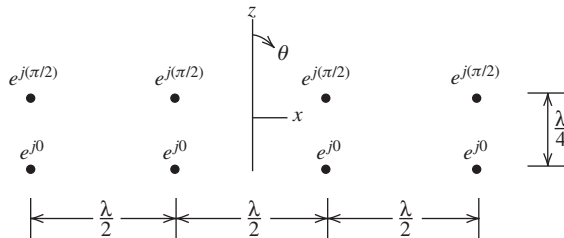
8.8-5 A planar array in the *xy*-plane has four isotropic elements in *x*- and *y*-directions with 0.5λ-spacings. Plot the radiation pattern (linear, polar plot) in the $45°$ plane. Then use the projection technique to find the equivalent linear array and then obtain the $45°$-plane pattern. Compare the patterns.

8.8-6 *Square and triangular grid planar arrays*. This problem compares square-grid and triangular-grid planar arrays in the *xy*-plane of the same area. The square-grid array has 25 isotropic elements consisting of five rows of five elements each with 0.5λ-spacings. The triangular grid has 16 isotropic elements consisting of four rows of four elements each with $0.625\ \lambda$ between all adjacent elements. Sketch the array geometries. Plot the radiation pattern (linear, polar plot) in the $45°$ plane for both arrays and comment on the results.

8.8-7 Verify the directivity values in Example 8-8 using numerical pattern integration for $d = 0.5\lambda$.

8.8-8 *Project*. Research the topic and prepare a report on the antenna aspects of one of the following technical programs. Include calculations of as many antenna parameter values as you can. (a) The Square Kilometer Array (www.skatelescope.org). (b) The Long Wavelength Array (lwa.nrl.navy.mil). (c) The High Frequency Active Auroral Research Program (www.haarp.alaska.edu). (d) The Low Band Antenna (www.lofar.org). (e) The Widefield Array (www.mwatelescope.org). (f) A Wullenweber array.

8.9-1 Plot the full array pattern for $0°$ and $30°$ scan angles off broadside for the array of Fig. 8-30. Also plot the array factor for 30 degrees scan to see the grating lobe reduction effect of the element pattern.

8.9-2 The four-port Butler matrix beamformer of Fig. 8-32 is used to drive four isotropic elements. (a) Write the interelement excitation phases for all four beams. (b) Plot all four beam patterns on one polar pattern plot. (c) Give the values of the beam peak locations and crossover levels. (d) Calculate the beam peak directions for Beam 1 and 2 using beam scanning principles.

8.9-3 For the AN/PPS-18 radar example of Fig. 8-33, find the aperture efficiency based on a specified gain of 27 dB.

8.9-4 Derive an expression for the line length of a delay-line phase shifter to steer the beam of a linear array to angle θ_o starting with the required phase for element *n* of $\alpha_n = -(2\pi/c)f\, nd \cos\theta_o$.

8.9-5 *Project*. Investigate quantization lobes due to digital phase shifting.

8.9-6 *Project*. Investigate and report on how the Butler and Blass beamforming networks operate.

8.9-7 Compute the pattern of Beams 1 and 2 in Fig. 8-32 for the Butler matrix BFN and determine the value and angle where beams crossover.

8.10-1 Find the element spacing in wavelengths, d/λ, that gives a maximum element gain value of 6 dB for a planar array with $d = d_x = d_y$.

8.11-1 This problem examines frequency effects on beam steering with a linear phased array. (a) Derive (8-122). Now consider an array with the phase shifters set to steer the beam 30° off broadside at 2 GHz. Find the angle off broadside of the beam peak at (b) 1.6 GHz, and (c) 2.4 GHz.

8.11-2 A phased array is required to have a bandwidth of 40% and 45° scan to off broadside. Find which antenna types in Fig. 8-36 are possible choices as elements.

Chapter 9

Aperture Antennas

As summarized in Table 1-4, there are four types of antennas: electrically small, resonant, broadband, and aperture. The first three have been discussed earlier. The basic electrically small antennas are treated in Secs. 3.1 and 3.4. Resonant antennas are covered in Secs. 3.2, 3.3, and Chap. 5. The discussion of broadband antennas is in Chap. 7. In this chapter, we treat the fourth and final antenna type, the **aperture antenna**. Part of the structure of an aperture antenna is an aperture, or opening, through which electromagnetic waves flow. An aperture antenna operating as a receiver "collects" waves via the aperture. Analogies in acoustics are the megaphone and the parabolic microphone, which uses a parabolic reflector to focus sound waves on a microphone at the focal point. Also, the pupil of the human eye is an aperture for optical frequency electromagnetic waves. At radio frequencies, horns and reflectors are examples of aperture antennas; see Table 1-4. Aperture antennas are in common use at UHF frequencies and above. They are the antenna of choice in applications requiring very high gain. A distinguishing feature of large aperture antennas is the increase in gain with the operating frequency. The gain of an aperture antenna increases with the square of frequency if aperture efficiency is constant with frequency; see (4-27). Another feature is the nearly real-valued input impedance.

Since all receiving antennas act as collectors of waves, an effective aperture can be defined for every antenna; see Sec. 4.4.1. However, this chapter deals with antennas that have an obvious physical aperture. In the first section, general principles are developed for calculating the radiation patterns from any aperture antenna. Subsequent discussions focus on rectangular and circular aperture shapes. The properties of specific antennas such as horns and circular parabolic reflectors then follow naturally. As in preceding chapters, theoretical derivations lead to an accurate description of the antenna parameters, as well as design techniques. Both rigorous and approximate methods of gain calculation are also presented in this chapter.

9.1 RADIATION FROM APERTURES AND HUYGENS' PRINCIPLE

Although aperture antennas were not widely used until the World War II period, the basic concepts were available in 1690 when Huygens explained, in a simple way, the bending (or diffraction) of light waves around an object. This was accomplished by viewing each point of a wave front as a secondary source of spherical waves. The next wave front is the envelope of these secondary waves in the forward direction. Some 150 years after Huygens' contribution, Fresnel recognized that the phase shift between wave fronts is computed from the distance between wave fronts ΔL by the familiar relation $\beta \Delta L$. Fig. 9-1 shows how a plane wave and a spherical wave can be constructed from secondary waves; also see Figs. 16-1 and 16-2. The envelope of secondary waves forms the new wave front. Geometrical optics (ray tracing) predicts that light shining through a slit in a screen will have a lit region and a completely dark shadow region with a sharp boundary
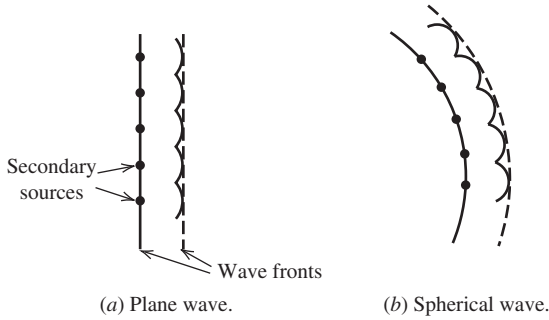
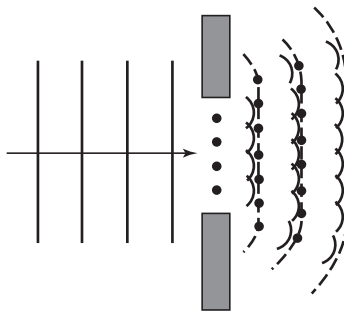Figure 9-1 Secondary waves used to construct successive wavefronts.

(*a*) Plane wave. (*b*) Spherical wave.

Figure 9-2 Plane wave incident on a slit in a screen. The edge diffraction leads to spreading of the radiation from the slit.

between them. Geometrical optics works well only for apertures that are very large relative to a wavelength. The secondary source concept shows that the secondary waves will spread out away from the aperture and there will be a smooth blending of the lit and shadow regions. This diffraction effect is illustrated in Fig. 9-2 for a slit in an opaque screen with a plane wave incident on it. The superposition (i.e., the sum) of the wavelets in the limit of an infinite number of secondary sources becomes a radiation integral over a continuous source that we are familiar with.

Huygens' principle evolved into a mathematical from referred to as the *equivalence principle* (or, *field equivalence principle*). The field equivalence principle replaces an aperture antenna with equivalent currents that produce radiation fields equivalent to those from the antenna. The equivalence principle is derived by observing that *a* solution to Maxwell's equations and the boundary conditions, which all electromagnetic problems must satisfy, is *the* solution. This follows from the *uniqueness theorem* in mathematics which states that a solution that satisfies a differential equation (e.g., Maxwell's equations) and the boundary conditions is unique. We now use this concept to set up equivalent current relations for use in analyzing aperture antennas.

In the original problem of Fig. 9-3*a*, the fields that satisfy Maxwell's equations in the region exterior to volume V and that satisfy boundary conditions along S are unique.[1] As long as the sources exterior to V and the boundary conditions along S are not changed, the solution ($\mathbf{E}$, $\mathbf{H}$) will not change. In the equivalent problem, the source exterior to V are not changed, since there are none. Also, the boundary conditions are not changed, as will now be explained. In the original problem, the fields along the boundary are $\mathbf{E}(S)$ and $\mathbf{H}(S)$. In the equivalent problem of Fig. 9-3*b*, the original sources (e.g., the antenna

---

[1] In this chapter, the uppercase symbols V and S will be used to denote volume and surface.
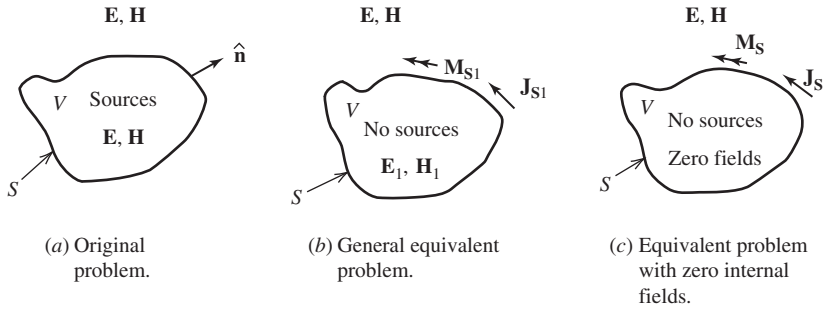
Figure 9-3 The equivalence principle.

structure) have been removed, altering the fields internal to S, denoted as $\mathbf{E}_1$ and $\mathbf{H}_1$. In order for the fields external to S to remain the same, equivalent currents must be introduced to satisfy the discontinuity of the fields across S. These equivalent currents are found from the boundary conditions of (2-22) and (2-23) as

$$\mathbf{J_{s1}} = \hat{\mathbf{n}} \times [\mathbf{H} - \mathbf{H}_1] \qquad \text{on } S \qquad (9\text{-}1a)$$

$$\mathbf{M_{s1}} = [\mathbf{E} - \mathbf{E}_1] \times \hat{\mathbf{n}} \qquad \text{on } S \qquad (9\text{-}1b)$$

Where $(\mathbf{E}_1, \mathbf{H}_1)$ and $(\mathbf{E}, \mathbf{H})$ are the fields internal and external to S; see Fig. 9-3b. These equivalent currents, which are obtained from only a knowledge of the tangential fields over S, can be used to find the fields external to S. However, the fields on S required to determine the equivalent currents are unknown. Also, we do not know how to find the external fields from the equivalent currents. We now address these difficulties, starting with the second one.

Since the internal fields $(\mathbf{E}_1, \mathbf{H}_1)$ are arbitrary, we choose them to be zero for simplicity; see Fig. 9-3c. Then (9-1) becomes

$$\mathbf{J_s} = \hat{\mathbf{n}} \times \mathbf{H}(S) \qquad (9\text{-}2a)$$

$$\mathbf{M_s} = \mathbf{E}(S) \times \hat{\mathbf{n}} \qquad (9\text{-}2b)$$

Where $\mathbf{E}(S)$ and $\mathbf{H}(S)$ are the fields over the surface S. This zero internal field formulation is referred to as *Love's equivalence principle*.

Since the fields inside S are zero in the equivalent problem of Fig. 9-3c, we are free to introduce materials inside S. If a perfect electric conductor is placed along S, $\mathbf{J_s}$ will vanish. The explanation is often given that the electric current is "shorted out" by the conductor. This leaves a magnetic current density $\mathbf{M_s}$ radiating in the presence of the electric conductor producing the external fields. Similarly, a perfect magnetic conductor can be introduced along S to eliminate $\mathbf{M_s}$, leaving only $\mathbf{J_s}$. Thus, we have two equivalent formulations. They are $\mathbf{M_s}$ in the presence of a perfect electric conductor over S and $\mathbf{J_s}$ in the presence of a perfect magnetic conductor over S. Both yield the correct fields external to S. However, these problems are difficult to solve as long as S is a general surface. Note that if the real antenna contains conducting portions, then the $\mathbf{J_s}$ equals the actual current density over that portion and the aperture portion contains both $\mathbf{J_s}$ and $\mathbf{M_s}$ (before any fictitious conductors are introduced).

If the surface S is large in terms of a wavelength and the curvature of S is small, image theory can be applied locally across the surface to solve for the currents operating in the presence of the introduced conductors; this is exploited in Chap. 16. However, since the selection of S is for convenience and we are interested in radiation problems, we can extend S to infinity. Since S must be a closed surface, S includes the infinite plane along $z = 0$ and closes at infinity to enclose sources in the region $z < 0$ of Fig. 9-4. In this case,

we can apply image theory with no approximation and easily solve many practical problems with planar apertures.

We can apply image theory to the planar surface S to simplify the solution procedure. To do this, we find the fields due to currents $\mathbf{J_s}$ and $\mathbf{M_s}$ on the $z = 0$ plane as shown in Fig. 9-4a; this will be followed by the application of image theory to reduce the solution formulation by one-half. Although not all antenna problems have a planar aperture that can be placed in the xy-plane, an equivalent planar aperture surface S can be set up. This will be fruitful if the tangential field over S can be obtained; more will be said about this later. First, we need to solve for fields in $z > 0$ due to the equivalent currents of Fig. 9-4a.

The fields ($\mathbf{E}, \mathbf{H}$) in the region $z > 0$, in general, are found by first evaluating $\mathbf{A}$ using (2-61) and finding $\mathbf{E}$ and $\mathbf{H}$ from (2-46) and (2-36). In this case, the equivalent currents $\mathbf{J_s}$ and $\mathbf{M_s}$ will yield the exact fields everywhere in $z > 0$. However, we restrict our solution to the far-field region appropriate to antenna problems. Then we can use the much simpler procedure of Sec. 2.4.4 that we have used many times to solve radiation problems. Now we slightly recast the formulation for the case of planar surface current densities in the xy-plane. First, the magnetic vector potential is found from the form of (2-101) appropriate to the geometry of Fig. 9-4a:

$$\mathbf{A} = \mu \frac{e^{-j\beta r}}{4\pi r} \iint_S \mathbf{J_S}(\mathbf{r}')e^{j\beta \hat{\mathbf{r}} \cdot \mathbf{r}'} dS' \tag{9-3}$$
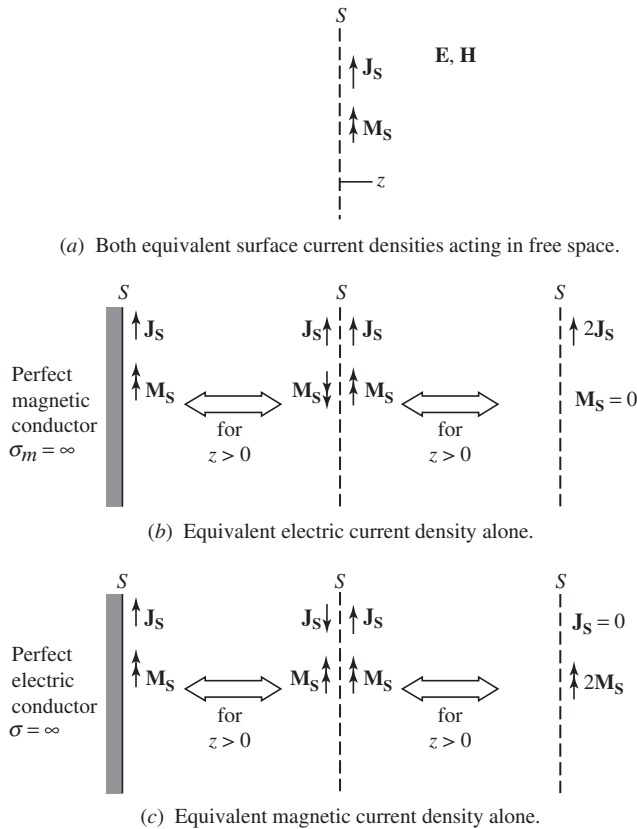

(a) Both equivalent surface current densities acting in free space.

(b) Equivalent electric current density alone.

(c) Equivalent magnetic current density alone.

Figure 9-4 Equivalent current configurations for a planar aperture surface. The antenna located in $z < 0$ has been removed and three different equivalent current sets introduced as shown.

The far-zone electric field from (2-105) is

$$\mathbf{E_A} = -j\omega(A_\theta\hat{\boldsymbol{\theta}} + A_\phi\hat{\boldsymbol{\phi}}) \tag{9-4}$$

The subscript A indicates that this field arises from the magnetic vector potential $\mathbf{A}$.

The electric vector potential $\mathbf{F}$ associated with the magnetic current density is found using the duality principle introduced in Sec. 3.4.1[2]:

$$\mathbf{F} = \varepsilon\frac{e^{-j\beta r}}{4\pi r}\iint\limits_{S}\mathbf{M_s}(\mathbf{r}')e^{j\beta\hat{\mathbf{r}}\cdot\mathbf{r}'}dS' \tag{9-5}$$

The far-zone magnetic field arising from $\mathbf{F}$ is the dual of (9-4):

$$\mathbf{H_F} = -j\omega(F_\theta\hat{\boldsymbol{\theta}} + F_\phi\hat{\boldsymbol{\phi}}) \tag{9-6}$$

Since the solution is in the far field, the electric field associated with $\mathbf{H_F}$ can be found from the TEM relationship of (2-107) as

$$\mathbf{E_F} = \eta\mathbf{H_F}\times\hat{\mathbf{r}} = -j\omega\eta(F_\phi\hat{\boldsymbol{\theta}} - F_\theta\hat{\boldsymbol{\phi}}) \tag{9-7}$$

The total electric field is then found by summing the contributions from each current:

$$\mathbf{E} = \mathbf{E_A} + \mathbf{E_F} = -j\omega[(A_\theta + \eta F_\phi)\hat{\boldsymbol{\theta}} + (A_\phi - \eta F_\theta)\hat{\boldsymbol{\phi}}] \tag{9-8}$$

The equivalent systems of Fig. 9-4a involves both the electric and magnetic current densities. Computations can be reduced considerably if image theory is used so that we only have to deal with one of the currents. First, we introduce a perfect magnetic planar conductor along surface S. The image currents shown in Fig. 9-4b are obtained by the duality of images in a perfect magnetic ground plane; that is, a magnetic current parallel to the plane has an oppositely directed image and a parallel electric current has a similarly directed image. The fields for $z > 0$ are unchanged after removing the conducting plane and introducing the images, as shown in Fig. 9-4b. Since the currents and their images are adjacent to the plane S, we can add them vectorially to obtain the final equivalent system, which has a doubled electric surface current density and no magnetic surface current density. The radiation electric field for $z > 0$ is $2\mathbf{E_A}$. In similar fashion, a perfect electric ground plane can be introduced along S as shown in the leftmost part of Fig. 9-4c. Image theory renders the images shown; see Fig. 3-8. These images acting together yield a zero total electric surface current density and a magnetic surface current density of $2\mathbf{M_S}$. Then the radiation electric field for $z > 0$ is $2\mathbf{E_F}$.

We can now summarize the equivalence theorem in terms most suitable to radiation pattern calculations. First, an aperture plane is selected for the antenna, this is usually the physical aperture of the antenna but need not be. Coordinates are set up such that the aperture plane is the xy-plane and the $+z$-axis is the forward radiation direction. Then the radiation field for $z > 0$ are found by one of the three equivalent systems of Fig. 9-4 as follows:

    a. $\mathbf{J_S}$ and $\mathbf{M_S}$ on S (xy-plane)

$$\mathbf{E} = \mathbf{E_A} + \mathbf{E_F} \tag{9-9a}$$

    with (9-3) in (9-4) and (9-5) in (9-6) and (9-7)

---

[2] The symbol $\mathbf{F}$ for magnetic vector potential should not be confused with the normalized radiation pattern function $F(\theta, \phi)$.

b. $2\mathbf{J_S}$ on S

$$\mathbf{E} = 2\mathbf{E_A} \qquad (9\text{-}9b)$$

c. $2\mathbf{M_S}$ on S

$$\mathbf{E} = 2\mathbf{E_F} \qquad (9\text{-}9c)$$

The procedures for finding radiation from equivalent aperture plane currents are now clear. It remains then to focus on determining those currents, which are established using (9-2).

So far, no approximations have been introduced other than the usual far-field approximations. Indeed, if the exact fields $\mathbf{E}(S)$ and/or $\mathbf{H}(S)$ are used in any of the above three procedures, exact far-field results will be obtained in the half-space $z > 0$. However, such exact knowledge of the fields over the entire plane S is rarely available. Usually, at best it is possible to obtain only an approximate knowledge of the fields over a finite portion of the infinite aperture plane. One such approach is the popular *physical optics approximation,* in which it is assumed that the aperture fields $\mathbf{E}_a$ and $\mathbf{H}_a$ are those of the incident wave. It is usually assumed that these fields exist over only some finite portion S_a of the infinite plane S and the fields elsewhere over S are zero. In most cases, the aperture surface S_a coincides with the physical aperture of the antenna. These approximations improve as the dimensions of the aperture relative to a wavelength increase.

The three solution procedures will now be simplified. Suppose that aperture fields $\mathbf{E}_a$ and $\mathbf{H}_a$, which exist over and are tangent to some portion of S_a of the infinite plane S, are known (perhaps by employing the physical optics approximation). The equivalent surface current densities follow from (9-2) as

$$\mathbf{J_S} = \hat{\mathbf{n}} \times \mathbf{H}_a \qquad (9\text{-}10)$$

$$\mathbf{M_S} = \mathbf{E}_a \times \hat{\mathbf{n}} \qquad (9\text{-}11)$$

on S_a and zero elsewhere. Using these in (9-3) and (9-5) gives

$$\mathbf{A} = \mu \frac{e^{-j\beta r}}{4\pi r} \hat{\mathbf{n}} \times \iint\limits_{S_a} \mathbf{H}_a e^{j\beta \hat{\mathbf{r}} \cdot \mathbf{r}'} dS' \qquad (9\text{-}12)$$

$$\mathbf{F} = -\varepsilon \frac{e^{-j\beta r}}{4\pi r} \hat{\mathbf{n}} \times \iint\limits_{S_a} \mathbf{E}_a e^{j\beta \hat{\mathbf{r}} \cdot \mathbf{r}'} dS' \qquad (9\text{-}13)$$

The integral the above two equations is a two-dimensional Fourier transform. The two-dimensional Fourier transform of an aperture field plays an important role in radiation calculations for aperture antennas, in a way similar to the Fourier transform of the current distribution for line sources (see Chap. 5). We therefore make the following definitions for the integrals:

$$\mathbf{P} = \iint\limits_{S_a} \mathbf{E}_a e^{j\beta \hat{\mathbf{r}} \cdot \mathbf{r}'} dS' \qquad (9\text{-}14)$$

$$\mathbf{Q} = \iint\limits_{S_a} \mathbf{H}_a e^{j\beta \hat{\mathbf{r}} \cdot \mathbf{r}'} dS' \qquad (9\text{-}15)$$

The far-zone electric field based on both aperture fields can be written in a single expression often encountered in the literature. The total electric field in terms of the potentials, form (9-4) and (9-7), is

$$\mathbf{E} = -j\omega\mathbf{A} - j\omega\eta\mathbf{F} \times \hat{\mathbf{r}} \tag{9-16}$$

Where the r-component of the first term is to be neglected. Substituting in (9-12) and (9-13) and performing some manipulations yield

$$\mathbf{E} = -j\beta\frac{e^{-j\beta r}}{4\pi r}\hat{\mathbf{r}} \times \iint_{S_a} [\hat{\mathbf{n}} \times \mathbf{E}_a - \eta\,\hat{\mathbf{r}} \times (\hat{\mathbf{n}} \times \mathbf{H}_a)]e^{j\beta\hat{\mathbf{r}} \cdot \mathbf{r}'}dS' \tag{9-17}$$

This gives the full vector form of the radiated electric field from the aperture fields and is often called a vector *diffraction integral*. The term "diffraction" is used because the field found using (9-17) represents the superposition of all elements of the source distribution; this is in contrast to geometrical optics that traces rays from points on the antenna directly to observation points (see Sec. 16.1). The subsequent developments here are cast in terms of the Fourier transforms **P** and **Q**; this provides a more procedural, as well as instructive, approach.

The aperture surface S_a is in the xy-plane, so $\mathbf{r}' = x'\hat{\mathbf{x}} + y'\hat{\mathbf{y}}$. This with $\hat{\mathbf{r}}$ in spherical coordinates from (C-4) in (9-14) and (9-15) yield

$$P_x = \iint_{S_a} E_{ax}(x', y')e^{j\beta(x'\sin\theta\cos\phi + y'\sin\theta\sin\phi)}dx'dy' \tag{9-18a}$$

$$P_y = \iint_{S_a} E_{ay}(x', y')e^{j\beta(x'\sin\theta\cos\phi + y'\sin\theta\sin\phi)}dx'dy' \tag{9-18b}$$

$$Q_x = \iint_{S_a} H_{ax}(x', y')e^{j\beta(x'\sin\theta\cos\phi + y'\sin\theta\sin\phi)}dx'dy' \tag{9-19a}$$

$$Q_y = \iint_{S_a} H_{ay}(x', y')e^{j\beta(x'\sin\theta\cos\phi + y'\sin\theta\sin\phi)}dx'dy' \tag{9-19b}$$

Now, (9-12) and (9-13) together with $\hat{\mathbf{n}} = \hat{\mathbf{z}}$ reduce to

$$\mathbf{A} = \mu\frac{e^{-j\beta r}}{4\pi r}(-Q_y\hat{\mathbf{x}} + Q_x\hat{\mathbf{y}}) \tag{9-20}$$

$$\mathbf{F} = -\varepsilon\frac{e^{-j\beta r}}{4\pi r}(-P_y\hat{\mathbf{x}} + P_x\hat{\mathbf{y}}) \tag{9-21}$$

Expressing $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ in spherical coordinates as in (C-l) and (C-2) and retaining only the θ- and ϕ-components give

$$\mathbf{A} = \mu\frac{e^{-j\beta r}}{4\pi r}\left[\hat{\boldsymbol{\theta}}\cos\theta(Q_x\sin\phi - Q_y\cos\phi) + \hat{\boldsymbol{\phi}}(Q_x\cos\phi + Q_y\sin\phi)\right] \tag{9-22}$$

$$\mathbf{F} = -\varepsilon\frac{e^{-j\beta r}}{4\pi r}\left[\hat{\boldsymbol{\theta}}\cos\theta(P_x\sin\phi - P_y\cos\phi) + \hat{\boldsymbol{\phi}}(P_x\cos\phi + P_y\sin\phi)\right] \tag{9-23}$$

Using these in (9-8) yields the final radiation field components

(a) $\quad E_\theta = j\beta\dfrac{e^{-j\beta r}}{4\pi r}\left[P_x\cos\phi + P_y\sin\phi + \eta\cos\theta(Q_y\cos\phi - Q_x\sin\phi)\right] \tag{9-24a}$

$\quad E_\phi = j\beta\dfrac{e^{-j\beta r}}{4\pi r}\left[\cos\theta(P_y\cos\phi - P_x\sin\phi) - \eta(Q_y\sin\phi + Q_x\cos\phi)\right] \tag{9-24b}$

In a similar fashion, the other two equivalent systems reduce to

(b) $E_\theta = j\beta\eta \dfrac{e^{-j\beta r}}{2\pi r} \cos\theta (Q_y \cos\phi - Q_x \sin\phi)$ (9-25a)

$E_\phi = -j\beta\eta \dfrac{e^{-j\beta r}}{2\pi r} (Q_y \sin\phi + Q_x \cos\phi)$ (9-25b)

(c) $E_\theta = j\beta \dfrac{e^{-j\beta r}}{2\pi r} (P_x \cos\phi + P_y \sin\phi)$ (9-26a)

$E_\phi = j\beta \dfrac{e^{-j\beta r}}{2\pi r} \cos\theta (P_y \cos\phi - P_x \sin\phi)$ (9-26b)

If the exact aperture fields over the entire aperture plane are used, the three formulations of (9-24) to (9-26) each yield the same result. Use of the exact aperture fields leads to equal contributions arising from the electric and magnetic currents. [H.3: Collin and Zucker, Part 1, p. 73]. Therefore, the equivalent system using both current types, as in (9-24), gives zero total field for $z < 0$ because $\cos\theta$ is negative for $\pi/2 \le \theta \le \pi$, and the contributions cancel as guaranteed by the equivalence theorem. However, the single current systems of (9-25) and (9-26) do not yield zero fields for $z < 0$. This is an expected result since image theory was involved in the development of these, and identical fields are obtained only in the region $z > 0$.

The trigonometric functions appearing in (9-24) to (9-26) actually describe the projections of the aperture equivalent surface current densities onto the plane containing the far-field components (i.e., perpendicular to $\hat{\mathbf{r}}$). For aperture field expressions, the trigonometric functions that multiply the radiation integrals are often referred to as *obliquity factors*. The element factor $\sin\theta$ for line sources along the z-axis is an obliquity factor. For apertures that are several wavelengths in extent, the obliquity factors do not reduce the main beam and first few side lobes by a significant amount. Then the Fourier transform adequately describes the pattern, and the aperture antenna problem reduces to first finding the (scalar) far-field pattern from the Fourier transform of the aperture electric field magnitude. Polarization is determined from the component(s) of the aperture electric field tangent to a far-field sphere by projecting $\mathbf{E_A}$ on to the far-field sphere.

In practice, only approximate information about the aperture fields is available, such as obtained from the physical optics approximation. Then the three formulations give different results. The accuracy of the three results depends on the accuracy of the aperture fields, but the differences might not be significant. For apertures mounted in a conducting ground plane, the aperture plane (except for the aperture itself) is well modeled as an infinite, perfectly conducting plane. Then the magnetic current (aperture electric field) formulation of (9-26) is preferred since the aperture electric field, and thus magnetic current, is zero outside the aperture because of the boundary condition of zero tangential electric field on the conductor. For apertures in free space, the dual current formulation of (9-24) is used. This is usually accompanied by the assumption that the aperture fields are related as a transverse electromagnetic (TEM) wave:

$$\mathbf{H}_a = \frac{1}{\eta} \hat{\mathbf{z}} \times \mathbf{E}_a$$ (9-27)

This implies that

$$\mathbf{Q} = \frac{1}{\eta} \hat{\mathbf{z}} \times \mathbf{P} \qquad \text{or} \qquad Q_x = -\frac{P_y}{\eta}, \quad Q_y = \frac{P_x}{\eta}$$ (9-28)

This assumption is valid for moderate- to high-gain antennas and is often applied with success even to apertures that are only a few wavelengths in extent. Using (9-28) in (9-24) leads to

$$E_\theta = j\beta \frac{e^{-j\beta r}}{2\pi r} \frac{1 + \cos\theta}{2} \left[P_x \cos\phi + P_y \sin\phi \right] \tag{9-29a}$$

$$E_\phi = j\beta \frac{e^{-j\beta r}}{2\pi r} \frac{1 + \cos\theta}{2} \left[P_y \cos\phi - P_x \sin\phi \right] \tag{9-29b}$$

which is a simpler form than (9-24). The factors in brackets are identical to those in (9-26). The obliquity factor of $(1 + \cos\theta)/2$ differs only slightly from the $\cos\theta$ obliquity factor in (9-26) for small values of θ, where radiation is significant for high-gain antennas. Unlike (9-26), (9-29) remains valid over all space (i.e., $0 < \theta < 180°$) because image theory was not employed and we can take surface S to enclose the antenna, since equivalent currents are zero except over the finite aperture. However, accuracy is likely to degrade for directions far out from the main beam. In summary, (9-26) should be used for apertures in ground planes and (9-29) should be used for aperture antennas in free space.

EXAMPLE 9-1   *Slit in an Infinite Conducting Plane*

The aperture antenna calculation procedures and the physical optics approximation can be illustrated rather simply for a plane wave normally incident on a slit in an infinite perfectly conducting plane as shown in Fig. 9-5. This is the same problem as in Fig. 5-7a, except for a coordinate system change. The physical optics approximation leads us to assume that the incident field $\mathbf{E}^i = \hat{\mathbf{y}} E_o e^{-j\beta r}$ associated with the plane wave propagating in the $+z$-direction renders the field over the physical aperture, so

$$\mathbf{E}_a = \begin{cases} \hat{\mathbf{y}} E_0 & |y| \le \dfrac{L}{2}, \qquad z = 0 \\[2mm] 0 & \text{elsewhere} \end{cases} \tag{9-30}$$

The magnetic current formulation is appropriate in this case because the aperture electric field is zero over the perfectly conducting portion of the aperture surface. This is essentially a one-dimensional problem because the aperture field is uniform in the x-direction; then the radiation fields will not change with position along the x-direction. We are thus concerned only with the yz-plane ($\phi = 90°$), and since the aperture field is only y-directed, (9-18) reduces to
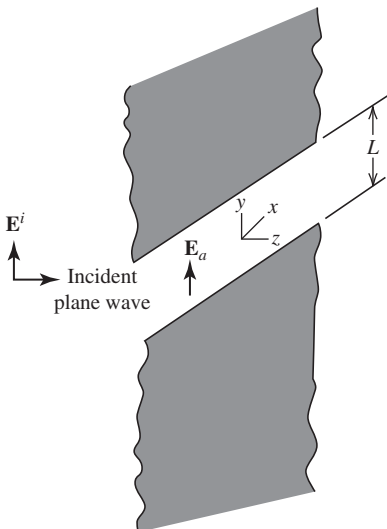

Figure 9-5 Plane wave incident on a slit in an infinite conducting plane. The slit is infinite in the x-direction and is L wide in the y-direction.

$$\mathbf{P} = \hat{\mathbf{y}}P_y = \hat{\mathbf{y}} \int_{-L/2}^{L/2} E_o e^{j\beta y' \sin\theta} dy' = \hat{\mathbf{y}}E_o L \, \frac{\sin[(\beta L/2)\sin\theta]}{(\beta L/2)\sin\theta} \qquad (9\text{-}31)$$

This result is similar to the line-source pattern of (5-2) except for the geometry difference of the z-axis now being normal to the current.

When normalized, the magnitude of this expression renders the following radiation pattern:

$$F(\theta) = \frac{\sin[(\beta L/2)\sin\theta]}{(\beta L/2)\sin\theta} \qquad (9\text{-}32)$$

The polarization of the far field (electric field) is the tangent of the aperture electric field direction $\hat{\mathbf{y}}$ onto the far-field sphere (in the $\phi = 90°$ observation plane), which is the θ-component and there is no ϕ-component. Notice that (9-32) is nonzero at $\theta = 90°$; this is acceptable since E_θ can be normal to the conducting plane. A note of caution is in order for this example, which has an aperture that is infinite in one dimension. The problem is really two-dimensional rather than three-dimensional (equivalently, the aperture is one-dimensional rather than two-dimensional). Therefore, the complete electric field will *not* be given by (9-26). The spherical wave behavior of $e^{-j\beta r}/r$ (e.g., free-space Green's function) is replaced by the cylindrical wave behavior $e^{-j\beta r}/\sqrt{r}$. However, the one-dimensional Fourier transform as presented here yields the correct angular variation (pattern). The pattern based on this simple approach agrees well with that of more rigorous techniques, even for a slit of $L = 0.45\lambda$, which is not large electrically. [1]

9.2   RECTANGULAR APERTURES

There are several antennas that have a physical aperture which is rectangular in shape. For example, many horn antennas have rectangular apertures. Another example is a rectangular slot in a metallic source structure such as a waveguide. In this section, we present some general principles about rectangular apertures that have uniform and tapered excitations. In Sec. 9.4, these principles are applied to rectangular aperture horn antennas.

9.2.1   Uniform Rectangular Apertures

A general rectangular aperture is shown in Fig. 9-6. It is excited in an idealized fashion such that the aperture fields are confined to the L_x by L_y region. If the aperture fields are uniform in phase and amplitude across the physical aperture, it is referred to as a *uniform rectangular aperture*. Suppose the aperture electric field is y-polarized; then the uniform rectangular aperture electric field is

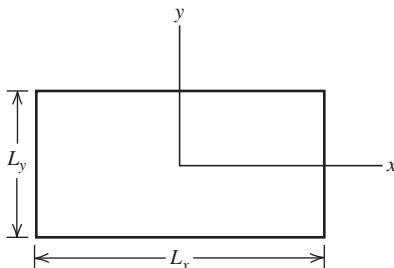$$\mathbf{E}_a = E_o \hat{\mathbf{y}}, \quad |x| \le \frac{L_x}{2}, \quad |y| \le \frac{L_y}{2} \qquad (9\text{-}33)$$

Figure 9-6 The rectangular aperture.

Then from (9-18b)

$$P_y = E_o \int_{-L_x/2}^{L_x/2} e^{j\beta x' \sin\theta \cos\phi} dx' \int_{-L_y/2}^{L_y/2} e^{j\beta y' \sin\theta \cos\phi} dy'$$

$$= E_o L_x L_y \frac{\sin[(\beta L_x/2)u]}{(\beta L_x/2)u} \frac{\sin[(\beta L_y/2)v]}{(\beta L_y/2)v}$$

(9-34)

where we have introduced the pattern variables

$$u = \sin\theta \cos\phi, \qquad v = \sin\theta \sin\phi$$

(9-35)

The complete radiation fields are found from (9-26) as

$$E_\theta = j\beta \frac{e^{-j\beta r}}{2\pi r} E_o L_x L_y \sin\phi \frac{\sin[(\beta L_x/2)u]}{(\beta L_x/2)u} \frac{\sin[(\beta L_y/2)v]}{(\beta L_y/2)v}$$

(9-36a)

$$E_\phi = j\beta \frac{e^{-j\beta r}}{2\pi r} E_o L_x L_y \cos\theta \cos\phi \frac{\sin[(\beta L_x/2)u]}{(\beta L_x/2)u} \frac{\sin[(\beta L_y/2)v]}{(\beta L_y/2)v}$$

(9-36b)

These fields are rather complicated functions of θ and ϕ, but fortunately they simplify in the principal planes. In the E-plane (yz-plane), $\phi = 90°$ (and 270°) and (9-36a) reduces to

$$E_\theta = j\beta \frac{e^{-j\beta r}}{2\pi r} E_o L_x L_y \frac{\sin[(\beta L_y/2)\sin\theta]}{(\beta L_y/2)\sin\theta} \qquad \text{E-plane}$$

(9-37)

In the H-plane (xz-plane), $\phi = 0°$ (and 180°) and (9-36b) becomes

$$E_\phi = j\beta \frac{e^{-j\beta r}}{2\pi r} E_o L_x L_y \cos\theta \frac{\sin[(\beta L_x/2)\sin\theta]}{(\beta L_x/2)\sin\theta} \qquad \text{H-plane}$$

(9-38)

Note that E_ϕ goes to zero at $\theta = 90°$ where it is tangent to the perfect conductor introduced in the equivalent magnetic current formulation. The normalized forms of these principal plane patterns are

$$F_H(\theta) = \cos\theta \frac{\sin[(\beta L_x/2)\sin\theta]}{(\beta L_x/2)\sin\theta}, \qquad \phi = 0°$$

(9-39)

$$F_E(\theta) = \frac{\sin[(\beta L_y/2)\sin\theta]}{(\beta L_y/2)\sin\theta}, \qquad \phi = 90°$$

(9-40)

For large apertures $(L_x, L_y \gg 1)$, the main beam is narrow, the $\cos\theta$ factor is negligible, and the principal plane patterns are both of the form $\sin(x)/x$ that we have encountered several times before, as, for example, with the uniform line source. By neglecting the obliquity factors in (9-36), the normalized pattern factor for the uniform rectangular aperture is

$$f(u, v) = \frac{\sin[(\beta L_x/2)u]}{(\beta L_x/2)u} \frac{\sin[(\beta L_y/2)v]}{(\beta L_y/2)v}$$

(9-41)

which is the normalized version of P_y in (9-34). The half-power beamwidths in the principal planes follow from the line source result in (5-12). In the xz- and yz-planes, the beamwidth expressions are

$$HP_x = 0.886 \frac{\lambda}{L_x} \text{ rad} = 50.8 \frac{\lambda}{L_x} \text{ deg} \qquad (9\text{-}42a)$$

$$HP_y = 0.886 \frac{\lambda}{L_y} \text{ rad} = 50.8 \frac{\lambda}{L_y} \text{ deg} \qquad (9\text{-}42b)$$

Finally, we derive an expression for the directivity of a uniform rectangular aperture. Such calculations are greatly simplified by using the variables u and v. The transformation from θ and ϕ to u and v given by (9-35) is essentially a collapsing of the spherical surface of unit radius onto a planar surface through the equator, giving a circular disk of unit radius. The u, v disk is analogous to the azimuthal map projection used in cartography to show, for example, the northern hemisphere on a planar map; the globe is projected with the North Pole at the center and the azimuth (radial) lines give true compass directions. The visible region in u and v corresponding to $\theta \leq \pi/2$ is

$$u^2 + v^2 = \sin^2 \theta \leq 1 \qquad (9\text{-}43)$$

which follows from (9-35).

The beam solid angle is found using

$$\Omega_A = \int_0^{2\pi} \int_0^{\pi/2} |F(\theta, \phi)|^2 d\Omega \qquad (9\text{-}44)$$

where only radiation for $\theta \leq \pi/2$ is considered. The beam solid angle can be evaluated by integrating over the entire visible region in terms of u and v. The projection of $d\Omega$ onto the u,v plane is given by $du\, dv = \cos\theta\, d\Omega$. From (9-43), it is seen that $\cos\theta = \sqrt{1 - u^2 - v^2}$. Therefore, $d\Omega = du\, dv/\sqrt{1-u^2-v^2}$ and (9-44) becomes

$$\Omega_A = \iint\limits_{u^2+v^2 \leq 1} |F(u,\, v)|^2 \frac{du\, dv}{\sqrt{1 - u^2 - v^2}} \qquad (9\text{-}45)$$

This is a general expression. For a large uniform phase aperture $(L_x$ and $L_y \gg \lambda)$, the radiation is concentrated in a narrow region about $u = v = 0$ $(\theta = 0)$. Then the square root in the denominator in (9-45) is approximately 1. Also, since the side lobes are very low, we can extend he limits to infinity without appreciably affecting the value of the integral.

Using these results and (9-41) for the uniform rectangular aperture in (9-45) yields

$$\Omega_A = \int_{-\infty}^{\infty} \frac{\sin^2[(\beta L_x/2)u]}{[(\beta L_x/2)u]^2} du \int_{-\infty}^{\infty} \frac{\sin^2[(\beta L_y/2)v]}{[(\beta L_y/2)v]^2} dv \qquad (9\text{-}46)$$

The following change of variables:

$$u' = \frac{\beta L_x}{2} u = \frac{\beta L_x}{2} \sin\theta \, \cos\phi \qquad (9\text{-}47a)$$

$$v' = \frac{\beta L_y}{2} v = \frac{\beta L_y}{2} \sin\theta \, \sin\phi \qquad (9\text{-}47b)$$

leads to

$$\Omega_A = \frac{2}{\beta L_x} \frac{2}{\beta L_y} \int_{-\infty}^{\infty} \frac{\sin^2 u'}{(u')^2} du' \int_{-\infty}^{\infty} \frac{\sin^2 v'}{(v')^2} dv' \qquad (9\text{-}48)$$

From (F-12) each integral above equals π, so

$$\Omega_A = \frac{4}{(2\pi/\lambda)^2 L_x L_y} \pi^2 = \frac{\lambda^2}{L_x L_y} \tag{9-49}$$

The directivity of the rectangular aperture with uniform amplitude and phase is then

$$D_u = \frac{4\pi}{\Omega_A} = \frac{4\pi}{\lambda^2} L_x L_y \tag{9-50}$$

From this expression, the physical area of the aperture can be identified as $A_p = L_x L_y$. Comparing this to $D = 4\pi A_{em}/\lambda^2$ from (4-23), we see that the maximum effective aperture A_{em} equals the physical aperture A_p for the uniform rectangular aperture. This is true for any shape aperture with uniform excitation. Also note that for ideal apertures, there are no ohmic losses (radiation efficiency $e_r = 1$), so gain equals directivity and $A_e = A_{em}$.

EXAMPLE 9-2   *A 20λ × 10λ Uniform Rectangular Aperture*

The complete pattern for a uniform rectangular aperture that has $L_x = 20\lambda$ and $L_y = 10\lambda$ is from (9-41):

$$f(u, v) = \frac{\sin(20\pi u)}{20\pi u} \frac{\sin(10\pi v)}{10\pi v} \tag{9-51}$$

The contour plot of this pattern is shown in Fig. 9-7. The principal plane patterns, which are profiles along the u and v axes of Fig. 9-7, are shown in Fig. 9-8. The aperture of Fig. 9-6 has a ratio $L_x/L_y = 2$ as in this example. Notice that the wide aperture dimension L_x leads to a narrow beamwidth in that direction (along the u-axis). The half-power beamwidth from (9-42) is $\mathrm{HP}_x = 0.0443$ rad $= 2.54°$ in the xz-plane, and $\mathrm{HP}_y = 0.0886$ rad $= 5.08°$ in the yz-plane. The directivity from (9-50) is $D = 4\pi(20\lambda)(10\lambda)/\lambda^2 = 2513 = 34$ dB.
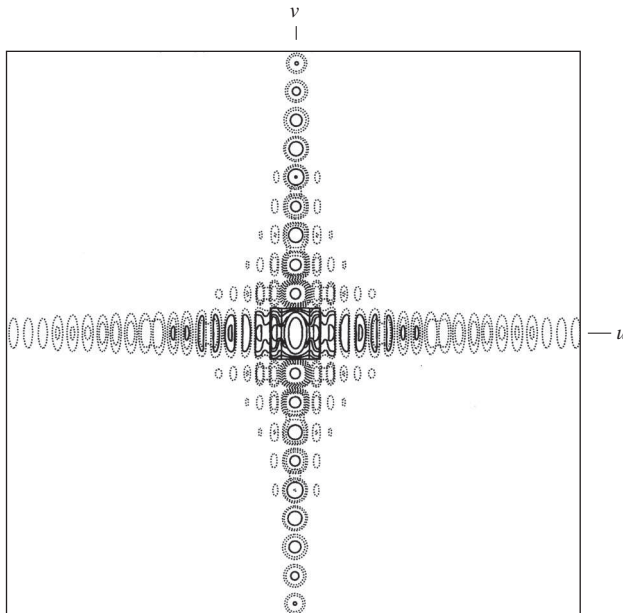

Figure 9-7 Contour plot of the pattern from a uniform amplitude, uniform phase rectangular aperture ($L_x = 20\lambda$, $L_y = 10\lambda$). The solid contour levels are 0, –5, –10, –15, –20, –25, –30 dB. The dashed contour levels are –35 and –40 dB. Principal plane profiles are shown in Fig. 9-8.

(a) The xz-plane pattern; $u = \sin\theta$.
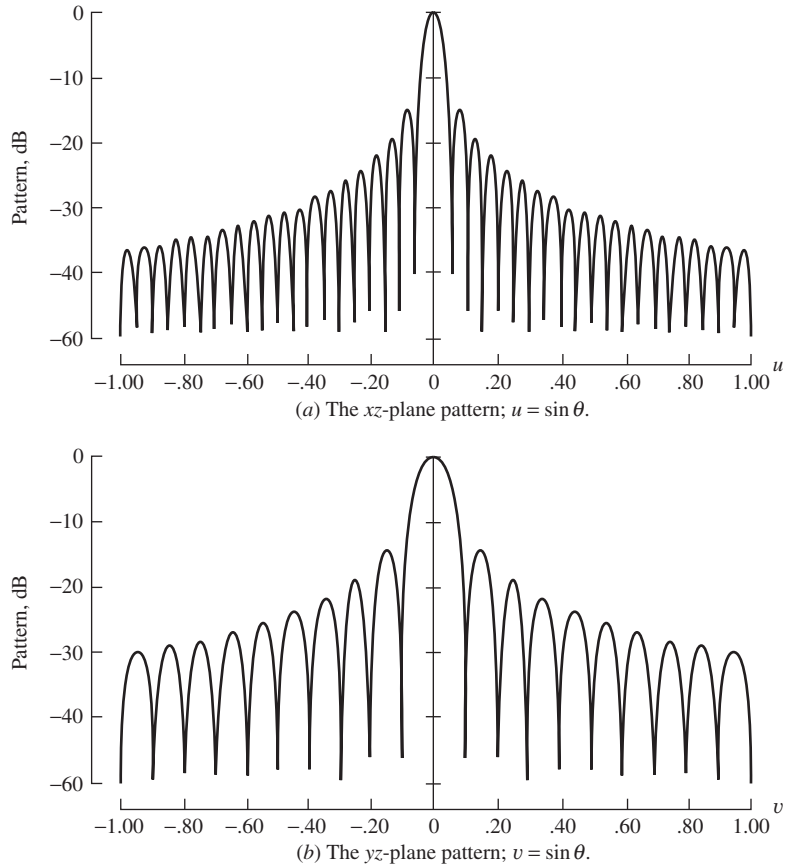

(b) The yz-plane pattern; $v = \sin\theta$.

Figure 9-8 Principal plane patterns for a uniform amplitude, uniform phase rectangular aperture ($L_x = 20\,\lambda$, $L_y = 10\,\lambda$). The complete pattern is shown in Fig. 9-7.

9.2.2 Tapered Rectangular Apertures

In the previous section, we saw that the uniform rectangular aperture has an effective aperture equal to its physical aperture. In other words, uniform illumination leads to the most efficient use of the aperture area. It will be shown in Sec. 9.3 that uniform excitation amplitude for an aperture gives the highest directivity possible for all constant phase excitations of that aperture. In the antenna design problem, high directivity is not the only parameter to be considered. Frequently, low side lobes are important. As we saw in Chap. 5, the side lobes can be reduced by tapering the excitation amplitude toward the edges of a line source. This is also true for two-dimensional apertures. In fact, many of the line source results can be directly applied to aperture problems.

To simplify our general discussion of rectangular aperture distributions, we omit the polarization of the aperture electric field, so that E_a can represent either the x- or y-component of the aperture field. Then (9-18) becomes

$$P = \iint\limits_{S_a} E_a(x', y')\, e^{j\beta u x'} e^{j\beta v y'}\, dx' dy' \tag{9-52}$$

Most practical aperture distributions are *separable* and can be expressed as a product of functions of each aperture variable alone:

$$E_a(x', y') = E_{a1}(x')E_{a2}(y') \tag{9-53}$$

Then (9-52) reduces to

$$P = \int_{-L_x/2}^{L_x/2} E_{a1}(x')e^{j\beta ux'}\,dx' \int_{-L_y/2}^{L_y/2} E_{a2}(y')e^{j\beta vy'}\,dy' \tag{9-54}$$

Each of these integrals is recognized as the pattern factor of a line source along the respective aperture directions. The normalized pattern factor for the rectangular aperture is then

$$f(u', v') = f_1(u')f_2(v') \tag{9-55}$$

where $f_1(u')$ and $f_2(v')$ arise from the first and second integrals in (9-54), which are essentially pattern factors of line source distributions along the x- and y-directions. Again, here we have neglected any obliquity factors. The uniform rectangular aperture result corresponding to (9-55) is (9-41). It is obtained directly from $\sin(u)/u$ of (5-7) by using u' of (9-47a) in place of u for $f_1(u')$ and v' of (9-47b) in place of u for $f_2(v')$. Note the different definition of u in Chap. 5 and this chapter.

Thus, the pattern expression for a rectangular aperture distribution that is separable, as in (9-53), is obtained by finding the patterns f_1 and f_2 corresponding to the distributions E_{a1} and E_{a2}, and then employing (9-55).

EXAMPLE 9-3   *The Open-Ended Rectangular Waveguide*

One of the smallest aperture antennas is the open-ended waveguide shown in Fig. 9-9. It requires no construction, since the antenna is the open end of a waveguide. It is often used as a probe, such as with a near-field measurement range (see Sec. 13.2), because of its compact size. When operated in the dominant TE_{10} mode, the aperture electric field is cosine-tapered in the x-direction with length $L_x = a$, similar to (5-21), and is uniform in the y-direction with length $L_y = b$. The radiation pattern then can be found from the corresponding line source
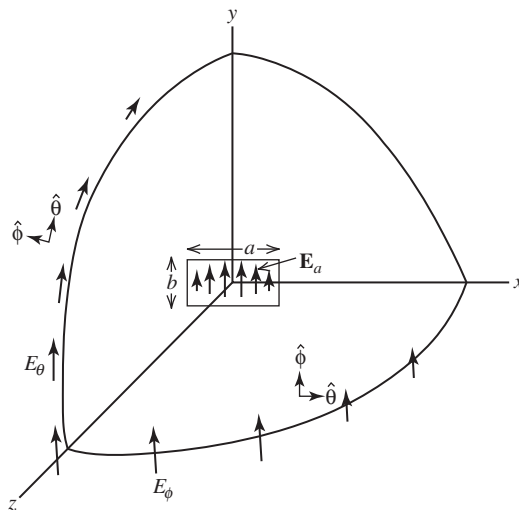

Figure 9-9 Geometry for an open-ended rectangular waveguide operating in the dominant TE_{10} mode as in Example 9-3. The aperture electric field E_a and radiated field components E_θ and E_ϕ are shown.

results using $f(u', v') = f_1(u')f_2(v')$, where $f_1(u')$ is obtained from (5-25) and $f_2(v')$ is obtained from (5-7):

$$f(u', v') = \frac{\cos u'}{1 - [(2/\pi)u']^2} \frac{\sin v'}{v'} = \frac{\cos[(\beta L_x/2)u]}{1 - [(2/\pi)(\beta L_x/2)u]^2} \frac{\sin[(\beta L_y/2)v]}{(\beta L_y/2)v} \tag{9-56}$$

The vertical linear polarization of this antenna is evident from Fig. 9-9, which shows the components of the electric field in the far field. The complete far-field component expressions are easily obtained from (9-56) using the equivalent current formulations of (9-24) or (9-26). If the open-ended waveguide is surrounded by a large ground plane, (9-56) in (9-26) yields the following principal plane pattern results by a process identical to the development presented for (9-39) and (9-40):

$$F_H(\theta) = \cos\theta \frac{\cos\left[\frac{\beta a}{2}\sin\theta\right]}{1 - \left[\frac{2}{\pi}\frac{\beta a}{2}\sin\theta\right]^2} \qquad \phi = 0° \tag{9-57a}$$

open-ended waveguide on a
ground plane, $0 < \theta < 90°$

$$F_E(\theta) = \frac{\sin\left[\frac{\beta b}{2}\sin\theta\right]}{\frac{\beta b}{2}\sin\theta} \qquad \phi = 90° \tag{9-57b}$$

The magnetic current formulation is chosen because the ground plane is well represented by the perfect conductor used in the image theory model of Fig. 9-4c. If the waveguide radiates into free space, the complete expression of (9-24) is more appropriate and yields

$$F_H(\theta) = \frac{1 + \cos\theta}{2} \frac{\cos\left[\frac{\beta a}{2}\sin\theta\right]}{1 - \left[\frac{2}{\pi}\frac{\beta a}{2}\sin\theta\right]^2} \qquad \phi = 0° \tag{9-58a}$$
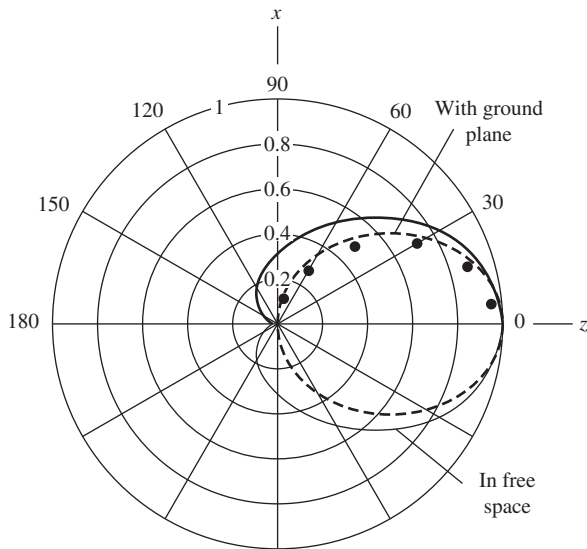
open-ended waveguide in
free space, $0 < \theta < 180°$

$$F_E(\theta) = \frac{1 + \cos\theta}{2} \frac{\sin\left[\frac{\beta b}{2}\sin\theta\right]}{\frac{\beta b}{2}\sin\theta} \qquad \phi = 90° \tag{9-58b}$$

Note the difference in obliquity factors between (9-57) and (9-58). For (9-57), the boundary conditions on the ground plane at $\theta = 90°$ are satisfied as explained in association with (9-38). For operation in free space, (9-58) is valid all the way to $\theta = 180°$ where the $(1 + \cos\theta)/2$ obliquity factor takes the pattern to zero. Fig. 9-10 shows the patterns in the E- and H-planes calculated using (9-57) and (9-58). Also shown are measured pattern data for a WR90 waveguide in free space operating at 9.32 GHz. [H.3: Silver, p. 345; 2] Agreement to the measured data is best in the E-plane for the free-space case of (9-58b), as expected. But in the H-plane the agreement is better for the ground plane case of (9-57a). This is due to the fact that the aperture theory used for the calculated patterns assumes an electrically large aperture, but the waveguide has dimensions that are not electrically large: $a = 2.286\,\text{cm} = 0.71\lambda$ and $b = 1.016\,\text{cm} = 0.31\lambda$. In addition, the fringe currents on the waveguide exterior are neglected, which will affect the H-plane more than the E-plane. Better results can be obtained by including the fringe currents on the waveguide walls, the phase constant β_g of the waveguide, and the reflection coefficient introduced by the discontinuity of the abrupt termination [2]; see Prob. 9.2-4. The directivity of the open-ended waveguide is discussed in Example 9-4.
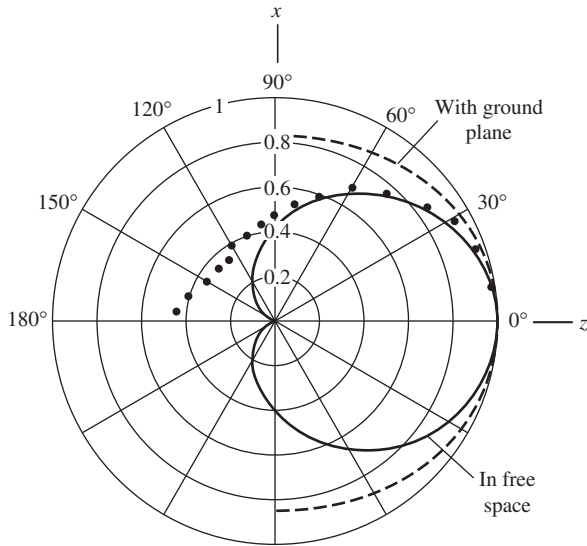
(a) The E-plane patterns.

(b) The H-plane patterns.

Figure 9-10 Radiation patterns of an open-ended *WR*-90 waveguide in free space operating at 9.32 GHz with the geometry of Fig. 9-9 and as described in Example 9-3. The computed patterns are for free space (solid curve) and with a ground plane in the *xy*-plane (dashed curve). Also shown are the measured patterns (dots).

9.3 TECHNIQUES FOR EVALUATING GAIN

Aperture antennas are often selected for use in applications requiring high gain. It is, therefore, important to be able to evaluate gain as accurately as possible. In this section, techniques are presented for evaluating gain based on pattern information and on aperture field information. In addition, simple formulas are presented that provide approximate gain values. These techniques apply to a wide variety of antenna types.

9.3.1 Directivity

Two useful forms for directivity from (2-144) and (2-140) are

$$D = \frac{4\pi}{\Omega_A} \qquad (9\text{-}59)$$

$$D = \frac{4\pi U_m}{P} \qquad (9\text{-}60)$$

where U_m is the maximum of the radiation intensity defined through

$$U(\theta, \phi) = \frac{1}{2\eta}\left[|E_\theta|^2 + |E_\phi|^2\right]r^2 = U_m|F(\theta, \phi)|^2 \qquad (9\text{-}61)$$

Ω_A is the beam solid angle:

$$\Omega_A = \int_0^{2\pi}\int_0^{\pi} F(\theta, \phi)|^2]\sin\theta\, d\theta\, d\phi \qquad (9\text{-}62)$$

The total radiated power P is evaluated by integrating (9-61) over all radiation space or from $P = U_m\Omega_A$. Accurate evaluation of directivity using (9-59) and (9-62) requires both a knowledge of the pattern over all angles θ, ϕ and integration of the pattern. If the pattern function is known, the integral in (9-62) can sometimes be evaluated analytically, but is usually found by numerical integration.

Pattern integration can be avoided when evaluating the directivity of aperture antennas. This approach is based on determining the radiated power in the aperture plane where it is easier to integrate. A knowledge of the aperture fields is required, of course. The formulation is simplified by assuming that the tangential aperture electric and magnetic fields are related as a TEM wave; see (9-27). This is justified by the good match to free space (e.g., low VSWR) that most aperture antennas exhibit, indicating real power flow as with a TEM wave. Using (9-27) in the general radiation field expression of (9-24) with (9-61) gives

$$U(\theta, \phi) = \frac{\beta^2}{32\pi^2\eta}\,(1+\cos\theta)^2\left[|P_x|^2 + |P_y|^2\right] \qquad (9\text{-}63)$$

The maximum value of this function, which corresponds to the main beam peak from (9-14), is

$$U_m = \frac{\beta^2}{8\pi^2\eta}\left|\iint_{S_a} \mathbf{E}_a\, dS'\right|^2 \qquad (9\text{-}64)$$

since $\hat{\mathbf{r}} \cdot \mathbf{r}' = 0$ in the broadside case ($\theta = 0$) because $\hat{\mathbf{r}} = \hat{\mathbf{z}}$ and $\mathbf{r}'$ is in the xy-plane.

Integration of (9-63) to obtain P is, in general, rather difficult. This can be avoided by observing that the total power reaching the far field must have passed through the aperture. Within the validity of (9-27), the power density in the aperture is $|\mathbf{E}_a|^2/2\eta$ and we can determine the radiated power from

$$P = \frac{1}{2\eta}\iint_{S_a} |\mathbf{E}_a|^2 dS' \qquad (9\text{-}65)$$

Substituting (9-64) and (9-65) in (9-60) gives a simplified, but powerful, directivity relationship:

$$D = \frac{4\pi}{\lambda^2} \frac{\left| \iint\limits_{S_a} \mathbf{E}_a \, dS' \right|^2}{\iint\limits_{S_a} |\mathbf{E}_a|^2 dS'} \tag{9-66}$$

This formula assumes the following: The pattern peak is directed broadside to the aperture, the aperture is large relative to a wavelength, and the aperture fields nearly form a plane wave. It turns out that the latter two conditions need not be strictly satisfied for good results to be obtained. Note the similarity of (9-66) to (8-85) for a half-wavelength spaced linear array.

If the aperture distribution is of uniform amplitude ($E_a = E_o$), then (9-66) reduces to

$$D_u = \frac{4\pi}{\lambda^2} A_p \tag{9-67}$$

where A_p is the physical aperture area. This was shown to be true for the rectangular aperture by direct evaluation; see (9-50). Further, (9-67) is a general result and implies that *the directivity of a uniform amplitude aperture is the highest obtainable from a uniform phase aperture*. This is true because the maximum of (9-66) occurs for a uniform illumination, which yields (9-67); see Prob. 9.3-2.

EXAMPLE 9-4   *Directivity of an Open-Ended Rectangular Waveguide*

To illustrate the aperture field integration method of determining directivity, we return to the open-ended waveguide operating in the TE_{10} mode as described in Example 9-3 and illustrated in Fig. 9-9. The aperture field distribution is

$$\mathbf{E}_a = \hat{\mathbf{y}} E_o \cos \frac{\pi x'}{a}, \qquad -\frac{a}{2} \le x' \le \frac{a}{2}, \qquad -\frac{b}{2} \le y' \le \frac{b}{2} \tag{9-68}$$

where the waveguide (and, thus, the aperture) has wide and narrow dimensions of a and b. Then

$$\left| \iint\limits_{S_a} \mathbf{E}_a \, ds' \right|^2 = \left(E_o \int_{-a/2}^{a/2} \cos \frac{\pi x'}{a} dx' \int_{-b/2}^{b/2} dy' \right)^2 = E_o^2 \left(\frac{2a}{\pi} \right)^2 b^2 \tag{9-69}$$

and

$$\iint\limits_{S_a} |\mathbf{E}_a|^2 \, dS' = E_o^2 \int_{-a/2}^{a/2} \cos^2 \frac{\pi x'}{a} dx' \int_{-b/2}^{b/2} dy' = E_o^2 \frac{a}{2} b \tag{9-70}$$

Substituting these into (9-66) gives

$$D = \frac{4\pi}{\lambda^2} \left(\frac{8}{\pi^2} ab \right) = \frac{4\pi}{\lambda^2} (0.81)ab \tag{9-71}$$

This directivity is reduced by a factor of 0.81 (the aperture taper efficiency, ε_t) from that of same aperture when uniformly illuminated as in (9-67). However, in order for this formula to provide a good approximation the aperture should be electrically large with dimensions of a few wavelengths. A more all-purpose approach is (9-59) that should be used for small apertures. This begins by finding the total normalized vector pattern by interpolating between the principal plane patterns:

$$\mathbf{F}(\theta, \phi) = \sin\phi\, F_E(\theta, \phi)\hat{\boldsymbol{\theta}} + \cos\phi\, F_H(\theta, \phi)\hat{\boldsymbol{\phi}} \tag{9-72}$$

Then the total pattern magnitude squared is

$$|F(\theta, \phi)|^2 = \mathbf{F}(\theta, \phi) \cdot \mathbf{F}(\theta, \phi) = \sin^2\phi\, F_E^2(\theta, \phi) + \cos^2\phi\, F_H^2(\theta, \phi) \tag{9-73}$$

The directivity is found from (9-59) and (9-62) with (9-73), (9-58b) for F_E and (9-57a) for F_H because of the better fit to the measured pattern. This process yields a directivity of 6 dB, which is close to the measured gain value at 9.32 GHz of 6.2 dB. [2] The directivity calculated from (9-71) is 3.6 dB, which is inaccurate because of the small aperture size.

Most rectangular aperture distributions are separable, that is,

$$E_a(x, y) = E_{a1}(x)E_{a2}(y) \tag{9-73}$$

It can then be shown (see Prob. 9.3-15) that the directivity is also separable:

$$D = \pi D_x D_y \cos\theta_o \tag{9-74}$$

where

> D_x, D_y = directivity of a line source with a relative current distribution of $E_{a1}(x)$, $E_{a2}(y)$
> θ_o = main beam pointing direction relative to broadside

The $\cos\theta_o$ factor represents the projection of the aperture physical area onto the plane normal to the main beam maximum direction θ_o. This approximation is valid if the beam is not scanned within several beamwidths of endfire. The directivity of a uniform rectangular aperture for broadside ($\theta_o = 0$) can be expressed as follows using (5-19):

$$D_u = \pi\frac{2L_x}{\lambda}\frac{2L}{\lambda} = \frac{4\pi}{\lambda^2}L_x L_y \tag{9-75}$$

which is (9-50).

9.3.2 Gain and Efficiencies

Gain equals directivity reduced by the amount of power lost on the antenna structure; see Secs. 2.5 and 4.4.1 for previous discussions of gain. This is expressed using radiation efficiency front (2-155) as

$$G = e_r D \tag{9-76}$$

Another form follows from (9-60) with input power in place of radiated power since $P = e_r P_{in}$; also see (2-152):

$$G = \frac{4\pi U_m}{P_{in}} \tag{9-77}$$

This form is often used when evaluating antennas by numerical computation.

Since the directivity of an aperture antenna is directly proportional to its physical aperture area A_p, gain will be also:

$$\boxed{G = \frac{4\pi}{\lambda^2}A_e = \frac{4\pi}{\lambda^2}\varepsilon_{ap}A_p = \varepsilon_{ap}D_u} \tag{9-78}$$

where (4-27), (4-29), and (9-67) were used. A_e is the effective aperture and can be calculated through this equation for any antenna, including arrays. From this, we see that

$$\boxed{A_e = \varepsilon_{\mathrm{ap}} A_p} \qquad 0 \leq \varepsilon_{\mathrm{ap}} \leq 1 \tag{9-79}$$

Aperture efficiency $\varepsilon_{\mathrm{ap}}$ is a measure of how efficiently the antenna physical area is utilized. If $\varepsilon_{\mathrm{ap}}$ is known, it is a simple matter to calculate the gain of an aperture antenna of aperture area A_p using (9-78).

There are several contributions to the overall aperture efficiency. The following form shows the factors separately and is appropriate for general use:

$$\boxed{\varepsilon_{\mathrm{ap}} = e_r \varepsilon_t \varepsilon_s \varepsilon_a} \tag{9-80}$$

All these factors have values from zero to unity. We discussed radiation efficiency e_r in Sec. 2.5; it represents all forms of dissipation on the antenna structure such as conductor losses. In most aperture antennas, these losses are very low, so $e_r \approx 1$ and

$$G \approx D \qquad \textit{most aperture antennas} \tag{9-81}$$

This may not hold if one of the following situations applies: The antenna size is less than a wavelength, a lossy transmission line or device is considered to be part of the antenna, or lossy materials are an integral part of the antenna such as a dielectric lens.

Aperture taper efficiency ε_t represents gain loss strictly due to the aperture amplitude distribution. Often, the amplitude is tapered from the center to the edges of an aperture intentionally to reduce sidelobes. ε_t is the ratio of directivity computed with only the amplitude taper present, D_t, to the directivity of the same aperture uniformly illuminated, D_u:

$$\varepsilon_t = \frac{D_t}{D_u} \quad \text{or} \quad D_t = \varepsilon_t D_u \tag{9-82}$$

Examples for line sources are given in Table 5-2. Also, in Example 9-4 we found $\varepsilon_t = 0.81$ for an open-ended waveguide.

Antennas that have a secondary radiating aperture illuminated by a primary (feed) antenna, such as a parabolic reflector, experience spillover loss due to power from the feed missing the radiating aperture. This *spillover efficiency* ε_s and aperture taper efficiency are the main sources of gain loss in most aperture antennas. The product $\varepsilon_t \varepsilon_s$ is called the *illumination efficiency* ε_i.

The remaining factor in (9-80), ε_a, is *achievement efficiency* and can include many subefficiencies. More subefficiencies will be treated with reflector antennas in Sec. 9.5, but the following two are usually dominant:

$$\varepsilon_a \approx \varepsilon_{\mathrm{cr}} \varepsilon_{\mathrm{ph}} \tag{9-83}$$

Cross-polarization efficiency, $\varepsilon_{\mathrm{cr}}$, represents loss due to power being radiated in a polarization state orthogonal to the intended polarization. *Phase efficiency*, $\varepsilon_{\mathrm{ph}}$, represents loss due to nonuniform phase across the aperture.

Any of the efficiency factors can be expressed as a gain factor in decibels as

$$\varepsilon_n(\mathrm{dB}) = 10 \log \varepsilon_n \tag{9-84}$$

Gain "loss" is negative of this. For example, the aperture taper efficiency for Example 9-4 is $\varepsilon_t = 0.81$, so $\varepsilon_t(\mathrm{dB}) = -0.91$ dB and the gain loss is $+0.91$ dB. This is the only source of loss in this case. In general, (9-78) and (9-80) can be written in dB form as

$$G(\mathrm{dB}) = 10 \log \left(\frac{4\pi}{\lambda^2} A_p \right) + e_r(\mathrm{dB}) + \varepsilon_t(\mathrm{dB}) + \varepsilon_s(\mathrm{dB}) + \varepsilon_a(\mathrm{dB}) \tag{9-85}$$

Recall that polarization mismatch factor p and impedance mismatch factor q are not included in aperture efficiency nor gain, but they play a role similar to the efficiency factors (as discussed in Sec. 4.4).

9.3.3 Simple Directivity Formulas

It is often necessary to estimate the gain of an antenna, especially in system calculations. If the gain cannot be measured, simple gain equations can be used. The most direct and simplest approach is to use (9-78). The operating wavelength and physical aperture area are easily obtained. Aperture efficiency can sometimes be determined by using a theoretical model, as will be discussed for horns and reflectors later in this chapter. In many cases, it can be estimated. In general ε_{ap} ranges from 30% to 80% with 50% being a good overall value. Optimum gain pyramidal horns have an aperture efficiency near 50%. Parabolic reflector antennas have an efficiency of 55% or greater. Gain can be found by estimating the aperture efficiency. For example, a 30-dB gain antenna with an actual efficiency of 55% will have a gain error of 0.38 dB when an estimated efficiency of 60% is used.

It is very useful to have an approximate directivity expression that depends only on the half-power beamwidths of the principal plane patterns. This is expected to yield good results since we know that directivity varies inversely with the beam solid angle ($D = 4\pi/\Omega_A$) and the beam solid angle is primarily controlled by the main beam. Thus, we expect to find that $D \propto (\text{HP}_E\text{HP}_H)^{-1}$, where the product of the principal plane beamwidths approximates the beam solid angle. We now derive such relations.

The topic of directivity and gain estimation was introduced in Sec. 4.5.3, and the full development along with references is found in [3]. The directivity of a rectangular aperture with a separable distribution given by (9-74) for broadside operation ($\theta_o = 0$) is

$$D = \pi D_x D_y \tag{9-86}$$

where D_x and D_y are the directivities of a line source (or linear array) associated with the x and y aperture distribution variations. But we know from studying several linear current distributions that these directivities are related to the aperture extents as

$$D_x = c_x \frac{2L_x}{\lambda}, \qquad D_y = c_y \frac{2L_y}{\lambda} \tag{9-87}$$

where directivity factors c_x and c_y are constants that vary slightly with the distributions $E_{a1}(x)$ and $E_{a2}(y)$. For uniform line sources, $c_x = c_y = 1$; see (5-19). Using (9-87) in (9-86) and rearranging give

$$D = \pi c_x \frac{2L_x}{\lambda} c_y \frac{2L_y}{\lambda} = \frac{4\pi c_x c_y k_x k_y}{\left(k_x \frac{\lambda}{L_x}\right)\left(k_y \frac{\lambda}{L_y}\right)} = \frac{4\pi c_x c_y k_x k_y}{\text{HP}_x \text{HP}_y} \tag{9-88}$$

The beamwidth factors k_x and k_y are constants associated with the following beamwidth formulas that we have used frequently (see Table 5-2):

$$\text{HP}_x = k_x \frac{\lambda}{L_x}, \qquad \text{HP}_y = k_y \frac{\lambda}{L_y} \tag{9-89}$$

For uniform line sources, $k_x = k_y = 0.886$. The numerator in (9-88) is the *directivity-beamwidth product*:

$$\text{DB} = 4\pi c_x c_y k_x k_y \cdot \left(\frac{180}{\pi}\right)^2 \quad [\text{deg}^2] \tag{9-90}$$

The general form of the directivity estimation formula based only on half-power beamwidths in degrees is

$$D = \frac{DB}{HP_{E^o} HP_{H^o}} \qquad (9\text{-}91)$$

It is similar to the gain-bandwidth product that is commonly used to characterize circuit devices. It remains relatively constant under a variety of operating circumstances because as the amplitude is tapered, constants c_x, c_y decrease due to aperture taper efficiency reduction, but the constants k_x, k_y increase due to beam broadening and nearly cancel the decrease in c_x, c_y.

If we could determine the value of DB, (9-91) would be our desired simple expression for directivity. For uniform distributions in x and y:

$$DB_u = 4\pi(1)(1)(0.886)(0.886) = 9.86 \text{ rad}^2 = 32{,}383 \text{ deg}^2 \qquad (9\text{-}92)$$

Then (9-88) becomes

$$D_{u_{\text{rect}}} = \frac{32{,}383}{HP_{E^o} HP_{H^o}} \quad \begin{array}{l} \textit{uniform rectangular aperture} \\ \textit{directivity approximation} \end{array} \qquad (9\text{-}93)$$

where HP_{E^o} and HP_{H^o} are the principal plane beamwidths in degrees. Although this is based on a uniform rectangular aperture, it produces accurate results for any pattern with a moderately narrow major lobe and with minor lobes present. This relation can be used for scanned beams if the beamwidths are those of the scanned beam. For uniform circular apertures,

$$\begin{aligned} DB_{u_{\text{cir}}} = D \cdot HP^2 &= \frac{4\pi}{\lambda^2} \pi a^2 \left(1.02 \frac{\lambda}{2a} \right)^2 \left(\frac{180}{\pi} \right)^2 \\ &= 33{,}709 \text{ deg}^2 \qquad \begin{array}{l} \textit{uniform circular aperture} \\ \textit{directivity approximation} \end{array} \end{aligned} \qquad (9\text{-}94)$$

where (9-171) and (9-172) were used. This is very close to the directivity-beamwidth product of 32,383 for uniform rectangular apertures. As the amplitude taper from the center to the perimeter of a circular aperture is changed, DB varies from 33,709 to about 39,000 deg^2 [4]. For a rectangular aperture with a cosine amplitude taper in the H-plane and uniform phase, as found in the open-ended waveguide of Example 9-4, DB is 35,231 $\text{deg}.^2$ The foregoing aperture distributions are included in Table 9-1, which lists the DB values for several cases.

Some simple antennas and theoretical patterns have no minor lobes. The Gaussian and $\cos^q \theta$ patterns are examples; see Probs. 9.3-18 and 9.3-19. The formula of Case 1 in Table 9-1 for no side lobe patterns with $DB = 41{,}253$ applies to both omnidirectional and unidirectional patterns. As mentioned in Sec. 4.5.3, the formula with 41,253 gives 1.47 dB for a half-wave dipole, which is close to the correct value of 1.63 dB. For the unidirectional $\cos^q \theta$ pattern (zero for $\theta > 90°$) the directivity is exactly $D = 2(2q + 1)$ which is 13.4 dB for $q = 5$. The half-power beamwidth is 42.2° in both planes for this pattern and the directivity is $41{,}2531(42.2)^2 = 13.6$ dB, which is good agreement with the exact value.

In practice, antennas produce radiation patterns with significant power content in side lobes that do not progressively decrease away from the main beam. In addition, the main beam rolls off more slowly than idealized patterns such as the $\sin u/u$ pattern for a uniform line source. Thus, the solid angle of the main beam is larger than that predicted by the product of the principal plane beamwidths. This together with the increased solid angle contribution from the side lobes reduces the directivity compared to models based

Table 9-1 Directivity-Beamwidth Product for Use in Directivity Estimation: $D \approx \frac{\text{DB}}{\text{HP}_{E^o}\text{HP}_{H^o}}$

Case	Pattern or Aperture Distribution Type	DB [deg^2]
1	No side lobe pattern, rectangular beam. See (4-56).	41,253
2	Uniform rectangular aperture. See (9-92).	32,383
3	Cosine-uniform rectangular aperture such as an open-ended waveguide. See Example 9-3 and Prob. 9.3-20a.	35,231
4	Gaussian pattern. See Prob. 9.3-18	36,407
5	Uniform circular aperture. See Prob. 9.3-20b and Sec. 9.5.1.	33,709
6	Circular aperture with parabolic-on-a -12 dB-pedestal distribution. See Prob. 9.3-20c and Table 9-2.	38,933
7	General use for practical patterns with side lobes and null fill. See Sec. 4.5.3 and (9-95).	26,000

on ideal patterns. A variation of (9-92) that yields accurate gain values for antennas used in practice is

$$G = \frac{26{,}000}{\text{HP}_{E^o}\text{HP}_{H^o}} \tag{9-95}$$

This is also (4-57). Gain is used here instead of directivity because it applies to commercial antennas for which vendors quote measured gain values. It is assumed that ohmic losses are very small so that $G \approx D$.

The simple formulas for finding directivity or gain are based only on half-power beamwidths and thus accurate results cannot be counted on. There is not enough information contained in the half-power beamwidths. The most important aspect of the beamwidth-based directivity formulas is to select the appropriate formula. The intended application for each DB value are included in Table 9-1. The following examples illustrate the power of the general formula of (9-95).

EXAMPLE 9-5 *Pyramidal Horn Antenna*

A pyramidal horn antenna (see Fig.9-18a), with a rectangular aperture of width A and height B, designed for optimum gain has an aperture efficiency of 51 %; so from (9-78)

$$G = 0.51 \frac{4\pi}{\lambda^2} AB \tag{9-96}$$

As a specific example; a "standard gain horn" operating from 33 to 50 GHz has a measured gain of 24.7 dB ($G = 295.1$) at 40 GHz ($\lambda = 0.75$ cm). The aperture dimensions of this horn are $A = 5.54$ cm and $B = 4.55$ cm: Using these values of λ, A, and B in (9-96) gives $G = 287.2 = 24.6$ dB. The gain can also be estimated from the principal plane half-power beamwidths, measured at 40 GHz to be $\text{HP}_{E^o} = 9°$ and $\text{HP}_{H^o} = 10°$. Then (9-95) yields $G = 288.9 = 24.6$ dB. The gain values from both of these methods agree very well with the measured gain of 24.7 dB.

EXAMPLE 9-6 *Circular Parabolic Reflector Antenna*

The aperture efficiency of a common circular parabolic reflector antenna with diameter D is 55%, so (9-78) becomes

$$G = \varepsilon_{ap} \frac{4\pi}{\lambda^2} A_p = 0.55 \frac{4\pi}{\lambda^2} \left(\pi \frac{D^2}{4} \right) = 5.43 \frac{D^2}{\lambda^2} \qquad (9\text{-}97)$$

For a specific example, a 3.66-m (12-ft) circular reflector operating at 11.7 GHz ($\lambda = 2.564$ cm) has a measured value of $G = 50.4$ dB and $HP_{E^o} = HP_{H^o} = 0.5°$. Again, we will check our estimation formulas. First, (9-97) gives $G = 5.43(366/2.564)^2 = 110,644 = 50.4$ dB. Next, (9.95) yields $G = 26,000/(0.5)^2 = 104,000 = 50.2$ dB. Both of these estimates are in good agreement with the measured gain.

9.4 RECTANGULAR HORN ANTENNAS

Horn antennas are extremely popular antennas in the microwave region above about 1 GHz. Horns provide high gain, low VSWR, relatively wide bandwidth, low weight, and they are rather easy to construct. As an additional benefit, the theoretical calculations for horn antennas are achieved very closely in practice.

The three basic types of horn antennas that utilize rectangular geometry are illustrated in Fig. 9-11. These horns are fed by a rectangular waveguide that is oriented with its broad wall horizontal. For dominant waveguide mode excitation, the *E*-plane is then vertical and the *H*-plane horizontal. If the horn serves to flare the broad wall dimension and leave the narrow wall of the waveguide unchanged, it is called an ***H*-plane sectoral horn antenna** as shown in Fig. 9-11*a*. On the other hand, if the horn serves to flare only in the *E*-plane dimension, it is called an ***E*-plane sectoral horn antenna** and is shown in Fig. 9-11*b*. When both waveguide dimensions are flared, it is referred to as a **pyramidal horn antenna**, which is shown in Fig. 9-11*c*.

The operation of a horn antenna can be viewed as analogous to a megaphone, which is an acoustic horn radiator providing directivity far sound waves. The electromagnetic horn acts as a transition from the waveguide mode to the free-space mode. This transition reduces reflected waves and emphasizes the traveling waves. This traveling wave behavior, as we have seen with other antennas, leads to low VSWR and wide bandwidth.

Aperture antennas are among the oldest antennas. Heinrich Hertz experimented with microwave parabolic cylinder antennas in 1888. The Indian Physicist J. Chunder Bose operated a pyramidal horn, which he called a "collecting funnel," at 60 GHz in 1897. The

(*a*) *H*-plane sectoral horn.   (*b*) *E*-plane sectoral horn.
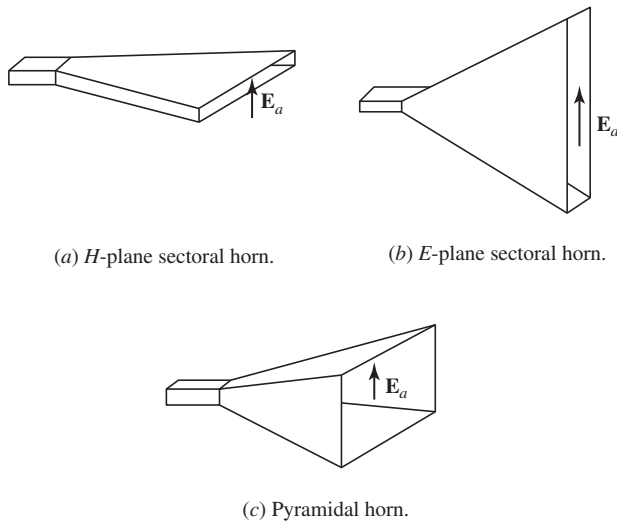
(*c*) Pyramidal horn.

Figure 9-11 Rectangular horn antennas.

horn antenna has been in widespread use since the 1940s; see Appendix H.8.6 books for on horn antennas.

A characteristic of the horn antenna that we have not encountered until now is that the longer path length from the connecting waveguide to the edge of the horn aperture compared to the aperture center in the flare plane introduces a phase delay across the aperture. This aperture "phase error" is not present in antennas such as an open-ended waveguide and complicates the analysis. Phase errors occur in several areas of antennas and they are treated in this section on rectangular horn antennas. In addition to the rectangular horns, conical horn antennas are common. There are also special-purpose horns including those with a dielectric or metallic plate lens in the aperture to correct for the phase error and those with metallic ridges inside the horn to increase bandwidth [H.8.5]. One of the most important applications for horn antennas is as a feed for a reflector antenna. Popular feed horns have corrugations on the inside walls; these are discussed in Sec. 9.7.

9.4.1 The *H*-Plane Sectoral Horn Antenna

The *H*-plane sectoral horn of Fig. 9-12*a* is fed from a rectangular waveguide of interior dimensions a and b, with a the broadwall dimension. The aperture is of width A in the *H*-plane and height b in the *E*-plane. The *H*-plane cross section of Fig. 9-12*b* reveals the geometrical parameters. The following relationships for the geometry will be of use in subsequent analysis:

$$\ell_H^2 = R_1^2 + \left(\frac{A}{2}\right)^2 \tag{9-98}$$

$$\alpha_H = \tan^{-1}\left(\frac{A}{2R_1}\right) \tag{9-99}$$

$$R_H = (A-a)\sqrt{\left(\frac{\ell_H}{A}\right)^2 - \frac{1}{4}} \tag{9-100}$$

The preceding relations follow directly from Fig. 9-12*b* and it is an exercise to prove (9-100). The dimensions A and R_H (or ℓ_H or R_1) must be determined to allow construction of the horn. We first investigate the principles of operation and then present design procedures for determining the horn dimensions.

The key to solving aperture antenna problems is to find the tangential fields over the aperture. The aperture plane for the *H*-plane sectoral horn shown in Fig. 9-12*a* is in

(*a*) Overall geometry.

(*b*) Cross section through the *xz*-plane (H-plane).
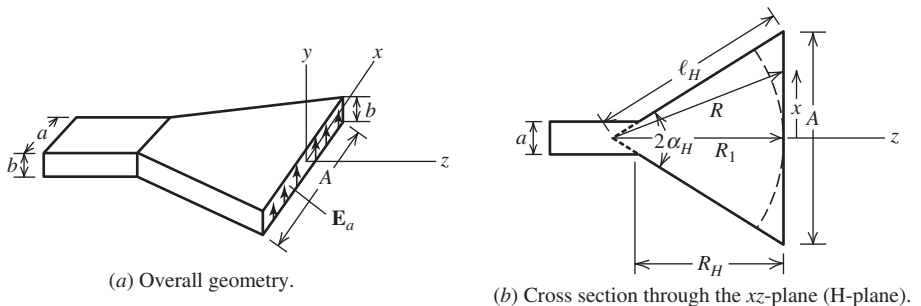
Figure 9-12 *H*-plane sectoral horn antenna.

the *xy*-plane. The aperture fields, of course, arise from the attached waveguide. As is usually the case in practice, we will assume that the waveguide carries the dominant TE_{10} rectangular waveguide mode. The transverse fields in the waveguide are then given by

$$E_y = E_{\text{og}} \cos \frac{\pi x}{a} e^{-j\beta_g z} \tag{9-101a}$$

$$H_z = -\frac{E_y}{Z_g} \tag{9-101b}$$

where $Z_g = \eta[(1-\lambda/2a)^2]^{-1/2}$ is the waveguide characteristic impedance. The fields arriving at the aperture are essentially an expanded version of these waveguide fields. However, the waves arriving at different points in the aperture are not in phase because of the different path lengths. We will now determine this phase distribution.

The path length R from the (virtual) horn apex in the waveguide to the horn aperture increases toward the horn mouth edges. Thus, waves arriving at aperture positions displaced from the aperture center lag in phase relative to those arriving at the center. The phase constant changes from that in the waveguide β_g to the free space constant β as waves progress down the horn. But for relatively large horns, the phase constant for waves in the vicinity of the aperture is approximately that of free space. The aperture phase variation in the *x*-direction is then given by

$$e^{-j\beta(R-R_1)} \tag{9-102}$$

The aperture phase is uniform in the *y*-direction. An approximate form for R using Fig. 9-12*b* is

$$R = \sqrt{R_1^2 + x^2} = R_1 \left[1 + \left(\frac{x}{R_1} \right)^2 \right]^{1/2} \tag{9-103a}$$

$$\approx R_1 \left[1 + \frac{1}{2} \left(\frac{x}{R_1} \right)^2 \right] \tag{9-103b}$$

for $x \ll R_1$ that holds if $A/2 \ll R_1$. Then

$$R - R_1 \approx \frac{1}{2} \frac{x^2}{R_1} \tag{9-104}$$

The amplitude distribution is an expanded version of that in the waveguide, so it is a cosine taper in the *x*-direction. Using this fact and (9-104) in (9-102) leads to the aperture electric field distribution

$$E_{ay} = E_o \cos \frac{\pi x}{A} e^{-j(\beta/2R_1)x^2} \tag{9-105}$$

inside the aperture and zero elsewhere. Note that $E_{ay}(x = \pm A/2) = 0$ as required by boundary conditions. The phase distribution is often referred to as a quadratic phase error, since the deviation from a uniform phase condition varies as the square of the distance from the aperture center. This result can be derived more rigorously by representing the horn as a radial waveguide [H.3: Collin and Zucker, Part 1, p. 636].

The quadratic phase error complicates the radiation integral evaluation; however, the result is worth the effort. Substituting (9-105) into (9-18*b*) yields

$$P_y = E_o \int_{-A/2}^{A/2} \cos \frac{\pi x'}{A} e^{-j(\beta/2R_1)x'^2} e^{j\beta u x'} dx' \int_{-b/2}^{b/2} e^{j\beta v y'} dy' \tag{9-106}$$

After considerable work, this reduces to

$$P_y = E_o \left[\frac{1}{2} \sqrt{\frac{\pi R_1}{\beta}} \, I(\theta, \phi) \right] \left\{ b \frac{\sin[(\beta b/2)\sin \theta \sin \phi}{(\beta b/2)\sin \theta \sin \phi)} \right\} \tag{9-107}$$

where the factors in brackets correspond to each of the integrals in (9-106). The second factor is that for a uniform line source. The first involves the function

$$\begin{aligned} I(\theta, \phi) &= e^{j(R_1/2\beta)(\beta \sin \theta \cos \phi + \pi/A)^2} [C(s_2') - jS(s_2') - C(s_1') + jS(s_1')] \\ &\quad + e^{j(R_1/2\beta)(\beta \sin \theta \cos \phi - \pi/A)^2} [C(t_2') - jS(t_2') - C(t_1') + jS(t_1')] \end{aligned} \tag{9-108}$$

where

$$s_1' = \sqrt{\frac{1}{\pi \beta R_1}} \left(-\frac{\beta A}{2} - R_1 \beta u - \frac{\pi R_1}{A} \right)$$

$$s_2' = \sqrt{\frac{1}{\pi \beta R_1}} \left(\frac{\beta A}{2} - R_1 \beta u - \frac{\pi R_1}{A} \right)$$

$$\tag{9-109}$$

$$t_1' = \sqrt{\frac{1}{\pi \beta R_1}} \left(-\frac{\beta A}{2} - R_1 \beta u + \frac{\pi R_1}{A} \right)$$

$$t_2' = \sqrt{\frac{1}{\pi \beta R_1}} \left(\frac{\beta A}{2} - R_1 \beta u + \frac{\pi R_1}{A} \right)$$

and the functions $C(x)$ and $S(x)$ are Fresnel integrals defined in (F-17) and tabulated in [5].

   The total radiation fields can now be obtained. Using (9-29) gives the far-zone electric field components

$$E_\theta = j\beta \frac{e^{-j\beta r}}{4\pi r} (1 + \cos \theta) \sin \phi \, P_y \tag{9-110a}$$

$$E_\phi = j\beta \frac{e^{-j\beta r}}{4\pi r} (1 + \cos \theta) \cos \phi \, P_y \tag{9-110b}$$

These together with (9-107) give the complete radiated electric field

$$\begin{aligned} \mathbf{E} &= j\beta E_o b \sqrt{\frac{\pi R_1}{\beta}} \frac{e^{-j\beta r}}{4\pi r} \left(\frac{1 + \cos \theta}{2} \right) (\hat{\theta} \sin \phi + \hat{\phi} \cos \phi) \\ &\quad \cdot \frac{\sin[\beta b/2)\sin \theta \sin \phi]}{(\beta b/2)\sin \theta \sin \phi} I(\theta, \phi) \end{aligned} \tag{9-111}$$

where $I(\theta, \phi)$ is still given by (9-108).

The complete radiation expression is rather cumbersome, so we will examine the principal plane patterns. In the E-plane, $\phi = 90°$ and the normalized form of (9-111) is

$$F_E(\theta) = \frac{1 + \cos\theta}{2} \frac{\sin[(\beta b/2)\sin\theta]}{(\beta b/2)\sin\theta} \tag{9-112}$$

The second factor is the pattern of a uniform line source of length b along the y-axis, as one would expect from the aperture distribution.

In the H-plane, $\phi = 0°$ and the normalized H-plane pattern is

$$F_H(\theta) = \frac{1 + \cos\theta}{2} f_H(\theta) = \frac{1 + \cos\theta}{2} \frac{I(\theta, \phi = 0°)}{I(\theta = 0°, \phi = 0°)} \tag{9-113}$$

The H-plane pattern can be displayed rather simply using universal radiation pattern plots that are based on the maximum phase error across the aperture. The aperture distribution phase error as a function of position x from (9-105) is

$$\delta = \frac{\beta}{2R_1} x^2 \tag{9-114}$$

Since the maximum value of x is $A/2$, the maximum phase error is

$$\delta_{\max} = \frac{\beta}{2R_1} \left(\frac{A}{2}\right)^2 = 2\pi \frac{A^2}{8\lambda R_1} = 2\pi t \tag{9-115}$$

where t is defined to be

$$t = \frac{A^2}{8\lambda R_1} = \frac{1}{8}\left(\frac{A}{\lambda}\right)^2 \frac{1}{R_1/\lambda} \tag{9-116}$$

The function $I(\theta, \phi = 0°)$ in (9-108) can be expressed in terms of t as

$$I(\theta, \phi = 0°) = e^{j(\pi/8t)[(A/\lambda)\sin\theta + 1/2]^2}[C(s_2) - jS(s_2) - C(s_1) + jS(s_1)]$$
$$+ e^{j(\pi/8t)[(A/\lambda)\sin\theta - 1/2]^2}[C(t_2) - jS(t_2) - C(t_1) + jS(t_1)] \tag{9-117}$$

where

$$s_1 = 2\sqrt{t}\left[-1 - \frac{1}{4t}\left(\frac{A}{\lambda}\sin\theta\right) - \frac{1}{8t}\right]$$

$$s_2 = 2\sqrt{t}\left[1 - \frac{1}{4t}\left(\frac{A}{\lambda}\sin\theta\right) - \frac{1}{8t}\right]$$

$$t_1 = 2\sqrt{t}\left[-1 - \frac{1}{4t}\left(\frac{A}{\lambda}\sin\theta\right) + \frac{1}{8t}\right] \tag{9-118}$$

$$t_2 = 2\sqrt{t}\left[1 - \frac{1}{4t}\left(\frac{A}{\lambda}\sin\theta\right) + \frac{1}{8t}\right]$$

This function is plotted in Fig. 9-13 for various values of t. It is normalized to the main beam peak for a zero phase error condition, which displays the directivity loss (reduction of the main beam peak) as the maximum phase error $2\pi t$ increases.

The curves in Fig. 9-13 are universal pattern plots from which antenna patterns can be derived for specific values of A, b, and λ. The H-plane plots (solid curves) are a function of $(A/\lambda) \sin \theta$. The E-plane plot (dashed curve) is the second factor of (9-112), and the abscissa for it is $(b/\lambda) \sin \theta$. The factor $(1 + \cos \theta)/2$ that appears in both pattern functions (9-112) and (9-113) is not included in Fig. 9-13. For most situations, it has a small effect on the total pattern and may be neglected. Its effect, however, is easily included by adding $20 \log[(1 + \cos \theta)/2]$ to the corresponding pattern value from the universal pattern. Note that the E-plane plot of Fig. 9-13 has the -13.3-dB side lobe level of a uniform line source pattern, and the H-plane constant phase $(t = 0)$ plot has the -23-dB side lobe level of a cosine-tapered line source pattern. As the phase error increases, the H-plane pattern beamwidth and side lobes increase.

The pattern of $f_H(\theta)$ in (9-113) can be evaluated with excellent results using a mathematics application computer package to perform the numerical integration:

$$f_H(\theta) \propto \int_{-A/2}^{A/2} \cos \frac{\pi x'}{A} e^{-j\beta \sqrt{R_1^2 + x'^2}} e^{j\beta \sin \theta \, x'} dx' \tag{9-119}$$

where (9-106) was used with the exact phase error expression of (9-103a) rather than the quadratic approximation of (9-103b). This avoids the foregoing complicated expressions and permits inclusion of the exact phase.
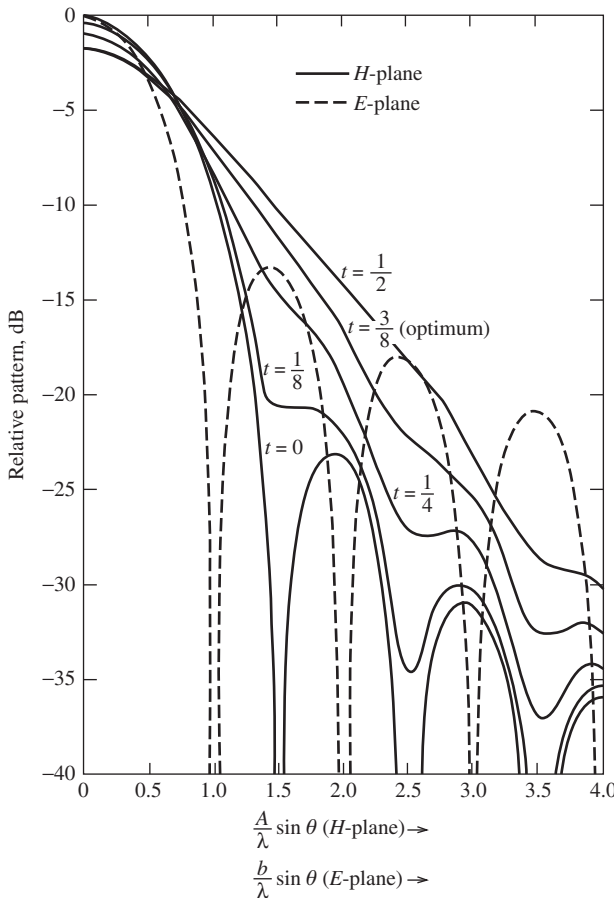
Figure 9-13 Universal radiation patterns for the principal planes of an H-plane sectoral horn as shown in Fig. 9-12. The factor $(1 + \cos \theta)/2$ is not included.

The directivity for an *H*-plane sectoral horn is obtained from the aperture integration method of (9-66) as

$$D_H = \frac{b}{\lambda} \frac{32}{\pi} \left(\frac{A}{\lambda} \right) \varepsilon_{\text{ph}}^H = \frac{4\pi}{\lambda^2} \varepsilon_t \, \varepsilon_{\text{ph}}^H \, Ab \tag{9-120a}$$

where

$$\varepsilon_t = \frac{8}{\pi^2} \tag{9-120b}$$

$$\varepsilon_{\text{ph}}^H = \frac{\pi^2}{64t} \left\{ [C(p_1) - C(p_2)]^2 + [S(p_1) - S(p_2)]^2 \right\} \tag{9-120c}$$

$$p_1 = 2\sqrt{t}\left[1 + \frac{1}{8t}\right], \qquad p_2 = 2\sqrt{t}\left[-1 + \frac{1}{8t}\right] \tag{9-120d}$$

Note that $p_1 = -s_1' = t_2'$ and $p_2 = s_2' = -t_1'$ from (9-108) for $u = 0$. This expression explicitly shows the two efficiency factors associated with aperture taper and phase, ε_t and $\varepsilon_{\text{ph}}^H$.

A family of universal directivity curves is given in Fig. 9-14, where $\lambda D_H/b$ is plotted versus A/λ for various values of R_1/λ. Notice that for a given axial length R_1, there is an optimum aperture width A corresponding to the peak of the appropriate curve. The values of A/λ corresponding to optimum operation plotted versus R_1/λ produce a smooth curve with the equation $A/\lambda = \sqrt{3R_1/\lambda}$, giving

$$A = \sqrt{3\lambda R_1} \qquad optimum \tag{9-121}$$

For example, the value of A/λ for the peak of the $R_1/\lambda = 30$ curve of Fig. 9-14 is 95, and from (9-121), $A/\lambda = \sqrt{3R_1/\lambda} = \sqrt{3(30)} = 9.49$.
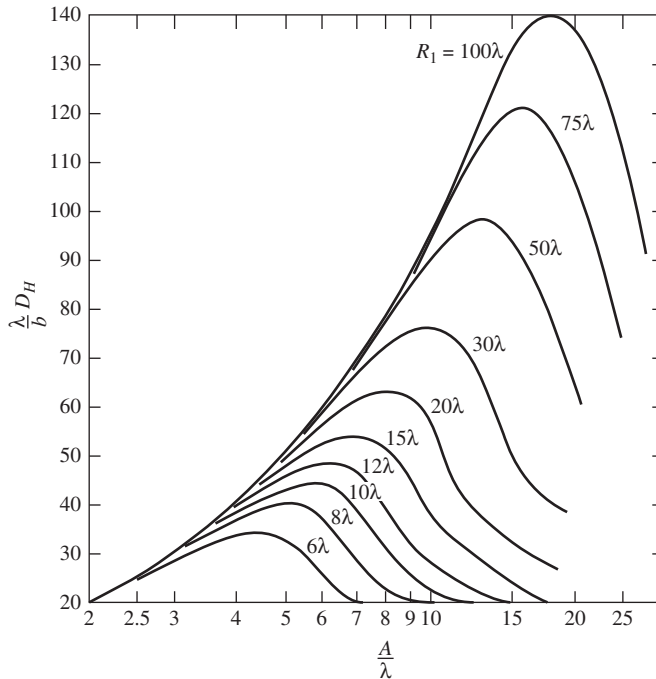
Figure 9-14 Universal directivity curves for an *H*-plane sectoral horn.

The optimum phase error parameter value corresponding to optimum directivity is found from (9-116) with (9-121) as

$$t_{\text{op}} = \frac{A^2}{8\lambda R_1} = \frac{3}{8} \quad \text{optimum} \tag{9-122}$$

The optimum behavior of the directivity curves can be explained rather simply. For a fixed axial length, as the aperture width A is increased from a small value, the directivity increases by virtue of the increased aperture area. Optimum performance is reached when $t = t_{\text{op}} = 3/8$, which corresponds to a phase lag at the aperture edges $(x = \pm A/2)$ of $\delta_{\text{max}} = 2\pi t_{\text{op}} = 3\pi/4 = 135°$. As A is increased beyond the optimum point, the phase deviations across the aperture lead to cancellations in the far field and decreased directivity, as can be seen from the pattern plots of Fig. 9-13.

The foregoing analysis can be performed without approximation by using numerical techniques together with the exact phase error (9-l03a), as with the radiation integral in (9-119), However, it is easier to use a corrected phase error parameter that includes the effects of the exact phase error. The exact pattern is obtained if the value of t in (9-118) is replaced by the following [6]:

$$t_{\text{e}} = \left(\frac{A}{\lambda}\right)^2 \frac{1}{8t} \left\{ \left[1 + \left(\frac{\lambda}{A}\right)^2 16\ t^2 \right]^{1/2} - 1 \right\} \quad \text{exact} \tag{9-123}$$

If the phase error is not large and the aperture is more than a few wavelengths, then $t_{\text{e}} \approx t$. For example, the optimum case with a 3λ aperture $(A/\lambda = 3)$ has an exact phase error parameter of $t_{\text{e}} = 0.354$, which is close to the approximate value of 0.375. Similarly, exact phase error conditions for directivity are obtained by replacing t with t_{e} in (9-120).

The half-power beamwidth for optimum performance can be determined from the pattern plot of Fig. 9-13 for $t = 3/8$. The 3-dB down point on the main beam occurs for $(A/\lambda)\sin\theta_H = 0.68$, so the H-plane beamwidth for an optimum H-plane sectoral horn is $2\theta_H = \sin^{-1}(0.68\lambda/A)$; and for $A \gg \lambda$,

$$\text{HP}_H \approx 1.36\frac{\lambda}{A} = 78°\frac{\lambda}{A} \quad \text{optimum} \tag{9-124}$$

9.4.2   The E-Plane Sectoral Horn Antenna

A rectangular horn antenna can also be formed by flaring the feed waveguide in the E-plane. The resulting horn is referred to as an **E-plane sectoral horn antenna** as shown in Fig. 9-15. The geometrical relationships for this horn are

$$\ell_E^2 = R_2^2 + \left(\frac{B}{2}\right)^2 \tag{9-125}$$

$$\alpha_E = \tan^{-1}\left(\frac{B}{2R_2}\right) \tag{9-126}$$

$$R_E = (B - b)\sqrt{\left(\frac{\ell_E}{B}\right)^2 - \frac{1}{4}} \tag{9-127}$$

A similar line of reasoning as employed for the H-plane horn leads to the following aperture electric field distribution for the E-plane horn:
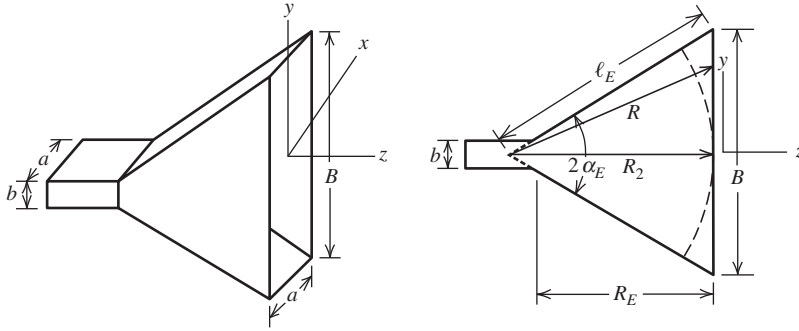
(a) Overall geometry.   (b) Cross section through the yz-plane (E-plane).

Figure 9-15 *E*-plane sectoral horn antenna.

$$E_{ay} = E_o \ \cos \frac{\pi x}{a} e^{-j(\beta/2R_2)y^2} \tag{9-128}$$

The same steps as used with the *H*-plane sectoral horn yield the radiation field:

$$\mathbf{E} = j\beta E_o \sqrt{\frac{\pi R_2}{\beta}} \frac{4a}{\pi} \frac{e^{-j\beta r}}{4\pi r} e^{j(\beta R_2/2)v^2} (\hat{\boldsymbol{\theta}} \sin \phi + \hat{\boldsymbol{\phi}} \cos \phi)$$

$$\cdot \frac{1 + \cos \theta}{2} \frac{\cos[(\beta a/2)u]}{1 - [(\beta a/\pi)u]^2} [C(r_2) - jS(r_2) - C(r_1) + jS(r_1)] \tag{9-129a}$$

where

$$r_1 = \sqrt{\frac{\beta}{\pi R_2}} \left(-\frac{B}{2} - R_2 v \right), \qquad r_2 = \sqrt{\frac{\beta}{\pi R_2}} \left(\frac{B}{2} - R_2 v \right) \tag{9-129b}$$

The normalized *H*-plane pattern follows from this with $\phi = 0°$ as

$$F_H(\theta) = \frac{1 + \cos \theta}{2} \frac{\cos[(\beta a/2) \sin \theta]}{1 - [(\beta a/\pi) \sin \theta]^2} \tag{9-130}$$

The second factor in this expression is the pattern of a uniform phase, cosine amplitude tapered line source of length a.

The aperture phase error in the *E*-plane is approximated with the quadratic phase error in (9-128) as $\delta = (\beta/2R_2)y^2$. The maximum phase error occurs for $y = \pm B/2$, giving $\delta_{\max} = (\beta/2R_2)(B/2)^2 = 2\pi(B^2/8\lambda R_2) = 2\pi s$, where we define the phase error parameter s as

$$s = \frac{B^2}{8\lambda R_2} = \frac{1}{8} \left(\frac{B}{\lambda} \right)^2 \frac{1}{R_2/\lambda} \tag{9-131}$$

The *E*-plane pattern magnitude from (9-129) with $\phi = 90°$ can be expressed in terms of *s* as

$$|F_E(\theta)| = \frac{1+\cos\theta}{2}|f_E(\theta)|$$

$$= \frac{1+\cos\theta}{2}\left\{\frac{[C(r_4)-C(r_3)]^2 + [S(r_4)-S(r_3)]^2}{4[C^2(2\sqrt{s}) + S^2(2\sqrt{s})]}\right\}^{1/2} \qquad (9\text{-}132a)$$

where

$$r_3 = 2\sqrt{s}\left[-1-\frac{1}{4s}\left(\frac{B}{\lambda}\sin\theta\right)\right], \qquad r_4 = 2\sqrt{s}\left[1-\frac{1}{4s}\left(\frac{B}{\lambda}\sin\theta\right)\right] \qquad (9\text{-}132b)$$

Similar to (9-119) for the *H*-plane sectoral horn, the pattern of an *E*-plane sectoral horn can be evaluated by direct numerical integration; $f_E(\theta)$ in (9-132a) is found from

$$f_E(\theta) \propto \int_{-B/2}^{B/2} e^{-j\beta\sqrt{R_2^2 + y'^2}} e^{j\beta\sin\theta\, y'} dy' \qquad (9\text{-}133)$$

The universal patterns for the *E*-plane sectoral horn are plotted in Fig. 9-16. The *E*-plane patterns (solid curves) for various values of *s* are not normalized to 0 dB at the maximum point, but rather are given relative to the no-phase error case, which is $s = 0$ corresponding to a uniform line source. The *H*-plane pattern (dashed curve) is that of a cosine-tapered line source, which is the second factor of (9-130). The factor $(1 + \cos\theta)/2$ is not included in these plots.

The directivity of the *E*-plane sectoral horn found from (9-66) is

$$D_E = \frac{a}{\lambda}\frac{32}{\pi}\frac{B}{\lambda}\varepsilon_{\text{ph}}^E = \frac{4\pi}{\lambda^2}\varepsilon_t\,\varepsilon_{\text{ph}}^E\, aB \qquad (9\text{-}134a)$$

where

$$\varepsilon_t = \frac{8}{\pi^2} \qquad (9\text{-}134b)$$

$$\varepsilon_{\text{ph}}^E = \frac{C^2(q) + S^2(q)}{q^2} \qquad (9\text{-}134c)$$

$$q = \frac{B}{\sqrt{2\lambda R_2}} = 2\sqrt{s} \qquad (9\text{-}134d)$$

A family of universal directivity curves $\lambda D_E/a$ for various values of R_2/λ is given in Fig. 9-17 as a function of B/λ. The peak of each curve corresponds to optimum directivity for the value of R_2. A curve fit to pairs of values of B/λ and R_2/λ for optimum conditions yields

$$B = \sqrt{2\lambda R_2} \quad optimum \qquad (9\text{-}135)$$

The corresponding value of *s* is

$$s_{\text{op}} = \frac{B^2}{8\lambda R_2} = \frac{1}{4} \quad optimum \qquad (9\text{-}136)$$
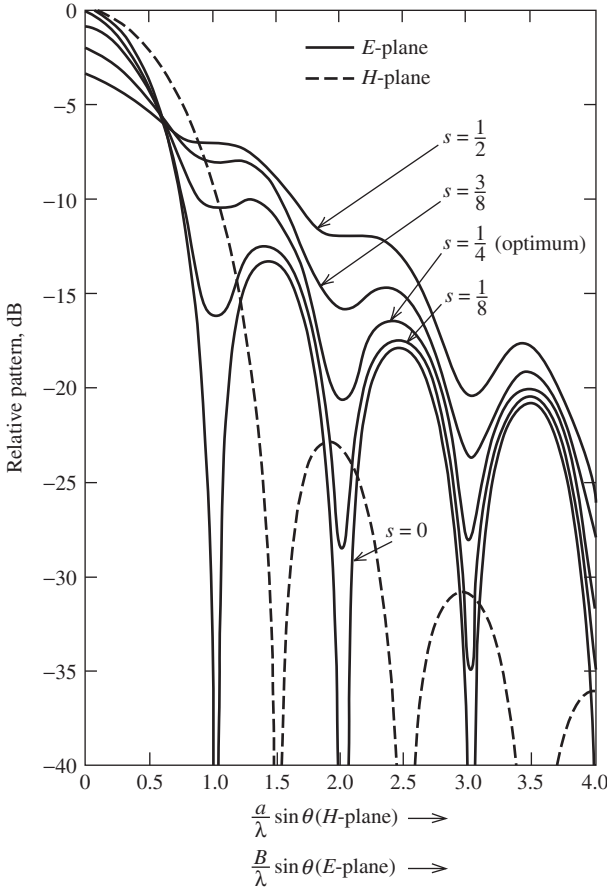
Figure 9-16 Universal radiation patterns for the principal planes of an E-plane sectoral horn antenna as shown in Fig. 9-15. The factor $(1 + \cos\theta)/2$ is not included.
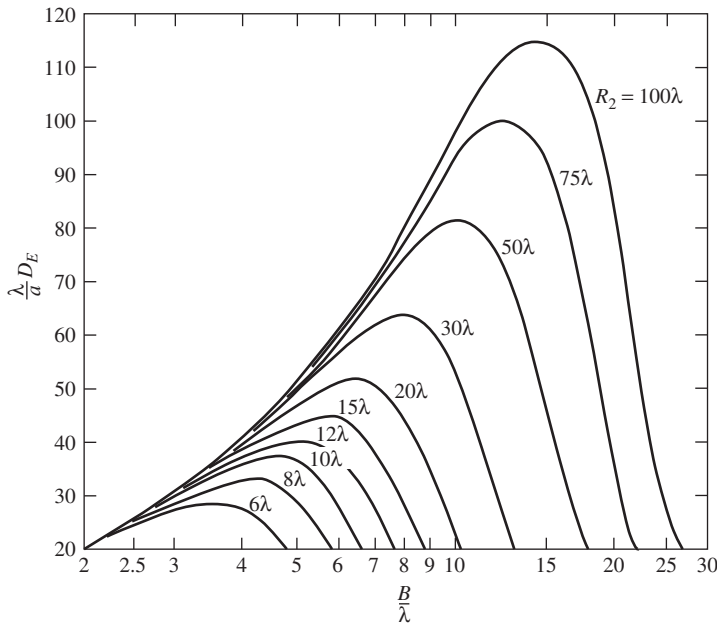
Figure 9-17 Universal directivity curves for an E-plane sectoral horn.

Exact phase error conditions corresponding to spherical wave fronts in the aperture plane can be included easily, replacing s by s_e [6]:

$$s_e = \left(\frac{B}{\lambda}\right)^2 \frac{1}{8s} \left\{ \left[1 + \left(\frac{\lambda}{B}\right)^2 16s^2 \right]^{1/2} - 1 \right\} \quad exact \qquad (9\text{-}137)$$

That is, the pattern and directivity expressions of (9-132) and (9-134) are made exact by using (9-137). However, in practice, accuracy cannot be expected when the aperture is small, because the aperture fields are not well approximated by free-space conditions.

The half-power beamwidth relationship for the optimum horn follows from the $s = 1/4$ plot in Fig. 9-16 and is

$$\text{HP}_E = 2 \sin^{-1} \frac{0.47}{B/\lambda} \approx 0.94 \frac{\lambda}{B} = 54° \frac{\lambda}{B} \quad optimum \qquad (9\text{-}138)$$

Gain for horn antennas nearly equals directivity—that is, $G_E \approx D_E$ and $G_H \approx D_H$. The gain of an E-plane sectoral horn has been shown to be more accurately given by [7]

$$G_E = \frac{16aB}{\lambda^2(1 + \lambda_g/\lambda)} \frac{C^2(q_1) + S^2(q_1)}{q_1^2} e^{\pi(a/\lambda)(1 - \lambda/\lambda_g)} \qquad (9\text{-}139)$$

where $\lambda_g = \lambda/\sqrt{1 - (\lambda/2a)^2}$ is the wavelength of the dominant mode in the waveguide feeding the antenna and $q_1 = B[\sqrt{2\lambda_g \ell_E} \cos(\alpha_E/2)]^{-1}$. This expression yields values that agree quite well with experimental results. The values from (9-134) are less than those of (9-139) by 20% or more.

9.4.3 The Pyramidal Horn Antenna

Probably the most popular form of the rectangular horn antenna is the **pyramidal horn antenna**. As shown in Fig. 9-18, it is flared in both the E- and H-planes. This configuration will lead to narrow beamwidths in both principal planes, forming a pencil beam. The aperture electric field is obtained by combining the results for H- and E-plane sectoral horns from (9-105) and (9-128) giving

$$E_{\text{ay}} = E_o \cos\left(\frac{\pi x}{A}\right) e^{-j(\beta/2)(x^2/R_1 + y^2/R_2)} \qquad (9\text{-}140)$$

Following a procedure similar to that used for the sectoral horns will yield a general radiation field expression. The principal plane patterns are the same as those obtained from the sectoral horn calculations because the aperture distribution is separable as in (9-72). To be precise, the E- and H-plane patterns of the pyramidal horn equal the E-plane pattern of the E-plane sectoral horn and the H-plane pattern of the H-plane sectoral horn. Therefore, the E-plane pattern of the pyramidal horn can be found from the universal pattern plots (solid curves) of Fig. 9-16, and the H-plane pattern can be found from the solid curves of Fig. 9-13.

Since pyramidal horns are used as gain standards at microwave frequencies, accurate gain evaluation is important. The directivity of the pyramidal horn is found rather simply from
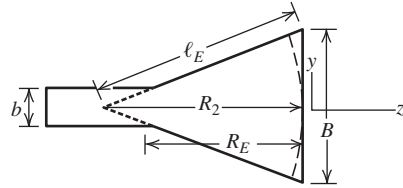
$$D_p = \frac{\pi}{32} \left(\frac{\lambda}{a} D_E\right) \left(\frac{\lambda}{b} D_H\right) \qquad (9\text{-}141)$$
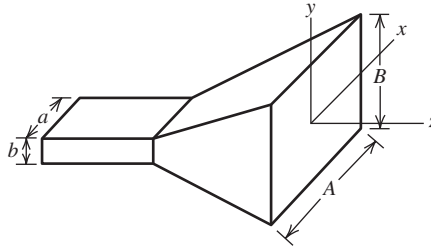
(a) Overall geometry.
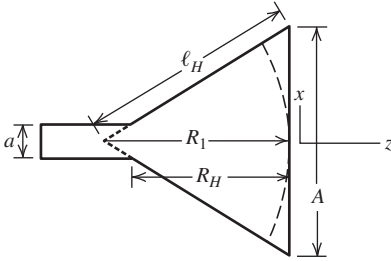
(b) Cross section through the xz-plane (H-plane).   (c) Cross section through the yz-plane (E-plane).

Figure 9-18 Pyramidal horn antenna.

See Prob. 9.4-15. The terms in parentheses are obtained directly from the directivity curves for sectoral horns of Figs. 9-14 and 9-17, respectively. Gain values computed with (9-141) agree very well with experiment for sufficiently large horns. It includes the geometrical optics fields and singly diffracted fields from the horn edges. The inclusion of multiple diffraction and diffraction at the edges arising from reflections from the horn interior leads to small oscillations in the gain about that predicted by (9-141) as a function of frequency, and in agreement with experimental results [8].

It is instructive to examine the aperture efficiency contributions for horns. The radiation efficiency e_r is close to unity, so we can take gain to be equal to directivity; see (9-81). The two efficiencies that must be considered are the aperture taper efficiency ε_t and phase efficiency ε_{ph}:

$$\varepsilon_{ap} = \varepsilon_t \varepsilon_{ph} = \varepsilon_t \varepsilon_{ph}^E \varepsilon_{ph}^H \tag{9-142}$$

where we decomposed total phase efficiency into factors due to phase errors in the E- and H-planes. Gain is then expressed from (9-77) as

$$G = \frac{4\pi}{\lambda^2} \varepsilon_{ap} AB = \frac{4\pi}{\lambda^2} \varepsilon_t AB \varepsilon_{ph}^E \varepsilon_{ph}^H = G_o \varepsilon_{ph}^E \varepsilon_{ph}^H \tag{9-143}$$

where G_o is the gain without a phase error effect and includes aperture taper efficiency, which was found in Example 9-4 to be $\varepsilon_t = 0.81$. The phase error efficiencies can be found by evaluating directivity for the sectoral horns and removing the known taper efficiency. The results of this process are plotted in Fig. 9-19 as a function of phase error parameters s and t. The aperture efficiencies for optimum sectoral horns with $s = 0.25$ and $t = 0.375$ are

$$\varepsilon_{ap}^E = 0.649, \quad \varepsilon_{ap}^H = 0.643 \quad optimum \tag{9-144}$$
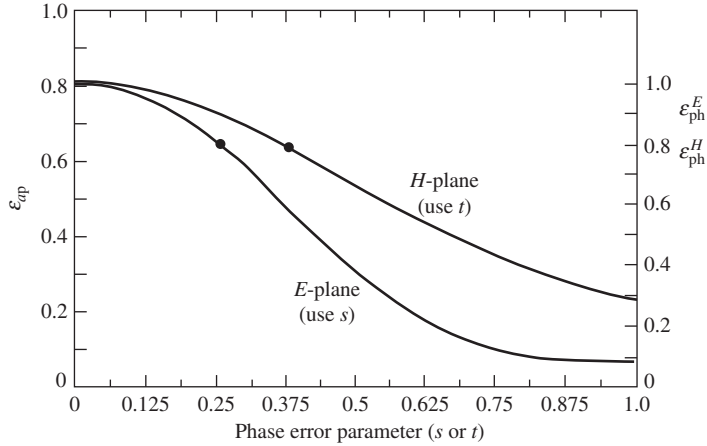
Figure 9-19 Aperture efficiencies for E- and H-plane sectoral horns (left ordinate) and phase efficiencies associated with E- and H-plane flares (right ordinate).

Both include $\varepsilon_t = 0.81$. So,

$$\varepsilon_{\text{ph}}^E = \frac{\varepsilon_{\text{ap}}^E}{\varepsilon_t} = 0.80, \quad \varepsilon_{\text{ph}}^H = \frac{\varepsilon_{\text{ap}}^H}{\varepsilon_t} = 0.79 \tag{9-145}$$

The aperture efficiency of an optimum pyramidal horn from (9-142) is

$$\varepsilon_{\text{ap}}^P = \varepsilon_t \varepsilon_{\text{ph}}^E \varepsilon_{\text{ph}}^H = 0.81(0.80)(0.79) = 0.51 \tag{9-146}$$

It is common to use an aperture efficiency value of 50% for optimum gain pyramidal horns.

The gain of an optimum gain pyramidal horn from (9-146) in (9-143) is

$$\boxed{G = 0.51 \frac{4\pi}{\lambda^2} AB} \quad \textit{optimum pyramidal horn} \tag{9-147}$$

It is popular to express horn gain in dB form by taking 10 log of (9-143):

$$G_{\text{dB}} = G_{o,\text{dB}} + \varepsilon_{\text{ph,dB}}^E(s) + \varepsilon_{\text{ph,dB}}^H(t) \tag{9-148}$$

The last two terms are gain reduction factors associated with the phase errors of (9-120c) and (9-134c). These phase efficiencies, before taking 10 log, can be approximated with simple formulas [9]:

$$\varepsilon_{\text{ph}}^E(s) = \frac{1}{4s} \left[C^2(2\sqrt{s}) + S^2(2\sqrt{s}) \right] \approx 1.00329 - 0.11911s - 2.75224s^2 \tag{9-149}$$

$$\varepsilon_{\text{ph}}^H(t) = \frac{\pi^2}{64t} \left\{ [C(p_1) - C(p_2)]^2 + [S(p_1) - S(p_2)]^2 \right\}$$
$$\approx 1.00323 - 0.08784t - 1.27048t^2 \tag{9-150}$$

The approximate formulas are valid from zero up to at least $s = 0.262$ and $t = 0.397$. For example, $s = 0.25$ and $t = 0.375$ in the approximate formulas give the values in (9-144), which are the points shown in Fig. 9-19. Increased accuracy is obtained if the exact phase error parameters in (9-123) and (9-137) are used.

Many applications for horns require a specified gain to be realized at a known operating frequency. Usually, the optimum gain design approach is used because it renders the shortest axial length for the specified gain. We now derive the single design equation that permits determination of the optimum horn geometry for the specified gain. The procedure includes the connecting waveguide internal dimensions a and b as well as the horn dimensions. There are three conditions that must be satisfied. The first two are that the phase error in the E- and H-planes be those associated with optimum performance. The third is that the structure of the pyramidal horn be physically realizable and properly mated to the connecting waveguide. This can be seen from Fig. 9-18 to be

$$R_E = R_H = R_P \tag{9-151}$$

From similar triangles in Fig. 9-18,

$$\frac{R_1}{R_H} = \frac{A/2}{A/2 - a/2} = \frac{A}{A - a} \tag{9-152}$$

$$\frac{R_2}{R_E} = \frac{B/2}{B/2 - b/2} = \frac{B}{B - b} \tag{9-153}$$

Imposing the optimum performance in the E-plane through (9-135) and substituting (9-153), we obtain

$$B = \sqrt{\frac{2\lambda R_E B}{B - b}} \quad \text{or} \quad B^2 - bB - 2\lambda R_E = 0 \tag{9-154}$$

which is a quadratic equation with one solution:

$$B = \frac{1}{2}(b + \sqrt{b^2 + 8\lambda R_E}) \tag{9-155}$$

The second solution yields the impossible case of negative B and is ignored. Similarly, the optimum performance condition for the H-plane of (9-122) together with (9-152) yields

$$R_H = \frac{A - a}{A} R_1 = \frac{A - a}{A}\left(\frac{A^2}{3\lambda}\right) = \frac{A - a}{3\lambda} A \tag{9-156}$$

Imposing the physical realization condition of (9-151) with (9-156) in (9-155) gives

$$B = \frac{1}{2}\left(b + \sqrt{b^2 + \frac{8A(A - a)}{3}}\right) \tag{9-157}$$

Linking this to the specified gain G gives

$$G = \frac{4\pi}{\lambda^2}\varepsilon_{\text{ap}} AB = \frac{4\pi}{\lambda^2}\varepsilon_{\text{ap}} A\frac{1}{2}\left(b + \sqrt{b^2 + \frac{8A(A - a)}{3}}\right) \tag{9-158}$$

Expanding to form a fourth-order equation in A gives the desired design equation [10]:

$$\boxed{A^4 - aA^3 + \frac{3bG\lambda^2}{8\pi\varepsilon_{\text{ap}}}A = \frac{3G^2\lambda^4}{32\pi^2\varepsilon_{\text{ap}}^2}} \quad \textit{pyramidal horn design equation} \tag{9-159}$$

It is possible to solve this quartic equation for its roots, but it is rather involved and the solution is easily obtained using a numerical equation solver routine. Alternatively, it can be solved by trial and error using a first guess approximation of

$$A_1 = 0.45\lambda\sqrt{G} \tag{9-160}$$

We now summarize the steps in the optimum horn design procedure:

Step 1: Specify the desired gain G at the operating wavelength λ and specify the connecting waveguide dimensions a and b.

Step 2: Solve (9-159) for A using $\varepsilon_{ap} = 0.5l$.

Step 3: Find the remaining horn dimensions as follows: B from (9-147); R_1 from (9-121); R_H from (9-152); l_H from (9-98); R_2 from (9-135); R_E from (9-153); and l_E from (9-125).

Step 4: The correct solution can be verified by checking to see if R_E equals R_H and by evaluating (9-131) and (9-116) to see if $s = 0.25$ and $t = 0.375$.

Horn antennas operate well over a bandwidth of about 50%. However, performance is optimum only at the design frequency. Fig. 9-20 is again curve for the "standard gain horn" in the 8.2 to 12.4 GHz band that is considered in Example 9-7. Note that gain increases with frequency, which is characteristic of aperture antennas. The curve is not a straight line as might seem to be the case from the explicit frequency-squared dependence in (9-78). This is because aperture efficiency decreases with frequency due to increasing phase errors, as shown in Fig. 9-20. Thus, an optimum gain horn is only "optimum" at its design frequency.

Before closing this section with an example of optimum horn design, we comment on the assumption that phase error arises from a spherical phase front in the aperture and the wavelength there equals that of free space. A solution technique for rectangular horns is available that uses a gradual change in phase velocity from the waveguide to the aperture by treating each point as a section of an infinitely long waveguide of that width [11]. However, for all but short horns with small apertures, gain does not differ noticeably from the foregoing design approach.
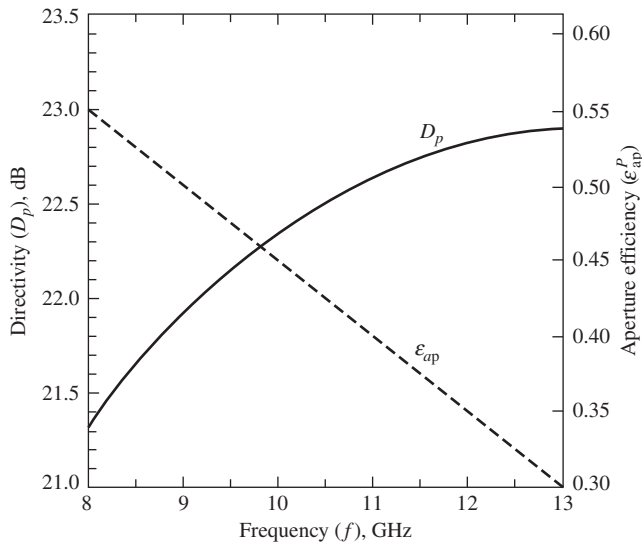
Figure 9-20 Directivity and aperture efficiency of the standard gain rectangular horn of Example 9-7.

EXAMPLE 9-7   *Design of an Optimum Gain Pyramidal Horn Antenna*

Commercial "standard gain" pyramidal horn antennas are available to cover the frequency band from 8.2 to 12.4 GHz (*X*-band). They are fed from a WR90 waveguide with $a = 0.9$ in. $= 2.286$ cm and $b = 0.4$ in. $= 1.016$ cm. As the gain curve in Fig. 9-20 indicates, the aperture efficiency decreases rapidly with frequency. Therefore, the optimum design point is chosen near the low end of the band to provide more uniform gain over the whole band. Efficiency reduction due to increased phase errors at the high end of the band is significant, but the aperture is much larger electrically so gain actually increases.

For this design example, we choose the optimum design point to be at 8.75 GHz, where aperture efficiency is 51%. The desired gain is $G = 21.75$ dB or $10^{2.175} = 149.6$ at 8.75 GHz ($\lambda = 3.43$ cm). The design equation of (9-159) is solved by trial and error beginning with $A_1 = 18.9$ cm using (9-160). Step 3 of the design procedure gives all remaining horn dimensions:

$$
\begin{array}{ll}
A = 18.61 \text{ cm}, & B = 14.75 \text{ cm} \\
R_1 = 33.67 \text{ cm}, & R_2 = 31.72 \text{ cm} \\
l_H = 34.93 \text{ cm}, & l_E = 32.56 \text{ cm} \\
R_H = 29.53 \text{ cm}, & R_E = 29.53 \text{ cm}
\end{array}
$$

These values are verified by noting that $R_p = R_E = R_H$ and by evaluating (9-131) and (9-116) to obtain the optimum values of $s = 0.25$ and $t = 0.375$. The gain value is verified using the universal directivity curves $R_2/\lambda = 9\text{-}3$ and $B/\lambda = 4.3$ with Fig. 9-17, giving $\lambda D_E/a = 36$, and $R_1/\lambda = 10.1$ and $A/\lambda = 5.4$ with Fig. 9-14, giving $\lambda D_H/b = 43$. Then from (9-141),

$$
D_p = \frac{\pi}{32}\left(\frac{\lambda}{a}D_E\right)\left(\frac{\lambda}{b}D_H\right) = \frac{\pi}{32}(36)(43) = 152 = 21.8 \text{ dB}
$$

which is very close to the design goal of 21.75 dB. Accurate evaluation of directivity using (9-120c) and (9-134c) in (9-148) with $s_{op} = 0.25$ and $t_{op} = 0.375$ gives a value of 21.79 dB. The exact phase errors for this geometry are $s_e = 0.247$ and $t_e = 0.368$ from (9-137) and (9-123); they lead to a directivity of 21.85 dB. The directivity as a function of frequency is plotted in Fig. 9-20; sec Prob. 9.4-17.

The complete radiation patterns at 8.75 GHz are plotted in Fig. 9-21 including the $(1 + \cos\theta)/2$ factor. The half-power beamwidths are

$$
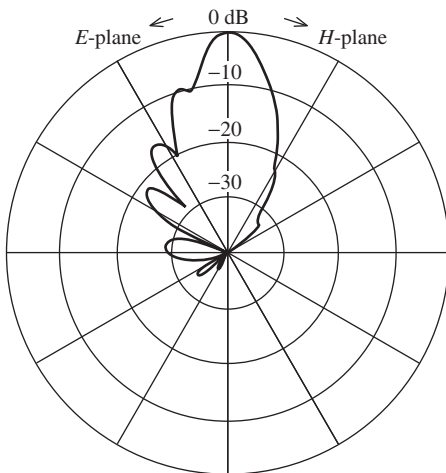\text{HP}_E = 12.4°, \text{HP}_H = 14.2°
$$

Figure 9-21 Principal plane patterns for the optimum pyramidal horn antenna of Example 9-7 at 8.75 GHz. The patterns include the $(1 + \cos\theta)/2$ factor. $\text{HP}_E = 12.4°$ and $\text{HP}_H = 14.2°$.

These agree exactly with the predicted values based on (9-138) and (9-124). Thus, the simple half-power beamwidth formulas of (9-138) and (9-124) give good results for optimum horns. The first side lobe of the E-plane and H-plane patterns in Fig. 9-21 are located at 16° and 44° with values of -9.4 and -32.5 dB, respectively. The E- and H-plane first side lobe values without the $(1 + \cos\theta)/2$ factor included are -9.2 and -31.2 dB, respectively, and can be found in Figs. 9-16 and 9-13. The gain (directivity) can also be estimated directly from beamwidths using (9-95):

$$G = \frac{26{,}000}{\mathrm{HP}_{E^o}\mathrm{HP}_{H^o}} = \frac{26{,}000}{(12.4°)(14.2°)} = 21.7 \text{ dB}$$

which is close to the design value of 21.75 dB.

9.5 CIRCULAR APERTURES

An antenna that has a physical aperture opening with a circular shape is said to have a *circular aperture*. Various forms of circular aperture antennas are encountered in practice. In this section, we discuss ideal circular aperture distributions with uniform and tapered amplitudes. This is followed in the next section by a study of parabolic reflector antennas that are the most popular circular aperture antennas.

9.5.1 The Uniform Circular Aperture

A general circular aperture is shown in Fig. 9-22. If the aperture distribution amplitude is constant, it is referred to as a *uniform circular aperture*. This is approximated by a circular hole in a conducting sheet with a uniform plane wave incident from behind. Suppose the aperture electric field is x-directed, or

$$\mathbf{E}_a = \hat{\mathbf{x}}E_o \quad \rho' \le a \tag{9-161}$$

Then (9-14) gives

$$\mathbf{P} = \hat{\mathbf{x}}E_o \iint\limits_{S_a} e^{j\beta\hat{\mathbf{r}}\cdot\mathbf{r}'}\,dS' \tag{9-162}$$

The general geometry for this radiation integral is shown in Fig. 2-7. It is applied to the circular aperture geometry of Fig. 9-22 using cylindrical coordinates to describe source
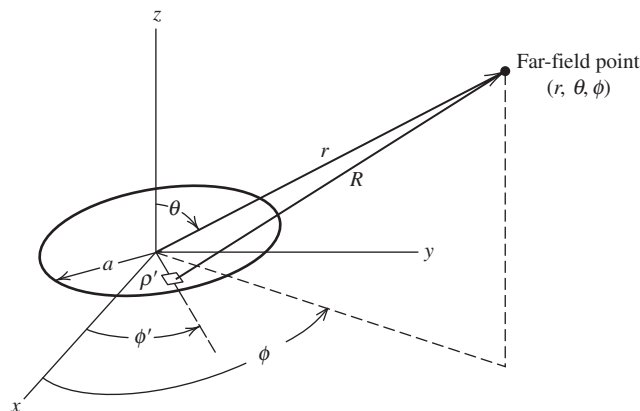
Figure 9-22 The circular aperture.

points in the xy-plane; see Appendix C.1 for a review of cylindrical coordinates. Then the vector from the origin to an arbitrary point in the aperture becomes

$$\mathbf{r}' = \boldsymbol{\rho}' = \rho' \cos \phi \, \hat{\mathbf{x}} + \rho' \sin \phi \, \hat{\mathbf{y}} \tag{9-163}$$

This with (C-4) yields

$$\begin{aligned} \hat{\mathbf{r}} \cdot \mathbf{r}' &= \rho' \sin \theta \, (\cos \phi \cos \phi' \, + \, \sin \phi \sin \phi') \\ &= \rho' \sin \theta \cos(\phi - \phi') \end{aligned} \tag{9-164}$$

Hence, (9-162) becomes

$$\begin{aligned} \mathbf{P} &= \hat{\mathbf{x}} E_o \int_0^a \left[\int_0^{2\pi} e^{j\beta\rho' \sin\theta \, \cos(\phi-\phi')} \, d\phi' \right] \rho' d\rho' \\ &= \hat{\mathbf{x}} E_0 2\pi \int_0^a \rho' J_0(\beta\rho' \, \sin\theta) d\rho' \end{aligned} \tag{9-165}$$

where (F-6) was used for the ϕ' integration. $J_0(x)$ is a Bessel function of the first kind and zero order, which is unity at $x = 0$ and is a decaying oscillatory function for increasing x. The ρ' integration can be performed using

$$\int x J_0(x) \, dx = x J_1(x) \tag{9-166}$$

which follows from (F-9). $J_1(x)$ is a Bessel function of the first kind and first order, which is zero for $x = 0$ and is a decaying oscillatory function for increasing x. Transforming variables as $x = \beta\rho' \sin\theta$ and using (9-166) in (9-165) yield

$$\mathbf{P} = \hat{\mathbf{x}} E_o \, 2\pi \frac{a}{\beta \sin\theta} J_1(\beta a \sin\theta) = \hat{\mathbf{x}} P_x \tag{9-167}$$

The equivalent magnetic current formulation of (9-26) gives

$$\mathbf{E} = (\hat{\boldsymbol{\theta}} \cos\phi - \hat{\boldsymbol{\phi}} \sin\phi \cos\phi) j\beta \frac{e^{j\beta r}}{2\pi r} P_x = \hat{\mathbf{p}} E_0 \pi a^2 j\beta \frac{e^{-j\beta r}}{2\pi r} f(\theta) \tag{9-168}$$

where the polarization unit vector is

$$\hat{\mathbf{p}} = \hat{\boldsymbol{\theta}} \cos\phi - \hat{\boldsymbol{\phi}} \sin\phi \cos\theta \tag{9-169}$$

and the relative variation of the radiation integral P_x normalized to unity maximum at $q = 0°$ is

$$f(\theta) = \frac{2J_1(\beta a \sin\theta)}{\beta a \sin\theta} \tag{9-170}$$

$f(\theta)$ is independent of ϕ due to the circular symmetry of the aperture distribution. In the E-plane, $\phi = 0°$ and (9-169) becomes $\hat{\mathbf{p}} = \hat{\boldsymbol{\theta}}$, and $f(\theta)$ represents the E_θ component. In the H-plane, $\phi = 90°$ and (9-169) reduces to $\hat{\mathbf{p}} = -\hat{\boldsymbol{\phi}} \cos\theta$, and $f(\theta)$ multiplied by $\cos\theta$ represents the E_ϕ component. This $\cos\theta$ factor ensures that the electric field goes to zero at $\theta = 90°$ as required by the boundary condition on the tangential electric field on the ground plane.

For large apertures, $f(\theta)$ gives a narrow main beam in the $\theta = 0°$ direction for the uniform phase aperture we are considering here. Thus, near the main beam, $\cos\theta = 1$ since θ is small and (9-169) gives $\hat{\mathbf{p}} = \hat{\boldsymbol{\theta}} \cos\phi - \hat{\boldsymbol{\phi}} \sin\phi$, which is the projection of the aperture electric field vector $\hat{\mathbf{x}}$ tangent to the far field sphere; see (C-1.). In this case, all θ dependence is contained in $f(\theta)$. An example of $f(\theta)$ is plotted in Fig. 9-23 in the
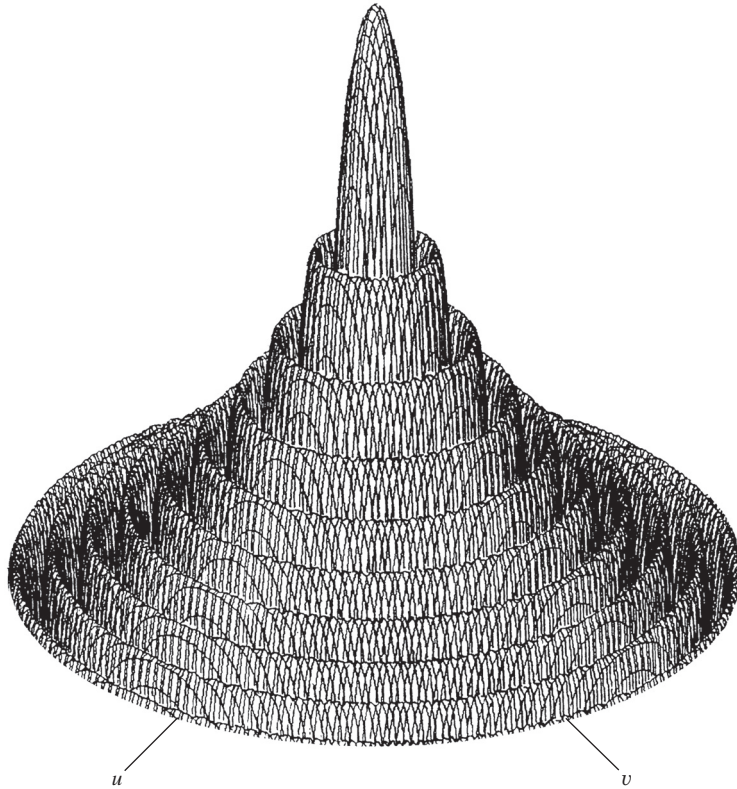
Figure 9-23 Radiation pattern of a uniform amplitude, uniform phase, 10-wavelength diameter circular aperture.
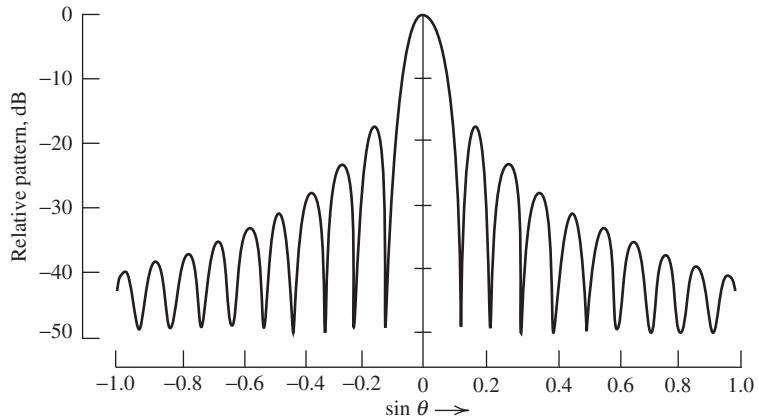

Figure 9-24 Pattern of a 10λ diameter uniform circular aperture. It is the pattern in any plane passing through the center of Fig. 9-23.

uv-plane for $a = 5\lambda$ out to the limit of the visible region $(\theta = 90°)$. A plot of the radiation pattern in any plane passing through the center of Fig. 9-23 is shown in Fig. 9-24. Note the similarity of this pattern, $2J_1(x)/x$, to the uniform line source pattern, $\sin x/x$.

The half-power point of (9-170) occurs at $\beta a \sin\theta = 1.6$, so the half-power beamwidth for $a \gg \lambda$ is

$$\text{HP} = 2\theta_{\text{HP}} = 2\sin^{-1}\frac{1.6}{\beta a} \approx 2\frac{1.6}{\pi}\frac{\lambda}{2a}$$

or

$$\text{HP} = 1.02\frac{\lambda}{2a}\text{rad} = 58.4\frac{\lambda}{2a}\text{deg} \tag{9-171}$$

For the 10λ diameter example, $\text{HP} = 0.102$ rad $= 5.84°$. The side lobe level of any uniform circular aperture pattern is -17.6 dB. This can be seen in Fig. 9-24. Since the uniform circular aperture has uniform excitation amplitude, it has unity aperture taper efficiency and the directivity, from (9-66), is

$$D_u = \frac{4\pi}{\lambda^2}A_p = \frac{4\pi}{\lambda_2}\pi a^2 \tag{9-172}$$

9.5.2 Tapered Circular Apertures

Many circular aperture antennas can be approximated as a radially symmetric circular aperture with an aperture field amplitude distribution that is tapered from the center of the aperture toward the edge. In practice, many circular aperture distributions are close to radially symmetric and do not vary with ϕ' (see Fig. 9-22). We shall assume this is the case, and again we will confine our attention to a broadside circular aperture that is large in terms of a wavelength. Then the pattern is well approximated by the unnormalized radiation integral

$$f_{\text{un}}(\theta) = \int_0^{2\pi}\int_0^a E_a(\rho')e^{j\beta\rho'\sin\theta\cos(\phi-\phi')}\rho'\,d\rho'\,d\phi' \tag{9-173}$$

Performing the integration over ϕ' with the aid of (F-6) leads to

$$f_{\text{un}}(\theta) = 2\pi\int_0^a E_a(\rho')\rho'J_0(\beta\rho'\sin\theta)\,d\rho' \tag{9-174}$$

This integral can be performed for various aperture tapers and normalized to obtain $f(\theta)$.

The properties of several common circular aperture tapers are given in Table 9-2. Similar data are available in the literature [4,12] including elliptical apertures [13]. Table 9-2 is analogous to Table 5-2 for line source distributions. The parabolic distribution ($n = 1$) of Table 9-2a provides a smooth taper from the aperture center to edge where the aperture field is zero. When $n = 0$, the distribution reduces to the uniform case where, of course, aperture taper efficiency is unity; see (9-82) and (9-172). The parabolic taper ($n = 1$) yields lower side lobes at the expense of wider beamwidth and reduced directivity compared to the uniform distribution. This effect is more pronounced for the parabolic squared ($n = 2$) distribution. The side-lobe level-beamwidth tradeoff can be customized by using the parabolic-on-a-pedestal aperture distribution in Table 9-2b. The pedestal height C is the edge (field) illumination relative to that at the center. This taper can be used to model illuminations commonly encountered with a circular reflector antenna, where the pedestal represents the fact that the feed antenna pattern is intercepted by the reflector only out to the reflector rim. Again, we observe that as the taper becomes more severe (n increases or C decreases), the side-lobe levels while the beam width increases and directivity decreases. The data in Table 9-2 provide canonical forms for use in modeling parabolic reflector antennas discussed in the next section.
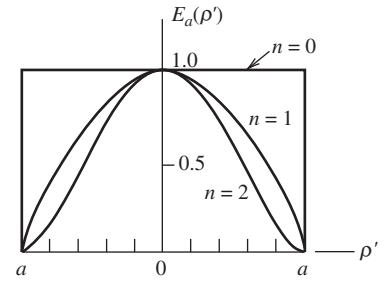
The directivity-bandwidth product DB_{cir} is found by a form of (9-90) appropriate to tapered circular apertures using $c_x c_y \to c^2 = \pi\varepsilon_t$ and $k_x k_y \to k^2$:

Table 9-2 Characteristics of Tapered Circular Aperture Distributions

a. Parabolic taper

$$E_a(\rho') = \left[1 - \left(\frac{\rho'}{a}\right)^2\right]^n$$

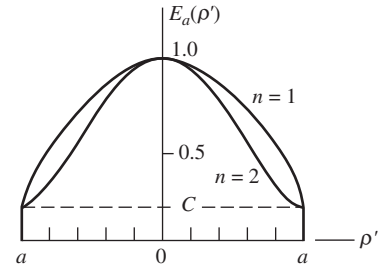$$f(\theta, n) = \frac{2^{n+1}(n+1)!J_{n+1}(\beta a \sin\theta)}{(\beta a \sin\theta)^{n+1}}$$

n	HP (rad)	Side Lobe Level (dB)	ε_t	Normalized Pattern $f(\theta, n)$	Distribution
0	$1.02\dfrac{\lambda}{2a}$	-17.6	1.00	$\dfrac{2J_1(\beta a \sin\theta)}{\beta a \sin\theta}$	Uniform
1	$1.27\dfrac{\lambda}{2a}$	-24.6	0.75	$\dfrac{8J_2(\beta a \sin\theta)}{(\beta a \sin\theta)^2}$	Parabolic
2	$1.47\dfrac{\lambda}{2a}$	-30.6	0.55	$\dfrac{48J_3(\beta a \sin\theta)}{(\beta a \sin\theta)^2}$	Parabolic squared

b. Parabolic taper on a pedestal

$$E_a(\rho') = C + (1-C)\left[1 - \left(\frac{\rho'}{a}\right)^2\right]^n$$

$$f(\theta, n, C) = \frac{Cf(\theta, n = 0) + \dfrac{1-C}{n+1}f(\theta, n)}{C + \dfrac{1-C}{n+1}}$$

Edge Illumination		$n = 1$			$n = 2$		
C_{dB}	C	HP (rad)	Side Lobe Level (dB)	ε_t	HP (rad)	Side Lobe Level (dB)	ε_t
-8	0.398	$1.12\dfrac{\lambda}{2a}$	-21.5	0.942	$1.14\dfrac{\lambda}{2a}$	-24.7	0.918
-10	0.316	$1.14\dfrac{\lambda}{2a}$	-22.3	0.917	$1.17\dfrac{\lambda}{2a}$	-27.0	0.877
-12	0.251	$1.16\dfrac{\lambda}{2a}$	-22.9	0.893	$1.20\dfrac{\lambda}{2a}$	-29.5	0.834
-14	0.200	$1.17\dfrac{\lambda}{2a}$	-23.4	0.871	$1.23\dfrac{\lambda}{2a}$	-31.7	0.792
-16	0.158	$1.19\dfrac{\lambda}{2a}$	-23.8	0.850	$1.26\dfrac{\lambda}{2a}$	-33.5	0.754

(*Continued*)

Table 9-2 Continued

-18	0.126	$1.20\dfrac{\lambda}{2a}$	-24.1	0.833	$1.29\dfrac{\lambda}{2a}$	-34.5	0.719
-20	0.100	$1.21\dfrac{\lambda}{2a}$	-24.3	0.817	$1.32\dfrac{\lambda}{2a}$	-34.7	0.690

Interpolation equations for finding HP and ε_t when C_{dB} is between -8 and -20 dB:

Quantity	$n = 1$	$n = 2$
$\mathrm{HP} = k\dfrac{\lambda}{2a}$	$k = -0.008C_{\mathrm{dB}} + 1.06$	$k = -0.015C_{\mathrm{dB}} + 1.02$
ε_t	$\varepsilon_t = 0.01C_{\mathrm{dB}} + 1.02$	$\varepsilon_t = 0.019C_{\mathrm{dB}} + 1.06$

$$\mathrm{DB}_{\mathrm{cir}} = \pi^2 \varepsilon_t k^2 \left(\frac{180}{\pi}\right)^2 = 32{,}400\,\varepsilon_t k^2 \ \ \mathrm{deg}^2 \tag{9-175}$$

For the range of values given in Table 9-2b, this product remains nearly constant, leading to

$$D \approx \frac{\mathrm{DB}_{\mathrm{cir}}}{\mathrm{HP}^2} = \frac{39{,}000}{\mathrm{HP}^2} \qquad \textit{tapered circular aperture} \tag{9-176}$$

where HP is the half-power beamwidth in all planes in degrees.

This section is closed by deriving the fundamental relation used to develop Table 9-2. The parabolic distribution used in (9-174) gives

$$f_{\mathrm{un}}(\theta) = 2\pi \int_0^a \left[1 - \left(\frac{\rho'}{a}\right)^2\right]^n \rho' J_0(\beta\rho' \sin\theta)\, d\rho' \tag{9-177}$$

The integral can be evaluated using

$$\int_0^1 (1 - x^2)^n x J_0(bx)\, dx = \frac{2^n n!}{b^{n+1}} J_{n+1}(b) \tag{9-178}$$

by letting $x = \rho'/a$ and $b = \beta a \sin\theta$. Then (9-177) reduces to

$$f_{\mathrm{un}}(\theta) = \frac{\pi a^2}{n + 1} f(\theta, n) \tag{9-179}$$

where

$$f(\theta, n) = \frac{2^{n+1}(n + 1)! J_{n+1}(\beta a \sin\theta)}{(\beta a \sin\theta)^{n+1}} \tag{9-180}$$

is the normalized pattern function. The patterns given in Table 9-2a follow from (9-180). The aperture tape efficiency is (see Prob. 9.5-4).

$$\varepsilon_t = \frac{\left[C + \dfrac{1 - C}{n + 1}\right]^2}{C^2 + \dfrac{2C(1 - C)}{n + 1} + \dfrac{(1 - C)^2}{2n + 1}} \tag{9-181}$$

9.6   REFLECTOR ANTENNAS

High-gain antennas are required for long-distance radio communication and high-resolution radar applications. Reflector systems are perhaps the most widely used high-gain antennas and routinely achieve gains far in excess of 30 dB in the microwave region. Such gains would be difficult to obtain with any other single antenna we have discussed thus far. Reflector, or "dish," antennas are seen on towers, store rooftops, home rooftops and lawns, and in many other locations. In this section, we consider the more important reflector antenna configurations with emphasis on those that have circular apertures. Before presenting the basic physics and mathematics leading to the reflector antenna design principles, we give a brief history of reflectors.

Reflector antennas have their origins in optical telescopes. Refracting optical telescopes were well known in the 17th century but suffered from aberrations. In 1663, Scottish mathematician James Gregory (1638–1675) conceived the reflecting optical telescope, but his construction attempts failed. It was Isaac Newton (1642–1727) who built the first such instrument in 1672. Also in 1672, Cassegrain[3] revealed his telescope design employing two reflectors like Gregory's but configured with the subreflector between the main reflector and focal point rather than beyond the focal point. The antenna counterparts of the Cassegrain and Gregorian reflectors are shown later in Fig. 9-32. The first reflector to be used as an antenna was the cylindrical-parabolic reflector of Hertz in 1888, shown in Fig. 1-2. The first large reflecting antenna was a 9-6-m prime focus reflector (i.e., no subreflector) antenna built by Grote Reber (1911–2002) in his backyard in Wheaton, Illinois, in 1937, which he used for radio astronomy. He may have been the first to use a waveguide cavity for a feed. He is often called the father of the large dish antenna. Reber used the instrument to confirm Jansky's initial discovery of extraterrestrial noise which was discussed in Sec. 4.3. Reber's actual antenna is now on display at the National Radio Astronomy Observatory in Green Bank, West Virginia; see [www .nrao.edu]. See [H.6: Hansen, Vol. I, Chap. 2] for more reflector history.

9.6.1   Parabolic Reflector Antenna Principles

The simplest reflector antenna consists of two components: A reflecting surface that is large relative to a wavelength and a much smaller feed antenna. The most popular form is the **parabolic reflector antenna** shown in Fig. 9-25a. The reflector (or "dish") is a paraboloid of revolution. The intersection of the reflector with any plane containing the reflector axis (z-axis) forms a curve of the parabolic type shown in Fig. 9-25b. The equation describing the parabolic reflector surface shape in the rectangular form using (ρ', z_f) is

$$(\rho')^2 = 4F(F - z_f), \quad \rho' \le a \tag{9-182}$$

The apex of the dish corresponds to $\rho' = 0$ and $z_f = F$, and the edge of the dish to $\rho' = a$ and $z_f = F - a^2/4F$. For a given displacement ρ' from the axis of the reflector, the point R on the reflector surface is a distance r_f away from the *focal point O*. The parabolic curve can also be expressed in polar coordinates (r_f, θ_f) as

$$r_f = \frac{2F}{1 + \cos\theta_f} = F \sec^2\frac{\theta_f}{2} \tag{9-183}$$

Then the projection of this distance r_f onto the aperture plane is

$$\rho' = r_f \sin\theta_f = \frac{2F \sin\theta_f}{1 + \cos\theta_f} = 2F \tan\frac{\theta_f}{2} \tag{9-184}$$

---

[3] Accurate historical information about Cassegain is not available. Even his first name is a subject of debate.
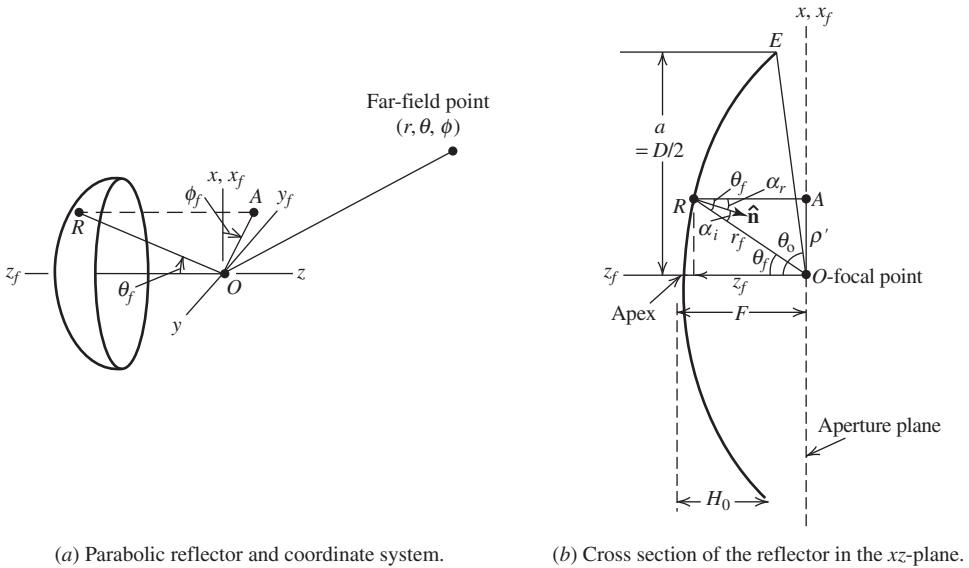
(*a*) Parabolic reflector and coordinate system. (*b*) Cross section of the reflector in the *xz*-plane.

Figure 9-25 The axisymmetric parabolic reflector antenna.

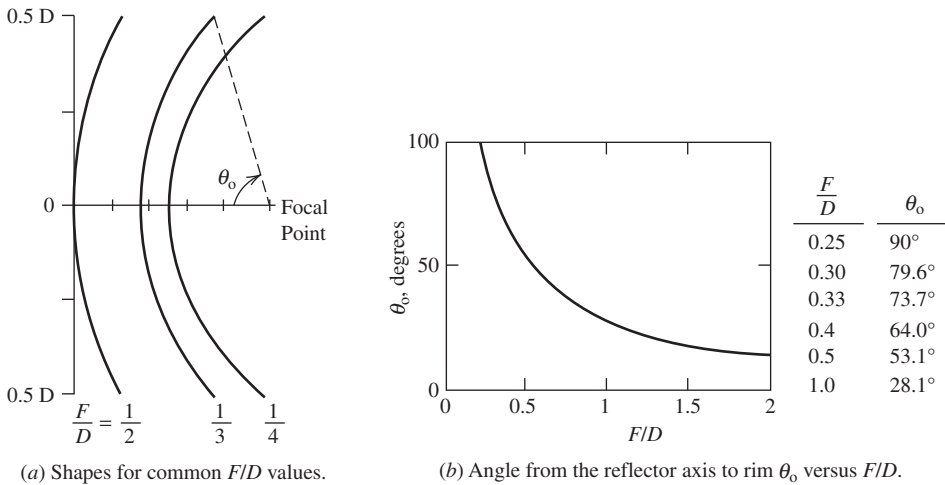(*a*) Shapes for common *F/D* values. (*b*) Angle from the reflector axis to rim θ_o versus *F/D*.

Figure 9-26 Curvatures of parabolic reflectors.

At the apex $(\theta_f = 0°)$, $r_f = F$, and $\rho' = 0$. At the reflector edge $(\theta_f = \theta_o)$, $r_f = 2F/(1 + \cos \theta_o)$ and $\rho' = a$.

The axisymmetric parabolic reflector is completely specified with two parameters, the diameter *D* and focal length *F*. Equivalently, the reflector is often stated in terms of *D* and *F/D*, which give the size and shape (curvature rate), respectively. The "focal-length-to-diameter" ratio *F/D* represents the curvature rate of the dish. In the limit as *F/D* approaches infinity, the reflector becomes planar. A flat reflector "focuses" at infinity, and a normally incident plane wave is reflected back as a plane wave (i.e., it is focused at infinity). Shapes associated with commonly used reflectors are shown in Fig. 9-26. When *F/D* is 0.25, the focal point lies in the plane passing through the rim. As indicated in Fig. 9-25*b*, the angle from the feed axis (the z_f-axis) to the reflector rim is related to *F/D* using (9-184) at point $E(\rho' = a, \theta_f = \theta_o)$ as

$$\frac{F}{D} = \frac{1}{4 \tan \dfrac{\theta_o}{2}} \tag{9-185a}$$

$$\theta_o = 2 \tan^{-1}\left(\frac{1}{4\left(\dfrac{F}{D}\right)}\right) \tag{9-185b}$$

The reflector design problem consists primarily of matching the feed antenna pattern to the reflector. The usual goal is to have the feed pattern about 10 dB down in the direction of the rim, for example, $F_f(\theta_f = \theta_o) = -10$ dB. Feed antennas with this property can be constructed for the commonly used F/D values of 0.3 to 1.0 and are discussed in Sec. 9.7. The F/D choice also impacts on cross-polarization performance, as we shall see.

The focal distance of a reflector is easily calculated using diameter D and height H_0. This practical relation is found by solving (9-182) at the rim, where $\rho' = D/2$ and $z_f = F - H_0$, giving

$$F = \frac{D^2}{16H_0} \tag{9-186}$$

For example, when $F/D = 1/4$, this gives $H_0 = D/4$; thus, $H_0 = F$, which is evident from Fig. 9-26.

The following two very important properties make the parabolic reflector useful as an antenna:

1. All rays leaving the focal point O are collimated after reflection from the reflector and the reflected rays are parallel to the reflector axis (z-axis).
2. All path lengths from the focal point to the reflector and on to the aperture plane are the same and are equal to $2F$.

The terminology used here is that of geometrical optics (GO) which treats wave propagation as rays that are normal to the equiphase surface. For a point source at the focus, the wave fronts are spherical and all rays are along r_f shown in Fig. 9-25. GO principles will now be used to verify the above two properties.

The first property follows directly from the enforcement of the law of reflection on the reflector surface; that is, $\alpha_r = \alpha_i$ in Fig. 9-25b. To show this, we first determine the surface normal $\hat{\mathbf{n}}$ by evaluating the gradient of the parabolic curve equation, $C_p = F - r_f \cos^2(\theta_f/2) = 0$, based on (9-183) in feed coordinates:

$$\begin{aligned}
\mathbf{N} = \nabla C_p &= \nabla\left(F - r_f \cos^2 \frac{\theta_f}{2}\right) \\
&= \left[\hat{\mathbf{r}}_f \frac{\partial}{\partial r_f} + \hat{\boldsymbol{\theta}}_f \frac{1}{r_f} \frac{\partial}{\partial \theta_f} + \hat{\boldsymbol{\phi}}_f \frac{1}{r_f \sin \theta_f} \frac{\partial}{\partial \phi_f}\right] C_p \\
&= -\hat{\mathbf{r}}_r \cos^2 \frac{\theta_f}{2} + \hat{\boldsymbol{\theta}}_f \cos \frac{\theta_f}{2} \sin \frac{\theta_f}{2} \tag{9-187}
\end{aligned}$$

Normalizing using $N^2 = \mathbf{N} \cdot \mathbf{N} = \cos^2 \dfrac{\theta_f}{2}$ gives

$$\hat{\mathbf{n}} = \frac{\mathbf{N}}{N} = -\hat{\mathbf{r}}_f \cos \frac{\theta_f}{2} + \hat{\boldsymbol{\theta}}_f \sin \frac{\theta_f}{2} \tag{9-188}$$

The angles between the surface normal and the incident and reflected rays are then easily found from

$$\cos \alpha_i = -\hat{\mathbf{r}}_f \cdot \hat{\mathbf{n}} = \cos \frac{\theta_f}{2} \tag{9-189a}$$

$$\cos \alpha_r = \hat{\mathbf{z}} \cdot \hat{\mathbf{n}} = (-\hat{\mathbf{r}}_f \cos \theta_f + \hat{\boldsymbol{\theta}}_f \sin \theta_f) \cdot \hat{\mathbf{n}}$$
$$= \cos \theta_f \cos \frac{\theta_f}{2} + \sin \theta_f \sin \frac{\theta_f}{2} = \cos \frac{\theta_f}{2} \tag{9-189b}$$

Comparing these two equations, we see that

$$\alpha_i = \alpha_r = \frac{\theta_f}{2} \tag{9-190}$$

proving that the law of reflection is satisfied. Also, note from Fig. 9-25b that $\alpha_i + \alpha_r = \theta_f$, which is consistent with (9-190).

The equal path length property follows from (9-183) as

$$\overline{OR} + \overline{RA} = \text{total path length from focal point to aperture}$$
$$= r_f + r_f \cos \theta_f = r_f(1 + \cos \theta_f) = 2F \tag{9-191}$$

Since the total path length is constant ($2F$), the phase of waves arriving in the aperture plane from a point source at the focus will also be constant. Thus, the *parabolic reflector with a feed that has a point phase center at the focal point will produce uniform phase across the aperture plane.* The aperture amplitude distribution, however, will not be uniform.

Reflector antennas are analyzed by tracing rays to the aperture and setting up an aperture distribution that can be integrated to obtain the far-field pattern. Alternatively, an equivalent surface current over the reflector can be integrated. In either case, GO principles are used to determine the current distribution. Application of GO requires the following to be true:

  a. The radius of curvature of the main reflector is large compared to a wavelength and the local region around each reflection point can be treated as planar.
  b. The radius of curvature of the incoming wave from the feed is large and can be treated locally at the reflection point as a plane wave.

In addition, with metallic objects we make the following assumption:

  c. The reflector acts as a perfect conductor so that the incident and reflected wave amplitudes are equal; in fact, $\Gamma = -1$.

The law of reflection applied to a reflector (e.g., see property 1 for parabolic reflectors) relies on these assumptions.

The parabolic reflector is inherently a very wideband antenna. The bandwidth of a reflector is determined at the low-frequency end by the size of the reflector; it should be at least several wavelengths in extent for GO principles to hold. At the high-frequency end, performance is limited by the smoothness of the reflector surface. Surface distortions must be much less than a wavelength to avoid phase errors in the aperture; see (9-235). In practice, the bandwidth of a reflector antenna system is usually limited by the bandwidth of the feed antenna rather than the reflector.

We now discuss techniques for analyzing reflector antennas. The techniques are not limited to reflectors of parabolic shape, but we consider only parabolic reflector cases.

GO/Aperture Distribution Method. The most basic method of analyzing reflector antennas is to use GO to determine the aperture field distribution and then find the far-field radiation pattern using the aperture theory developed in Sec. 9.5. This is done by tracing rays from the feed antenna to the aperture. First, we assume the feed is an isotropic

radiator; the influence of the radiation pattern of a real feed antenna will he included later. Since all rays from the feed travel the same physical distance to the aperture, the aperture distribution of a parabolic reflector will be of uniform phase (this is true for all frequencies). However, there is a nonuniform amplitude distribution introduced. This is due to the fact that the power density of the rays leaving the isotropic feed falls off as $1/r_f^2$ since the wave is spherical. After reflection, there is no longer any spreading loss since the rays are parallel (i.e., focused at infinity), forming a section of plane wave. Hence, the aperture field intensity varies as $1/r_f$. This is proved more formally below.

Geometrical optics (see Sec. 16.1) assumes that power density in free space follows straight-line paths. Applied to this case, the power in a conical wedge of solid angle $d\Omega$ with cross-sectional angle $d\theta_f$ as shown in Fig. 9-27 will remain confined to that conical wedge as it progresses out from the feed. After reflection, the power associated with the increment $d\theta_f$ arrives at the aperture plane in a thin ring of thickness $d\rho'$ and area dA. The power leaving the feed, assumed to be isotropic, and arriving at the aperture is proportional to $P_t\, d\Omega$, where P_t is the transmit power. This power is distributed over area dA in the aperture plane. Thus, the power density in the aperture plane varies as

$$S_a(\rho') \propto \frac{P_t\, d\Omega}{dA} \propto \frac{d\Omega}{dA} \tag{9-192}$$

since P_t is a constant. After integration over ϕ_f, $d\Omega = 2\pi \sin\theta_f\, d\theta_f$ and $dA = 2\pi\rho'\, d\rho'$. So,

$$S_a(\rho') \propto \frac{2\pi \sin\theta_f\, d\theta_f}{2\pi\rho'\, d\rho'} = \frac{\sin\theta_f}{\rho'} \frac{d\theta_f}{d\rho'} \tag{9-193}$$

From (9-184),

$$\frac{d\rho'}{d\theta_f} = \frac{d}{d\theta_f}\left(2F\tan\frac{\theta_f}{2}\right) = F\sec^2\frac{\theta_f}{2} = r_f \tag{9-194}$$

From (9-183) was used for the last equality. Then
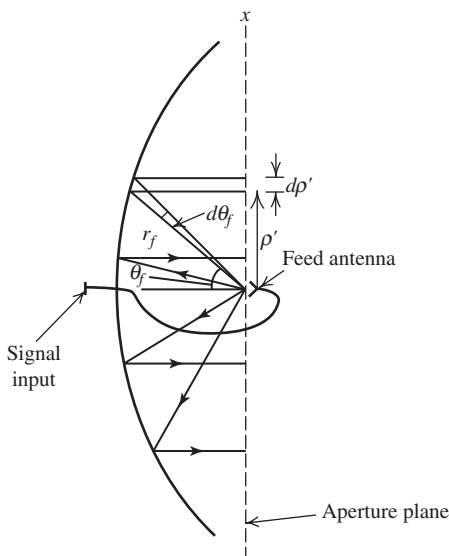
Figure 9-27 Axisymmetric, focus-fed parabolic reflector antenna in cross section.

$$\frac{d\theta_f}{d\rho'} = \frac{1}{r_f} \tag{9-195}$$

Hence, (9-192) with (9-184) and (9-195) becomes

$$S_a \propto \frac{\sin\theta_f}{r_f \sin\theta_f}\frac{1}{r_f} = \frac{1}{r_f^2} \tag{9-196}$$

This proves the spherical wave nature of the feed radiation and is referred to as *spherical spreading loss*. And since $E_a \propto \sqrt{S_a}$,

$$E_a(\theta_f) \propto \frac{1}{r_f} \tag{9-197}$$

Thus, there is a natural amplitude taper in the aperture caused by the curvature of the reflector.

If the primary (or feed) antenna is not isotropic, the effect of its normalized radiation pattern $F_f(\theta_f, \phi')$ can be included, using the coordinate system of Fig. 9-25 as

$$\mathbf{E}_a(\theta_f, \phi') = V_o e^{-j\beta 2F}\frac{F_f(\theta_f, \phi')}{r_f}\hat{\mathbf{u}}_r \tag{9-198}$$

where $\hat{\mathbf{u}}_r$, is the unit vector of the aperture electric field. The phase shift associated with the $2F$ path length from the focal point to the aperture plane is also included. The coordinates ρ' and ϕ' are appropriate for describing the aperture electric field. Feed coordinates r_f and θ_f are expressed in terms of the aperture coordinate ρ' using

$$r_f = \frac{4F^2 + \rho'^2}{4F} \tag{9-199a}$$

$$\theta_f = 2\tan^{-1}\frac{\rho'}{2F} \tag{9-199b}$$

which follow from (9-183) and (9-184). These transformations can be used with (9-198) to obtain the aperture distribution at points (ρ', ϕ') from the feed antenna radiation pattern F_f. It remains only to find the polarization of the aperture electric field vector by determining $\hat{\mathbf{u}}_r$ in (9-198). This follows by using the approximations that at the point of reflection, the reflector behaves as if planar [assumption (a)] and is perfectly conducting [assumption (c)]. Then the tangential component of the total electric field formed by the sum of the incident and reflected wave electric fields, $\mathbf{E_i} + \mathbf{E_r}$, is zero at the reflector. The law of reflection requires that $\hat{\mathbf{n}}$ bisect the incident and reflected rays; then, $\mathbf{E_i} + \mathbf{E_r} = 2(\hat{\mathbf{n}} \cdot \mathbf{E_i})\hat{\mathbf{n}}$, or

$$\mathbf{E_r} = 2(\hat{\mathbf{n}} \cdot \mathbf{E_i})\hat{\mathbf{n}} - \mathbf{E_i} \tag{9-200}$$

Since the reflector is assumed to be perfect [assumption (c) above], $|\mathbf{E_r}| = |\mathbf{E_i}|$; using this to normalize the above equation gives

$$\hat{\mathbf{u}}_r = 2(\hat{\mathbf{n}} \cdot \hat{\mathbf{u}}_i)\hat{\mathbf{n}} - \hat{\mathbf{u}}_i \tag{9-201}$$

where $\hat{\mathbf{u}}_r = \mathbf{E_r}/|\mathbf{E_r}|$ and $\hat{\mathbf{u}}_i = \mathbf{E_i}/|\mathbf{E_i}|$.

We can now write the radiation pattern function for the entire reflector system. The Fourier transform of the aperture distribution follows from (9-164) and (9-198) in (9-14):

$$\mathbf{P} = V_o \int_0^{2\pi} \int_0^a \frac{F_f(\theta_f, \phi')}{r_f} \hat{\mathbf{u}}_r \, e^{j\beta\rho' \, \sin\theta \, \cos(\phi-\phi')} \rho' \, d\rho' \, d\phi' \qquad (9\text{-}202)$$

for a circular projected aperture of radius $a = D/2$. The complete radiation pattern then follows from (9-29). A single reflector fed from a feed antenna at the focal point is often called a *focus-fed* or *prime-focus* reflector antenna. The feed is the *primary antenna* and the reflector forms the *secondary antenna*. The feed pattern is then the *primary pattern* and the pattern from the antenna system as a whole is called the *secondary pattern*.

The uniform aperture phase associated with the GO formulation and the use of a real-valued feed pattern function $F_f(\theta_f, \phi')$ leads to a symmetric pattern function since the Fourier transform of a real function is symmetric. Thus, the GO formulation always renders a symmetric far-field pattern. However, for general situations, such as offset reflectors to be discussed in Sec. 9.6.3, reflector antennas have asymmetric patterns. A more accurate analysis technique is now introduced for this purpose.

PO/Surface Current Method. The theory developed in Sec. 2.4 indicates that we can integrate over a current distribution to obtain the far field. For reflector antennas, we use the current on the metallic reflector generated by the incident fields from the feed antenna. Using the general expression for the magnetic vector potential of (2-101) appropriate to a surface current in the general far-field electric field expression of (2-105) gives

$$\mathbf{E} = -j\omega\mu \frac{e^{-j\beta r}}{4\pi r} \iint\limits_{S_r} [\mathbf{J_s} - (\mathbf{J_s} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}] e^{j\beta\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}} \, dS' \qquad (9\text{-}203)$$

where S_r is the surface of the reflector. This approach, of course, is viable only if the surface current $\mathbf{J_s}$ is known. The current is found using the *physical-optics* (PO) approximation that makes use of assumptions (a) to (c) used in GO analysis to relate the surface current to the incident field from the feed. That is, the wave arriving from the feed behaves locally as a plane wave and reflects from a locally-plane reflector that behaves as a perfect reflector. Then the incident magnetic field from the feed $\mathbf{H_i}$ and the magnetic field associated with the reflected wave $\mathbf{H_r}$ are related to the surface current from (2-26) as $J_s = H_{\text{tan}}$, where H_{tan} is the tangential component of the total magnetic field that is given by $\hat{\mathbf{n}} \times (\mathbf{H_i} + \mathbf{H_r})$. But for a perfect conductor, the reflected magnetic field equals the incident magnetic field, so we are led to

$$\mathbf{J_s} = \begin{cases} 2\hat{\mathbf{n}} \times \mathbf{H_i} & \text{over the front of the reflector} \\ 0 & \text{on the shadowed side of the reflector} \end{cases} \qquad (9\text{-}204)$$

Thus, the PO approximation interprets the reflector surface S_r as having a nonzero current only over portions illuminated by the feed using ray tracing. Also, the discontinuity at the rim of the reflector separating the illuminated and shadowed regions is neglected. This effect as well as direct radiation and blockage/scattering effects due to the feed assembly can be treated separately.

The integral in (9-203) can be evaluated analytically for arbitrary symmetrical reflectors. Known as *Rusch's method,* it ushered in the era of modern reflector antenna analysis and remains the most popular approach for reflector analysis [H.8.5: Scott]. However, the integration is usually performed numerically [14]. In addition, a Jacobian transformation is usually employed to evaluate the integral using aperture plane coordinates, avoiding direct integration over the curved reflector surface [H.6: Lo & Lee, Eds., Chap. 15]. Series expansions are also employed for efficient integral evaluation [15]. A number of codes are available to evaluate reflector antenna performance. The powerful code GRASP [Appendix G.2] includes PO evaluation. Web PRAC (Parabolic Reflector Antenna Code) is a user-friendly web-based program, and is described in Appendix G.

For axisymmetric reflectors, the GO/aperture integration (GO/AI) and PO/surface integration (PO/SI) methods yield identical results [16]. However, the two methods, as conventionally applied to offset reflectors, yield different results, with the GO/AI method being slightly inferior. In addition, the pattern accuracy of both formulations degrades beyond the main beam and first few side lobes; the pattern effects in this region are dominated by diffraction from the reflector rim.

Diffraction Effects. The GO/AI and PO/SI techniques just discussed produce accurate results for the main beam and first few side lobes. The pattern in the far-out side-lobe region is found by including diffraction (i.e., scattering) from the rim of the reflector (and subreflector in dual systems) and any other sharp edges. This is done by augmenting GO with diffraction effects through the use of the geometrical theory of diffraction (GTD) or by augmenting PO with a fringe current on the rim using the physical theory of diffraction (PTD). GTD and PTD are discussed in Chap. 16 and an example of diffraction from a reflector is treated in Sec. 16.18.

The level of back-directed radiation is quantified with *front-to-back* ratio (*F/B*) which is the ratio of the main beam peak to the pattern value 180° away. For good reflector, this is greater than 40 dB.

9.6.2   The Axisymmetric Parabolic Reflector Antenna

In this and the remaining subsections, we discuss commonly used reflector geometries and their properties. The most popular reflector antenna is the focus-fed, axisymmetric parabolic reflector illustrated in Fig. 9-27. The feed is located at the focal point and its main beam peak is directed toward the reflector center. Usually, the feed is some type of horn antenna as discussed in Sec. 9.7. A simple dipole feed can be used at UHF frequencies and below.

Consider a feed antenna that is linearly polarized along the x-axis (coincident with the x_f-axis) and pointed toward the reflector apex with E- and H-plane patterns $C_E(\theta_f)$ and $C_H(\theta_f)$, respectively. The aperture field produced by a feed that is represented by (9-239a) is found from (9-198) and (9-201) with (9-188) as

$$\mathbf{E}_a = V_o \frac{e^{-j\beta 2F}}{r_f} \{-\hat{\mathbf{x}}[C_E(\theta_f)\cos^2\phi_f + C_H(\theta_f)\sin^2\phi_f]$$
$$+ \hat{\mathbf{y}}[C_E(\theta_f) - C_H(\theta_f)]\sin\phi_f\cos\phi_f\} \tag{9-205}$$

If the feed is balanced (i.e., has a rotationally symmetric normalized pattern $F_f(\theta_f)$), (9-205) simplifies to $\mathbf{E}_a \sim -\hat{\mathbf{x}}F_f(\theta_f)$. So, cross-polarization (i.e., y-polarized field content) arises from the imbalance between the E- and H-plane copolarized patterns, C_E and C_H. Eq. (9-205) also shows that cross-polarization in the aperture is maximum in the $\phi_f = 45°$, 135° plane. This aperture cross-polarization causes far-field cross-polarization. The aperture distribution of (9-205) can be integrated to find the radiation. It is instructive to examine (9-205) for an x_f-polarized short dipole, which has $C_E = \cos\theta_f$ and $C_H = 1$:

$$\mathbf{E}_a = V_o \frac{e^{-j\beta 2F}}{r_f} \{-\hat{\mathbf{x}}(\cos\theta_f\cos^2\phi_f + \sin^2\phi_f)$$
$$- \hat{\mathbf{y}}(1 - \cos\theta_f)\sin\phi_f\cos\phi_f\} \tag{9-206}$$

The field components are shown in Fig. 9-28. The bracketed expression reduces to $-\hat{\mathbf{x}}\cos\theta_f$ in the E-plane ($\phi_f = 0°$) and $-\hat{\mathbf{x}}$ in the H-plane ($\phi_f = 90°$). Thus, the aperture electric field is polarized parallel to the short dipole feed in the principal planes. Note that the aperture fields are inverted relative to the incident fields due to the reflection process. For nonprincipal planes, field components orthogonal to that of the feed
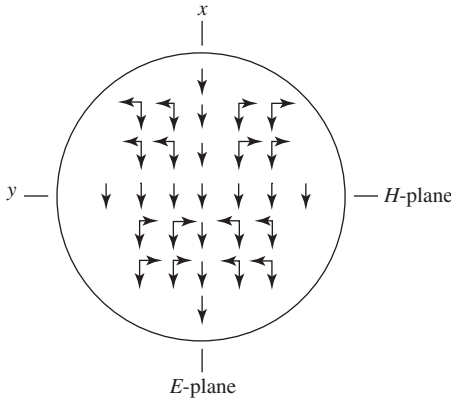
Figure 9-28 Electric field distribution in the aperture of a parabolic reflector for an x-polarized short dipole feed antenna. The electric field is decomposed into its x- and y-components. See (9-206).

(e.g., y-components) are present, giving cross-polarization. These cross-polarization properties are also true for the reflector radiation. The cross-polarization content in the aperture fields cancels in principal planes. This follows from the opposite phase of the cross-polarized components on opposite halves of the aperture. However, in nonprincipal planes, complete cancellation in the far field does not occur. The largest cross-polarized components introduced by the reflector occur in the 45° planes. We also note a very important conclusion: As *F/D increases, cross-polarization decreases.* This follows by first noting from (9-185b) that as F/D increases, the maximum feed angle $\theta_f = \theta_o$ decreases and thus the second term in the (9-206) decreases, leading to reduced cross-polarization. In the limit of large F/D, the reflector becomes flat and does not introduce cross-polarization.

Fig. 9-29 presents pattern data computed with the GRASP commercial reflector antenna code [Appendix G] using physical optics, surface current integration. A half-wave dipole, as is frequently encountered in practice, is used as the feed and behaves similar to the short dipole discussed above. Cross-polarization for reflectors is defined as ratio of the cross-polarization relative to the co-polarized pattern peak value. The peak cross-polarization is denoted as XPOL. Note the lack of cross-polarization in the principal planes in Fig. 9-29 and a cross-polar peak of XPOL = −26.3 dB in the $\phi = 45°$ plane. Ignoring the polarization vector and substituting (9-199) into (9-198) give an expression for the normalized *aperture illumination* in aperture coordinates:

$$E_{an}(\rho') = F_f\left(\theta_f = 2\tan^{-1}\frac{\rho'}{2F}\right)\frac{1}{1 + \left(\dfrac{\rho'}{2F}\right)^2} \tag{9-207a}$$

$$= 20\ \log|F_f| - 20\ \log\left[1 + \left(\frac{\rho'}{2F}\right)^2\right]\ \text{[dB]} \tag{9-207b}$$

This permits direct evaluation of the electric field variation as a function of aperture radius ρ'. The second factor in (9-207a) is called the spherical spreading factor and represents the fact that the distance r_f from the focal point to the reflector increases with ρ'. As noted earlier, rays leaving the feed at the focal point spread out in all directions, leading to r_f^{-1} field variation. After reflection from the main reflector, the rays are collimated and no longer experience amplitude decay. The exact form of the aperture distribution has less influence on the pattern and directivity of a reflector than the *edge illumination EI*, or *edge taper ET*, that are found from (9-207b) for $\rho' = a$ as

$$\text{EI} = 20\ \log[E_{an}(\rho' = a)] = -\text{FT} - L_{sph} = -\text{ET}\quad\text{[dB]} \tag{9-208}$$

where

EI $\;=\;$ edge illumination (dB) $= 20 \log C$

ET $\;=\;$ edge taper (dB) $= -$EI

FT $\;=\;$ feed taper (at aperture edge) (dB) $= -20 \log[F_f(\theta_o)]$

$L_{\text{sph}} =$ spherical spreading loss at the aperture edge (dB)

$$= 20 \log \left[1 + \frac{1}{16 \left(\dfrac{F}{D} \right)^2} \right] = -20 \log \left[\frac{1 + \cos \theta_o}{2} \right]$$
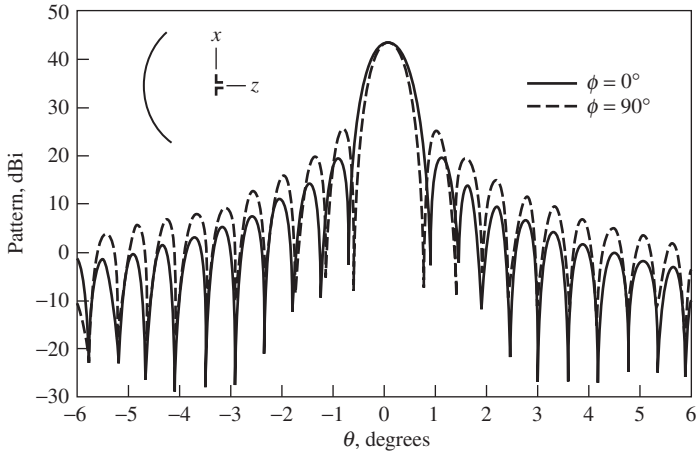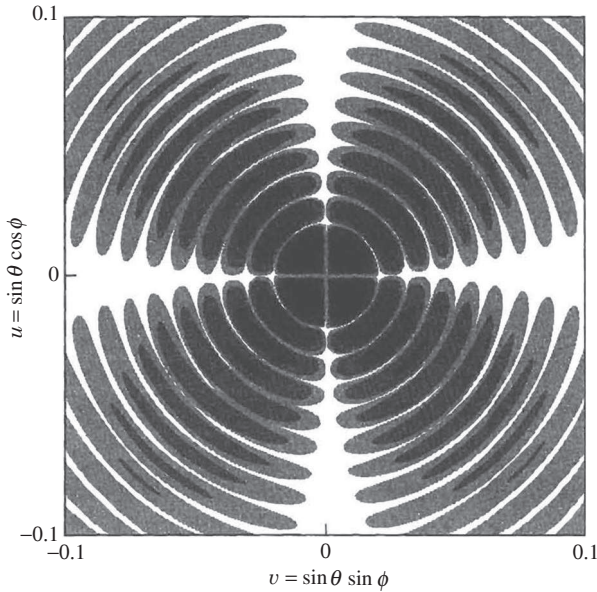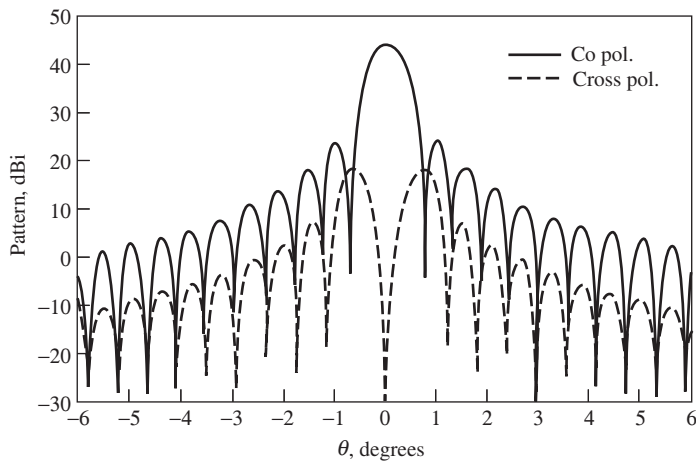

(a) Principal plane patterns.
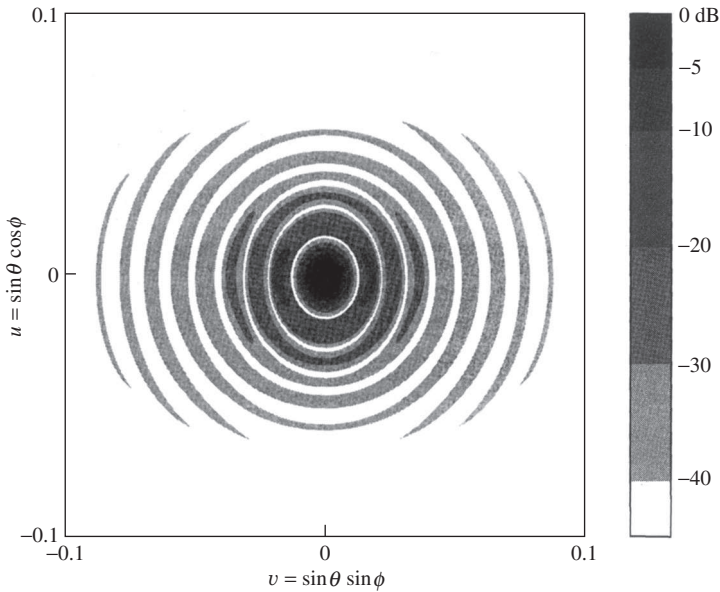
(b) Co-polar contours (normalized).

Figure 9-29 An axisymmetric parabolic reflector with diameter $D = 100\lambda$ and $F/D = 0.5$ fed by a half-wave dipole located at the focus. All data were computed using GRASP. (Reproduced by permission from [H.11.1: Stutzman]. © 1993 by Artech House, Inc., Norwood, MA.)

(c) Cross-polarization contours (normalized).

(d) Co- and cross-polarization patterns in $\phi = 45°$, $135°$ plane.
The peak cross polarization is 26.3 dB below the co-polar peak.

Figure 9-29 (continued)

The above expression for L_{sph} shows that F/D influences the amount of spherical spreading loss. It varies from 0.5 to 6.0 dB for F/D, ranging from 1.0 down to 0.25; see Fig. 9-26.

Reflector antenna performance can be estimated by a simple process. First, the aperture distribution is obtained using (9-207). Next, a canonical distribution, such as presented in Table 9-2, is selected so that it approximates the aperture distribution. Then the performance parameters of Table 9-2 such as HP, SLL, and ε_t are evaluated. Interpolation can be used for intermediate values. This canonical distribution method is illustrated in Example 9-8. The approach is described in detail in [4], which also contains useful data for canonical distributions; however, [4] failed to include spherical spreading loss.

EXAMPLE 9-8   *A 28-GHz Parabolic Reflector Antenna*

An axisymmetric parabolic reflector antenna 1.22 m (4 ft) in diameter was used at Virginia Tech to receive signals at 28.56 GHz ($\lambda = 1.05$ cm) from a geostationary satellite. It has an $F/D = 0.5$. It is constructed of epoxy fiberglass with a reflective metal layer and has an rms surface accuracy of 0.2 mm. The feed antenna is a circular corrugated horn positioned at the focal point and supported with four thin-profile spars. The feed pattern is slightly asymmetric, but analysis using a canonical distribution yields good results. We assume the feed to be rotationally symmetric and equal to the measured *E*-plane feed pattern with the following beamwidths:

$$HP = 56°, \qquad BW_{-10dB} = 104° \tag{9-209}$$

The angle from the center to the edge of the reflector from (9-185b) is $\theta_o = 53.1°$. The 10-dB down point will fall inside the reflector since $BW_{-10dB}/2 = 52°$. We assume the feed pattern has fallen to 11 dB down at the rim, that is, FT = 11 dB. The spherical spreading loss from (9-208b) is $L_{sph} = 1.9$ dB. The edge taper from (9-208a) is ET = $11 + 1.9 = 12.9$ dB. The corresponding edge illumination $EI = -12.9$ dB (C = 0.2265) falls between the values in Table 9-2. Linearly interpolating for the parabolic-squared taper yields the following results:

$$HP = 1.214\frac{\lambda}{D} = 0.599° \quad (0.605° \text{ measured}) \tag{9-210a}$$

$$SLL = -30.5 \text{ dB} \quad (-28.5 \text{ dB measured}) \tag{9-210b}$$

Note the very good agreement to measured values. The complete pattern using $f(\theta, n, C)$ from Table 9-2b is plotted in Fig. 9-30. Also shown for comparison is the measured pattern; it is identical to the computed pattern over the main beam. The gain is evaluated in Prob. 9.6-6.

9.6.3 Offset Parabolic Reflectors

The blockage of the main reflector aperture by the feed assembly and associated support structure can be reduced or eliminated by using the offset reflector shown in Fig. 9-31.
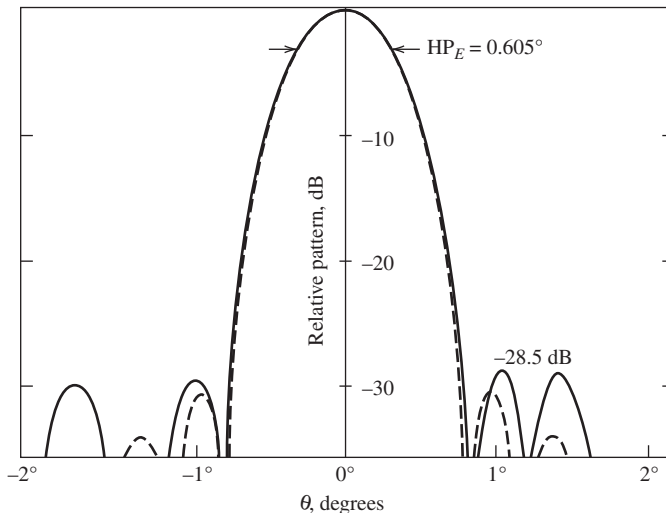
Figure 9-30 Measured (solid) *E*-plane pattern for the 1.22-m-diameter axisymmetric parabolic reflector at 28.56 GHz in Example 9-8. The computed (dashed) pattern is for a parabolic-squared circular aperture distribution on a −12.9-dB pedestal.
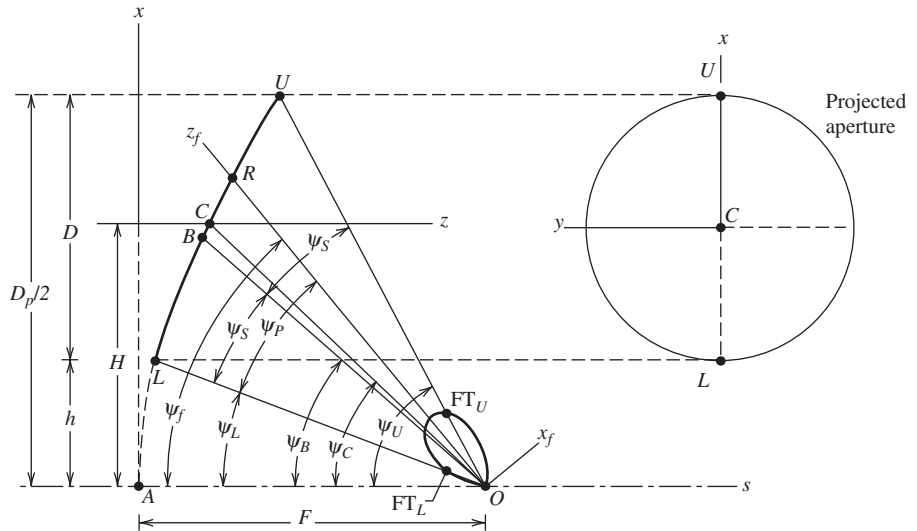
Figure 9-31 Geometry of the offset parabolic reflector of diameter D and focal length F. The axis of symmetry s bisects the parent parabolic curve of diameter D_p. Note that the axisymmetric case occurs when $H = 0$.

The properties of the offset reflector are similar to the axisymmetric counterpart formed by using the diameter of the *parent reflector D_p*. That is, the appropriate focal-length-to-diameter ratio to use for electrical performance is F/D_p. This degrades cross-polarization performance because $F/D_p < F/D$ for the offset reflector and cross-polarization levels rise with decreasing F/D_p for an unbalanced feed pointed toward the apex, that is, $\theta_f = 0°$. More will be said about cross-polarization in Sec. 9.6.5.

The analysis techniques explained in the previous section for axisymmetric geometries are general and are also used for offset reflectors. The GO/AI and PO/SI methods yield identical results if the integration surface is chosen to cap the reflector [16]. This is the natural choice for axisymmetric reflectors where the integration plane is selected to coincide with the physical aperture plane that contains the rim of the reflector, thus capping it. However, if the projected aperture of an offset reflector (as shown in Fig. 9-31) is used for the integration surface, GO/AI and PO/SI yield different results [17]. PO/SI is thought to produce more accurate patterns, especially for cross-polarization. In addition, the pattern accuracy of both formulations degrades beyond the main beam and first few side lobes. Pattern effects in this region are dominated by diffraction from the reflector rim; see Sec. 16.18.

Pencil beams are required in communication applications for high gain and in remote sensing for scene resolution. Offset reflectors are used not only to produce pencil beam patterns, as discussed so far, but also for *contoured beams*. Offset reflectors produce contour beams by using an array of feed horns or by shaping the main reflector (e.g., using a non-parabolic shape). Offset reflectors avoid blockage caused by hardware in the feed region created by a cluster of feed horns. A popular application for contour beams is on geostationary satellites that have antennas which produce a *footprint* conforming to a desired Earth region such as a country or continent. The multiple feed antennas in the focal region, each creating a scanned beam according to the displacement from the focal point, are combined with amplitude and phase weighting to produce a custom-shaped main beam.

9.6.4   Dual Reflector Antennas

A subreflector can be introduced between the feed and main reflector of a single reflector antenna to form a *dual reflector*. The most popular dual reflector is the axisymmetric
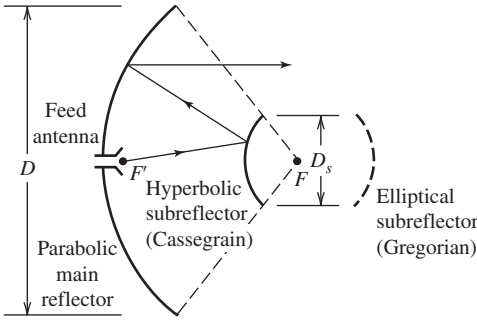
Figure 9-32 Classical axisymmetrical dual reflectors. The main reflector is parabolic and the subreflector is hyperbolic (elliptical) for the Cassegrain (Gregorian) reflector system.

Cassegrain reflector antenna shown in Fig. 9-32. The main reflector is parabolic and the subreflector is hyperbolic. This geometry again produces a focused system. That is, rays associated with an incoming plane wave parallel to the axis of symmetry reflect from both reflectors and intersect at a point, the focal point F'. The *virtual focal point F* shown in Fig. 9-32 is the point from which transmitted rays appear to emanate with a spherical wave front after reflection from the subreflector. That is, the feed is mirrored in the subreflector.

A second form of the dual reflector antenna that offers perfect focusing is the *Gregorian reflector*. It has a concave rather than a convex subreflector that is located beyond the virtual focal point as shown in Fig. 9-32 and has an elliptical cross section. Both Cassegrain and Gregorian systems have their origins in optical telescopes and are named for their inventors. The subreflector for the Gregorian system being more distant from the main reflector requires more support structure. Both types of dual reflectors offer the major advantage of having the feed conveniently located near the apex of the main reflector. This provides convenient access to the feed region, reduces the support problem for feed hardware, and eliminates the long transmission line run, and associated losses, often used to reach the focal region of a prime focus reflector. Another advantage of dual reflectors over single reflectors is in low-noise Earth terminal applications. The feed radiation not intercepted by the subreflector of a dual reflector (e.g., spillover) is directed toward the low-noise sky region rather than the more noisy ground as seen by the spillover of a single reflector.

The subreflector shapes used in the classical dual reflector configurations are described by conic sections. Fig. 9-33 gives the geometry of the subreflectors in subreflector coordinates x_s and z_s; the complete subreflector surface is obtained by rotating the curve about the z_s-axis. The subreflector is determined by its diameter D_s and *eccentricity e*. The shape is controlled by the eccentricity, which is defined as defined as

$$e = \frac{c}{a} \begin{cases} > 1 & \text{hyperbola (Cassegrain)} \\ < 1 & \text{ellipse (Gregorian)} \end{cases} \tag{9-211}$$

Example shapes are $e = \infty$, planar; $e = 0$, circle (sphere); and $e = 1$, parabola. The equations of the subreflector surfaces are given by

$$\frac{z_s^2}{a^2} - \frac{x_s^2}{b^2} = 1 \qquad b^2 = c^2 - a^2 \qquad \text{hyperbola}$$

$$\frac{z_s^2}{a^2} + \frac{x_s^2}{b^2} = 1 \qquad b^2 = a^2 - c^2 \qquad \text{ellipse} \tag{9-212}$$

The distances c and a are shown in Fig. 9-33. The required hyperbolic shape will be proved for the Cassegrain dual reflector. This derivation also illustrates how the subreflector operates.
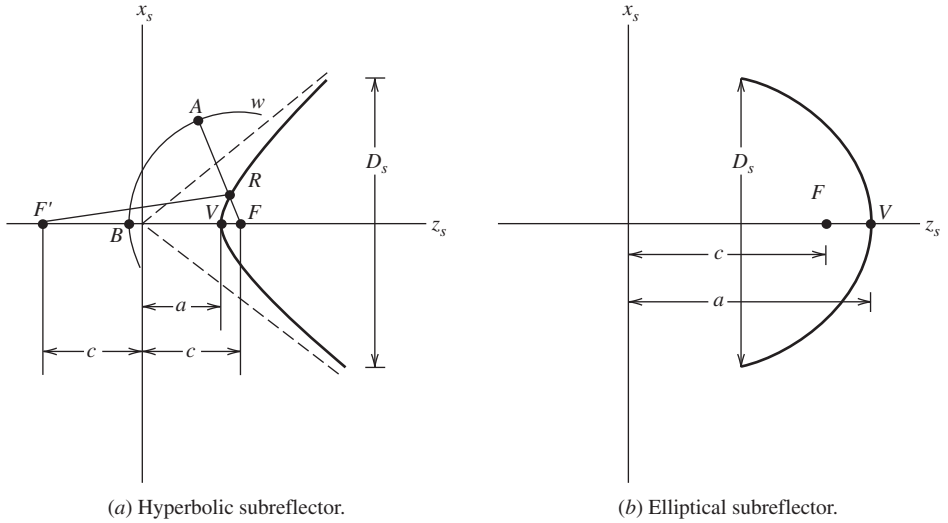
(a) Hyperbolic subreflector.

(b) Elliptical subreflector.

Figure 9-33 Geometry of classical subreflectors.

The function of the hyperbolic subreflector is to convert the incoming wave from a feed antenna located at the focal point F' to a spherical wave front w that appears to originate from the virtual focal point F. For this to be true, the optical path (total distance) from F' to wavefront w must be constant. Enforcement of this condition determines the subreflector shape. As seen from Fig. 9-33a, the total distance from F' to A including reflection from the subreflector is

$$\overline{F'R} + \overline{RA} = \overline{F'V} + \overline{VB} = c + a + \overline{VB} \tag{9-213}$$

But

$$\overline{RA} = \overline{FA} - \overline{FR} = \overline{FB} - \overline{FR} \tag{9-214}$$

where $\overline{FB} = \overline{FA}$ was based on the exiting wave being spherical. This in (9-213) leads to

$$\overline{F'R} - \overline{FR} = c + a - (\overline{FB} - \overline{VB}) = c + a - (c - a) = 2a \tag{9-215}$$

This result coincides with the following definition of a hyperbola: A hyperbola is the locus of a point that moves such that the difference of its distances, $\overline{F'R} - \overline{FR}$, from two fixed points, F' and F, is equal to a constant, $2a$.

Dual reflectors can be modeled with a single *equivalent parabolic reflector* as shown in Fig. 9-34 for the axisymmetric Cassegrain reflector. The equivalent parabola has the same diameter $(D_e = D)$, but a focal length (F_e) longer than that of the main reflector (F):

$$F_e = \frac{e + 1}{e - 1} F = MF \tag{9-216}$$

where M is called the *magnification*. From (9-211), $e > 1$ for the hyperbolic subreflector of a Cassegrain reflector, so $M > 1$ and $F_e > F$. This increased effective focal length provides several advantages, which are noted by examining the equivalent (single) parabolic reflector with diameter D and focal length F_e, and thus larger focal-length-to-diameter ratio than the actual dual reflector. First, as noted in Sec. 9.6.2, cross-polarization in the far-field pattern improves with larger focal length-to-diameter ratio. Second, there is less spherical spreading loss at the rim of the main reflector; see (9-207). Finally, main beam scan performance is improved. This follows because the larger the focal-length-to-diameter ratio is,
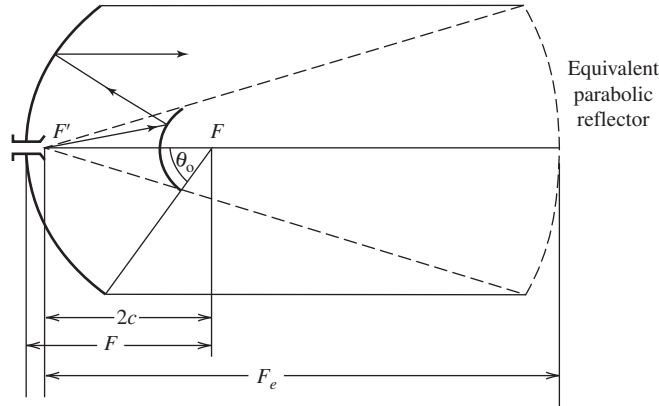
Figure 9-34 The equivalent single parabolic reflector for a Cassegrain dual reflector.

the less the radiation pattern deteriorates as the feed antenna is laterally displaced (in the $x_s y_s$-plane). This is explained by examining the limiting case of an infinite focal length-to-diameter ratio case (e.g., a flat main reflector) in which no deterioration occurs for reflection off normal incidence. The equivalent parabola concept applies to dual offset reflectors also [18].

In a single reflector antenna system, the phase front from the feed antenna is converted to the desired exiting phase front. This is usually a spherical to planar conversion as accomplished with a parabolic reflector. Limited aperture amplitude control is accomplished through feed taper and F/D; see (9-208). This is true for traditional single and dual reflectors with a parabolic main reflector. However, if both reflector shapes in dual reflectors are allowed to be "shaped," both the aperture amplitude and phase can be controlled. In the usual synthesis case, the subreflector has a highly tapered illumination to reduce spillover and is shaped to spread the reflected rays out for uniform amplitude. This requires shaping of the main reflector. The topic of dual reflector synthesis is rather advanced, but can be understood with the following simplified explanation [19–22]. In principle, the shapes of both the main reflector and the subreflector can be determined exactly for axisymmetric systems [21] and for offset dual reflectors [22]. As in the above-mentioned case, the shapes of dual reflectors can be determined to yield uniform aperture amplitude and phase, giving maximum aperture utilization. The design problem is to convert the rather broad feed antenna radiation pattern to a nearly uniform amplitude and phase aperture distribution while keeping spillover acceptable. The concept is to under-illuminate the subreflector in order to reduce its spillover and then increase its curvature over classical shape to direct reflected rays to edges of the main reflector. Within the limit of geometrical optics, main reflector spillover is avoided with proper reflector sizing. Shaping of the subreflector changes the total path length from the feed to the aperture. But that can be nearly compensated for by shaping the main reflector to correct for phase errors introduced by subreflector shaping. The amount of main reflector shape change is on the same order as subreflector shape change because both introduce about the same phase error. However, subreflector shaping almost completely controls the aperture amplitude distribution. This sequential shaping does not yield an exact solution but avoids the difficult mathematics associated with the exact approach, which requires simultaneous solution for the reflector shapes.

Dual-shaped reflectors are in wide use for high-gain axisymmetric systems such as satellite communications Earth terminals with main reflectors larger than a few meters in diameter. Shaped offset dual reflectors for smaller systems are also being used. One of the highest reported measured aperture efficiencies (85%) was achieved with a 1.5-m main reflector in an offset dual shaped reflector configuration operated at 31 GHz. [23]

Theoretical predictions for larger main reflectors indicate that more than 90% efficiency is possible. [23] Dual reflectors can also be designed for low side lobes. [24]

For the synthesis of dual-shaped reflectors, it is common to use geometrical optics (GO) during the design process. The previous discussion of shaped reflectors for high efficiency is a good example. After the reflector shapes are determined using GO-based synthesis, a computer code is then employed to accurately determine performance parameters such as gain, side-lobe level, and cross-polarization. There are two popular approaches for the computer analysis of dual reflectors: GTD/PO that uses the geometrical theory of diffraction (GTD is treated in Chap. 16) on the subreflector and PO on the main reflector, and PO/PO that uses PO on both the subreflector and main reflector. For small subreflectors less than about 10λ, PO/PO is thought to give more accurate results. GTD/PO is usually used for electrically large reflector systems because PO/PO computations are very time consuming.

9.6.5   Cross-Polarization and Scanning Properties of Reflector Antennas

Reflectors are used in many situations, including dual polarized operation that permits two communication channels on the same frequency. This requires the two polarizations to be nearly orthogonal. A second popular use for reflectors is to scan a single main beam or to form multiple narrow beams with the same large main reflector by displacing the feed(s) from the focal point. Both cross-polarization and beam scanning are discussed in this section.

Analysis of the axisymmetric parabolic reflector in Sec. 9.6.2 showed that for a purely polarized but unbalanced feed, such as a dipole, reflector-induced cross-polarization is zero in the principal planes and is maximum in the 45° planes; see Fig. 9-29. A balanced feed with a rotationally symmetric pattern (see Sec. 9.6.2) positioned with its perfect phase center at the focus produces no far-field cross-polarization based on GO analysis. However, PO analysis of an axisymmetric reflector with a balanced feed does yield a small cross-polarized pattern. In practice, the feed is usually responsible for the dominant contribution to cross-polarization from axisymmetric reflectors.

As noted in Sec. 9.6.2, the cross-polarization of axisymmetric reflectors decreases with increasing *F/D*. That is, as the dish curvature reduces, less cross-polarization is introduced. This is easily remembered because as *F/D* becomes larger, the reflector becomes flatter and a flat reflector does not depolarize. If the feed antenna pattern is rotationally symmetric (e.g., balanced) and purely polarized, cross-polarization is reduced significantly, but a residual level remains due to axial (*z*-directed) currents on the reflector.

Cross-polarization behavior for offset reflectors is more complex. The general geometry is shown in Fig. 9-31. As explained in Sec. 9.6.3, the parent reflector diameter must be used for electrical performance. Since *F/D* of the actual offset reflector is greater in F/D_p, for the parent reflector, cross-polarization will be worse in an offset reflector. This assumes that the feed is directed toward the apex; that is, $\psi_f = 0°$. This would lead to considerable spillover. Instead, the feed is usually pointed so that its axis bisects the angle subtended by the reflector ($\psi_f = \psi_B$) or the ray along the feed axis arrives in the center of the projected aperture ($\psi_f = \psi_C$). The influence of feed pointing angle is illustrated in Fig. 9-35 for an offset reflector with a diameter of 85.5λ, offset height $h = D/8$, and $F/D = 0.3$. The feed is linearly polarized with a circularly symmetric pattern (e.g., balanced feed) and a 10-dB beamwidth of 70°. This geometry yields results that are representative of offset reflectors used in practice. In this example, the bisector angle is $\psi_B = 45.1°$ and the angle of the central ray is $\psi_C = 49.7°$. Note in Fig. 9-35 that cross-polarization performance degrades as the feed pointing angle increases. However, gain shows a broad peak centered on $\psi_f = 47°$, which is between ψ_B and ψ_C. Side-lobe level is fairly constant with pointing angle as long as $\psi_f > 30°$. It turns out that the feed pointing angle $\psi_f = \psi_E$ that leads to equal edge illuminations ($\text{EI}_L = \text{EI}_U$) produces very low side lobes with only small penalties in gain and cross-polarization [25].
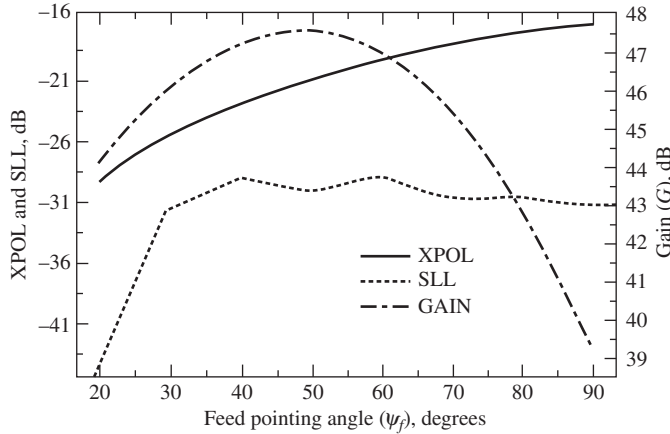
Figure 9-35 Peak cross-polarization level (XPOL), side-lobe level (SLL), and gain (G) as a function of feed pointing angle (ψ_f) for an offset reflector (see Fig. 9-31) with $D = 85.5\lambda$, $F/D_p = 0.3$, and $h = D/8$. The bisector angle is $\psi_B = 45.1°$. The central ray angle is $\psi_C = 49.7°$. The feed is balanced and linearly polarized. (Used with permission from Terada and Stutzman [25] © 1993 IEEE.)

In this example, $\psi_E = 49.6°$, where SLL = −30.1 dB. An important fundamental limitation on offset reflectors is apparent from Fig. 9-35. For maximum gain (i.e., aperture efficiency), cross-polarization is on the order of −23 dB, which is unacceptably high in many applications.

Cross-polarization performance of single offset reflectors is summarized in Fig. 9-36. The cross-polarization for a balanced feed linearly polarized will be maximum in the plane of asymmetry (*yz*-plane) and zero in the plane of symmetry (*xz*-plane). [H.11.1:Stutzman; p.102; 26; 27] The polarization in the (secondary) pattern of a reflector antenna is influenced by both the cross-polarization of the (primary) feed pattern, XPOL$_F$, as well as that introduced by the reflector, XPOL$_R$. So far, we have neglected any feed cross-polarization, that is,
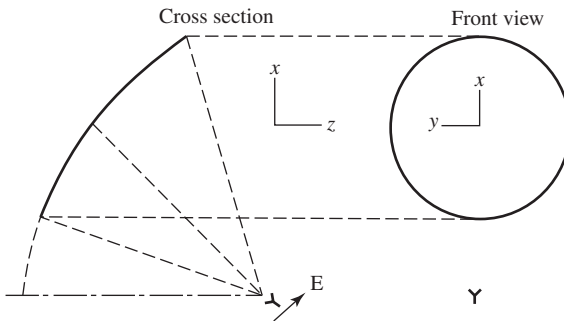
	Cross-polarization properties	
	xz-plane (plane of symmetry, plane of offset)	*yz*-plane (plane of asymmetry)
Feed		
Ideal *LP*	No cross-polarization	Cross-polarization
Ideal *CP*	No cross-polarization	No cross-polarization Beam squint

Figure 9-36 Summary of cross-polarization level properties of an offset parabolic reflector with a balanced feed. (Reproduced by permission from [H.11.1: Stutzman]. © 1993 by Artech House, Inc., Norwood, MA.)

$XPOL_F = 0$. Since cross-polarization requires a code such as a physical optics code for exact evaluation, an approximate technique is very useful. A worst-case estimate of cross-polarization of the reflector system is found by adding the contributions [28]:

$$XPOL_S = XPOL_F + XPOL_R \tag{9-217}$$

For example, a feed with -30-dB cross-polarization ($XPOL_F = 0.0316$) used in a reflector with -23-dB cross-polarization ($XPOL_R = 0.0708$) gives $XPOL_S = 0.1024 = -19.8$ dB.

The situation is much different for a circularly polarized (CP), balanced feed as indicated in Fig. 9-36. There is no cross-polarization; however, the main beam "squints" off axis in the yz-plane [26, 27, 29]. The main beam rotates (or squints) to opposite sides of the reflector axis for opposite senses of CP. Note that with CP feeds, the sense of CP radiated from the reflector will be opposite that of the feed due to the sense reversal encountered during reflection.

The advantages of the offset reflector of Sec. 9.6.3 and the dual reflector of Sec. 9.6.4 can be combined in the form of a *dual offset reflector*. Aperture blockage is eliminated and the subreflector introduces a second design variable for reducing cross-polarization far below that of a single offset reflector. The performance of a dual offset reflector can be evaluated in a manner similar to that for the axisymmetric dual reflector by using an equivalent single paraboloid with the same diameter as the main reflector and a focal length given by $F_e = MF$ from (9-216). If the dual reflector geometry is chosen such that the feed axis of the original system is coincident with the equivalent paraboloid axis, cross-polarization will be minimum.

Reflector antennas designed for high gain are focused systems. That is, an incoming wave parallel to the axis of the main reflector will be focused to a small region near the focal point. This leads to a simple antenna system, but limits beam scanning possibilities if rotating the entire reflector system is to be avoided. Some beam scanning is possible by displacing the feed off the focal point. This can be understood by considering a planar reflector with a small feed antenna that is displaced (and tilted back to aim at the reflector) from the axis perpendicular to the reflector. The reflected wave, or main beam from the reflector, will exit at an angle equal to the displacement angle. A similar effect applies to parabolic reflectors as shown in Fig. 9-37, where the feed antenna is laterally displaced a distance δ in the focal plane. If the reflector is flat ($F/D = \infty$), the angle of the beam scan angle θ_B equals the feed tilt angle θ_F. For curved reflectors ($F < \infty$), the beam scan angle will be less than the feed tilt angle. Scanning is quantified with *beam deviation factor* (BDF):

$$BDF = \frac{\theta_B}{\theta_F} \tag{9-218}$$

BDF is maximum at unity for a flat reflector and decreases with decreasing F/D or F/D_p for axisymmetric and offset reflectors, respectively. The following approximate expression can be used for small displacements δ [30]:

$$BDF = \frac{1 + 0.36\left[4\dfrac{F}{D}\right]^{-2}}{1 + \left[4\dfrac{F}{D}\right]^{-2}} \tag{9-219}$$

Lateral feed displacement introduces a planar phase front tilted with respect to the aperture plane that is responsible for scanning the beam in a direction opposite to the displacement, as indicated in Fig. 9-37. However, nonlinear phase as a function of position in the aperture plane is also introduced, leading to pattern distortion, including beam broadening, and gain loss [31]. These effects worsen with increasing displacement
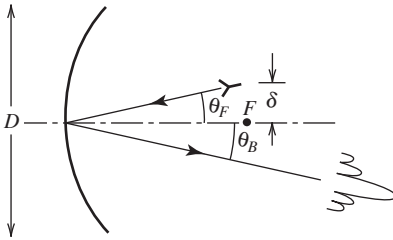
Figure 9-37 Beam scanning of a reflector antenna by feed displacement.

and lower *F/D* values. One characteristic pattern distortion is a high first side lobe called the *coma lobe* on the reflector axis side of the main beam.

Multiple reflectors can also be scanned by displacement of the feed off the focal point. Dual reflectors offer the advantage of a longer effective focal length through the equivalent parabolic reflector; see (9-216). This leads to better scan performance. Application of an equivalent single paraboloid provides approximate results for small scan angles with dual offset reflectors [32]. Advanced techniques can be used with tri-reflectors to minimize motion during scan [33, 34].

9.6.6 Gain Calculations for Reflector Antennas

Aperture antennas usually have an obvious physical aperture of area A_p through which energy passes on its way to the far field. The maximum achievable gain for an aperture antenna from (9-67)

$$G_{\max} = D_u = \frac{4\pi}{\lambda^2} A_p \tag{9-220}$$

This gain is possible only under the ideal circumstances of a uniform amplitude, uniform phase antenna with no spillover or ohmic losses present. In practice, these conditions are not satisfied and gain is decreased from ideal, as represented through the following:

$$G = \varepsilon_{\mathrm{ap}} D_u = \varepsilon_{\mathrm{ap}} \frac{4\pi}{\lambda^2} A_p \tag{9-221}$$

where $\varepsilon_{\mathrm{ap}}$ is aperture efficiency and $0 \le \varepsilon_{\mathrm{ap}} \le 1$; see (9-78) and (9-79). Since wavelength and physical aperture area are easily found, the study of gain reduces to one of aperture efficiency, which can be expressed as a product of subefficiencies:

$$\varepsilon_{\mathrm{ap}} = e_r \varepsilon_t \varepsilon_s \varepsilon_a \tag{9-222}$$

where

e_r = radiation efficiency
ε_t = aperture taper efficiency
ε_s = spillover efficiency
ε_a = achievement efficiency

We now explain each of these efficiencies.

Aperture taper efficiency is obtained by working with that portion of the power that reaches the aperture. That is, if we ignore achievement and spillover losses, (9-66) for a circular reflector aperture of radius a leads to

$$\varepsilon_t = \frac{1}{\pi a^2} \frac{\left| \int_0^{2\pi} \int_0^a E_a(\rho', \phi') \rho' \, d\rho' \, d\phi' \right|^2}{\int_0^{2\pi} \int_0^a |E_a(\rho', \phi')|^2 \rho' \, d\rho' \, d\phi'} \tag{9-223}$$

This expression can be written directly in terms of the known feed antenna pattern by transforming to feed angles:

$$\varepsilon_t = \frac{4F^2}{\pi a^2} \frac{\left| \int_0^{2\pi} \int_0^{\theta_o} F_f(\theta_f, \phi') \tan \frac{\theta_f}{2} \, d\theta_f \, d\phi' \right|^2}{\int_0^{2\pi} \int_0^{\theta_o} |F_f(\theta_f, \phi')|^2 \sin \theta_f \, d\theta_f \, d\phi'} \qquad (9\text{-}224)$$

The aperture taper efficiency can be evaluated from the feed pattern using this expression.

The feed antenna radiation pattern has the greatest influence on reflector antenna gain by its control over the aperture distribution and aperture taper efficiency, as discussed in Sec. 9.6.2. Since the feed pattern will extend beyond the rim of the reflector, the associated power will not be redirected by the reflector into the main beam and consequently gain is reduced. This is referred to as *spillover* and the associated efficiency factor is called *spillover efficiency* ε_s, which is defined as the fraction of power radiated by the feed that is intercepted by the main reflector of a single reflector or the subreflector of a dual reflector.

Spillover efficiency measures that portion of the feed pattern that is intercepted by the main reflector (and redirected through the aperture into the main beam) relative to the total feed power:

$$\varepsilon_s = \frac{\int_0^{2\pi} \int_0^{\theta_o} |F_f(\theta_f, \phi')|^2 \sin \theta_f \, d\theta_f \, d\phi'}{\int_0^{2\pi} \int_0^{\pi} |F_f(\theta_f, \phi')|^2 \sin \theta_f \, d\theta_f \, d\phi'} \qquad (9\text{-}225)$$

Notice that the numerator involves an integral over the feed pattern only out to the angular extent of the reflector, whereas the denominator integral extends over the entire feed pattern.

The reflector design problem reduces to a tradeoff between aperture taper and spillover through feed antenna choice. A broad feed pattern introduces little amplitude taper across the aperture, but there will be a significant spillover as illustrated in Fig. 9-38a. The spillover problem is solved by using a feed with a narrow pattern as illustrated in Fig. 9-38b. However, now the feed pattern taper is large, leading to low aperture taper efficiency.

Taper and spillover efficiencies can be combined to form *illumination efficiency* ε_i to completely account for feed pattern and main reflector effects. That is, ε_i yields total aperture efficiency under ideal circumstances of no ohmic losses ($e_r = 1$) and no achievement losses ($\varepsilon_a = 1$). Multiplying (9-224) and (9-225) and using $a = 2F \tan(\theta_o/2)$ from (9-185a) lead to
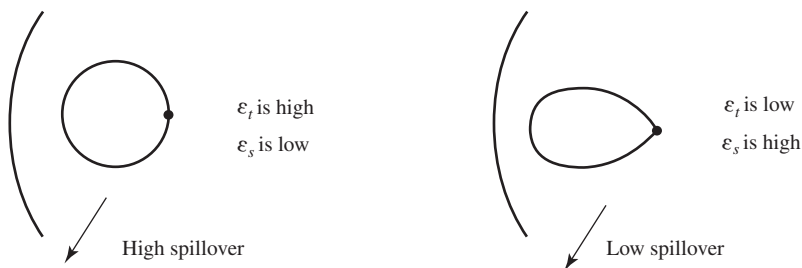
(*a*) Broad feed pattern giving high aperture taper efficiency but low spillover efficiency.

(*b*) Narrow feed pattern giving high spillover efficiency but low aperture taper efficiency.

Figure 9-38 Illustration of the influence of the feed antenna pattern on reflector aperture taper and spillover.

$$\varepsilon_i = \varepsilon_t \varepsilon_s = \frac{G_f}{4\pi^2} \cot^2 \frac{\theta_o}{2} \left| \int_0^{2\pi} \int_0^{\theta_o} F_f(\theta_f, \phi') \tan \frac{\theta_f}{2} \, d\theta_f \, d\phi' \right|^2 \tag{9-226}$$

Here, we have made use of the following expression for the gain of the feed:

$$G_f = \frac{4\pi}{\int_0^{2\pi} \int_0^{\pi} |F_f(\theta_f, \phi')|^2 \sin \theta_f \, d\theta_f \, d\phi'} \tag{9-227}$$

This is actually feed directivity, but feed ohmic losses are included in e_r for the entire reflector system.

An ideal, and unrealizable, feed antenna pattern would compensate for spherical spreading loss by increasing with angle off axis and then abruptly falling to zero in the direction of the rim to avoid spillover. This pattern follows from (9-196) and (9-194) as

$$F_f(\theta_f, \phi_f) = \begin{cases} \cos^2 \frac{\theta_o}{2} \sec^2 \frac{\theta_f}{2} & \theta_f \leq \theta_\sigma \\ 0 & \theta_f > \theta_\sigma \end{cases} \quad \textit{ideal feed} \tag{9-228}$$

which is normalized to a peak of unity at $\theta_f = \theta_o$. Using this in (9-226) yields an efficiency of $\varepsilon_i = 1$; see Prob. 9.6-12. Thus, the ideal feed pattern of (9-228) will lead to 100% aperture efficiency if no ohmic or achievement losses are present. The ideal feed must, however, be infinitely large in order to produce the discontinuous pattern.

Usually, (9-226) cannot be evaluated analytically and must either be estimated based on canonical distributions or evaluated numerically. However, there is one feed pattern function that is used to model the patterns of real feeds such as conical corrugated horns and that can be handled analytically. This pattern, which is discussed in detail in Sec. 9.7.4, is the rotationally symmetric pattern of

$$F_f(\theta_f) = \begin{cases} \cos^2 \theta_f & \theta_f \leq \frac{\pi}{2} \\ 0 & \theta_f > \frac{\pi}{2} \end{cases} \tag{9-229}$$

The evaluation of (9-225) and (9-227) using this feed model is straightforward and yields

$$\varepsilon_s = 1 - \cos^{2q+1} \theta_o \tag{9-230}$$

$$G_f = 2(2q + 1) \tag{9-231}$$

The evaluation of ε_t in (7-226) is more difficult. Expressions for ε_i follow for a few q values [H.3: Silver, p. 425].

$$\varepsilon_i = \cot^2 \frac{\theta_o}{2} \begin{cases} 24 \left[\sin^2 \frac{\theta_o}{2} + \ln \left(\cos \frac{\theta_o}{2} \right) \right]^2 & q = 1 \\[3mm] 40 \left[\sin^4 \frac{\theta_o}{2} + \ln \left(\cos \frac{\theta_o}{2} \right) \right]^2 & q = 2 \quad (9\text{-}232) \\[3mm] 14 \left[\frac{1}{2} \sin^2 \theta_o + \frac{1}{3} (1 - \cos \theta_o)^3 + 2 \ln \left(\cos \frac{\theta_o}{2} \right) \right]^2 & q = 3 \end{cases}$$

The edge illumination EI for this distribution from (9-208a) is

$$\text{EI} = \frac{1 + \cos\theta_o}{2}\cos^q\theta_o \tag{9-233}$$

The q-value of 2 is representative of situations encountered in practice. The taper and spillover efficiencies and their product found from (9-230) and (9-232) are plotted in Fig. 9-39. The tradeoff between taper and spillover is evident. The peak value is about $\varepsilon_i = 82\%$ and occurs for an edge illumination of about $\text{EI}_{\text{dB}} = -11\text{dB}$. Thus, we arrive at a general rule: *Peak aperture efficiency of a parabolic reflector occurs for an edge illumination of about* -11 dB, *or* $E_{\text{an}}(\rho' = a) = 0.28$. It turns out that the peak illumination efficiency for q-values of 1 to 4 is near 82%; see Prob. 9.6-16. In practice, the highest achievable aperture efficiency for a single reflector using a nearly rotationally symmetric feed pattern is about 75%. If simple feeds such as an open-ended waveguide are used, the aperture efficiency is about 60%. We now examine the remaining efficiencies responsible for gain reduction.

The several factors that reduce gain for practical implementation reasons are lumped together into achievement efficiency, which is expressed using subefficiencies as

$$\varepsilon_a = \varepsilon_{\text{rs}}\varepsilon_{\text{cr}}\varepsilon_{\text{blk}}\varepsilon_{\phi\text{r}}\varepsilon_{\phi\text{f}} \tag{9-234}$$

where

ε_{rs} = random surface error efficiency
ε_{cr} = cross-polarization efficiency
ε_{blk} = aperture blockage efficiency
$\varepsilon_{\phi\text{r}}$ = reflector phase error efficiency
$\varepsilon_{\phi\text{f}}$ = feed phase error efficiency

All these efficiencies can range from 0 to 1, but for properly designed systems they are just slightly less than unity. We now discuss them.

Random surface deviations from the ideal shape of a reflector cause gain reduction and side-lobe increase. This is due to the distortions in the aperture phase because of the consequent departure from equal ray path lengths of a focused reflector system. *Random surface error efficiency* ε_{rs} is the efficiency factor associated with gain loss from random reflector surface errors. This efficiency can be expressed in terms of the rms surface deviation δ, which is approximately one-third of the peak-to-peak error. For
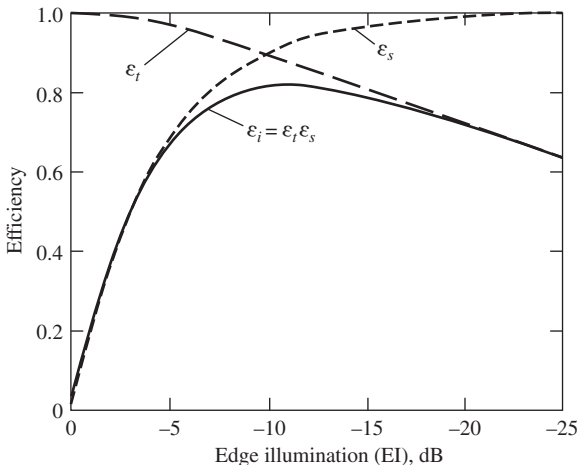
Figure 9-39 Aperture taper ε_t, spillover ε_s and illumination ε_i efficiencies for a $\cos^2\theta_f$ feed pattern ($q = 2$) as a function of edge illumination EI.

surface errors that are not large and that have a correlation length that is small compared to the reflector size,

$$\varepsilon_{rs} = e^{-(4\pi\delta/\lambda)^2} = 685.8(\delta/\lambda)^2 \text{ dB} \tag{9-235}$$

$2\pi/\lambda$ converts the surface errors to phase and the additional factor of 2 accounts for the two-way path of the reflected rays. This result was derived for a flat reflector with Gaussian distributed errors, but it works well in practice [35]. It can be seen from (9-235) that for $\delta \ll \lambda$, the efficiency is nearly 100%. For a fixed random error δ, as frequency increases such that δ varies from 0.01λ and 0.1λ, the efficiency decreases from 0.98 to 0.21. The corresponding gain loss from (9-235) is 0.07 to 6.9 dB. We conclude that random error loss is in transition for δ near 0.01λ. Smaller errors are negligible, whereas larger errors can be a significant problem. The manufacturing techniques for a reflector determine its surface accuracy. Machined metal reflectors are the most accurate with δ near 0.04 mm (0.001 in.). Mass production of reflectors that are a few meters in diameter or less using presses and molds yields slightly larger errors and this accuracy degrades for larger reflectors.

Cross-polarization efficiency ε_{cr} has contributions due to the reflector(s) and the feed antenna. The former is usually small (except for offset reflectors) and is neglected. Feed antennas have a component that is orthogonal to the desired polarization. The associated power ends up in the far field and is wasted—hence, a gain loss. Typical feeds yield ε_{cr} values from 96 to 99% [36], corresponding to gain losses of from 0.2 to 0.04 dB.

Structures placed in front of a reflector such as the feed, subreflector, and support hardware will block rays exiting the aperture and scatter power into the side-lobe region. A simple approximate formula is available for *aperture blockage efficiency* [36]:

$$\varepsilon_{blk} = \left[1 - \frac{1}{\varepsilon_t} \frac{A_b}{A_p} \right]^2 \tag{9-236}$$

where A_b is the blockage area projected onto the physical aperture of area A_p. The square is present because of gain loss due to a decrease of on-axis power by blockage and due to the increase in off-axis power by redirection of the same power into the side lobes. The aperture taper efficiency ε_t is included to weight the central area more heavily where blockage is usually present. For optimum operation, ε_t is about 0.89; see Prob. 9.6-16. Then for blockages of $A_b/A_p = 1$, 5, and 10%, $\varepsilon_{blk} = 0.98$, 0.89, and 0.79, respectively.

Under ideal circumstances, reflector antennas have uniform aperture phase. As with horn antennas, phase errors in the aperture plane lead to gain loss and pattern deterioration [36]. Phase errors arise for the following reasons:

a. **Displacement of the phase center of the feed antenna off the focal point**. The reflector is said to be defocused. Lateral displacement causes beam scanning as discussed in Sec. 9.6.5. Often, these errors can be corrected by repositioning the phase center of the feed antenna to the focal point.

b. **Deterministic deviations of the reflector(s) from design shapes**. For example, a single reflector that deviates from a paraboloidal shape with a "potato chip" distortion over the entire reflector will produce a smooth phase error over the aperture. Forces such as wind, temperature gradients, and differential gravity effects in addition to manufacturing defects cause deterministic errors whose efficiency is represented with $\varepsilon_{\phi r}$. Only sophisticated techniques such as array feeds are capable of correcting for deterministic errors [37].

c. **An imperfect feed antenna phase center**. The loss is represented by $\varepsilon_{\phi f}$ and can be partially compensated by feed repositioning [36].

d. **Random surface error effects**. These effects cannot be corrected. The associated efficiency factor ε_{rs} is given in (9-235).

The first three listed effects can be combined into phase-error efficiency ε_{ph}. Since random phase errors are usually the dominant effect, ε_{rs} is shown explicitly in (9-234).

The diffraction effects mentioned in Sec. 9.6.1 also cause gain loss, but are usually small compared to spillover loss [36]. This and other sources of gain loss not specifically mentioned are included in ε_a.

It is important to remember that a reflector antenna usually includes some processing components such as an orthomode transducer (OMT) to separate orthogonal polarizations at the feed. These components are lossy and reduce the gain. Their losses, along with other losses such as radome loss, are all included in e_r. Systems using offset reflectors usually place the upconverter/downconverter hardware immediately behind the feed horn because aperture blockage is not a problem. This greatly reduces the RF transmission line loss compared to an axisyminetric reflector with a transmission line running from the feed to the rear of the reflector.

Although highly approximate, it is helpful to have a "typical" aperture efficiency value for a reflector antenna. Usually most efficiency loss comes from using an unbalanced feed. For example, a dipole feed is often used at UHF frequencies, leading to an aperture efficiency of about 50%. As mentioned earlier, a well-designed, prime-focus reflector antenna with a balanced feed (to be discussed in detail in Sec. 9.7.1) has up to 75% efficiency. The range of typical values for many applications is

$$\varepsilon_{\text{ap}} \approx 0.55 \text{ to } 0.65 \tag{9-237}$$

9.6.7   Other Reflector Antennas

Single and dual parabolic reflectors, as described in this chapter, were developed roughly from World War II through 1960. Since that period, modifications to the basic reflector types have been introduced for the purpose of increasing aperture efficiency or for special antenna pattern-shaping applications to produce a pencil beam, a fan beam, a shaped main beam, low side lobes, or multiple main beams. In this section, we introduce a few of the many types of reflector antennas that are in common use.

A parabolic reflector with a circular perimeter and a simple feed at the focal point as in Fig. 9-27 is used to produce a pencil-beam pattern that is rotationally symmetric. As we have seen, the configuration can be axisymmetric or offset, and subreflectors can be used to form a multiple reflector. There are many applications for a high-gain reflector antenna with different beamwidths in the principal planes. An example is shown in Fig. 9-40a, which is a single parabolic reflector with wider horizontal than vertical aperture extent. This produces a narrower main beam in the horizontal plane as needed for VSAT (Very Small Aperture Terminal) satellite communications. The narrow beam is in the geostationary satellite arc to avoid interference between adjacent satellites. The feed antenna must have a broader pattern in the horizontal plane for proper dish illumination. A pattern with different principal plane beamwidths can also be produced with a *parabolic cylinder* as shown in Fig.9-40b, which has a parabolic cross section in one plane and a line cross section parallel to the reflector axis. The narrow beamwidth is in the plane containing the reflector axis and requires a feed that extends along the focal lime. The corner reflector antenna discussed in Sec. 6.6.2 is a simplified version of the parabolic cylinder that uses flat metallic sides. The *parabolic torus* of Fig. 9-40c is, in a sense, a curved version of the parabolic cylinder, having a parabolic and circular cross sections in the principal planes. A popular application for the parabolic torus employs multiple feeds located along the focal arc to produce separate beams for receiving different satellites with a single Earth terminal antenna. Aperture efficiency is sacrificed, but there is a cost savings over using several antennas. The *spherical reflector* of Fig. 9-40d, with a circular cross section in all planes containing the reflector axis, produces a pencil beam but with lower aperture efficiency than a parabolic version due to nonuniform aperture phase; equivalently, there is a focal region rather than a focal point. However, the feed can be moved over the focal
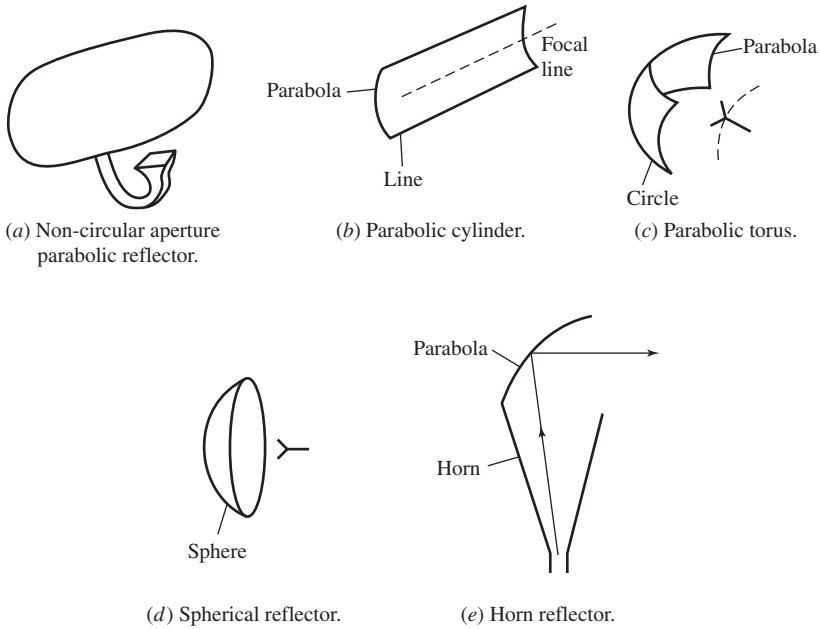
(*a*) Non-circular aperture   (*b*) Parabolic cylinder.   (*c*) Parabolic torus.
     parabolic reflector.

(*d*) Spherical reflector.   (*e*) Horn reflector.

Figure 9-40 Other reflector antenna types.

region to scan the beam with lower gain loss than experienced when displacing a feed from the focal point of a parabolic reflector. The *horn reflector* antenna of Fig.9-40*e* is formed by joining a horn to an offset parabolic reflector. It is very popular for terrestrial microwave communication links because of its low side and back lobes.

   Finally, we mention that *shaped reflectors* are used to produce *shaped beams* for either optimum power distribution in desired directions or to reduce power in directions of interference. Geometrical optics-based techniques are usually used for synthesizing shaped beams [H.8.5: Wescott].

9.7   FEED ANTENNAS FOR REFLECTORS

A reflector antenna must be fed properly in order to realize maximum performance, such as high aperture efficiency. This section presents design principles and the types of commonly used feed antennas.

9.7.1   Field Representations

The electric field from a feed antenna can be expressed in general, following the geometry of Fig. 9-25, as

$$\mathbf{E}_f = V_o \frac{e^{-j\beta r_f}}{r_f} \left[U_f(\theta_f, \phi_f)\hat{\boldsymbol{\theta}}_f + V_f(\theta_f, \phi_f)\hat{\boldsymbol{\phi}}_f \right] \tag{9-238}$$

Rarely are U_f and V_f known for all angles. Instead, usually only the principal plane patterns are available: $U_f(\theta_f, \phi_f = 0) = C_E(\theta_f)$ in the *E*-plane and $V_f(\theta_f, \phi_f = 90°) = C_H(\theta_f)$ in the *H*-plane. Then the field from the feed is found approximately for any angle ϕ_f by interpolation. If the feed is purely linearly polarized, it can be modeled in terms of its principal plane patterns as

$$\mathbf{E}_f = V_o \frac{e^{-j\beta r_f}}{r_f} \left[C_E(\theta_f) \cos \phi_f \, \hat{\boldsymbol{\theta}}_{\mathbf{f}} - C_H(\theta_f) \sin \phi_f \, \hat{\boldsymbol{\phi}}_{\mathbf{f}} \right] \quad x_f\text{-polarized} \qquad (9\text{-}239a)$$

or

$$\mathbf{E}_f = V_o \frac{e^{-j\beta r_f}}{r_f} \left[C_E(\theta_f) \sin \phi_f \, \hat{\boldsymbol{\theta}}_{\mathbf{f}} - C_H(\theta_f) \cos \phi_f \, \hat{\boldsymbol{\phi}}_{\mathbf{f}} \right] \quad y_f\text{-polarized} \qquad (9\text{-}239b)$$

As an example, the E-plane and H-plane of an x_f-polarized short dipole are

$$C_E(\theta_f) = \cos \theta_f, \qquad C_H(\theta_f) = 1 \qquad x_f\text{-polarized short dipole} \qquad (9\text{-}240)$$

As illustrated in Fig. 9-28 and with (9-206), cross-polarization is present in the aperture of an axisymmetric reflector fed with a dipole antenna and the resulting far-field patterns contain cross-polarization except in the principal planes.

An axisymmetric reflector will have a rotationally symmetric secondary pattern and very low cross-polarization if it is fed with a rotationally symmetric feed pattern:

$$C_E(\theta_f) = C_H(\theta_f) \qquad balanced\ feed \qquad (9\text{-}241)$$

A feed that creates such a pattern is referred to as a *balanced feed*. The field representations of (9-239) for a balanced feed reduce to

$$\mathbf{E}_{\mathbf{f}}^{\mathbf{v}} = V_o \frac{e^{-j\beta r_f}}{r_f} F_f(\theta_f) \left[\cos \phi_f \, \hat{\boldsymbol{\theta}}_f - \sin \phi_f \, \hat{\boldsymbol{\phi}}_f \right] = V_o \frac{e^{-j\beta r_f}}{r_f} F_f(\theta_f) \hat{\mathbf{v}}$$

$$= E_v \hat{\mathbf{v}} \qquad x_f\text{-polarized} \qquad (9\text{-}242a)$$

$$\mathbf{E}_{\mathbf{f}}^{\mathbf{h}} = V_o \frac{e^{-j\beta r_f}}{r_f} F_f(\theta_f) \left[\sin \phi_f \, \hat{\boldsymbol{\theta}}_f + \cos \phi_f \, \hat{\boldsymbol{\phi}}_f \right] = V_o \frac{e^{-j\beta r_f}}{r_f} F_f(\theta_f) \hat{\mathbf{h}}$$

$$= E_h \hat{\mathbf{h}} \qquad y_f\text{-polarized} \qquad (9\text{-}242b)$$

These correspond to vertical (v) and horizontal (h) feed polarizations with pure linear polarizations in the $x_f z_f$- and $y_f z_f$-planes, respectively. Note that they have a rotationally symmetric pattern $F_f(\theta_f)$. Also, there is no cross-polarization since, for example, with the vertically polarized feed $\mathbf{E}_f^v \cdot \hat{\mathbf{h}} = E_v \hat{\mathbf{v}} \cdot \hat{\mathbf{h}} = 0$.

The aperture electric field for a balanced x_f-polarized feed from (9-205) is

$$\mathbf{E}_a = -\hat{\mathbf{x}} V_o \frac{e^{-j\beta 2F}}{r_f} F_f(\theta_f) \qquad (9\text{-}243)$$

This corresponds to the GO model of (9-198) and has no cross-polarization. However, there will be a small amount of off-axis cross-polarization in the secondary pattern that is not accounted for here and arising from the axial currents (z-directed) on the surface of the reflector.

9.7.2   Matching the Feed to the Reflector

There are two equivalent viewpoints that can be used to select a feed for proper illumination of a reflector for high aperture efficiency: Matching the feed pattern to the reflector or matching of the feed antenna aperture distribution to the focal field distribution. We discuss these approaches in this section.

As noted previously, the feed pattern is matched to the reflector when its pattern gives about a –11-dB edge illumination. The governing equation for axisymmetric reflectors,

(9-208), can be solved using the popular $\cos^q \theta_f$ feed pattern model to determine the required half-power and –10-dB beamwidths. The result is plotted in Fig. 9-41. These curves are very useful in reflector design. Also shown in Fig. 9-41 is the reflector edge angle θ_o from Fig. 9-26*b*.

The focal field matching approach involves a plane wave incident on the reflector. In the limit of infinite frequency, the rays converge to the focal point. In practice, the received fields extend over a finite region near the focal point, resulting in a *focal plane distribution* (FPD). It turns out that the FPD is approximately the Fourier transform of the aperture plane distribution (APD), with increasing accuracy with larger F/D [38]. So, the uniform APD created by the incident plane wave leads to a $\sin(u)/u$ form FPD and 100% aperture efficiency. The purpose of the feed is to capture the FPD. In fact, if the aperture distribution of the feed antenna placed in the focal plane matches the FPD, the aperture efficiency will be 100%. However, a feed of infinite extent would be required to collect all the fields. This "ideal feed" for an axisymmetric parabolic reflector has a point phase center and pattern that is rotationally symmetric, extending over a cone only out to the reflector rim and compensating for spherical spreading loss. This pattern is expressed functionally in normalized form in (9-228). The pattern discontinuity at the reflector rim ($\theta_f = \theta_o$) gives the required uniform APD and zero outside. This pattern is impossible to achieve. The Fourier transform gives a FPD with infinite extent, requiring an infinite-sized feed to realize.

A classical feed for producing the purely linearly polarized aperture distribution as in (9-243), yielding low cross-polarization in the secondary pattern, is the *Huygens' source*. Its rotationally symmetric pattern leads to high efficiency when feeding an axisymmetric parabolic reflector. The development of the Huygens' source begins by reexamining the aperture fields created by a short dipole feeding a parabolic reflector. The electric fields of an x_f-polarized short dipole in (9-206) have cross-polarized components as indicated in Fig. 9-28. This means that the total aperture electric field has outward curvature as shown in Fig. 9-42*a*. Opposite curvature fields as in Fig. 9-42*b* are created by a y_f-directed magnetic dipole (see Sec. 3.4.2) at the focal point. The combination of crossed electrically small electric and magnetic dipoles produces the purely linearly polarized field of Fig. 9-42*c*. The Huygen's source aperture fields can be derived using (9-243) for a short dipole and their dual form for a magnetic dipole feed; see Prob. 9.7-4. The magnetic current required to produce an electric field from the magnetic dipole equaling that for the electric dipole is found by equating the components of (2-74*a*) and (3-40), yielding $I^m = \eta I$. Practical Huygens' sources are discussed in Sec. 9.7.4.
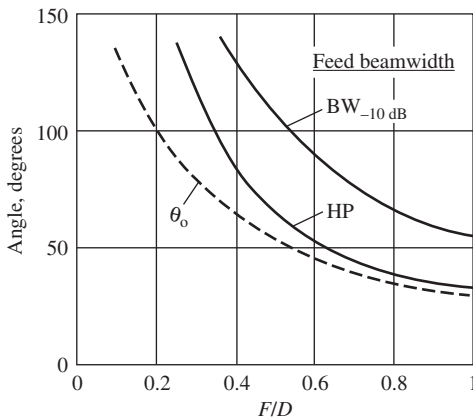
Figure 9-41 Angle from the axis of an axisymmetric parabolic reflector to the rim θ_o (dashed curve) and the required feed pattern beamwidths (solid curves) to produce an –11-dB edge illumination as a function of F/D.

(*a*) Aperture electric field when fed
with a short dipole.

(*b*) Aperture electric field when fed with an
electrically small magnetic dipole.

(*c*) Aperture electric field when fed with
crossed electric and magnetic dipoles.
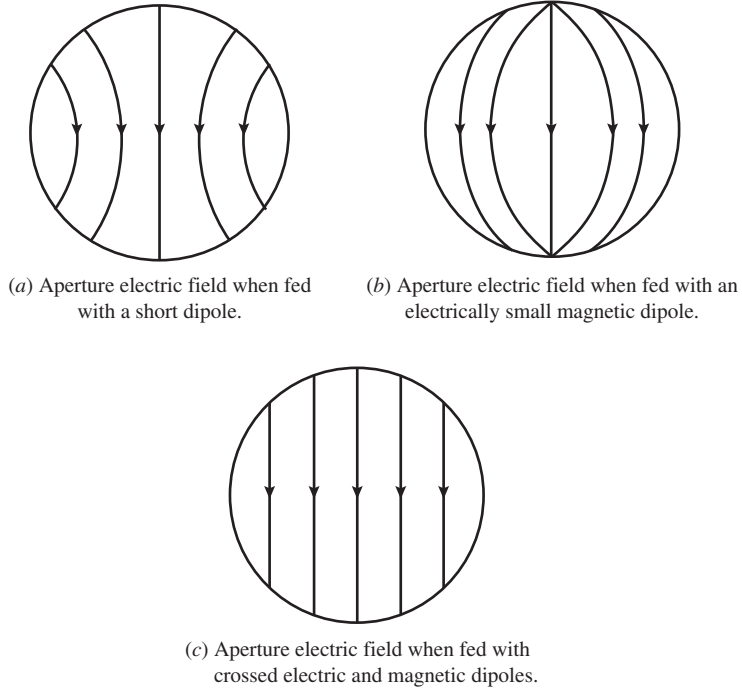
Figure 9-42 Aperture electric fields of an axisymmetric parabolic antenna with various feeds.

9.7.3 A General Feed Model

A popular representation for feed patterns is the $\cos^q \theta_f$ pattern given by

$$C_E(\theta_f) = \cos^{q_E} \theta_f, \qquad C_H(\theta_f) = \cos^{q_H} \theta_f, \qquad \theta_f < \pi/2 \tag{9-244}$$

which for a balanced feed reduces to

$$C_E(\theta_f) = C_H(\theta_f) = F_f(\theta_f) = \cos^q \theta_f, \qquad \theta_f < \pi/2 \tag{9-245}$$

The value of q (or q_E and q_H) is chosen to match the pattern of a real feed antenna at one point in addition to the unity beam peak:

$$q = \frac{\log[F_f(\theta'_f)]}{\log(\cos \theta'_f)} \tag{9-246}$$

where θ'_f is the match point, such as the -3- or -10-dB pattern point or θ_o. An advantage of using the simple pattern form of (9-245) is that it can be used to evaluate important parameters such as the feed antenna directivity:

$$G_f = \frac{2(2q_E + 1)(2q_H + 1)}{q_E + q_H + 1} \tag{9-247}$$

which reduces to $2(2q + 1)$ as in (9-231) for a balanced feed.

We now have all the tools to formulate a simple procedure for designing an axisymmetric reflector using the following steps:

1. **Determine the reflector diameter.** The diameter to achieve a required gain is found using (9-78) if an aperture efficiency value can be assumed. If the beamwidth is specified, the diameter is found by solving the following for D:

$$\text{HP} = 1.18\frac{\lambda}{D} \quad rad \tag{9-248}$$

which is a good approximation for reflectors with a -11-dB edge illumination.

2. **Choose F/D.** The normal range of F/D values is 0.3 to 1.0. Higher values lead to better cross-polarization performance, but require a narrower feed pattern and, hence, physically larger feed antenna.

3. **Determine the required feed pattern.** The edge illumination is specified for a desired performance and the q-value for a $\cos^q \theta_f$ feed model is found by solving (9-233) for q:

$$q = \frac{\log\left[\text{EI}\left(1 + \dfrac{1}{16(F/D)^2}\right)\right]}{\log\left[\cos\left(2 \tan^{-1}\dfrac{1}{4(F/D)}\right)\right]} \tag{9-249}$$

El $= 0.28(-11$ dB$)$ is used for optimum gain.

The final step in the complete design process is to select a feed antenna that approximates the $\cos^q \theta_f$ pattern with the q-value found from (9-249). The next two subsections address feed design. This subsection is closed with a comprehensive example.

EXAMPLE 9-9   *Design of an Axisymmetric Reflector Antenna*

Suppose that a high-gain, narrow beam antenna is required at 10 GHz. The axisymmetric parabolic reflector antenna is a good choice. To achieve a $1°$ half-power beamwidth, the required diameter follows from (9-248) as

$$D = \frac{1.18\lambda}{\text{HP}\dfrac{\pi}{180°}} = \frac{1.18(0.03 \text{ m})}{1 \cdot \dfrac{\pi}{180°}} = 2.0 \text{ m}$$

The *F/D* is chosen to be 0.5 for low cross-polarization. Solving (9-249) for the optimum case of EI $= 0.28$ gives a value of q near 2. The edge illumination value is verified using (9-208) with $\theta_o = 53.1°$.

$$\text{EI} = -\text{FT} - L_{\text{sph}} = 20 \log(\cos^q \theta_o) + 20 \log[(1 + \cos \theta_o)/2] = -8.86 - 1.93$$
$$= -10.79 \text{ dB} \approx -11 \text{ dB}$$

The aperture distribution based on (9-208) is plotted in Fig. 9-43 along with a parabolic-squared taper on a pedestal with $C = 0.28$. The agreement suggests that the parabolic-squared tapered circular aperture model works well for reflectors. The illumination efficiency follows from (9-232) for $q = 2$ as $\varepsilon_i = 0.82$. The spillover efficiency from (9-230) is

$$\varepsilon_s = 1 - \cos^{2q+1}\theta_o = 1 - \cos^5 53.1° = 0.92$$

So,

$$\varepsilon_t = \varepsilon_i/\varepsilon_s = 0.82/0.92 = 0.89$$
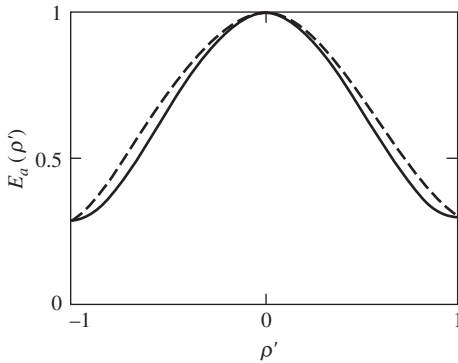
Figure 9-43 Aperture field distribution for axisymmetric parabolic reflector of Example 9-9 (dashed curve) along with the parabolic-squared-on-a-pedestal distribution with $C = 0.28$ (solid curve).
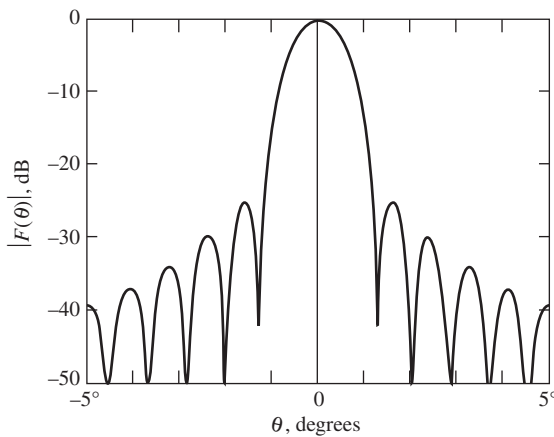
Figure 9-44 Pattern for the 2-m axisymmetric parabolic reflector antenna of Example 9-9 computed using the PRAC code.

This is consistent with Table 9-1*b* for the parabolic-squared on a -11-dB pedestal distribution. The radiation pattern in the 45° plane computed using PRAC (see Appendix G) is shown in Fig. 9-44. See Prob. 9.7-3 for gain evaluation.

9.7.4   Feed Antennas Used in Practice

The ideal feed of (9-228) produces a uniform amplitude and phase distribution which compensates for spherical spreading loss and does not have spillover. However, it cannot be realized in practice. A practical feed is smaller than a few wavelengths in diameter and, thus, has a broad pattern. The feed pattern is usually modeled with the $\cos^q \theta_f$ pattern of (9-245). In this section we present the guidelines for design of practical feed antennas along with examples of popular feeds. Further details are found in [H.6: Balanis, Ed., Chap. 18] and the books in [H.8.6].

For the usual case of high aperture efficiency designs, the feed should have the following characteristics:

1. The feed pattern should be rotationally symmetric, or balanced, as in (9-245).
2. The feed pattern should be such that the reflector edge illumination is about -11 dB, as discussed in Sec. 9.7.2.
3. The feed should have a point phase center and the phase center should be positioned at the focal point of the reflector.

4. The feed should be small in order to reduce blockage; it is usually on the order of a wavelength in diameter.
5. The feed should have low cross-polarization, usually below -30 dB.
6. The above characteristics should hold over the desired operational frequency band.

Usually, the feed is responsible for limiting the performance of a reflector antenna system.

The simplest feed antenna is a dipole, which is often combined with some type of metallic backing to reduce direct feed radiation in the direction of the reflector main beam. This not only reduces aperture efficiency but also leads to significant cross-polarization since it is the unbalance in the principal plane patterns that most strongly influences cross-polarization, as seen in (9-205). Dipoles are widely used as feeds for reflectors operating in the UHF range. Aperture efficiency is, however, low. For example, the illumination efficiency of the dipole-fed reflector in Fig. 9-29 is only 24%. At frequencies above a few GHz, waveguide and small horn antennas are used.

The open-ended rectangular waveguide and rectangular horn antennas operating in the dominant TE_{10} mode, discussed in Secs. 9.2 and 9.4, respectively, are used as feeds. Circular waveguides and conical horns operating in the dominant TE_{11} mode are also used as feeds and provide more symmetric principal plane patterns. Next, we discuss these feed antennas, followed by a discussion of multimode feeds.

The *open-ended circular waveguide* has cross-polarization below -30 dB. It is small in size, with diameters from 0.8 to 1.15λ and –10-dB beamwidths of about 140 to $104°$, respectively. The E- and H-plane beamwidths are not greatly different. Equal principal plane beamwidths occur for a diameter of 0.96λ where $BW_{-10\,dB} = 118°$. This provides a good match to a reflector with $\theta_o \approx 59°$, or $F/D \approx 0.44$. An axisymmetric parabolic reflector with this feed has $\varepsilon_i = 0.74$ [H.4: Collin, p. 231].

Conical horn antennas behave similarly to pyramidal horn antennas and display optimum gain with 48 to 56% aperture efficiency [H.6: *Ant. Eng. Hdbk.*, 4th ed., p. 14–32]. The half-power beamwidths under the condition of optimum gain are $HP_E = 1.05\lambda/d_f$ and $HP_H = 1.22\lambda/d_f$ [39], which can be used with the $\cos^q \theta_f$ feed model to perform reflector design.

The simple feeds discussed above operate in their dominant mode (TE_{10} for rectangular and TE_{11} for circular) and have unbalanced principal plane patterns. This is due to the markedly different amplitude distributions that are uniform in the E-plane and taper to zero in the H-plane. Since the aperture phase errors caused by the spherical phase fronts are strongly frequency-dependent, equal principal plane patterns can be obtained only over a narrow frequency range. Wider bandwidth balanced feeds with low cross-polarization can be achieved by introducing higher-order modes using a *multimode horn feed*.

There are several forms of multimode feed horns. Here, we consider the most popular form, the *dual mode* (*conical*) *horn* or *Potter horn* [40.]. The operating principle of the dual mode horn is similar to the Huygens' source of Fig. 9-42. In addition to the dominant TE_{11} mode of the conical horn, a TM_{11} mode is generated internal to the horn that has little effect on the H-plane pattern, but with proper amplitude and phase will alter the TE_{11} mode field distribution in the E-plane to be nearly like that in the H-plane. The electric fields of the separate modes as well as their combination are shown in Fig. 9-45a. Note that the modes reinforce in the central region of the feed aperture and cancel around the aperture perimeter, giving the desired circular symmetry and pure linear polarization. Conversion from the TE_{11} mode to the TM_{11} mode can be accomplished with an iris, dielectric ring, flare, or, as shown in Fig. 9-45b, with a step. Proper TM_{11} mode amplitude is controlled by the step size and phase is controlled by the distance d. The horn diameter must be greater than 1.3λ and has $HP \approx 1.26\lambda/d_f$; thus, $HP < 55°$ and the feed is usually used with large F/D reflectors [H.6: *Ant. Eng. Hdbk.*, 3rd ed., p. 15–21]. Bandwidths of 10% are possible.

The limited bandwidth of the dual-mode horn can be overcome while still achieving axial beam symmetry, low side lobes, and low cross-polarization by using a *hybrid mode*

TE$_{11}$ TM$_{11}$ TE$_{11}$ and TM$_{11}$
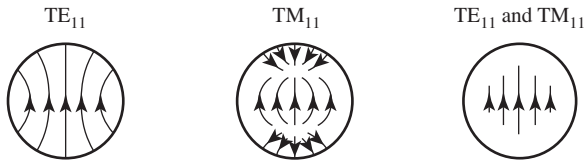
(*a*) Horn aperture electric field distribution.
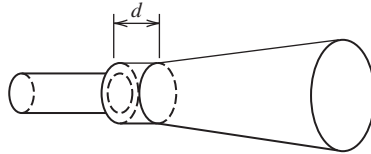
(*b*) A dual-mode horn that uses a step to generate the TM$_{11}$ mode.

Figure 9-45 The dual-mode feed horn antenna.
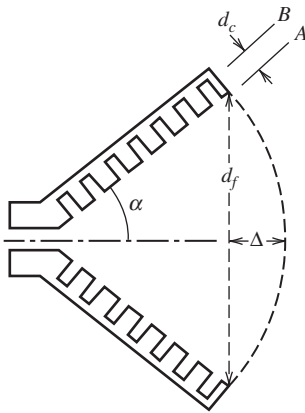
Figure 9-46 The corrugated conical feed horn antenna.

feed. The mixture of TE$_{11}$ and TM$_{11}$ modes occur in a natural way and propagate with a common phase velocity, forming what is known as a hybrid HE$_{11}$ mode. This leads to bandwidths of 1.6:1 or more. The most popular hybrid mode feed is the *corrugated conical horn*. Some variation of the corrugated horn is used as the feed for most of today's microwave reflector antennas. There is no exact formulation for the corrugated horn, but considerable design data are available: [41], [H.8.6: Clarricoats], and [H.3: Milligan, 2nd ed., Sec. 7.3] The basic principle is to provide the same boundary conditions around the inside of the horn. This is accomplished using corrugations (or grooves or teeth) as shown in Fig. 9-46. When the corrugation depths d_c are a quarter wavelength the short circuit at the bottom (*B*) is transformed to an open circuit at the surface (*A*), choking off current. Then the content of the TE$_{11}$ and TM$_{11}$ modes is the same, forming a balanced HE$_{11}$ mode. If there are several corrugations per wavelength, the surface appears to be uniform. The axial current conditions are equivalent to no azimuthal magnetic field. Similarly, an azimuthal electric field is not possible due to the shorting effect of the teeth edges. Thus, all fields decay to zero at the walls, yielding symmetric horn aperture fields and, consequently, a far-field pattern from the horn that is symmetric down to as low as -25 dB. This symmetry along with low side lobes leads to low spillover from the reflector when the horn is used as a feed.

   The phase center of a feed is placed at the focal point of reflector, so it is important to know its location. For long horns (i.e., Δ small) the phase center is near the aperture and moves along the axis toward the apex as Δ increases, reaching the apex for $\Delta > 0.7\lambda$.
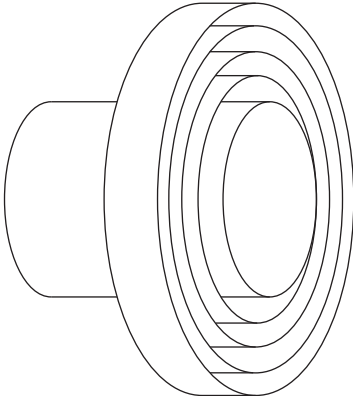
Figure 9-47 The choke horn.

That is, for wide flare angle horns (i.e., α large), the phase center is near the apex. The corrugated horn provides the desireable feature for a feed antenna of a phase center that is stable with frequency.

The corrugated horn in Fig. 9-46 is often called a *scalar horn* because of its field direction independence. Horns with flare angles α from $0°$ to $90°$ are used in practice, but the term scalar horn is usually reserved for the large flare angle cases, which we consider further. A simple model is not available for the patterns of corrugated horns, but the following is a rule of thumb useful in matching a scalar horn to a reflector [H.3: Milligan, 2nd ed., p. 357]:

$$\text{BW}_{-3\text{ dB}} = 0.74\ \alpha \quad \text{and} \quad \text{BW}_{-10\text{ dB}} = 1.51\ \alpha \tag{9-250}$$

Computer-aided design tools for multimode horn antennas are available. [42]

A popular feed horn is the corrugated horn with $\alpha = 90°$, shown in Fig. 9-47, which is called a *choke horn* because the circular waveguide has one or more choke rings around it to "choke off" currents running on the outside of the waveguide and to provide a uniform boundary condition.

9.8   LENS ANTENNAS

Lens antennas are of two major types, dielectric or constrained; these for the most part can also be called non-metallic or metallic, respectively. Both types are introduced in this section. Further details can be found in [H.8.5: Sletten, Ed.] [H.3: Kraus, 3rd ed., Chap. 17], and [H.3: Milligan, 2nd ed., Chap. 9].

9.8.1   Dielectric Lens Antennas

Advantages of the horn antenna over the reflector antenna are low cross-polarization and little or no back lobe. However, the quadratic phase error in the aperture of a conventional horn antenna causes gain loss and high side lobes. This can be prevented using a *dielectric lens* in the aperture to slow down the waves in the central region, allowing edge rays to "catch up" and produce a nearly uniform phase condition for high gain. This is analgous to the lens in the human eye and the lens in eyeglasses. Lenses can also be used to correct for *astigmatism*, which is the condition of non-coincident phase centers in the principal planes. The horn phase error parameters of s and t in (9-131) and (9-116), repectively, for the E- and H-planes, show that phase error decreases as the inverse of horn length. This leads to a horn of large axial length for low phase error. A lens in the aperture that corrects phase error permits a much shorter horn. To reduce lens thickness, a *zoned lens* can be used, which removes material greater than a wavelength from the lens, producing a stepped geometry.
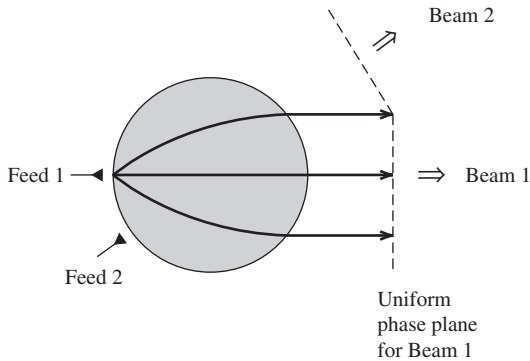
Figure 9-48 The Luneberg lens antenna. The sphere is a dielectric with decreasing dielectric constant from the center to the surface of the sphere.

A dielectric lens antenna of special design is the *Luneberg lens antenna* shown in Fig. 9-48. It consists of a dielectric sphere with a dielectric constant that smoothly varies from 2 at the center to 1 at the surface, following the relation $\varepsilon_r = 2-(r/R)^2$. In practice, the sphere is constructed of nested hemispherical shells of uniform dielectric constant, producing a discrete approximation to the gradient. The classic and simplest Luneberg lens has two foci, one at the surface of the sphere and one at infinity. The ray paths are illustrated in Fig. 9-48 for Feed 1 and Beam 1. Each feed antenna produces a beam directly across the sphere. Feeds are added to produce additional beams. Thus, this is a multibeam antenna. A beam can be electronically scanned by switching between feeds or can be mechanically scanned by rotating a single feed around the sphere.

9.8.2 Constrained Lens Antennas

Direct-radiating electronically scanned antennas provide compact size compared to reflector antennas and, thus, are preferred for many applications. Phased arrays have the disadvantages of narrow bandwidth in many cases, loss, and high cost. The *constrained lens antenna* is a direct radiating antenna that focuses the radiated wave by using arrays of small elements to collect signal from one direction and re-radiate in another. The space feed of Fig. 8-31 is a constrained lens antenna with the phase shifters acting to correct the phase error of the spherical wave from the primary horn. The lens can also electronically steer the beam and provide amplitude control for low side lobes and beam shaping. With cables in the lens providing a uniform output phase front, the bandwidth is unlimited because of the true time delay. When multiple feeds are placed along a focal arc, the antenna is called a *bootlace lens antenna*. Each feed produces a beam in a fashion similar to the Luneberg lens.

A special case of the bootlace lens, but a popular configuration, is the *Rotman lens antenna*. Rotman [43] found the lens parameters that produce three perfect focal points for equal path lengths to the far field. The phase errors for feeds at points in between the foci are small. Parallel plate and printed circuit forms of the Rotman lens are common implementations. Bandwidths greater than 2:1 have been achieved.

REFERENCES

1. M. S. Leong, P. S. Kooi, and Chandra, "Radiation from a Flanged Parallel-Plate Waveguide: Solution by Moment Method with Inclusion of Edge Condition," *IEE Proc.*, Pt. H, Vol. 135, pp. 249–255, Aug. 1988.
2. A. D. Yaghjian, "Approximate Formulas for the Far Field and Gain of Open-Ended Rectangular Waveguide," *IEEE Trans. on Ant. & Prop.*, Vol. AP-32, pp. 378–384, April 1984.
3. W. L. Stutzman, "Estimating Directivity and Gain of Antennas," *IEEE Trans. on Ant. & Prop. Mag.*, Vol. 40, pp. 7–11, Aug. 1998.

4. A. F. Sciambi, "Instant Antenna Patterns," *Microwaves*, Vol. 5, pp. 48–60, June 1966.

5. M. Abramowitz and I. Stegun, Eds., *Handbook of Mathematical Functions,* NBS Applied Mathematics Series 55, U.S. Printing Office, 1964, Chap. 7.

6. M. J. Maybell and P. S. Simon, "Pyramidal Horn Gain Calculation with Improved Accuracy," *IEEE Trans. on Ant. & Prop.*, Vol. 41, pp. 884–889, July 1993.

7. E. V. Jull and L. E. Allan, "Gain of an *E*-Plane Sectoral Horn — A Failure of the Kirchhoff Theory and a New Proposal," *IEEE Trans. on Ant. & Prop.*, Vol. AP-22, pp. 221–226, March 1974.

8. E. V. Jull, "Errors in the Predicted Gain of Pyramidal Horns," *IEEE Trans. on Ant. & Prop.,* Vol. AP-21, pp. 25–31, Jan. 1973.

9. J. Aurand, "Pyramidal horns, Part 1: Analysis of Directivity as a Function of Aperture Phase Errors," *IEEE Trans. on Ant. & Prop. Society Newsletter*, pp. 33–34, June 1989.

10. J. Aurand, "Pyramidal horns, Part 2: Design of Horns for any Desired Gain and Aperture Phase Error," *IEEE Trans. on Ant. & Prop. Society Newsletter*, pp. 25–27, Aug. 1989.

11. D. Hawkins and F. Thompson, "Modifications to the Theory of Waveguide Horns," *IEE Proc.-H*, Vol. 140, pp. 381–386, Oct. 1993.

12. J. F. Ramsey, "Lambda Functions Describe Antenna/Diffraction Patterns," *Microwaves*, Vol. 6, pp. 69–104, June 1967.

13. D-W Duan and Y. Rahmat-Samii, "A Generalized Three-Parameter (3-P) Aperture Distribution for Antenna Applications," *IEEE Trans. on Ant. & Prop.*, Vol. 40, pp. 697–713, June 1992.

14. W. L. Stutzman, S. W. Gilmore, and S. H. Stewart, "Numerical Evaluation of Radiation Integrals for Reflector Antenna Analysis Including a New Measure of Accuracy," *IEEE Trans. on Ant. & Prop.*, Vol. 36, pp. 1018–1023, July 1988.

15. Y. Rahmat-Samii and V. Galindo-Israel, "Shaped Reflector Antenna Analysis Using the Jacobi-Bessel Series," *IEEE Trans. on Ant. & Prop.*, Vol. AP-28, pp. 425–435, July 1980.

16. A. D. Yaghijan, "Equivalence of Surface Current and Aperture Field Integration for Reflector Antennas," *IEEE Trans. on Ant. & Prop.*, Vol. AP-32, pp. 1355–1358, Dec. 1984.

17. Y. Rahmat-Samii, "A Comparison between GO/Aperture-Field Integration and Physical-Optics Methods for Offset Reflectors," *IEEE Trans. on Ant. & Prop.*, Vol. AP-32, pp. 301–306, March 1984.

18. W. V. T. Rusch, A. Prata, Y. Rahmat-Samii, and R. A. Shore, "Derivation and Application of the Equivalent Paraboloid for Classical Offset Cassegrain and Gregorian Antennas," *IEEE Trans. on Ant. & Prop.*, Vol. 38, pp. 1141–1149, Aug. 1990.

19. V. Galindo, "Design of Dual-Reflector Antennas with Arbitrary Phase and Amplitude Distributions," *IEEE Trans. on Ant. & Prop.*, Vol. AP-12, pp. 403–408, July 1964.

20. W. F. Williams, "High Efficiency Antenna Reflector," *Microwave J.*, Vol. 8, pp. 79–82, July 1965.

21. R. Mittra and V. Galindo-Israel, "Shaped Dual Reflector Synthesis," *IEEE Trans. on Ant. & Prop. Soc. Newsletter*, pp. 5–9, Aug. 1980.

22. P-S Kildal, "Synthesis of Multireflector Antennas by Kinematic and Dynamic Ray Tracing," *IEEE Trans. on Ant. & Prop.*, Vol. 38, pp. 1587–1599, Oct. 1990.

23. A. Cha, "An Offset Dual Shaped Reflector with 84.5 Percent Efficiency," *IEEE Trans. on Ant. & Prop.*, Vol. AP-31, pp. 896–903, Nov. 1983.

24. S. V. Parekh and J. H. Cook, "Reshaped Subreflectors Reduce Antenna Sidelobes," *Microwaves*, Vol. 16, pp. 170–173, May 1977.

25. M. A. Terada and W. L. Stutzman, "Design of Offset Parabolic Reflector Antennas for Low Cross Polarization and Low Side Lobes," *IEEE Ant. & Prop. Mag.*, Vol. 35, pp. 46–49, Dec. 1993.

26. M. A. Terada and W. L. Stutzman, "Cross Polarization and Beam Squint in Single and Dual Offset Reflector Antennas," *Electromagnetics*, Vol. 16, pp. 633–650, Nov./Dec. 1996.

27. T-S Chu and R. H. Turin, "Depolarization Properties of Offset Reflector Antennas," *IEEE Trans. on Ant. & Prop.*, Vol. AP-21, pp. 339–345, May 1973.

28. M. A. Terada and W. L. Stutzman, "Computer-Aided Design of Reflector Antennas," *Microwave J.*, Vol. 38, pp. 64–73, Aug. 1995.

29. D. W. Duan and Y. Rahmat-Samii, "Beam Squint Determination in Conic-Section Reflector Antennas with Circularly Polarized Feeds," *IEEE Trans. on Ant. & Prop.*, Vol. 39, pp. 612–619, May 1991.

30. Y. T. Lo, "On the Beam Deviation Factor of a Parabolic Reflector," *IEEE Trans. on Ant. & Prop.*, Vol. AP-8, pp. 347–349, May 1960.

31. J. Ruze, "Lateral-Feed Displacement in a Paraboloid," *IEEE Trans. on Ant. & Prop.*, Vol. AP-13, pp. 660–665, Sept. 1965.

32. Y. Rahmat-Samii and V. Galindo-Israel, "Scan Performance of Dual Offset Reflector Antennas for Satellite Communications," *Radio Science*, Vol. 16, pp. 1093–1099, Nov./Dec. 1981.
33. P. C. Werntz, W. L. Stutzman, and K. Takamizawa, "A High-Gain Trireflector Antenna Configuration for Beam Scanning," *IEEE Trans. on Ant. & Prop.*, Vol. 42, pp. 1205–1214, Sept. 1994.
34. B. Shen and W. L. Stutzman, "A Scanning Spherical Tri-Reflector Antenna with a Moving Flat Mirror," *IEEE Trans. Ant. & Prop.*, Vol. 43, pp. 270–276, March 1995.
35. J. Ruze, "Antenna Tolerance Theory—A Review," *Proc. IEEE*, Vol. 54, pp. 633–640, April 1966.
36. A. W. Rudge et al., Eds., *The Handbook of Antenna Design*, Vol. I, Van Norstrand Reinhold, New York, 1988, Chap. 3.
37. W. T. Smith and W. L. Stutzman, "A Pattern Synthesis Technique for Array Feeds to Improve Radiation Performance of Large Distorted Reflector Antennas," *IEEE Trans. on Ant. & Prop.*, Vol. 40, pp. 57–62, Jan. 1992.
38. A. W. Rudge and M. J. Withers, "New Techniques for Beam Steering with Fixed Parabolic Reflectors," *Proc. IEE*, Vol. 118, pp. 857–863, July 1971.
39. A. P. King, "The Radiation Characteristics of Conical Horn Antennas," *Proc. of IRE*, Vol. 38 pp. 249–251, March 1950.
40. P. D. Potter, "A New Horn Antenna with Suppressed Side Lobes and Equal Beamwidths," *Microwave J.*, Vol. 6, pp. 71–78, June 1963.
41. B. M. Thomas, "Design of Corrugated Conical Horns," *IEEE Trans. on Ant. & Prop.*, Vol. AP-26, pp. 367–372, Mar. 1978.
42. Mician GMBH, "Horn Antenna Synthesis CAD Tool," *Microwave J.*, Vol. 51, pp. 150–158, Mar. 2008.
43. W. Rotman and R. F. Turner, "Wide-Angle Microwave Lens for Line Source Applications," *IEEE Trans. on Ant. & Prop.*, Vol. AP-11, pp. 623–632, Nov. 1963.

PROBLEMS

9.1-1 An ideal dipole with an infinitesimal current centered along the *z*-axis has only a θ-component of **E**. If this uniform current element is now rotated to line up with the *x*-axis, there will be both θ- and ϕ-components. Make the necessary changes to the far-zone **E** expression for the *z*-directed current case to obtain the far-zone **E** expression for the *x*-directed current case. Sketch the **E** and **H** field orientations (not the pattern) for the *x*-directed ideal dipole in the *xz*- and *yz*-planes.

9.1-2 Use the principle of duality to derive (9-5) and (9-6) from (9-3) and (9-4), respectively.

9.1-3 Show that (9-24) follows from (9-22) and (9-23).

9.1-4 Show how (9-26) follows from (9-6).

9.1-5 Show how (9-24) follows from (9-17).

9.1-6 If the incident field in Example 9-1 is *x*-polarized, write radiation field E_ϕ and the pattern $F(\theta)$. Your answer will be that of (5-15) with a coordinate change. Is the appropriate boundary condition for E_ϕ satisfied on the conducting plane?

9.2-1 Derive (9-34).

9.2-2 Use geometric arguments to prove that $du\,dv = \cos\theta\,d\Omega$ where $d\Omega = \sin\theta\,d\theta\,d\phi$; that is, sketch a hemisphere and project the intersection of the differential $d\Omega$ with the hemisphere onto the *uv*-plane.

9.2-3 For the open-ended waveguide of Example 9-3:
   (a) Numerically evaluate the *E*-plane pattern expressions to verify the results in Fig. 9-10*a*. Plot the two computed patterns.
   (b) Repeat (*a*) for the *H*-plane patterns of Fig. 9-10*b*.

9.2-4 The open-ended waveguide of Fig. 9-9 can be more accurately modeled by including the reflection coefficient at the aperture Γ and the waveguide phase constant $\beta_g = 2\pi/\lambda_g$, where $\lambda_g = \lambda/\sqrt{1 - (\lambda/2a)^2}$. The aperture fields are given by [2]

$$E_y = E_o\,(1 + \Gamma)\,\cos\frac{\pi x}{a}\,e^{-j\,\beta_g\,z} \qquad \text{and} \qquad H_x = -\frac{E_o}{Z_o}\,(1 - \Gamma)\,\cos\frac{\pi x}{a}\,e^{-j\,\beta_g z}$$

where $Z_o = \omega\,\mu/\beta_g$.
   (a) Derive complete expressions for the far-field electric field components. Use the magnetic and electric equivalent current formulation.

(b) Write the normalized *E*- and *H*-plane pattern expressions.

(c) Write the normalized obliquity factor in the principal planes for the case when the waveguide is matched ($\Gamma = 0$) and the guide phase velocity is that of free space.

9.2-5 Estimate the horizontal and vertical beamwidths in degrees of the Hertz cylindrical-parabolic reflector of Fig. 1-2 by assuming a uniform aperture distribution in the horizontal direction and cosine in the vertical direction approximating a vertically oriented dipole feed. The beam will be narrower than Hertz found because of the ideal excitation assumed.

9.3-1 Prove (9-63).

9.3-2 Prove that the uniform amplitude aperture excitation yields the highest directivity of all uniform phase excitations. *Hint*: Use (9-66) and the Schwarz inequality

$$\left| \int\int_S fg \, dS' \right|^2 \leq \int\int_S f^2 \, dS' \int\int_S g^2 \, dS'$$

for any functions f and g. Let $g = 1$ and f equal the aperture field.

9.3-3 Show that the aperture taper efficiency is $\frac{2}{3}$ for a rectangular aperture with a uniform amplitude distribution in one direction and a cosine-squared distribution in the other.

9.3-4 A rectangular aperture (L_x by L_y) has a field distribution that is cosine-tapered in both the *x*- and *y*-directions. Derive the directivity expression. What is the aperture taper efficiency?

9.3-5 Compute the directivity in decibels for a rectangular aperture with $L_x = 10\lambda$ and $L_y = 20\lambda$ for (a) A completely uniform aperture illumination and (b) A cosine amplitude taper in one direction and a uniform taper in the other aperture direction.

9.3-6 Evaluate the aperture taper efficiency for a triangular tapered, rectangular aperture distribution:

$$\mathbf{E}_a(x', y') = \left[1 - \frac{2|x'|}{L_x} \right] \hat{\mathbf{x}}, \qquad |x'| \leq \frac{L_x}{2}, \qquad |y'| \leq \frac{L_y}{2}$$

9.3-7 Strictly speaking, is the uniform aperture distribution physically realizable? Why?

9.3-8 An antenna operating at 150 MHz has a physical aperture area of 100 m², a gain of 23 dB, and a directivity of 23.5 dB. Compute: (a) Effective aperture, A_e, (b) Maximum effective aperture, A_{em}, (c) Aperture efficiency, ε_{ap}, (d) Radiation Efficiency e_r, and (e) Aperture taper efficiency ε_t.

9.3-9 Verify (4-56).

9.3-10 The general antenna gain relation of (9-78) includes a frequency-squared dependence. However, the class of frequency-independent antennas display nearly constant gain over large frequency variations. Explain this apparent paradox.

9.3-11 A horn antenna with a 185.5×137.4 cm rectangular aperture has the following measured parameter values at 0.44 GHz: $HP_{E^\circ} = 30°$, $HP_{H^\circ} = 27°$, and $G = 15.5$ dB.

(a) Compute the aperture efficiency.

(b) Estimate the gain from the measured half-power beamwidths.

9.3-12 Repeat Prob. 9.3-11 for a horn with a 28.85×21.39 cm aperture and $HP_{E^\circ} = 12°$, $HP_{H^\circ} = 13°$, and $G = 22.1$ dB at 6.3 GHz.

9.3-13 A 3.66-m (12-ft) diameter circular parabolic reflector operates at 460 MHz. The measured parameters of this antenna are $G = 22.2$ dB and $HP_{E^\circ} = HP_{H^\circ} = 12.5°$. Estimate the gain using both (9-97) and (9-95).

9.3-14 Estimate the gain of a circular parabolic reflector operating at 28.56 GHz in two ways:

(a) Using only its size, which is 1.22 m (4 ft) in diameter.

(b) Using only the measured half-power beamwidths, which are $HP_{E^\circ} = 0.605°$ and $HP_{H^\circ} = 0.556°$.

9.3-15 (a) Prove (9-86) for a separable distribution using (9-73) and Prob. 5.2-11.

(b) Using $\varepsilon_{ap} = \varepsilon_{apx} \varepsilon_{apy}$ and assuming $e_r = 1$, write expressions for ε_{apx} and ε_{apy}. using the results from (a).

(c) For a general aperture distribution show that

$$\varepsilon_t = \frac{1}{A_p} \frac{\left[\displaystyle\iint E_a \, dS\right]^2}{\displaystyle\iint |E_a|^2 dS}$$

9.3-16 Show that the directivity-beamwidth product for a uniform phase rectangular aperture with a cosine amplitude taper in the *H*-plane and uniform amplitude in the *E*-plane is 35,230 deg^2.

9.3-17 A geostationary satellite is 42,000 km from the center of the Earth. If the -3-dB pattern points fall near the edge of the Earth, find an approximate value for the spacecraft antenna gain. Note that the result is independent of frequency.

9.3-18 A Gaussian power pattern of half-power beamwidth HP in degrees is

$$P_n(\theta) = e^{-(4 \ln 2)(\theta/\text{HP})^2}$$

Derive the following approximate directivity expression for narrow beam Gaussian patterns:

$$D \approx \frac{36,407}{\text{HP}^2}$$

9.3-19 Horn antennas used as feeds for reflectors have patterns that are well approximated by

$$F(\theta) = \cos^q(\theta), \qquad 0 \le \theta \le \pi/2$$

  (a) Derive the directivity expression $D = 2(2q + 1)$.
  (b) Compare directivity values based on (a) to those using (9-94) and (4-56) for $q = 0, 1, 5, 10, 50$; tabulate the results.

9.4-1 Derive the expression for R_H in (9-100).

9.4-2 Derive the *H*-plane sectoral horn radiation field expression (9-107) and (9-108) by changing to complex exponentials and then completing the square in the exponents in the integrand.

9.4-3 In the *H*-plane pattern expression of (9-117) and (9-118) for an *H*-plane sectoral horn:
  (a) Show that s_1 follows from s'_1 of (9-109).
  (b) Show that the phase term $(\pi/8t)[(A/\lambda) \sin \theta + 1/2]^2$ follows from the corresponding term in (9-108).

9.4-4 Derive the directivity formula of (9-120) for an *H*-plane sectoral horn from (9-66). The numerator in (9-66) can be evaluated using (9-106) to (9-108).

9.4-5 Use a computer program to evaluate Fresnel integrals. Compute $C(x)$ and $S(x)$ for $x = 0, 1, 2, 3, 4,$ and 5. Tabulate the values along with those from a math table, giving the deviation from the known values.

9.4-6 The *H*-plane pattern for an *H*-plane sectoral horn arises from the first integral in (9-106).
  (a) First evaluate this integral for a no phase error condition.
  (b) Show that the on-axis value of the *H*-plane pattern relative to the on-axis value of the zero phase error case is given by

$$\frac{\pi}{16\sqrt{t}} I(\theta = 0^\circ, \phi = 0^\circ)$$

  (c) Evaluate this for t $= \frac{1}{8}, \frac{1}{4}, \frac{3}{8},$ and $\frac{1}{2}$ and compare to the values from Fig. 9-13.

9.4-7 An *H*-plane sectoral horn antenna has an axial length of 5λ and a flare half-angle α_H of 12.6°.
  (a) Plot the polar plot of the *H*-plane radiation pattern in decibels.
  (b) Compute the directivity function $\lambda D_H/b$ using (9-120) and compare to that obtained from Fig. 9-14.
  (c) Since the aperture is not large relative to a wavelength, use the zero phase error directivity formula of (9-71) to compute $\lambda D_H/b$.

9.4-8 Design an optimum *H*-plane sectoral horn antenna with 12.15-dB gain at 10 GHz. It is fed with a WR90 waveguide.

(a) Find the horn dimensions employing Fig. 9-14.

(b) Draw the *H*-plane horn geometry to scale.

(c) Use (9-120) to compute the directivity as a check.

9.4-9 Repeat Prob. 9.4-8(a), except use (9-120) and (9-21) instead of Fig. 9-14.

9.4-10 Derive the *E*-plane sectoral horn far-zone electric field expression of (9-129)

9.4-11 Show how the *E*-plane pattern magnitude expression for an *E*-plane sectoral horn of (9-132) follows from (9-129).

9.4-12 Use physical reasoning to explain why the phase error parameters of optimum *E*-plane and *H*-plane sectoral horns are different.

9.4-13 An *E*-plane sectoral horn antenna is attached to a WR90 waveguide. Determine the horn dimensions for a half-power beamwidth of 11° in the *E*-plane and an optimum gain of 14.9 dB at 10 GHz.

9.4-14 An *E*-plane sectoral horn has an *E*-plane aperture height of 24.0 cm and a half-flare angle of 16.5°. It is attached to a WR284 waveguide. Compute the gain at 3.75 GHz (a) using (9-134a) and (b) using (9-139).

9.4-15 Derive (9-141) by starting with (9-78) and using $\varepsilon_{\mathrm{ap}} = \varepsilon_t \varepsilon_{\mathrm{ap}}^E \varepsilon_{\mathrm{ap}}^H$.

9.4-16 Start with $A = 18.61$ cm and verify all horn dimensions given in Example 9-7.

9.4-17 The aperture efficiency in Fig. 9-20 for the pyramidal horn of Example 9-7 is based on the aperture quadratic phase error approximation.

(a) Find the aperture efficiency by direct evaluation of (9-120c) and (9-134c) from 8 to 13 GHz to verify Fig. 9-20. Then repeat using the exact phase errors (9-123) and (9-137). Compare these results.

(b) Evaluate aperture efficiency at 8, 10, and 13 GHz using the approximate formulas in (9-149) and (9-150). Compare to results from (a); tabulate values.

9.4-18 Explain why an optimum horn is designed for about 50% aperture efficiency at a frequency near the low end of its operating band as in Example 9-7.

9.4-19 Design an optimum gain pyramidal horn antenna connected to a WR90 waveguide with 20-dB gain at 10 GHz. (a) Give all horn dimensions. (b) Evaluate the directivity at 10 GHz using the exact phase errors.

9.4-20 A commercially available standard gain horn antenna operates from 18 to 26.5 GHz and has a WR42 waveguide input. The gain is 24.7 dB at 24 GHz.

(a) Use optimum gain design principles to determine the horn geometry values in centimeters.

(b) Plot the *E*- and *H*-plane polar patterns in decibels including the $(1 + \cos \theta)/2$ factor.

(c) Determine the half-power beamwidths from pattern calculations and compare to the simple formula values.

(d) Evaluate the gain at the design frequency using the exact phase errors. Give the aperture efficiency values.

(e) Compare the gain to that calculated using the approximation in (9-95) for both beamwidth values found in (c).

9.4-21 Repeat Prob. 9.4-20 for a pyramidal horn designed for optimum operation at 1 GHz and a gain of 15.45 dB. It is connected to WR975 waveguide with $a = 9.75$ in. $= 24.765$ cm and $b = 4.875$ in. $= 12.3825$ cm.

9.4-22 For Example 9-7, compute the half-power beamwidths using line source models from Chap. 5 for the same amplitude tapers as in the horn aperture. Explain why there are deviations from the values in Example 9-7.

9.4-23 Derive the following relationship that must be satisfied for a physically realizable pyramidal horn antenna:

$$\frac{R_1}{\lambda} = \frac{1 - \dfrac{b/\lambda}{B/\lambda}}{1 - \dfrac{a/\lambda}{A/\lambda}} \frac{R_2}{\lambda}$$

9.4-24 *A square main beam horn antenna.* It is often desirable to have equal principal plane half-power beamwidths. This problem develops a design technique for a so-called square main beam

pyramidal horn. If optimum design techniques under the condition of a square main beam are used, the resulting horn dimensions will render a horn that cannot be constructed. To avoid this problem, we can design for a square main beam and aim for *near* optimum conditions. To do this we first determine the aperture dimensions that give the desired beamwidths and optimum operation. Then the axial lengths are adjusted to provide a physically realizable structure. This will probably not move the operating point too far from optimum. Follow this procedure to design a square main beam horn at 8 GHz with 12° beamwidths and fed by a WR90 waveguide.

   (a) Determine A/λ and B/λ.
   (b) Use the results of Prob. 9.4-23 for adjusting the axial lengths. Do this to keep the fractional increase or decrease of both the same—that is, use

$$\frac{R_1}{\lambda} = \frac{R_{1op}}{\lambda} F \qquad \text{and} \qquad \frac{R_2}{\lambda} = \frac{R_{2op}}{\lambda} = \frac{1}{F}$$

and solve for the constant F.
   (c) Evaluate the final phase error parameters t and s.
   (d) Give the horn dimensions in centimeters.
   (e) Evaluate the gain.
   (f) Compute the aperture efficiency.

9.5-1 Write the radiated electric field expression analogous to (9-168) using the equivalent current formulation that includes both electric and magnetic surface current densities.

9.5-2 Verify that the uniform circular aperture pattern of (9-170) is unity for $\theta = 0$.

9.5-3 Derive the pattern expression $f(\theta, n, C)$ in Table 9-2b for a parabolic taper on a pedestal.

9.5-4 For a parabolic-on-a-pedestal circular aperture distribution: (a) Derive (9-181) using (9-66). (b) Evaluate ε_t for $n = 1$ and 2 for a -10-dB edge taper.

9.5-5 For a tapered circular aperture: (a) Prove that $c = c_x c_y = \sqrt{\pi \varepsilon_t}/2$. (b) Show by examining the values in Table 9-2b that $DB_{cir} \approx 39{,}000$ $\deg^2$.

9.5-6 Derive the directivity-beamwidth product, DB_{cir}, for a circular aperture with a parabolic-on-a -12-dB pedestal distribution (Case 6 in Table 9-1).

9.6-1 Derive (9-199a).

9.6-2 Using sketches, show that the components of the incident and reflected electric fields tangent to a parabolic reflector cancel and that (9-200) holds.

9.6-3 Plot the edge illumination in decibels for a circular parabolic reflector due to spherical spreading loss only (i.e., the feed is isotropic) as a function of F/D from 0 and 1.

9.6-4 A commercially available parabolic reflector antenna operating at 2.1 GHz has an aperture diameter of 1.83 m (6 ft). Compute the gain in decibels.

9.6-5 A commercially available parabolic reflector antenna operating at 11.2 GHz has an aperture diameter of 3.66 m (12 ft). Compute the gain in decibels.

9.6-6 Analyze the reflector in Example 9-8 using a reflector computer code (see Appendix G). Model the feed using a $\cos^q \theta_f$ pattern with a -10-dB beamwidth of 104°. Tabulate the values for HP, SLL, G, and ε_{ap}. Include values for the canonical distribution approach. Plot the pattern in decibels.

9.6-7 A commercial axisymmetric reflector antenna used for Ku-band satellite reception (11.95 GHz midband) is 2.4 m in diameter and has an $F/D = 0.37$. Assuming a $\cos^q \theta_f$ feed pattern: (a) Use a canonical aperture distribution to determine reflector performance, (b) Use a reflector code (see Appendix G) to evaluate performance. Tabulate results from (a) and (b) including G, HP, and SLL.

9.6-8 Use a reflector code (see Appendix G) to determine the following performance parameters for an offset parabolic reflector with $D = 100\lambda$, $H = 70\lambda$, $F/D_p = 0.466$, $\psi_f = 34.72°$, and a $\cos^q \theta_f$ feed with $q = 13.0897$: (a) Gain, (b) SLL, (c), XPOL peak location, (d) XPOL peak value in decibels relative to the main beam peak. (e) Plot the pattern in decibels out to 3° in the plane normal to the offset plane (i.e., $\phi = 90°$).

9.6-9 A popular commercial offset parabolic reflector antenna for receiving direct broadcast television (12.45-GHz midband) has the following geometric parameters: $D = 45.70$ cm (18 in.), $F = 26.23$ cm, $D_p = 94.00$ cm, and $H = 24.15$ cm. The beamwidth between -10 dB points on the feed pattern is 80.8° and the peak of the feed pattern is aimed 49.5° from the

reflector axis. Use a computer program to evaluate the radiation pattern in the principal planes. Tabulate the values of G, HPs, SLLs, and XPOLs.

9.6-10 Derive (9-224) and (9-226).

9.6-11 Derive (9-230) and (9-231).

9.6-12 Prove that the ideal feed of (9-228) produces 100% aperture efficiency.

9.6-13 (a) Derive an expression for the aperture efficiency of an axisymmetric reflector fed with an isotropic feed antenna, and (b) Evaluate for $F/D = 0.25, 0.5,$ and 1.

9.6-14 Derive the illumination efficiency expression (9-232) for a $\cos^q \theta_f$ feed pattern for $q = 2$.

9.6-15 A geostationary satellite transmits at 4 GHz using a parabolic reflector antenna. The peak of the beam is directed toward the center of the Earth disk and the -3 dB pattern points fall on the edge of the Earth. Find the gain in decibels. (Earth radius = 6,400 km; distance from center of Earth to orbit = 42,000 km.)

9.6-16 This problem serves to verify the claim that the -11-dB edge illlumination yields about $\varepsilon_i \approx 0.82$ under a variety of axisymmetric reflector system cases. For $\cos^2 \theta_f$ feed patterns as in (9-229) and values of $q = 1, 2,$ and 3, find the F/D value of the optimum gain axisymmetric reflector. Tabulate the following for each q value: F/D, $2\theta_o$, feed $\text{BW}_{-10\,\text{dB}}$, ε_s, ε_i found using (9-232), and ε_t.

9.6-17 Compute the blockage efficiency for a reflector of optimum gain for $A_b/A_p = 0.1, 1, 2, 5,$ and 10%.

9.6-18 A subreflector in a Cassegrain dual reflector has a diameter which is 10% of the main reflector diameter. Find the aperture blockage efficiency assuming optimum operation.

9.7-1 Find the half-power and -10-dB beamwidths of a $\cos^q \theta_f$ feed pattern required to produce an edge illumination of -11 dB in an axisymmetric reflector with $F/D = 0.4$. Give the value of q.

9.7-2 Plot the aperture electric field amplitude distribution for an axisymmetric reflector with $F/D = 0.3$ and a $\cos^2 \theta_f$ feed pattern. Also, show on the same plot the parabolic-squared aperture distribution with the same edge illumination.

9.7-3 For the reflector of Example 9-9, (a) Calculate the gain in decibels using aperture efficiency, and (b) Use a reflector code to find the gain and compare to the value from (a).

9.7-4 Derive the aperture electric field expression for the Huygens' source of Fig. 9-42.

9.7-5 A commercial offset parabolic reflector antenna has a diameter of 1.8 m is used for Ku-band satellite communications. It is just fully offset (that is, $h = 0$) and $F/D_p = 0.305$. The feed has a -10-dB beamwidth of $76.8°$. For the middle of the transmit band at 14.25 GHz: (a) Determine the feed pointing angle that produces nearly equal edge illumination at the upper and lower reflector edges. (b) Use a reflector code to evaluate the reflector performance. (c) Find a canonical distribution that approximates the aperture distribution. Tabulate values from (b) and (c) for as many of the following parameters as possible: G, HP, SLL, XPOL, and ε_{ap}.

9.7-6 An optimum gain conical horn is used to feed an axisymmetric parabolic reflector with $F/D = 0.44$. Using HP $\approx 1.14 \, d_f/\lambda$ as an average beamwidth expression, find the d_f/λ value for maximum illumination efficiency.

9.7-7 Derive (9-247).

Chapter **10**

Antenna Synthesis

10.1 THE ANTENNA SYNTHESIS PROBLEM

The study of antennas thus far in this book has been one of analysis, where antenna performance is determined once the antenna type and physical parameter values are given. Antenna engineering, like most engineering fields, has both analysis and synthesis problems. *Antenna analysis* uses analytical formulation, simulation, and/or measurement to understand how an antenna operates and to determine its performance characteristics. *Antenna synthesis* is the reverse of analysis: an antenna structure is derived to meet a given set of performance specifications, most often the radiation pattern. *Antenna design* could incorporate analysis or synthesis but usually applies engineering judgment in the following areas: formulating the electrical and mechanical specifications, selecting the antenna type, and performing iterative analysis to arrive at an antenna that meets specifications.

Antenna synthesis in general does not require the *a priori* selection of an antenna type. As an analogy, in circuit theory it is possible to determine the needed circuit components and their values for achieving a specified frequency response characteristic. There is no single antenna synthesis method that will yield an optimum antenna for the given specifications. However, computing packages are emerging that include a suite of antenna types that can be simulated using a single user interface to greatly speed up the design process. In this chapter, we present synthesis methods for the most popular antenna type variables, shown in the left column of Table 10-1. The time domain is shown along with frequency domain as antenna variables. Here we continue to work in the frequency domain. Time domain synthesis using the singularity expansion method is treated in [H.6: Balanis, Sec. 13.5].

10.1.1 Formulation of the Synthesis Problem

We will pose the antenna synthesis problem as one of determining the excitation of a given antenna type that leads to a radiation pattern which suitably approximates a desired pattern. The desired pattern can vary widely depending on the application and has the variables listed in the right-hand column of Table 10-1. To illustrate, consider a communication satellite in synchronous orbit that is required to generate separate antenna beams for the western United States and for Alaska. Two main beams are required, both shaped for nearly uniform illumination of each region. Also low side lobes may be specified to minimize interference over other regions of the earth, but higher side lobes could be permitted for directions not toward the earth. This type of pattern has multiple shaped main beams and a shaped side lobe envelope.

Table 10-1 Antenna Synthesis Variables

Antenna Type Variables	Radiation Pattern Variables
• Antenna Continuity 　　Continuous 　　Discrete—array • Antenna Shape 　　Linear 　　Planar 　　Conformal 　　Three-dimensional • Antenna Size • Domain 　　Frequency 　　Time	• Main Beam 　　Narrow main beam 　　　Single beam 　　　Multiple beams 　　Shaped main beam • Pattern Nulls at Specified Angles • Side Lobes 　　Nominal side lobes 　　Low side lobes 　　Shaped envelope side lobes

The antenna itself can take many forms, as listed in the left-hand column of Table 10-1. Antenna type refers to the geometry of the antenna and consists of continuity, shape, and size. The performance of an antenna is used to define the antenna types in Table 1-4 and includes the performance parameters listed in Table 1-3. Performance parameters other than pattern shape can be included in the synthesis problem specifications. In this chapter, we consider the problem of pattern synthesis. The remaining performance parameters are considered elsewhere in the book. A general synthesis procedure would yield the antenna type and its excitation that produces the best approximation to specified performance values including the desired pattern shape. No such general synthesis method exists. Instead, synthesis methods have been developed for each antenna type. The discussion of synthesis in this chapter is divided between continuous and discrete (array) antenna types. Before addressing these methods, we present further general remarks.

If the radiation electric field components E_θ and E_ϕ are specified in the synthesis problem, a secondary synthesis problem can be formulated in terms of antenna aperture field transform components. For example, the aperture magnetic equivalent surface current solution of (9-26) can be solved, giving

$$\begin{bmatrix} P_x \\ P_y \end{bmatrix} = \left(j\beta \frac{e^{-j\beta r}}{2\pi r} \right)^{-1} \begin{bmatrix} \cos\phi & \sin\phi \\ -\cos\theta\sin\phi & \cos\theta\cos\phi \end{bmatrix}^{-1} \begin{bmatrix} E_\theta \\ E_\phi \end{bmatrix} \qquad (10\text{-}1)$$

This can be used to obtain P_x and P_y from specified functions E_θ and E_ϕ. The problem is then of synthesizing desired functions P_x and P_y, which are Fourier transforms of the aperture electric field components; see (9-18). The process is similar for each of P_x and P_y. Therefore, we let $f(\theta, \phi)$ be the normalized pattern factor for either and frame our discussions using $f(\theta, \phi)$. As another example, consider a line source along the z-axis. If $F_d(\theta)$ is the normalized desired radiation pattern, then the desired pattern factor is

$$f_d(\theta) = \frac{F_d(\theta)}{\sin\theta} \qquad (10\text{-}2)$$

This chapter discusses synthesis of the pattern function $f(\theta, \phi)$ that provides an approximation to the desired pattern $f_d(\theta, \phi)$. The pattern synthesis techniques will be presented for one-dimensional formulations with a geometry yielding $f(\theta)$. That is, the continuous form (line sources) and the discrete form (linear arrays) will be treated. However, these results can be applied to two-dimensional antennas such as planar aperture and planar array antennas. Direct application of the methods is possible if the

two-dimensional aperture distribution is separable (see Sec. 9.2.2). Then the synthesized pattern function f is used to represent each principal plane pattern. Synthesis methods can be classified by antenna or pattern type. Only a few methods exist that can be applied to a variety of antenna and pattern types, and one such method is presented in Sec. 10.5.

Usually, synthesis methods for shaped beam patterns are completely different from those for low side lobe, narrow beam patterns, so we will classify the methods by pattern type. Line source and linear array synthesis principles with applications to shaped beam patterns are detailed in Secs. 10.2 and 10.3. Low side lobe, narrow main beam methods are presented in Sec. 10.4.

10.1.2   Synthesis Principles

The radiation pattern of a line source of current (actual or equivalent) along the z-axis and of length L is given by (5-28) for the geometry of Fig. 2-9a. The element factor $g(\theta) = \sin\theta$ is accounted for separately; for narrow-beam, broadside line sources, it is negligible. The normalized pattern factor of a line source follows from (5-29) as[1]

$$f(\theta) = \frac{1}{\lambda}\int_{-L/2}^{L/2} i(z)e^{i\beta z\,\cos\theta}dz \tag{10-3}$$

where $i(z)$ is the normalized form of the current function $I(z)$, and it is usually normalized such that (10-3) produces a pattern $f(\theta)$ that is unity at its maximum. The linear phase shift that scans the main beam is contained in $i(z)$; for example, see (5-1). For convenience, we define

$$w = \cos\theta \quad \text{and} \quad s = \frac{z}{\lambda} \tag{10-4}$$

and w is related to the u in Chap. 5 through $u = (\beta L/2)w$. Then (10-3) becomes

$$f(w) = \int_{-L/2\lambda}^{L/2\lambda} i(s)e^{j2\pi ws}\,ds \tag{10-5}$$

This equation forms the relationship between the relative current distribution $i(s)$ and the normalized pattern factor $f(w)$.

Since the current distribution $i(z)$ extends only over the length L, (that is, $i(s)$ is zero for $|s| > L/2\lambda$), the limits of the integral in (10-5) can be extended to infinity, giving

$$f(w) = \int_{-\infty}^{\infty} i(s)e^{j2\pi ws}\,ds \tag{10-6}$$

This is recognized as a Fourier transform. The corresponding inverse Fourier transform is

$$i(s) = \int_{-\infty}^{\infty} f(w)e^{-j2\pi sw}\,dw \tag{10-7}$$

(See Prob. 10.1-1). Several common line-source currents and their associated pattern Fourier transform pairs are shown in Table 5-3.

It is important to understand the requirements on a current distribution in order to achieve a pattern shape. This is useful in synthesis and in finding explanations for pattern

---

[1] Frequently, the z-axisis selected to be normal to the line source, in which case $\cos\theta$ in (10-3) becomes $\sin\theta$.

<div align="center">**Table 10-2** Symmetry Properties of Current Distributions and Patterns</div>

Definitions:

$i(s) = i_r(s) + ji_i(s)$	Real and imaginary
$i(s) = |i(s)|e^{j\phi(s)} = A(s)e^{j\phi(s)}$	Amplitude and phase
$i(s) = i_e(s) + i_o(s)$	Even and odd
$\qquad i_e(-s) = i_e(s), \; i_o(-s) = -i_o(s)$	
$|i(-s)| = |i(s)|$	Symmetric

Properties:

Pattern	Required Current Distribution
1. Real pattern $\quad f(w) = f_r(w) + j0$	$i(-s) = i*(s):$ $\quad A(-s) = A(s) \quad$ Sym. amp. $\quad \phi(-s) = -\phi(s) \quad$ Odd phase
2. Symmetric pattern $\quad |f(-w)| = |f(w)|$	$A(s)$ symmetric and $\phi(s)$ even; that is, $i(s)$ even *or* $A(s)$ asymmetric and $\phi(s)$ constant
3. Asymmetric pattern $\quad |f(-w)| \neq |f(w)|$	$A(s)$ symmetric and $\phi(s)$ odd, nonzero *or* $A(s)$ a symmetric and $\phi(s)$ nonconstant

abnormalities. The (linear) Fourier transform relationship between the current and pattern developed in Sec. 5.3 can be used to infer the general properties shown in Table 10-2. These principles also apply to arrays. The current distribution and the pattern functions can be described mathematically in terms of either real and imaginary, or amplitude and phase, or even and odd parts, as shown in the definitions section of Table 10-2. A current or pattern function is said to be symmetric if its magnitude is mirror-imaged about the origin.

The pattern properties and the associated required current distribution in Table 10-2 are explained next.

1. *Synthesis Property 1: A real-valued pattern is achieved if and only if the current distribution amplitude is symmetric and the phase is odd.*

    Real patterns are often used in synthesis for mathematical simplicity, but in general a pattern can be complex-valued. Although allowing pattern phase to vary adds another synthesis variable, it turns out to add little to synthesis capability. Thus, we only consider real-valued patterns.

2. *Synthesis Property 2: A real-valued current distribution produces a symmetric pattern.*

    A symmetric pattern is obtained if either of the two conditions shown in Property 2 of Table 10-2 is satisfied. The first condition states that a symmetric current amplitude distribution and an even phase distribution produce a symmetric pattern. The second condition states that an asymmetric current amplitude distribution and a constant phase distribution produce a symmetric pattern. Both of these conditions indicate that a zero current phase yields a symmetric pattern. A current phase distribution that is zero, of course, means a real-valued current.

3. *Synthesis Property 3: An asymmetric pattern can be realized only through the use of current phase control.*

    Both conditions shown in Table 10-2 for achieving an asymmetric pattern require control of the current phase. For example, a linear phase taper in the current, which is an odd phase function, will steer the beam off axis ($w \neq 0$) and produces a pattern that is not symmetric about $w = 0$, although it can be symmetric about the beam peak.

These properties provide guidelines for what is required of the current distribution to achieve the specified pattern shape. A complicated current distribution may be difficult to realize with a continuous aperture antenna. The designer often looks to an array in such cases.

10.2   LINE SOURCE SHAPED BEAM SYNTHESIS METHODS

Shaped beam synthesis is used when the main beam of an antenna pattern is required to conform to a specified shape and cannot be of nominal shape. This section focuses on one-dimensional continuous current distributions (i.e., line sources), but the results apply to two-dimensional current distributions that are separable (i.e., factorable into a product of two functions, each depending only on the associated source coordinate in the corresponding plane).

10.2.1   The Fourier Transform Method

The Fourier transform pair relationship for the pattern and current of (10-6) and (10-7) suggests a synthesis method. If $f_d(w)$ is the *desired pattern,* the corresponding current distribution $i_d(s)$ is found rather easily from (10-7) as

$$i_d(s) = \int_{-\infty}^{\infty} f_d(w) e^{-j2\pi sw}\, dw \qquad (10\text{-}8)$$

This is very direct, but unfortunately, the resulting $i_d(s)$ will *not,* in general, be confined to $|s| \leq L/2\lambda$ as required; it will usually be, in fact, of infinite extent. An approximate solution can be obtained by truncating $i_d(s)$, giving the synthesized current distribution as follows:

$$i(s) = \begin{cases} i_d(s) & |s| \leq \dfrac{L}{2\lambda} \\[3mm] 0 & |s| > \dfrac{L}{2\lambda} \end{cases} \qquad (10\text{-}9)$$

The current $i(s)$ produces an approximate pattern $f(w)$ from (10-6). The current $i_d(s)$ extending over all s produces the pattern $f_d(w)$ exactly.

The Fourier transform synthesized pattern yields the least mean-square error (MSE), or least mean-squared deviation from the desired pattern, over the entire w-axis. The mean-squared error

$$\text{MSE} = \int_{-\infty}^{\infty} |f(w) - f_d(w)|^2\, dw \qquad (10\text{-}10)$$

with $f(w)$ corresponding to $i(s)$ in (10-9), is the smallest of all patterns arising from line sources of length L. The Fourier transform synthesized pattern, however, does not provide minimum mean-squared deviation in the visible region.

EXAMPLE 10-1   *Fourier Transform Synthesis of a Sector Pattern*

A *sector pattern* is a shaped beam pattern that, ideally, has uniform radiation over the main beam (a sector of space) and zero side lobes. Such patterns are popular for search applications where vehicles are located by establishing communications or by a radar echo in the sector of space occupied by the antenna pattern main beam. As a specific example, let the desired pattern be

$$f_d(\theta) = \begin{cases} 1 & \cos^{-1}c \le \theta \le \cos^{-1}(-c) \\ 0 & \text{elsewhere} \end{cases} \qquad (10\text{-}11a)$$

or, equivalently,

$$f_d(w) = \begin{cases} 1 & |w| \le c \\ 0 & \text{elsewhere} \end{cases} \qquad (10\text{-}11b)$$

$f_d(w)$ is shown in Fig. 10-1a by the dashed curve. Using (10-11b) in (10-8) and (10-9) gives

$$i(s) = 2c\frac{\sin(2\pi cs)}{2\pi cs}, \quad |s| \le \frac{L}{2\lambda} \qquad (10\text{-}12)$$

If this $\sin(x)/x$ function were not truncated, its Fourier transform (its pattern) would be exactly the sector pattern of (10-11). The actual pattern from (10-6) using (10-12) is

$$f(w) = \frac{1}{\pi}\left\{ \text{Si}\left[\frac{L}{\lambda}\pi(w+c)\right] - \text{Si}\left[\frac{L}{\lambda}\pi(w-c)\right] \right\} \qquad (10\text{-}13)$$

where Si is the sine integral of (F-13). Alternate means of evaluating $f(w)$ include direct numerical integration or numerical Fourier transform. This synthesized sector pattern is plotted in Fig. 10-1a for $c = 0.5$ and $L = 10\lambda$. The pattern is plotted in linear form, rather than in decibels, to emphasize the details of the main beam. Note the oscillations about the desired pattern on the main beam, called ripple, and the nonzero side lobes. This appearance of main beam ripple and overshoots in the synthesized pattern occur because the desired pattern has a discontinuity. This Gibbs phenomenon is a familiar result from Fourier theory and signal theory. The current distribution of (10-12) is plotted in Fig. 10-1b, and it is very similar to the current in Table 5-3b, as expected.

10.2.2 The Woodward–Lawson Sampling Method

A particularly convenient way to synthesize a radiation pattern is to specify values of the pattern at various points, that is, to sample the pattern. The Woodward–Lawson method's is the most popular of the sampling methods [1, 2]. It is based on decomposition of the source current distribution into a sum of uniform amplitude, linear phase sources:

$$i_n(s) = \frac{a_n}{L/\lambda}e^{-j2\pi w_n s}, \qquad |s| \le \frac{L}{2\lambda} \qquad (10\text{-}14)$$

(a) The synthesized pattern (solid curve) and the desired sector pattern (dashed curve).

(b) The current distribution.
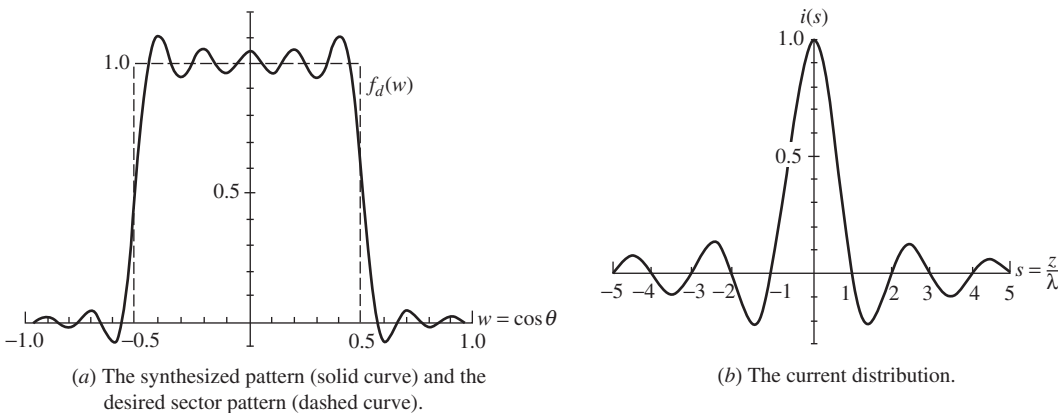
Figure 10-1 Fourier transform synthesis of a sector pattern using a 10λ line source (Example 10-1).

The pattern corresponding to this component current, from (10-6), is

$$f_n(w) = a_n \ \text{Sa}\left[\pi \frac{L}{\lambda}(w - w_n)\right] \tag{10-15}$$

where the sampling function $\text{Sa}(x)$ is defiend as $\text{Sa}(x) = \sin(x)/x$. This component pattern has a maximum of a_n centered at $w = w_n$. The current component phase coefficient w_n in (10-14) controls the location of the component pattern maximum, and the current component amplitude coefficient a_n controls the component pattern amplitude.

In the Woodward–Lawson method, the total current excitation is composed of a sum of $2M + 1$ component currents as

$$i(s) = \sum_{n=-M}^{M} i_n(x) = \frac{1}{L/\lambda} \sum_{n=-M}^{M} a_n e^{-j2\pi w_n s} \tag{10-16a}$$

where

$$w_n = \frac{n}{L/\lambda}, \quad |n| \le M, \quad |w_n| \le 1.0 \tag{10-16b}$$

The pattern corresponding to this current is

$$f(w) = \sum_{n=-M}^{M} f_n(w) = \sum_{n=-M}^{M} a_n \ \text{Sa}\left[\pi \frac{L}{\lambda}(w - w_n)\right]$$

$$= \sum_{n=-M}^{M} a_n \ \text{Sa}\left[\pi \left(\frac{L}{\lambda} w - n\right)\right] \tag{10-17}$$

At pattern points $w = w_n = n\lambda/L$, this reduces to $f(w = w_n) = a_n$. Thus, the pattern has specified values a_n, called pattern *sample values,* at the pattern locations w_n of (10-16b), called *sample points*. The pattern sample values are chosen to equal the values of the desired pattern at the sample points:

$$a_n = f_d(w = w_n) \tag{10-18}$$

The current distribution required to produce a pattern with values a_n at locations w_n is that of (10-16).

The Woodward–Lawson sampling method can be made more flexible by noting that as long as adjacent samples are separated by the sampling interval $\Delta w = \lambda/L$, the pattern values at the sample points are still uncorrelated, that is, (10-18) holds. The total number of samples is chosen such that the visible region is just covered; samples located outside the visible region could lead to superdirective results. Since the visible region is of extent 2 and $\Delta w = \lambda/L$, the number of samples $2M + 1$ is on the order of $2/(\lambda/L)$, or M is on the order of L/λ. The Woodward–Lawson synthesis procedure is very easy to visualize, as illustrated in the following example.

EXAMPLE 10-2   *Woodward–Lawson Line Source Synthesis of a Sector Pattern*

The sector pattern of Example 10-1 is now to be synthesized with a 10-wavelength-long line source using the Woodward–Lawson method. Sampling this pattern according to $a_n = f_d(w = w_n)$ with sample locations $w_n = n\lambda/L = 0.1n$ gives the values in Table 10-3. The sample value at the discontinuity ($w = 0.5$) could be selected as 1, 0.5, or 0 according to the specific application. Using $a_{\pm 5} = 1$ gives the widest main beam, whereas $a_{\pm 5} = 0$ gives the narrowest. In this case, we choose $a_{\pm 5} = 0.5$ as a compromise. The synthesized pattern is

Table 10-3 Sample Locations and Sample Values for a
10λ Woodward–Lawson Sector Pattern (Example 10-2)

n	Sample Location w_n	Pattern Sample Value a_n
0	0	1
±1	±0.1	1
±2	±0.2	1
±3	±0.3	1
±4	±0.4	1
±5	±0.5	0.5
±6	±0.6	0
±7	±0.7	0
±8	±0.8	0
±9	±0.9	0
±10	±1.0	0

computed using the sample values and locations of Table 10-3 in (10-17) and is plotted in Fig. 10-2*a*. The sample points are indicated by dots.

To illustrate the sampling nature of the Woodward–Lawson method, two sampling functions from the sum in (10-17) are shown in Fig. 10-2*b* for sample locations $w_{-1} = -0.1$ and $w_0 = 0$. Note that when one sampling function is maximum, the other is zero, thus making the samples independent. Further, each sampling function is zero at all sample locations $w_n = n\lambda/L$, except at its maximum. When all samples are included, the value of the total synthesized pattern at locations w_n is completely determined by the Sa function centered at that location. This is the beauty of the Woodward–Lawson sampling method.

Note that the Woodward–Lawson pattern of Fig. 10-2*a* is a better approximation to the desired pattern (in the visible region) than that of the Fourier transform method in Fig. 10-1*a*, both generated from a 10-wavelength line source. Detailed comparisons of all the sector pattern examples are presented in sec. 10.3.3.

The current distribution corresponding to the sector pattern of this example is plotted in Fig. 10-2*c*. It was obtained from (10-16). Note the similarity to the current distribution in Fig. 10-1*b* for the Fourier transform method. This occurs because the Fourier transform of any pattern is the antenna current distribution. Since the patterns in Examples 10-1 and 10-2 are both close to a sector pattern, their Fourier transforms (currents) must be close to that of an ideal sector pattern, which is $\sin(\pi s)/\pi s$ in these examples.

10.3   LINEAR ARRAY SHAPED BEAM SYNTHESIS METHODS

In this section we consider linear array synthesis methods. But, just as for line sources, shaped beam synthesis for linear arrays can be applied to two-dimensional arrays with distributions that are separable. The applicable equation expressing this result for planar arrays is (8-109), which gives the two-dimensional array pattern as a product of two one-dimensional array patterns. Fourier series and Woodward–Lawson methods are presented in this section. They are analogous to the Fourier transform and Woodward–Lawson methods of the previous section. Before presenting these methods, we model the array general configuration to be used with any linear array synthesis method.

Consider an equally spaced linear array along the *z*-axis with interelement spacing *d*. For simplicity, the physical center of the array is located at the origin. The total number of elements in the array P can be either even (then let $P = 2N$) or odd (then let $P = 2N + 1$). For an odd element number, the element locations are given by

(a) The synthesized pattern (solid curve) and the desired
pattern (dashed curve). The dots indicate
the sample values and locations.

(b) Two component patterns at sample
locations $w_{-1} = -0.1$ and $w_0 = 0$.

(c) The current distribution corresponding to the
synthesized pattern.
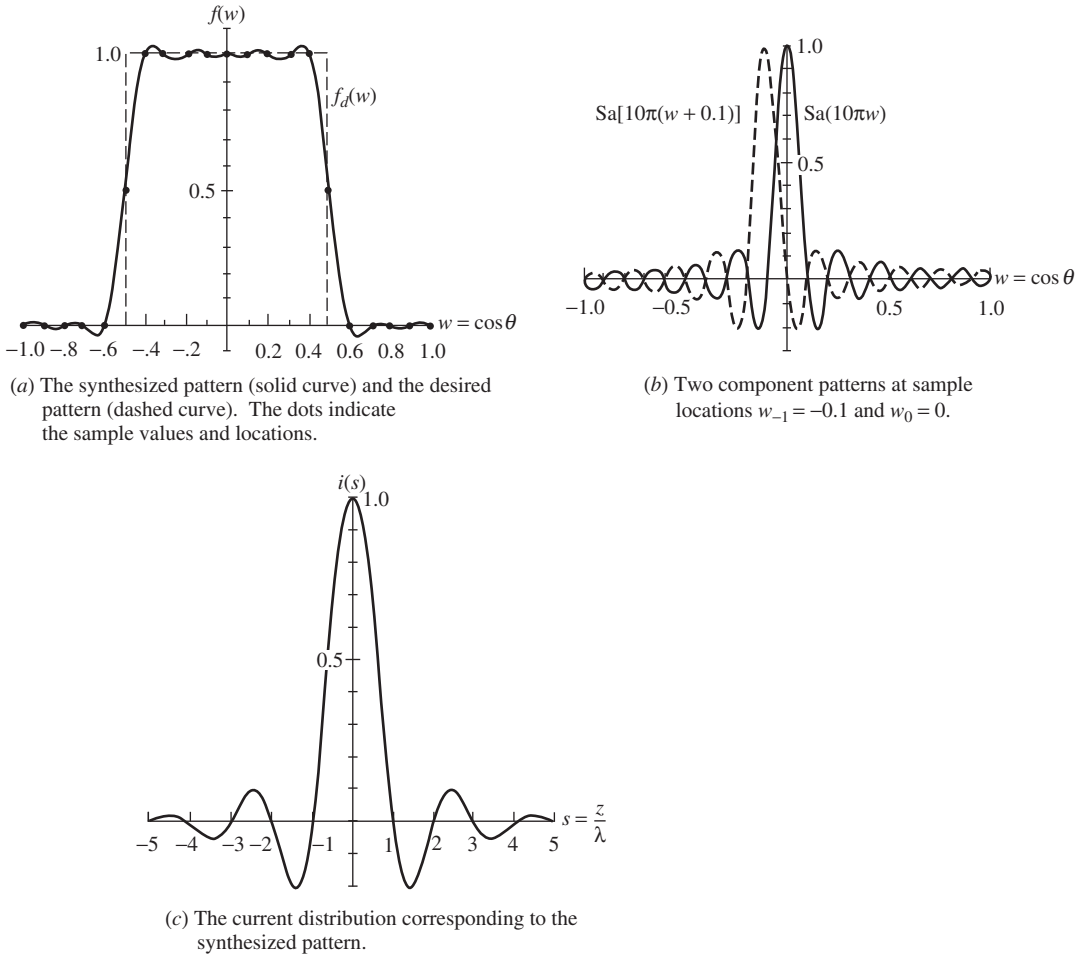
Figure 10-2 Woodward–Lawson synthesis of a sector pattern using a 10λ line source (Example 10-2).

$$z_m = md, \qquad |m| \leq N \qquad (10\text{-}19)$$

and $P = 2N + 1$. The corresponding array factor is

$$f(w) = \sum_{m=-N}^{N} i_m e^{j2\pi m(d/\lambda)w} \qquad (10\text{-}20)$$

where i_m are the element currents and again $w = \cos\theta$. This expression is similar to
(8-42).

For an, even number of elements, the element positions are

$$z_m = \frac{2m-1}{2}d, \qquad 1 \leq m \leq N$$

$$z_{-m} = -\frac{2m-1}{2}d, \qquad -N \leq -m \leq -1$$

$$(10\text{-}21)$$

and $P = 2N$. The corresponding array factor is

$$f(w) = \sum_{m=1}^{N} \left(i_{-m} e^{-j\pi(2m-1)(d/\lambda)w} + i_m e^{j\pi(2m-1)(d/\lambda)w} \right) \tag{10-22}$$

for P even.

For comparison to a line source, the total array length is defined as

$$L = Pd \tag{10-23}$$

This definition applies to both the even and odd element cases, and it includes a distance $d/2$ beyond each end element. Similar to (10-2) for continuous sources, the desired array factor is found from (8-54) by dividing the desired complete pattern, $F_d(w)$, by the element pattern, $g_a(w)$: $f_d(w) = F_d(w)/g_a(w)$.

10.3.1   The Fourier Series Method

The array factor resulting from an array of identical discrete radiators (elements) is, of course, the sum over the currents for each element weighted by the spatial phase delay from each element to the far-field point. This array factor summation can be made to be of a form that is very similar to a Fourier series, just as the radiation integral for a continuous source resembles a Fourier transform (see Sec. 10.2.1). To see how this correspondence comes about, we first observe that a function $f_d(w)$, the desired pattern function, can be expanded into a Fourier series in the interval $-\lambda/2d < w < \lambda/2d$ as

$$f_d(w) = \sum_{m=-\infty}^{\infty} b_m e^{j2\pi m(d/\lambda)w} \tag{10-24}$$

where

$$b_m = \frac{d}{\lambda} \int_{-\lambda/2d}^{\lambda/2d} f_d(w) e^{-j2\pi m(d/\lambda)w} \, dw \tag{10-25}$$

If we identify d as the spacing between elements of an equally spaced linear array and $w = \cos\theta$ where θ is the angle from the line of the array, the sum in (10-24) is recognized as the array factor of an array with an infinite number of elements with currents b_m.

An infinite array is, of course, not practical, but truncating the series (10-24) to a finite number of terms produces the following approximation to $f_d(w)$:

$$f(w) = \sum_{m=-N}^{N} b_m e^{j2\pi m(d/\lambda)w} \tag{10-26}$$

If we let the currents of each element in the array equal the Fourier series coefficients, that is,

$$i_m = b_m, \quad |m| \le N \tag{10-27}$$

then (10-26) is identical to (10-20), the array factor for an array with an odd number of elements.

The Fourier series synthesis procedure is, then, to use element excitations i_m equal to the Fourier series coefficients b_m calculated from the desired pattern f_d, as in (10-25). The array factor f arising from these element currents is an approximation to the desired pattern. This Fourier series synthesized pattern provides the least mean-squared error [see (10-10)] over the region $-\lambda/2d < w < \lambda/2d$. If the elements are half-wavelength

spaced $(d = \lambda/2)$, this region is exactly the visible region $(-1 < w < 1$, or $0 < \theta < \pi)$.

A similar line of reasoning leads to the results for an even number of elements. In this case, the Fourier series coefficient currents are

$$i_m = b_m = \frac{d}{\lambda} \int_{-\lambda/2d}^{\lambda/2d} f_d(w)e^{-j\pi(2m-1)(d/\lambda)w}\,dw, \qquad m \geq 1$$

$$(10\text{-}28)$$

$$i_{-m} = b_{-m} = \frac{d}{\lambda} \int_{-\lambda/2d}^{\lambda/2d} f_d(w)e^{j\pi(2m-1)(d/\lambda)w}\,dw, \qquad -m \leq -1$$

for P even. The synthesized pattern is given by (10-22). Note that if N is infinite, (10-22) together with (10-28) is the Fourier series expansion of f_d; that is, $f(w) = f_d(w)$. Generally the element spacing is equal to or greater than a half wavelength, so $\lambda/2d < 1$.

EXAMPLE 10-3   *Fourier Series Synthesis of a Sector Pattern*

For an equally spaced linear array with an even number of elements, the sector pattern of (10-11) in (10-28) yields excitation currents

$$i_m = i_{-m} = 2\frac{d}{\lambda}c \ \text{Sa}\left[\pi(2m-1)\frac{d}{\lambda}c\right], \qquad 1 \leq m \leq N \qquad (10\text{-}29)$$

The spacing should satisfy $d < \lambda/(1+c)$ to avoid secondary main beams' folding back into the visible region. Since these currents are symmetric, the array factor of (10-22) reduces to

$$f(w) = 2\sum_{m=1}^{N} i_m \ \cos\left[\pi(2m-1)\frac{d}{\lambda}w\right] \qquad (10\text{-}30)$$

which is a real function. Note this is a special case of symmetry Property 1 in Table 10-2.

The specific case of $c = 0.5$, $d/\lambda = 0.5$, and 20 elements $(N = 10)$ has an array length $L = Pd = 10\lambda$ and excitation currents from (10-29) given by

$$i_m = i_{-m} = \frac{1}{2} \ \text{Sa}\left[\frac{\pi}{4}(2m-1)\right], \quad 1 \leq m \leq 10 \qquad (10\text{-}31)$$

These excitation values are listed in Table 10-4, together with the element positions from (10-21). When these are used in the pattern expression (10-30), the pattern shown in Fig. 10-3 is produced.

10.3.2   The Woodward–Lawson Sampling Method

The Woodward–Lawson sampling method for linear arrays is analogous to the Woodward–Lawson sampling method for line sources (see Sec. 10.2.2). In the array case, the synthesized array factor is the superposition of array factors from uniform amplitude, linear phase arrays:

$$f(w) = \sum_{n=-M}^{M} a_n \frac{\sin[(P/2)(w - w_n)(2\pi/\lambda)\,d]}{P \ \sin\left[\frac{1}{2}(w - w_n)(2\pi/\lambda)\,d\right]} \qquad (10\text{-}32)$$

Table 10-4 Array Positions and Currents for a Fourier Series Synthesized Linear Array of 20 Half-Wavelength Spaced Elements for a Sector Pattern (Example 10-3)

Element Number m	Element Position z_m	Excitation Current i_m
± 1	$\pm 0.25\lambda$	0.4502
± 2	$\pm 0.75\lambda$	0.1501
± 3	$\pm 1.25\lambda$	-0.0900
± 4	$\pm 1.75\lambda$	-0.0643
± 5	$\pm 2.25\lambda$	0.0500
± 6	$\pm 2.75\lambda$	0.0409
± 7	$\pm 3.25\lambda$	-0.0346
± 8	$\pm 3.75\lambda$	-0.0300
± 9	$\pm 4.25\lambda$	0.0265
± 10	$\pm 4.75\lambda$	0.0237
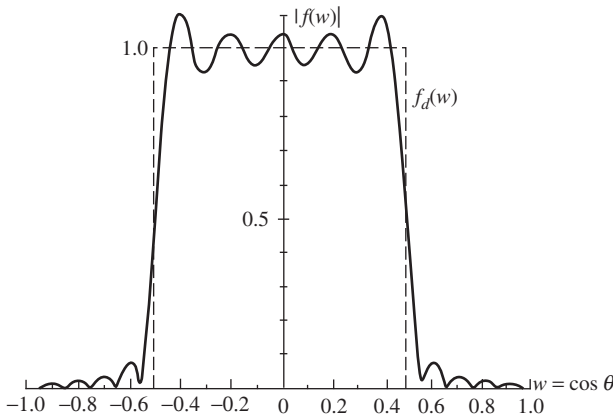
Figure 10-3 Fourier series synthesized array factor for a 20-element, $\lambda/2$ spaced linear array (Example 10-3). The desired pattern (dashed curve) is a sector pattern.

where the sample values are

$$a_n = f_d(w = w_n) \tag{10-33}$$

and the sample points are

$$w_n = n\frac{\lambda}{Pd} = \frac{n}{L/\lambda}, \quad |n| \le M, \quad |w_n| \le 1.0 \tag{10-34}$$

The element currents required to give this pattern are found from

$$i_m = \frac{1}{P} \sum_{\pi=-M}^{M} a_n e^{-j2\pi(z_m/\lambda)w_n} \tag{10-35}$$

These results hold for arrays with either an even or odd number of elements.

EXAMPLE 10-4 **Woodward–Lawson Array Synthesis of a Sector Pattern**

Again, the sector pattern of (10-11) with $c = 0.5$ is to be synthesized, this time with a 20-element, half-wavelength spaced linear array using the Woodward–Lawson method. The sample locations from (10-34) are $w_n = 0.1n$. Thus, the sample locations and values are the same as for Example 10-2 and are given in Table 10-3. Using these and element

Table 10-5 Array Element Currents and Positions
Synthesized from the Woodward–Lawson Method
for a Sector Pattern (Example 10-4)

Element Number m	Element Position z_m	Excitation Current i_m
± 1	$\pm 0.25\lambda$	0.44923
± 2	$\pm 0.75\lambda$	0.14727
± 3	$\pm 1.25\lambda$	-0.08536
± 4	$\pm 1.75\lambda$	-0.05770
± 5	$\pm 2.25\lambda$	0.04140
± 6	$\pm 2.75\lambda$	0.03020
± 7	$\pm 3.25\lambda$	-0.02167
± 8	$\pm 3.75\lambda$	-0.01464
± 9	$\pm 4.25\lambda$	0.00849
± 10	$\pm 4.75\lambda$	0.00278
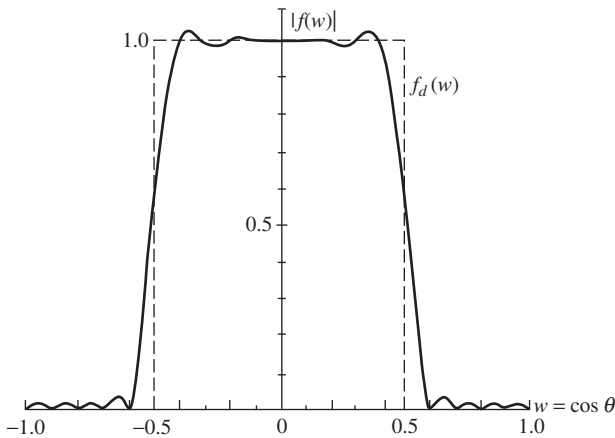
Figure 10-4 Woodward–Lawson synthesized array factor for a 20-element, $\lambda/2$ spaced linear array (Example 10-4). The desired pattern (dashed curve) is a sector pattern.

positions z_m from (10-21) in (10-35) yields the array currents of Table 10-5. The pattern can be generated from either the Woodward–Lawson pattern expression of (10-32), or by direct array computation using (10-30), which is the version of (10-22) for the symmetric case, and the array parameters of Table 10-5. The pattern is plotted in Fig. 10-4.

The Woodward–Lawson array can be realized in practice using a Butler matrix beam forming network that produces independent beams. The Butler matrix was discussed in Sec. 8.9.2, where it was mentioned that the component beam crossover value is about 4 dB below the peak.

A topic that relates to array synthesis is that of *synthetic aperture radar,* which uses platform motion to create an effective aperture that is much larger than the real aperture. An example is a down-looking antenna on an airplane or spacecraft moving at a constant velocity in a straight line that collects signal data returned from the transmitted radar pulses. An image of the Earth's surface is formed by combining the data using the real aperture as an element in a synthesized aperture along the motion path. Applications are also found in radio astronomy.

10.3.3   Comparison of Shaped Beam Synthesis Methods

Most shaped beam antenna patterns have three distinct types of pattern regions: side lobe, main beam, and transition. The side lobe region is easily recognized, and the *side lobe level,* SLL, is defined from

Table 10-6 Comparison of Synthesized Sector Patterns ($c = 0.5$, $L = 10\lambda$)

Method	Type	Example Number	Figure Number	Side Lobe Level, SLL (dB)	Ripple R (dB)	Transition Width, T
Fourier transform	10λ line source	10-1	10-1a	−21.9	0.83	0.0893
Woodward-Lawson	10λ line source	10-2	10-2a	−29.8	0.27	0.1303
Fourier series	20-element $\lambda/2$ spaced array	10-3	10-3	−22.6	0.87	0.0891
Woodward-Lawson	20-element $\lambda/2$ spaced array	10-4	10-4	−29.6	0.27	0.1343

$$\text{SLL} = 20 \log \left| \frac{\text{value of the highest side lobe peak}}{\text{maximum of desired pattern}} \right| \qquad (10\text{-}36)$$

over the side lobe region. The quality of fit to the desired pattern $f_d(w)$ by the synthesized pattern $f(w)$ over the main beam is measured by the *ripple R*, which is defined as

$$R = 20 \left\{ \text{maximum} \left| \log \left| \frac{f(w)}{f_d(w)} \right| \right| \right\} \qquad [\text{dB}] \qquad (10\text{-}37)$$

over the main beam. Also of interest is the region between the main beam and side lobe region, referred to as the transition region. In many applications, such as direction finding, it is desirable to have the main beam fall off very sharply into the side lobe region. To quantify this, *transition width T* is introduced and defined as

$$T = |w_{f=0.9} - w_{f=0.1}| \qquad (10\text{-}38)$$

where $w_{f=0.9}$ and $w_{f=0.1}$ are the values of w where the synthesized pattern f equals 90 and 10% of the local discontinuity in the desired pattern. For unsymmetrical, single beam patterns, there are two transition regions with different transition widths. Transition width is analogous to rise time in time-signal analysis.

The shaped beam synthesis methods we have discussed in this and the previous section can be compared rather easily using SLL, R, and T. The sector pattern results of Examples 10-1 to 10-4 are presented in Table 10-6. A few general trends can be extracted from the table. The Woodward–Lawson methods (for both line sources and arrays) tend to produce low side lobes and low main beam ripple at some sacrifice in transition width. On the other hand, Fourier methods yield somewhat inferior side lobe levels and ripples. The Fourier series synthesized pattern gives very sharp rolloff from the main beam to the side lobe region; that is, small transition width.

10.4 LOW SIDE LOBE, NARROW MAIN BEAM SYNTHESIS METHODS

Many applications call for a pattern with a single narrow beam and low side lobes to reject unwanted signals in directions outside the main beam. Point-to-point communications and imaging are example uses for narrow beam, low side lobe patterns. Extreme side lobe reduction is used to cope with high-power jamming transmitters. Dynamic (adaptive) pattern control is required to counter interference or jamming signals that change arrival direction. In this section, we present the two most popular methods to synthesize narrow beam, low side lobe patterns: the Dolph-Chebyshev method for linear arrays and the Taylor line source method. The methods are closely related, and the Dolph-Chebyshev method is presented first to simplify the mathematical development.

10.4.1   The Dolph-Chebyshev Linear Array Method

A uniform line source has a first side lobe level of -13.3 dB, and a uniform linear array has slightly higher side lobes and depends on the number of elements. As we learned with line sources in Sec. 5.2 and arrays in Sec. 8.6, the current amplitude must taper from the center to the edge of a source for low side lobes. The side lobes lower as the element current amplitude taper from the center to the edge of an array is increased, as illustrated in Figs. 8-20 and 8-21. The tradeoff to achieve low side lobes is main beam widening and a reduction in directivity. This tradeoff is optimum (highest directivity for a given side lobe level) when all the side lobes are of equal level. This optimum case will give the narrowest possible beamwidth for a specified side lobe level, or vice versa. In this section, we present a method for achieving the optimum broadside linear array with equal element spacings that are equal to or greater than a half-wavelength.

Chebyshev[2] polynomials have the desired property of equal ripples. They are commonly used with equal-ripple filters. Optimum side lobe performance occurs when there are as many side lobes as possible in the visible region that are of equal level. Dolph [3] applied Chebyshev polynomials in this manner to arrays using what is now called the *Dolph-Cheyshev method*. We begin by examining the properties of Chebyshev polynomials, which are defined as:

$$T_n(x) = \begin{cases} (-1)^n \cosh(n \cosh^{-1}|x|), & x < -1 \\ \cos(n \cos^{-1} x), & -1 < x < 1 \\ \cosh(n \cosh^{-1} x), & x > 1 \end{cases} \qquad (10\text{-}39)$$

A few of the lower-order polynomials are

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_2(x) &= 2x^2 - 1 \\ T_3(x) &= 4x^3 - 3x \\ T_4(x) &= 8x^4 - 8x^2 + 1 \end{aligned} \qquad (10\text{-}40)$$

Higher-order polynomials can be generated from the recursive formula

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \qquad (10\text{-}41)$$

or by letting $\delta = \cos^{-1} x$ and expanding $\cos m\delta$ in powers of $\cos \delta$. For example, $T_3(x) = \cos(3 \cos^{-1} x) = \cos 3\delta = 4 \cos^3 \delta - 3 \cos \delta$ from (D-13). Hence, $T_3(x) = 4x^3 - 3x$. A few polynomials are plotted in Fig. 10-5.

Some important general properties of Chebyshev polynomials follow from (10-39) or Fig. 10-5. The even-ordered polynomials are even, that is, $T_n(-x) = T_n(x)$ for n even, and the odd-ordered ones are odd, that is, $T_n(-x) = -T_n(x)$ for n odd. All polynomials pass through the point (1, 1). In the range $-1 \leq x \leq 1$, the polynomial values lie between -1 and 1, and the maximum magnitude is always unity there. All zeros (roots) of the polynomials also lie in $-1 \leq x \leq 1$.

The equal amplitude oscillations of Chebyshev polynomials in the region $|x| \leq 1$ is the desired property for equal side lobes. Also, the polynomial nature of the functions makes them suitable for array factors since an array factor can be written as a polynomial. The

---

[2] The polynomials were introduced by Russian mathematical Pafnuty Chebyshev (1821–1894). A common alternate spelling is Tchebyscheff.
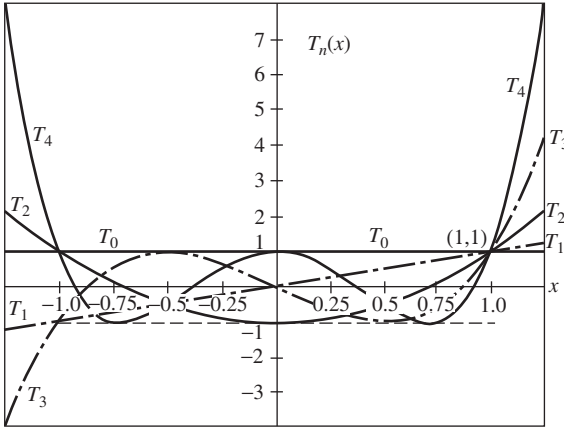
Figure 10-5 Chebyshev polynomials $T_0(x)$, $T_1(x)$, $T_2(x)$, $T_3(x)$, and $T_4(x)$.

connection between arrays and Chebyshev polynomials is established by considering a symmetrically excited, broadside array for which (see Table 10-2).

$$i_{-m} = i_m \tag{10-42}$$

where the mth element is located at $z_m = md$ and current i_0 is at the origin ($z = 0$) for an odd number of elements. Symmetrical excitation leads to a real-valued array factor that, from (10-20) and (10-22), is given by

$$f(\psi) = \begin{cases} i_0 + 2 \sum_{m=1}^{N} i_m \cos m\psi & P \text{ odd} \\ 2 \sum_{m=1}^{N} i_m \cos\left[(2m-1)\dfrac{\psi}{2}\right] & P \text{ even} \end{cases} \tag{10-43}$$

where $\psi = 2\pi(d/\lambda)w$. This array factor (for odd or even P) is a sum of $\cos(m\psi/2)$ terms for m up to $P-1$. But each term containing $\cos(m\psi/2)$ can be written as a sum of terms with powers of $\cos(\psi/2)$ up to m, through the use of trigonometric identities. Therefore, the array factor is expressible as a sum of terms with powers of $\cos(\psi/2)$ up to $P-1$.

By choosing an appropriate transformation between x and ψ, the array factor and Chebyshev polynomial will be identical. The transformation

$$x = x_o \cos\frac{\psi}{2} \tag{10-44}$$

and the correspondence

$$f(\psi) = T_{P-1}\left(x_o \cos\frac{\psi}{2}\right) \tag{10-45}$$

will yield a polynomial in powers of $\cos(\psi/2)$ matching that of the array factor. The main beam maximum value of R occurs for $\theta = 90°$, or $\psi = 0$, for a broadside array.[3] Then

---

[3] The symbol f is usually reserved for a pattern normalized to a maximum value of unity, but for the Dolph-Chebyshev array, it is more convenient to normalize the array factor f to a maximum value of R.
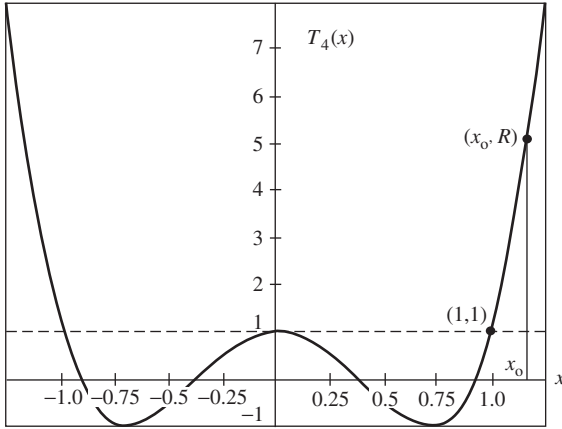
Figure 10-6 Chebyshev polynomial $T_4(x)$.

(10-44) indicates that $x = x_o$ at the main beam maximum. The visible region extends from $\theta = 0°$ to $180°$, or $\psi = 2\pi(d/\lambda)$ to $-2\pi(d/\lambda)$. These limits correspond to $x = x_o \cos(\pi d/\lambda)$; for half-wavelength spacing, the limits are $x = 0$. Thus, for $d = \lambda/2$, the visible region begins at $x = 0$, or $\theta = 0°$, and x increases θ does until x_o (the main beam maximum point) is reached and retraces back to $x = 0$, or $\theta = 180°$ (see Fig. 10-6).

The main beam-to-side lobe ratio R is the value of the array factor at the main beam maximum, since the side lobe lvel magnitude is unity (see Fig. 10-6). The side lobe level is thus $1/R$, or

$$\text{SLL} = -20 \log R \quad [\text{dB}] \tag{10-46}$$

Evaluating (10-45) at the main beam maximum gives

$$f(\psi = 0) = R = T_{p-1}(x_o) = \cosh[(P-1)\cosh^{-1} x_o] \tag{10-47}$$

from (10-39). Or, solving for x_o, we get

$$x_o = \cosh\left(\frac{\cosh^{-1}R}{P-1}\right) \tag{10-48}$$

The design procedure can now be summarized. For a given side lobe ratio, R can be determined from (10-46), leading to x_o from (10-48). The array factor is then given by (10-45), or it can be computed from (10-43) directly from the current values. The excitation currents are found by comparison between the array factor of (10-43) and the Chebyshev polynomial of (10-45). This synthesis procedure will be illustrated in Example 10-5. The currents are difficult to find by this method for arrays larger than a few elements. A direct computational method is available [4] for finding the currents and was used to calculate the current magnitude values in Table 10-7 for some common side lobe levels for Dolph-Chebyshev arrays of from 3 to 10 elements. For example, Table 10-7 gives the currents for an array with $-30\,\text{dB}$ side lobes as $1:2.41:3.14:2.41:1$, which agree with the values given in the caption of Fig. 8-20.

The Dolph-Chebyshev array design procedure provides the lowest side lobe pattern for a specified beamwidth. However, these solutions only depend on the number of elements, not their spacing (and thus total array length). Practical design seeks the narrowest beamwidth solution within the Dolph-Chebyshev family. This result is obtained by including as many side lobes as possible in the visible region without letting the grating lobe emerge to a level

Table 10-7 Current Magnitudes for Equally Spaced Dolph-Chebyshev Linear Arrays (current values are listed for the central element first and the edge elements have unity magnitude)

	SLL (dB)				
P	**−20**	**−25**	**−30**	**−35**	**−40**
3	1.6364	1.7870	1.8774	1.9301	1.9604
4	1.7357	2.0699	2.3309	2.5265	2.6688
5	1.9319	2.5478	3.1397	3.6785	4.1480
	1.6085	2.0318	2.4123	2.7401	3.0131
6	1.8499	2.5876	3.3828	4.1955	4.9891
	1.4369	1.8804	2.3129	2.7180	3.0853
7	1.8387	2.7267	3.7846	4.9811	6.2731
	1.6837	2.4374	3.3071	4.2625	5.2678
	1.2764	1.7081	2.1507	2.5880	3.0071
8	1.7244	2.6467	3.8136	5.2208	6.8448
	1.5091	2.2296	3.0965	4.0944	5.1982
	1.1386	1.5464	1.9783	2.4205	2.8605
9	1.6627	2.6434	3.9565	5.6368	7.6989
	1.5800	2.4751	3.6516	5.1308	6.9168
	1.3503	2.0193	2.8462	3.8279	4.9516
	1.0231	1.4036	1.8158	2.2483	2.6901
10	1.5585	2.5318	3.8830	5.6816	7.9837
	1.4360	2.2770	3.4095	4.8740	6.6982
	1.2125	1.8265	2.5986	3.5346	4.6319
	0.9264	1.2802	1.6695	2.0852	2.5182

above the design side lobe level. A general expression for the optimum spacing of the Dolph–Chebyshev array, with isotropic elements, that gives the narrowest beamwidth possible for a specified side lobe level and given number elements is [4]

$$d_{\text{opt}} = \lambda \left[1 - \frac{\cos^{-1}\frac{1}{\gamma}}{\pi} \right] \quad \text{broadside} \tag{10-49a}$$

$$d_{\text{opt}} = \frac{\lambda}{2} \left[1 - \frac{\cos^{-1}\frac{1}{\gamma}}{\pi} \right] \quad \text{endfire} \tag{10-49b}$$

where

$$\gamma = \cosh\left[\frac{1}{P-1} \ln(R + \sqrt{R^2 - 1})\right] \tag{10-49c}$$

An interesting result is that the directivity is identical for broadside and endfire operation for both half-wavelength and optimum spacings [5]. Although the beamwidth is broader at endfire than at broadside, the main beam is a fan beam at broadside, leading to the same solid angle (and, thus, the same directivity) as for the pencil-shaped endfire beam.

The half-power beamwidth of a Dolph-Chebyshev array, in general, is given by

$$\text{HP} = \pi - 2\cos^{-1}\frac{\psi}{\beta d} \quad \text{broadside} \tag{10-50a}$$

$$HP = 2\cos^{-1}\left(1 - \frac{\psi_h}{\beta d}\right) \quad endfire \tag{10-50b}$$

where

$$\psi_h = 2\cos^{-1}\left\{\frac{\cosh\left[\frac{1}{P-1}\cosh^{-1}\frac{R}{\sqrt{2}}\right]}{\cosh\left[\frac{1}{P-1}\cosh^{-1}R\right]}\right\} \tag{10-50c}$$

The beamwidth of a Dolph-Chebyshev array will be wider than a uniform line source of the same length $L = Pd$, which is $HP \approx 0.886\ \lambda/L$.

The directivity of Dolph-Chebyshev arrays can be calculated using the appropriate formula of (8-83) to (8-85), or the method in [5]. The directivity value is identical for broadside and endfire cases for spacing values that are multiples of a half-wavelength. [5] At broadside, the beamwidth is narrow, but the three-dimensional pattern is a fan beam. The main beam is broader at endfire, but is a pencil beam. The directivity of an endfire Dolph-Chebyshev array equals that of the broadside array with the same number of elements and side lobe level, but with twice the spacing [5]: $D_{\text{endfire}}(d) = D_{\text{broadside}}(2d)$. The following formula provides an approximation to directivity for spacings greater than a half-wavelength but not so large that grating lobes appear:

$$D \approx \frac{2\,R^2}{1 + R^2\,HP} \quad broadside \tag{10-51}$$

where HP is the half-power beamwidth in radians. It should be remembered that the foregoing results are for isotropic elements. The element pattern in most cases reduces the beamwidth and increases directivity.

Before presenting two examples, we point out that arbitrarily high directivity and narrow beamwidth can be obtained from an array of fixed length. This is a *superdirective array*, which is defined as an array with more directivity than with an array of uniform excitation. The concept and quantitative definition of superdirectivity was introduced with line sources in Sec. 5.5, and superdirective arrays were mentioned in Sec. 8.6. Superdirective arrays generally require the element spacing to be less than a half-wavelength. Modest superdirectivity is accomplished with the Hansen-Woodyard endfire array of Sec. 8.3.4. Extremely high directivity is achieved through widely varying amplitudes and phases of the elements. [7] However, directivity much higher than for uniform, excitation carries the disadvantages of narrow bandwidth and high sensitivity to the accuracy of the excitation currents, rendering the array impractical in most situations.

EXAMPLE 10-5   *A Five-Element, Broadside, −20-dB Side Lobe, Half-Wavelength Spaced Dolph-Chebyshev Array*

For a five-element array ($P = 5$, $N = 2$), the array factor from (10-43) is

$$f(\psi) = i_o + 2i_1\cos\psi + 2i_2\cos 2\psi \tag{10-52}$$

where $\psi = 2\pi(d/\lambda)\cos\theta = \pi\cos\theta$ for $d = \lambda/2$. Using $\cos(2\psi/2) = 2\cos^2(\psi/2) - 1$ from (D-12) and $\cos(4\psi/2) = 8\cos^4(\psi/2) - 8\cos^2(\psi/2) + 1$ from (D-14), the array factor can be written as

$$f(\psi) = (i_0 - 2i_1 + 2i_2) + (4i_1 - 16i_2)\ \cos^2\frac{\psi}{2} + 16i_2\cos^4\frac{\psi}{2} \qquad (10\text{-}53)$$

And from (10-40),

$$T_4(x) = 1 - 8x^2 + 8x^4 = 1 - 8x_o^2\cos^2\frac{\psi}{2} + 8x_o^4\cos^4\frac{\psi}{2} \qquad (10\text{-}54)$$

where (10-44) was used in the second step. Now, the currents are found by successively equating the coefficients of like terms of (10-53) and (10-54). From the $\cos^4(\psi/2)$ term,

$$i_2 = \frac{1}{2}x_o^4 \qquad (10\text{-}55)$$

The $\cos^2(\psi/2)$ term yields

$$i_1 = 4i_2 - 2x_o^2 = 2x_o^4 - 2x_o^2 \qquad (10\text{-}56)$$

using (10-55). The final term gives

$$i_0 = -2i_2 + 2i_1 + 1 = 3x_o^4 - 4x_o^2 + 1 \qquad (10\text{-}57)$$

using (10-55) and (10-56). The current values will be completely determined when x_o is evaluated. This is accomplished by first finding the main beam-to-side lobe ratio from (10-46) using the specified -20-dB side lobe level;

$$R = 10^{-\text{SLL}/20} = 10 \qquad (10\text{-}58)$$

Then from (10-48) with $P = 5$ and $\text{R} = 10$,

$$x_o = 1.293 \qquad (10\text{-}59)$$

The element currents from (10-55) to (10-57) with (10-59) are

$$i_2 = i_{-2} = 1.3975, \quad i_1 = i_{-1} = 2.2465, \quad i_0 = 2.6978 \qquad (10\text{-}60)$$

These currents yield a main beam maximum of $R = 10$ and unity side lobes. Normalizing these to unity edge currents gives a $1:1.61:1.93:1.61:1$ current distribution. These values agree with those in Table 10-7. The currents of (10-60) in (10-52) lead to the pattern in Fig. 10-7, which was normalized to 0 dB on the main beam maximum. The half-power beamwidth from (10-50a) is 23.7°, which was also found by direct evaluation of the polar
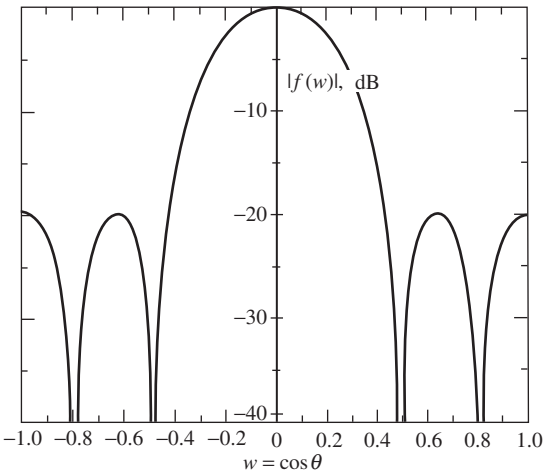
Figure 10-7 Dolph-Chebyshev synthesized array factor for a five-element, $\lambda/2$ spaced, broadside array with -20-dB side lobes (Example 10-5).
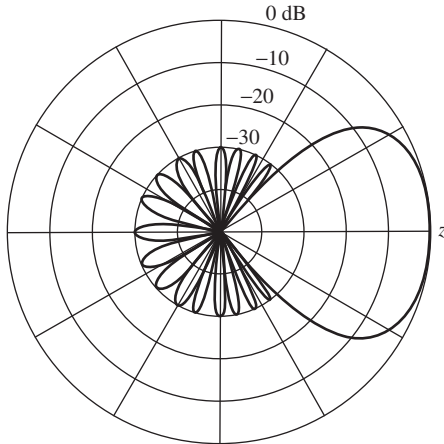
Figure 10-8 Polar pattern for the optimum endfire Dolph-Chebyshev, 10 element -30-dB side lobe array of Example 10-6.

pattern in Fig. 8-20d. The directivity for equi-phased, half-wavelength spaced arrays can be obtained from (8-85). For this example, the directivity is

$$D = \frac{\left| \sum\limits_{m=-2}^{2} i_m \right|^2}{\sum\limits_{m=-2}^{2} i_m^2} = 4.69 \tag{10-61}$$

The directivity found from (10-51) is 4.72.

EXAMPLE 10-6   *Optimum 10-Element, -30-dB Side Lobe Dolph-Chebyshev Endfire Array*

Using $P = 10$ and $R = 10^{-(-30)/20} = 31.62$ in (10-49a) yields

$$d_{opt} = 0.429\lambda \tag{10-62}$$

The element current amplitudes are [4]

$$1 : 1.67 : 2.60 : 3.41 : 3.88 : 3.88 : 3.41 : 2.60 : 1.67 : 1 \tag{10-63}$$

These with interelement phase shift $\alpha = -\beta d \ \cos(0°) = -154.5°$ yield the pattern in Fig. 10-8. The half-power beamwidth from (10-50) is $59.6°$. The directivity from (8-83) is 14.5.

10.4.2   The Taylor Line Source Method

Although the Dolph-Chebyshev array does yield the highest directivity and narrowest beamwidth, the constant side lobe envelope leads to a high reactive energy condition, especially for large arrays. This means high-Q (low bandwidth) operation. This situation can be avoided by first designing a line source with nearly constant side lobes and using the current values at element locations in an array configuration that produces a very similar pattern; see Prob. 10.4-11.

The optimum narrow beam pattern from a line source antenna occurs when all side lobes are of equal level, just as in the array case. The required functional form, as we have

seen, is that of the Chebyshev polynomial. The Chebyshev polynomial $T_N(x)$ has $N - 1$ equal level "side lobes" in the region $-1 < x < 1$, and for $|x| > 1$, its magnitude increases monotonically. A change of variables will transform the Chebyshev polynomial into the desired pattern form; that is, with a zero slope main beam maximum at $x = 0$ and equal level side lobes. The new function resulting from the variable change is

$$P_{2N}(x) = T_N(x_o - a^2 x^2) \tag{10-64}$$

where a is a constant and

$$x = \frac{L}{\lambda} \cos \theta = \frac{L}{\lambda} w \tag{10-65}$$

At the pattern maximum,

$$P_{2N}(w = 0) = T_N(x_o) = R \tag{10-66}$$

which is the main beam-to-side lobe ratio. A plot of (10-64) for $N = 4$ is shown in Fig. 10-9; it is the transformed version of Fig. 10-6.

From (10-39), we have in the side lobe region

$$P_{2N}(x) = \cos[N \cos^{-1}(x_o - a^2 x^2)], \quad |x_o - a^2 x^2| < 1 \tag{10-67}$$

The zeros of this function occur when the cosine argument equals $(2n - 1)\pi/2$, or when the values of x are as follows:

$$x_n = \pm \frac{1}{a} \sqrt{x_o - \cos \frac{(2n - 1)\pi}{2N}}, \quad |n| \geq 1 \tag{10-68}$$

where the plus sign is used for zero locations on the positive x-axis and $x_{-n} = -x_n$. In the main beam region, from (10-39),

$$P_{2N}(x) = \cosh[N \cosh^{-1}(x_o - a^2 x^2)], \quad |x_o - a^2 x^2| > 1 \tag{10-69}$$

The main beam maximum value of P_{2N} is R and occurs for $x = 0$; see (10-65) and (10-66). Solving (10-69) for x_o at the main beam maximum yields

$$x_o = \cosh\left(\frac{1}{N} \cosh^{-1} R\right) \tag{10-70}$$
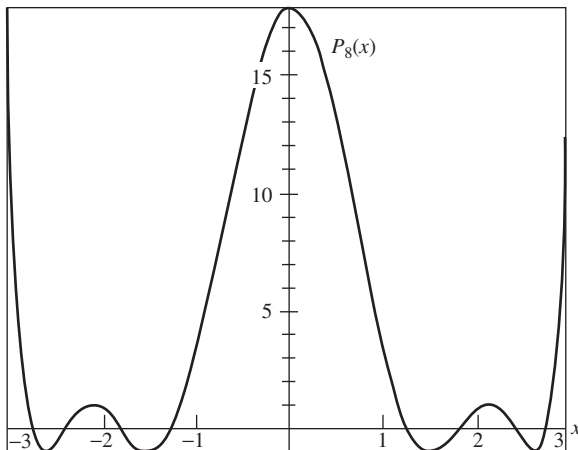
Figure 10-9 Transformed Chebyshev polynomial $P_8(x) = T_4(x_o - a^2 x^2)$. Values of $a = 0.55536$ and $x_o = 1.42553$ corresponding to Example 10-6 were used.

It is convenient to introduce A such that

$$A = \frac{1}{\pi} \cosh^{-1} R \tag{10-71}$$

so then

$$x_o = \cosh \frac{\pi A}{N} \tag{10-72}$$

In order to have all side lobe levels equal, we let N approach infinity, but simultaneously the argument of P_{2N} is changed to keep the first nulls stationary, thus leaving the beam-width unchanged. For large N, $x_o = \cosh(\pi A/N) \approx 1 + \frac{1}{2}(\pi A/N)^2$ and $\cos \frac{(2n-1)\pi}{2N} \approx 1 - \frac{1}{2}\left[\frac{(2n-1)\pi}{2N}\right]^2$ and using these in (10-68) gives

$$x_n = \pm \frac{1}{a} \frac{\pi}{\sqrt{2N}} \sqrt{A^2 + \left(n - \frac{1}{2}\right)^2} \qquad N \to \infty \tag{10-73}$$

By letting

$$a = \frac{\pi}{\sqrt{2N}} \tag{10-74}$$

the first zero location remains fixed as N increases. Then

$$x_n = \pm \sqrt{A^2 + \left(n - \frac{1}{2}\right)^2} \tag{10-75}$$

The pattern factor is a polynomial in x with an infinite number of roots x_n and can be expressed as a product of factors $(x - x_n)$ for n from $-\infty$ to $+\infty$. And since $x_{-n} = -x_n$, the pattern is

$$\prod_{n=1}^{\infty} (x^2 - x_n^2) = \prod_{n=1}^{\infty} \left[x^2 - A^2 - \left(n - \frac{1}{2}\right)^2\right] \tag{10-76}$$

Normalizing this to unity at $x = 0$ gives

$$f(x) = \frac{\displaystyle\prod_{n=1}^{\infty} \left[1 - \frac{x^2 - A^2}{(n-\frac{1}{2})^2}\right]}{\displaystyle\prod_{n=1}^{\infty} \left[1 + \frac{A^2}{(n-\frac{1}{2})^2}\right]} = \frac{\cos(\pi\sqrt{x^2 - A^2})}{\cosh \pi A} \tag{10-77}$$

The last step above utilizes the closed-form expression for the infinite products. Using (10-65) and (10-71) in (10-77) gives the pattern in w as

$$f(w) = \frac{\cos\left\{\pi\sqrt{[(L/\lambda)w]^2 - A^2}\right\}}{R} \tag{10-78}$$

Note that this is normalized to unity at the maximum ($w = 0$) and oscillates between $-1/R$ and $1/R$ in the side lobe region. For large w, the argument of the cosine function in (10-78) is approximately $\pi w L/\lambda$, so the zero locations of the pattern are $w_n \approx \pm \lambda(n - \frac{1}{2})/L$ or $x_n \approx \pm(n - \frac{1}{2})$, and thus they are regularly spaced. Also note that for $w < \lambda A/L$, the cosine argument of (10-78) is imaginary and since $\cos(j\theta) = \cosh \theta$, (10-78) is more conveniently expressed as

$$f(w) = \frac{\cosh\left\{\pi\sqrt{A^2 - [(L/\lambda)w]^2}\right\}}{\cos \pi A} \tag{10-79}$$

This pattern is that of the *ideal Taylor line source* [8]. It is a function of A that is found from the side lobe level; see (10-46) and (10-71). The line source is "ideal" in the sense that equi-level side lobes extend to infinity in pattern space, thus leading to infinite power. The required source excitation, in turn, must have infinite power and, in fact, will have singularities at each end of the line source.

An approximate realization of the ideal Taylor line source, referred to as the *Taylor line source,* nearly equals the first few side lobes but has decreasing far-out side lobes [8]. The decaying side lobe envelope removes the infinite power difficulty encountered with the ideal Taylor line source. The Taylor line source pattern is again a polynomial in x, but with zero locations given by

$$x_n = \begin{cases} \pm\sigma\sqrt{A^2 + \left(n - \dfrac{1}{2}\right)^2} & 1 \le n < \bar{n} \\[2ex] \pm n & \bar{n} \le n < \infty \end{cases} \tag{10-80}$$

The zeros for $n < \bar{n}$ are those of the ideal line source in (10-75) scaled by the factor σ. The far-out side lobes for $n \ge \bar{n}$ are located at the integer x positions. The zero arrangement for a $\sin(\pi x)/\pi x$ pattern is $x = \pm n$ for $n \ge 1$, so the Taylor pattern far-out side lobes are those of the $\sin(\pi x)/\pi x$ pattern. The scaling parameter σ is determined by making the zero location expressions in (10-80) identical for $n = \bar{n}$, which yields

$$\sigma = \frac{\bar{n}}{\sqrt{A^2 + (\bar{n} - \frac{1}{2})^2}} \tag{10-81}$$

From the zero locations of (10-80), we write the approximate Taylor line source pattern as

$$f(x, A, \bar{n}) = \frac{\sin \pi x}{\pi x} \prod_{n=1}^{\bar{n}-1} \frac{1 - (x/x_n)^2}{1 - (x/n)^2} \tag{10-82}$$

The side lobes are nearly constant at the value $1/R$ out to $x = \bar{n}$ and decay as $1/x$ beyond $x = \bar{n}$. The pattern in terms of $w = \cos\theta$ is

$$f(w, A, \bar{n}) = \frac{\sin(\pi L w/\lambda)}{\pi L w/\lambda} \prod_{n=1}^{\bar{n}-1} \frac{1 - (w/w_n)^2}{1 - (Lw/\lambda n)^2} \tag{10-83}$$

where the pattern zero locations on the w-axis are

$$w_n = \begin{cases} \pm\dfrac{\lambda}{L}\sigma\sqrt{A^2 + \left(n - \dfrac{1}{2}\right)^2} & 1 \le n < \bar{n} \\[2ex] \pm\dfrac{\lambda}{L}n & \bar{n} \le n < \infty \end{cases} \tag{10-84}$$

with σ given by (10-81).

The Taylor line source is actually a pattern of the Woodward–Lawson family. We show how this comes about and also determine the sample values and locations. First, assume that the required source excitation can be expanded in a Fourier series as

$$i(s) = \frac{\lambda}{L} \sum_{n=-\infty}^{\infty} a_n e^{-j2\pi(\lambda/L)ns} \qquad |s| \le \frac{L}{2\lambda} \tag{10-85}$$

The corresponding pattern from (10-17) is

$$f(w) = \sum_{n=-\infty}^{\infty} a_n \ \mathrm{Sa}\left[\left(w - \frac{\lambda}{L}n\right)\frac{L}{\lambda}\pi\right] \tag{10-86}$$

where the sample locations are identified as

$$w_n^s = \frac{\lambda}{L}n \tag{10-87}$$

The infinite expansion of (10-86) gives the exact pattern if the sample values are (see Prob. 10.4-7)

$$a_n = f(w = w_n^s) = f(n, A, \bar{n}) \tag{10-88}$$

But the pattern zeros correspond to the sample locations of (10-87) for $|n| \ge \bar{n}$ since $x_n = n$, or $w_n = (\lambda/L)n$ for $|n| \ge \bar{n}$ from (10-80). Thus,

$$a_n = 0 \quad \text{for} \quad |n| \ge \bar{n} \tag{10-89}$$

Using (10-88) and (10-89) in (10-86) gives the pattern expression

$$f(w) = \sum_{n=-\bar{n}+1}^{\bar{n}-1} f(n, A, \bar{n}) \ \mathrm{Sa}\left[(w - w_n^s)\frac{L}{\lambda}\pi\right] \tag{10-90}$$

The required current distribution from (10-85) is

$$i(s) = \frac{\lambda}{L}\left[1 + 2\sum_{n=1}^{\bar{n}-1} f(n, A, \bar{n}) \cos\left(2\pi\frac{\lambda}{L}ns\right)\right] \tag{10-91}$$

The coefficients $f(n, A, \bar{n})$ are the samples of Taylor line source pattern for $x = n$ and $n < \bar{n}$. They are found from

$$f(n, A, \bar{n}) = \begin{cases} \dfrac{[(\bar{n}-1)!]^2}{(\bar{n}-1+n)!(\bar{n}-1-n)!} \displaystyle\prod_{m=1}^{\bar{n}-1}\left(1 - \frac{n^2}{x_m^2}\right) & |n| < \bar{n} \\ \\ 0 & |n| \ge \bar{n} \end{cases} \tag{10-92}$$

and $f(-n, A, \bar{n}) = f(n, A, \bar{n})$. Also $f(0, A, \bar{n}) = 1$. Tables of the coefficient values are also available. [H.6: Hansen, Vol. I, Appendix I] These coefficients together with (10-90) and (10-91) determine the Taylor line source pattern and current distribution.

The half-power beamwidth expression is obtained rather easily for the ideal pattern. Evaluating (10-79) at the half-power points yields

$$\frac{1}{\sqrt{2}} = \frac{1}{R} \cosh\left[\pi\sqrt{A^2 - \left(\frac{L}{\lambda}w_{\mathrm{HP}}\right)^2}\right] \tag{10-93}$$

Solving this gives the two solutions

$$w_{\mathrm{HP}} = \pm\frac{\lambda}{L\pi}\left[(\cosh^{-1}R)^2 - \left(\cosh^{-1}\frac{R}{\sqrt{2}}\right)^2\right]^{1/2} \tag{10-94}$$

The half-power beamwidth in w is then

$$\text{HP}_{wi} = 2|w_{\text{HP}}| = \frac{2\lambda}{L\pi}\left[(\cosh^{-1}R)^2 - \left(\cosh^{-1}\frac{R}{\sqrt{2}}\right)^2\right]^{1/2} \tag{10-95}$$

The angle from broadside is $\gamma = \theta - 90°$, so $w = \cos\theta = \cos(\gamma + 90°) = -\sin\gamma$ and $\gamma = -\sin^{-1} w$. The half-power beamwidth based on the ideal Taylor line source is

$$\begin{aligned}\text{HP}_i &= |\theta_{\text{HP}_i \text{ left}} - \theta_{\text{HP}_i \text{ right}}| = |\gamma_{\text{HP}_i \text{ left}} - \gamma_{\text{HP}_i \text{ right}}| \\ &= |\sin^{-1} w_{\text{HP}+} - \sin^{-1} w_{\text{HP}-}| = 2|\sin^{-1}w_{\text{HP}}| \\ &= 2\sin^{-1}\left\{\frac{\lambda}{L\pi}\left[(\cosh^{-1}R)^2 - \left(\cosh^{-1}\frac{R}{\sqrt{2}}\right)^2\right]^{1/2}\right\} \tag{10-96}\end{aligned}$$

where $w_{\text{HP}+}$ and $w_{\text{HP}-}$ are the two solutions of (10-94). The beamwidth for the approximate Taylor line source is given approximately by

$$\text{HP}_w \approx \sigma\text{HP}_{w_i} \tag{10-97}$$

and in θ by

$$\text{HP} \approx 2\sin^{-1}\left\{\frac{\lambda\sigma}{L\pi}\left[(\cosh^{-1}R)^2 - \left(\cosh^{-1}\frac{R}{\sqrt{2}}\right)^2\right]^{1/2}\right\} \tag{10-98}$$

EXAMPLE 10-7   *A 10-Wavelength Taylor Line Source with* -25-*dB Side Lobes and* $\bar{n} = 5$

The side lobe ratio is

$$R = 10^{-\text{SLL}/20} = 10^{1.25} = 17.7828 \tag{10-99}$$

From (10-71),

$$A = \frac{1}{\pi}\cosh^{-1} R = 1.13655 \tag{10-100}$$

Then from (10-81),

$$\sigma = \frac{\bar{n}}{\sqrt{A^2 + (\bar{n} - \frac{1}{2})^2}} = 1.07728 \tag{10-101}$$

If we use values of A and σ, the zero locations x_n can be calculated from (10-80), and then the sample coefficients follow from (10-92) as given in Table 10-8. The sample locations from (10-87) are also tabulated. The pattern and current distribution can now be computed from (10-90) and (10-91) with the sample values and locations of Table 10-8. The resulting pattern and current distribution are plotted in Fig. 10-10. The side lobe decay envelope for the far-out side lobes of the pattern is shown in Fig. 10-10a. The half-power beamwidths from (10-95) to (10-98) are

$$\text{HP}_{wi} = 0.0978, \qquad \text{HP}_i = 5.606° \tag{10-102}$$

and

Table 10-8 Sample Values and Locations for the Taylor
Line Source of Example 10-7 ($L = 10\lambda$, $\bar{n} = 5$)

n	$a_n = f(n, A, \bar{n}) = f(n, 1.13655, 5)$	w_n^s
0	1.000000	0
± 1	0.221477	± 0.1
± 2	-0.005370	± 0.2
± 3	-0.006621	± 0.3
± 4	0.004917	± 0.4
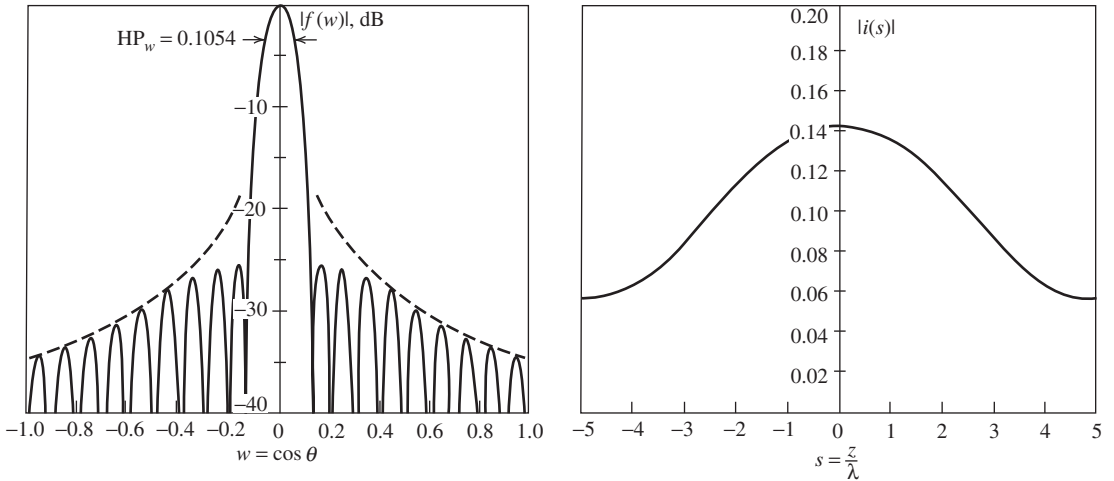
(a) The synthesized pattern.

(b) The current distribution.

Figure 10-10 A 10λ Taylor line source with -25-dB side lobes and $\bar{n} = 5$ (Example 10-7).

$$\text{HP}_w \approx 0.1054, \quad \text{HP}_i \approx 6.039° \tag{10-103}$$

In this case, the ideal Taylor line source beamwidth is very close to that of the approximate Taylor line source. The half-power beamwidth HP_w is indicated in Fig. 10-10a.

10.5 THE ITERATIVE SAMPLING METHOD

Several other synthesis methods are presented along with many references in [H.6: Balanis, Ed., Sec. 13.5]. One method that merits mention involves manipulation of the array polynomial of (8-76). [9] Varying the placement of the $P-1$ zeros (roots) in the complex Z-plane covers the range of possible synthesized patterns. A set of rules using the roots for pattern control is presented in [10]. In this section, we present a method called the *iterative sampling method* that can be used for both continuous and array antennas that have one or multiple dimensions.

Classical synthesis methods such as Woodward–Lawson or Fourier methods produce patterns that often do not come sufficiently close to the desired pattern over at least some parts of the pattern. In addition, once a pattern is synthesized with a classical method, there is limited ability to change the synthesized pattern. In essence, classical methods are "one-shot" methods in the sense that if the synthesized pattern is not a good enough fit to the desired pattern, a different method must be tried. The iterative sampling method, on

the other hand, starts with a pattern that roughly approximates the desired pattern, and then adjustments are made to the pattern (and associated current distribution) to improve the fit to the desired pattern. Although convergence to the desired pattern is not guaranteed, good results are obtained in practice. If the desired pattern is over specified such that the specifications cannot be realized, then as with any method the closeness of fit to the desired pattern must be relaxed or the aperture size increased. The iterative sampling method is appealing because of its simplicity and because it is applicable to nearly any antenna type, including linear arrays, linear apertures, nonuniformly spaced arrays, and muliti-dimensional antennas. In the following, we illustrate the iterative sampling method for line sources. More details and applications are found in [11].

The Fourier transform method provides lowest mean squared error, which is an integrated error, between the desired and synthesized pattern. The iterative sampling method, on the other hand, can minimize the maximum deviation from the desired pattern, which is called a minimax error and is the most appropriate quality-of-fit measure for most applications. The maximum allowed deviation is specified over regions of the far field where pattern fit is critical. The iterative sampling method is based on the Woodward–Lawson formulation of (10-16) to (10-18). [12] But instead of requiring the individual pattern samples to be uncorrelated, they can be weakly correlated. This gives the method flexibility and permits iterative application for a custom fit to the desired pattern.

The iterative sampling method begins with any initial pattern $f^{(0)}(w)$. The initial pattern is usually obtained by a classical synthesis method such as the Woodward–Lawson method but could be any pattern, such as a measured pattern. Next, correction patterns $\Delta f^{(i)}(w)$ are added to the initial pattern $f^{(0)}(w)$ to adjust the total pattern toward the desired pattern. The general expression for the synthesized pattern after K iterations is

$$f^{(K)}(w) = f^{(0)}(w) + \sum_{i=1}^{K} \Delta f^{(i)}(w) = f^{(0)}(w) + \sum_{i=1}^{K} \sum_{n=-M}^{M} a_n^{(i)} G_n^{(i)}(w) \qquad (10\text{-}104)$$

Each iteration has a maximum of $2M+1$ component correction patterns $G_n^{(i)}(w)$ and usually fewer are used. The component correction can be any function, but the uniform line source sampling function pattern, Sa, used in the Woodward–Lawson method is usually adequate. The weighting coefficient, $a_n^{(i)}$, is usually taken to be the difference between the previous iteration pattern value and the desired pattern value at the sample point:

$$a_n^{(i)} = f_d(w_n^{(i)}) - f^{(i-1)}(w_n^{(i)}) \qquad (10\text{-}105)$$

The correction pattern locations, $w_n^{(i)}$, can be anywhere, but the algorithm that works best is to place the new samples at the mean location between previous samples:

$$w_n^{(i)} = \frac{w_n^{(i-1)} + w_{n-1}^{(i-1)}}{2} \qquad (10\text{-}106)$$

The corrections are applied symmetrically about the coordinate origin to maintain any pattern symmetry that exists and to maintain a real-valued pattern and current. The source distribution corresponding to the synthesized pattern is found by the Fourier transform of (10-104).

$$i^{(K)}(s) = i^{(0)}(s) + \sum_{i=1}^{K} \sum_{n=-M}^{M} a_n^{(i)} g_n^{(i)}(s) \qquad (10\text{-}107)$$

where $i^{(K)}(s)$ is the current responsible for the pattern $f^{(K)}(w)$ and $g_n^{(i)}(s)$ is the current responsible for the pattern $G^{(i)}(w)$.

Many examples of patterns synthesized with the iterative sampling method are presented in [11–13], including line sources, two-dimensional continuous sources, linear

(equally or unequally spaced) arrays, and planar arrays for a shaped-beam, low side lobes, or both. Thus, the iterative sampling method can be used for all antenna types and desired patterns within one simple framework. The method can even be used with reflector antennas to compensate for surface distortions. [14] For example, when an array is used to feed a mesh reflector that deploys in space with distortions, the iterative sampling method can be used to adjust the feed element excitations to reduce deleterious pattern effects such as high side lobes.

REFERENCES

1. P. M. Woodward, "A Method for Calculating the Field over a Plane Aperture Required to Produce a Given Polar Diagram," *J. IEE,* Vol. 93, Part IIIA, pp. 1554–1558, 1946.
2. P. M. Woodward and J. D. Lawson, "The Theoretical Precision with Which an Arbitrary Radiation Pattern May Be Obtained from a Source of Finite Extent," *J. IEE*, Vol. 95, Part III, pp. 363–370, Sept. 1948.
3. C. L. Dolph, "A Current Distribution for Broadside Arrays Which Optimizes the Relationship between Beam Width and Side-lobe Level," *Proc. IRE*, Vol. 34, pp. 335–348, June 1946.
4. A Safaai-Jazi, "A New Formulation of the Design of Chebyshev Arrays," *IEEE Trans. on Ant. & Prop.*, Vol. 42, pp. 439–443, March 1994.
5. A Safaai-Jazi, "Directivity of Chebyshev Arrays with Arbitrary Element Spacing, *Electronics Lett.*, Vol. 31, pp. 772–774, 11 May 1995.
6. R. S. Elliott, "Beamwidth and Directivity of Large Scanning Arrays," *Microwave J.*, Part 1, Vol. 6, pp. 53–60, Dec. 1963.
7. N. Yaru, "A Note on Super-Gain Antenna Arrays," *Proc. IRE*, Vol. 39, pp. 1081–1085, Sept. 1951.
8. T. T. Taylor, "Design of Line-Source Antennas for Narrow Beamwidth and Low Sidelobes," *IRE Trans. Ant. & Prop.,* Vol. AP-3, pp. 16–28, Jan. 1955.
9. S. A. Schelkunoff, "A Mathematical Theory of Linear Arrays," *Bell System Tech. J.*, Vol. 22, pp. 80–107, Jan. 1943.
10. T. T. Taylor and J. R. Whinnery, "Application of Potential Theory to the Design of Linear Arrays," *J. Appl. Phy.*,Vol. 22, pp. 19–29, Jan. 1951.
11. W. L. Stutzman and E. L. Coffey, "Radiation Pattern Synthesis of Planar Antennas Using the Iterative Sampling Method," *IEEE Trans. Ant. & Prop.,* Vol. AP-23, pp. 764–769, Nov. 1975.
12. W. L. Stutzman, "Synthesis of Shaped-Beam Radiation Patterns Using the Iterative Sampling Method," *IEEE Trans. on Ant. and Prop.,* Vol. AP-19, pp. 36–41, Jan. 1971.
13. W. L. Stutzman, "Sidelobe Control of Antenna Patterns," *IEEE Trans. on Ant. and Prop.*, Vol. AP-20, pp. 102–104, Jan. 1972.
14. W. T. Smith and W. L. Stutzman, "A Pattern Synthesis Technique for Array Feeds to Improve Radiation Performance of Large Distorted Reflector Antennas," *IEEE Trans. on Ant. and Prop.*, Vol. 40, pp. 57–62, Jan. 1992.

PROBLEMS

10.1-1   If $g(t)$ and $G(\omega)$ are a Fourier transform pair, then

$$G(\omega) = \int_{-\infty}^{\infty} g(t)e^{-j\omega t}dt \quad \text{and} \quad g(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(\omega)e^{j\omega t}\, d\omega$$

If g, t, G, and ω are replaced by f, w, i, and $2\pi s$, respectively, show that (10-6) and (10-7) follow from the above equations.

10.1-2 Decomposing a linear current distribution $i(s)$ into real/imaginary and even/odd parts and applying the Fourier transform, give the following pattern expression:

$$f(w) = 2\int_0^{\infty} [i_{re}(s)\ \cos 2\pi ws\ - i_{io}(s)\sin 2\pi ws]\ ds$$

$$+ j2\int_0^{\infty} [i_{ie}(s)\ \cos 2\pi ws + i_{ro}(s)\sin 2\pi ws]\, ds$$

Use this to prove the following: (a) Property 1 in Table 10-2, and (b) Property 2 in Table 10-2 by first forming the magnitude of $f(w)$.

10.2-1 A narrow pencil beam pattern represented by a delta function is scanned to the location $\theta = 53.1°$. Find the required current distribution using Fourier line source synthesis.

10.2-2 Use (10-6) to find the pattern from a uniform amplitude, zero phase line source of length L centered on the z-axis.

10.2-3 Derive the current distribution $i_d(s)$ in (10-12) required to exactly produce the sector pattern of (10-11b).

10.2-4 (a) Derive the Fourier transform synthesis pattern of (10-13) for a sector pattern.
(b) Plot this pattern, thus verifying Fig. 10-1a. Numerical integration of the Fourier transform via computer may be easier than using (10-13).

10.2-5 Derive (10-15).

10.2-6 Repeat the Woodward–Lawson synthesis of the sector pattern of Example 10-2, but this time for a five-wavelength line source.
   (a) Plot the pattern in linear, rectangular form as a function of w.
   (b) Plot the current distribution.

10.2-7 A cosecant pattern (see Prob. 2.5-7 for a discussion of the cosecant pattern) is given by

$$
f_d(w) = \begin{cases} 1 & 0 \le w \le 0.1 \\[2mm] \dfrac{0.1}{w} & 0.1 \le w \le 0.5 \\[2mm] 0 & \text{elsewhere} \end{cases}
$$

Use the Woodward–Lawson method to synthesize an approximation to this pattern for a 10λ line source.
   (a) Plot the pattern in linear, rectangular form together with the desired pattern as a function of w.
   (b) Plot the required current amplitude and phase.

10.3-1 Discuss the conditions on $f_d(w)$ such that it can be represented by the Fourier series in (10-24).

10.3-2 Derive the Fourier series coefficient expression in (10-25).

10.3-3 Derive the element current expression (10-29) for the Fourier series synthesis of a sector pattern.

10.3-4 Synthesize a sector pattern with $c = 0.5$ using the Fourier series method as in Example 10-3 for an array of 20 elements that are spaced 0.6λ apart.
   (a) Determine the element locations and current values.
   (b) Plot the radiation pattern in linear, rectangular form as a function of w.

10.3-5 Repeat Prob. 10.3-4 for an array of 10 elements and half-wavelength spacings.

10.3-6 Use the Fourier series synthesis method to synthesize a sector pattern with $c = 0.5$ for an array of 21 half-wavelength spaced elements. Evaluate the element currents. Plot the pattern. Compare pattern parameters to the 20-element array result of Example 10-3.

10.3-7 Repeat the cosecant pattern synthesis as in Prob. 10.2-7a, using the Fourier series method for a 20-element, half-wavelength spaced array. Tabulate the element current values.

10.3-8 Show that the Woodward–Lawson sampling method pattern of (10-32) arises from the array factor with the currents of (10-35) for:
   (a) An odd number of elements. *Hint*: Use (10-19) and (10-20).
   (b) An even number of elements. *Hint*: Use (10-21) and (10-22).

10.3-9 Verify the array element positions and currents of Table 10-5 for the Woodward–Lawson synthesized sector pattern of Example 10-4.

10.3-10 Repeat the Woodward–Lawson synthesis as in Example 10-4 but for a 10-element, half-wavelength spaced array.

10.3-11 A linear array can be synthesized by sampling the current distribution of a line source with an acceptable pattern, and the resulting array pattern will be close to that of the line source. Create a linear array of 20, half-wavelength spaced elements by sampling the current distribution of Example 10-2. Plot the array factor and comment on how it compares to the pattern of Fig. 10-2a.

10.3-12 A collinear array of 18 half-wave dipole antennas is to be used to synthesize a sector pattern with a main beam sector over the region $70° \leq \theta \leq 110°$, that is, $F_d(\theta) = 1$ over this region and zero elsewhere.

   (a) For 0.65λ spacings, determine the input currents required for Woodward–Lawson synthesis of the complete pattern. Account for the element pattern.
   (b) Plot the total array pattern in linear, polar form as a function of θ.

10.3-13 Repeat Prob. 10.3-12 for a cosecant desired pattern, where $F_d(\theta)$ is 1 for $80° \leq \theta \leq 90°$, cos $80°$/cos θ for $0° \leq \theta \leq 80°$, and zero elsewhere. Use 18 pattern samples.

10.4-1 For the five-element, broadside, –20 dB side lobe, half-wavelength spaced Dolph-Chebyshev array of Example 10-5:

   (a) Calculate and plot the pattern in rectangular, logarithmic form as a function of w.
   (b) Find the half-power beamwidth in two ways: from the pattern calculation and using the formula of (10-50a).
   (c) Compare the beamwidth to that of a uniform line source of the same length as the array.

10.4-2 Design a Dolph-Chebyshev broadside array of five, half-wavelength spaced isotropic elements for –30 dB side lobes.

   (a) Calculate elements currents to verify the values in Table 10-7.
   (b) Compute the directivity using (8-84) and using the approximate directivity formula of (10-51) with half-power beamwidth from (10-50a).

10.4-3 Design a Dolph-Chebyshev array of 10 isotropic elements for -20 dB side lobes for the following cases:

   (a) Broadside and optimum spacing. Find the spacing value. Plot the pattern in rectangular, logarithmic form as a function of w. Calculate the beamwidth and directivity.
   (b) Repeat (a) for endfire.

10.4-4 Design a low-side lobe, broadside collinear array of half-wave dipoles. Design a Dolph-Chebyshev array of eight isotropic elements with the narrowest beamwidth for -20 dB side lobes. Then plot the patterns in polar-dB form with and without the element pattern.

10.4-5 Derive the ideal Taylor line source pattern results of (10-76) and (10-77).

10.4-6 Show how the approximate Taylor line source pattern of (10-82) follows from the zero locations.

10.4-7 The sampling theorem from time-signal analysis states that a signal $g(t)$ is exactly reconstructed from the time samples $g(m/2B)$ as

$$g(t) = \sum_{m=-\infty}^{\infty} g\left(\frac{m}{2B}\right) \text{Sa}\left[2\pi B\left(t - \frac{m}{2B}\right)\right]$$

where B is the highest frequency component of the signal. Draw the appropriate analogies to antenna theory to obtain the sampled data pattern expression of (10-86).

10.4-8 Verify (10-94).

10.4-9 Compute the sample values a_n of Table 10-8 for the Taylor line source of Example 10-7.

10.4-10 Compare the half-power beamwidth values for the Taylor line source of Example 10-7. Compare your answers to those of (10-102) and (10-103).

10.4-11 An array antenna can be designed by choosing the element current excitations at the corresponding points of the continuous current from a line source synthesized for the desired pattern. This is illustrated in this problem with a narrow main beam, low side lobe pattern. The Taylor line source of Example 10-7 has current values appropriate for a 20-element array given in the table:

Array Excitations for Problem 10.4-11

m	z_m/λ	i_m
±1	±0.25	0.14234
±2	±0.75	0.13833
±3	±1.25	0.13127
±4	±1.75	0.12175
±5	±2.25	0.10935
±6	±2.75	0.09429
±7	±3.25	0.07891
±8	±3.75	0.06676
±9	±4.25	0.05980
±10	±4.75	0.05720

(a) Use these current values to obtain the array factor of the corresponding linear array.
(b) Compare and comment on the half-power beamwidths and side lobe levels of the array and line source patterns.

10.4-12 Design an eight-wavelength Taylor line source $(\bar{n} = 7)$ with -30-dB side lobes.
(a) Obtain and tabulate the sample values and locations.
(b) Plot the pattern in rectangular-logarithmic form as a function of w.
(c) Plot the current distribution.

10.4-13 Evaluate σ for several values of $\bar{n}$ for the case of a -25-dB side lobe level. Using $HP_w \approx \sigma HP_{wi}$, explain the half-power beamwidth behavior as a function of $\bar{n}$.

10.4-14 Design an optimum broadside Dolph–Chebyshev array with 10 elements and -20-dB side lobes. With the same array geometry, find the element currents by sampling a -20-dB Taylor line source $(\bar{n} = 8)$ current distribution with the same length. Plot the polar-dB pattern for both arrays and compare for: (a) Isotropic elements, and (b) Collinear half-wave dipoles along the line of the array.

10.4-15 *Effect of mutual coupling on array synthesis.* Use a moment method code to evaluate the array of Example 10-5; see Chap. 14. and Appendix G. Use resonated half-wavelength dipoles parallel to the *x*-axis with centers along the *z*-axis with voltage sources proportional to the desired currents. Compute the *yz*-plane pattern in polar form and compare to that of the example. Give the side lobe levels and currents for the two approaches.

Chapter 11

Low-Profile Antennas and Personal Communication Antennas

11.1 INTRODUCTION

Wireless communication was discussed in Sec. 4.5 from a system perspective. This chapter focuses on antennas for *wireless personal communications* (*WPC*), which employ a communication device that a person can carry or easily move from place to place. Of necessity, these antennas should be low-profile—that is, of small physical thickness. Antennas in or protruding from a wireless terminal are needed to support several wireless personal communication applications, many simultaneously. Some example applications are cellular telephone communications; Wi-Fi, Bluetooth, and UWB communications; RFID (radio frequency identification); position location (such as GPS) and asset tracking; and body area networks (BAN). A few of the applications permit the use of tunable antennas that move a narrow frequency band over a large frequency range; examples are TV and radio reception in which only one channel is received at a time. Demand is increasing for smaller radio terminals used by warfighters, first responders, and law enforcement personnel. Not all applications are conventional two-way communication uses, but we treat them together because of the similar antenna requirements of low profile and compact volume. The need for a physically small terminal leads to the requirement for an even smaller antenna. The essence of the handset antenna challenge is to accommodate an ever increasing number of wireless services and frequency bands within a shrinking allocated volume.

Handset-based communication devices such as cellular telephones and personal digital assistants form a large part of WPC, and we will highlight this application. In a handset there are certain regions of the internal volume that are designated as "keep-out" zones, where an antenna cannot be located, thus restricting the possible antenna volume. A typical volume maximum is about $5 \, cm^3$ with dimensions of $35 \times 35 \times 4 \, mm$. The printed circuit board size for a bar shape phone is $100 \times 40 \, mm$ ($3.9 \times 1.6 \, in.$), but smart phones often are larger. The antenna is expected to support several cellular telephone bands as well as Wi-Fi, TV and FM radio reception, MIMO modes, and so on, in this small volume. One technique used to relax the size constraint is to make use of other handset parts for the antenna ground plane, such as a printed circuit board. However, handset antenna performance requirements are less demanding than in many other areas of antenna design. Typical performance metrics for handset antennas are

Impedance match:	VSWR <3 (return loss $>6\,$dB)
Antenna gain:	0 dBi
Radiation efficiency:	50%
Bandwidth:	8 to 12% (depending on the band)

The low gain specification is due to the need for a very low directivity pattern to minimize signal variations as handset position is varied. Handset antennas can be external or internal, or both, but multiple internal antennas is the most popular approach. As more antennas and antenna functions are added to a handset, antenna isolation is more difficult to achieve. [1] Also, the close proximity to active devices on the platform increases noise and self-interference. Although most handset antennas are unbalanced (i.e., monopole type), balanced antennas (i.e., dipole type) are desirable because of reduced influence of the chassis and human operator. While balanced antennas are larger than their unbalanced counterpart, they are more efficient, are less susceptible to noise, and provide more isolation. Diversity features can also be incorporated such as polarization diversity or spatial diversity. In the latter, a second antenna on the handset provides a second copy of the signal via a different propagation path that may offer improved reception (see Sec. 12.4 for more on diversity). The handsets themselves are generally one of two shapes, clam shell (that fold in the middle) or bar (single block). The antenna design differs for these two form factors. The perfect handset antenna, albeit unobtainable, would be one that can be made as small as desired to fit into any allocated volume; presents no biohazard; and experiences no frequency shifting, pattern distortion, gain loss, or impedance mismatch due to the presence of the human operator. However, the antenna size limitations are physics-based and not technology-based, as will be discussed in Sec. 11.5, which treats the fundamental limits on antenna size.

In this chapter we feature several antennas that can be used in personal communication devices. Many of the antenna types that have been covered in previous chapters are revisited here with added emphasis on how the antennas are configured for common applications. We introduce microstrip antennas and arrays, and leaky wave antennas, as well as other topics related to personal communications such as antenna near fields, human body effects, and radiation hazards. The next chapter addresses antennas for larger terminals such as base stations, satellites, and vehicles.

11.2   MICROSTRIP ANTENNA ELEMENTS

The microstrip antenna is a special type of **printed antenna**, which is constructed using methods similar to those used for printed circuits. Example printed antennas are the printed dipole of Fig. 8-35a, tapered slot antennas of Fig. 8-35c and d, and the printed bow-tie antenna array of Fig. 8-33. The **microstrip antenna** (MSA) consists of a metallic *patch* printed on top of a thin substrate with a ground plane on the bottom of the substrate, as illustrated in Fig. 11-1. The MSA is usually less than $0.05\,\lambda$ thick, where λ is the free space wavelength. The basic feeds for the patch are a *probe feed* using a coaxial transmission line below the ground plane or an *edge feed* using a coplanar microstrip transmission line connected to an edge of the patch. The edge-fed patch is a very low profile antenna that also can include other components using microwave integrated circuit techniques and the feed network when arrayed. This offers the advantage of low-cost, controlled-dimension construction. Printing on flexible substrates allows the microstrip antenna to be wrapped conformally onto a vehicle, for example. The radiation pattern is a single, broad unidirectional beam broadside to the patch due to the ground plane greatly reducing the back radiation (zero for an infinitely large ground plane). Because the MSA is a resonant antenna (see Table 1-4) it is narrowband, leading to its biggest design challenge—achieving adequate bandwidth. Other disadvantages include spurious radiation from the feed, poor cross-polarization purity, limited power handling, and adjustment difficulty after fabrication.
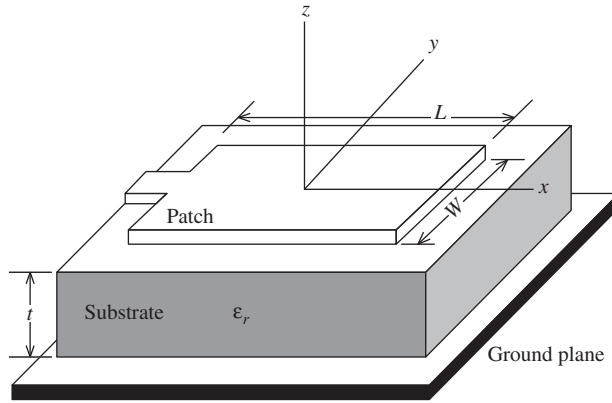
(*a*) Geometry for analyzing the edge-fed microstrip patch antenna.
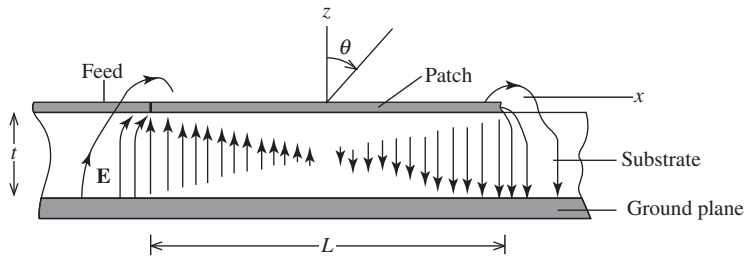
(*b*) Side view showing the electric fields.
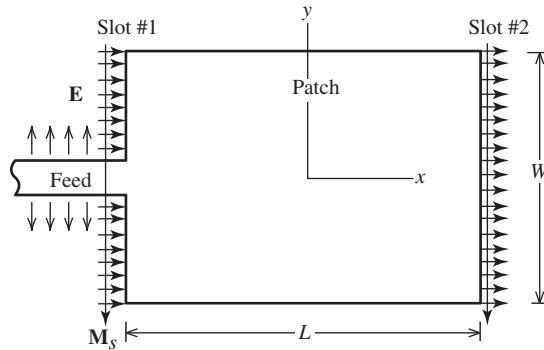
(*c*) Top view showing the fringing electric fields. The equivalent magnetic surface currents are also shown.

Figure 11-1 The half-wavelength rectangular patch microstrip antenna: $L \approx 0.49\lambda_d$.

The microstrip antenna is a recent innovation in the field of antennas. It was first proposed in 1950s, and intense research in the 1970s on the basic MSA resulted in the first comprehensive publication on the MSAs in 1981 [2]. Although the radiation pattern of the basic MSA varies slowly with frequency, the impedance does not. Thus, impedance bandwidth is the limiting factor of the basic MSA, because it is only a few percent. In the 1980s, new feed methods led to enhanced bandwidth. In the 1990s, commercial systems began widely adopting MSAs. Now MSAs are commonly used as isolated antennas and as elements in arrays. A popular application is in base station antennas, usually as a linear array. Perhaps the widest application is in position location devices. Fixed mounted and portable GPS units in cars usually use a microstrip antenna. The MSA is not as popular

for use in non-GPS bands in handsets because of its large physical size. But GPS signals are narrowband, so a MSA with a high dielectric constant substrate to reduce size is small enough to find wide use for GPS reception. Laptop computers also use MSAs for Wi-Fi links. The MSA as an isolated antenna or an array usually appears on a list of candidate antenna types for a new antenna design project. It is easy to predict that there will be many new applications for microstrip antennas that will emerge in the future.

In this section, we present several single microstrip antenna types, followed in the next section by arrays of microstrip elements. More than four decades of research and development devoted to the MSA has led to many books on the topic, which are listed in Appendix H.8.4.

11.2.1   Rectangular Microstrip Patch Antennas

The simplest MSA is a rectangular patch on top of a substrate material with relative dielectric substrate ε_r of thickness $t \ll \lambda$ and backed by a large ground plane as shown in Fig. 11-1a. The edge feed is a microstrip transmission line on the left side of the patch. The patch is L long (in the plane of the feed line) and W wide. When the length is approximately a half wavelength in terms of the wavelength in the dielectric λ_d, the patch becomes resonant (input reactance is zero) and is called a *half-wave rectangular patch antenna* (or *half-wave patch*). The electric field distribution for a half-wave patch shown in Fig. 11-1b displays fringing fields that extend beyond the patch. This fringing makes the patch effectively longer than its physical length, so the resonant length is less than a half-wavelength. In addition to ε_r, the amount of length reduction to achieve resonance also depends on t, and W, and formulas including these variables are presented later in this section. But, for many purposes the following simple formula is adequate:

$$L \approx 0.49\lambda_d = 0.49\frac{\lambda}{\sqrt{\varepsilon_r}} \qquad \text{half-wave patch} \qquad (11\text{-}1)$$

where λ is the free-space wavelength at resonance and λ_d is the wavelength of a wave in an infinitely large block of dielectric with ε_r. In practice, this formula is used as a starting point for an experimental hardware model or for simulation. Next, the input impedance is measured or calculated as a function of frequency spanning the design frequency, and the resonant frequency is noted. Then, the model is scaled in size to shift the resonance to the desired frequency, followed by a second measurement or simulation for verification. For laboratory models, the patch size is reduced by trimming with a knife or increased by adding conductive tape. The width W is used to adjust the input resistance, as we will see next.

A popular material for MSA substrate is PTFE (polytetrafluorethylene) woven glass laminate with $\varepsilon_r = 2.50$; Teflon® is an example. Micowave PTFE-based materials such as Duroid® are commercially available in various thicknesses and dielectric constants (e.g., $\varepsilon_r = 2.22$, 2.33, 6.15, 10.2). FR-4 with a dielectric constant of 4.6 is fiberglass material with epoxy resin binder that offers low cost, but with high loss, so it is used less often for frequencies above a few GHz. Ceramics such as Barium Titanate are used for high dielectric constant MSAs like GPS patches. The resonant frequency of narrow band patches is sensitive to variations of material thickness and dielectric constant. Thus, it is important to have a material that is manufactured with proper quality control over material uniformity (especially ε_r), has low loss, and has a selection of thicknesses and sheet sizes. As frequency increases, these are even more important considerations.

Approximate analytical models were developed in the early years of MSA development to evaluate parameters such as input impedance, pattern, and bandwidth because simulation codes were inadequate and slow. With today's efficient computational algorithms and fast computing platforms, accurate simulations of MSAs are now commonly performed and several commercial packages are available. The analytical models yield

formulas, often including empirically based parameters, that are used as a first step in simulations. The models presented by Carver and Mink [2] remain an excellent starting point. Recent developments found in [H.6: *Ant. Eng. Hdbk.*, 4th ed., Chap. 7; H.6: Balanis, Ed., Chap. 4; H.8.4: Lee and Chen, Chap. 5] augment the original work and provide more accurate formulas. The two main analytical methods are the transmission-line model and the cavity model. The *transmission-line model* is essentially a one-dimensional model that assumes uniform fields in *y*-direction in the patch and treats the patch as two parallel radiating slots, each having an equivalent parallel slot admittance. The *cavity model* is more generally applicable than the transmission-line model because it is not restricted to rectangular patches and is not limited to one-dimensional variations. The cavity model includes all patch internal field variations through a summation of modal fields. The TM_{10} mode is identical to the transmission-line mode, which has been used to develop the formulas to follow that include dependencies on the several rectangular patch parameters. Accuracy degrades for *t* greater than about $0.02\,\lambda$ and for high dielectric constants. The design formulas begin with a more accurate equation for the resonant patch length:

$$L = 0.5\frac{\lambda}{\sqrt{\varepsilon_r}} - 2\Delta L \tag{11-2}$$

where ΔL is *fringing length* given by

$$\Delta L = 0.412\,\frac{(\varepsilon_{re} + 0.3)\left(\dfrac{W}{t} + 0.264\right)}{(\varepsilon_{re} - 0.258)\left(\dfrac{W}{t} + 0.8\right)}\,t \tag{11-3}$$

where ε_{re} is the *effective dielectric constant* given by

$$\varepsilon_{re} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2}\left(1 + \frac{10\,t}{W}\right)^{-0.5} \tag{11-4}$$

The total fringing length, $2\Delta L$, due to the fringing on both edges is the amount the patch length should be reduced below a half wavelength to achieve resonance. Viewed another way, ΔL is the amount to be added to each edge to form an effective patch length. It is evident from (11-3) that ΔL increases with substrate thickness approximately linearly. This is because as the thickness increases, the electric fields extend farther and "fringe" more. As an example, at 10 GHz ($\lambda = 3$ cm) for a 0.01-λ-thick substrate with $\varepsilon_r = 2.2$ the preceding three formulas yield $L = 0.98$ cm, which is in good agreement with the approximate formula of (11-1) that gives 0.99 cm.

   Because the radiation pattern of a microstrip antenna is unidirectional broadside to the patch, the rectangular patch of Fig. 11-1 will have a beam peak in the *z*-direction. We now apply antenna theory principles to derive approximate pattern expressions for the rectangular patch. First, we note that the microstrip feed line will not radiate because the width of the upper conductor of the microstrip transmission line is small compared to a wavelength, whereas the patch is large relative to a wavelength. Thus, the electric field lines on the two sides of the upper conductor of the microstrip transmission line are oppositely directed (in the *y*-direction) as shown in Fig. 11-1*c*. The radiation from these fringing fields cancels in the far field because the distance to the far field from each opposing arrow in a pair is nearly the same, leaving a phase difference of 180° due to the opposing directions. But the situation is different for the patch. The transmission-line model viewpoint of two radiating slots is useful in finding the radiation pattern of a rectangular patch with the region between the patch and the ground plane acting as a half-wavelength transmission line that is open circuited at the end. Fig. 11-1*b* shows the

electric field lines associated with the standing wave mode inside the patch. The electric field lines are perpendicular to the conductors as required by boundary conditions and look much like those in a parallel plate capacitor. The fringing fields at the ends are exposed to the upper half space ($z > 0$). Unlike any other antenna, it is these fringing fields that are responsible for the radiation. The half-wavelength (in the dielectric) separation between the ends leads to electric fields that are of opposite sign on the left and right halves of the patch as shown in Fig. 11-1b. Therefore, the fringing fields at the ends are 180° out of phase and equal in magnitude. But, as the top view in Fig. 11-1c shows, the half-wavelength separation has brought the fringing fields back in-phase, leading to a reinforcement of radiation broadside to the patch. The radiation can be calculated by either treating the two edges as radiating slot apertures with electric fields extending beyond the edges, or by employing equivalent magnetic currents. We will use the latter approach.

The equivalent magnetic surface currents are found from (2-23) with the fringing electric fields $\mathbf{E}_a$ using $\mathbf{M_S} = 2\mathbf{E}_a \times \hat{\mathbf{n}}$. The factor of 2 comes from the image of the magnetic current in the conducting ground plane (see Fig. 9-4c) for small t. The far-field components for this geometry with the z-axis normal to the patch (see Prob. 11.2-12) as

$$E_\theta = E_o \cos \phi f(\theta, \phi) \tag{11-5a}$$

$$E_\phi = -E_o \cos \theta \sin \phi f(\theta, \phi) \tag{11-5b}$$

where

$$f(\theta, \phi) = \frac{\sin\left[\dfrac{\beta W}{2} \sin \theta \sin \phi\right]}{\dfrac{\beta W}{2} \sin \theta \sin \phi} \cos\left(\frac{\beta L}{2} \sin \theta \cos \phi\right) \tag{11-5c}$$

and β is the usual free-space phase constant. The first factor in (11-5c) is the pattern factor for a uniform line source of width W in the y-direction. The second factor is the array factor for a two-element array along the x-axis corresponding to the edge slots; see (3-71). The radiated electric field is linearly polarized in the plane containing the feed, because the source electric fields shown in Fig. 11-1c are polarized in the xz-plane, making it the E-plane. The principal plane patterns follow from (11-5) as

$$F_E(\theta) = \cos\left(\frac{\beta L}{2} \sin \theta\right) \qquad E\text{-plane}, \phi = 0° \tag{11-6a}$$

rectangular patch antenna principal plane patterns

$$F_H(\theta) = \cos \theta \frac{\sin\left[\dfrac{\beta W}{2} \sin \theta\right]}{\dfrac{\beta W}{2} \sin \theta} \qquad H\text{-plane}, \phi = 90° \tag{11-6b}$$

This simple H-plane pattern expression is based on a symmetric standing wave with uniform fields along the radiating edges. As a result there are no cross-polarized fields, but in reality cross-polarization can be significant in MSAs.

The input impedance at the edge of the resonant rectangular patch ranges from 100 to 400 Ω. An approximate expression for the input resistance (reactance is zero at resonance) of a resonant edge-fed patch is [3].

$$Z_A = 90 \frac{\varepsilon_r^2}{\varepsilon_r - 1} \left(\frac{L}{W} \right)^2 \Omega \quad \textit{half-wave rectangular patch antenna} \tag{11-7}$$

This formula reveals that input resistance decreases as the patch is widened. For example, for a dielectric with $\varepsilon_r = 2.2$, a patch width-to-length ratio of $W/L = 1$ gives an input resistance of $363\,\Omega$. A very wide patch with $W/L = 2.7$ is required to obtain an input impedance of $50\,\Omega$.

A narrow band antenna such as a microstrip antenna is limited by impedance match to the connecting transmission line. As operating frequency moves off resonance, the patch input impedance increases rapidly, causing large mismatch. The efficiency of the MSA improves as the impedance mismatch is reduced, which is accomplished by widening the bandwidth. An empirical formula for the bandwidth is [3]:

$$B = 3.77 \frac{\varepsilon_r - 1}{\varepsilon_r^2} \frac{W}{L} \frac{t}{\lambda} \qquad t \ll \lambda \qquad \textit{rectangular patch bandwidth} \tag{11-8}$$

where B is the fractional bandwidth relative to the resonant frequency (see (7-1)) for VSWR ≤ 2. For example, for a $0.02\text{-}\lambda$ thick dielectric with $\varepsilon_r = 2.2$, a square patch ($W/L = 1$), the bandwidth is 0.019, or 1.9%. For thicknesses larger than about $t = 0.15\,\lambda$, (11-8) loses accuracy and more complete models must be used as in [3]. Evident from (11-8) is that bandwidth can be increased by increasing the dielectric thickness or the patch width.

The final important parameter is radiation efficiency e_r (often referred to as just efficiency) which for MSAs represents dissipation from dielectric and conductor losses, surface waves, and cross-polarization content. Not included in radiation efficiency is loss encountered due to impedance mismatch, but usually one of the impedance matching techniques discussed below is used to eliminate this loss at resonance. Radiation efficiency is as high as 95% for thin, low dielectric constant substrates and decreases with increasing substrate thickness; see [3] or [4] for evaluation formulas and [H.6: *Ant. Eng. Hdbk.*, 4th ed., Fig. 7-8] for efficiency values as a function of substrate properties.

The design process for a MSA focuses on optimizing performance values for the application at hand such as having an input impedance of a certain value (often $50\,\Omega$), achieving a specified bandwidth, or having high efficiency. The first step is to find the patch length L for resonance at the center frequency of the desired operating band using (11-1). Next, the patch width is found by back solving (11-7) for achieving a desired input impedance or (11-8) for achieving a desired bandwidth. If high efficiency is the goal, the following width is selected [H.8.4: Bahl & Bhartia, p. 57]:

$$W = \frac{\lambda}{2} \left[\frac{\varepsilon_r + 1}{2} \right]^{-1/2} \quad \textit{patch width for high efficiency} \tag{11-9}$$

which is calculated at resonance. The length is then recalculated using the more accurate formulas in (11-2) to (11-4) or a simulation code. MSA computer simulation algorithms use so-called full-wave solutions because they account for the full wave behavior contained in Maxwell's equations. The method of moments (see Chap. 14) is very popular for MSA simulations; however the finite element method and the FDTD method (see Chap. 15) are also used; see [H.6: Balanis, Ed., p. 175] for the available commercial microstrip antenna codes. Full-wave solutions have more capabilities than models and can handle irregular geometries, feed details, and material characteristics. The outputs include the input impedance and the patterns (co- and cross-polarized). However, even with the latest computing hardware, simulations can be time consuming, especially for arrays. Thus, approximate models remain valuable as first-cut solutions and for

validations of simulation results. The design phase continues with iterations on parameter values of ε_r, t, W, and L, attempting to converge on the desired operating band, impedance, gain, and so on. Simulations and measurements with prototype hardware are usually required for validation of the final design.

EXAMPLE 11-1   *Half-Wave, Square Microstrip Patch Antenna*[1]

A square, half-wave patch was designed to be resonant at 3.03 GHz ($\lambda = 9.9$ cm) on a $t = 0.114$ cm thick substrate of $\varepsilon_r = 2.35$ dielectric. The patch dimensions are found from (11-1):

$$L = W = 0.49 \frac{\lambda}{\sqrt{\varepsilon_r}} = 0.49 \frac{9.9}{\sqrt{2.35}} = 3.165 \text{ cm} \qquad (11\text{-}10)$$

This value is close to that from the more accurate formulas of (11-2) to (11-4) which yield $L = 3.11$ cm. Fig. 11-2 shows the measured input impedance as a function of frequency. The resonant frequency (where the reactance is zero in Fig. 11-2a) is 3.01 GHz, which is close to the design frequency of 3.03 GHz. The measured input resistance from Fig. 11-2b at the resonant frequency of 3.01 GHz is 316 Ω, which compares to 369 Ω computed from (11-7). Also shown in Fig. 11-2 are values computed using the cavity model of Carver and Mink [2] and using method of moments simulations. Note the excellent agreement between simulations and measurements. The bandwidth from (11-8) is 1.1 %. The radiation patterns computed from (11-6) are plotted together with measured values in Fig. 11-3, showing that agreement of measured data to the simple theory results is very good. The beamwidths in the *E*- and *H*-planes are 103° and 81°, leading to a directivity of 6.9 dB using (4-56).

In addition to the direct, edge feed of Fig. 11-1, the main techniques for feeding patches are shown in Fig. 11-4 and can be classified into three groups: directly coupled (*a, b, c*), electromagnetically coupled (*d, e, f*), or aperture coupled (*g*). Direct coupling methods have limited freedom to adjust the input impedance but are simple to implement. The edge feed techniques of Figs. 11-1, 11-4b and 11-4c are used for planar, low-profile realizations. This approach is especially well suited to arrays of MSAs, where the patch, feed, and other feed network components can be fabricated on a single layer. The probe feed of Fig. 11-4a allows access from below the patch. The rectangular patch is normally fed along the patch centerline in the *E*-plane as shown in Fig. 11-1. This avoids excitation of a second resonant mode orthogonal to the desired mode, which would lead to excessive cross-polarization levels.

The only means of input impedance control with the edge-fed patch of Fig. 11-1 is by changing the patch shape (*L/W*). The feeds in Fig. 11-4 offer additional degrees of freedom in controlling the impedance. The edge-fed patch of Fig. 11-4b includes a *quarter-wave transformer*, which is widely used in microwave systems. The antenna input impedance of Z_A is matched to a transmission line of characteristic impedance Z_o (often 50 Ω) with a section of a transmission line that is a quarter-wavelength long based on the wavelength in the transmission line and has characteristic impedance given by

$$Z_o' = \sqrt{Z_A Z_o} \qquad \textit{quarter-wave transformer} \qquad (11\text{-}11)$$

where the impedances are real-valued. The characteristic impedance of a microstrip transmission line is controlled by the width of the strip, much as the loss resistance of a

---

[1] Measurements and simulations were performed by the Virginia Tech Antenna Group.

(a) Input reactance.

(b) Input resistance.
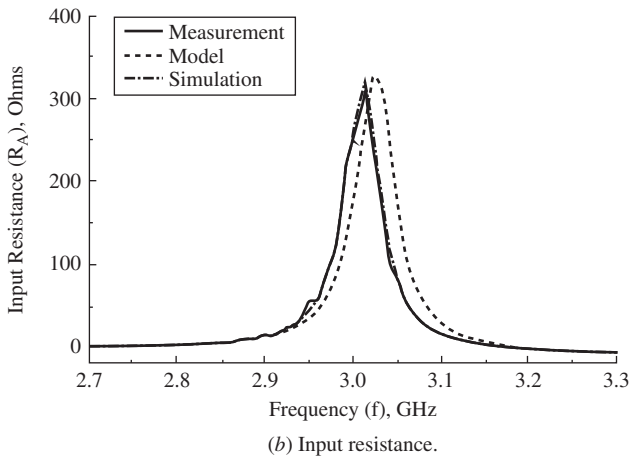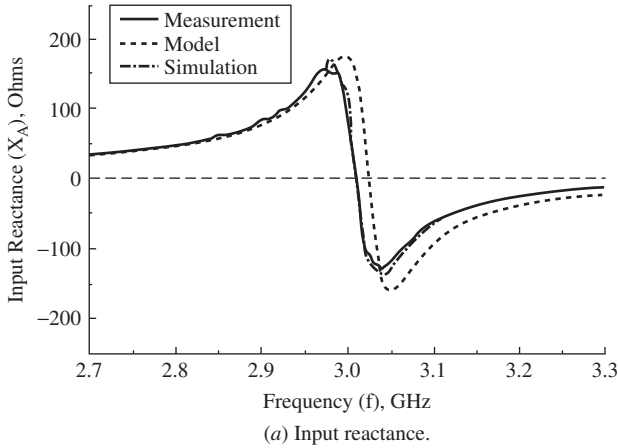
Figure 11-2 Measured input impedance for the square-patch microstrip antenna of Example 11-1 as a function of frequency. Also shown are computed values from the cavity model of Carver and Mink [2] and from method of moments (MoM) simulation.

(a) E-plane. (b) H-plane.
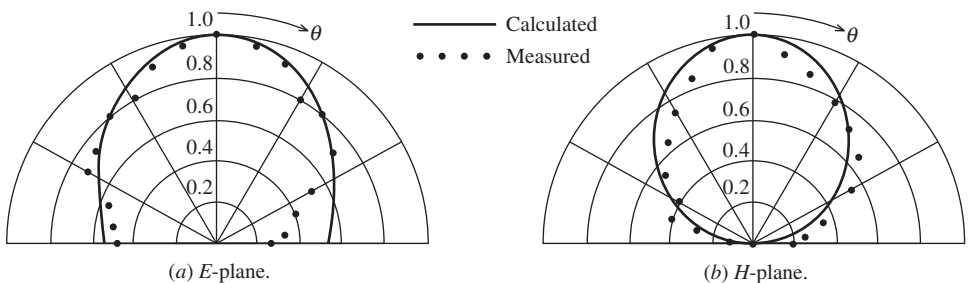
Figure 11-3 Radiation patterns for the square mictrostrip patch of Example 11-1: calculated (curves) using (11-6), and measured (points).

wire is inversely proportional to the wire diameter in (2-170). That is, the wider the strip, the lower the characteristic impedance.

The direct coaxial probe feed of Fig. 11-4a is simple to implement by extending the center conductor of the connector attached to the ground plane upward to the patch. Impedance can be adjusted by proper location of the probe attachment point. As the probe

(*a*) Probe feed.   (*b*) Microstrip edge feed with quarter-wave transformer.   (*c*) Microstrip edge feed with inset.

(*d*) Probe feed with a gap.   (*e*) Microstrip edge feed with gap.   (*f*) Two-layer feed.

(*g*) Aperture coupled feed.
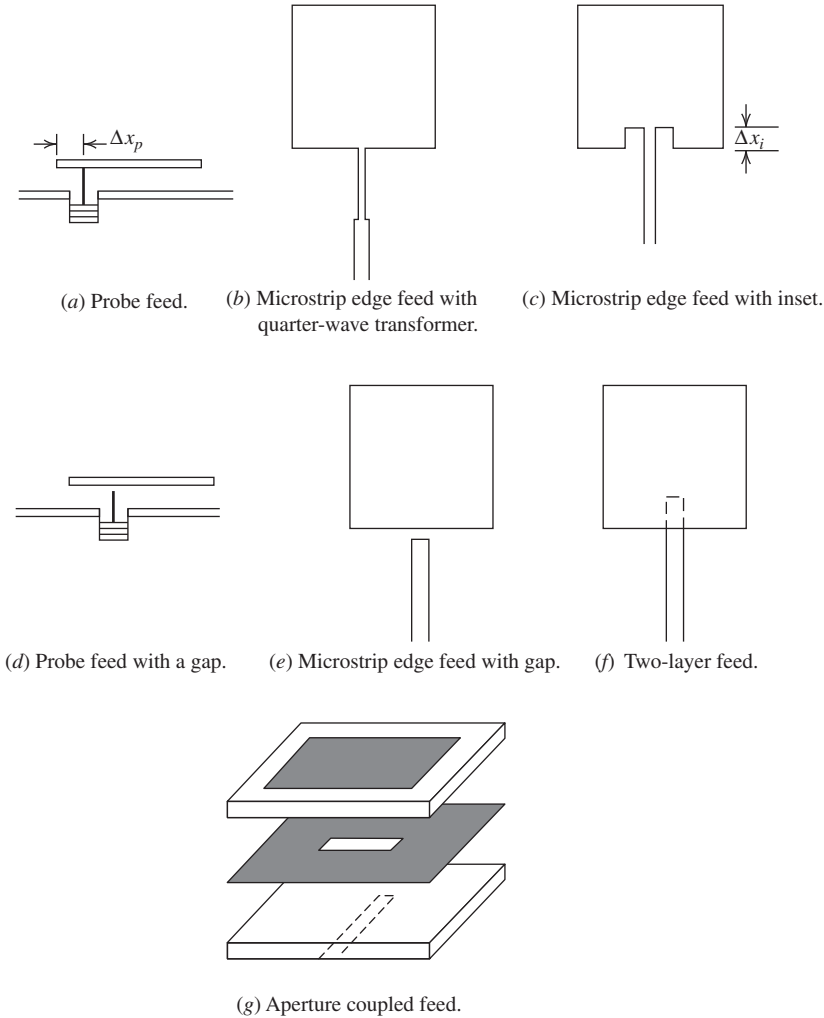
Figure 11-4 Techniques for feeding microstrip patch antennas. Along with Fig. 11-1, (*a*), (*b*), and (*c*) are direct feeds; (*d*), (*e*), and (*f*) are electromagnetic coupled feeds; and (*g*) is an aperture coupled feed.

distance from the patch edge, Δx_p, in Fig. 11-4*a* is increased, the input resistance decreases and is expressed as [3].

$$Z_A(\Delta x_p) = Z_A(\Delta x_p = 0) \cos^2 \frac{\pi \Delta x_p}{L} \qquad \textit{probe-fed patch} \qquad (11\text{-}12)$$

where $Z_A(\Delta x_p = 0)$ is given in (11-7). A disadvantage of the probe feed is that it introduces an inductance that can prevent the patch from resonating if t is 0.1λ or more. Also, the probe itself will radiate, adding to the cross polarization. The inset feeding of Fig. 11-4*c* also offers input impedance control by varying the inset distance Δx_i. The input impedance expression is the same as (11-12) except that the power on the cosine function is 4 instead of 2 and Δx_i is the argument in place of Δx_p. [5] A more accurate formula using a cosine squared function with a shifted argument and a family of empirically derived constants is available in [6]. Large input impedance changes that are required for

high-permittivity substrates demand a significant inset depth, which affects cross polarization and the radiation pattern shape.

Direct feeds have narrow bandwidth, which can be increased by increasing the substrate thickness, but with the penalty of increasing the power trapped in surface waves. An *electromagnetically coupled feed* (also called proximity, noncontacting, or gap feed) does not contact the patch and has at least two design parameters. It has the advantage of being less sensitive to dimensional errors from the etching process. For each direct feed in Figs. 11-4a, b, and c, there is a gap feed counterpart in Figs. 11-4d, e, and f. The probe feed with a gap in Fig. 11-4d is fed from below with a probe, often formed by extending the center conductor of a coaxial cable. The gap between the end of the probe and the patch introduces capacitance that partially cancels the probe inductance, permitting the use of thicker substrates. The microstrip feed line with a gap of Fig. 11-4e is entirely planar and easy to etch. However, in high-permittivity substrate designs, the gap distance can become small. The two-layer feed of Fig. 11-4f is especially useful in microstrip arrays with a top layer for the patches and a second layer for the microstrip feed network.

There are conflicting goals for the antenna and its connecting transmission line. The goal with an antenna is to encourage radiation, and with a transmission line the goal is to bind the wave to the line. The *aperture coupled feed* of Fig. 11-4g addresses this conflict by using an upper substrate of low dielectric constant for better radiation and a lower substrate of high-dielectric constant for containing the fields to the feed line. This also provides wider bandwidth. Another advantage is that the central ground plane acts to isolate the feed system from the patch.

11.2.2 Other Microstrip Patch Antennas and Their Applications

Many patch shapes and configurations have been investigated. Following the rectangular patch, the circular patch is probably is the most common shape. It can be made slightly smaller than its rectangular counterpart, but with a small penalty in gain and bandwidth. [H.8.4: Waterhouse, p. 27] See [H.8.4: Garg et al., Chap. 5] for analysis details on circular patches.

The half-wave rectangular patch MSA is too large for many applications, especially in arrays. An alternative is the *quarter-wave patch antenna* which has a patch length of $L \approx \lambda_d/4$ at resonance, which is half that of the half-wave patch. The principle of operation is understood by first noting from Fig. 11-1b that the electric fields in the center of the half-wave patch are zero. Thus, a conducting wall can be placed at $x = 0$ with little effect on the fields for $x < 0$. The quarter-wave patch is then formed by removing the patch half to the right of the shorting wall (for $x > 0$). In practice, the conducting wall is usually implemented with shorting pins or vias between the ground plane and the patch. With only the open edge to radiate, the *E*-plane pattern is that of a single slot and the arraying effect of two slots with the half-wave patch is lost, giving a broad pattern. Such a broad pattern makes the antenna useful as an element in a wide scan angle phased array. The *H*-plane pattern is similar to that of the half-wave patch. The input radiation resistance at resonance is about twice that of the similar half-wave patch. The bandwidth is smaller than as its counterpart half-wave patch, and cross-polarization is higher.

MSAs with a single feed point in the center of the radiating edge will produce linear polarization. Circularly polarized (CP) radiation can be created using a square patch with two feed points on orthogonal edges and an external means to excite the ports with equal amplitude and 90° out of phase (leading or lagging phase determine the sense of left or right CP); see Sec. 2.8 for the basics on antenna polarization. This is accomplished with either edge or probe feeds. The two feed point approach can also be used to generate dual orthogonal linear polarizations with a single patch. Single feed point CP patch implementations are shown in Fig. 11-5. These CP patches operate by internally generating orthogonal modes of equal amplitude and quadrature phase by slight perturbations in the symmetric geometry. Although patches of many different shapes are capable of CP

(*a*) Nearly square corner-fed patch.

(*b*) Slotted square patch.

(*c*) Corner trimmed square patch.

(*d*) Offset probe-fed square patch.
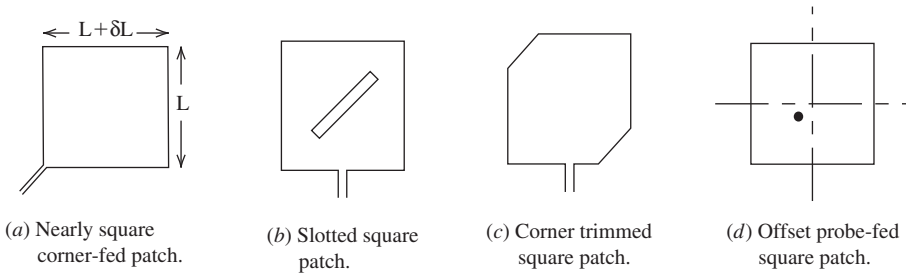
Figure 11-5 Microstrip antennas for generating circular polarization.

operation, usually modified square or circular patches are used in practice. The price paid for having a simple geometry is narrow CP bandwidth (usually specified as under 3 dB axial ratio) that is much narrower than the patch impedance bandwidth, which itself is usually narrow. The principle of operation for a single-patch CP antenna is understood by modeling the nearly square patch ($L + \delta L$ by L) of Fig. 11-5*a* as two orthogonal dipoles, one slightly longer than half-wave resonant and the other slightly less. The input reactances will differ in phase as noted from Fig. 6-6 on each side of resonance. Similarly, the orthogonal patch mode impedances will be out of phase and, with proper dimensions, the radiated fields will be 90° out of phase. Inserting a slot into the patch as in Fig. 11-5*b* or trimming the corners of a square patch as in Fig. 11-5*c* are also used to create CP. The first three CP patches in Fig. 11-5 are edge fed and the final one in Fig. 11-5*d* is an off-center, probe-fed patch, which is very popular due to its compact feed coming from below the patch. See [H.9.1: Hirasawa and Haneishi, Sec. 4.5] for more details on the operation and for design data on CP patches.

Microstrip antennas are the main antenna used for reception of global navigation satellite system (GNSS) signals, which are right-hand CP. They are ideally suited because of their small size and low cost. The narrow bandwidth is not a limitation because of the narrowband signal. For example, the GPS L1 frequency spans 1565–1585 MHz for a 20 MHz, or 1.3%, bandwidth. To achieve small size the substrate is typically a ceramic with a dielectric constant of 20 or 36. The probe-fed, nearly square patch is popular. Using $\varepsilon_r = 36$ in (11-1) yields a patch side length of about 15 mm. If a substrate with $\varepsilon_r = 2.2$ had been used, the side length would be 63 mm, which is a factor of 4.2 larger in side length and unacceptably large for most applications. The high-dielectric constant patch in this case has an area that is less than 6% of the conventional patch area. The gain of a GPS patch antenna is lower than that of a conventional patch for two reasons. First, the smaller antenna size increases the beamwidth in the principal planes with the benefit of better coverage of satellites in off-zenith directions, but lower the directivity. Second, ceramics are lossy, leading to a radiation efficiency around 70%. Thus, a typical gain value for a ceramic patch GPS antenna is 5 dB. GPS antennas in handset wireless phones present a special challenge because of the size constraint. One solution is to print a monopole onto the PCB, but the gain is very low and there is an additional loss of 3 dB due to the LP-to-CP polarization mismatch. MSAs are used in handsets also, but suffer from several dB gain loss due to the small ground plane available. GNSS systems transmit multiple frequency bands, so multiband receiving antennas are often required. The common solution is to stack a smaller patch for the upper band on top of a patch resonant at the low band; see [H.9.1: Kumar, Sec. 5.5] for stacked patch designs.

The MSA is a ground-plane dependent antenna, and our discussions to this point have assumed an infinitely large, good conductor ground plane. In many applications, the available ground plane size is limited and performance will be compromised. A general guideline is that if the ground plane is a few wavelengths in size, the pattern is unaffected. If the ground plane is more than twice the patch size, the impedance is unaffected. If the ground plane is larger than the patch size by $\lambda/20$ on all sides, the resonant frequency will

not be increased. [H.8.4: Garg et al., p. 295] With finite size ground planes, the pattern will have ripples due to interfering diffractions from the ground plane edges. The pattern deviates the most from its infinite ground plane pattern in the endfire directions, and there will be significant radiation below the ground plane. Cross-polarization levels are increased with finite ground planes as well.

Microstrip antennas, elements, and arrays are also popular in non-planar applications. A widely used example is the so-called *wrap-around antenna* in which the MSA is printed on a flexible material that is then applied to a curved surface such as the skin of a vehicle. [7] Wrap-around MSAs are ideally suited for use on aircraft, missiles, and rockets where conformal antennas are a must. See [H.8.4: Wong, *Design of Nonplanar Microstrip Antennas and Tr. Lines*] for several topics in non-planar MSAs. Finally, we mention that MSAs are now being used in millimeter-wave applications (above 30 GHz); see [H.8.4: Bhartia et al.].

11.2.3   Broadband Microstrip Patch Antennas

The intensive research and development efforts into overcoming the main limitation of MSAs, narrow bandwidth, have provided several broadband designs. Because MSAs are inherently narrowband, we define broadband patch antennas to be about 10% or more. Here we review the techniques for realizing a broadband MSA; the following references provide comprehensive treatments: [H.8.4: Bancroft; Kumar and Ray; Wong, *Compact and Broadband Micro. Ant.*; Garg et al., Chap. 9; Waterhouse, Chap. 3; Pozar and Schaubert, Chap. 4], [H.8.3: Chia and Chen], and [H.6: Godara, Sec. 6.4]. It is evident from (11-8) that the bandwidth of a rectangular patch element can be increased by increasing the dielectric thickness. Bandwidth as high as 20% is possible using this simple method with substrates about 0.15 λ_d thick. [8] However, thick MSAs suffer from high feed radiation (probe or edge fed), leading to excessive cross polarization, and high feed inductance (probe fed). In array applications, surface wave generation leads to increased side lobes and cross-polarization.

The inductance introduced by a thick substrate can be canceled by adding capacitance. The gap-fed patch of Fig. 11-4a and the slotted patch of Fig. 11-5b provide this tuning effect. A small slot in a rectangular patch can provide 7% bandwidth. [H.8.4: Bancroft, Sec. 4.2] Patch shape optimization yields 8% bandwidth. [H.8.4: Bancroft, Sec. 4.3] Adding a shorting pin or low-resistance chip resistor between the patch edge and the ground plane to a conventional patch increases its bandwidth to 10%. [H.8.4: Wong, *Compact and Broadband Microstrip Ant.*, Sec. 1.3] Shorting pins can also be used to reduce the size of an MSA. [H.8.4: Waterhouse, Chap. 5] Meandering the patch to give an apparent longer current path also reduces patch size. [H.8.4: Wong, *Compact and Broadband Microstrip Ant.*, Sec. 2.3].

Very broadband operation is usually achieved with a double tuning effect realized by adding parasitic elements such as coplanar or stacked patches. Although 20% bandwidth is possible using parasitic patches coplanar to the radiating patch [H.8.4: Waterhouse, Sec. 3.3], the design is not feasible if space is limited. Numerous stacked patch designs, however, are widely used. The two popular stacked patch antennas shown in Fig. 11-6 are electromagnetically coupled and aperture-coupled stacked patches. Both have a feed layer on the bottom, usually of high dielectric constant, a lower patch on a substrate for the next layer, and an upper patch on a substrate for the top layer. Stacked MSAs increase bandwidth over a single patch because of the increased overall thickness, the decrease in effective dielectric constant, and the multiple resonances. The *electromagnetically coupled stacked patch antenna* of Fig. 11-6a consists of two substrate layers (lower with ε_{r1} and t_1, and upper with ε_{r2} and t_2) with a gap between of thickness g filled with air or foam material. The lower patch is directly fed and the upper patch is parasitically fed from the lower patch. With the upper patch slightly smaller than the lower patch, the two associated resonant frequencies are close together, yielding a broadband effect.

(*a*) Electromagneticially coupled stacked patch.

Figure 11-6 Stacked-patch microstrip antennas for broadband operation.
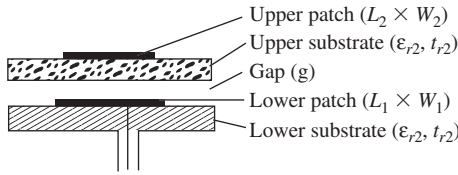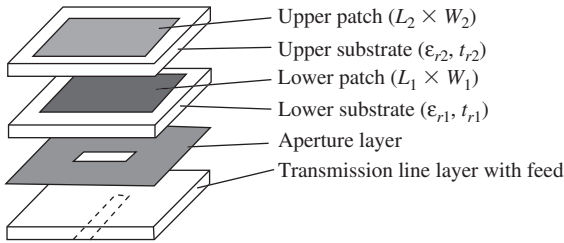
(*b*) Aperture-coupled stacked patch (viewed with layers separated).

With no gap ($g = 0$), the bandwidth can reach 8%. The presence of the gap increases the thickness and thus the bandwidth. Designs with over 20% bandwidth and gains with and without the gap of 6.8 dB and 8.3 dB, respectively, have been reported. [H.8.4: Kumar and Ray]. These square-shaped, stacked patches have cross-polarization below −25 dB.

The aperture-coupled patch antenna of Fig. 11-4*g* uses a metal layer with an aperture to feed the radiating (upper) patch. The *aperture-coupled stacked patch* (also called an aperture-stacked patch) of Fig. 11-6*b* makes use of the aperture layer as another resonator, in addition to the two patches, to further broaden bandwidth. With careful design, it is capable of almost 70% bandwidth. [9] A thorough treatment of parameter tradeoffs is found in [H.8.4: Kumar and Ray]. In summary, we see that stacked patch antennas offer greater bandwidth and gain than single patch antennas at the expense of increased overall thickness.

Many applications do not require broad bandwidth, but instead multiple narrow bands. GNSS satellite navigation satellite receiving terminals, for example, use two or more narrow bands. Land mobile radio (cellular telephone) handset terminals require multiple band compact antennas. WLAN terminals do also. Microstrip antennas with multiple resonances offer a compact, inexpensive solution. The simplest way to dual band a MSA is to use two feeds on adjacent edges that excite orthogonal modes on the same patch, such as a rectangular patch which has a resonant frequency ratio approximately equal to the ratio of the patch side dimensions. Low cross-polarization will, however, be difficult to achieve. Other means of multi-banding a patch are to use multiple patches or to use reactive loading. See the following for design details: [H.8.4: Bancroft, Chap. 5; Kumar and Ray, Chap. 7] and [10].

11.3   MICROSTRIP ARRAYS

Low-profile single antenna elements, such as a microstrip antenna, are capable of no more than about 10 dB gain. Many applications require a low-profile antenna with moderate to high gain, and an array of microstrip antennas is an attractive solution. The **microstrip array antenna** offers the opportunity to print the feed network at the same time as the radiating elements, often fabricating them on the same single-layer printed circuit board. Antenna technology is following the evolutionary path toward full integration similar to that of low-frequency electronics. Largely as the result of pioneering work for sophisticated military radar, phase-scanned microstrip arrays can be produced with monolithic

microwave integrated circuit (MMIC) techniques that fabricate amplifiers, phase shifters, and other devices on the same substrate. Replacing older array technology using discrete components with integrated subassemblies, especially for large arrays, yields savings in hardware volume and cost.

The microstrip array antenna has many possible configurations. Microstrip arrays are very popular for fixed-phase arrays applications because the radiating elements and feed network can be fabricated on a single-layer printed circuit board using low-cost lithographic techniques. The inter-element spacing for fixed-beam arrays is usually selected to be less than a free-space wavelength (λ) to avoid grating lobes and greater than $\lambda/2$ to provide sufficient room for the feed lines, to achieve higher gain for a given effects number of elements, and to reduce mutual coupling. Mutual coupling effects depend on the array architecture, the spacings, and the scan range. The active-element patterns for the microstrip elements shown in Fig. 8-27 are very similar, indicating that mutual coupling effects may not be a problem for those patches that are spaced $0.57\,\lambda$ apart. Element spacings in scanned arrays need to be reduced from that for fixed-phase arrays, with more reduction as the maximum scan angle off broadside increases; see Sec. 8.9.1.

As with any array, the microstrip array geometry can be linear, planar, or conformal and the feed system can be parallel, series, or hybrid; see Fig. 8-31. Series feeds are either resonant or traveling wave. The resonant-series-fed microstrip array has elements spaced multiples of a guide wavelength apart along the feed line, which makes both the incident and reflected waves in-phase, resulting in a broadside beam. The resonant array bandwidth is very narrow, usually less than 1%. [H.6: Balanis, Ed., p. 170] A traveling wave microstrip array absorbs power arriving at the end of the array in either a terminal element or a load, avoiding a reflected wave. As a consequence, any spacing is possible and the bandwidth is wider than for the resonant array. The feed network for the series-fed array is not only more compact than a parallel-fed array with the same number of elements, but has lower dissipative losses because of the smaller total line length in the feed network.

Parallel-fed arrays with their equal path lengths to each element have wider bandwidth than series-fed arrays. The successive branching used in parallel feed networks (without special power dividers) lowers the input impedance. But proper impedance can be maintained by using microstrip transmission lines in the feed network. In fact, a microstrip feed network enables impedance matching, phase control, and amplitude tapering, all in a planar feed network, which is a major advantage over other array implementations. Here we illustrate a simple parallel feed network and associated impedance matching using the 2×2 array shown in Fig. 11-7a that can be used as a fundamental subarray unit to build a planar array of any number of elements divisible by 4 such as the 16-element example in Fig. 11-7b. The four patches in the subarry are edge fed with microstrip transmission lines of equal length from the subarray center (point C) to preserve equal excitation phase. To illustrate, the patch input impedance is taken to be $Z_A = 200\,\Omega$. The microstrip feed lines attached to the patches (points A) are of characteristic impedances $Z_{o1} = 200\,\Omega$. These lines are matched to the patch loads, and thus present $200\,\Omega$ at each branch at points B and D. The parallel combination gives a $100\,\Omega$ impedance looking into points B and D. Again matched microstrip lines are used, this time with $Z_{o2} = 100\,\Omega$. Paralleling at point C gives $50\,\Omega$. If only a four-element array were desired, a probe feed attached to a 50-Ω coaxial line could be connected through ground plane to point C. For larger arrays, the 50-Ω impedance at point C is transformed using a quarter-wave transformer, producing $100\,\Omega$ at point E. The required characteristic impedance is found from (11-11) as $Z_{o3} = \sqrt{Z_C Z_{o4}} = \sqrt{50 \cdot 100} = 70\,\Omega$. That is, the quarter-wavelength-long section of $Z_{o3} = 70\,\Omega$ line is followed by a line of $Z_{o4} = 100\,\Omega$ that can be of any length. This impedance matching pattern is repeated as subarrays are added to build larger arrays as shown in Fig. 11-7b. This design produces equal amplitude element excitations. If amplitude control is needed, for example to lower side lobes, the two-way dividing branch points are followed by different characteristic impedance lines to feed element pairs. [H.8.4: Bancroft, p. 156]

(*a*) The four-element subarray.   (*b*) A 16-element array formed from subarrays.
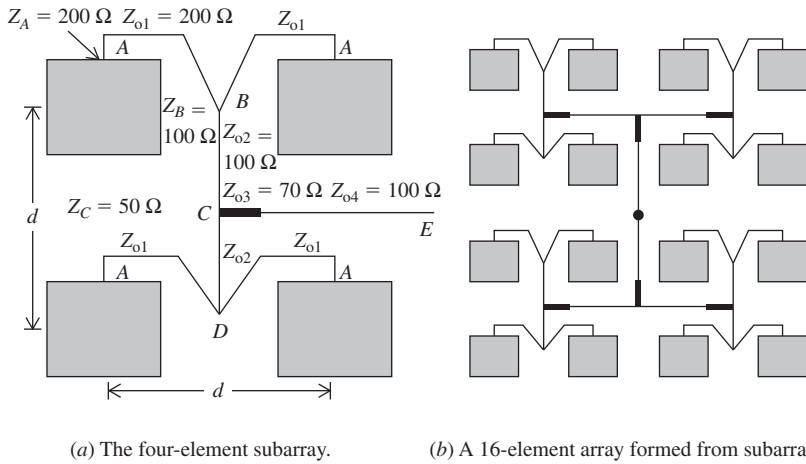
Figure 11-7 A planar microstrip array with a feed network that produces equal amplitude and phase element excitations.

EXAMPLE 11-2   *Microstrip Patch Array of 16 Elements*

The goal for this array design is to achieve high gain. This requires equal amplitude and phase to each element and a simple feed network with low loss. The architecture of Fig. 11-7*b* is an ideal solution that also provides a low-profile geometry. The operating frequency is 10 GHz ($\lambda = 3$ cm) and the substrate is a dielectric with $\varepsilon_r = 2.2$. From (11-1), the patch length is $L = 0.49\lambda/\sqrt{\varepsilon_r} = 0.49 \cdot 3/\sqrt{2.2} = 0.99$ cm. The desired patch input impedance is 200 Ω, so (11-7) is used to solve for the required patch width:

$$W = \sqrt{\dfrac{90\dfrac{\varepsilon_r^2}{\varepsilon_r - 1}}{Z_A}}L = \sqrt{\dfrac{363}{200}} \cdot 0.99 = 1.335 \text{ cm} \tag{11-13}$$

The element spacing is chosen to be $d = 0.8\ \lambda = 2.4$ cm, which is relatively large to provide large array area and large directivity for the fixed number of elements. This spacing provides sufficient separation between elements that have dimensions of 0.99 cm $\times$ 1.33 cm.

The design goal for a microstrip array with the highest possible gain is mainly one of controlling the several sources of loss, which are surface wave loss, spurious radiation loss, dielectric loss, conductor loss, and connector loss. As the number of elements in the array increases, most of these losses increase as well. The usual competitor to an array for moderate- to high-gain applications is an aperture antenna like a parabolic reflector. The gain from (2-155) is directivity multiplied by the effect of all these losses, the radiation efficiency e_r, or $G = e_r D$. Here we are assuming that the array is uniformly illuminated so that the aperture efficiency ε_{ap} equals the radiation efficiency; see (9-79). A reflector antenna has unavoidable aperture taper and spillover losses, but radiation efficiency is nearly 100%. The aperture efficiency of a simple focus-fed parabolic does not depend on aperture size. As the radiating aperture size increases, the efficiency of an array decreases due to increasing dielectric and conductor losses, and will eventually fall below that for an equivalently sized reflector antenna. The array size for which the efficiencies of an array and a reflector are the same is about 256 elements (16 $\times$ 16). This was determined from a theoretical study supported by measurements. [11] Arrays similar to the one shown in Fig. 11-7 were constructed and measured for $N = 16$, 64, 256, and 1024

elements, and the efficiencies were 72%, 63%, 50%, and 34%, respectively. Comparing to a reflector antenna of the same aperture area and 50% aperture efficiency, a fixed-phase, uniformly illuminated array provides more gain for the same aperture size up to about 256 elements. The reflector antenna is a better solution for larger apertures. The gain of the array can be calculated using the maximum directivity from (8-113) as

$$G_{\text{Array}} = e_r D_{\text{Max Array}} = e_r 4\pi N \frac{d_x}{\lambda} \frac{d_y}{\lambda} \tag{11-14}$$

For the case of $d_x = d_y = d = 0.8\,\lambda$ as in Example 11-2 where uniform illumination is used, $G_{\text{Array}} = 8.04\,e_r N$. Using the efficiencies for the study, the gain values are $G_{\text{Array}} = 19.7$, 25.1, 30.0, and 34.5 dB for arrays of $N = 16$, 64, 256, and 1024 elements, respectively. The area of these arrays increases by a factor of 4 for each size increment, which corresponds to a directivity increase of 6 dB, but the increase in gain in going from 256 to 1024 elements is only 4.5 dB due to the loss build-up.

As would be expected, an array of microstrip elements presents bandwidth challenges. Increasing the substrate thickness to increase element bandwidth unfortunately leads to surface waves along the substrate that reduce the usable scan range. [12] Fixed-phase MSA arrays are not affected as much as phased arrays. The deleterious effects of the surface waves in phased arrays are rapid variation in the active impedance with scan angle, increased side lobes, and scan blindness moving toward broadside. Thus, there is tradeoff in scan range and bandwidth in microstrip arrays. When widely spaced operating bands are required, it is usually preferred to use a dual-band array rather than a broadband array. This approach also affords interference rejection of unwanted signals at frequencies between the bands. Dual-band arrays can be constructed using two layers of patches, each layer made of patches resonant at the desired frequency band. The lower layer contains the low-frequency patches and the upper layer contains the high-frequency patches with separate feed networks for each layer. A fixed-phase array operating at 2.44 and 5.78 GHz using a 4×4 array for the low band and a 8×8 array for the high band is described in [13].

Microstrip arrays are often required to be circularly polarized. One popular application is for Earth-space propagation where CP is used because of Faraday rotation experienced by linear polarization. Circularly polarized elements as in Fig. 11-5 can be used in array applications. An especially attractive way to achieve CP in a microstrip array that was mentioned in Sec. 8.8 is to use 2×2 subarrays of linearly polarized elements fed with sequential phases of 0°, 90°, 180°, and 270° clocked around the subarray. [H.6: Balanis, Ed., Sec. 4.5] This special configuration not only reduces the feed network complexity compared to dual-feeds to each element, but it also has wider bandwidth. Microstrip arrays are also used when dual polarization is called for. A dual-polarized microstrip array can be realized using orthogonal feeds in a patch and connecting the same polarization feed lines together, forming essentially orthogonally polarized arrays in the same aperture.

Microstrip antennas are also excellent candidates for omnidirectional arrays that are formed by wrapping the linear array around a vehicle, for example. [H.6: *Ant. Eng. Hdbk.*, 3rd ed., Sec. 7.3]

11.4   MICROSTRIP LEAKY WAVE ANTENNAS

The microstrip leaky wave antenna illustrated in Fig. 11-8*a* and 11-8*b* is an inexpensive low-profile antenna. It consists of a microstrip transmission line operating in the first higher-order mode, as we will see later. The objective of this section is to introduce the general characteristics of leaky wave antennas and then to illustrate those characteristics with a more detailed look at a microstrip leaky wave antenna.

11.4.1   Characteristics of Leaky Wave Antennas

A leaky wave antenna is a traveling wave antenna composed of a transmission line structure that leaks energy along its entire length. Transmission lines do not normally

radiate unless they are perturbed or unbalanced in some fashion. If a transmission line is perturbed such that it radiates a leaky wave, the wave will have a complex guided wave number, k_g, of the form $k_g = \beta_g - j\alpha_g$ where α_g is the attenuation constant due to leakage. Leaky wave antennas are designed such that at least 90% of the power on the structure has leaked away before the traveling wave reaches the end of the antenna. The remaining power can be absorbed by a matched load to prevent a strong minor lobe in the backward direction. When α_g is known, the 10% power at the end of the antenna, $P(L)$, farthest from the input can be calculated according to

$$\frac{P(L)}{P(0)} = (e^{-\alpha L})^2 = (e^{-2\alpha L}) = 0.1 \tag{11-15}$$

where $P(0)$ is the power at the input end of the antenna.

There are two types of leaky wave geometries: *uniform* along its length as are the geometries in Fig. 11-8; and *periodic* wherein the geometry has some periodicity along its length. Examples of the periodic type are shown in Fig. 11-9. Both types are fast wave structures. The periodic microstrip antenna in Fig. 11-9 is capable of backward as well as forward radiation [14].

The radiation pattern of leaky wave antennas is strongly dependent upon the length of the antenna (typically 5–20 wavelengths long) and its phase constant, β_g. Since the length is long and β_g changes with frequency, the main beam frequency scans within the forward range of about 10° to 80° from endfire. The beamwidth (and gain) normally remains relatively constant with scan angle. The pattern factor of a typical leaky wave antenna is that of a line source (see Chap. 5) with current of the form

$$I = I_o e^{-\alpha_g z} e^{-j\beta_g z} \tag{11-16}$$

If the effect of α_g is neglected, the position of the main lobe at θ_o is then

$$\theta_o \approx -\cos^{-1}\frac{\beta_g}{\beta} = -\cos^{-1}\frac{c}{v} \tag{11-17}$$
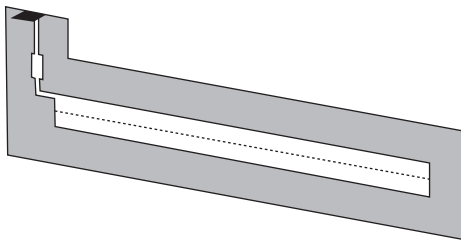
(*a*) With vias along the center line simulating a metallic
wall to prevent fundamental mode propagation.
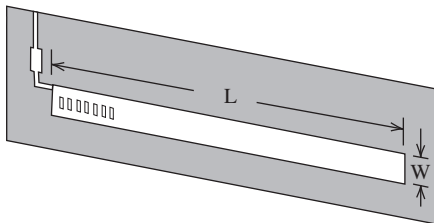
(*b*) With rectangular slots to prevent fundamental mode
propagation.

Figure 11-8 Uniform microstrip leaky
wave antennas.
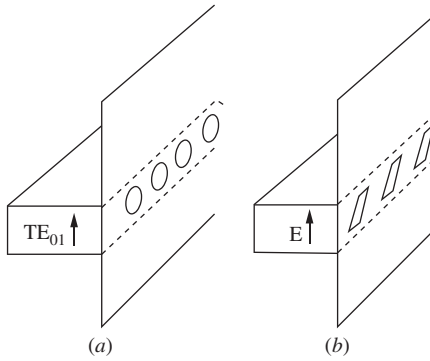
(*a*), (*b*) Metallic waveguide leaky wave antennas.
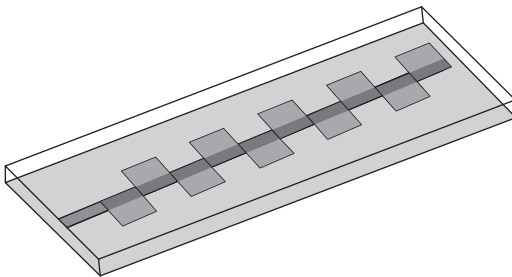
(*c*) Microstrip leaky wave antenna. [14]

Figure 11-9 Periodic leaky wave antennas.

as expressed in (5-35). When α_g is taken into account, there will be some small modifying effect on θ_o by α_g, but (11-17) will usually have satisfactory accuracy [H.8.8: Walter, p. 69].

The side lobe level of a leaky wave antenna tends to be high and can be higher than the -13.3 dB level of a uniform aperture. Most leaky wave antennas are easily impedance matched. Bandwidths typically are in the range of 15–30% but can be higher [15]. For example, the antennas in Fig. 11-8 have a pattern bandwidth of about 28% whereas that in [15] is more than 2:1. Leaky wave antennas have high excitation efficiency and low losses other than that due to radiation. If high power is not a requirement, printed circuit techniques can be used for fabrication as in the microstrip leaky wave antenna in Sec. 11.4.3.

11.4.2   Microstrip Modes

A microstrip transmission line normally does not radiate. The fields produced by the fundamental EH_0 mode do not decouple from the structure. It is only when the fundamental mode is blocked or restricted within the antenna that the microstrip transmission line can operate in a higher-order mode (specifically odd-numbered modes). It is in these odd-numbered higher-order modes that the fields are able to decouple from the surface and, thus, radiate from the structure. When this happens, the antenna is said to be operating as a leaky wave antenna.

Fig. 11-10*a* illustrates the fundamental mode and Fig. 11-10*b* shows the next higher-order mode. The fundamental mode for microstrip is a slow wave quasi-TEM mode, usually annotated, EH_0, in which both the electric and magnetic fields have a component in the direction of propagation. Examination of the electric field lines in Fig. 11-10*a* indicates that there will be no net radiation in the far field. Likewise, a higher-order mode in microstrip is not purely TE or TM, but a hybrid combination of the two. The first

E-field lines

(*a*) Field distribution associated with the fundamental microstrip mode, EH_0.

E-field lines

(*b*) Field distribution associated with the first higher-order mode, EH_1.
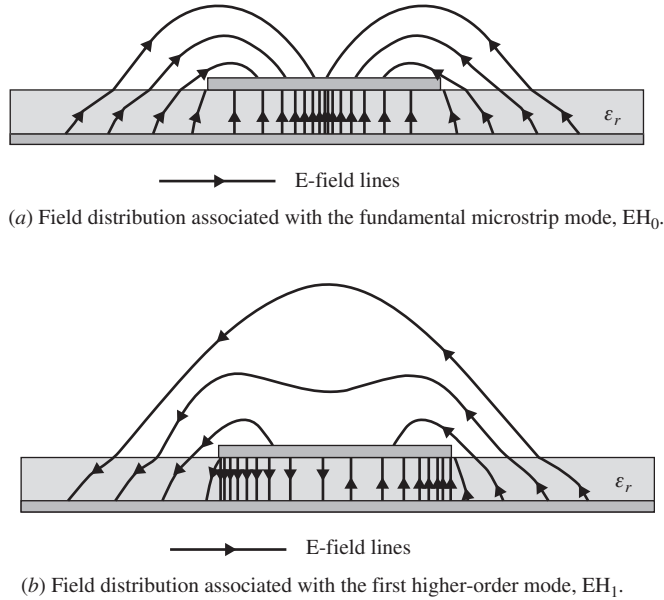
Figure 11-10 Two microstrip modes.

higher-order mode is termed the EH_1 mode shown in Fig. 11-10*b*. It has a phase reversal, or null, along the centerline, allowing the fields to decouple and radiate. EH_1 is a fast wave. Examination of the electric field lines in Fig. 11-10*b* indicates that this mode has the potential to produce far field radiation.

In Fig. 11-8*a* the dominant mode is prevented from existing by the metal wall of vias separating the left half of the transmission line from the right half. In Fig. 11-8*b* the dominant mode is prevented from existing by the rectangular slots in the microstrip line which interrupt the current of the EH_0 mode [H.6: Lo, p. 32–14].

The EH_1 mode on the microstrip transmission line is not a leaky wave mode. Instead, the EH_1 mode makes it possible for the microstrip structure to support a leaky wave. Leaky waves are not modes in the usual sense; in fact, leaky waves are non-modal. Sometimes leaky waves are called improper or nonspectral because the forward leaky wave increases in the *y*-direction vertically away from the guiding structure, thereby seemingly in violation of the condition that the radiation vanish at infinity. Fig. 11-11 illustrates this. In Fig. 11-11, the strength of the leaking field is depicted by the thickness of the rays at angle θ_o, the angle of the main beam. The fields increase exponentially to a distance above the antenna, $y_{\max}$, given by

$$y_{\max} = y \ \tan \theta_o \qquad (11\text{-}18)$$
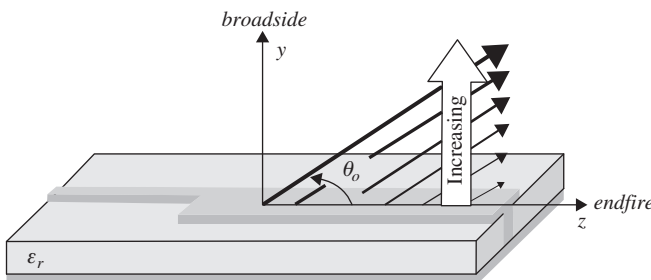
Figure 11-11 The intensity of fields radiating from a leaky wave antenna.

and then quickly decay, with the result that the fields do not continue to increase indefinitely.

Before the microstrip leaky wave antenna operation can be better understood, an understanding of the propagation characteristics of the antenna is required. Once the propagation characteristics are known, other quantities of interest can readily be obtained.

11.4.3   Propagation Regimes

The EH_1 mode exhibits four distinct propagation regimes. Below the cutoff frequency, f_c, is the reactive regime and above cutoff are the leaky wave, surface wave, and bound wave regimes, as illustrated in Fig. 11-12. These regimes can be understood by analyzing the general dispersion relation

$$k_g^2 = k_{gx}^2 + k_{gy}^2 + k_{gz}^2 \qquad (11\text{-}19)$$

where each component is complex (i.e., $k_{gx} = \beta_{gx} - j\alpha_{gx}$, $k_{gy} = \beta_{gy} - j\alpha_{gy}$, $k_{gz} = \beta_{gz} - j\alpha_{gz}$). For notational simplification, the subscript "g" will be dropped from the x, y, and z components.

There are numerous methods for finding the various propagation constants in (11-19). These methods include techniques such as FDTD (Chap. 15), MoM (Chap. 14), full-wave Green's function analysis, the method of Steepest Descent, mode matching, and an approximate somewhat simpler method, transverse resonance. Each method has different advantages and limitations for analyzing the microstrip leaky wave antenna. None of these methods yield propagation constants without a fair amount of effort. However, once the propagation constants are known, the design of a leaky wave antenna is a straightforward procedure involving, in part, (11-15). The main beam may be shaped by altering the geometry so that α_z and β_z are changed to meet design goals [H.6: *Ant. Eng. Hdbk.*, 3rd ed., Chap. 11] and the sidelobe level can be controlled by the choice of α_z.

Usually, the transverse propagation constants k_x and k_y must be found first and then the longitudinal propagation constant, k_z, found from (11-19). In the radiation regime, between the microstrip line and the ground plane,

$$k_z^2 = \sqrt{k_g^2 - k_x^2} \qquad (11\text{-}20)$$

since k_y is zero. In a method like FDTD which calculates fields, k_z can be found directly from a field component [16, 17] as indicated in Fig. 11-13. Note that

$$E_y \propto e^{-jkz} = e^{-(\alpha_z - j\beta_z)} \qquad (11\text{-}21)$$

and then

$$\ln E_y = \ln e^{-(\alpha_z z - j\beta_z z)} = -\alpha_z z - j\beta_z z \qquad (11\text{-}22)$$

Thus $\alpha_z z$ is simply the slope of the peaks in Fig. 11-13 and β_z is found from the separation of nulls since $\beta_z = 2\pi/\lambda_z$.

Once the longitudinal propagation constant is known, the far-field pattern can be determined by using the complex propagation constant in the appropriate line-source expression similar to (5-2) or (2-103).

The results that follow were obtained by FDTD (see Sec. 15.9) and also by Transverse Resonance [16, 17, 18]. Additionally, results in the radiation regime were verified by near-field measurements. The parameters of the antenna were length, $L = 220$ mm; width, $w = 15$ mm; $\varepsilon_r = 2.33$ (see also Fig. 15-34). The configuration used was that in Fig. 11-8a with the vias. This is called the half-width antenna because it takes advantage of the symmetry plane longitudinally down the middle of the microstrip such that only
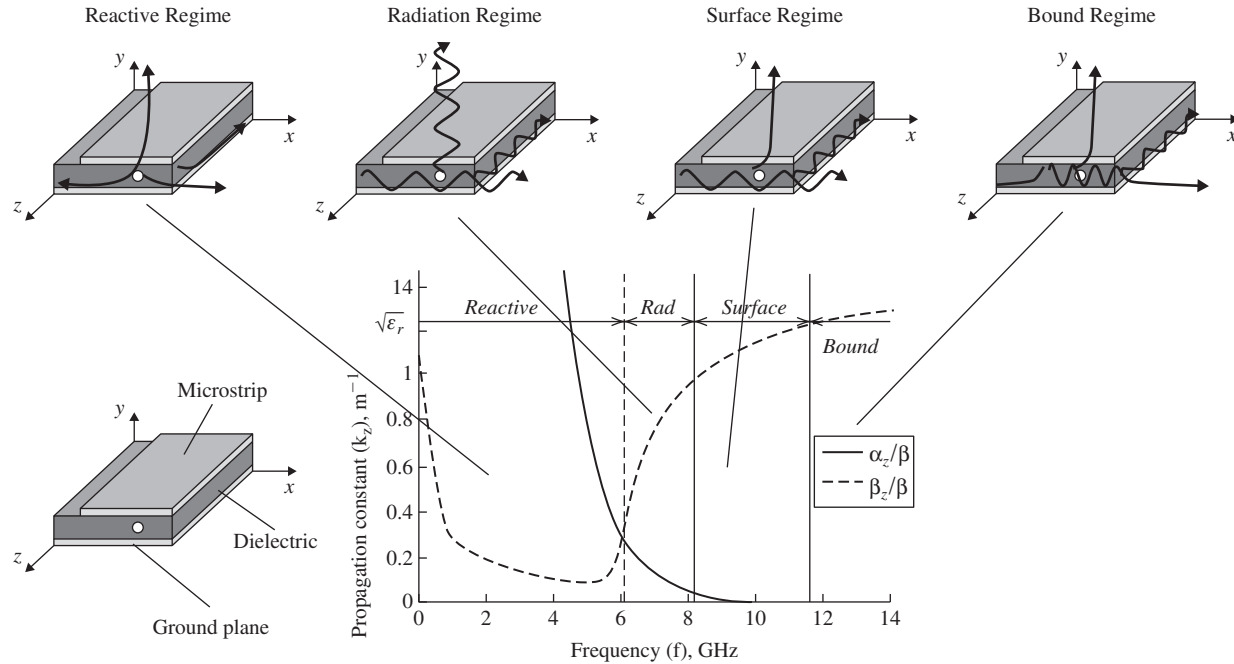
Figure 11-12 Normalized attenuation constant, $\dfrac{\alpha_z}{\beta}$, and phase constant, $\dfrac{\beta_z}{\beta}$, in the direction of propagation of the first higher-order mode, EH_1. (From [18]. Courtesy of the Air Force Institute of Technology.). The frequency plot is for the antenna in Fig. 15-34.
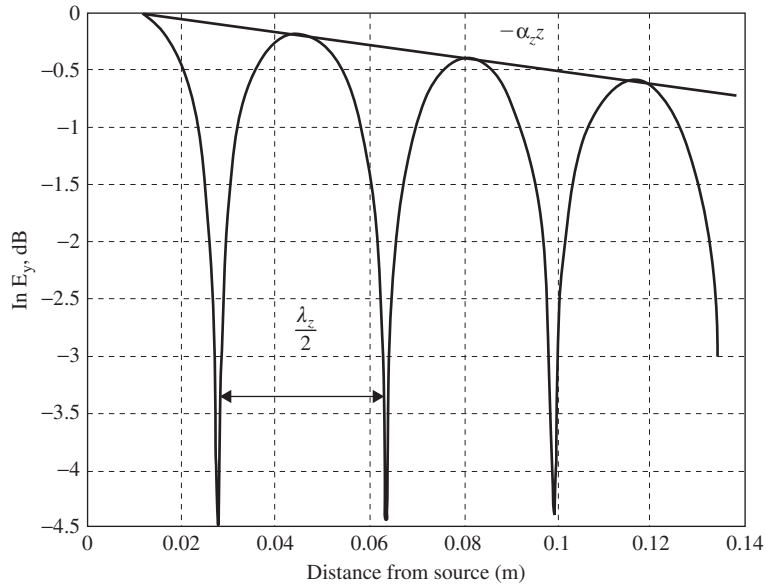
Figure 11-13 Natural log of FDTD y-directed electric field data (From [17]. Courtesy of the Air Force Institute of Technology.) for the antenna in Fig. 15-34.

half of the 15 mm width is excited and thus the other half can be eliminated. The results that follow can be obtained with either the half-width or full-width configurations.

Fig. 11-12 shows the plot of normalized (with respect to β) β_z and α_z versus frequency. From zero frequency up to the cutoff frequency, f_c, the region is called the *reactive* regime. It is characterized by a very large α_z component, causing the microstrip to appear as a reactive load at the input of the microstrip. At the cutoff frequency, the phase and attenuation constants are equal ($\beta_z = \alpha_z$). This is the start of the *radiation* regime. In this regime, propagation outside the antenna is occurring in all directions. The radiation regime continues until the phase constant is that of free space (β). At this point, there is no attenuation in the longitudinal direction ($\alpha_z = 0$) and large attenuation in the radiated direction (α_y). Because of this, fields no longer radiate. However, they continue to propagate in the longitudinal and transverse directions. This is called the *surface* regime. This regime continues until $\beta_z > \beta$. At this point, there is a large attenuation in the transverse direction (α_x) causing all fields to be bound under the top conductor and only propagate in the longitudinal direction. This is called the *bound* regime. Clearly the computation of the various propagation constants is crucial to understanding the operational regimes of the microstrip leaky wave antenna.

Fig. 11-14a shows an H-plane pattern at 6.25 GHz, and Fig. 11-14b shows an H-plane pattern at 7.2 GHz obtained using the FDTD and Transverse Resonance–derived complex longitudinal propagation constants as part of a line-source model to calculate the far-field pattern. Only the forward traveling wave is used for the pattern calculations, there being a matched load assumed at the end opposite the feed to absorb power reaching the far end. The operating band of the leaky wave antenna is from about 6 GHz to 8.2 GHz. At 6.25 GHz the main beam of the pattern is at about 61° and would rise higher if the frequency was decreased toward cutoff at 6 GHz. At 7.2 GHz the main beam is at about 37° and would approach endfire if the frequency was increased toward 8.2 GHz. The lack of deep nulls in the 6.25 GHz pattern is due to the value of α_z being higher at the lower frequencies.

In conclusion, α_z determines the length of the antenna and controls side lobe levels, β_z determines the main beam direction, and together α_z and β_z determine the radiation bandwidth. The relative permittivity influences α_z and β_z. The higher ε_r is, the larger is α_z and the smaller is the bandwidth. The most difficult part of microstrip leaky wave
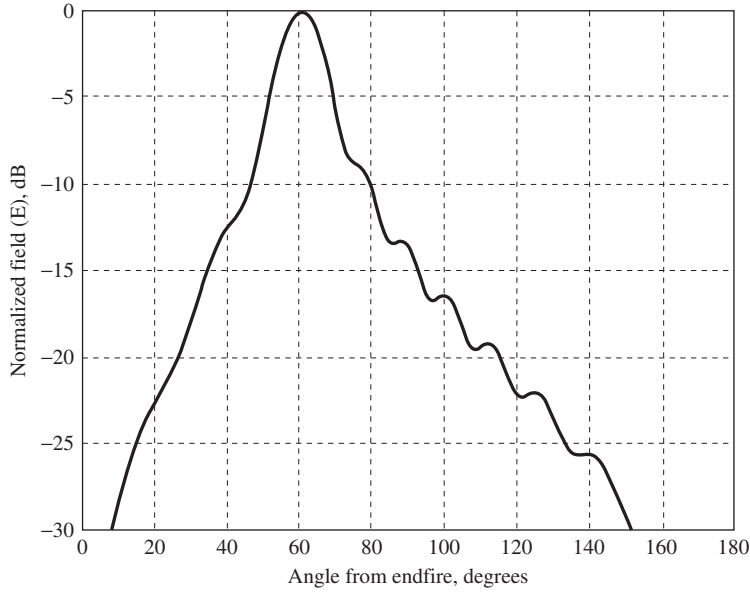
(*a*) Normalized far-field pattern at 6.25 GHz.
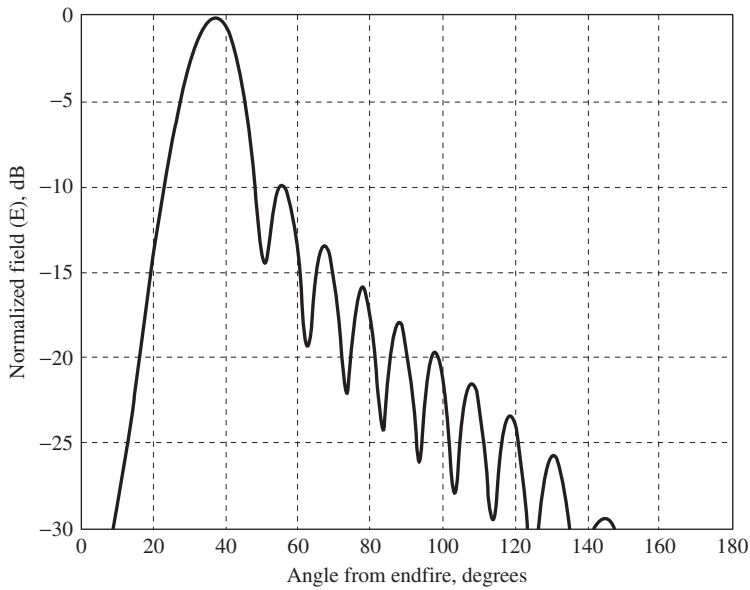
(*b*) Normalized far-field pattern at 7.2 GHz.

Figure 11-14 Normalized far-field patterns at two frequencies for the antenna in Fig. 15-34.

design is the determination of the complex propagation constants. However, once they are known, the design procedure is straightforward. A discussion of the FDTD simulation of the microstrip leaky wave antenna is in Sec. 15.9.3.

11.5 FUNDAMENTAL LIMITS ON ANTENNA SIZE

The topic of small antennas is of great interest in the wireless communications community because a primary goal in many applications is to have the antenna volume as
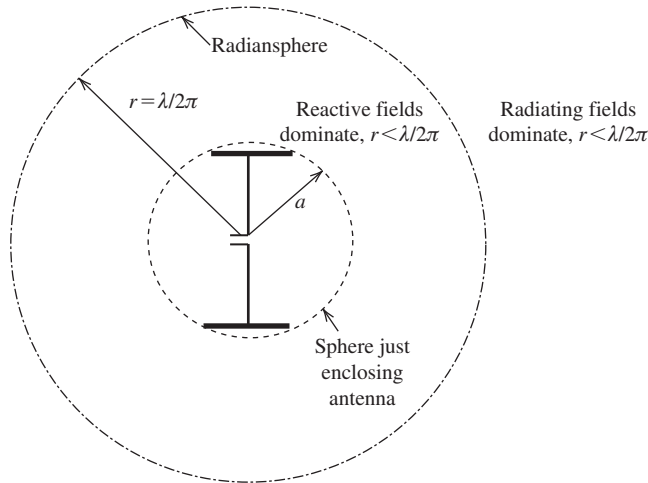
Figure 11-15 The radiansphere and the antenna sphere. The antenna sphere is of radius a and just encloses the antenna, which in this case is a capacitor-plate dipole antenna. The radiansphere is of radius r where $\beta r = 1$. If $\beta a < 1$, the antenna is electrically small.

small as possible. One of the most challenging areas is wireless handset design in which size reduction is limited by the size of two main components: batteries and antennas. Electronic devices in handsets have increased steadily in functional volume density for many years, whereas antenna size has not decreased significantly. This is because antennas (and batteries) do not follow Moore's law for integrated circuits, which predicts that component density on a chip increases by a factor of 2 every 18 months. Alternatively, Moore's law states that size can be reduced by one-half every 18 months for the same performance. This size reduction is not possible for antennas because they are directly limited by physics principles rather than manufacturing techniques as is the case with integrated circuits. In this section, we examine the theoretical limit on antenna size and how close it can be approached in practice. Also included are the consequences of reducing antenna size too much. Emphasis is primarily on electrically small antennas, but many results apply to resonant and broadband antennas as well.

In Sec. 1.4, we defined an electrically small antenna as one whose size is much less than the operating wavelength. The precise definition of electrically small begins by enclosing the antenna with an imaginary sphere of radius a as shown in Fig. 11-15. An electrically small antenna, as defined by Wheeler [19] and widely adopted, is one that is smaller than a *radiansphere*, which is an imaginary sphere of radius r with $\beta r = 1$.[2] So the radiansphere radius r is $\lambda/2\pi$, for a diameter of about 1/3 wavelength. Thus, an electrically small antenna (ESA) satisfies

$$\beta a \leq 1 \qquad \textit{electrically small antenna} \qquad (11\text{-}23)$$

The physical significance of the radiansphere is that it marks the boundary between the near-field and far-field regions for an ESA; Prob. 11.5-2 guides the proof of this. Inside the radiansphere, reactive stored energy dominates over radiated energy. Small antennas are usually mounted over a ground plane. If the ground plane is large, it is well approximated as an infinitely large conductor, and then the antenna is considered to include its image for determining the antenna sphere radius. For example, the monopole

---

[2] Wheeler [20] originally used diameter instead of radius, giving $\beta r \leq 0.5$ and some references use this. But (11-23) has become the most common criterion for defining the size limit.

version of the capacitor-plate dipole shown in Fig. 3-10*b* fits in a hemisphere of radius *a*, and the monopole and its image fit in the full sphere as in Fig. 11-15 (and in Fig. 3-3).

In Chaps. 2 and 3, we examined several simple radiators. Some of these were electrically small radiators (e.g., ideal dipole, short dipole, and small loop) smaller than resonant size. The consequence of electrically small size, as we saw in Example 2-4, is high input reactance, low radiation resistance, low efficiency, and narrow bandwidth. An electrically small antenna is characterized by a radiation resistance that is much less than its reactance and by a radiation pattern that is essentially independent of the antenna size. Although the radiation pattern and the directivity of an electrically small antenna are independent of size or frequency, the radiation resistance and especially the reactance are not. This is apparent from Figs. 6-5 and 6-6 for the dipole and Fig. 6-68 for the loop when their sizes are electrically small. This sensitivity of the input impedance to frequency makes the transfer of power from a generator into the antenna or from the antenna into a receiver a challenge for the antenna engineer. ESAs also can have low efficiency because of ohmic losses.

The key parameters for evaluating an ESA of size, bandwidth, and efficiency (gain) are highly interrelated and cannot be simultaneously optimized. That is, *antenna size is reduced at the expense of decreased bandwidth or decreased efficiency, or both*. An alternate way to view this is that there is an upper limit to gain-bandwidth product for an antenna, and maximum possible gain is increased only at the expense of bandwidth. [21] In spite of this barrier, the wireless industry continues to place great emphasis on reducing antenna size as much as possible without large compromise in performance. There are some system design approaches that allow for improved performance of ESAs, as we will note in this section, but these require an understanding of the fundamental limits on antenna size, which is our first topic.

11.5.1   The Fundamental Limit on Antenna Size

Antenna size limit theory started with Wheeler in 1947 [20] and remains an active area of investigation. There are three approaches used: field theory analysis, equivalent circuit models, and radiation pattern integration. Books on ESAs are found in Appendix H.8.11, with [H.8.11: Volakis et al.] providing the most comprehensive summary. In this section, we present an overview with emphasis on practical results and relating antenna size to its bandwidth.

As with any resonant electrical device, a basic performance parameter of an ESA is *antenna Q (quality factor)*, Q_A. The antenna is assumed to be resonant (tuned) to a resonant frequency and to be lossless (loss effects are accounted for in the radiation efficiency). Under these assumptions, antenna Q approximately equals radiation Q, which deals only with the radiation portion of power accepted by the antenna. Further assumptions in the theory to follow are that the antenna is passive, linear, and time-invariant. We start with Q because, unlike bandwidth, it can be rigorously defined. Following Q discussion, we present approximate relationships for bandwidth. Antenna Q is defined as the ratio of stored to radiated power:

$$Q_A = \frac{2\,\omega\,\max\{W_{e_{av}}, W_{m_{av}}\}}{P} \tag{11-24}$$

where $W_{e_{av}}$ and $W_{m_{av}}$ are the time-average, non-radiating (stored) electric and magnetic energies, and *P* is the time-average radiated power. At the resonant frequency, if there is one, the electric and magnetic energies are equal, just as in a resonant circuit. As done in circuit theory, it is logical to express Q in terms of input impedance as the ratio of reactive impedance, X_A, to real impedance, R_A:

$$Q_A = \frac{|X_A|}{R_A} \tag{11-25}$$

This is an approximation that is accurate for a small dipole with $\beta a < 0.5$ and for a small loop with $\beta a < 0.3$. [H.8.11: Volakis et al., p. 52] The Q of a monopole enclosed by a hemisphere is the same as its dipole counterpart because the reactive power in the numerator and the radiated power in the denominator of (11-24) are both halved, yielding the same ratio.

When handsets first found widespread use, they only operated in a single frequency band. Currently, the antenna system in a handset has to function in several bands, supporting many applications. Usually a single antenna in a handset must cover several bands. Communication engineers include each communication band in counting the number of bands, leading to the common designations for handsets of quadband (four bands), pentaband (five bands), hexaband (six bands), and so on. Some of these communication bands are contiguous in frequency. From an antenna operational standpoint, contiguous bands are one band, and only bands separated in frequency are counted in a multiband antenna. We use the latter definition, but the reader should be aware of the different ways to count bands. Unless otherwise noted, the theories here do not apply to multi-resonant or broadband antennas.

Q_A is approximately inversely related to fractional bandwidth, B, which was defined in (7-1a), giving [21]:

$$B \approx \frac{1}{Q_A} \qquad (11\text{-}26)$$

This is also a classical result for an RLC circuit, as Chu [21] modeled the small antenna. It is accurate for $Q_A > 10$. After completing the discussion of Q, we will be more specific in the bandwidth relationship.

In 1948, Chu [21] published his classic paper in which he expressed the fields outside the radiansphere as a summation of spherical waves. He then formed an equivalent RLC ladder network to separate the non-propagating energy from the total energy that led to an approximate expression for minimum radiation Q [23]:

$$Q_{\text{Chu}} = \frac{1 + 2(\beta\,a)^2}{(\beta\,a)^3 \left[1 + (\beta a)^2\right]} \qquad (11\text{-}27)$$

where a is the radius of the just-enclosing sphere centered on the antenna. Chu showed that an antenna exciting only the TM_{10} mode (equivalent to a small dipole) or only the TE_{10} mode (equivalent to a small loop) has the lowest possible Q for a linearly polarized antenna. Subsequent work by Collin and Rothschild [22] and McLean [23] based on full-wave analysis for the TM_{10} mode gave an exact expression for the minimum radiation Q:

$$Q_{\text{McLean}} = \frac{1}{\beta^3 a^3} + \frac{1}{\beta a} \qquad (11\text{-}28)$$

This result is very close to that of (11-27) for a smaller than the radiansphere, as in (11-23). Based on variation of energy velocity inside the radiation sphere, Davis et al. [24] corrected the derivation for the false assumption of previous investigators that radiated energy propagates radially outward at the speed of light, giving

$$Q_A = \frac{1}{\beta^3 a^3} \qquad \textit{minimum Q for all antennas} \qquad (11\text{-}29)$$

This equation yields values nearly identical to those from the Chu formula of (11-27) for $\beta a < 1$ and is only slightly lower for other βa values. As we discuss next, this has general application and is the lowest possible Q for any antenna. Enclosing shapes other than spherical (that are convex) leads to a smaller volume and higher minimum Q. [25]
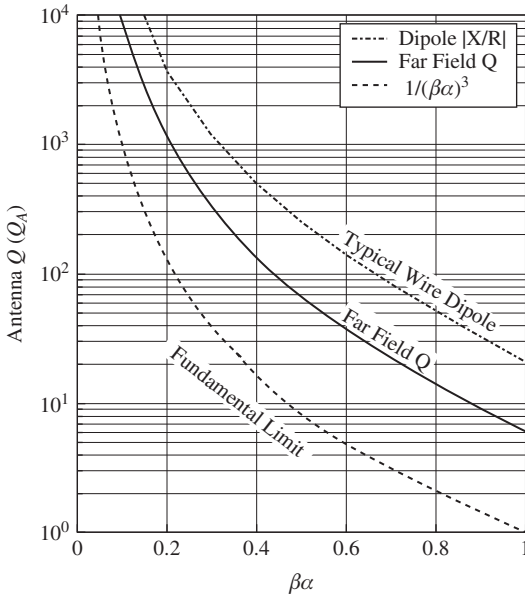
Figure 11-16 Antenna Q versus βa, where a is the radius of a sphere that just encloses the antenna. The bottom curve is the *minimum* Q_A for any antenna from (11-29). The middle curve is the *lower bound* for a dipole with a sinusoidal current distribution. The top curve is a typical result: a thin wire dipole using (11-25) and moment method simulations were used to find the input reactance and radiation resistance.

Antenna size limits are best understood by plotting Q versus antenna size. The lowest curve in Fig. 11-16 is a plot of the limit in (11-29). Increasing abscissa values, βa, represent increasing antenna electrical size, a/λ. Increasing ordinate values, Q_A, correspond to decreasing bandwidth because of the inverse relationship in (11-26). A point $(Q_A, \beta a)$ can be plotted on this graph for an antenna of size a operating at wavelength λ and having Q value Q_A using (11-38) below. The fundamental limit theory requires that the point cannot be below the curve. The curve also shows that smaller size (i.e., βa) is achieved by increasing Q_A (i.e., lowering bandwidth). Approaching this lower bound on Q has been elusive to achieve in practice. Usually it is not possible to merely lower the Q for an antenna while maintaining its size. The antenna engineer has to look for an inherently lower-Q design. Antennas that approach the limit on Q_A tend to efficiently utilize the spherical volume—that is, fill out the antenna sphere. A linear dipole, for example, fills very little of a spherical volume and is far above the fundamental limit curve. However, filling the volume does not guarantee reaching the lower limit.

11.5.2   Practical Aspects of Antenna Size Limits

The obvious question is how close an antenna can come to the minimum Q limit of (11-29). There is no general answer to this question, and we address it through examples for practical antennas. The middle curve in Fig. 11-16 is for a short dipole. It was derived by integration of the far-field pattern using (5-38) to obtain Q_A for the short dipole with a sinusoidal current distribution. The curve, as expected, is above the bottom curve, which is the lowest possible Q_A. The top curve is for a thin wire dipole based on the ratio $|X_A|/R_A$ in (11-25) and calculated using the method of moments for a dipole of radius 0.005λ. Although the top curve is for an electrically small dipole similar to that for the middle curve, the finite wire size and all other aspects of a real antenna are included. Summarizing Fig. 11-16, the lower curve is the minimum theoretically possible Q_A; the middle curve is the lowest Q_A for an electrically small linear dipole with a sinusoidal current distribution; and the top curve is actual Q_A for a real wire dipole that is electrically small.

It is interesting to note that the Q_A curve for an ideal dipole with uniform current varies as $1/(\beta a)^3$ for small a and the near fields (reactive) of an ideal dipole vary as $1/(\beta r)^3$. This implies that as an antenna becomes electrically smaller there is a sharp increase in the stored energy close to the antenna. This can be interpreted as radiating energy into the invisible region. Also as electrical size is decreased the directivity remains equal to 1.5. Thus, ESAs are superdirective, which means there is more directivity than would be expected based on size. Superdirectivity and Q_A are closely related, as explained in Sec. 5.5.

In addition to limited bandwidth, the electrically small antenna also suffers from low radiation efficiency. This is due to the radiation resistance decreasing faster than the ohmic loss resistance as the size decreases. From (2-172), the radiation resistance R_r of a short dipole varies as f^2 and from (3-54), the radiation resistance of a small loop varies as f^4. However, from (2-170), ohmic resistance R_o is constant with frequency. Thus, the radiation efficiency, given in (2-178) as $e_r = R_r/(R_r + R_o)$, of an ESA decreases rapidly as frequency is decreased. Radiation efficiency can be improved by reducing losses on the antenna by, for example, using planar strips instead of thin wires. Radiation resistance can be increased by filling more of the volume (inside a) with antenna structure. Various loading schemes and geometry perturbations can be used. [26] Increasing the radiation resistance also improves input impedance. For example, the high capacitive reactance of a short electric dipole can be reduced by inductive loading. An example of this is the normal mode helix, to be discussed in more detail in Sec. 11.6.1.

An electrically small dipole or loop antenna has a directivity of $D = 1.5$. Other types of ESAs will have directivities near this value. Gain is found from directivity as $G = e_r D$, and radiation efficiency e_r (referred to here as efficiency) will decrease with decreasing antenna size and can become very small. Thus, the attention with ESAs is focused on efficiency. Impedance mismatch depends on how the antenna is loaded and is accounted for by multiplying impedance mismatch factor q times gain to find realized gain; see Sec. 4.4.4. Here efficiency includes only internal antenna loss.

An insightful way to display efficiency versus frequency is to use a Bode plot to reveal asymptotic behavior. This requires expressing the efficiency in (2-178) in a form suitable for interpretation on a Bode diagram. Thus,

$$e_r = \frac{1}{1 + \dfrac{R_o}{R_r}} \tag{11-30}$$

For a short dipole, (2-171), (2-172), and (2-175) give

$$\frac{R_r}{R_o} = \left(\frac{\pi^{3/2}}{c^2 \sqrt{\varepsilon}} \right) \left(a \Delta L \sqrt{\sigma} \right) f^{3/2} \tag{11-31}$$

where c is the speed of light. Fig. 11-17 is the Bode diagram obtained by combining these two equations and plotting efficiency in dB form using $e_r(\text{dB}) = 10 \log(e_r)$ for the dipole case of the AM car radio antenna in Example 2-4 with $a = 0.159$ cm and $\Delta z = 1.57$ m. [26] When $R_o = R_r$, the radiation efficiency is 50% (-3 dB), and when $R_o \gg R_r$ the electrically small asymptotic behavior is R_r/R_o, which falls off with decreasing frequency at 15 dB/decade due to the $f^{3/2}$ frequency dependence when $e_r \ll 0.5$. The efficiency value for the car antenna of Example 2-4 at 1 MHz of 6.85% $= -11.6$ dB can be found on the curve in Fig. 11-17. The break point where the low- and high-frequency asymptotes meet, which is where $R_o = R_r$ and the efficiency is -3 dB, in Fig. 11-17 occurs at 5.68 MHz. The plot can be modified to display gain instead of efficiency by adding the directivity for the short dipole of $1.5 = 1.76$ dB.

The small loop antenna can be developed in a similar fashion. Using (2-171), (3-53), and (3-60) in (11-30) gives the asymptotic ratio:
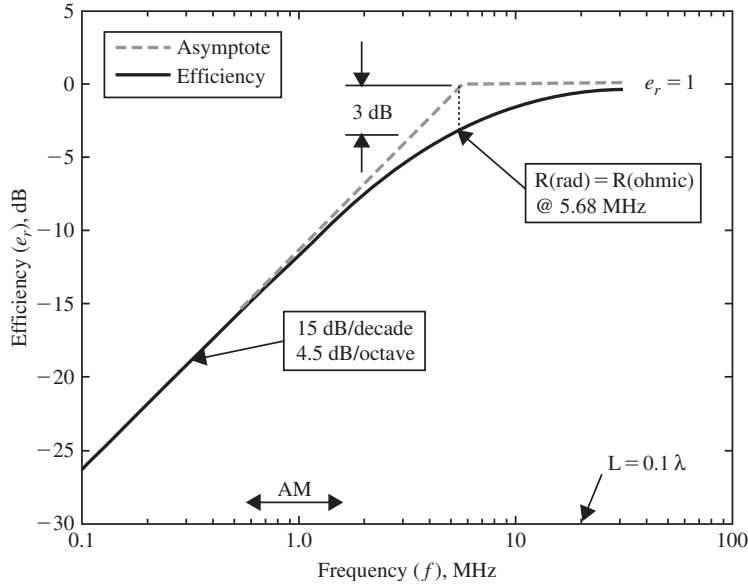
Figure 11-17 Efficiency in dB as a function of frequency for a short dipole antenna.

$$\frac{R_r}{R_o} = 31,200 \left(\frac{\pi^{3/2}}{c^4 \sqrt{\mu}}\right) (\sqrt{\sigma}) (ab^3) f^{7/2} \tag{11-32}$$

for a circular loop of mean radius b and wire of radius a and conductivity σ. This is the low-frequency asymptote for an electrically small loop antenna of wire radius a and mean loop radius b. The expression is valid up to about a loop perimeter of $0.3\,\lambda$. Consider a 1-meter circumference circular loop made of 0.5-inch diameter aluminum wire. Fig. 11-18 plots the efficiency in dB using (11-32) in (11-30) versus frequency on a logarithmic scale for this small loop example, including the Bode asymptotes. The break point of the asymptotes occurs at 38.2 MHz where the efficiency is -3 dB. The roll-off for low frequencies is at the rate of 35 dB/decade, or 10 dB/octave, for $e_r \ll 0.5$ due to the $f^{7/2}$ frequency dependence (11-32). An example including measured values for a multi-turn loop antenna is given in [27, 28].

The Bode asymptotic diagram can be used to extend measured data for gain or efficiency to frequencies outside the range of measurement. There are two scenarios, depending on whether the corner frequency is known or unknown. If the corner frequency is known and both asymptotes are also known, it is straightforward to either extend high-frequency data to low frequencies or vice versa. If the corner frequency is unknown, the situation is more difficult. Ideally at least one piece of low-frequency data allows the location of the low-frequency asymptote to be determined, and this allows the corner frequency to be fixed. If there is no low-frequency data, it may be possible to extend high-frequency data to low frequencies following the basic behavior of the Bode diagram of

$$\frac{1}{1 + \alpha x^{-\gamma}} \tag{11-33}$$

where x is a relative frequency, γ determines the slope of the asymptotic low-frequency dependence, and α is a constant that determines the location of the low-frequency asymptote. The low-frequency asymptotic frequency dependence is likely to be $f^{7/2}$ in
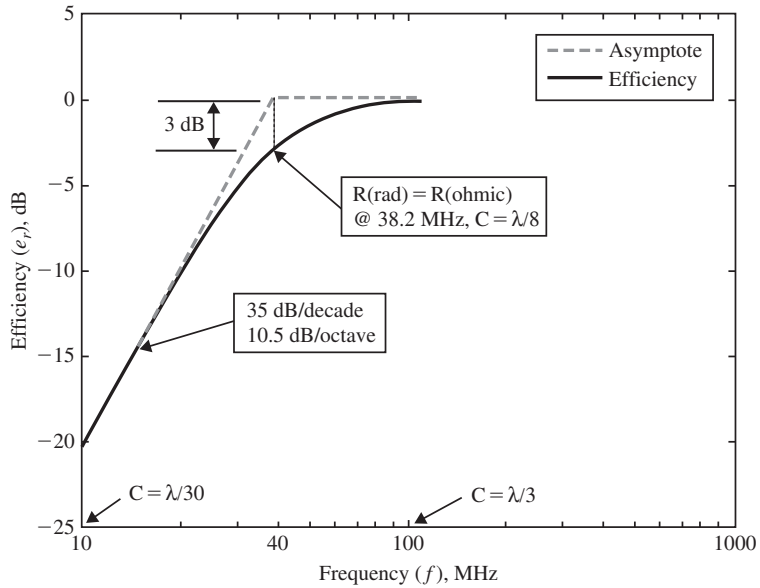
Figure 11-18 Efficiency in dB as a function of frequency for a one-turn, 1-meter diameter circular loop antenna made of 0.5-inch diameter aluminum wire.
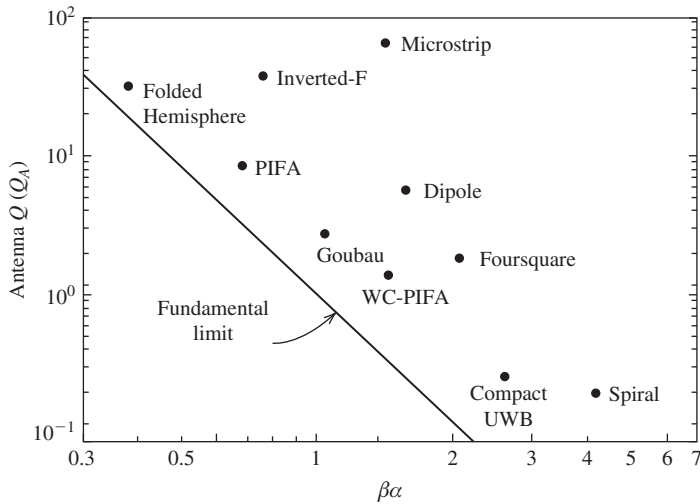
Figure 11-19 Several practical antenna examples. Each point $(Q_A, \beta a)$ is for the mid-band operating point of the antenna. The solid line is the minimum Q_A for any antenna from (11-29).

magnetic cases and may be $f^{3/2}$ in electric cases if the antenna behaves like a thin linear dipole or monopole.

   Fig. 11-19 shows the Q_A values for several practical antennas where each point $(Q_A, \beta a)$ is for the mid-band operating point of the antenna. Most antenna names on the plot are self-explanatory. The "Patch" is a microstrip patch antenna. It is far above the minimum Q_A curve because it has very narrow bandwidth. The "Dipole" is a half-wave dipole. The "Goubau" antenna is a specially fed top-loaded monopole that is 0.09λ tall

with an octave of bandwidth. Note that it is close to the minimum curve. [29] The "Foursquare" antenna is a broadband antenna that was discussed in Sec. 8.11. The "Folded Hemisphere" antenna is a small, folded spherical helix that efficiently uses the radian-sphere, getting very close to the minimum Q_A. [30] The "Compact UWB" antenna is shown in Fig. 7-44d and has a fractional bandwidth of 3.5. It is close to the limit, which is difficult to achieve for a broadband antenna. The folded hemisphere and compact UWB antennas are close to the minimum Q_A because they come close to filling the surrounding hemisphere. The remaining antennas will be discussed later.

11.5.3   Antenna Loading and Impedance Matching

Electrically small antennas have low radiation resistance and high reactance, capacitive for an electric antenna (such as a small dipole) and inductive for a magnetic antenna (such as a small loop). A matching network can always be found to cancel the reactance and increase the input resistance to a convenient value such as 50 Ω. The antenna is then tuned at the matched frequency. However, a matching network has disadvantages. First, adding reactance to the whole antenna system increases the Q, thus lowering the band-width. Second, loss in the matching network will lower the efficiency. The usual goal for realizing impedance matching, at least for ESAs, is to use internal loading to avoid the added volume and expense of an external matching network. When internal loading is used, the antenna is said to be *self resonant*. There are two forms of internal loading: using lumped circuit elements within the antenna or by changing antenna geometry. Lumped loads were discussed in Sec. 6.5. Increasing the length of the antenna conductors without making it taller provides a longer current path, which increases the reactance and radiation resistance. Top-hat loading, as with the monopole of Fig. 2-10b, accomplishes this. The normal mode helix antenna (NMHA) is very popular as a self-resonant small monopole. It is essentially a wire monopole coiled into a helix, giving it a longer current path compared to straight wire of the same height. The meander line monopole antenna shown in Fig. 11-21 is a low-profile version of the NMHA that is used in handsets.

Now we return to the evaluation of bandwidth that was approximated as $B \approx 1/Q_A$ in (11-26). An antenna operating band spans the limit frequencies from f_L to f_U. Antenna performance is acceptable between these frequencies, which for impedance-limited antennas is for a VSWR less than or equal to a specified value S. In the Q_A calculation, the frequency limits are where the power drops to half (-3 dB) peak power in the band, corresponding to $S = 5.83$. The bandwidth is the fractional bandwidth defined in (7-1a) as $B = (f_U - f_L)/f_C$, and in this application the center frequency is the geometric mean of $f_C = \sqrt{f_L f_U}$ rather than the normal arithmetic mean. Geometric mean agrees closely with arithmetic mean for low bandwidths and is 6% greater than arithmetic mean when B reaches 0.7. Geometric mean frequency is needed for accurate calculation of Q for broadband antennas. [H.8.3: Schantz, Sec. 5.5] Bandwidth is often quoted for a VSWR value of S that is different than the half-power level value of 5.83. To convert, the following is used [H6: *Ant. Eng. Hdbk.*, 4th ed., p. 6-13]:

$$B_{\text{VSWR}=S} = \frac{S-1}{2\sqrt{S}} B_{-3\,\text{dB}} \tag{11-34}$$

As a check, if VSWR $= 5.83$, $B_{-3\text{dB}} = B_{\text{VSWR}=5.83}$, as it should. A common bandwidth limit is for $S = 2$, and then $B_{-3\text{dB}} = 2\sqrt{2}B_{\text{VSWR}=2}$. Q_A is related to bandwidth, but Q_A is unloaded Q, meaning the antenna is assumed to be unaffected by any attached load. Then

$$Q_A = \frac{1}{B_{-3\text{dB, unloaded}}} \qquad \textit{small antenna} \tag{11-35}$$

where the unloaded bandwidth, $B_{-3\text{dB, unloaded}}$, must also be used. However, when measurements are made, the measuring instrument loads the antenna. In general, when a matching network is conjugate matched to the antenna, $Z_M = Z_A{}^* = R_A - jX_A$; see Sec. 4.4.4. When the load is matched to the antenna, the bandwidth is doubled, so the unloaded bandwidth is found as

$$B_{\text{unloaded}} = \frac{1}{2} B_{\text{loaded}} \qquad \textit{match loaded antenna} \qquad (11\text{-}36)$$

Combining this with (11-35) gives

$$Q_A = \frac{2}{B_{-3\text{dB, loaded}}} \qquad \textit{match loaded small antenna} \qquad (11\text{-}37)$$

Expressing this equation in terms of VSWR = 2 bandwidth using (11-34) with $S = 2$ gives

$$Q_A = \frac{1}{\sqrt{2}B_{\text{VSWR}=2,\ \text{loaded}}} \qquad \textit{match loaded small antenna} \qquad (11\text{-}38)$$

The foregoing development applies to tuned electrically small antennas and may be inaccurate for other applications. For broadband antennas, Q_A can be roughly approximated using

$$Q_A = \frac{1}{B_{\text{VSWR}=2,\ \text{loaded}}} \qquad \textit{broadband antenna} \qquad (11\text{-}39)$$

Ohmic loss effects can be included in the evaluation of Q rather simply by noting from (11-25) that Q_A is inversely proportional to dissipation and thus Q_A decreases with increasing loss. Then minimum Q expression of (11-29) is modified to include loss as follows:

$$\text{Q}_A = \frac{e_r}{\beta^3 a^3} \qquad \textit{minimum Q including loss} \qquad (11\text{-}40)$$

The effect is to lower the fundamental limit curve in Fig. 11-16. Of course, as efficiency decreases, Q_A decreases and thus bandwidth increases. Sometimes loss is intentionally added to an antenna to widen bandwidth. This result also shows that if an antenna has wide bandwidth, it may be due to loss. Matching network losses can be included in the efficiency too.

Loading antennas with materials is also used for size reduction. Materials act to lower the wave velocity $c = 1/\sqrt{\mu\varepsilon} = f\lambda$ which reduces the wavelength, making the antenna appear to be electrically larger. If the materials also satisfy $\mu_r = \varepsilon_r$, then the intrinsic impedance of the medium will remain equal to that of free space and $\eta = \sqrt{\mu_r\mu_o/\varepsilon_r\varepsilon_o} = \sqrt{\mu_o/\varepsilon_o}$, implying a good match to free space. As would be expected, there is a gap between this theory and reality. First, the equations are based on an infinite medium filled with material. Using finite volumes of loading material can reduce the electrical size, although size reduction will not be as high as for the infinite medium case with a size reduction factor of $F = \sqrt{\mu_r\varepsilon_r}$. The amount of size reduction depends on the size, shape, and material parameters of the loading materials. For example, a resonant quarter-wave monopole (7.6 cm tall) operating at 900 MHz can be reduced in height by a factor of $F = 4.6$ by embedding it in the center of a cylinder of dielectric with $\varepsilon_r = 90$, radius of 10 cm, and height equal to that of the monopole of 1.65 cm. [H.8.11: Fujimoto et al., Chap. 3] This is almost half of the infinite-medium reduction factor of $F = \sqrt{\varepsilon_r} = \sqrt{90} = 9.5$. This large height reduction is achieved with

some penalties in bandwidth and impedance. The bandwidth is reduced from 26% to 16% and input resistance at resonance is lowered from 39 Ω to 5 Ω. Hybrid magneto-dielectric loading materials provide the greatest size reduction for $\mu_r = \varepsilon_r$ but with a penalty in efficiency. It has been shown for a spherical helix filled with a magneto-dielectric core that the fundamental size limit can be approached. [31] The challenge with constructing material-loaded antennas is finding low-loss, affordable materials. Magnetic materials, such as ferrites, are more lossy than dielectrics and can become very lossy at RF frequencies.

Artificial materials are being developed for antenna loading in an attempt to avoid the loss problem and to create characteristics not found in nature. Popular artificial materials are *metamaterials* (also called *electromagnetic bandgap structures*). Metamaterials are constructed with small pieces of common material and arranged in a spatially ordered lattice. Interestingly, negative values of μ_r and ε_r are possible. An *artificial magnetic conductor* is well suited as a closely spaced ground plane to back electric antennas. A shell of negative permittivity material surrounding an ESA has been proposed as a way to approach the fundamental size limit, but there are dispersion and loss issues. [32] Double negative materials with both μ_r and ε_r negative have been studied, but even the possibility of building them is open to question.

Electrically small antennas are ideal candidates for using superconducting materials because of the high ohmic loss when antenna size is much less than a wavelength. A *superconductor* has close to zero DC resistance. There are some superconductors that are suitable for constructing antennas and matching networks, but they are often required to be cooled to the liquid nitrogen temperature of $77°$K. Usually the greatest benefit is in the matching network where high losses are reduced greatly with superconductors. However, bandwidth is often sacrificed. Superconductor antenna systems are costly and usually only used for special critical applications. See [H.8.11: Hansen, Chap. 3] for more details.

The prior discussion assumed a linear, time-invariant medium. It is possible to miniaturize an antenna below the fundamental limit using nonlinear, time-variant techniques. One uses *impedance modulation* where an antenna's tuning is time varied. The resonant frequency is changed periodically, creating the effect of two resonant frequencies that can be used simultaneously. The method has been demonstrated using a varactor diode in the antenna. [33] However, spurious radiation and phase nonlinearity are issues.

Recently it has been shown that the fundamental limit of (11-29) applies to all antennas including broadband antennas such as ultra-wideband antennas. [34] As already mentioned, the bandwidth and Q relationships for small antennas are not accurate for broadband antennas. The inaccuracy of bandwidth formulas has been shown to be the case for doubly resonant antennas. [35] Broadband antennas can be size-reduced using similar techniques as for ESAs such as efficient volume filling and loading; however, the challenge is even greater. Examples of efficient volume filling antennas are the folded hemisphere antenna [30] and the compact UWB antenna [36] shown in Fig. 7-44*d*. As shown in Fig. 11-19, the compact UWB antenna is a broadband antenna ($B_{VSWR = 2} = 3.5$) that is close to fundamental size limit, which is difficult to accomplish in practice. Size reduction of frequency independent antennas such as spirals using continuous loading can be implemented in the following ways: coiling the arms similar to the meander monopole or zigzagging the arms (see Sec. 7.7.4); dielectric loading; and resistive loading. [H.8.11: Volakis et al., Chap. 5]

11.6   ANTENNAS FOR COMPACT DEVICES

Compact antennas are required in a wide variety of wireless devices, ranging from handsets for cellular telephone, to modules for asset tracking, to implantable medical devices. The evolution in handset antennas has been dramatic. When public cellular radio started, the antenna protruded from the handset and was an "eye-poker," similar to the antenna shown in Fig. 11-20*b*. Gradually the external antenna size was reduced,

eventually being replaced by an internal antenna. So how has it been possible for the handset antenna to shrink from a long antenna protruding from the handset to an invisible internal antenna? The preceding section explained how antenna size is limited by fundamental physics principles. As explained in Sec. 4.5 and shown in Table 4-1, compact internal antennas were enabled by reducing the cell size, replacing base station omni-directional antennas with more directive sector antennas, and making other system improvements. In this section, we consider several types of compact antennas. Many of these antennas operate as monopoles and require a ground plane.

11.6.1   Normal Mode Helix Type Antennas

The theory for the normal mode helix antenna (NMHA) was presented in Sec. 7.3.1. In this section, we examine practical realizations of the NMHA. The NMHA is essentially a wire monopole coiled into a helix as shown in Fig. 7-9. The total wire length for N turns is

$$L_w = NL = N\sqrt{C^2 + S^2}$$

where C is the circumference of the helix and S is the spacing between turns. This gives greater wire length compared to a straight wire of the same axial length (height, h), which is desired because of the increase in radiation resistance. The *length increase ratio* is defined as

$$R = \frac{L_w}{h} = \frac{N\sqrt{C^2 + S^2}}{h} \qquad \textit{NMHA length increase ratio} \qquad (11\text{-}41)$$

For a helix of fixed height, increasing the number of turns by decreasing the spacing between turns, $N = h/S$, increases length L_w. Using a fatter helix (increasing the diameter, $D = C/\pi$) also increases the wire length.

   An example compact NMHA operating at 150 MHz in the VHF band is 10 cm tall and 2 cm in diameter and has 13.25 turns. The total wire length of $L_w = N\sqrt{C^2 + S^2} =$ 83.9 cm $= 0.42\lambda$. Simulations with a moment method code give for copper wire $Z_A = 4.9 - j0.7$ Ohms, indicating resonance (see Prob. 11.6-1). The value of radiation resistance predicted by (7-20) with $h = 0.05\,\lambda$ is 1.6 Ω, which is close to the 2 Ω based on simulation with perfectly conducting wire that gives $Z_A = 2 - j3.5\,\Omega$. The length increase ratio is $R = L_w/h = 83.9/10 = 8.4$. These results are in close agreement to the simulations of Best. [H.6: Balanis, Ed., p. 488]

   Example 7-1 gave a NMHA operating at 883 MHz ($\lambda = 34$ cm) with a quarter-wave wire length of $L_w = 8.5$ cm. The βa value is about 1, so the antenna is electrically small. The height is $h = 5.7$ cm $= 0.17\lambda$, giving a length increase ratio of $R = 1.49$. The radiation resistance from (7-20) is $R_r = 11\,\Omega$. As we saw in the previous example, larger length increase ratios are used in the lower frequency bands such as VHF to avoid large physical heights, but as the electrical height of the antenna is reduced, the radiation resistance is lowered. A commercial version of this handset-NMHA antenna, called a **stubby antenna**, is shown in Fig. 11-20a. Generally, if the total wire length, L_w, is less than a quarter-wavelength it is called a stubby. If the wire length is approximately a quarter-wavelength to achieve resonance, it is called a quarter-wave helix, or simply a quarter-wave. The particular implementation of Fig. 11-20 operates in one of two antenna modes, stubby antenna only or whip antenna only. In the stubby antenna mode of Fig. 11-20a, the extendable section is stowed and has a non-metallic portion at its tip that is inside the helix so as not to affect stubby operation. The stubby has a feed connection inside the handset attached to the lower end of the helix. The stubby antenna is also used without the whip antenna and is mounted either external or internal to the handset. When mounted externally it is molded in polycarbonate or polyurethane. When the extendable wire is deployed as in Fig. 11-20b, the stubby is used to feed the whip antenna, either by direct
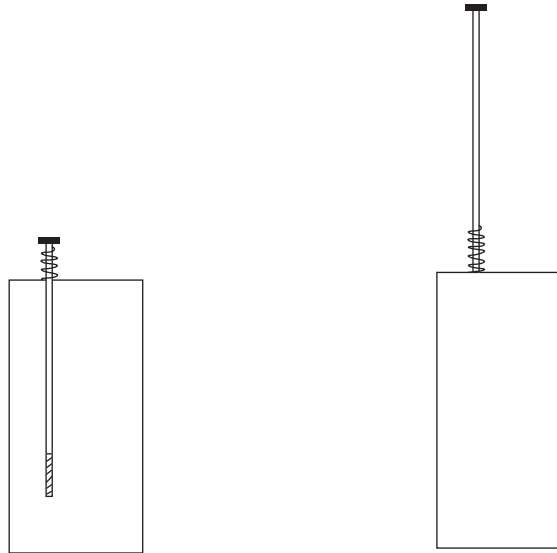
(*a*) Stowed configuration.
Only the stubby antenna is operational.

(*b*) Deployed configuration. The stubby
antenna is used to feed the whip antenna.

Figure 11-20 The stubby antenna (NMHA) in a wireless handset. The NMHA is used alone as in (*a*) or with an extendable wire antenna ("whip") as in (*b*). When extended, the stubby is used to feed the whip antenna.

contact or through capacitive coupling. Operation is then as a quarter-wave straight wire antenna. Both antennas are monopoles and are grounded to the PCB in the handset. The human operator also can act as a ground. Ground plane independence is possible if the extendable section is an end-fed half-wave dipole, but at 850 MHz it is rather long at 17.5 cm (7 in). The whip antenna mode provides gain of around 2 dB and the stubby antenna gain is less than 0 dB. The extendable antenna has largely disappeared from cellular handsets. Early cordless telephones used NMHAs that also have been replaced by internal antennas. The NMHA can provide dual-band operation in separated frequency bands.

If ribbon conductors are used for an NMHA in place of wires and at the same time it is flattened, a planar version results that is easily manufactured by printing or stamping. The **meander line antenna** (MLA) is a low-profile version of the NMHA that can be configured as either a dipole or a monopole. Meander dipoles are sometimes used as in-glass antennas on vehicles. The meander monopole in Fig. 11-21 has 7 vertical sections and 6 horizontal sections. If the section lengths are $s_V = 1.43$ cm and $s_H = 3.33$ cm, the total wire length is $L_w = 7 \cdot 1.43 + 6 \cdot 3.33 = 10 + 20 = 30$ cm. Best [H.6: Balanis, Ed., Sec. 10.6] simulated this monopole over a perfect ground plane and found the resonant frequency to be 361 MHz ($\lambda = 83$ cm), giving an electrical height of 0.12λ and wire length of 0.361λ for a length increase ratio from (11-41) of $R = 0.361/0.12 = 3$, which is a significant increase. The meander monopole is used in handsets because of its compact geometry, but because it does not fill much of the enclosing sphere, it has lower radiation resistance, efficiency, and bandwidth than a NMHA of the same height. As in the NHMA, the current along the wire of the meander monopole reduces from a peak at the feed to zero at the top end. Due to its small electrical size, the phase is nearly constant along the length. Because of this, the adjacent horizontal sections, which have oppositely directed currents as shown in Fig. 11-21, lead to canceled radiated fields. Thus, the vertical sections are responsible for the radiation just as in a straight wire monopole and control the radiation resistance. The adjacent horizontal sections add inductance and loss. Thus, the pattern of the meander antenna is similar to a straight monopole of the same height with a null in the vertical
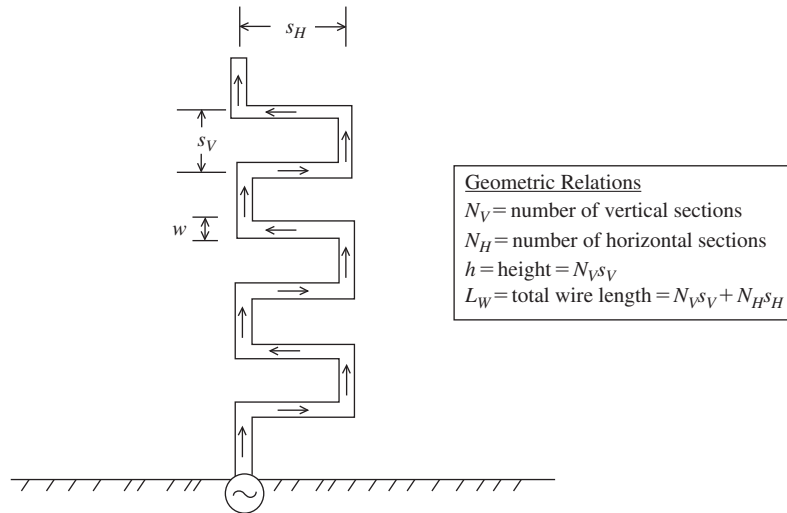
Figure 11-21 The meander line monopole antenna. The arrows indicate current direction.

direction, a half-donut 3D shape, and linearly polarized parallel to the long axis. The NMHA behaves in the same fashion. Comparisons of NHMA to meander monopoles of the same height and conductor length adjusted to give the same resonant frequency show that the NMHA has higher radiation resistance, wider bandwidth, and higher efficiency. [H.6: Balanis, Ed., Sec. 10.6] In spite of these shortcomings, the meander monopole is a popular antenna because it is compact and low-cost. Also found was that decreasing wire diameter lowers the resonant frequency, but this will lead to more loss.

Calculations with meander monopoles over finite ground planes in [H.6: Godara, Ed., Sec. 12.2] provide some general trends. First, the meander monopole has multiple resonances and thus is useful in dual-band handsets. Second, as ground plane size is reduced from infinite to fractional wavelength sides, the impedance decreases and the resonant frequency increases. The ground plane (PCB) of a typical handset is 100 mm × 50 mm which at 850 MHz is $0.28\,\lambda \times 0.14\,\lambda$ and the PCB itself participates in the radiation, leading to low directivity. At 2 GHz the PCB is larger electrically ($0.67\,\lambda \times 0.3\,\lambda$) and the current is more localized to the antenna, giving a pattern that is more like the infinite ground plane case. As a consequence the gain will be higher at 2 GHz compared to 850 MHz.

11.6.2   Quadrifilar Antennas

In the family of compact helix antennas is the **quadrifilar helix antenna** (QHA), also called the *quadrifilar antenna*, shown in Fig. 11-22. Quadrifilar antennas provide a circularly polarized unidirectional beam without need for a ground plane. The quadrifilar helix antenna is a popular choice for satellite communications both for the spacecraft and the ground terminal. The four wire helical arms are spaced equally around a circular cylinder surface (i.e., 90° apart) and are fed at the top with equal amplitudes and sequentially 90° out of phase. One way to construct a phasing network is with one 180° hybrid and two 90° hybrid dividers. The QHA can be thought of as two orthogonal bifilar helical loops. If the phase progression from turn to turn matches the sense of the circular polarization, the wave is forward fire. If not, a backfire mode results. Unlike the axial mode helix, circular polarization is not dependent on the circumference being about a wavelength, but instead can be much smaller because the CP is generated by the feed phases to the arms, which is a significant advantage. Both high-gain and broad shaped-beam patterns
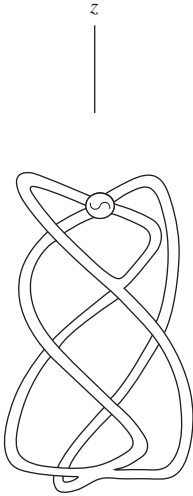
Figure 11-22 The quadrifilar helix antenna.

are possible. Quadrifilar antennas are capable of moderately high gain by using several turns and a long axial length. Generally, large pitch angles, few turns, and short lengths are used for broad beams. Being ground-plane independent because of its balanced feed arrangement, the QHA is not greatly affected by nearby objects such as a human.

When the bifilar loops are a fractional turn (1/4 to 1) the quadrifilar is called a *volute antenna*, and it is resonant when each arm is an integer multiple of a quarter wavelength. The ends of the volute are opened (shorted) at the base if the helical arms are an odd (even) multiple of a quarter-wavelength long. The volute antenna can produce a broad beam with a cardioid-shape pattern off its top end ($+z$-axis) with low back lobes and good circular polarization, but with narrow bandwidth of a few percent. [H.6: *Ant. Eng. Hdbk*, 4th ed., Sec. 12.5] The QHA in Fig. 11-22 has left-hand wound half-turn arms, but produces right-hand CP due to the backfire nature of the antenna. The beam peak is in the z-direction. An example QHA of Prob. 11.6-4 has half-turn arms, a diameter of $D = 0.174\lambda$ and height of $h = 0.243\lambda$. Using the helix geometry relations in Sec. 7.3, we find $S = 0.486\lambda$, $\alpha = 41.6°$, and $L = 0.732\lambda$. The wire length of one loop is the length of one turn (two half-turns on each side) plus the end lengths, so $L_{loop} = L + 2D = 1.08\lambda$. Thus, each arm is about a half-wavelength long. Simulations show that the endfire, unidirectional beam has a half-power beamwidth, HP, of 120° and the bandwidth is 3.2%. [H.3: Milligan, 2nd ed., p. 266] See [H.9.1: Kumar, Chap. 5] for derivations of QHA radiation properties and design data.

The QHA becomes a compact antenna suitable for use with mobile and portable terminals through size reduction techniques such as meandering or folding of the arms, or by loading with high dielectric constant ceramic material. It is popular for GPS reception where narrow bandwidth is acceptable. A typical GPS QHA printed onto a ceramic dielectric core with $\varepsilon_r = 40$ is 18-mm tall and 10 mm in diameter with a gain of -2dBic at 1.575 GHz. [H.6: *Ant. Eng. Hdbk.*, 4th ed., p. 39–26] In the GPS application, it is important to have a phase center that is stationary over the pattern main beam and bandwidth, which a well-designed QHA provides.

11.6.3   Planar Inverted-F Type Antennas

The quarter-wave monopole is an excellent antenna with the desirable characteristics of 5.16 dB directivity, omnidirectional coverage, and $35 + j0\,\Omega$ input impedance. However, at frequencies in the UHF band and below, the antenna often becomes unacceptably tall. The simple solution is to bend the vertical wire over, forming an upside-down L shape.
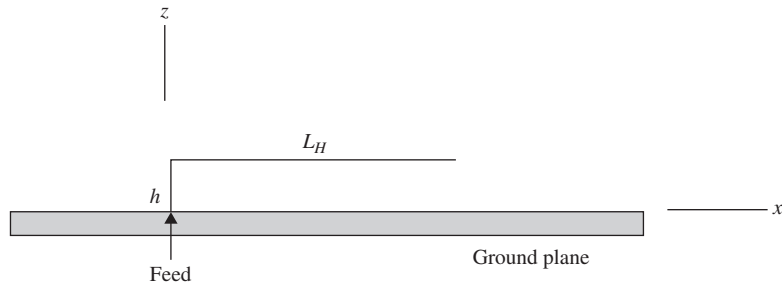
Figure 11-23 The inverted-L antenna (ILA). The arrow indicates the feed point.

This **inverted-L antenna** (ILA) we introduced as the transmission line monopole in Fig. 3-10*c*. The ILA finds many applications where a low-profile antenna is needed and narrow bandwidth is acceptable. It also forms the starting point for the evolution of the planar inverted-F antenna.

The image current in the ground plane (if large) of the ILA has a horizontal component that is directed opposite to the current on the horizontal wire, and the vertical image current is in the same direction as the current on the vertical wire. The current distribution is approximately sinusoidal over the whole wire and is zero at the open end. The distance above the ground plane, h, in Fig. 11-23 is much smaller than a wavelength, so contributions from the horizontal currents of the wire and its image remain nearly out of phase in the far field, leading to substantial cancellation. On the other hand, the vertical wire forms an effective single vertical radiating element with nearly uniform current of length $2h$. Thus, the vertical wire controls the radiation resistance and the total wire length of $L_w = h + L_H$ influences the input impedance. The ILA resonates when $L_w \approx \lambda/4$, as would be expected. The pattern is a half-donut shape in three dimensions similar to a quarter-wave monopole, except along the z-axis there is dip rather than a perfect null. Due to the ground plane effects, the radiation resistance is lower than a quarter-wave monopole. A straight quarter-wave monopole has a radiation resistance of about 36 Ω and the quarter-wave ILA is about 8 Ω, depending on wire details. When the ILA is size-reduced to form an electrically small ILA, the radiation resistance falls even more and the reactance is highly capacitive. The radiation resistance for a small ILA can be approximated with the following formula derived in [37]:

$$R_r = 160\pi^2 \left(\frac{h}{\lambda}\right)^2 \left[1 - \frac{h}{2(h + L_H)}\right]^2 \quad ILA, \;\; h \ll \lambda \qquad (11\text{-}42)$$

EXAMPLE 11-3   *Garage Door Opener Transmitter ILA Antenna*

Many garage door opener remote control units transmit on a frequency of 315 MHz ($\lambda = 95.2$ cm). A commonly used antenna for this application is a small ILA that has $h = 1.2$ cm and $L_H = 6.7$ cm, for a total wire length of $L_w = 7.9$ cm $= 0.08\,\lambda$, making it electrically small. Simulations for perfect conducting wire of 1-mm diameter give $Z_A = 0.21 - j364$ Ω and (11-42) confirms this radiation resistance value, which is very low. Simulations with copper wire give a total resistance of 0.25 Ω, which means the efficiency is $e_r = 0.21/0.25 = 84\%$. For comparison, if the same wire were a straight monopole, simulations show that $Z_A = 2.7 - j415$ Ω and near 100% efficiency. A resonated quarter-wave monopole would have a good input impedance of $Z_A = 35 + j0$ Ω but would be 22.7 cm (9 in) tall and not suitable for the application. The receiving antenna in the garage is usually a piece of wire that is close to a quarter-wavelength in length.
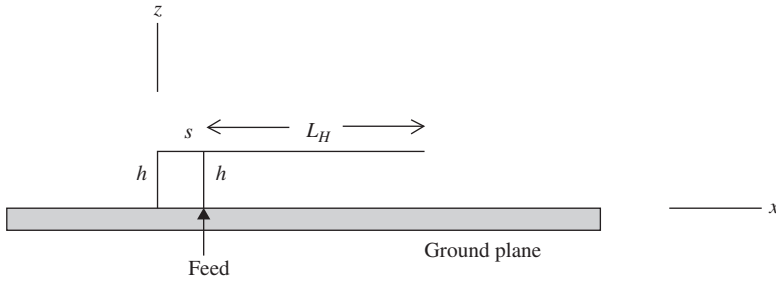
Figure 11-24 The inverted-F antenna (IFA). The arrow indicates the feed point.

The input impedance of the ILA can be altered by simply adding two wires to form a closed conducting loop as shown in Fig. 11-24. The second vertical wire is grounded and of the same height (h) as the driven wire and is separated from it by distance s. This antenna, called an **inverted-F antenna** (IFA), adds only a small amount of length over the ILA but provides more design parameters and wider bandwidth. The IFA is resonated just like for the ILA and the length $h + L_H$ should be about a quarter-wavelength. The currents on the vertical wires are in phase, leading to reinforced radiation and higher radiation resistance. The loop can be thought of as adding inductance to partially cancel the capacitance of the horizontal wire segment. Another way to think of the added section is that it behaves like the gamma match shown in Fig. 6-35c. The pattern of the IFA is similar to the ILA. When an IFA is mounted at the edge of a device-size ground plane such as a handset PCB, the pattern becomes similar to a dipole pattern with nulls near the long axis of the ground plane and close to omnidirectional around the long axis. [H.9.1: Sanchez-Hernandez, Ed., p. 29]

EXAMPLE 11-4   *Inverted-F Antenna for the 2.45 GHz ISM Band*

The unlicensed ISM band at 2.45 GHz spans 2.4 to 2.5 GHz for a bandwidth of 4%, which can be covered with an IFA. Simulation trials using wire 1 mm in diameter lead to a resonance for $h = 9.8$ mm, $L_H = 21.8$ mm, and $s = 4$ mm. This size is compact and suitable for many applications. The total wire length is $h + s + L_H = 35.6$ cm $= 0.29\,\lambda$, which is close to a quarter-wavelength. The input resistance is 28 Ω (real). The ILA form of this antenna using the same h and L_H values has $Z_A = 12 + j11$ Ω, showing that the added section on the IFA increases the radiation resistance significantly.

The resonated ILA has narrow bandwidth (a few percent) and low input resistance, which are increased with the IFA. Further increase in bandwidth is possible by adding a second, parasitic ILA wire parallel to an IFA, forming a *dual inverted-F antenna* (DIFA) that is capable of 14% bandwidth. [38]

Employing the technique used many times for increasing bandwidth, the bandwidth of an IFA is increased by replacing the wires with metallic plates forming the **planar inverted-F antenna** (PIFA) shown in Fig. 11-25. The basic PIFA consists of a ground plane and an upper plate fed by a coaxial type feed line with the center conductor passed through the ground plane and connected to the top plate while the outer conductor is connected to the ground plane. There is also a shorting plate that shorts the top plate to the ground plane. In addition to the explanation of the PIFA evolving from an IFA by using wide conductors, the PIFA can also be thought of as a quarter-wave microstrip patch, which was discussed in Sec. 11.2.2. As the shorting plate width, w, of the quarter-wave patch is reduced the resonant frequency lowers, making the antenna more compact. This effect is explained by the capacitance of the top plate being reduced by the inductance of
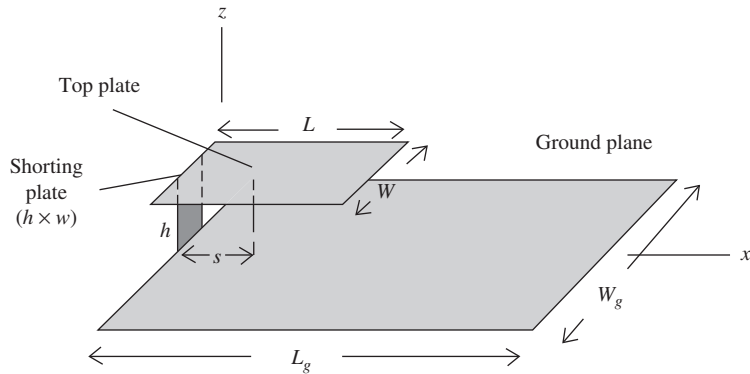
Figure 11-25 The planar inverted-F antenna (PIFA). The arrow ($\uparrow$) indicates the feed point.

the shorting plate that increases with decreasing width. The greatest frequency reduction occurs for placement of the shorting plate (or pin) in the corner of the top plate to provide a longer current path compared to placement at the center of the edge. The PIFA forms an effective quarter-wave monopole at first resonance when the longest current path is about a quarter-wavelength. The lowest resonant frequency configuration is for a shorting pin located at the corner, giving a current path of one side plus one end length totaling a quarter-wavelength [H.9.1: Waterhouse, p. 199]:

$$\lambda_r = 4(L + W) \quad PIFA \ resonant \ wavelength \tag{11-43}$$

This makes the PIFA about half the size of a microstrip patch, which has $\lambda_r = 4L$. The input impedance is controlled by the spacing of the feed from the edge, s, and as the resonant frequency is reduced, s is decreased to obtain a 50-Ω input impedance.

The bandwidth of conventional PIFA is about 8%. [39] The bandwidth can be increased by raising the top plate height h or altering the top plate geometry. [H.9.1: Waterhouse, Chap. 7] Using a specially shaped feed plate instead of a straight feed wire, a bandwidth of 25% is possible. [40] Parasitic elements near the PIFA can be used to increase bandwidth, but add to the antenna size. PIFA size can be reduced by increasing the effective current path length by altering the top plate geometry through meandering, notches, or slots. Dielectric loading can also be used but at the expense of bandwidth reduction.

The foregoing assumes an infinite ground plane. The impedance of a PIFA with a finite ground plane does not differ greatly from the infinite ground plane case, but the resonant frequency, bandwidth, and pattern do change. A investigation [39] of ground plane size influence on the PIFA's resonant frequency and bandwidth using simulations and measurements showed that as ground plane size ($L_g = W_g$) is reduced, resonant frequency is unaffected unless smaller than $0.2\,\lambda$, but bandwidth reduces below the infinite ground plane value of 8% for ground plane side lengths below $0.8\,\lambda$, with 5% bandwidth for ground plane sides of $0.5\,\lambda$. Gain is approximately equal to the infinite ground plane value of 4.5 dB for ground planes at least $0.8\,\lambda$ in size, but decreasing to 3 dB for $0.8\,\lambda$ side lengths.

Patterns of a PIFA with various size ground planes are shown in Fig. 11-26. All patterns have an overhead ($+z$-directed) null, indicating monopole behavior. In fact, for the infinite ground plane case the pattern is very similar to a quarter-wave monopole with pattern peaks in the $\theta = 90°$ directions. As the ground plane is reduced in size, the pattern peaks rise above the horizontal plane and radiation also increases in the lower hemisphere. When the ground plane size reaches the top plate size of $0.156\,\lambda$ square the pattern becomes dipole-like. The values mentioned so far were for the PIFA centered on the ground plane, and optimum performance is obtained by placing it near a corner of the ground plane. [39]

Bandwidths of a PIFA mounted in a typical handset can reach 17%. [41]
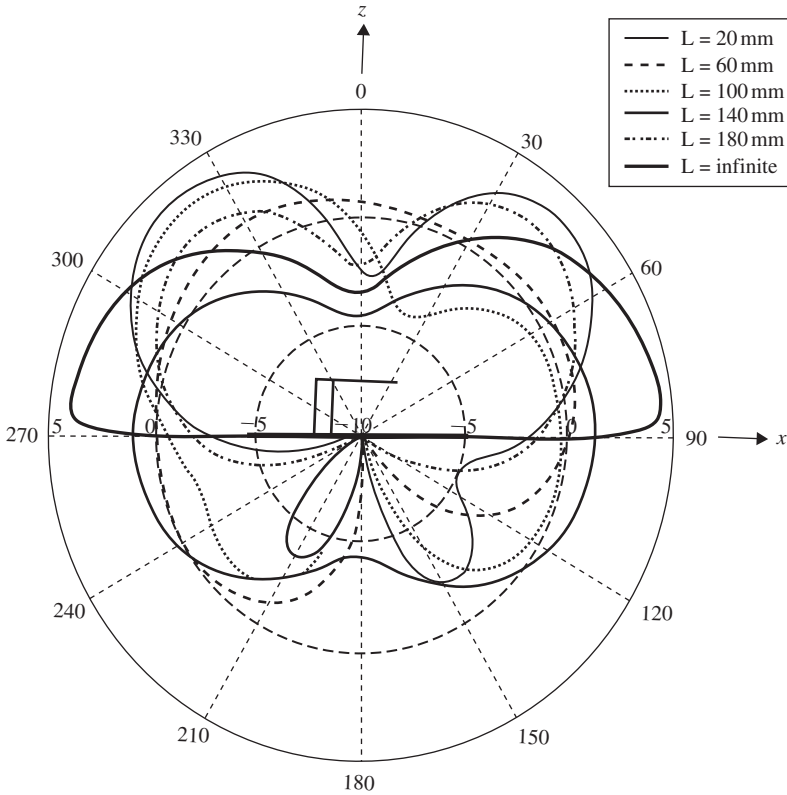
Figure 11-26 Radiation patterns of the PIFA antenna of Fig. 11-25 on finite ground planes. The top plate dimensions are $L = W = 0.156\lambda$ and the ground plane dimensions $L_g = W_g$ are infinite, 0.940λ, 0.643λ, 0.391λ, and 0.156λ. (From [39] © 2003. Reprinted with permission from IET.)

11.6.4 Other Compact Antennas, Including Multiband/Broadband Handset Antennas

Handsets have unique operational circumstances. They can be held in most any orientation. For best reception from a base station, the handset should respond best to signals in the horizontal plane parallel to the ground, called the azimuth plane. So handsets are designed for a pattern peak in the azimuth plane. But the directivity (and thus gain) are intentionally low to have response to signals from any direction because of the random orientations that are possible, and the random incoming signal directions due to multipath propagation. Thus, the specification on gain for a handset antenna is near 0 dB. Similarly, cross-polarization response is usually high for the same reasons. This is counter to the usual antenna requirement situation in that low gain and poor polarization purity are desired, not avoided. Typical specifications on handset mobile phone antennas are listed in Table 11-1. While impedance match (usually to 50 Ω) for most antennas is specified as VSWR ≤ 2, handset antennas often have a degraded match of VSWR ≤ 3.

Handset antennas are evaluated using the usual antenna parameters of gain, pattern, and efficiency. These are measured with the antenna mounted on its intended handset platform. But the special operating conditions and antenna requirements of handsets necessitate additional evaluation parameters, which are directed toward evaluating the antenna as it will be used, often including the operating environment and propagation conditions. Here, we give a simple description of the commonly used parameters and a

Table 11-1 Typical Handset Antenna Specifications

Gain	0 dBi
Bandwidth	12% @ 800 MHz (824−960 MHz)
	25% @ 1900 MHz (1710−2170 MHz)
Impedance match	VSWR ≤ 3 (RL ≥ 6 dB)
Efficiency	$e_r > 50\%$ (in free space)
Polarization	Usually linear, but high cross-polarization is acceptable
Power handling	900 MHz bands[3]:
	24 dBm (UMTS/WCDMA)
	33 dBm (GSM/CDMA)
	1900 MHz bands:
	24 dBm (UMTS/WCDMA)
	30 dBm (GSM/CDMA)
Handset size (bar shape)	$100 \times 50 \times 15$ mm^3
Antenna size	$40 \times 20 \times 10$ mm^3

full quantitative development is found in Sec. 13.6. *Total radiated power* (*TRP*) is the total power radiated from a handset when transmitting by summing up power density received at sample points (both polarizations) over a sphere around the handset. *Total isotropic sensitivity* (*TIS*) is the minimum input signal power that yields an acceptable system performance parameter such as bit error rate, sampled at points over the sphere around the handset in both polarizations. TRP and TIS permit the system engineer to determine the effectiveness of the handset, with emphasis on the maximum range it will operate.

The common cellular telephone 800 MHz and 1900 MHz bands require the antenna to operate at frequencies about an octave apart and have bandwidths in each band of at least 15% and 24%, as indicated in Table 11-1. Antennas used in modern handsets have no protruding parts, but instead are entirely internal and have a single feed point. A possible antenna is a stubby antenna consisting of a normal mode helix with two sections of differing pitch angle or diameter to support the two bands. A meander line antenna with two sections is also used. But perhaps the most popular internal handset antenna is some variation of the PIFA antenna for dual band operation. A common implementation is to have two sections (or arms) of the PIFA top plate operating in parallel, the longer one for the low band and the shorter one for the high band, and each arm is about a quarter-wavelength long in its operating band. Often one or both arms are folded around a supporting plastic mold or printed onto it. Sometimes the antenna is printed onto a flexible printed circuit board material and then positioned inside the handset. Slots are often inserted in the top plate to widen bandwidth. Fig. 11-27 shows an example of a commercial handset antenna based on PIFA design principles. [H.9.1: Fujimoto and James, Chap. 5] gives several PIFA configurations for handsets. The PIFA can be positioned over the PCB ground plane or off the end. Separate additional antennas are used for diversity that may or may not be the same type as the primary antenna. Also found in a handset are separate antennas for other functions such as Wi-Fi, GPS, and so on.

The *wideband compact PIFA* (*WC-PIFA*) shown in Fig. 11-28*a* makes use of a capacitive feed to broaden bandwidth. [42] The return loss plotted in Fig. 11-28*b* is for an experimental model covering several popular frequency bands with VSWR < 2 (RL > 10 dB). [43] The physical dimensions are $L = 25$ mm, $W = 10$ mm, $h = 6.3$ mm, and $G = 63$ mm. A bandwidth of 49% is achieved in a very low profile that is just 6 mm in height. The narrow but long ground plane requirement is not critical when the antenna is

---

[3] See Appendix for frequency bands.
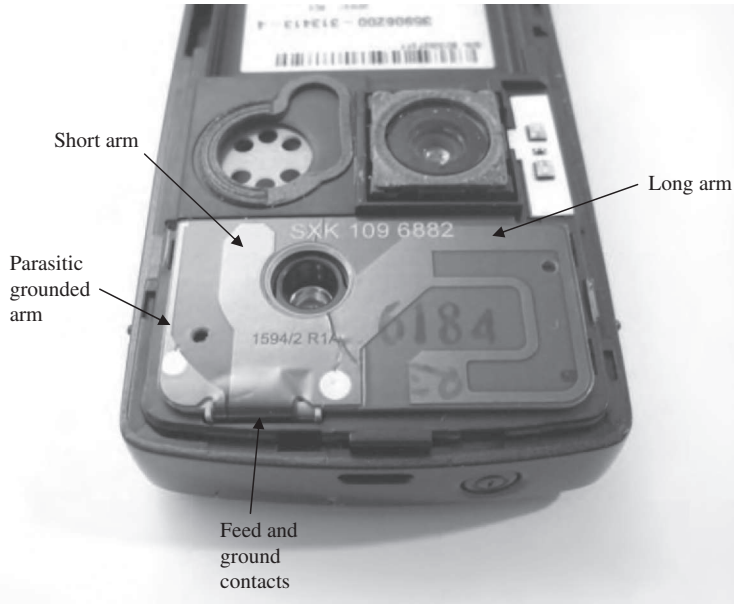
Figure 11-27 A commercial handset (Sony Ericsson W810®) showing the antenna for covering the 850/900 MHz and 1800/1900 MHz bands. It is based on PIFA design principles. (Reprinted with permission from Sony Ericsson Mobile Communications. W810® is a registered trademark used for cellular or mobile phones and owned by Sony Ericsson Mobile Communications AB.)

mounted to make use of available metal in the device as a ground plane. Another feature of this antenna is that it is relatively insensitive to its surroundings.

The most compact antenna is the so-called *chip antenna*, which is an electrically small antenna such as a NMHA or meander line antenna encased in a high-dielectric constant ceramic. The chip antenna is usually designed to make use of the handset PCB as a ground plane to assist in radiating, which makes the antenna sensitive to platform details, and tuning may be required for the specific application. A typical commercial chip antenna covering the 800/1900 MHz and GPS bands has −4 dB gain and dimensions of about $20 \times 4 \times 5$ mm^3.

All of the low-profile antennas we have discussed have been unbalanced with one feed terminal and a ground connection to a large ground plane. The currents on the ground plane make the antenna system sensitive to the ground plane details and surrounding objects, especially the human operator. This effect is strongest for the low-frequency bands. If a balanced antenna is used, antenna performance will be less sensitive to the handset dimensions and to a human operator. Also, there is more flexibility in the placement of the antenna in the handset. Examples of balanced antennas are dipole and quadrifilar antennas. A balanced antenna is often fed with a differential amplifier to reduce RF circuit complexity. [H.9.1: Chen, p. 48]

In the design of antennas for handsets, allowance must be made for the shift of the operating band when the antenna is placed in a case of low-loss plastic. Generally the band shifts down in frequency by a few percent.

11.6.5 Radio Frequency Identification (RFID) Antennas

A *radio frequency identification* (*RFID*) system identifies an object without direct contact through digital wireless techniques. The RFID system has the functional blocks shown in Fig. 11-29 consisting of a reader (or interrogator) and a tag that are separated by a short distance. The reader emits a radio-frequency signal that the tag processes and then sends a

(*a*) The antenna geometry. The dimensions are: $L = 0.17\lambda_c$, $W = 0.068\lambda_c$, $h = 0.0425\lambda_c$, and $G = 0.425\lambda_c$.

(*b*) The measured return loss for a WC-PIFA designed for center frequency of 2.02 GHz.
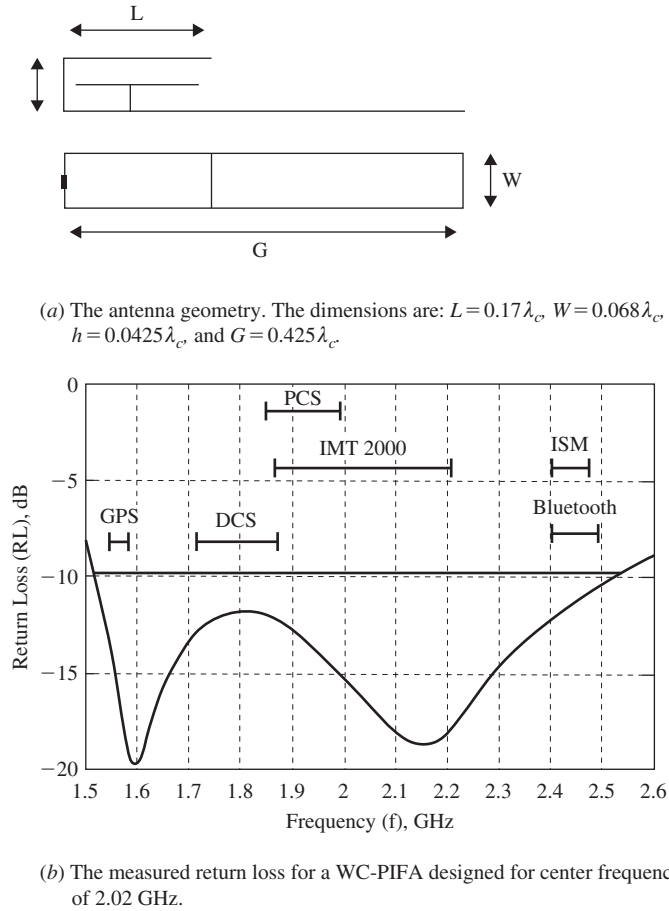
Figure 11-28 The wideband-compact PIFA antenna.
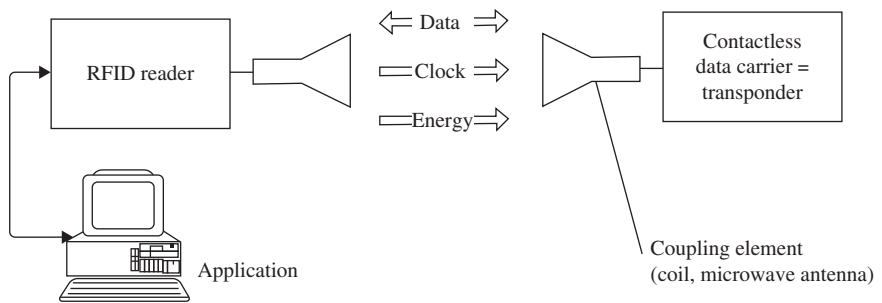
Figure 11-29 The basic elements of an RFID system. (From [45] © 2003. Reprinted with permission from John Wiley & Sons, Inc.)

return signal back to the reader containing identification, location, status, and/or other information about the tagged object. The number of applications for RFID systems is growing rapidly. Uses include security systems, vehicular applications, personal identification (e.g., passports), personnel access control (e.g., door entry authorization), and asset tracking (e.g., baggage, cargo, inventory, and animals). Table 11-2 lists the major categories of RFID applications.

Table 11-2 Some RFID Applications

Tracking and identification
    Large assets (e.g., shipping containers, military vehicles)
    Inventory control
    Retail checkout
    Pets with implanted tags
Payment and stored value systems
    Subway and bus passes
    Electronic toll systems
    Contactless credit cards
Access control
    Automobile ignition systems
    Building access proximity cards
    Ski lift passes

Table 11-3 RFID Allocated Frequency Bands

Frequency Band Name	Frequencies	Passive Read Distance
Low frequency (LF)	120–140 kHz	10–120 cm
High frequency (HF)	13.56 MHz	10–20 cm
Ultra-high frequency (UHF)	868–928 MHz	<10 m
Microwave (MW)	2.45 & 5.8 MHz	<3 m
Ultra-wideband (UWB)	3.1–10.6 GHz	10 m

Tags are distinguished by how they generate power when interrogated by the reader, and are one of three types: passive, semi-active, and active. Tags are usually small, simple devices consisting of an inexpensive electrically small antenna, an IC chip, and in the case of a semi-active or active tag, a battery to supply power. A passive tag utilizes the power in the electromagnetic field from the reader to energize a chip and respond. Passive tags are the least expensive and therefore are the most common type of tag. Passive tags potentially have an indefinitely long life. Passive tags contain a rectifier to convert the RF power from the reader into DC power to activate tag circuitry.

Tags must operate in one of the allocated frequency bands listed in Table 11-3, including the *passive read distance* for each band, which is the maximum distance from the reader that the tag can be read. The two lowest frequency systems, LF and HF, operate through inductive coupling. The tag is in the near field of the reader, and power transfer is via magnetic fields. When the tag detects the reader signal, it switches onto and off of a load resistor, which changes the impedance of the reader loop. The higher-frequency bands operate based on backscattering. The tag antenna receives the reader signal and turns on and off a load resistor, varying the antenna reflection characteristics. [H.9.1: Fujimoto, 2nd ed., Chap. 13]

The lower the frequency, the less a tag is affected by its operating environment and the more reliable the system is. For example, the LF band, which is the oldest RFID band, is well suited for applications requiring the reading of small amounts of data at low speed within short distances. This frequency range is also relatively free of regulatory limitations. Magnetic fields at these frequencies penetrate through water, tissue, wood, glass, and thin sheets of metal. Fig. 11-30 shows the shielding effectiveness of a 0.02-inch-thick aluminum plate at a near field distance of 18 inches and also for a plane wave (far-field source). Shielding effectiveness is defined for magnetic fields as
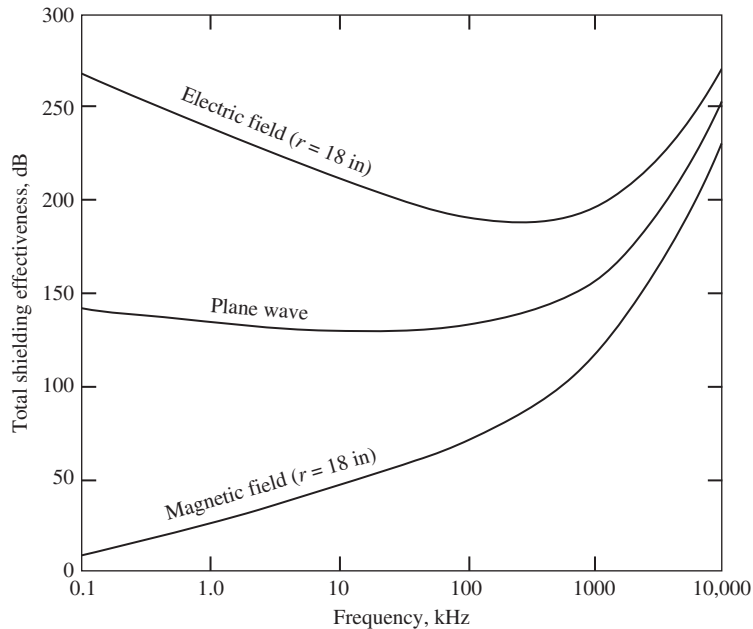
Figure 11-30 Electric field, plane wave, and magnetic field shielding effectiveness of a 0.02 in. solid aluminum plate. (From [44] © 2003. Reprinted with permission from John Wiley & Sons, Inc.)
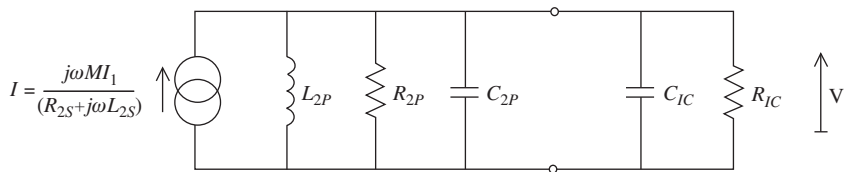
Figure 11-31 Parallel equivalent circuit of a low-frequency tag and IC. (From [52]. Y. Huang and K. Boyle, "Antennas from Theory to Practice," Wiley, 2008.)

$$S(\text{dB}) = 20\log\frac{H^i}{H^t} \tag{11-44}$$

where H^i is the field incident on the metal plate and H^t is the field transmitted on the opposite side of the plate. The definition for the shielding effectiveness for electric fields is similar. Fig. 11-30 shows that the shielding effectiveness of the aluminum plate is vastly less effective for the magnetic field compared to that for the electric field. Further, for the low-frequency magnetic field, very little reflection loss occurs, and absorption loss is the primary shielding mechanism [44]. For materials with much less conductivity than aluminum, like wood, plastic, or glass, the magnetic field will penetrate quite easily. The better penetrating qualities of magnetic fields aid the performance of the LF and HF RFID systems that operate using magnetic coupling. The Norton equivalent circuit for a mutual coupling-based tag is shown in Fig. 11-31 [H.3: Huang & Boyle, p. 338–348]. The circuit to the left of the terminals is the tag antenna (e.g., coil) and that to the right is the IC chip. The maximum voltage across the IC resistance, R_{IC}, occurs at the resonant frequency, which is determined by all the circuit reactances [H.3: Huang & Boyle, p. 338–348]. The level of the resonant maximum voltage is largely determined by the resistance of the tag antenna or coil, R_{2P}. This means that the read range over which there is voltage in excess of the IC threshold is determined largely by the quality of the tag antenna (i.e., coil). Further details can be found in [45]. In general, tag antennas are the limiting factor in many RFID systems.
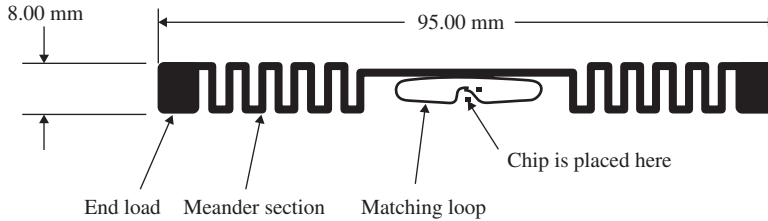
Figure 11-32 An example UHF RFID tag antenna. (Reprinted with permission from Impinj, Inc.)

In the higher RFID frequency bands, there is potential for greater passive read distances, but reliability, the EM environment, interference, and regulatory limitations become issues. Both passive and active tags are used at the higher frequencies and are far-field systems governed by the Friis transmission formula in (4-33). In the popular application of asset tracking, either low-frequency or high-frequency tags are used, depending on the object being tracked. Pallet-level and container-level RFID systems use UHF and above bands. Item-level tagging systems use the HF band because there are many more tags used and the tag cost is low.

As expected, different RFID antennas are used in different bands. The tags always use thin, low-cost antennas, whereas the size and cost constraints on reader antennas are not as great. Because RFID systems in the LF and HF bands operate using mutual inductance, multi-turn loops (coils) are used in both the reader and tag. LF tags use loops with about 40 turns and HF tags use loops of about 7 turns. Although called antennas, the loops are really coupling devices designed to enhance the mutual coupling between the reader and tag. Tags use planar multi-turn loops printed using low-cost conductive ink. As an example, the typical tag antenna size used in a retail medicine container box is about 2 cm square, corresponding to about $0.0001\,\lambda$ at 13.56 MHz.

Tag antennas for the UHF and above bands are usually a reduced-size, half-wavelength-long dipole that is meandered or spiraled in order to obtain sufficient bandwidth for the wide RFID bands at UHF and above. Wider bandwidth is needed because the allocated bands are wider than for the lower bands, and worldwide allocations do not always cover the same frequencies. Thus, a tag intended for use in multiple regions has to have wide enough bandwidth to operate in those regions. Also, objects near the tag can detune the antenna, so the antenna needs to cover more than the allocated band(s) with a moderately broadband antenna. The most popular solution is to use a meander line dipole with a perturbation to broaden the bandwidth. The principles of meander line antenna operation were presented in Sec. 11.6.1. The commercial UHF tag from Impinj shown in Fig. 11-32 is 9.5 cm long, corresponding to about $0.28\,\lambda$ at 900 MHz. Impedance matching sections are often printed with the tag antenna. One of the more popular matching configurations is the Tee match shown in Fig. 6-35*b*. The matching loop in Fig. 11-32 resembles the Tee match. The antennas used in UHF readers include dipoles, PIFAs, and microstrip patches, especially for circular polarization implementations. Advantages to UHF tags are that they have wide bandwidth for handling high amounts of data; they can be optimized for the item they are attached to; and the tags on items can be read while in a stacked configuration. See [H.8.12: Volakis et al., Chap. 9] for more details on RFID antennas.

11.7   DIELECTRIC RESONATOR ANTENNAS

A low-profile antenna, whose height can be substantially less than a wavelength, is the dielectric resonator antenna (DRA). The dimensions of a DRA are on the order of $\lambda/\sqrt{\varepsilon_r}$ or less. The most basic forms of dielectric resonator antennas are that of a short, resonating section of circular or rectangular dielectric waveguide placed vertically on a metal
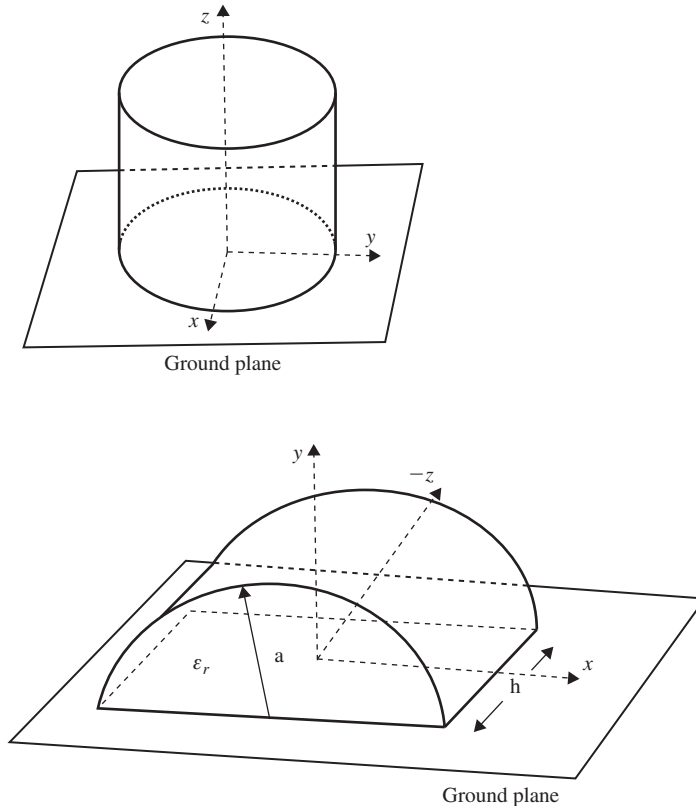
Figure 11-33 Circularly cylindrical DRAs on a ground plane.

plane, as illustrated by Fig. 11-33 for the circular cylinder. The fields excited in the dielectric radiate as explained by the Volume Equivalence Principle. The Volume Equivalence Principle states that a dielectric object (ε_d, μ_d) can be replaced by equivalent currents in free space (ε_o, μ_o). These currents then radiate in free space. The electric and magnetic volume equivalent currents $\mathbf{J}_V$ and $\mathbf{M}_V$ are [46]

$$\mathbf{J}_V = j\omega(\varepsilon_d - \varepsilon_o)\mathbf{E} \quad \text{and} \quad \mathbf{M}_V = j\omega(\mu_d - \mu_o)\mathbf{H} \qquad (11\text{-}45)$$

and exist only in the dielectric (ε_d, μ_d) itself. $\mathbf{E}$ and $\mathbf{H}$ are the electric and magnetic fields in the dielectric. Although any amount of field in the dielectric has the potential to radiate according to the Volume Equivalence Principle, resonance provides the necessary concentration of fields for strong radiation. However, the fields E and H are generally unknown and must be found (e.g., Moment Method, Finite Element, FDTD, etc.) before the volume equivalent currents can be used to calculate the radiation. The volume equivalent currents are useful for radiation by dielectric volumes just as equivalent surface currents are useful for radiation by conducting objects.

Most dielectrics are non-magnetic and so the electric volume equivalent is the current of interest. It indicates that the strength of the current $\mathbf{J}_V$ is proportional to the relative dielectric constant since $(\varepsilon_d - \varepsilon) = \varepsilon_o(\varepsilon_r - 1)$, and also proportional to the electric field, $\mathbf{E}$, in the dielectric, which is a maximum at the resonant frequency. When ε_r is high, the plane wave reflection coefficient at the dielectric-air interface becomes large, resulting in the confinement of energy within the dielectric, which makes resonance possible. For example, the plane wave reflection coefficient exceeds ½ when ε_r is greater than only 9 (see Prob. 11.7-1).
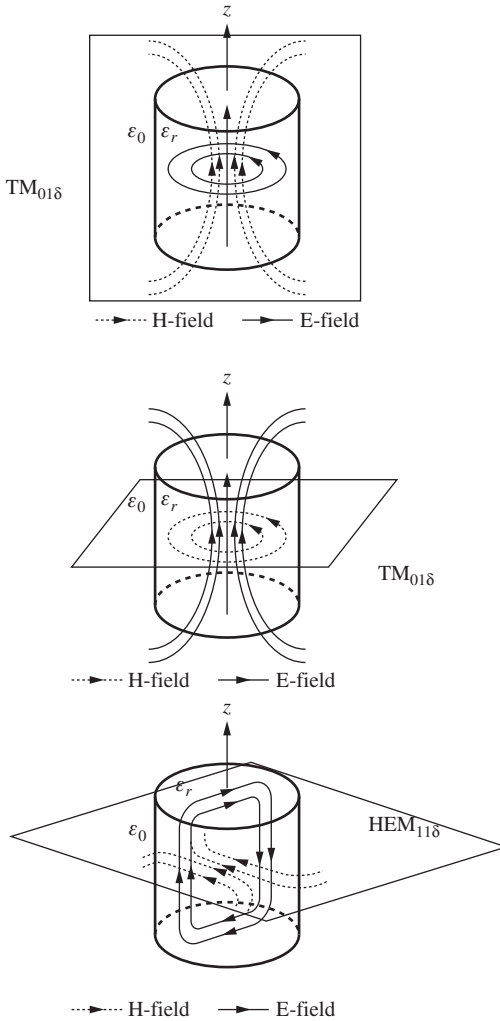
Figure 11-34 Mode patterns for the lowest-order modes of the circularly cylindrical DRAs.

Dielectric Resonant antennas are potentially quite efficient since they are not enclosed by metal walls, thereby allowing all the energy inputted to the DRA to be radiated except for modest amounts of power consumed by the ground plane and feed line conduction losses, the loss in the dielectric itself typically being quite small. Thus the stored energy inside the dielectric can be quite high but quite small outside, making it difficult for external objects to couple to and/or detune the DRA. Because of this, mutual coupling in an array is low, there are no surface waves, no scan blind angles, and thus the scanning range can be quite wide.

The DRA exhibits well-defined resonances. The most commonly used modes in a circular cylinder geometry that exhibit resonances are the lower-order TE, TM, and hybrid modes. These are shown in Fig. 11-34. The $TM_{01\delta}$ produces a monopole pattern as shown in Fig. 11-35 while the $TE_{01\delta}$ and hybrid $EH_{11\delta}$ produce a pattern like that of a narrow slot on a ground plane. The mode subscripts denote the variation of fields in radial, azimuth, and z-directions inside the DRA, respectively. The directivity of these modes is similar to that of an ideal electric or magnetic dipole (i.e., 1.76 dBi + 3 dB ground plane effect) for the lower-order modes when ε_r is high.
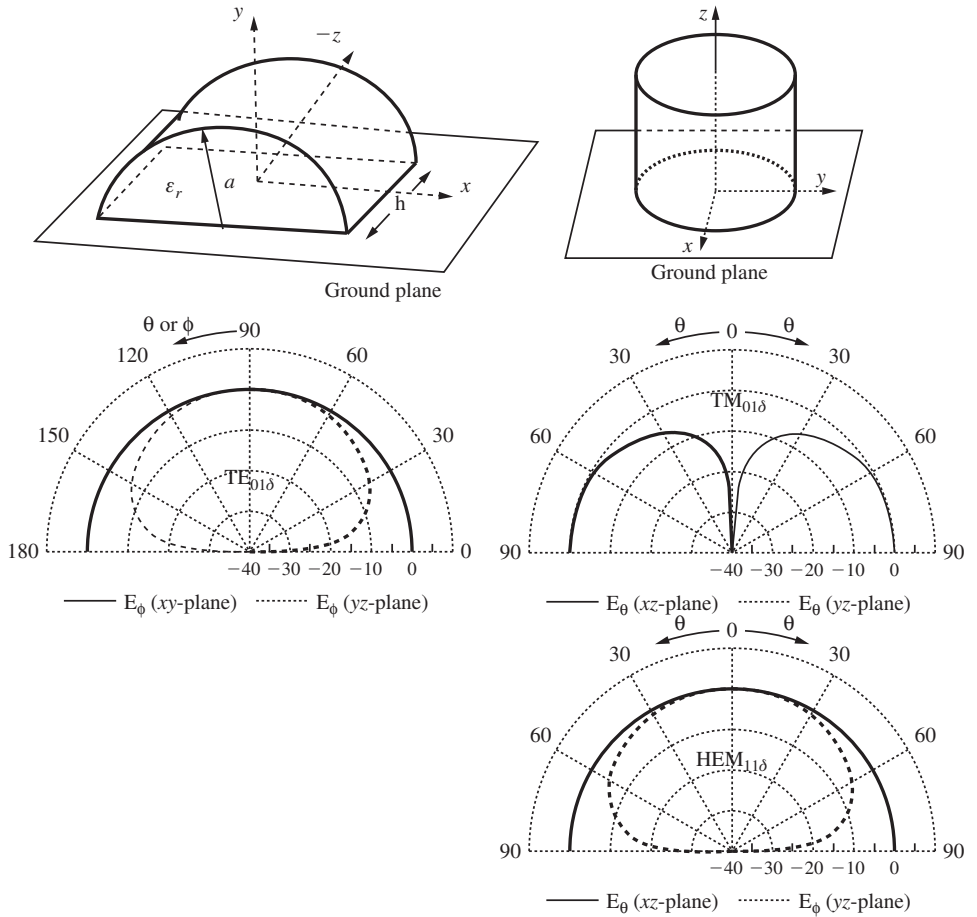
Figure 11-35 Radiation patterns for the lowest-order modes of the circularly cylindrical DRAs.

The well-defined resonances of the various DRA modes are indicative of somewhat high Q devices with bandwidths in the 5–10% range. However, these basic bandwidths can readily be expanded to 20% and beyond, which is an advantage over patch antennas. Let βa be a normalized phase constant or wavenumber (i.e., β normalized with respect to a). The resonant frequency of a circularly cylindrical DRA can be found from the normalized resonant phase constant, which is of the form

$$\beta a = \frac{1}{\sqrt{\varepsilon_r + \chi}} \kappa_1 \qquad (11\text{-}46)$$

where the values of χ and κ_1 are found empirically. χ was found to be approximately in the range $1 \leq \chi \leq 2$. Table 11-4 gives values for χ and κ_1 for the three lowest-order modes [47; H.8.7: Petosa, Sec. 2.2]. Alternative empirical formulations may be found in [H.6: *Ant. Eng. Hdbk.*, 4th ed. Sec. 17.1].

Solving (11-46) for the resonant frequency, f_r, gives

$$f_r \simeq \frac{c}{\sqrt{\varepsilon_r + \chi}} \frac{1}{2\pi a} \kappa_1 \qquad (11\text{-}47)$$

Table 11-4 Cylindrical DRA Resonance Parameters

Mode	χ	κ_1	Limitations
$TE_{01\delta}$	1	$2.327\left[1.0 + 0.2123\left(\dfrac{a}{H}\right) - 0.00898\left(\dfrac{a}{H}\right)^2\right]$	$0.33 \leq \dfrac{a}{H} \leq 5$
$TM_{01\delta}$	2	$\sqrt{3.83^2 + \left(\dfrac{\pi a}{2H}\right)^2}$	$0.33 \leq \dfrac{a}{H} \leq 5$
$HE_{11\delta}$	2	$6.342\left[0.27 + 0.36\left(\dfrac{a}{H}\right) + 0.02\left(\dfrac{a}{H}\right)^2\right]$	$0.4 \leq \dfrac{a}{H} \leq 6$
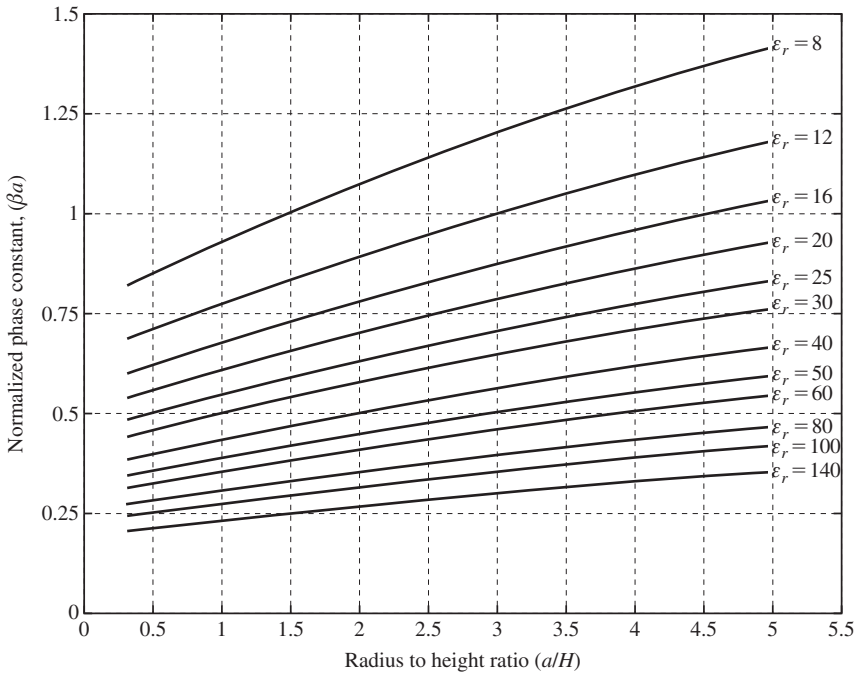
Figure 11-36 Normalized phase constant of the $TE_{01\delta}$ mode of the circularly cylindrical DRA.

The resonance frequency in (11-47) is mostly determined by ε_r and a, and is inversely proportional to these quantities as one would expect. It follows that f_r is modified by the ratio a/H as given by the relationships for κ_1 in Table 11-4. The value for κ_1 in Table 11-4 the $TE_{01\delta}$ mode varies monotonically from about 2.5 to 4.25. Fig. 11-36 shows the $TE_{01\delta}$ mode normalized wave number for $8 \leq \varepsilon_r \leq 140$.

Empirically derived relationships for the radiation Q for two of the lowest-order modes are given in [47; H.8.7: Petosa, Sec. 2.2] and for a third ($TM_{01\delta}$) in [H.6: *Ant. Eng. Hdbk.*, 4th ed. Sec. 17-1]. These relationships are of the general form

$$Q_A = (\varepsilon_r)^p \kappa_2 \qquad (11\text{-}48)$$

Table 11-5 Cylindrical DRA Q_A Parameters

Mode	p	κ_2	Limitations
$TE_{01\delta}$	1.27	$0.078192\left[1.0 + 17.31\left(\dfrac{H}{a}\right) - 21.57\left(\dfrac{H}{a}\right)^2 + 10.86\left(\dfrac{H}{a}\right)^3 - 1.98\left(\dfrac{H}{a}\right)^4\right]$	$0.5 \leq \dfrac{a}{H} \leq 5$
$TM_{01\delta}$	0.468	$2.933\left\{1 - \left[0.075 - 0.1\left(\dfrac{a}{H}\right)\right](\varepsilon_r - 10)/28\right\} \left\{1.048 + 0.734\left(\dfrac{a}{H}\right) - 0.142\left(\dfrac{a}{H}\right)^2\right\}$	$0.33 \leq \dfrac{a}{H} \leq 5$
$HE_{11\delta}$	1.30	$0.01007\left(\dfrac{a}{H}\right)[1 + 100e^{-2.05x}],$ $\times = \left\{0.5\left(\dfrac{a}{H}\right) - 0.0125\left(\dfrac{a}{H}\right)^2\right\}$	$0.4 \leq \dfrac{a}{H} \leq 6$
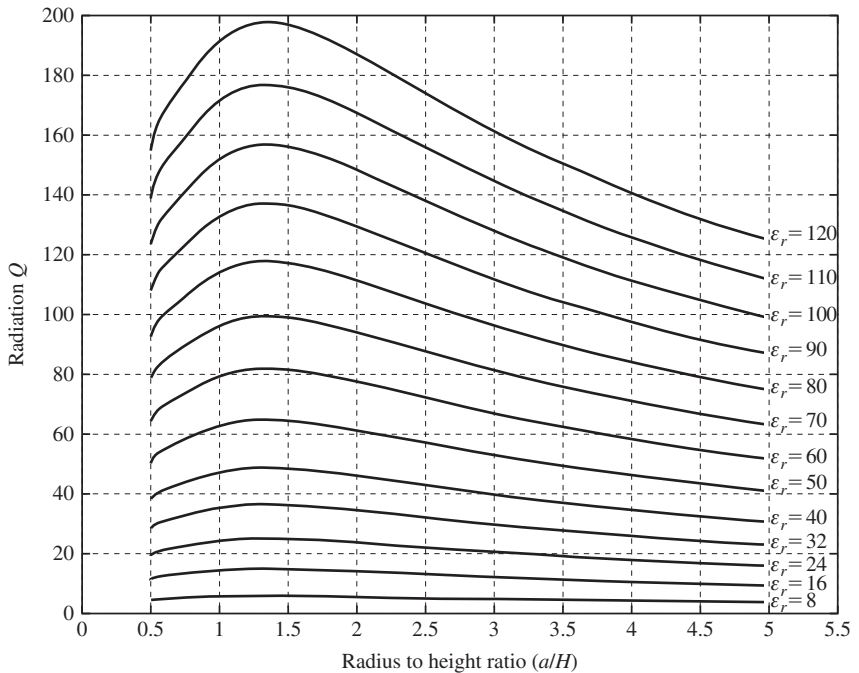
Figure 11-37 Radiation Q of the $TE_{01\delta}$ mode of the circularly cylindrical DRA.

Values for p and κ_2 are given in Table 11-5. Fig. 11-37 plots the $TE_{01\delta}$ mode radiation Q versus the height to radius ratio for $8 \leq \varepsilon_r \leq 120$. Here κ_2 varies from about 0.46 to 0.28.

The graphs in Figs. 11-36 and 11-37 may be used for design purposes. The steps are to find the Q, then chose ε_r, next determine βa, and finally determine the radius. Example 11-5 illustrates the procedure.

EXAMPLE 11-5   **DRA Design**

A design for a cylindrical DRA in the $TE_{01\delta}$ mode is to have a bandwidth of at least 3.5% for a VSWR of 2 at 10 GHz.

1. Determine the radiation Q. The relationship between VSWR, Q, and fractional bandwidth, B_f, is $B_f = \frac{VSWR-1}{Q\sqrt{VSWR}}$, which yields a Q of 20 in this case.
2. To determine the dielectric constant, draw a straight horizontal line on at $Q = 20$ in Fig. 11-37. The desired design must fall on or below this line. If $\varepsilon_r = 16$ is chosen, a wide range of radius-to-height ratios is possible for the designer, whereas if $\varepsilon_r = 24$ is chosen, for example, the choice for (a/H) is more limited. We will chose $\varepsilon_r = 16$.
3. To determine βa, note that $\beta a = 0.20958 \cdot \left(\dfrac{a}{H}\right) \cdot f(\text{GHz}) \cdot H(\text{cm})$

   which in this case is $\beta a = 2.0958 \cdot \left(\dfrac{a}{H}\right) \cdot H(\text{cm})$

4. To find the radius, plot the expression immediately above for βa on the curves in Fig. 11-36 as shown in Fig. 11-38. This will result in a series of straight lines for various values of $H(\text{cm})$. Any $H(\text{cm})$ line that intersects the $\varepsilon_r = 16$ curve is a possible solution for βa and (a/H). Some possible solutions are shown in Fig. 11-38.
5. Using (11-47) to check the resonant frequency, we chose $H = 0.3$ cm. Using the cursor on the MATLAB plot of Fig. 11-38 to locate coordinates, it is determined that $\beta a = 0.6898$ and $(a/H) = 1.097$ from which it is found that $a = 0.3291$ cm. Using these values and the cursor on Fig. 11-37 shows that the Q is 15.11. Thus $B_f = 4.68\%$, which is greater than the 3.5% requirement due to using $\varepsilon_r < 20$. A calculation of κ_1 gives $\kappa_1 = 2.844$. Then using these various values in (11-47) confirms that $f_r = 10.0$ GHz exactly.
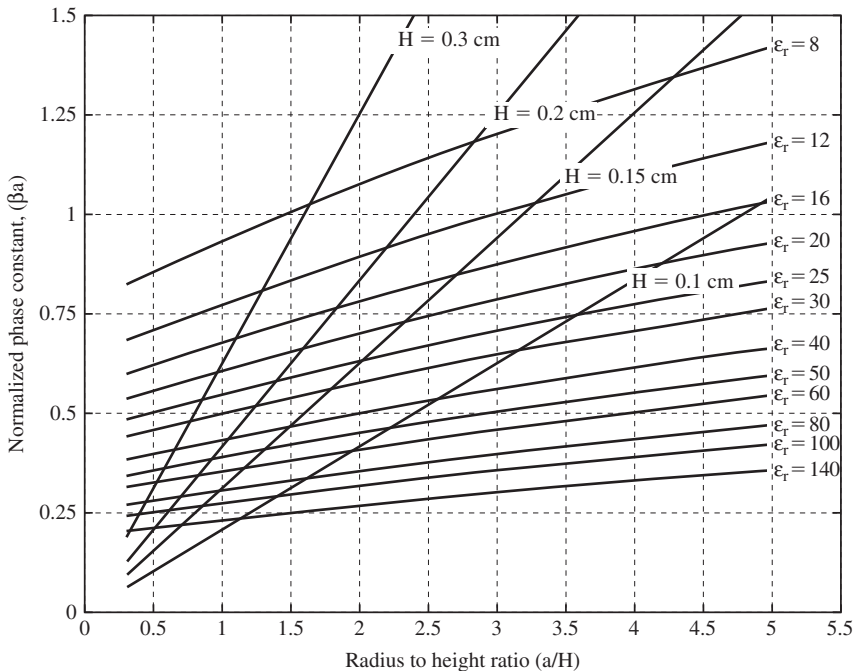
Figure 11-38 $TE_{01\delta}$ DRA designs at 10 GHz for $\varepsilon_r = 16$.

Tables 11-1 and 11-2 use empirically derived expressions because exact analytically derived results are not tractable for all but the hemispherical DRA. The usual approximate method of analysis for most shapes involves assuming that the z-component of the magnetic field vanishes at all surfaces parallel to the z-axis [48]. This is called the magnetic wall or perfect open circuit condition and is known to be accurate for high values of ε_r, but holds reasonably well for low values as well. Mainly for fabrication reasons, the hemispherical DRA is seldom used.

There are a variety of DRA shapes that have been studied other than the sphere, circular cylinder, and rectangular cylinder. It is beyond the scope of this text to explore those here. Many other shapes have been investigated that provide improvements in bandwidth. A number of geometries provide bandwidths on the order of 20%, and several on the order of 50%, and some more. An excellent reference for DRAs is found in [H.8.7: Petosa].

DRAs can be designed to have a low profile (e.g., $H < 0.03\lambda$). This is usually done with high-dielectric constant material, that is ε_r on the order of 100. The reported bandwidths are between 1 and 4% because of the high Q. This is an example of the tradeoff between ε_r and Q or bandwidth. By way of contrast, microstrip antennas experience a decrease in radiation efficiency as ε_r is increased due to the increased excitation of surface waves whereas DRAs do not suffer from this problem.

A number of excitation methods are used for DRAs. Some of these are the coaxial probe, aperture-coupling with a microstrip feed line, direct microstrip feed line, co-planar feed, soldered-through probe, slotline, stripline, and dielectric image guide. Details of these may be found in [H.8.7: Luk & Leung].

DRAs have been mostly used from 2 to 40 GHz, but some usage has been as low as 1.3 GHz for GPS applications. The lower frequency limit is due to large size and weight constraints. The higher frequency limit is due to issues associated with small size, particularly difficulties with fabrication. Small size also presents challenges with feeding the DRA. Another fabrication issue is bonding the DRA to its metal ground plane. This must be done without air gaps between the DRA and the ground plane.

11.8   NEAR FIELDS OF ELECTRICALLY LARGE ANTENNAS

In Chap. 2 the boundary between the near field and the far field was examined. Several criteria were presented in (2-100). These criteria depend on the size of the antenna and the wavelength. Similarly there were criteria for the boundary between regions within the near field itself. The near field regions are called the reactive near field and radiating near field, as Fig. 2-8 indicates. The boundary between the reactive near field and radiating near field also depends on the size of the antenna and the wavelength. In Chap. 2 the near fields of the electrically small antenna ideal dipole were examined. In this section, we will examine the near fields of some electrically large antennas.

11.8.1   Near Field of a Uniform Rectangular Aperture

Fig. 11-39 shows the behavior of the power density along a line perpendicular to the center of a large uniform rectangular aperture. The graph starts on the right at the far-field boundary of $2D^2/\lambda$ and begins to separate from the $1/r^2$ curve. Moving toward the aperture, the power density cannot vary as $1/r^2$ since this would imply an infinite power density at the aperture.

If one visualizes the aperture as being made up of a large number (e.g., 100) of small, equal size rectangles, the oscillation in Fig. 11-39 can be explained by the differences in path length from each small rectangle to the observation point along the line perpendicular to the center of the aperture. At infinity, the path lengths are all the same. Moving toward the aperture, but still well out in the far field, the path lengths are not all quite the same, but the differences are only a few degrees in phase. In the near field, the path lengths can be different enough that some are nearly $180°$ different in phase from others. This constructive and destructive process produces the oscillation in Fig. 11-39. The closer the observation point comes to the aperture, the more the net contribution to the
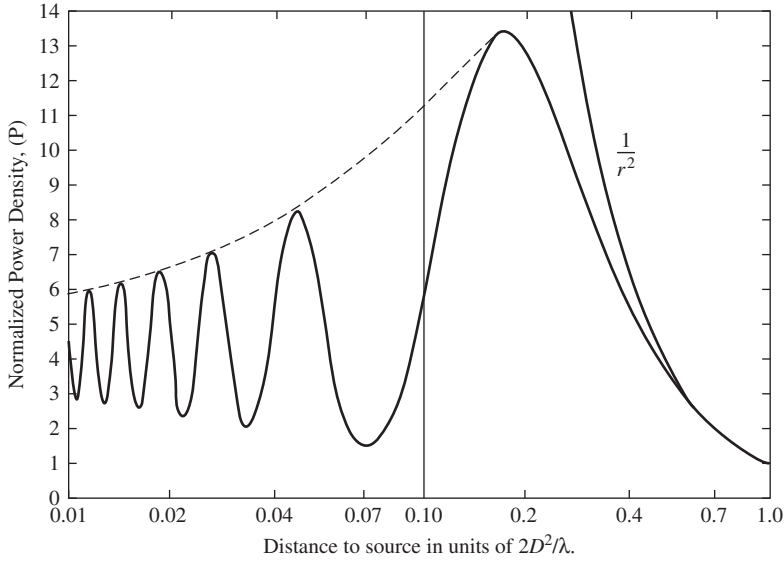
Figure 11-39 Near-field behavior of a uniform rectangular aperture. (From [H.6: Hansen] © 1964 Academic Press. Used with permission.)

power density comes from the small rectangles at the center region of the aperture. This concept is similar to the stationary phase principle in Sec. 16.15.

11.8.2 Calculating Near Fields

In Chap. 2 the common far-field criterion of $2D^2/\lambda$ was derived by retaining the third term in the expansion for R in (2-86). The retention of the third term is sometimes referred to as the Fresnel approximation, since the radiating near field may also be referred to as the Fresnel region. One approach to calculating fields in the radiating near field is to use the Fresnel approximation, as Fig. 11-40 illustrates.

Fig. 11-40*a* shows a comparison between the two-term, far-field approximation the three-term Fresnel approximation. The difference between the two curves increases as the distance to the aperture decreases. Close to the aperture, the far-field approximation predicts deep nulls, whereas the more accurate Fresnel approximation does not. Also apparent in Fig. 11-40*a* is an oscillation like that in Fig. 11-39.

Fig. 11-40*b* shows a comparison between the Fresnel approximation and probe-measured values of the field strength. The agreement between calculation and experimental measurements is generally quite good.

Another method of calculating near fields is obtained by quantizing the source into small segments such that the observation point is in the far field of each small segment while being in the near field of the entire source. The far fields of all the segments are summed at the point of observation, with the proper amplitudes and phases determined by the original source distribution. We will illustrate the method with a line source, but the method is extendable to rectangular apertures as well. The following discussion is adopted from [H.8.8: Walter, pp. 40–44].

Consider Fig. 11-41, where a line source is comprised of N segments. Let each segment have a constant amplitude I_n, a constant absolute phase ψ_n, and a constant phase constant β_n all of whose values are determined by the source. Then

$$I_n(z'_n) = I_n e^{j\psi_n} e^{j\beta_n z'_n} \tag{11-49}$$
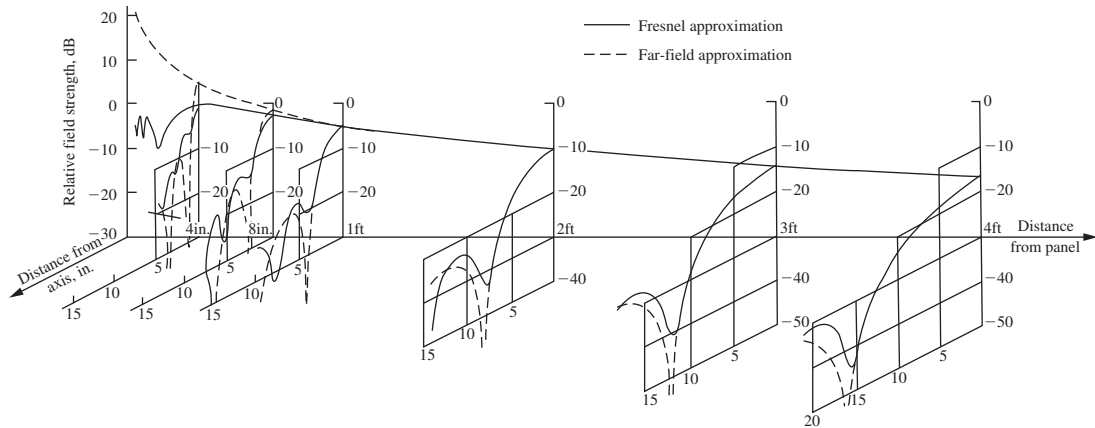
(*a*) Comparison of Fresnel and far-field computations for a 4-inch-square (10.16 cm) aperture at a wavelength of 1.18 inches (3 cm).
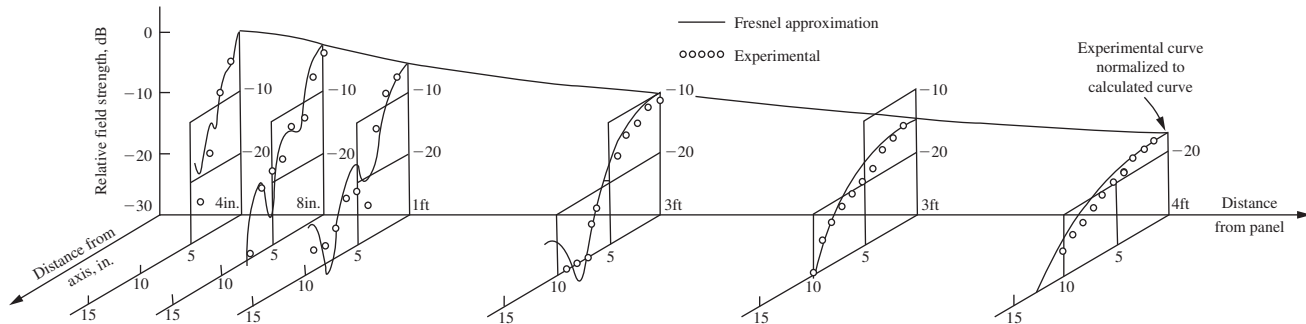
(*b*) Comparison of Fresnel computations and measurements for a 4-inch-square (10.16 cm) aperture at a wavelength of 1.18 inches (3 cm).

Figure 11-40 Fresnel calculations compared to the far-field approximation and to experiment. From [H.8.8: Walter] © 1965 The McGraw-Hill Companies © 1990 Peninsula Publishing.
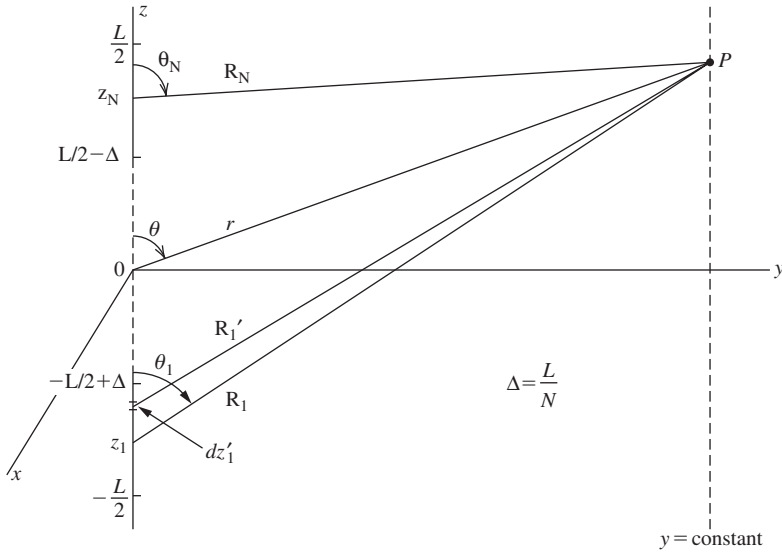
Figure 11-41 Near-field observation point in the far field of individual segments.

The observation point in Fig. 11-41 is in the far field of the individual segments but in the near field of the overall source. In this situation it is desirable to work with the z-component of the electric field on the line $y = $ constant in the figure. The derivation in [H.8.8: Walter, pp. 40–44] employs the equivalent of (2-80) and (2-95), but here we only present the final formulation.

For N segments, the minimum distance for which the result is a valid representation of the source fields is

$$y = \frac{2L^2}{\lambda N^2} \tag{11-50}$$

The electric field is given as

$$E_z = \frac{-j\omega\mu L}{4N\pi} \sum_{n=1}^{N} \frac{I_n e^{j(\psi_n - \beta R_n)}}{R_n} \sin^2\theta_n \frac{\sin u_n}{u_n} \tag{11-51}$$

where now u_n becomes

$$u_n = \left[\frac{\beta(z - z_n)}{R_n} + \beta_n\right] \frac{L}{2N} \tag{11-52}$$

The nearer the observation point to the source, the more the source needs to be divided into smaller uniform segments.

The y-component of the electric field may be found in the same manner as the z-component, thus giving the total near field of the source. This method can be applied in the reactive near field as well as the radiating near field and has been used to investigate the quiet zone in a compact range.

Fig. 11-42 shows calculations using the method of this section for a 250λ-long, uniform-amplitude constant-phase line source. The z-component of the electric field is calculated along a line $y = $ constant shown in Fig. 11-42. The variation in z is $-125\lambda < z < 125\lambda$. The distances $y = 300\lambda$ and 600λ are in the region where there would be a "quiet zone" for approximately $-60\lambda < z < 60\lambda$. The distance $y = 300\lambda$ offers less amplitude and phase variation than $y = 600\lambda$, and the best quiet zone location is probably
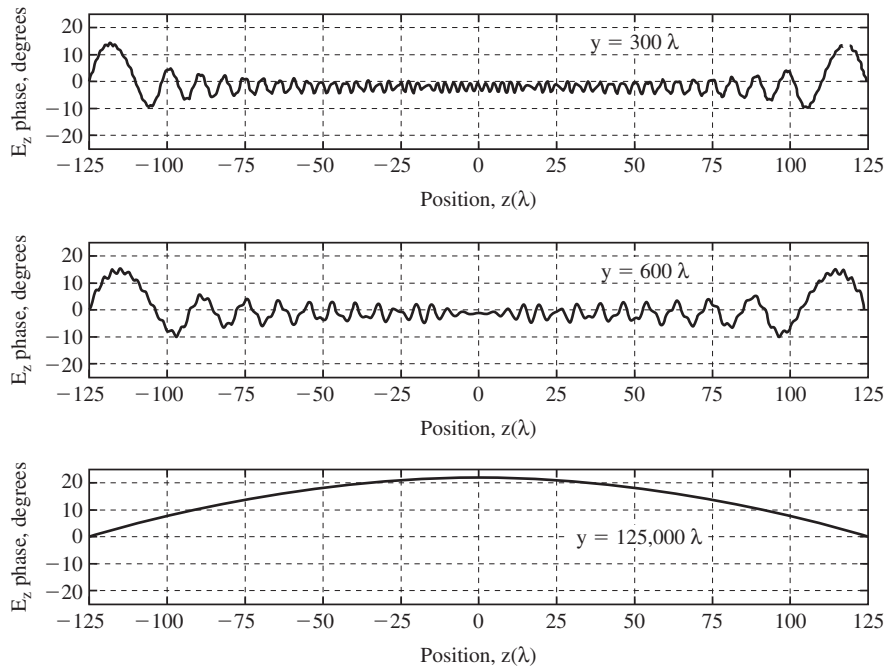
Figure 11-42 Near-field normalized phase variations at various distances from a uniform line source along the *z*-axis.

near $y = 300\lambda$. The distance $y = 125,000\lambda$ is the $2D^2/\lambda$ boundary between the near and far fields. Note that the phase plot shows the 22.5° phase variation used in the derivation of $2D^2/\lambda$. The 1 dB amplitude variation is due to the half-power beamwidth being about 0.2°, well outside the 0.057° angle at the origin subtended by $z = 125\lambda$ and $y = 125,000\lambda$ (i.e., $(90° - \theta)$ in Fig. 11-41).

11.9   HUMAN BODY EFFECTS ON ANTENNA PERFORMANCE

The effect of the human body on the electric and magnetic fields around it can be substantial. The electric field in close proximity to the human body is reduced significantly due to the conductivity of the human tissue, whereas the magnetic field is increased, in accordance with the boundary conditions on tangential E and tangential H. This body-field interaction has received increasing attention in recent years, particularly because of the rising use of mobile phone handsets. There are two reasons for understanding the interaction of electromagnetic fields with the human body. The first reason is to understand the effect on mobile phone performance (mainly the radiation pattern and impedance). This will be examined in this section. The second reason is to understand the interaction for compliance with safety standards. This will be examined in the next section on radiation hazards.

The effect of the human body on antenna performance can be done via experimental measurements or via computer simulation using the FDTD method discussed in Chap. 15. Early computer models of the human head, for instance, consisted of a layer of bone surrounded by a layer of muscle. But the head contains many different kinds of tissue, so later models used data from magnetic resonance imaging to provide very accurate models of the various tissues in the body, perhaps more accurate than necessary for many investigations. Experimental investigations employ models called "phantoms," wherein the interior of the phantom is filled with one or more liquids having the required dielectric

constant at the frequency under consideration. The dielectric constant of the human body is both tissue specific and frequency sensitive. Extensive data for the electrical properties of the human body may be found on the FCC website: www.fcc.gov.

The main effects of the human body on the radiation from a handset is to decrease the overall radiated power, decrease the radiation in the direction of the head relative to the opposite direction, and in some cases detune the handset antenna. Fig. 11-43 shows the simulated far-field pattern of a monopole antenna on a rectangular box model of a handset located 1.5 cm from the head [49] at 900 MHz. The pattern is compared to the omnidirectional pattern of the monopole without the head present. Clearly the head alters the pattern of the monopole alone. The head also causes a marked decrease in the radiated power level due to the absorption of power by the head. Thus the radiation efficiency of (2-177) is reduced for the monopole antenna system according to

$$e_r = \frac{P}{P + P_o + P_{\text{loss}}} \tag{11-53}$$

where P_{loss} is the power absorbed by the human body. As one would expect from the discussion in Sec. 11.8, P_{loss} is reduced if the phone is positioned farther away from the head. In general, the tissues of the human body are lossy dielectrics at RF frequencies, as discussed in the next section.

The presence of the human body also can have an effect on the handset antenna impedance. Consider the handset model Fig. 11-44 but with a monopole on the handset. Fig. 11-45 shows the computed return loss for the monopole antenna on the box model.
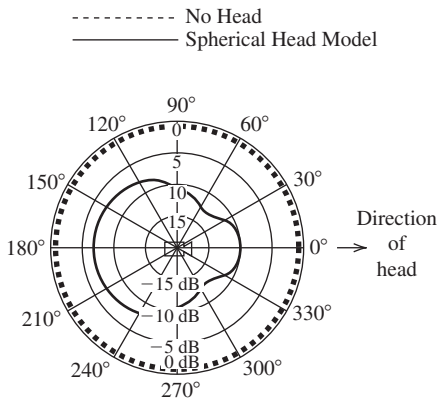
Figure 11-43 Computed radiation patterns in the azimuthal plane for a vertically oriented monopole at 900 MHz. (From [49] © 2004. Reprinted with permission from *Microwave Journal*.)
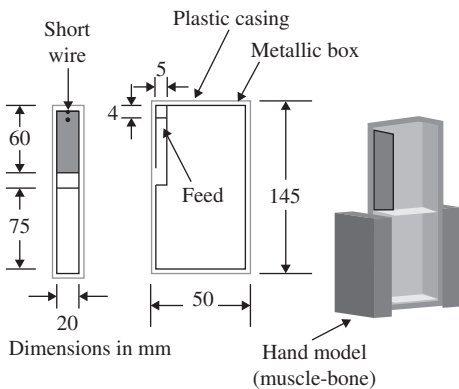
Figure 11-44 Handset model showing the PIFA antenna location and the hand model used in the return loss computations. (From [49] © 2004. Reprinted with permission from *Microwave Journal*.)

Monopole only
Monopole with hand
Monopole with hand and head
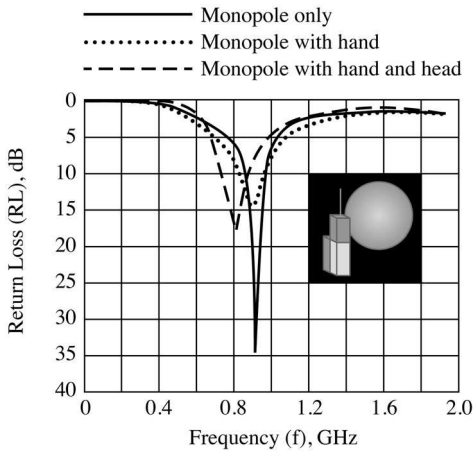
Figure 11-45 Computed return loss values of a monopole on a handset for three cases. (From [49] © 2004. Reprinted by permission of Artech House.)

No hand
d = 40 mm
d = 60 mm
d = 70 mm
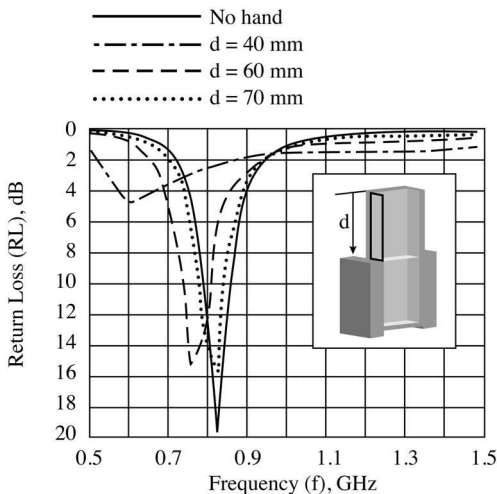
Figure 11-46 Computed return loss of the side-mounted PIFA handset for different hand locations. (From [49] © 2004. Reprinted with permission from *Microwave Journal*.)

Three cases are shown: monopole on the box model only; hand on the monopole-box model in the presence of the head; and hand on the monopole-box model but not in the presence of the head. The results indicate that there is relatively little change in the antenna input impedance as evidenced by the low level of the return loss in all three cases. This is a desirable characteristic of the monopole antenna. There is some detuning in Fig. 11-45, however, when the head is present.

Fig. 11-46 shows the effect of the hand on a planar inverted-F antenna (PIFA) for several positions of the hand. The handset model is shown in Fig. 11-44. When the hand covers an appreciable portion of the PIFA (i.e., $d = 40$ mm), the return loss curves show that the input impedance of the PIFA is noticeable affected. When $d = 40$ mm and 60 mm there is detuning of the PIFA. When $d = 70$ mm, the hand is not covering any part of the PIFA and performance is almost the same as if no hand were present. The results suggest the importance of minimizing antenna masking through proper placement of the antenna in the handset. A desirable feature of the PIFA is its ability to direct radiation away from the head due to the ground plane upon which the PIFA resides. This feature helps reduce the power absorption by the head.

Another device that employs an antenna close to the human body is a pager. Pagers are typically worn on the belt or in a shirt pocket, placing the antenna quiet close to the body,
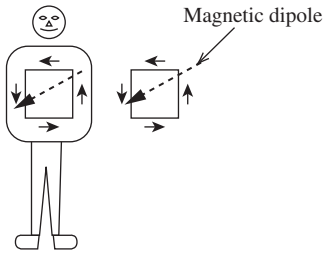
Figure 11-47 A small loop antenna on the side of a person and its image in the body. The small loop is perpendicular to the body. The magnetic dipole is parallel to the body and perpendicular to the loop.

if not for all practical purposes right on the body. In the VHF–UHF regions of the spectrum where most pager services are found, the human body acts predominately like a reflector to a degree dependent upon the frequency. The wave impedance will be low, even for plane wave incidence, suggesting a magnetic antenna be used for good on-body performance.

The magnetic antenna of choice is, of course, a small loop. A small loop positioned such that the axis of the loop is tangential to the body acts like a magnetic dipole (see Sec. 3.4) parallel to the body surface as indicated in Fig. 11-47. As such it will always receive the maximum magnetic field available [H.9.1: Fujimoto, 1st ed., p. 178], whatever the orientation of the body wearing it. The proximity of the body distorts the idealized uniform phase of the current on the small loop such that the perfect nulls of the magnetic dipole are partially filled in, a beneficial proximity effect of the human body in this instance.

11.10   RADIATION HAZARDS

Electromagnetic waves can be divided into two categories according to their ionizing capability. Ionization is the process by which electrons are stripped from atoms and molecules by high-energy photons.[4] The human body is largely water, so the water molecule is used to define the minimum quantum energy necessary to produce ejection. Ionization in humans can produce molecular changes that damage biological tissue and even DNA. Ionization occurs at wavelengths shorter than most ultraviolet rays (e.g., x-rays, gamma rays) as indicated in Fig. 11-48. The photon energy associated with x-rays is a billion times more energetic than that of a 1 GHz microwave photon.

The photon energies in RF electromagnetic fields are not great enough to cause the ionization of atoms and molecules. Therefore, RF radiation is generally non-ionizing radiation. An exception to this is radio transmitters using very high supply voltages in excess of 5 kV, which can produce x-rays. When considering biological effects of electromagnetic radiation, it is important to distinguish between ionizing radiation and non-ionizing radiation. Unfortunately, the general public does not distinguish between ionizing and non-ionizing radiation.

Radiation hazards to humans can be divided into two categories: direct and indirect. First, direct effects include both thermal effects and non-thermal effects. Thermal effects involve heating of the human body and include shocks and burns. Non-thermal or athermal effects are biological effects without significant heating. Athermal effects are not well understood and are the subject of ongoing research. There are also non-human effects such as when strong fields interact with flammable vapors. This effect is obviously a hazard. Not all biological effects constitute a hazard.

---

[4] Electromagnetic phenomena can be described either in terms of waves or particles. This is known as the wave-particle duality of electromagnetic energy. The particle in an electromagnetic wave is the photon whose energy is $\hbar f$ expressed in electron-volts where $\hbar$ is Plank's constant. There is an electron-volt scale in Fig. 11-48.
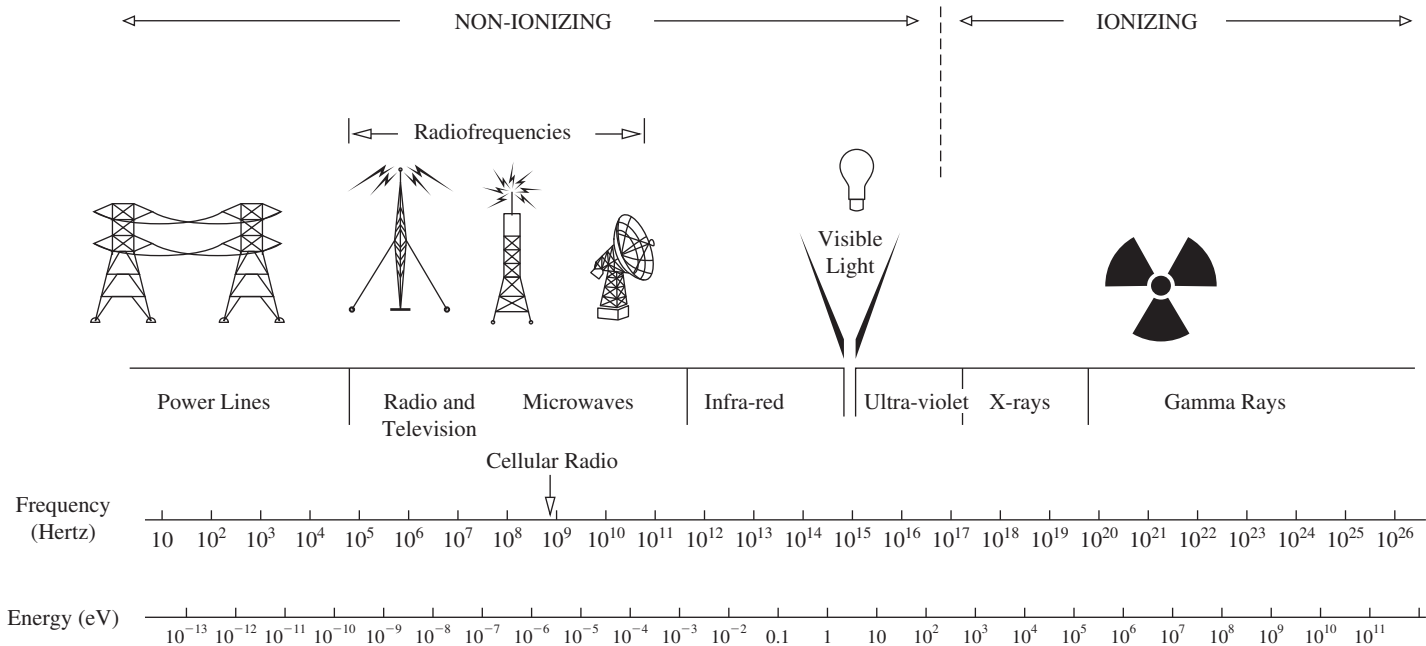
Figure 11-48 Regions of the spectrum (From [50] © 1972 IEEE.)
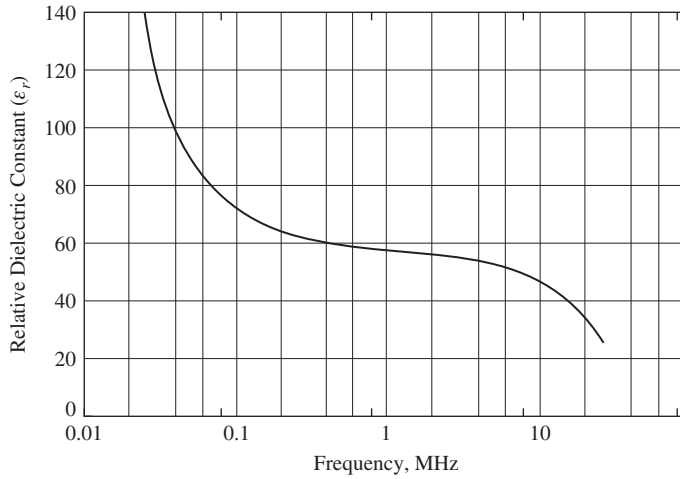
Figure 11-49 Relative dielectric constant of representative human tissue. (Figure drawn from data in [51].)
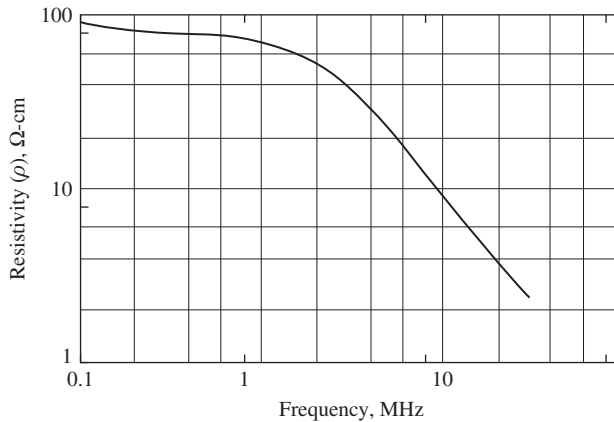
Figure 11-50 Resistivity, ρ, of representative human tissue. (Figure drawn from data in [51].)

Recall that generally a dielectric constant is complex, with the imaginary part determined by the conductivity and frequency as in (2-15). Fig. 11-49 shows the relative dielectric constant of some human tissue versus frequency. The decline in ε_r with increasing frequency at the lower frequencies is due to capacitance inherent in the cellular membranes. From about 100 MHz to about 10 GHz those membranes are short-circuited. The decline above 10 GHz is due to the water content and its frequency characteristics.

Fig. 11-50 shows the resistivity, ρ, of human tissue, which is the inverse of the conductivity, σ. The resistivity of high water content human tissues changes slowly at low frequencies but rapidly at high frequencies due to the conductivity of water changing strongly there. At these higher frequencies there is little penetration of the human body, as shown in Fig. 11-51, and surface absorption determines bio-effects and hazards except for the more sensitive eye tissue. The skin depth, δ, used for the curve in Fig. 11-51, is given by [H.4: Siwiak].

$$\delta(\omega) = \left\{ \frac{\beta^2}{2} \left[\sqrt{\varepsilon_r^2(\omega) + \left[\frac{\sigma(\omega)}{\omega\varepsilon_o}\right]^2} - 1 \right] \right\}^{-1/2} \tag{11-54}$$
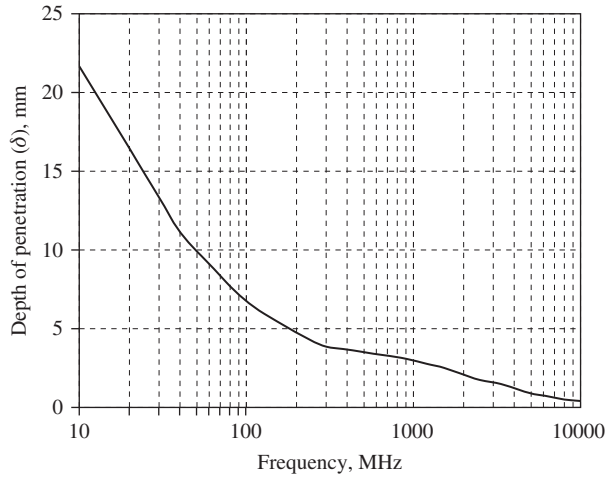
Figure 11-51 Depth of penetration of RF energy in tissues with high water content. (Data from [51] © 1972 IEEE.)

The skin depth or penetration depth is, of course, taken to be the depth where the incident power density has been reduced by a factor of e^{-2}—that is, down to about 13.5% of the incident power density.

RF exposure standards have been developed by various organizations and governments. These standards recommend safe levels of exposure for both the general public (i.e., an uncontrolled environment) and also for workers in a controlled environment. The recommended safe levels have been revised downward several times in recent decades, and not all organizations agree on exactly on what the safe levels are, but the various implemented safe levels are all similar. The various standards put in place have done a good job of protecting humans.

Table 11-6 shows the maximum permissible exposure (MPE) in terms of power density for an uncontrolled environment as specified by the ANSI/IEEE C95.1-1992 guidelines and adopted by the FCC. The standard is based on whole body absorption. Fig. 11-52 graphs the MPE FCC standard. The most restrictive limits on exposure are in the frequency range of 30–300 MHz, where the human body absorbs RF energy most efficiently, partly due to a possible half-wave dipole-like resonance around 100 MHz. The low-frequency MPE is higher than the high-frequency MPE since the depth of penetration is greater at low frequencies.

Most RF safety limits are defined in terms of the electric and magnetic field strengths as well as in terms of power density. For lower frequencies, limits are usually expressed in terms of field values, and the power densities are far-field equivalent values. At higher frequencies, where a person is likely to actually be in a far field, it is usually only necessary to consider power densities as is done in Table 11-6.

There are also limits for localized or partial body absorption, which is expressed in terms of the specific absorption rate. The specific absorption rate, or SAR, is the absorbed power, P_a, per unit volume divided by the material density, p, so as to obtain the absorption per unit mass. Thus, if P_a is expressed as

$$P_a = \frac{1}{2}\sigma|E|^2 \quad \text{W/m}^3 \tag{11-55}$$

then SAR is

$$SAR = \frac{\sigma}{2p}|E|^2 \quad \text{W/kg} \tag{11-56}$$

Table 11-6 FCC Limits for Maximum Permissible Exposure (MPE)

a. Limits for Occupational/Controlled Exposure [50]

Frequency Range (MHz)	Electric Field Strength (E) (V/m)	Magnetic Field Strength (H) (A/m)	Power Density (S) (mW/cm^2)	Averaging Time $|E|^2$, $|H|^2$ or S (minutes)
0.3–3.0	614	1.63	(100)[*1]	6
3.0–30	1842/f	4.89/f	(900/f^2)[*2]	6
30–300	61.4	0.163	1.0	6
300–1500	–	–	f/300	6
1500–100,000	–	–	5	6

b. Limits for General Population/Uncontrolled Exposure [50]

Frequency Range (MHz)	Electric Field Strength (E) (V/m)	Magnetic Field Strength (H) (A/m)	Power Density (S) (mW/cm^2)	Averaging Time $|E|^2$, $|H|^2$ or S (minutes)
0.3–1.34	614	1.63	(100)[*3]	30
1.34–30	824/f	2.19/f	(180/f^2)[*4]	30
30–300	27.5	0.073	0.2	30
300–1500	–	–	f/1500	30
1500–100,000	–	–	1.0	30

f = frequency in MHz   *Plane-wave equivalent power density

NOTE 1: Occupational/controlled limits apply in situations in which persons are exposed as a consequence of their employment provided those persons are fully aware of the potential for exposure and can exercise control over their exposure. Limits for occupational/controlled exposure also apply in situations when an individual is transient through a location where occupational/controlled limits apply provided he or she is made aware of the potential for exposure.

NOTE 2: General population/uncontrolled exposures apply in situations in which the general public may be exposed, or in which persons that are exposed as a consequence of their employment may not be fully aware of the potential for exposure or can not exercise control over their exposure.

SAR represents the ratio of an infinitesimal amount of RF power absorbed in an infinitesimal mass surrounding a specific point. SAR is a measure of the local heating rate of tissue. A SAR of 1W/kg equates to a heating rate of less than 0.0003 C per second in muscle tissue. This corresponds to a rise in temperature of 1°C in about one hour.

At a SAR of 4 W/kg, reversible tissue disruption has been observed, and above 5 W/kg permanent adverse effects have been found. Thus, most standards are based on a fraction of 4 W/kg to account for biological uncertainty and to add an additional safety factor. Exposure is averaged over a 6-minute period of time, and peak exposure cannot exceed 8 W/kg. Consider an adult human weighing 75 kg or 165 lbs. For continuous exposure lasting 6 minutes or more, the maximum RF power absorbed by the human is limited to 0.4 W/kg × 75 kg = 30 watts. An absorbed power of 30 watts is equivalent to approximately one-quarter of the resting metabolic rate of an adult. This standard was not developed to protect against possible long-term exposure to low-level RF radiation because such hazards, if any, are not well understood.

In specifying safety exposure limits for humans, both the local tissue SAR and whole body average are important. [50] There is no practical way of measuring SAR in humans. The determination of SAR is either done via computer modeling or experimental
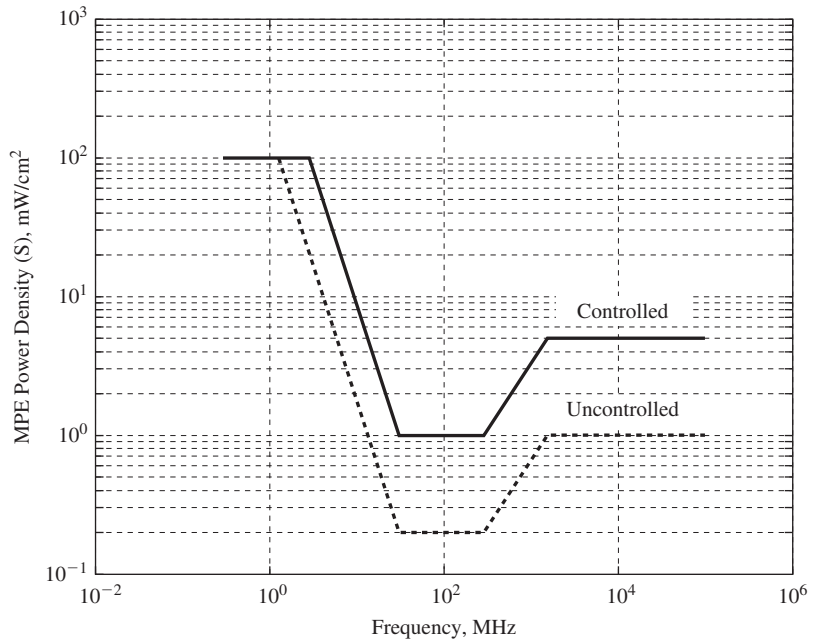
Figure 11-52 Maximum permissible exposure (MPE) in terms of power density.

measurements done with "dummy" models of the human anatomy comprised of materials which mimic the electrical characteristics of human tissues. The favorite computer modeling technique is the FDTD, discussed in Chap. 15. It should be noted that such modeling does not take into account the thermo-regulatory system of the human body.

There have been numerous studies of the SAR distribution for hand-held cellular phones done by both experimental models and software simulation models. All studies show that about 50% of the power transmitted by the cellular phone is absorbed by the body, mostly in the head on the side of the phone. These studies indicate that, in general, the 1.6 mW/g partial body limit is unlikely to be exceeded under normal phone use conditions. Various other studies, while finding that electro-stimulation in the brain does occur, find no link between cell phone use and brain tumors. Nevertheless, prudence might dictate that hand-held cell phone use be in moderation until this and other issues are more definitively resolved.

Electromagnetic fields can have a positive effect on the human body. It has been known for several decades that pulsed DC used to produce a magnetic field at the site of a bone fracture can promote healing that otherwise might not occur due to weak blood supply at the site of the fracture. A "Jones fracture" in the foot is one example where the blood flow is weak and healing is problematic. The pulsed magnetic field of a bone stimulator applied at the site of a Jones fracture promotes a flow of current in the bone similar to the currents the body would naturally produce to signal the bone to heal. While this effect is not an RF effect and certainly not a hazard, it is an example a beneficial interaction of electromagnetic fields with the human body.

REFERENCES

1. S-Y. Suh and S. Ooi, "Challenges on Multi-Radio Antenna System for Mobile Devices," *Proc. IEEE Trans. on Ant. & Prop. Inter. Symp.*, June 2007.
2. K. R. Carver and J. Mink, "Microstrip Antenna Technology," Special Issue on Microstrip Antennas, *IEEE Trans. on Ant. & Prop.*, Vol. AP-29, pp. 2–24, Jan. 1981.

3. D. R. Jackson and N. Alexopoulos, "Simple Approximate Formulas for Input Resistance, Bandwidth, and Efficiency of a Resonant Rectangular Patch," *IEEE Trans. on Ant. & Prop.*, Vol. 39, pp. 407–410, Mar. 1991.

4. M. Kara, "A Novel Technique to Calculate the Bandwidth of Rectangular Microstrip Antenna Elements with Thick Substrates," *Micro. & Optical Tech. Letters*, Vol. 12, pp. 59–64, June 5 1996.

5. L. Basilio, M. Khayat, J. Williams, and S. Long, "The Dependence of the Input Impedance on Feed Probe and Microstrip Line-Fed Patch Antennas," *IEEE Trans. on Ant. & Prop.*, Vol. 49, pp. 45–47, Jan. 2001.

6. Y. Hu et al., "Characterization of the Input Impedance of the Inset-Fed Rectangular Microstrip Antenna," *IEEE Trans. on Ant. & Prop.*, Vol. 56, pp. 3314–3318, Oct. 2008.

7. R. E. Munson, "Conformal Microstrip Antennas and Microstrip Phased Arrays," *IEEE Trans. on Ant. & Prop.,* Vol. AP-22, pp. 74–78, Jan. 1974.

8. E. Chang, S. Long, and W. Richards, "An Experimental Investigation of Electrically Thick Rectangular Microstrip Antennas,"*IEEE Trans. on Ant. & Prop.*, Vol. AP-34, pp. 767–772, June 1986.

9. S. Targonski, R. Waterhouse, and D. Pozar, "Design of Wide-Band Aperture-Stacked Patch Microstrip Antennas," *IEEE Trans. on Ant. & Prop.*, Vol. 46, pp. 1245–1251, Sept. 1998.

10. S. Maci and B. Gentilli, "Dual-Frequency Patch Antennas," *IEEE Trans. on Ant. & Prop. Mag.*, Vol. 39, pp. 13–20, Dec. 1997.

11. E. Levine, G. Malamud, S. Shtrikman, and D. Treves, "A Study of Microstrip Array Antennas with the Feed Network," *IEEE Trans. on Ant. & Prop.*, Vol. 37, pp. 426–433, April 1989.

12. J. Aberle and D. Pozar, "Phased Arrays of Probe-Fed Stacked Microstrip Patches," *IEEE Trans. on Ant. & Prop.*, Vol. 42, pp. 920–927, July 1994.

13. C. McCarrick, "Dual-Frequency Antenna Aids WLAN Applications," *Microwaves & RF*, Vol. 32, pp. 174–178, Sept. 1993.

14. Y. X. Li, et al., "The Periodic Half-Width Microstrip Leaky Wave Antenna with a Backward to Forward Scanning Capability," *IEEE Trans. on Ant. & Prop.*, Vol 58, pp. 963–966, March 2010.

15. W. Hong, et al., "Broadband Tapered Microstrip Leaky Wave Antenna," *IEEE Trans. on Ant. & Prop.* Vol. AP-51, pp. 1922–1928, Aug. 2003.

16. G. M. Zelinski, G. A. Thiele, M. L. Hastriter, M. J. Havrilla, and A. J. Terzuoli, "Half Width Leaky Wave Antennas," *IET Micro. Ant. Prop.*,Vol. 1, pp 341–348, Apr. 2007.

17. G. M. Zelinski, "FDTD Analysis of a Leaky Traveling Wave Microstrip Antenna," Thesis, Air Force Institute of Technology, March 2005.

18. J. A. Girard, "Material Perturbations to Enhance Performance of the Half-Width Leaky Mode Antenna," Dissertation, Air Force Institute of Technology, June 2008.

19. H. A. Wheeler, "The Radiansphere around a Small Antenna," *Proc. IRE*, Vol. 47, pp. 1325–1331, Aug. 1959.

20. H. A. Wheeler, "Fundamental Limitations of Small Antennas," *Proc. IRE*, Vol. 35, pp. 1479–1484, Dec. 1947.

21. L. J. Chu, "Physical Limitations on Omni-Directional Antennas," *J. Applied Physics*, Vol. 19, pp. 1163–1175, Dec. 1948.

22. R. Collin and S. Rothschild, "Evaluation of Antenna Q," *IEEE Trans. on Ant. & Prop.*, Vol. AP-12, pp. 23–27, Jan. 1964.

23. J. S. McLean, "A Re-Examination of the Fundamental Limits on the Radiation Q of Electrically Small Antennas," *IEEE Trans. on Ant. & Prop.*, Vol. 44, pp. 672–675, May 1996.

24. W. A. Davis, T. Yang, E. Caswell, and W. Stutzman, "Fundamental Limits of Antenna Size: A New Limit," *IET Micro. Ant. Prop.,*Vol. 5, No. 11, pp. 1297–1302, Aug. 19, 2011.

25. M. Gustafsson, C. Sohl, and G. Kristensson, "Physical Limitations on Antennas of Arbitrary Shape," *Proc. Royal Society*, Vol. 463, pp. 2589–2607, 2007.

26. G. A. Thiele, P. Detweiler, and R. Penno, "On the Lower Bound of the Radiation Q for Electrically Small Antennas,"*IEEE Trans. on Ant. & Prop.*, Vol. AP-51, pp. 1263–1268, June 2003.

27. G. A. Thiele, "On the Gain and Efficiency of Electrically Small Antennas," *Proc. Ant. Measurements Tech. Assoc.* (Atlanta, GA), Nov. 2005.

28. G. A. Thiele and C. Donn, "Design of a Small Conformal Array," *IEEE Trans. on Ant. & Prop.*, Vol. 22, pp. 64–70, Jan. 1974.

29. G. Goubau, "Multi-Element Monopole Antennas," *Proc. ECOM-ARO Workshop on Elec. Small Ant.*, pp. 63–67, May 1976.

30. S. Best, "The Radiation Properties of Electrically Small Folded Spherical Helix Antennas," *IEEE Trans. on Ant. & Prop.*, Vol. 52, pp. 953–960, Apr. 2004.
31. O. Kim, O. Breinbjerg, and A. Yaghjian, "Electrically Small Magnetic Dipole Antennas with Quality Factors Approaching the Chu Lower Bound," *IEEE Trans. on Ant. & Prop.*, Vol. 58, pp. 1898–1906, June 2010.
32. R. Ziolkowski and A. Eretok, "Metamaterial-Based Efficient Electrically Small Antennas," *IEEE Trans. on Ant. & Prop.*, Vol. 54, pp. 2113–2129, July 2006.
33. Majid Manteghi, "Non-LTI Systems, a New Frontier in Electromagnetics Theory," *IEEE Inter. Symp. On Ant. & Prop. Proc.* (Toronto), Paper 425, July 2010.
34. T. Yang, W. A. Davis, and W. Stutzman, "Fundamental-Limit Perspectives on Ultrawideband Anennas," *Radio Science*, Vol. 44, 2009.
35. H. Stuart and S. Best, "Limitations on Relating Quality Factor to Bandwidth in a Double Resonance Small Antenna," *IEEE Trans. on Ant. & Wireless Prop. Letters*, Vol. 6, pp. 450–463, 2007.
36. T. Yang, W. A. Davis, W. Stutzman, S. Hasan, R. Nealy, C. Dietrich, and J. Reed, "Antenna Design Strategy and Demonstration for Software-Defined Radio," *Analog Integrated Circuits and Signal Processing - An International Journal*, Vol. 69, pp. 161–171, 2011.
37. A. D. Wunsch, "A Closed-Form Expression for the Driving-Point Impedance of the Small Inverted L Antenna," *IEEE Trans. on Ant. & Prop.*, Vol. 44, pp. 236–242, Feb. 1996.
38. A. Gobien, "Investigations of Low Profile Antenna Designs for Use in Hand-Held Radios," Master's Thesis, Virginia Tech, 1997.
39. M-C Huynh and W. Stutzman, "Ground Plane Effects on Planar Inverted-F Antenna (PIFA) Performance," *IEE Proc.-Microw. Ant. Prop.*, Vol. 150, pp. 209–213, Aug. 2003.
40. R. Feick, H. Carrasco, M. Olmos, and H. Hristov, "PIFA Input Bandwidth Enhancement by Changing Feed Plate Silhouette," *Elec. Letters*, Vol. 40, 22 July 2004.
41. H. Morishita, Y. Kim, and K. Fujimoto, "Design Concept of Antennas for Small Mobile Terminals and the Future Perspective," *IEEE Trans. on Ant. & Prop. Mag.*, Vol. 44, pp. 30–43, Oct. 2002.
42. M-C Huynh and W. L. Stutzman, "Wideband Compact Planar Inverted-F Antenna," U.S. Patent 6,795,028, Sept. 21, 2004.
43. M-C Huynh and W. L. Stutzman, "A Low-Profile Compact Multi-Resonant Antenna for Wideband and Multi-Band Personal Wireless Applications," *IEEE Trans. on Ant. & Prop. Soc. Inter. Symp. Proc.*, pp. 1879–1882, June 2004.
44. H. W. Ott, *Compatibility Engineering*, Wiley, New York, 2009.
45. K. Finkenzeller, *RFID Handbook*, 2nd ed. Wiley, New York, 2003.
46. C. A. Balanis, *Advanced Engineering Electromagnetics*, Wiley, New York, 1989, pp. 327–328.
47. R. K. Mongia and P. Bhartia, "Dielectric Resonator Antennas—A Review and General Design Relations for Resonant Frequency and Bandwidth," *Inter. J. Microwave and Millimetre-Wave Computer-Aided Eng.*, Vol. 4, pp. 230–247, 1994.
48. S. A. Long, M. W. McAllister, and L. C. Shen, "The Resonant Cylindrical Dielectric Cavity Antenna," *IEEE Trans. on Ant. & Prop.*, Vol. 31, pp. 406–412, March 1983.
49. M. Huynh and W. Stutzman, "A Review of Radiation Effects on Human Operators of Hand-Held Radios," *Microwave J.*, Vol. 47, pp. 23–42, June 2004.
50. FCC OET Bulletin 56, August 1999.
51. C. Johnson and A. Guy, "Non-ionizing Electromagnetic Radiation," *Proc. IEEE*, Vol. 60, pp. 692–718, June 1972.
52. Y. Huang and K. Boyle, "Antennas from Theory to Practice," Wiley, 2008.

PROBLEMS

11.2-1 Show how the normalized principal plane patterns of (11-6) follow from (11-5). Discuss the electric field polarization in each plane.

11.2-2 A square microstrip patch with $L = W = 4.02$ cm is printed on an 0.159-cm-thick substrate with $\varepsilon_r = 2.55$. Find the resonant frequency, input impedance at resonance for an edge feed, and the bandwidth.

11.2-3 For the rectangular MSA of Example 11-1: (a) Verify the patch length. (b) Calculate the input impedance. (c) Calculate the bandwidth. (d) Plot the E- and H-plane polar patterns. (e) Estimate the directivity. (f) Use (11-2) to (11-4) to find the resonant patch length; compare to the value in (a).

11.2-4 Calculate the length of a rectangular patch at resonance at 3 GHz for a substrate of dielectric constant 2.2 and variable thickness from 0.001 to 0.05 λ. The patch width is a half-wavelength. Do this twice for the two levels of approximation in (11-1) and (11-2). Compare using a graph.

11.2-5 An edge-fed rectangular ($W = 1.5\ L$) microstrip patch antenna has a substrate of thickness 0.01λ. (a) Vary the dielectric constant to find the value which gives the highest bandwidth and state the values of ε_r and B. Using these values: (b) Calculate the patch length and width at a resonant frequency of 3 GHz, and (c) Calculate the input impedance.

11.2-6 A square microstrip patch antenna has measured resonant frequency of 2.28 GHz and has a substrate of dielectric constant 2.55 and thickness of 0.159 cm. (a) Calculate the resonant frequency for a patch length of 4.02 cm and compare to the measured value. (b) Calculate the input impedance at the measured resonant frequency and compare to the measured value of 400 Ω.

11.2-7 A probe-fed rectangular patch antenna of 0.3-cm-thick dielectric of 4.53 dielectric constant is 1.74 cm long and 2.31 cm wide. The resonant frequency is 3.72 GHz. (a) Calculate input impedance in Ohms for an edge feed. (b) Calculate input impedance for a probe feed 0.55 cm in from the edge.

11.2-8 A probe-fed rectangular patch antenna of 0.127-cm thick dielectric of 4.53 dielectric constant is 13.97 cm long and 20.45 cm wide. (a) Calculate the resonant patch length at 660 MHz using the simple formula. (b) Calculate the input impedance for an edge feed. (c) Calculate the input impedance for a probe feed 0.635 cm in from the edge.

11.2-9 A probe-fed square microstrip patch antenna on a 0.3-cm thick substrate of $\varepsilon_r = 4.53$ is to be designed to operate at 3.72 GHz. (a) Find the patch length at resonance. (b) Find the probe location for 50-Ω input impedance.

11.2-10 An inset-fed rectangular microstrip patch antenna with dimensions $L = 4.04$ cm, $W = 5.94$ cm, and $t = 0.127$ cm has a substrate of $\varepsilon_r = 2.42$. Find the inset distance required for an input impedance of 50 Ω.

11.2-11 A commercial microstrip patch antenna has HP beamwidths in both principal plane of $100°$ and a radiation efficiency of 70%. Calculate the gain in dB.

11.2-12 Derive the pattern of a rectangular MSA in (11-5) starting with image theory for a magnetic current above an electric ground plane as in Fig. 9-4c and (9-2b), as well as (9-26).

11.3-1 For the 16-element (4×4) array of Example 11-2. (a) Plot one of the principal plane patterns using (11-6a) for the element pattern. (b) Compute the directivity by pattern integration assuming the element pattern applies to all space, and then find the gain.

11.3-2 A linear array of four quarter-wave microstrip patches spaced 4.2 cm apart are to be excited to produce an endfire beam. The operating frequency is 2.73 GHz and the substrate relative dielectric constant is 2.45. Sketch the corporate feed network required to accomplish this, indicating on it the extra microstrip transmission line lengths in cm required for proper phasing.

11.4-1 Calculate the power remaining at the end of the leaky wave antenna used in Fig. 11-14a. Repeat for Fig. 11-14b. Assume the length of the antenna is 5λ at 6.25 GHz.

11.5-1 (a) An electrically small antenna has $Q_A = 60$. The smallest sphere that can enclose the antenna has a radius of $0.0159\ \lambda$. Find the antenna gain in dB. (b) Repeat (a) if the antenna Q_A is 40 and the radius is $0.0318\ \lambda$. (c) What is the directivity of these antennas in dB? (d) Explain what negative dB gain means.

11.5-2 Show that the radius of a radiansphere, $\lambda/2\pi$, corresponds to the distance from an ideal dipole where the power density contained in the near field equals that in the far field in the direction of maximum radiation, $\theta_o = 90°$.

11.5-3 Derive the formulas for R_r/R_o for (a) Short dipole, (11-31), and (b) Small loop, (11-32).

11.5-4 Use (11-27), (11-28), and (11-29) to plot curves of Q_A vs. βa on a single graph with a log-scale ordinate from 1 to 10^4 and a linear-scale abscissa from 0 to 1.

11.5-5 Use (11-40) to plot curves of Q_A vs. βa on a single graph with a log-scale ordinate from 1 to 10^4 and a linear-scale abscissa from 0 to 1 for radiation efficiency values of $e_r = 1$, 0.5, 0.2, and 0.1.

11.5-6 For the dipole of Fig. 6-7 with a 0.1-mm radius, calculate the ratio of the half-power (VSWR = 5.83) bandwidth to the VSWR = 2 bandwidth and compare to that predicted using (11-34).

11.6-1 *Simulation*. Use a moment method code to simulate the 13.25-turn NMHA discussed in Sec. 11.6.1 that is 10 cm tall and 2 cm in diameter constructed with copper wire 0.18 mm in diameter. Give the input impedance and gain values at 150 MHz.

11.6-2 *Simulation*. Use a moment method code to simulate the NMHA of Example 7-1 at 883 MHz using a wire diameter of 1 mm and operated over a perfect ground plane. Give the input impedance and gain.

11.6-3 *Simulation*. The meander line monopole of Fig. 11-21 for the following dimensions: $s_V = 1.43$ cm, $s_H = 3.33$ cm, and $w = 2$ mm. Use a simulation code to compute the input impedance to see if the resonant frequency is 344 MHz. Model the flat conductor with equivalent (perfectly conducting) wires of diameter equal to half the conductor width (see Sec. 6.1.1).

11.6-4 Consider the example QHA discussed in Sec. 11.6.2 that has half-turn arms, a diameter of $D = 0.174\lambda$, and height of $h = 0.243\lambda$. (a) Calculate the helix parameters of S, α, L, and L_{loop}. (b) Compute the physical dimensions in cm for operation at 2.2 GHz.

11.6-5 *Simulation*. Use a simulation code to verify the values in Example 11-3 for the ILA garage door opener transmitter. Plot the elevation pattern and give the gain for copper wire.

11.6-6 *Simulation*. Use a simulation code to verify the values in Example 11-4 for the IFA antenna. Plot the elevation pattern and give the gain for 1-mm diameter copper wire.

11.6-7 The PIFA of [39] has $L = W = 2$ cm. Calculate the resonant frequency for an infinite ground plane and compare to the measured value of 1.885 GHz.

11.6-8 The RFID tag using the antenna of Fig. 11-32 can respond to -17 dBm of power. If the reader has 36 dBm EIRP and the tag antenna has 0 dB gain, calculate the passive read distance. Assume free space propagation conditions and a frequency of 900 MHz.

11.7-1 Calculate the plane wave reflection coefficient for and air-dielectric interface when $\varepsilon_r = 9$, 16, 25, 36, 49, 64, 81, and 100. Plot the results.

11.7-2 Determine the radius and Q of the design in Example 11-2 when $H = 0.2$ cm.

11.7-3 Design an $HE_{11\delta}$ mode circular cylinder DRA with 5% bandwidth at 10 GHz.

Chapter 12

Terminal and Base Station Antennas for Wireless Applications

An overview of wireless communications was presented in Sec. 4.5, including the conceptual drawing in Fig. 4-7 of the integration of wireless systems. In this chapter, we discuss wireless communication antennas used in practice. The types of wireless communication networks are shown in Fig. 12-1. If there are just two terminals as in Fig. 12-1*a*, it is called a *point-to-point link*. An example is two-way, high data rate, point-to-point communications using microwave frequencies. Microwave point-to-point links are used in backhaul systems to connect cellular base stations together. A simple example of a one-way, point-to-point link is a wireless garage door opener remote control system. A *base station* services multiple terminals. If the terminals are only receivers as in Fig. 12-1*b*, the system is referred to as *broadcast* communications. Examples include AM and FM broadcast radio and broadcast TV. If both the base station and the terminals transmit and receive as shown in Fig. 12-1*c*, the system is called *point-to-multipoint* communications. Cellular telephone is an example. If the base station serves a small region such as a building, it is referred to as a *wireless access point (WAP)*. Fig. 12-1*c* indicates the terminology used to describe directional links. The terms *uplink* and *downlink* are referenced to the base station, indicating that signals are traveling up to or down from it. The terms *forward link* and *reverse link* (also called a return link) are referenced to the user terminal and indicate signal directions that are forward to the base station from the user or reverse from the base station to the user. Another type of network consists of just multiple terminals, each capable of communicating with any other terminal that is within range, which is referred to as a peer-to-peer (or mesh) network. Walkie-talkie radios are an example.

When setting up a new wireless communication system, there is a set of steps followed in the design process. After the system and antenna specifications are clearly defined, the operating frequency range is selected. Regulatory agencies restrict uses by frequency and also limit the power output and sometimes the antenna pattern and gain.[1] The next

---

[1] A chart of all allocated frequencies for the U.S. is found at http://www.ntia.doc.gov/osmhome/allochrt .pdf, and the FCC rules and regulations are found at http://fcc.gov/.

(*a*) Point-to-point.

(*b*) Broadcast.

(*c*) Point-to-multipoint.
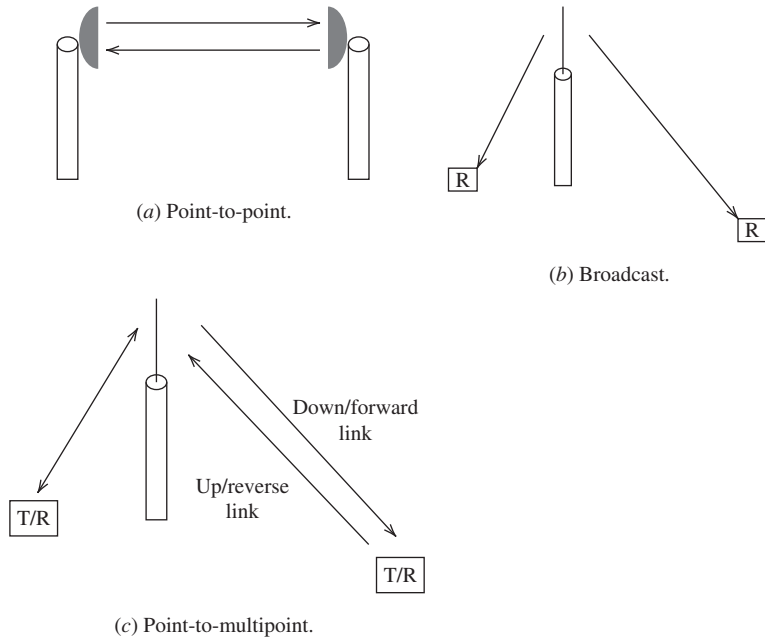
Figure 12-1 Wireless communication network configurations.

decision is whether to use a frequency band that requires a license from the regulatory agency ("licensed") or a band that is license exempt ("unlicensed"). *Licensed operation* requires an application and approval but affords protection against unauthorized users. *Unlicensed operation* avoids the licensing process, and affordable equipment is available, but interference and competition for bandwidth can be issues. Propagation is also a large factor in the frequency selection decision. Frequencies at VHF and below offer long-distance communication, even over *non-line-of-sight* (NLOS) paths. Frequencies at UHF and above are used for applications requiring wide bandwidth. Above 10 GHz, a clear *line-of-sight* (LOS) path is required, and the atmosphere can cause attenuation and depolarization; see Sec. 4.5.2. Once a frequency is chosen, the antenna selection process begins. This starts by identifying the relevant system specifications listed in Table 1-3. The first choice to be made is between an array and a single antenna. Next, Table 1-4 is used as a guide in finding the appropriate antenna or array element. A list of candidate antennas is prepared, followed by a down selection process that eliminates candidates after preliminary investigations. The final candidate antennas are evaluated in detail using simulations and experimental models. Final selection is made based on comparisons of the results from the evaluation phase that often includes full system tests.

12.1   SATELLITE TERMINAL ANTENNAS

Although more than 90% of the world's population has cellular telephone service available, 82% of the world's surface does not have cellular coverage. Whereas terrestrial-based communications are not practical for reaching remote locations and oceans, satellite-based communications are. Due to the long distances between Earth and space, satellite links require a clear line-of-sight to operate. Earth terminals can be fixed, portable, or mobile. For wide bandwidth communications, a fixed Earth terminal is used that most often has a reflector antenna and sometimes a mechanically steered fixed-phase array or a phased array. If the satellite is not geostationary, the terminal requires a steering mechanism to continuously maintain the antenna main beam direction aimed toward the satellite, or a broad beam antenna is used.

Reflector antennas are very popular for fixed terminals but are also used on vehicles such as ships and aircraft with a means of tracking the satellite during vehicle motion. Cassegrain reflectors are commonly used for Earth terminals because of their low noise due to lower ground pickup from spillover compared to a prime focus reflector (see Secs. 4.3 and 9.6). Also, the antenna must have low side lobes to avoid interference with and from adjacent satellites that typically are 2° apart. Gateway terminals (see Fig. 4-7) communicate many channels of information occupying wide bandwidth using reflector antennas several meters in diameter. User terminals have narrower bandwidth, allowing smaller antennas.

Fixed two-way Earth user terminals use large reflector or array antennas. However, *direct broadcast satellite* (DBS) or *direct-to-home* (DTH) systems have special features that allow the antenna to be small and the whole Earth terminal to be low cost. First, the geostationary broadcast satellites are spaced about 9° apart instead of 2°. This permits a wider beam antenna, which allows an antenna diameter as small as 0.5 m in diameter. Low antenna gain is possible because of the high satellite transmit power and video compression, which allows for reduced bandwidth (and thus more channels). See Examples 4-3 and 4-4 for link calculations.

Circular polarization is popular for satellite communications for two reasons. First, at UHF frequencies and below, Faraday rotation effects lead to variable tilt angle of linearly polarized waves as conditions in the ionosphere change, requiring the Earth terminal to track the polarization angle of the incoming wave. Circular polarization is unaffected by Faraday rotation. Second, circular polarization used on satellites that spin for stabilization would not have a polarization state alignment problem due to satellite body rotation.

Mobile satellite terminals present challenges. Two-way mobile terminals must track the satellite using a medium-sized antenna to point in the satellite direction to compensate for vehicle motion. Mobile Earth terminals on small vehicles or that are hand carried must have small antennas, so an antenna with a broad beam pointed upward is used. Earth terminals that only receive can have very small antennas. Position location such as GPS and digital audio radio are examples. The quadrifilar antenna of Fig. 11-22 is popular for satellite communication terminals with circular polarization. For non-geostationary satellite constellations such as GPS, the pattern is often shaped to have more gain at an angle off zenith ($+z$-direction) for better coverage of satellites that are not directly overhead and more distant. See [H.9.1: Fujimoto, 2nd ed., Chap. 11] for more details on mobile satellite antennas.

An example of a commercial system using portable satellite Earth terminals is Omni-TRACS, used by long-distance freight vehicles for two-way transfer of data. The Ku-band (14/12 GHz) antenna is a sectoral horn (see Sec. 9.4) with a broad beam in elevation so that elevation tracking to the geostationary satellite is not required. Azimuth tracking is by mechanical rotation of the horn attached to a rotary joint. The white bubble radome housing the antenna and RF system can be seen on truck cabs.

The antenna choice for a satellite depends on the type of satellite. For spinning satellites, omnidirectional patterns are required. Candidate antennas are monopoles, quadrifilars, or elements disposed around the perimeter of the body. High-gain antennas can be used on spinning satellites with a de-spun platform or on body-stabilized spacecraft. Reflectors, horns, slots, microstrip elements, and arrays are used in the higher frequency bands, whereas dipoles and helices are used at lower frequencies.

12.2   BASE STATION ANTENNAS

A *base station antenna* carries multiple signal channels and has an antenna pattern that covers a defined region around the base station.[2] The type of antenna used depends on the size scale of the cell. Fig. 4-7 shows the various cell sizes. In cellular communications, channels of the same frequency can be reused in cells that are spatially separated. Fig. 12-2

---

[2] The term base station is used here to mean a terrestrial base station and not include air, space, etc. platforms.

shows spatial coverage (i.e., where the radiation is strong enough for effective communication) of the main types of base stations for a single cell. The omnidirectional pattern of Fig. 12-2a provides uniform radiation around the tower in azimuth (the horizontal plane). Early cellular communications systems used these exclusively, but most have been replaced by antennas which cover a specific angular region in the horizontal plane, except for rural areas and in-fill applications. The most common cellular base station antenna has the sector shape pattern of Fig. 12-2b, which enables serving more user terminals. The most flexible configuration is the smart base station antenna of Fig. 12-2c that adjusts the directions and shapes of beams according to traffic needs. Smart antennas are discussed further in Sec. 12.4.

The omnidirectional antenna used in cellular telephone base stations has the appearance of a tall monopole, but in reality it is a vertical collinear array of dipole elements. Often there are three or more such antennas on the same mount to provide for spatial receive diversity and for separating transmit and receive functions. The elements in the array are fed in parallel, often with uniform amplitude to achieve high gain. Although the pattern is uniform in azimuth, it is directive in the elevation plane. A longer array will produce a narrower elevation pattern and thus higher gain and longer communication distance. However, a pattern that is extremely narrow will lead to loss of coverage near the tower. A linear phase taper across the array is used to produce what is referred to as *electrical downtilt* that tilts the beam downward slightly to reduce interference into adjacent cells. If the entire antenna is tilted mechanically (called *mechanical downtilt*), the donut pattern tilts down on one side and up on the opposite side, whereas electrical downtilt produces a cone around the antenna with the same angle of downtilt in all azimuth directions. Special construction techniques are used to realize the parallel feed. The feed lines are routed through the center of the vertical support mast to avoid having the cable disturb the radiated fields. The array analysis techniques in Chap. 8 are readily applied to base station arrays, as illustrated in following examples.

(a) Omnidirectional.

(b) Sectorized.

(c) Smart.
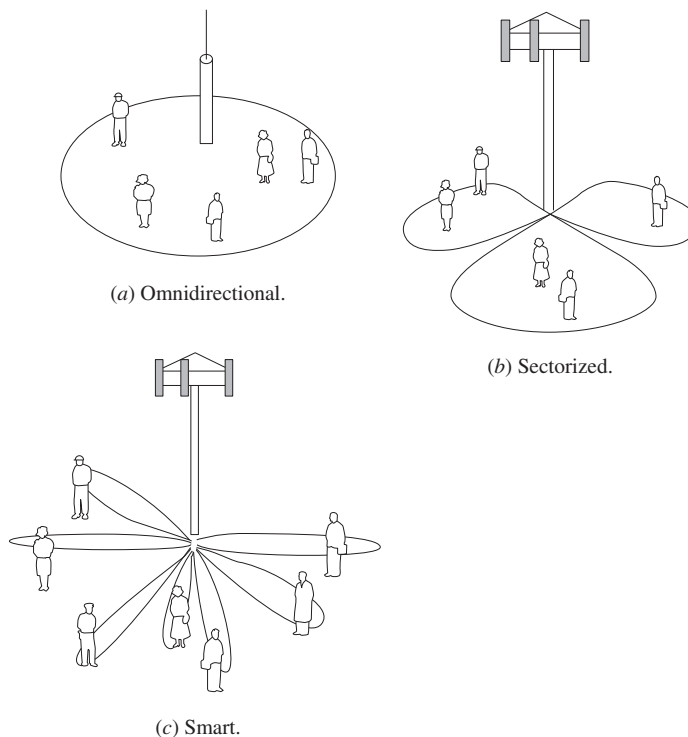
Figure 12-2 Base station antenna types.

(a) Elevation radiation pattern in dB. (b) Directivity versus element spacing.
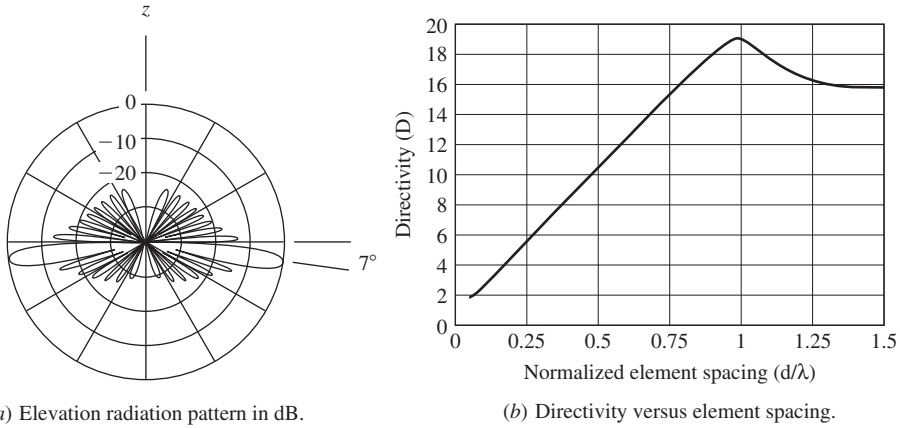
Figure 12-3 An omnidirectional base station antenna for 800–900 MHz that is a collinear array of 10 half-wave dipoles phased for 7° beam downtilt. See Example 12-1.

EXAMPLE 12-1   *An Omnidirectional Base Station Antenna for Cellular Wireless Communications*

A commercial omnidirectional base station antenna for operation from 800–900 MHz is constructed using in a collinear array of 10 half-wave dipoles spaced 28.5 cm apart ($d = 0.81\,\lambda$ at mid-band). The array factor is found from (8-22) for a uniformly excited, equally spaced linear array. The required inter-element phase shift for a 7° downtilt from (8-25) is $\alpha = -\beta d \cos(\theta_o) = -360 \times 0.81 \cos(97) = 35.5°$. The array factor for this array is multiplied by the element pattern of a half-wave dipole from (3-4), giving the full array pattern, which is plotted in Fig. 12-3a in the elevation plane. It agrees closely with the measured pattern. Note that the beam peak is 7° down from the horizontal plane as designed. The directivity of the array as a function of element spacing computed using (8-62) is plotted in Fig. 12-3b. Note that the directivity peaks for about one-wavelength spacing. The choice of $0.81\,\lambda$ spacing gives lower side lobes than when one-wavelength spacing is used. Also, at the top end of the band (900 MHz) the spacing is closer to one wavelength $d = 0.87\lambda$. The mid-band directivity from the plot is 16.5, or 12.2 dB, and equals the vendor's quoted gain value.

Omnidirectional antennas are in wide use in private mobile radio systems, including delivery trucks and municipal (police and rescue) vehicles. VHF and UHF collinear arrays of exposed folded dipoles mounted on towers or buildings are easily identified. An example of a popular base station antenna design follows.

EXAMPLE 12-2   *A Near-Omnidirectional Base Station Antenna for Land Mobile Radio*

Fig. 12-4a is a base station antenna with four half-wave folded dipoles; see Sec. 6.2 to review folded dipoles. This example considers a commercial UHF-band antenna operating at 460 MHz, where the folded dipoles are 32.6 cm long ($L = 0.5\,\lambda$) and the spacing is $d = 47$ cm $= 0.72\,\lambda$. With the elements mounted on one side of the tower, the tower acts as a reflector and enhances radiation in the direction of the elements, as indicated in the pattern of Fig. 12-4b. More directivity is achieved by decreasing the standoff distance from the tower, s. A directive pattern is preferred over an omnidirectional pattern for many applications. An example where a directive pattern is desired is when the tower is located on a hill just outside of a town. The antenna is oriented so that the beam peak is directed toward town, where most users are located. In this example, the pattern has beamwidths of $HP_v = 220°$ and $HP_h = 18°$.

(a) The array of four collinear
folded-dipole elements.

(b) The azimuth patterns for stand-off distances of
the elements from the tower of $s = \lambda/4$ (solid curve)
and $s = \lambda/2$ (dashed curve).
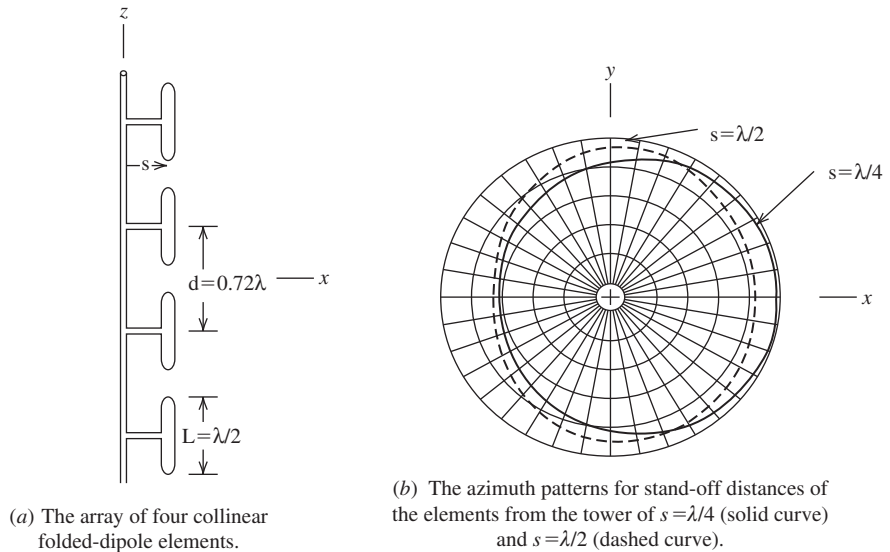
Figure 12-4 A base station antenna for land mobile radio in the UHF band. See Example 12-2.

Using the directivity estimation formula from (4-56), $D \approx 41{,}253/(220° \times 18°) = 10.2$ dB, which equals the vendor's quoted value. If the elements are replaced by pairs of opposing folded dipoles and alternate pairs are rotated 90°, the structure is symmetrical and thus the pattern is nearly omnidirectional. The directivity of the omnidirectional, four-element array obtained by pattern integration is $D = 7.8$ dB. With an elevation pattern beamwidth of $HP_v = 18°$ and using the directivity estimation formula from (4-56), $D \approx 41{,}253/(360° \times 18°) = 8.0$ dB, which equals the vendor's quoted value.

The most common cellular base station uses a *sector pattern* to cover a sector of angular space rather than having uniform coverage around the tower. A typical sector antenna arrangement with a triangular tower-top structure uses *panel antennas* on each face that cover one 120° sector, as shown in Fig. 12-3b. Common values for the half-power beamwidth in the azimuth plane of a 120°-sector antenna are 65° and 90°. Fig. 12-5 shows typical panel antenna azimuth radiation patterns with 65° and 90° beamwidths. Note that both patterns substantially cover a 120° sector and that the front-to-back ratios are 30 and 25 dB. Unequal sectors are used when users are not uniformly distributed around the base station, such as near a major roadway where users are concentrated over limited angular
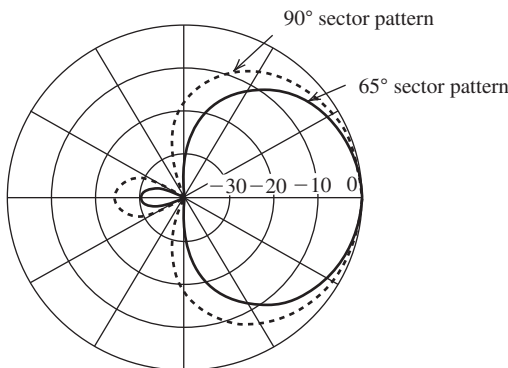
Figure 12-5 Representative azimuth radiation patterns of a 120° sector antenna with 65° (solid line) and 90° (dashed line) beamwidth cases.
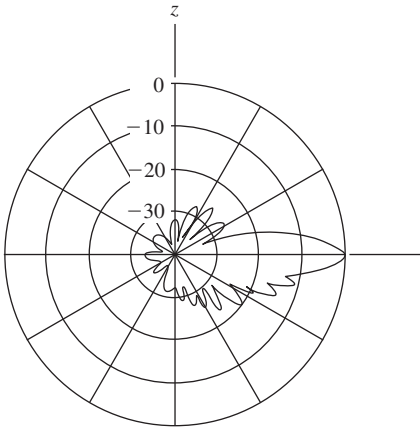
Figure 12-6 Elevation pattern for a sector antenna with nonuniform excitation to shape the beam for improved coverage below horizontal and lower side lobes above.

regions. Similar to the omnidirectional base station antenna, a sector antenna also uses a vertical array of elements to produce a narrow elevation beam to increase gain. The smallest vertical beamwidth used in practice is about $HP_v = 4°$. The corresponding practical upper bound on gain for a $65°$-panel antenna from (4-57) is $26,000/HP_h HP_v = 26,000/(65 \times 4) = 100 = 20$ dB. Mechanical downtilt is used in combination with electrical downtilt with panel antennas. The downtilt angle adjustment is a balancing process. Small downtilt angle with the beam peak nearly horizontal provides longer distance communication, and large downtilt angle reduces interference into cells that reuse the same frequencies. There are other pattern considerations as well. Reduced radiation above the horizontal direction is desired to reduce interference and increase directivity. Uniform excitation produces the highest directivity but will have a pattern with deep nulls that could produce lost communication. Nonuniform excitation is used to shape the beam, especially below the horizontal direction to fill in the nulls (referred to as "null-fill"). Pattern shaping is also used to lower the radiation above the horizontal where coverage is not needed. The synthesis techniques of Sec. 10.3 are used for beam shaping, often employing a nonlinear phase component of the element currents in addition to the linear phase shift for downtilting. Fig. 12-6 shows an example of a shaped elevation pattern that is typical for a sector antenna. Comparing to Fig. 12-3a, which is a typical uniform amplitude and phase pattern, the deep nulls below horizontal are filled in and the side lobes above horizontal are lowered.

Several types of antennas are used as elements in panel antennas. General specifications on panel antennas that apply over the bandwidth are nearly constant azimuth pattern, good impedance match (VSWR < 1.4 or RL > 16 dB), and *front-to-back ratio* (pattern peak/pattern value in opposite direction) >30 dB. There are several forms of broadband dipoles that are commonly used as elements, including dropping dipoles and dipoles with four wide, metallic wings (called a *twin dipole* antenna). [H.6: Balanis, Ed., p. 1177] Patch antennas are used without dielectrics for high-power handling and low loss. Log-periodic antennas are also used as elements in both planar and trapezoidal forms (see Sec. 7.8). All panel antennas have a ground plane to reduce back directed radiation. For dipoles and patches, the ground plane is required to produce a unidirectional beam, as explained in Example 3-5 for a dipole that is a quarter-wavelength away from a ground plane. Ground planes are also used in panel antennas to back elements that would not normally require a ground plane, such as a log-periodic antenna to achieve very low back radiation; i.e., high front-to-back ratio. The ground plane size and shape in the azimuth plane are varied to control the azimuth pattern shape. [H.6: Balanis, Ed., p. 1163]

Electrical downtilt is achieved using phase shifters that are implemented by changing the feed line length or by varying the amount of dielectric in the transmission path.

Probably the most common phase shifter uses concentric arcs, each with two outputs to a pair of elements on each side of the array in a parallel configuration with a sliding contact to vary the electrical path length. The path length through each arc changes equally to both elements but as an increase to one element and as a decrease to the other, giving a differential phase shift; the phase change is more for the larger conducting arc path. [H.9.1: Chen and Luk, p. 75] Another popular differential phase shifter that simultaneously increases the line length on one side of the phase shifter while decreasing it on the other uses an inline mechanical device. Also used is a phase shifter that inserts a dielectric slug between the conductors of a transmission line. Phase is adjusted manually at the antenna or, more commonly, using remote control means. In addition to elevation beam downtilt adjustment, phase control is also used for azimuth beam steering and shaping. To respond to rapidly varying user traffic conditions (essentially, the number of users in each sector), antenna pattern changes can be made dynamically, which falls into the area of smart antennas (discussed in Sec. 12.4).

A special consideration for base station antennas is self-interference between the multiple channels that are transmitting at the same time. This *passive intermodulation* (*PIM*) results from the mixing of two carriers at frequencies f_1 and f_2 due to the presence of a nonlinearity. The third-order intermodulation products often produce the worst PIM. An example third-order product occurs at the frequency of $2f_2 - f_1$, which can fall within the operating band because the f_1 and f_2 carriers are close in frequency. PIM not only occurs within a single antenna but also between nearby antennas. For example, when transmit antennas for different wireless services are mounted on the same tower (or a nearby tower), they can create PIM, which is referred to as *co-site interference*, more generally called electromagnetic interference (see Sec. 4.7). There are several sources of PIM. At RF frequencies, current is confined to the surface of a conductor because of the skin effect. Thus, oxide layers (aluminum is a common offender) can cause rectification leading to PIM, which is called the "rusty bolt" effect. Protection of the antenna from weathering (i.e., rusting) is a key consideration in base station installations. Other sources of PIM are dissimilar metal contacts, poor solder joints, and ferromagnetic materials (steel and nickel). Metal-to-metal contacts can produce PIM and should be minimized, but when used, solid compression of contacting surfaces is required. Non-contacting (i.e., capacitive) coupling and insulation between metals are solutions. Details on good practice principles for mechanical joints are found in [H.9.1: Chen and Luk, p. 71]. Connectors commonly cause PIM problems. In general applications, gold-plated mating surfaces in connectors are used to reduce PIM and loss because of gold's low oxidation property. Manufacturing techniques that employ stamping and folding of antenna and feed line parts are used to avoid metallic joints that generate PIM. Some environments are very challenging, such as shipboard installations, where there are many metallic structures near a transmitting antenna. Manufacturers usually test all base station antennas for PIM specification compliance before shipping. Testing is performed by injecting two 20-watt carriers at f_1 and f_2 and using a network analyzer to sweep over the operating band to make sure all spurious signals are -153 dB below the carrier level. Thus, with the total input power of 43 dBm (2×20 W), the spurious signal levels should be below -110 dBm.

Dual-polarized antennas allow for diversity using one dual-polarized receiving panel antenna instead of two separate antennas using spatial diversity. One of the most popular configurations is dual orthogonal slant $45°$ linear polarizations. Dual-band panel antennas add a further significant savings in antenna hardware and associated wind loading and visual impact. The most compact deployment uses panels that are both dual polarized and dual band. Dual-band arrays can be constructed using two interleaved arrays, one covering the low bands and one covering the high bands. The element count can be reduced further using wideband elements in an array covering both bands with interleaved (and smaller) elements to cover only the high bands that are required, because a closer physical spacing is needed to achieve about $0.8\,\lambda$ electrical spacing. One such candidate dual-polarization, dual-band element is the Fourpoint antenna that was discussed in Sec. 8.11

and is shown in Fig. 12-7a. The orientation in Fig. 12-7a gives the most popular dual-polarization mounting of +/−45°. The antenna is 11.4 cm wide, which is much smaller than a conventional panel antenna of width $12'' = 30.5$ cm that covers the low band of 824–960 MHz. The Fourpoint antenna is fed against a ground plane that is 6.4 cm away (about 0.2λ and 0.4λ at the center of the 824–960 MHz and 1710–2170 MHz bands, respectively). Simulation and measured results in Fig. 12-7b show a 2.7:1 bandwidth for VSWR < 2 and two very low-VSWR bands centered on the low and high cellular bands. [1, 2]

Base station antennas are mounted in a variety of ways. So-called monopole towers are popular because of low visual impact. Nearly invisible installations are finding acceptance with zoning administrations. These include mountings inside wooden building structures, such as church steeples, and towers that replicate coniferous trees with placement near a coniferous tree stand. Coincident installation with power and light utility poles are also well received. It is important to hide cable runs to the tower top, too.
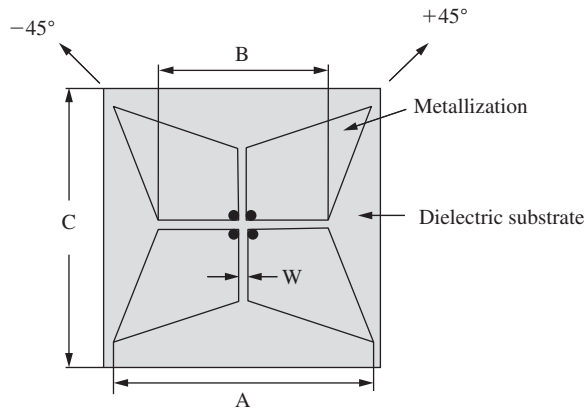
(a) Top view of the Fourpoint antenna. The two slant linear polarizations of +45° and −45° are oriented as shown, each generated by the feed terminal pairs at the center in the respective planes.
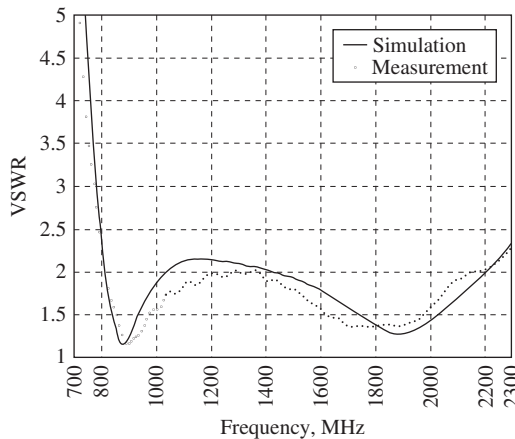
(b) Measured and simulated VSWR for the Fourpoint element.

Figure 12-7 The Fourpoint antenna is a candidate element for dual-polarized, dual-band base station panel antennas. The dimensions are $A = 11.4$ cm, $B = 9.5$ cm, $C = 11.7$ cm, $W = 2$ mm, and 6.4 cm deep. (Used with permission from Suh et al. [2] © 2004 IEEE.)

12.3   MOBILE TERMINAL ANTENNAS

Mobile communications developed around 1900, soon after Marconi demonstrated that long-distance wireless communications was possible. At first, only trains and ships could handle the large mobile communications equipment. A prominent example is the top-loaded monopole in the shape of a T used on the Titanic ocean liner. The top-loading wires spanned most of the length of the ship and were 62 m above the water. The shipboard radio was able to send out distress signals to summon aid before sinking after hitting an iceberg on April 14, 1912. Today, of course, antennas are found on mobile vehicles and devices of all types. In this section, we treat antennas used on moving platforms.

The size constraints placed on antennas are usually significantly less for vehicle-mounted antennas compared with handheld devices. Many vehicles have a natural ground plane available in the form of a roof, fender, or other metal body panel. This makes monopoles attractive solutions. At low frequencies (VHF and below), portable and mobile terminals often use an electrically small antenna (see Sec. 11.5) such as the stubby and quarter-wave helical (see Sec. 11.6.1). These two normal mode helix antennas (NMHA) are always options at any frequency band where small height is needed. These provide a comparison point to start examining simple wire antennas. We begin by reviewing the ideal dipole with length much less than a wavelength. It has directivity from (2-158) of 1.76 dB. The ideal monopole of height $\ll \lambda$ with a perfect ground plane has 3 dB more directivity because the radiation is confined to half the radiation sphere; so the directivity is 4.76 dB. The stubby antenna, which is an NMHA with a wire length $< \lambda/4$, has a directivity of 4.5 dB. The quarter-wave NMHA (see Example 7-1) has a directivity of 4.7 dB. These antennas are used on portable handheld devices with minimal ground planes and the directivity is 3 dB smaller in theory. In practice, the gains are around 0 dB. More gain[3] is obtained by increasing the antenna size.

Table 12-1 summarizes popular vehicular antennas in order of increasing size. The quarter-wave whip is one of the leading candidates because of its good input impedance of about $35 + j0 \ \Omega$ when resonated and above a perfect ground plane, and for its simple structure. Its height is acceptable in many cases: 7.5 cm at 1 GHz and 75 cm at 100 MHz. It has a gain of 5.15 dB, which is 2.15 dB for the half-wave dipole formed with the image in the ground plane plus 3 dB due to use of only half the solid angle. The quarter-wave whip is frequently seen on the roofs of fleet vehicles such as police and emergency cars. In the higher frequency bands (high UHF and above), physical size limitations are less of a problem, opening up more antenna options such as half-wave wire antennas. The end-fed half-wave whip and center-fed half-wave dipole are relatively independent of the ground

Table 12-1 Land Mobile Vehicular Antennas

| | | | Gain (dBi) | |
| | | | Perfect ground plane | Typical vehicle mount |
Antenna	Figure No.	Height, h		
Quarter-wave whip	6-47b	$\lambda/4$	5.15	2
Half-wave end-fed whip		$\lambda/2$	N/A	2.4
Half-wave center-fed dipole	3-5	$\lambda/2$	N/A	2.4
Three-quarter-wave monopole with stub	6-45	$3\lambda/4$	8.6	5
Five-eighths-over-quarter-wave monopole	12-8	$7/8\lambda$	8.1	5

---

[3] Gain includes ohmic losses. Except for electrically small antennas, gain can be taken to equal directivity for wire antennas.

plane, making them good choices for small (or no) ground plane applications, such as on handheld radios. The obvious disadvantage is that they are twice as long as a quarter-wave whip. Also, the *end-fed half-wave whip* has high input impedance and thus requires a matching network. The center-fed half-wave dipole is well matched but has to have a feed line from the lower end to the center feed point, or realized as a sleeve dipole (see Fig. 7-22c). The three-quarter wave monopole with stub of Fig. 6-45 has in internal phase-reversing stub that brings the current of the upper section in phase with that of the lower wire segment as well as loading coil at the base. The added wire length increases the gain to 8.6 dB when over an infinite ground plane. The most common variation of this antenna is *five-eighths-over-quarter-wave monopole*, shown in Fig. 12-8, which has a longer upper section, and the coil is moved to the stub location and acts as the phase

(a) The antenna geometry.

(b) The current distribution with the coil unwound for plotting simplicity.
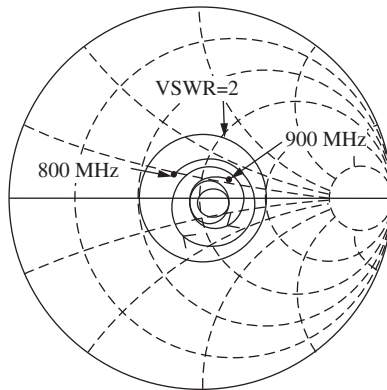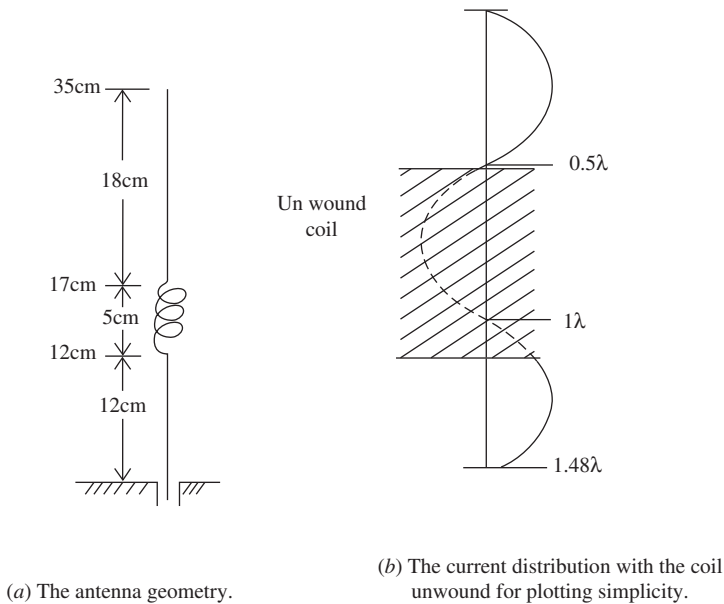
(c) The measured input impedance swept from 800 to 900 *MHz* on a Smith chart. The center of the chart is 50 + j0 Ω.

Figure 12-8 The 5/8-over-1/4 wavelength wire antenna of Example 12-3.

reverser. This antenna was popular for cellular phone antennas mounted on the rear window glass of automobiles. It remains popular as a rooftop mount vehicular antenna, with the following being an example.

EXAMPLE 12-3   *A 5/8-over-1/4 Wavelength Monopole Antenna*

Fig. 12-8 shows a commercial 5/8-over-1/4 wavelength monopole antenna for the 824–894 MHz cellular band with a mid-band wavelength of $\lambda_c = 35$ cm at 859 MHz. The commercial model of Fig. 12-8*a* is composed of the three sections with a total height of 35 cm $= 1\,\lambda_c$, and the electrical lengths of the upper and lower sections are $h_U = 0.51\,\lambda_c$ and $h_L = 0.34\,\lambda_c$. The antenna is somewhat longer than the nominal design height of $0.875\,\lambda_c = 30.6$ cm. The approximate current distribution, assuming a sinusoidal shape, is shown in Fig. 12-8*b* with the coil unwound for illustration purposes. The cross hatching denotes the non-radiating current in the coil. The fields created by opposite sides of the coil cancel in the far field. Note that the upper and lower straight wire section currents are in phase. This design produces an excellent impedance match (VSWR < 2) over a greater bandwidth than the operating band of 70 MHz, or 8.1%, as shown in the Smith chart in Fig. 12-8*c*. This impedance was measured with a spectrum analyzer swept from 800 to 900 MHz and spirals around the chart center of $50 + j0\ \Omega$. The gain is 5.2 dB. Simulations give an impedance of $52 + j5\Omega$ at mid-band and a gain of 8.1 dB.

Antennas mounted on vehicles are affected by the vehicle itself, the location on the vehicle, and frequency. At low frequencies, the antenna simply acts to excite the vehicle as a radiator. Broadcast AM/FM radio reception at 1 MHz in a car is a good example. The 31-inch fender-mount short monopole was the standard AM car antenna for decades, but has been replaced by in-glass antennas. Example 2-4 showed that the fender whip antenna is only 7% efficient at AM; however, it is a quarter-wavelength long at FM and is very efficient. Early rear-window units used the defroster element for an antenna, but current practice is to have the AM/FM antenna separate from the defroster. [H.6: *Ant. Eng. Hdbk.*, 4th ed., Chap. 39] Whereas AM antenna performance is not sensitive to vehicle shape details, FM broadcast radio reception can depend on vehicle details such as the passenger compartment, which is near resonance at 100 MHz. Thus, it is important to simulate or measure the antenna on a full vehicle model. Fig. 12-9 shows such results for an FM rear window glass antenna. [3] The side glass of sport utility vehicles is used for in-glass antennas. Antenna placement on a vehicle at HF and VHF frequencies follows the general principle of the horizontal-plane pattern elongating in the directions of the most metal; see [4] for measured patterns of antennas on vehicles from 30 to 830 MHz. The pattern of a roof-mounted antenna is slightly forward-rearward elongated, and a trunk deck mount antenna pattern is forward elongated. At high frequencies (UHF and above) such as the cellular band, only the materials near the antenna are important, so a center roof-mounted antenna behaves much like an infinite ground plane. Automobile-mounted cellular telephone antennas perform well. Handheld phones, however, experience about 10 dB loss when used inside a vehicle. [5]

Land vehicles have many other wireless applications besides radio reception and cellular telephone. The term *telematics* is used to describe all such applications and includes the following:

- Satellite navigation (e.g., GPS)
- Remote keyless entry (RKE)
- Tire pressure monitoring system (TPMS)
- Satellite radio reception (SDARS)
- Terrestrial TV reception
- Short-range communication (DSRC)
- In-vehicle short-range data (e.g., Bluetooth)

- Electronic toll collection
- Collision avoidance
- Intelligent cruise control

Antennas for used for satellite navigation (for example, GPS that uses RHCP) are the microstrip patch of Sec. 11.2 and the quadrifilar antenna of Sec. 11.6.2. Experiments have shown that rear roof mounting for a GPS antenna provides a good compromise for all satellite view angles. [6] There are many mobile applications for satellite navigation besides the most familiar one of driving guidance. Driver assistance communications services like OnStar send on-vehicle GPS-derived location information over the cellular network to a central processing location, from where directions and commands are sent back to the vehicle via cellular telephone link. One interesting GPS application is in agriculture, where harvesting and planting equipment are instrumented for sensing and position location. This assists in avoiding gaps and overlaps in the field work. Also, as corn, for example, is harvested in the fall, the picker records GPS-derived location information along with the amount of corn picked at each location. In the spring during planting, fertilizer and seed are applied in amounts based on the harvest data.

RKE and TPMS both operate at 315 MHz. The RKE key fob antenna is usually on the PCB and couples to the operator's hand and arm, extending the effective radiator to near resonance. There can be as many as five receiving antennas on the vehicle. The TPMS
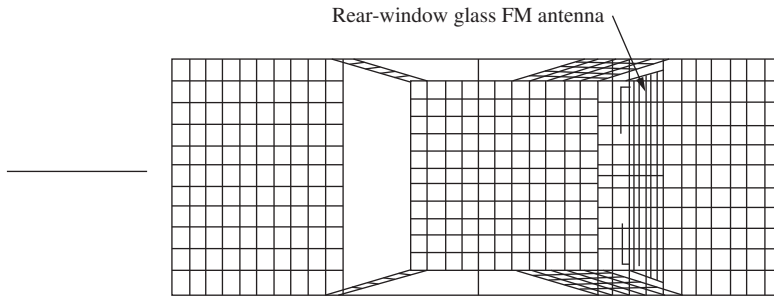
(*a*) Wire mesh model used for the NEC-2 method of moments simulations.
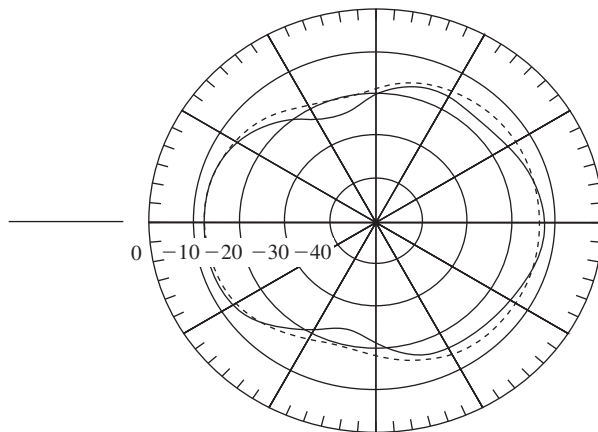
(*b*) Measured (solid curve) and simulated (dashed curve) radiation patterns.

Figure 12-9 A rear window glass FM broadcast car radio antenna. (From [3] © 2001. Reprinted with permission of IET.)

system uses a miniature transmitter unit in the valve stem of each tire including an antenna, often a loop, that is 0.01-λ wide. Receiving antennas can be in the dashboard, trunk, or side glass. The very popular entertainment application of satellite radio (known as satellite digital audio radio service, SDARS, in North America, with XM Radio as an example) operates at S-band frequencies and uses LHCP antennas similar to those used for GPS reception. Often antennas for cellular, GPS, and SDARS are located together at the center-rear of the roof. Antennas, usually in-glass, are also needed for the emerging application of terrestrial TV that usually includes multiple antennas for diversity reception. Another emerging application is dedicated short range communications (DSRC) operating at 5.9 GHz for roadside-to-vehicle ("talking signs") and vehicle-to-vehicle communications (for driver alerts to hazards ahead). DSRC is used in electronic toll collection systems that transmit a signal from the toll plaza to an RFID transponder in the vehicle, which returns identification information. The mobile wireless applications of lane change assistance, forward collision warning, and adaptive cruise control use radar, many systems operating at 24 or 76 GHz. More information on automotive antennas is found in [H.6: *Ant. Eng. Hdbk.*, 4th ed., Chap. 39] and [H.9.1; Fujimoto and James, 2nd ed., Chap. 8].

Antennas for aircraft must have low drag and thus be close to conformal to the super structure. Because the antenna is essentially integrated into the aircraft, modeling and measurements must include some (or, at low frequencies all) of the plane's body. Antennas used on aircraft include monopoles, dipoles, loops, slots, and even reflector antennas housed inside a radome. A popular aircraft antenna is the *blade antenna*, which has many antenna forms, including monopole, and often has multiple octave bandwidth. For more details on aircraft antennas and their modeling and measurement see [H.9.1: Macnamara] and [H.6: *Ant. Eng. Hdbk.*, 4th ed., Chap. 40].

Ships are able to support large antenna systems and associated hardware but often have an environment filled with many metal structures that complicate antenna design and placement. In addition, ship motion makes pointing of narrow beam antennas difficult, and sea-surface reflections introduce multipath interference. Modern military ship designs are moving toward a streamlined topside shape composed of four faces, each containing wideband phased arrays. Such multifunctional phased arrays are able to support broadband communications, radar, and electronic warfare tasks, and the arrays can be reconfigured to accommodate new applications instead of adding another antenna on the deck. The reference [H.9.1: Law] is a book devoted to the topic of shipboard antennas.

Small platforms that are mobile require small antennas and usually have multiple antennas for diversity to combat multipath propagation effects. Laptop computers commonly have wireless services (Wi-Fi, cellular data, TV) and are used while in motion. Antennas that are used in laptops include sleeve dipoles, helical monopoles, inverted-Fs, and slots. [H.9.1: Chen, Chap. 4] Special antenna designs are available that have low isolation to reduce coupling [7].

12.4   SMART ANTENNAS

The antenna historically has been considered as a separate component in a system. However, as processing power has been added to wireless systems, the role of the antenna has expanded from being a separate device into a subsystem with functions distributed through several sections of the radio. Digital signal processing (i.e., computing) hardware and software in the system has allowed intelligence to be added to the antenna subsystem, leading to an intelligent antenna, usually referred to as a **smart antenna**. A smart antenna is used for expanding frequency reuse and for combating propagation problems and interference. [8] The propagation problems are multipath fading and delay due to multiple signals arriving at the receiver from different directions and with different arrival times created by reflection and scattering of the transmitted

signal. An example of multipath fading is the fluctuation of the signal received from a distant radio station in a moving car.

Smart antennas are also capable of improving performance in the presence of interference caused by other nearby transmitters. As the sophistication increases, the smart antenna incorporates more support software to control its functions. An antenna that adapts by responding dynamically to its operating conditions (such as changing surroundings due to platform motion or to changing terminal locations) is called an **adaptive antenna**. Adaptive antennas are discussed in the next section and are used in radar as well as communications. A smart antenna can include adaptive capabilities, and the term smart antenna is not clearly defined. An example of an adaptive feature is steering a pattern null in the direction of an interference source as the interferer moves. The limiting case of a smart antenna is to have a single beam track each user, similar to that shown in Fig. 12-1c. Smart antennas are used both for transmitting and receiving, although more commonly implemented to receive at the base station in wireless communications because the uplink link usually limits performance. Smart antennas use multiple antenna elements. In the simplest form, they perform only a diversity function to improve performance and do not increase communication capacity. We begin with the topic of antenna diversity.

The dimensions used for *antenna diversity* (or simply, *diversity*) are similar to the dimensions used in frequency reuse (see Sec. 4.5.1): spatial, angle, and polarization. However, diversity is employed to improve performance, whereas frequency reuse is for increasing capacity by increasing effective bandwidth. Antenna diversity is effective in situations where the terminal or environment is time-changing, such as with a mobile terminal. Often there is no line-of-sight, and the terminal is said to be shadowed or blocked (for the path partially or totally hidden, respectively), as with vegetation or buildings in the direct path. Diversity is based on the concept that multiple copies of the signal are received by different branches (associated with different antenna elements, angles, or polarizations) that have experienced different propagation conditions and are not correlated or only partially correlated. We begin discussion of these diversity dimensions for the case of a transmitting mobile terminal and a receiving base station antenna with diversity. The classic diversity configuration is *spatial diversity*, in which two or more receiving antennas are widely separated, as illustrated by the two panels at the ends of the tower top face in Fig. 12-1b. The separation distance of around 10 or more wavelengths provides sufficiently different angles of arrival for the fields arriving at each antenna to have differing phases and thus low correlation. Fig. 12-10a shows the signals received by two panel antennas 3-m apart (8.5 λ) measured at Virginia Tech [10] as a function of the distance of the terminal moving at walking speed and transmitting at 842 MHz. The two dashed curves (S1 and S2) are the signals received at the base station. They are similar but not correlated because the peaks and nulls do not occur at the same time. The diversity curve was obtained by selecting the stronger of the two signals at each instant (position of the terminal). Note that most deep nulls are eliminated with diversity. The time variation is quantified on a statistical basis using the *cumulative distribution function* (CDF) shown in Fig. 12-10b. The ordinate value is the probability that the signal equals or exceeds the abscissa signal level in dB. The two curves S1 and S2 are the CDFs for the single antennas. Note that they are nearly identical, indicating that having just a different location for a single antenna provides no improvement. The curve on the right is the spatial diversity CDF. For 1% of the time, the single-antenna signals equal or exceed −20 dB, whereas the diversity signal equals or exceeds −10 dB, giving a *diversity gain* of 10 dB. At 10% probability, the diversity gain is 5 dB. Of course, the goal is to reduce loss of signal due to momentary deep fades. Diversity provides a solution. There are several algorithms for combining diversity branch signals, but the simplest one conceptually is selection combining, that we assume here, which selects the strongest signal at each instant.

Angle diversity uses an antenna with beams pointed in different directions over the sector. Angle diversity relies on having a different multipath component structure in
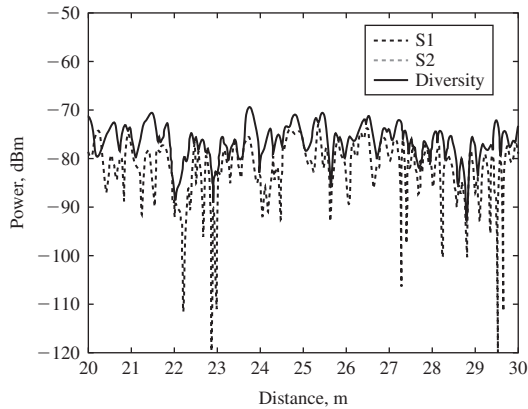
(*a*) Signals received by the two antennas (S1 and S2) along with the diversity-combined signal.
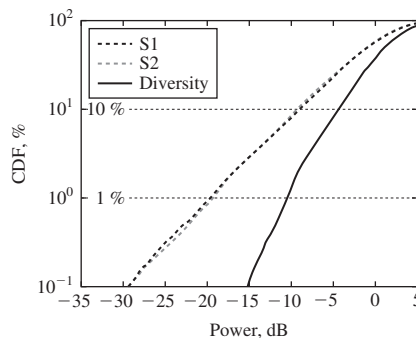
(*b*) The cumulative probability distribution. (Used with permission
from Dietrich et al. [10] © 2000 IEEE.)

Figure 12-10 Spatial antenna diversity data measured at Virginia Tech for two receiving 95°-sector panel antennas separated by 3 m. The terminal transmitted at 842 MHz and moved at walking speed.

each beam within the sector. The beams are often formed using a Butler matrix feed (see Sec. 8.5.2) that connects all elements in an array and provides one output port for each beam. Each beam spatial weights the multipath components differently, providing diversity.

Polarization diversity uses dual orthogonal polarizations in a panel antenna with one output port for each polarization. Polarization diversity at the base station is particularly effective for handset terminals, which are held in various orientations, changing the polarization of the transmitted wave. Polarization diversity works because different multipath components arrive at the receiver in different polarization states and each path has different depolarization effects. For example, a vertical utility pole along one path scatters incident vertical polarization differently than a horizontal bridge along a different path. The orthogonal polarizations in the receiving antenna respond differently to these two signals.

It has been shown experimentally that all three types of diversity—spatial, angle, and polarization—perform about the same. This research was done at Virginia Tech using a base station antenna with three antenna subsystems, one for each of the three diversities, and measuring the received signals in each diversity subsystem simultaneously. [9, 10] Tests were performed with the terminal in all possible principal orientations for various

distances and propagation scenarios, mostly non-line-of-sight, including indoor and outdoor locations. The results showed that spatial diversity has the most diversity gain with $G_D = 8$ to 11 dB at 1% probability, and with CDFs similar to that in Fig. 12-10b for nearly all cases. [10] For light shadowing, spatial diversity provides more diversity gain. Polarization diversity had almost as much gain with $G_D = 6$ to 10 dB at 1% probability. Angle diversity with $G_D = 5$ to 9 dB has the least diversity gain, but still a significant improvement over single antenna reception. Polarization diversity offers the significant advantage that the number of antennas per sector can be reduced to just one panel with dual polarization, reducing cost and visual impact. Often such a panel antenna has a top part with a single vertical polarization for transmitting and a lower part with dual slant-45° polarizations for receiving. Further compactness is obtained by using dual-band panels instead of separate bays of antennas for each frequency band.

Antenna diversity has been found to also be effective at the terminal level. This is because a handset operating in a building, for example, experiences wide-angle multipath due to reflections from walls and objects. In fact, an antenna element spacing as small as 0.1λ is effective. [11] This can be understood from a two-ray model for wide-angle multipath consisting of two oppositely directed waves of equal amplitude (one representing the direct wave and the other the reflected wave). This produces a standing wave with perfect nulls a half-wave apart and a null-to-peak distance of a quarter-wavelength. In a real propagation environment, the summation of all multipath components will not produce the same perfect standing wave pattern, but there will be peaks and nulls spaced fractional wavelengths apart. Experiments using a moving terminal with four antennas operating at 2 GHz yielded CDFs similar to that for base station diversity. [1, 11] Measurements with a human operator next to the handset showed that the diversity gain is only reduced by about 2 dB. [11] Adaptive beam forming showed that mean signal-to-interference plus noise-ratio is improved from 0 dB to 30 to 40 dB for an interferer separated as little as 2° away from the signal direction. [10]

Smart antennas used in wireless communications focus on recovering the signal of interest in a cluttered environment with interference present. A common feature of a smart antenna beyond spatial processing is to include time-domain delay equalization to properly combine the multipath signal components. As capabilities increase, the excitations of each element (often called weights) are adjusted based on a complex algorithm crafted to optimize certain performance parameters and/or to expand capacity. Smart antennas make use of the digital beamforming array techniques discussed in Sec. 8.9.2. Smart antennas in wireless implement space division multiple access (SDMA), which is a technique to spatially focus the transmitted signal in the direction of intended users, as shown in Fig. 12-1c. Adaptive beamforming techniques can be used to simultaneously steer the main beam toward a user (or, target) and to steer nulls toward interferers that are not in the same direction as the user. Dual (or even multiple) polarized adaptive smart arrays can also be used to reject interferers that differ in polarization from the user's polarization.

Smart base stations can be expanded to include transmit diversity alone or as a part of *multiple-input multiple-output* (*MIMO*) systems which have multiple transmit antennas and multiple receive antennas. Spatial diversity is the special case of two antennas at one end and one antenna at the other end. MIMO relies on the fact that a transmit beam is spread over a wide angular sector and that the receive antenna collects signals arriving from many angles. So the multiple paths from the transmitter to the receiver lead to signals that are spread in time due to their different path lengths traveled. This is exploited by sending the same data stream out on different beams from the transmit antenna. The receive antenna beams are directed toward the dominant scatterers and collect the data streams for subsequent combining. Although counterintuitive, the more multipath, the more effective is MIMO. MIMO uses space-time coding to increase performance and capacity. [H.6: Balanis, Ed., Chap. 26] See [H.9.1: Sanchez-Hernandez, Ed., Chap. 6] for material on handset antennas for MIMO systems. For more information

on smart antennas, see the following sources: [H.9.2], [H.6: Godara, Ed., Chap. 22], and [H.9.1: Fujimoto, 3rd ed., Chap. 15].

12.5   ADAPTIVE AND SPATIAL FILTERING ANTENNAS

We have seen in previous sections that antenna arrays can steer their main beams and nulls in prescribed directions. This process is sometimes referred to as spatial filtering. Spatial filtering is of particular importance in systems where performance is interference limited. These include, for example, mobile wireless communications systems as well as radar systems in a jamming environment. Spatial filtering can be achieved using the following implementations:

Phased array: We have already seen that by adjusting the phase of each element, the main beam maximum can be pointed as desired.

Sectorized: As in Fig. 12-2*b*, multiple antenna elements are used such that each provides coverage over a sector. In second generation mobile phone systems, three base station antennas, each covering 120° sectors, are used to cover an entire cell.

Switched beam: A software algorithm is used to select from multiple predefined beams of an antenna array such that the strongest signal is chosen. This will be discussed in the next section.

Adaptive array: Current generation adaptive arrays use a software algorithm to maximize the desired signal in one or more directions while directing nulls toward interferers. The software is more complex than for a switched beam system. This will also be discussed in the next section.

12.5.1   Switched Beam Antenna Systems

There are several ways of implementing a switched beam system according to how the antenna outputs are chosen or combined. [H.4: Saunders, Chap. 16] In *selection combining*, a separate receiver is employed at each antenna output, and the switch selects the receiver output which instantaneously has the best signal-to-noise plus interference ratio (SNIR). This is a complicated and expensive scheme because the switch must monitor all receiver outputs simultaneously, and each antenna has its own receiver. An alternative is *switched combining* that uses just one receiver, with the switch selecting the receiver input with a SNIR above a predefined threshold. The performance of this scheme is inferior to that of selection combining because there may be other receiver inputs with a better SNIR that are not selected due to other input(s) being above the threshold. Switched combining wastes the signal power residing on the unused receiver inputs. This can be overcome by *equal-gain combining*, which adds the signals of the available receiver inputs. Care is taken to add the signals such that they are co-phased for coherent combining. Noise at each receiver input is randomly phased and adds incoherently. When equal gain combining is employed, there is the possibility that one of the inputs has a much lower SNIR than the others, thereby degrading the overall SNIR. This negative possibility can be overcome by *maximum ratio combining*, in which the various inputs are weighted according to their individual SNIR so as to maximize the overall SNIR.

A relatively new type of spatial filtering antenna is a *reconfigurable antenna* that can alter its radiation pattern by opening and closing switches. Current switch technologies include PIN diodes, FETs, and MEMS devices, the latter in their commercial infancy. [H.4: Huang, p. 352] For example, a number of patch antenna elements can be interconnected by MEMS devices controlled by a microprocessor controller. The antenna performance is measured by a sensor system measuring signal strength. By using many interconnected elements, the reconfigurable antenna is able to change itself into a large number of different electrical configurations in real time as it adapts to changing conditions.

12.5.2   Adaptive Antennas in General

Signal reception in a conventional array is subject to degradation in the signal-to-noise plus interference ratio (SNIR) due to the reception of undesired signals, as well as the unwanted reception of natural noise. This undesired signal energy enters the conventional array via side lobes and portions of the main beam. To overcome this degradation, arrays can be designed to adapt to the signal environment. Such arrays are called adaptive arrays because they sense the presence of unwanted energy in the signal environment and suppress it, thereby enhancing the SNIR.

The earliest example of an adaptive antenna, although it was not called that, is the *side lobe canceller* (SLC) employed in radar systems as a protection against jamming. In this scheme, a secondary array of low-gain elements works in conjunction with the main array to null out interfering signals by processing the outputs of the two arrays such that the jamming signal or signals are canceled. This is actually a partial adaptive array because not all the elements contribute equally to the operation of the SLC processor.

The term adaptive antenna was first used by Van Atta in the late 1950s to describe a self-phasing array that automatically and passively reradiates a signal back in the direction from which it came. [H.9.2: Monzingo, p. 5] No prior knowledge of the signal direction is required. The Van Atta array may be regarded as an adaptive *transmitting* array. This is in contrast to an adaptive *receiving* array, which employs a computer algorithm to control the signal levels and phasing at the elements such that, when the output of the elements are combined or selected individually, the SNIR is improved. We will consider both kinds of adaptive arrays.

12.5.3   Van Atta Retrodirective Array

A simple example of the Van Atta array is shown in Fig. 12-11. This array concept is based upon the principle of phase conjugacy, which states: "*the phase of the transmitted signal from any element in the retrodirective array must bear a conjugate relationship to the phase of the received signal at that element when compared to a common reference signal.*" Referring to Fig. 12-11, this means that each element in the array must have an outgoing wave which is delayed (with respect to a given reference element) by exactly as much as the incoming wave was advanced. When this requirement is satisfied, the total path length from the source to an array element and back to the source will be the same for all elements in the array as it is in Fig. 12-11.
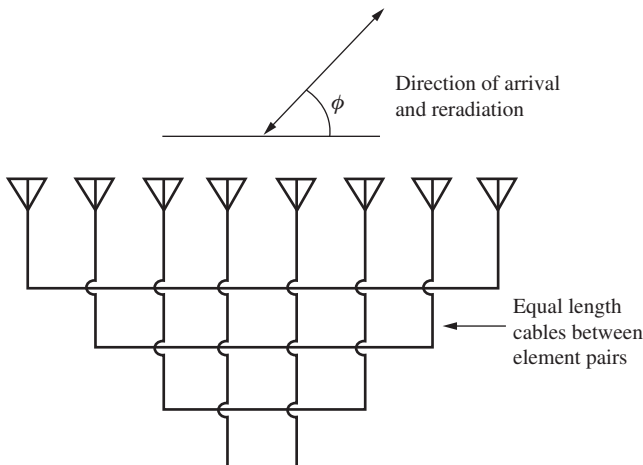
Figure 12-11 Passive Van Atta array.
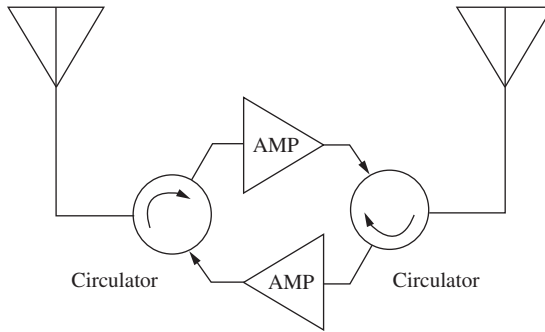
Figure 12-12 Two-element Van Atta array with bilateral amplification. (From [H.6: Hansen, Vol. III, p. 370] © 1966 Academic Press. Used with permission.)
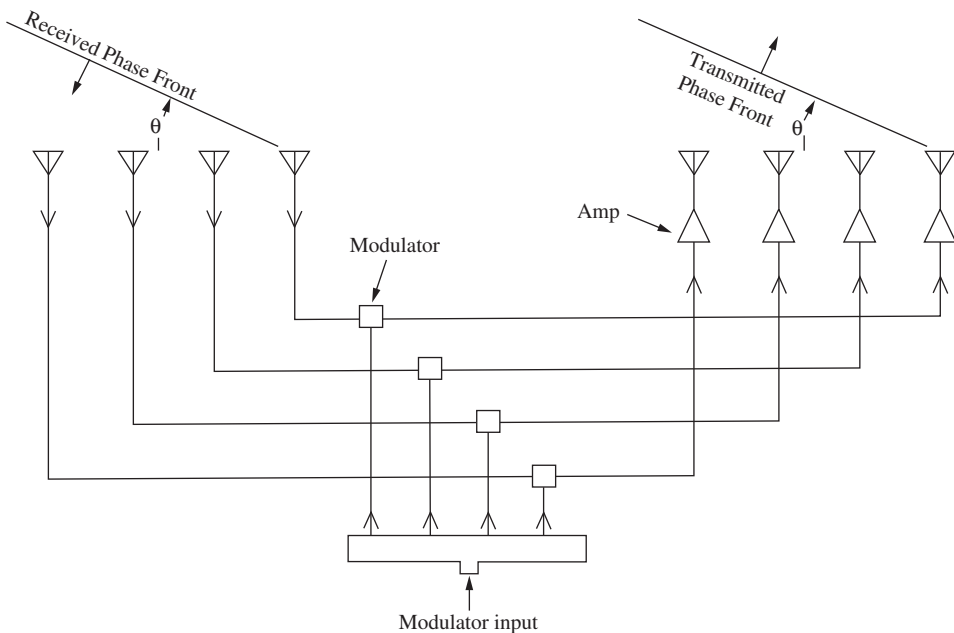
Figure 12-13 Van Atta repeater array with amplification and modulation. After Leroy B. Van Brunt, "Applied ECM," Vol. I (E. W. Engineering), 1980.

Fig. 12-11 is that of a passive array. Its gain is determined by the size of the array. Additional gain and/or modification of the received signal may be achieved with an active scheme as shown in Fig. 12-12. The amplification in Fig. 12-12 is usually limited to about 15 dB due to mutual coupling. This limitation can be overcome by using a Van Atta repeater array, as shown in Fig. 12-13. In addition to amplification, Fig. 12-13 shows signal modification by the modulators in the interconnecting lines. Care must be taken to phase-match the amplifiers and the modulators as well as the cable lengths.

12.5.4   Adaptive Receiving Arrays

An adaptive receiving array is one that coherently receives desired signals and suppresses undesired signals. In theory, it does this by steering the main beam to the direction of the desired signal source and simultaneously steering nulls toward the undesired signal sources. Since the outputs of the array elements coherently combine the desired signal, but incoherently combine noise (i.e., thermal noise and interfering signals), the output signal to

noise plus interference ratio of an *N*-element adaptive array can be *N* times the SNIR of an individual element. The number of degrees of freedom (i.e., main beam plus null directions) in the adaptive array has is equal to the number of elements in the array minus one.

An adaptive array has four basic components: the RF array itself, the beam forming network, the signal processor, and the adaptive algorithm. Beamforming can be analog or digital. Analog beamforming is usually accomplished with a Butler matrix (see Fig. 8-32). The fast Fourier transform (FFT) is equivalent to the Butler matrix. The FFT is used in digital beam forming.

There are two common basic adaptive algorithms: the Applebaum algorithm and the Widrow algorithm (see Fig. 12-14). The *Applebaum algorithm* requires that the signal bearing be known and seeks to maximize the SNIR at the output of the array system under the assumption that the desired signal is absent most of the time (e.g., radar). Because it is simple to implement, it is used extensively for clutter and interference rejection in radar systems.

The *Widrow algorithm* minimizes the mean square error (LMS) signal, which is the difference between the array system output and a reference signal. In a communication system, the desired signal is usually present and is used as the reference signal. In other kinds of systems (e.g., radar) it is pointless to generate a fictitious desired signal. In both
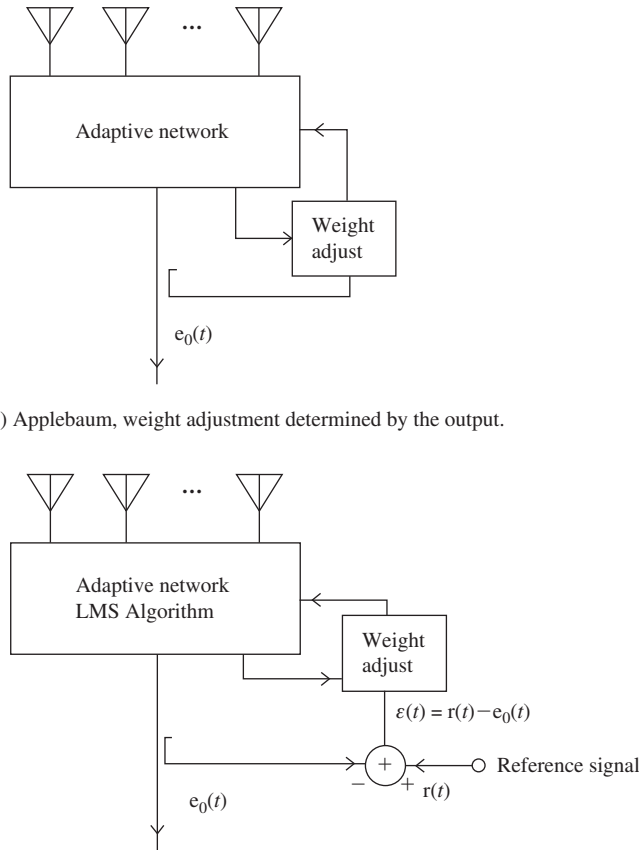
(*a*) Applebaum, weight adjustment determined by the output.

(*b*) Widrow, weight adjustment is determined by the error signal, $\varepsilon(t)$.

Figure 12-14 Two types of adaptive arrays.

of these algorithms, if there is no interference, the optimum weights are the complex conjugate of the desired signal (i.e., principle of phase conjugacy).

The use of adaptive arrays in mobile communications systems helps to overcome the problems of single antenna systems. The use of adaptive antenna arrays helps improve the system's performance by increasing spectrum efficiency and channel capacity, by extending range, steering multiple beams to track many mobiles, and electronically compensating for aperture distortion. The use of adaptive arrays also reduces multipath fading, co-channel interference, bit-error rate, and outage probability.

REFERENCES

1. C. Dietrich, R. Barts, W. Stutzman, and W. Davis, "Trends in Antennas for Wireless Communications," *Microwave J.*, Vol. 46, pp. 22–44, Jan. 2003.
2. S-Y Suh, W. Stutzman, W. Davis, A. Waltho, K. Skeba, and J. Schiffer, "A Novel Low-profile, Dual-polarization, Multi-band Base-station Antenna Element—The Four-point Antenna," *IEEE Vehicular Technology Conference Proc.* (Los Angeles), Vol. 1, pp. 225–229, Sept. 2004.
3. J. Batchelor, R. Langley, and H. Endo, "On-Glass Mobile Antenna Performance Modeling," *IEE Proc. Micro. Ant. Prop.*, Vol. 148, pp. 233–238, Aug. 2001.
4. D. Horn, "Selection of Vehicular Antenna Configuration and Location through Use of Radiation Pattern," *IEEE Veh. Tech. Conf. Rec.* (Cleveland), Dec. 1973.
5. D. Horn, "Vehicular Cellular Passive Repeaters," *Mobile Radio Technology*, Vol. 9, July 1991.
6. K. Yegin, "On-Vehicle GPS Antenna Measurements," *IEEE Ant. and Wireless Prog. Let.*, Vol. 6, pp. 488–491, 2007.
7. S-Y Suh, V. Nair, D. Souza, and S. Gupta, "High Isolation Antenna for Multi-Radio Antenna System Using a Complementary Antenna Pair," *Proceedings of IEEE Inter. Symp. On Ant. and Prop.*, pp. 1229–1232, June 2007.
8. C. Dietrich and W. Stutzman, "Smart Antennas Enhance Cellular/PCS Performance," *Microwaves and RF*: Vol. 36: Part I, pp. 76–86, April 1997; and Part II, pp. 164–168, May 1997.
9. B-K Kim, W. Stutzman, and D. Sweeney, "Indoor and Outdoor Measurements of Space, Polarization, and Angle Diversity for Cellular Base Stations in Urban Environments," *Proc. Vehicular Tech. Conf.* (Boston), pp. 22–29, Sept. 2000.
10. C. Dietrich, W. Stutzman, B-K Kim, and K. Dietze, "Smart Antennas in Wireless Communications: Base-Station Diversity and Handset Beamforming," *IEEE Trans. on Ant. & Prop. Mag.*, Vol. 42, pp. 142–151, Oct. 2000.
11. C. Dietrich, K. Dietze, J. Nealy, and W. Stutzman, "Spatial, Polarization, and Pattern Diversity for Wireless Handheld Terminals," *IEEE Trans. on Ant. & Prop.*, Vol. 49, pp. 1271–1281, Sept. 2001.
12. Leroy B. Van Brunt, "Applied ECM," Vol. 1 (E. W. Engineering), 1980.

PROBLEMS

12.2-1 Plot the pattern for Example 12-1 of an omnidirectional cellular base station antenna for 7° downtilt and comment on it compared to the pattern of the same array but for 1.25° of downtilt.

12.2-2 For the base station of Example 12-2: (a) Calculate and plot the elevation plane pattern, and (b) Evaluate the directivity assuming an omnidirectional element pattern in the azimuth plane.

12.2-3 The azimuth pattern of a sector base station antenna can be modeled by generalizing the cardioid pattern of (3-73):

$$F(\theta) = \left\{ \cos \left[a \, \frac{\pi}{4} \left(\cos \left(\theta \right) - b \right) \right] \right\}^{c}$$

(a) Adjust the coefficients a, b, and c in the above formula to find pattern that approximates the 65° pattern of Fig. 12-5.
(b) Search antenna product literature to find a commercial base station antenna with a 65° azimuth pattern. Compare the manufacturer-provided pattern to that in (a).

12.3-1 Use a simulation code to model the three-quarter-wavelength monopole with a stub of Fig. 6-45 over an infinite ground plane at 860 MHz, but with no inductive load at the base. Include the input impedance, elevation plane pattern, and gain.

12.3-2 Use a simulation code to model the five-eighths over quarter-wave monopole of Fig. 12-8 over an infinite ground plane. Give the input impedance, gain, and elevation pattern.

12.4-1 *Project*. Investigate the topic of antenna diversity on handset terminals and write a report including the kinds of diversities used, the antennas used, quantitative information about diversity gain, and effects on the human operator.

Chapter 13

Antenna Measurements

Accurate antenna measurement requires specialized space and instrumentation. The space can be an outdoor range with towers and positioners or an indoor room (often very large) configured with probes, positioners and usually absorber-lined walls. Instrumentation is needed for generating test signals, controlling the positioners, collecting data, and processing and displaying data. This usually includes a vector network analyzer. All of the foregoing items are expensive to purchase and require trained personnel to operate. Thus, most engineers involved in antennas and wireless probably will never make an antenna measurement. However, all antenna engineers and personnel involved with wireless technologies will have a need to understand measured antenna data. In the past, the antenna design cycle was based on measuring hardware models and iterating on the models and measurements until specifications were satisfied. Now it is common to use simulation packages (see Appendix G) developed from numerical methods (see Chaps. 14–16) for preliminary investigations of antennas during the design phase, followed by measurements for validation and for compliance with specifications and regulations. The study of antenna measurements provides a deeper understanding of antennas, allows one to interpret measured data, and serves as an introduction to those who desire to actually be directly involved with antenna measurements. The principles introduced here also apply to scattering measurements, but we focus on antenna measurements. The primary measured antenna characteristics are radiation pattern, gain, polarization, and impedance. The first three of these are discussed in this chapter. Impedance is often measured in conjunction with the other characteristics and is usually performed with a network analyzer. We begin the chapter with the principles that pattern measurements are based on. Reference books on both general and specialized topics in antenna measurements are listed in Appendix H.7.

13.1 RECIPROCITY AND ANTENNA MEASUREMENTS

We have noted several times that usually the pattern of an antenna is the same for either transmitting or receiving. If this is the case, the antenna is said to be *reciprocal*. This is true as long the antenna materials and the surrounding medium satisfy conditions for reciprocity. In this section, we use Maxwell's equations to derive reciprocity theorems, which are then used to find the conditions under which an antenna is reciprocal. These results have significant implications on antenna measurements. In particular, if an antenna is reciprocal, the performance parameters (pattern, gain, etc.) can be measured with the antenna operating in transmit or receive and the results are identical. This avoids measuring a parameter twice, once for transmit and once for receive, and allows the test antenna to be used in either transmit or receive based on which is easier to do or more likely to give greater accuracy. Of course, for using an antenna in operational settings,
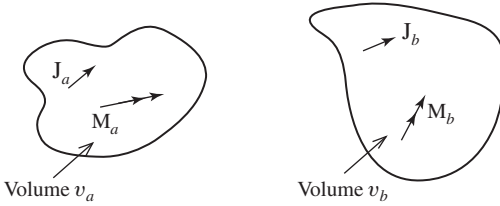
Figure 13-1 Source configuration for the Lorentz reciprocity theorem.

other factors come into play in choosing the type of antenna for transmit or receive applications. For example, an electrically small antenna with high ohmic loss and high input reactance would not be suitable for high-power transmitting because the loss would generate heat and impedance mismatch would cause significant power to be reflected back to the transmitter. Both of these effects could damage the transmitter. On the other hand, an inefficient antenna is often acceptable for receive applications such as an AM car radio antenna (see Example 2-4). Inefficient antennas are also used at HF frequencies and below for receiving because the internal noise generated in the antenna loss resistance is usually small compared to external atmospheric noise.

There are several forms reciprocity theorems take for electromagnetic field problems. We consider two forms of reciprocity for use in antenna problems. The Lorentz reciprocity theorem is discussed first. Let sources $\mathbf{J}_a$ and $\mathbf{M}_a$ produce fields $\mathbf{E}_a$ and $\mathbf{H}_a$ and sources $\mathbf{J}_b$ and $\mathbf{M}_b$ produce fields $\mathbf{E}_b$ and $\mathbf{H}_b$. See Fig. 13-1. The frequencies of all quantities are identical. The Lorentz reciprocity theorem that is derivable from Maxwell's equations (see Prob. 13.1-1) states that for isotropic media.

$$\iiint_{v_a} (\mathbf{E}_b \cdot \mathbf{J}_a - \mathbf{H}_b \cdot \mathbf{M}_a)\, dv' = \iiint_{v_a} (\mathbf{E}_a \cdot \mathbf{J}_b - \mathbf{H}_a \cdot \mathbf{M}_b)\; dv' \qquad (13\text{-}1)$$

The left-hand side is the reaction (a measure of the coupling) of the fields from sources b on sources a, and the right-hand side is the reaction of the fields from sources a on sources b. This is a very general expression, but it can be put into a more usable form. Let sources b consist of only an ideal electric dipole of vector length $\mathbf{p}$ located at point (x_p, y_p, z_p). Since the ideal dipole can be represented as an infinitesimal source and $\mathbf{M}_b$ is zero, (13-1) becomes[1]

$$\mathbf{E}_a(x_p, y_p, z_p) \cdot \mathbf{p} = \iiint_{v_a} (\mathbf{E}_b \cdot \mathbf{J}_a - \mathbf{H}_b \cdot \mathbf{M}_a)\; dv' \qquad (13\text{-}2)$$

This expression allows calculation of the electric field from sources a by evaluating the integral using known sources $\mathbf{J}_a$ and $\mathbf{M}_a$ and known ideal dipole fields $\mathbf{E}_b$ and $\mathbf{H}_b$ of (2-72) and (2-71), evaluated at the location of sources a. This can be performed for various orientations $\mathbf{p}$ of the ideal dipole, which is acting as a field probe.

The Lorentz reciprocity theorem can also be used to derive a second reciprocity theorem using terminal voltages and currents. Suppose sources a and b are antennas excited with ideal (infinite impedance) current generators I_a and I_b. Since no magnetic sources are present, (13-1) reduces to

---

[1] The ideal dipole current could be written as $\mathbf{J}_b = \delta(x - x_p)\delta(y - y_p)\delta(z - z_p)\mathbf{p}$. This together with $\mathbf{M}_b = 0$ in (13-1) yields (13-2).

$$\iiint\limits_{v_a} \mathbf{E}_b \cdot \mathbf{J}_a \, dv' = \iiint\limits_{v_b} \mathbf{E}_a \cdot \mathbf{J}_b \, dv' \tag{13-3}$$

For perfectly conducting antennas, the electric fields will be zero over the antennas; however, voltages will be produced across the terminals. Taking the current to be constant in the terminal region and using the concept of $\int \mathbf{E} \cdot \mathbf{d}\ell = -V$, we see that (13-2) becomes

$$V_a^{oc} I_a = V_b^{oc} I_b \tag{13-4}$$

where V_a^{oc} is the open circuit voltage across the terminals of antenna a due to the field $\mathbf{E}_b$ generated by antenna b and, similarly, V_b^{oc} is the open circuit voltage at antenna b due to antenna a. Open circuit voltages have been used because of the infinite impedances of the generators. Rearranging (13-4) leads to a statement of reciprocity in circuit form

$$\frac{V_a^{oc}}{I_b} = \frac{V_b^{oc}}{I_a} \tag{13-5}$$

Several factors affect the voltage appearing at one antenna due to another antenna that is excited: the specific antennas used, the medium between the antennas with perhaps other objects present, and the relative orientation of the antennas. We can represent the general situation entirely in terms of circuit parameters using the following, which holds for any linear passive network:

$$V_a = Z_{aa}I_a + Z_{ab}I_b \tag{13-6a}$$

$$V_b = Z_{ba}I_a + Z_{bb}I_b \tag{13-6b}$$

where V_a, V_b, I_a, and I_b are the terminal voltages and currents of antennas a and b. If antenna a is excited with a generator of current I_a, the open circuit voltage appearing at the terminals of antenna b is $V_b|_{I_b=0}$. The transfer impedance Z_{ba} from (13-6b) with I_b zero is

$$Z_{ba} = \frac{V_b}{I_a}\bigg|_{I_b=0} \tag{13-7}$$

If antenna b is excited with a generator of current I_b, the open circuit voltage appearing at the terminals of antenna a is $V_a|_{I_a=0}$. The transfer impedance Z_{ab} is, from (13-6a) with I_a zero,

$$Z_{ab} = \frac{V_a}{I_b}\bigg|_{I_a=0} \tag{13-8}$$

Comparing (13-7) and (13-8) to (13-5), we see that[2]

$$Z_{ab} = Z_{ba} = Z_m \tag{13-9}$$

where Z_m is the transfer (or mutual) impedance between the antennas. This can also be shown from the circuit formulation of (13-6) if the individual impedances are linear, passive, and bilateral (see Probs. 13.1-3 and 13.1-4). This, in turn, is true if the medium and the antennas are linear, passive, and isotropic.

---

[2] See [1] for a proposed slight modification of (13-9).

(a) Two-port representation of a
two-antenna system.

(b) Antenna a excited with current source I.

(c) Antenna b excited with current source I.
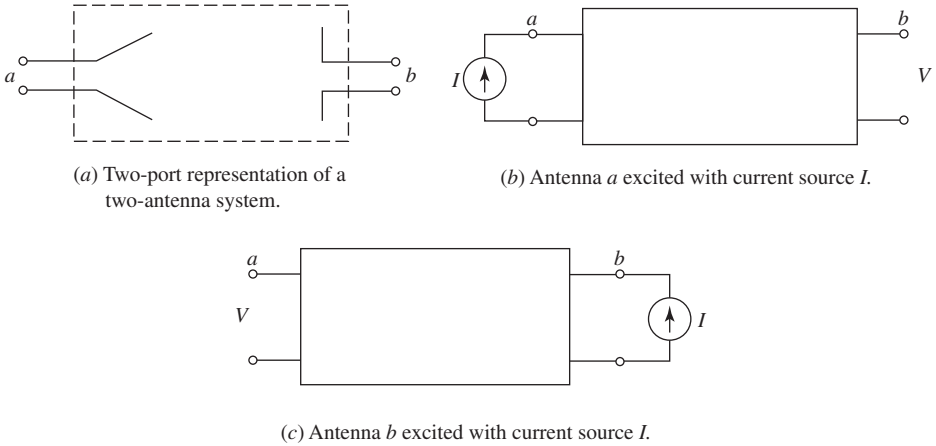
Figure 13-2 Reciprocity for antennas. The output voltage V is the same in (b) and (c) for the
same input current I.

The significance of these results is now explained using the model of Fig. 13-2. If an
ideal current source of current I excites antenna a, the open circuit voltage at the term-
inals of b from (13-7) is

$$V_b\big|_{I_b=0} = I\,Z_{ba} \tag{13-10}$$

If the same source is now applied to the terminals of antenna b, the open circuit voltage
appearing at the terminals of antenna a from (13-8) is

$$V_a\big|_{I_a=0} = I\,Z_{ab} \tag{13-11}$$

But $Z_{ab} = Z_{ba}$, so the preceding two equations yield

$$V_a\big|_{I_a=0} = V_b\big|_{I_b=0} = V \tag{13-12}$$

Thus, the same excitation current will produce the same terminal voltage independent
of which port is excited, as illustrated in Fig. 13-2. In other words, reciprocity states that
the source and the measurer can be interchanged without changing the system response.
The same is true of an ideal voltage source and short circuit terminal currents. These are
familiar results from network theory.

The self-impedances of the antennas from (13-6) are

$$Z_{aa} = \frac{V_a}{I_a}\bigg|_{I_b=0} \tag{13-13}$$

$$Z_{bb} = \frac{V_b}{I_b}\bigg|_{I_a=0} \tag{13-14}$$

If antennas a and b are widely separated, which is the usual operating situation, Z_{aa} and
Z_{bb} are much greater than $Z_{ab} = Z_{ba} = Z_m$. Thus, the input impedance to antenna a, for
example, from (13-6a) is

$$Z_a = \frac{V_a}{I_a} = Z_{aa} + Z_{ab}\frac{I_b}{I_a} \approx Z_{aa} \tag{13-15}$$

Thus, if an antenna is isolated so that all objects including other antennas are far away and the antenna is lossless, the self-impedance equals its input impedance.

Suppose antenna a is excited (i.e., acting as a transmitter) and the voltage produced at the terminals of antenna b is measured with an ideal voltmeter. If the antennas are separated so that they are in each other's far field, the transfer impedance Z_{ba} is actually the far-field (or radiation) pattern of antenna a if antenna b is moved around a on a constant radius as shown in Fig. 13-3a. As antenna b is moved, it is maintained with the same orientation and polarization relative to antenna a. The output voltage of b as a function of angle around antenna a gives the relative angular variation of the radiation from antenna a, that is, its radiation pattern. Examining (13-7), we see that this is really Z_{ba} (I_a is constant). Thus, Z_{ba} as a function of angle is the transmitting pattern of antenna a. If now antenna b is excited and antenna a acts as a receiver, the terminal voltage of antenna a is the receiving pattern of antenna a as antenna b is again moved around at a constant distance from antenna a; see Fig. 13-3b. Thus, Z_{ab} as a function of angle is the receiving pattern of antenna a. Since the transfer impedances are identical, we can conclude that *the transmit and receive patterns of an antenna are identical*. This is an important consequence of reciprocity.

The equality of the transmit and receive patterns of an antenna is not an unexpected result. This can be seen through the relation $G(\theta, \phi) = 4\pi A_e(\theta, \phi)/\lambda^2$ of (4-28) which relates the receiving characteristic of the antenna $A_e(\theta, \phi)$ for an incoming plane wave from angle (θ, ϕ) to the gain pattern value $G(\theta, \phi)$ in the direction (θ, ϕ) when the antenna transmits. The reciprocal property is of major practical importance. It permits the test antenna to be used in either a receive or transmit mode during pattern measurements. In practice, pattern measurements are usually made with the test antenna used in reception.

It is important to note that reciprocity, as illustrated in Fig. 13-2 along with (13-9), is a general result. Also, in the case when the antennas are far removed from each other,
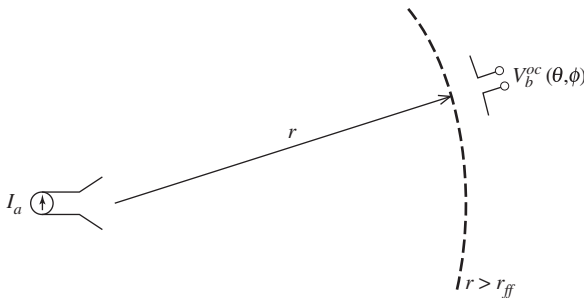
(a) The transmitting pattern of antenna a is $Z_{ba}(\theta,\phi) = V_b^{oc}(\theta,\phi)/I_a$.
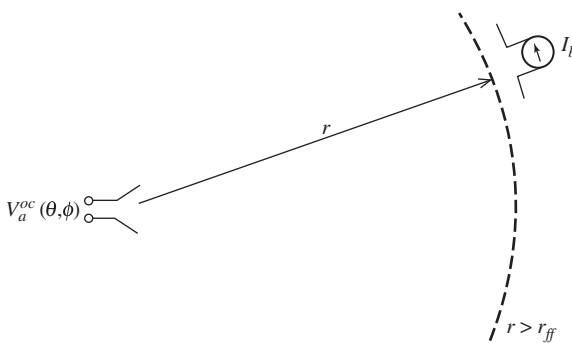
(b) The receiving pattern of antenna a is $Z_{ab}(\theta,\phi) = V_a^{oc}(\theta,\phi)/I_b$.

Figure 13-3 Antenna pattern reciprocity. The transmitting and receiving patterns of an antenna are identical because $Z_{ab}(\theta, \phi) = Z_{ba}(\theta, \phi) = Z_m(\theta, \phi)$.

$Z_m(\theta, \phi)$ is the far-field pattern. These results hold under the conditions that the antenna and transmission medium are linear, passive, and isotropic. These conditions are satisfied in most cases. A notable exception is if the antenna system includes a non-reciprocal device such as a circulator or isolator, which use ferrite materials.

13.2   PATTERN MEASUREMENT AND ANTENNA RANGES

An antenna pattern is a graphical representation of the field magnitude at a fixed distance from an antenna as a function of direction. With the antenna at the origin of a spherical coordinate system, radiation fields **E** and **H** are perpendicular to each other and both are transverse to the direction of propagation $\hat{\mathbf{r}}$. Also, the field intensities vary as r^{-1}. In antenna pattern discussions, electric field is used, but magnetic field behavior follows directly since its intensity is proportional to the electric field and its direction is perpendicular to **E** and $\hat{\mathbf{r}}$; see (2-107).

The radiated electric field is both a vector and a phasor. In general, it has two orthogonal components, E_θ and E_ϕ. These components are complex-valued and their relative magnitude and phase determine the polarization; see Sec. 2.8. For simple antennas, only one component is present. For example, the ideal dipole parallel to the z-axis has only an E_θ component as shown in Fig. 2-4. Measurement of the radiation pattern in this case is conceptualized by moving a receiving probe around the antenna as it transmits a constant signal a fixed distance away, r. The probe's orientation is maintained parallel to E_θ as shown in Fig. 13-4. The output of the probe varies in direct proportion to the intensity of the received field component arriving from direction (θ, ϕ). The pattern of the ideal dipole is $\sin \theta$; see Fig. 2-4. In general, antennas will have both E_θ and E_ϕ components and patterns are cut twice, once with the probe oriented parallel to E_θ and once with it parallel to E_ϕ.

Although we have conceptualized the measurement of a radiation pattern by moving a receiver over a sphere of constant radius, this is obviously an impractical way of making such measurements. The important feature is to maintain a constant large distance between the antennas and to vary the observation angle. This is accomplished by rotating the *test antenna,* or *antenna under test* (AUT), as illustrated in Fig. 13-5. By reciprocity, it makes no difference if the test antenna is operated as a receiver or transmitter, but usually
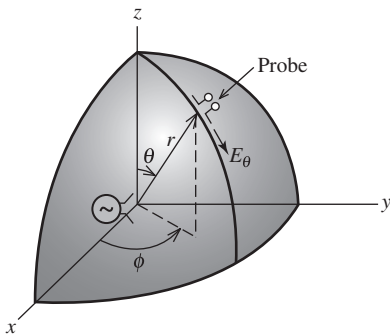
Figure 13-4 Pattern measurement conceptualized by movement of a probe antenna over the surface of a sphere in the far field of the antenna.
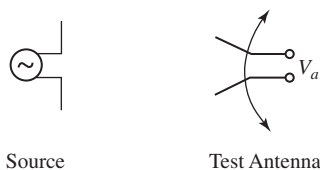
Figure 13-5 Radiation pattern measurement. The pattern of antenna a is proportional to the terminal voltage V_a, which is a function of the angular position of antenna a during rotation.

the test antenna is used as a receiving antenna and we adopt this convention. The fields from the motionless source antenna provide a constant illumination of the test antenna whose output varies with its angular position. This leads to the rule that *it is the pattern of the rotated antenna that is being measured.*

A complete representation of the radiation properties of an antenna would, of course, require measuring the radiation at all possible angles (θ, ϕ). This is rarely attempted and fortunately is not necessary. For most applications, the principal plane patterns are sufficient. See Fig. 2-4 for an illustration of the principal plane patterns using an ideal dipole.

There are many ways of displaying antenna patterns. For example, a principal plane pattern could be plotted in polar or rectangular form. In addition, the scale could be either linear or logarithmic (decibel). All combinations of plot type and scale type are used: polar-linear, polar-log, rectangular-linear, and rectangular-log. Fig. 13-6 shows the same radiation pattern plotted in these four ways. Generally speaking, log plots are used for high-gain, low side-lobe patterns and linear plots are used when the main beam details are of primary interest. These antenna pattern representations can be recorded directly using commercially available measuring and recording equipment. When more detailed information is required, the results of several planar cuts can be put together to make a contour plot. It is important to appreciate that measured patterns are usually not perfectly symmetric even though the antenna structure appears to be symmetric and also nulls are often partially filled.

A facility used to measure antenna radiation characteristics is referred to as an *antenna range*. Often, the same, range can also be used to measure scattering characteristics such as radar cross section. The entire measurement facility consists of the space (indoor or outdoor) for the source and test antennas, antenna positioners, a transmitter, a receiving system, and data display/recording equipment. In this section, we discuss the basic range layouts. For more details on antenna ranges, see [H.7: Evans], [H.9.4: Macnamara, Chap. 8], and chapters in handbooks of Appendix H.6 for a complete discussion of antenna measurement techniques.

(*a*) Polar-linear.

(*b*) Polar-logarithmic.

(*c*) Rectangular-linear.

(*d*) Rectangular-logarithmic.
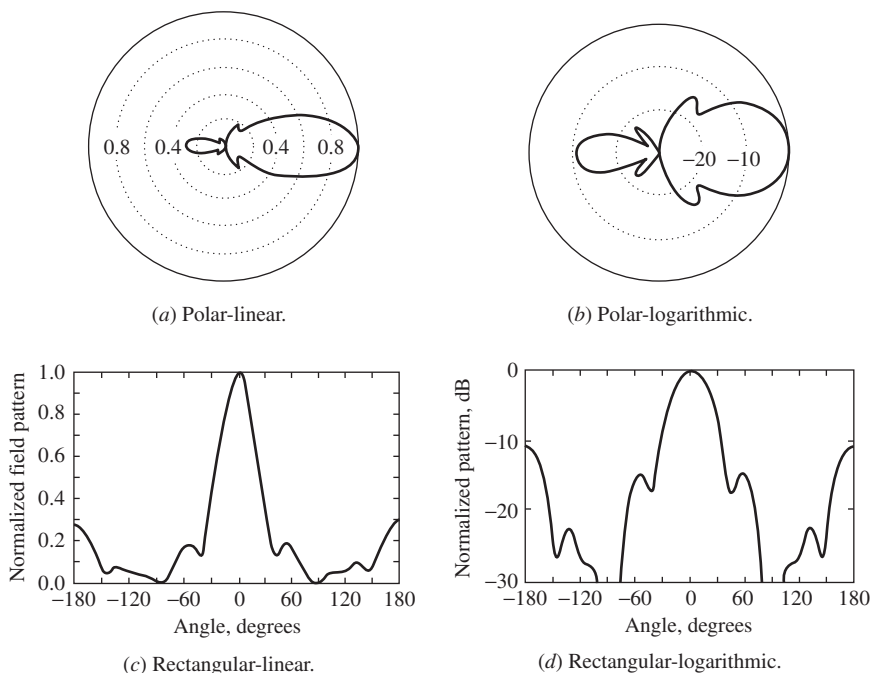
Figure 13-6 Illustration of the four antenna pattern plot types using the same pattern.

Table 13-1 lists the types of antenna ranges together with their characteristics and advantages and disadvantages. Most ranges are *free space ranges* that are designed to have strong direct illumination of the test antenna with weak indirect illumination. First, we consider *far-field ranges* in which the source antenna is far from the test antenna. This can be accomplished by elevating both (or one of) the source and test antennas giving an *elevated range* (or *slant range*). For all antenna ranges, the site around the test antenna affects pattern measurement accuracy. The guiding principle is to have the line of sight (direct) path between the source and test antenna unblocked and as high above the ground (or floor) as practical. This yields large values for the angles α_t and α_r shown in Fig. 13-7. Then directive antennas will have indirect rays arising from specular reflection from the ground of reduced level because angles α_t and α_r usually correspond to side-lobe directions. In the elevated range of Fig. 13-7, the source and test antennas are approximately the same height, $h_t \approx h_r$. The slant range is similar to the elevated range except that only the source is elevated, leaving the test antenna conveniently located near the ground.

In far-field ranges, the test antenna is located in the far field of the source antenna so that the incoming waves are nearly planar as indicated in Fig. 13-7. In fact, a common goal

Table 13-1 Characteristics of Antenna Ranges

Range Type	Description	Advantages	Disadvantages
FREE SPACE RANGES	Effects of all surroundings are suppressed to acceptable levels		
Far-field Ranges			
Elevated range	Source and test antenna are placed on towers, buildings, hills, etc	Low cost	Requires real estate May require towers Outdoor weather
Slant range	Either the source or test antenna is elevated	Low cost	Requires real estate May require a tower Outdoor weather
Anechoic chamber	A room is lined with absorber material to suppress reflections	Indoors	Absorber and large room are costly
Compact Range	The test antenna is illuminated by the collimated near field of a large reflector	Small space	A large reflector is required
Near-field Range	The magnitude and phase of the near field of the test antenna are sampled and the far field is computed	Very small space	Accurate probe positioning is required Accurate amplitude and phase are required Time-consuming measurements Computer-intensive
GROUND REFLECTION RANGE	The ground between the source and test antennas is reflective, enhancing the indirect ray that interferes with the direct ray, giving a smooth test antenna illumination	Test tower is short Operates well at low frequencies (VHF)	Outdoor weather
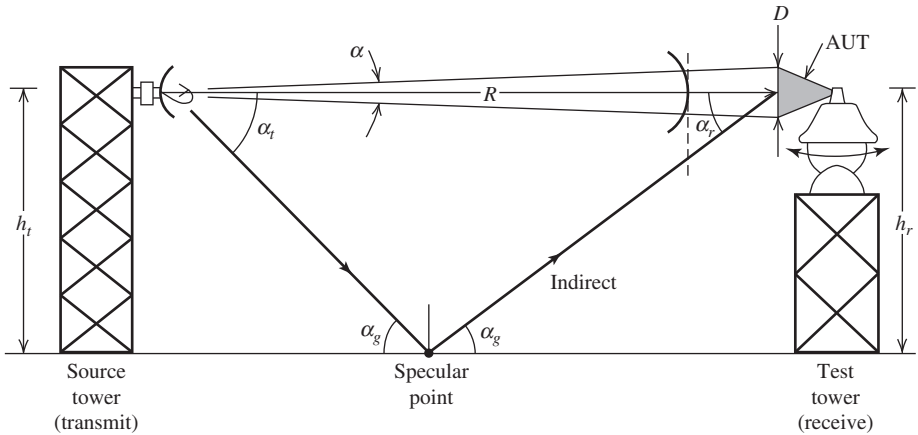
Figure 13-7 The elevated antenna range.

of all antenna ranges is to provide plane wave illumination of the test antenna. Deviations from uniform field illumination amplitude (i.e., magnitude) and phase across the test antenna aperture add to the inherent aperture taper of the test antenna, causing pattern measurement errors. In far-field ranges, the illumination field amplitude variation is determined by the radiation pattern of the source antenna. The effect of increased amplitude taper imposed on the test antenna aperture is to reduce the measured gain and change the side lobes close to the main beam. If the source antenna pattern peak is centered on the test antenna, as it should be in all cases, and the amplitude taper created by the source antenna pattern is -0.25 dB at the edges of the test antenna aperture, there will be a directivity (and thus, gain) reduction of 0.1 dB. [H.7: IEEE Std. 149-1979, p. 19] That is, the pattern point at angle $\alpha/2$ is 0.25 dB down from the peak; see Fig. 13-7. This is difficult to achieve for the wide variety of measurement situations on an antenna range, but in all cases the source antenna should be directed toward the test antenna and have a beamwidth that is as small as possible to reduce illumination of the surroundings and to increase the received signal for adequate dynamic range. At the same time, the source antenna should not have a beamwidth that is narrow enough to impose an amplitude taper across the test antenna.

Phase errors are due to the fact that to achieve a planar phase front from a finite-sized source antenna, the source must be an infinite distance away from the test antenna. The spherical waves from the source antenna cause a phase error across the test antenna extent of D that behave exactly as the far-field distance phase error discussed in Sec. 2.4.3. There we found that spherical waves deviate from parallel rays with a 22.5° phase error ($\lambda/16$ distance error) at a distance of $2D^2/\lambda$. This distance is often taken as the minimum separation distance between the source and test antennas, or

$$R > r_{\text{ff}} = \frac{2D^2}{\lambda} \tag{13-16}$$

The size of the source antenna does not enter into this distance calculation. [2]

The traditional minimum far-field measurement distance in (13-16) is adequate for moderate- to high-gain antennas if high accuracy in the side lobes is not a requirement. A measurement distance that is much greater than $2D^2/\lambda$ produces a nearly planar phase front and thus small phase error. However, very large distances are impractical because of the low power arriving at the test antenna. So it is desirable to keep the measurement distance as small as possible. The effects of decreasing the measurement distance are to

fill in the nulls between side lobes, increase the peak of the side lobes (mainly near the main beam), broaden the main beam, and reduce the main beam peak causing an apparent gain loss. The same effects were noted with horn antennas that experience phase error due to a spherical wave in the aperture. The patterns of a horn in Fig. 9-13 show the null filling, side lobe rise, and gain due to phase error. The quantitative impact of range-distance phase-error effects has been thoroughly studied. [3, 4] For example, a first side lobe of -30 dB relative to the main beam peak with no phase error rises to -27 dB for a $2D^2/\lambda$ measurement distance. The measured gain value is about 0.1 dB low for typical high-gain antennas measured at the $2D^2/\lambda$ distance. For broad main beam antennas, the measurement distance should be at least $2D^2/\lambda$ to ensure accurate measurement of any ripple on the main beam. [5]

When far-field ranges are used indoors, the walls of the room must be lined with absorbing material to reduce reflections generated by the flat wall surfaces. The range is then referred to as an *anechoic chamber*. Most absorbing materials are carbon-impregnated urethane foam. A fire retardant chemical is also included in the absorber. The foam provides a low impedance mismatch to the incident wave, giving a high coefficient of transmission. After a wave enters the absorber, it is attenuated by the carbon. A popular shape for absorber is pyramidal. This shape directs the small fraction of the incident wave reflected from one pyramid toward the next pyramid, giving another opportunity for transmission into the absorber to be attenuated. In some commercial products the faces of the pyramids are custom-shaped to optimize its properties, which often include bandwidth. A second improvement is to use different-sized pyramids in an aperiodic configuration that gives random phase to reflections, which lowers the net amount reflected. Absorber performs better with increasing electrical thickness. Absorber is specified using reflectivity (ratio of reflected power to incident power) for normal incidence. For example, one-wavelength-thick pyramidal absorber has -33 dB reflectivity and -51 dB for eight-wavelength thickness. [H.7: Hemming, p. 31] Using Fig. 13-7 to examine indoor chambers, the ground is now a wall of a chamber and $\alpha_g = 90°$ for normal incidence. Absorber performance degrades (i.e., reflectivity increases) for incidence farther off normal, that is, decreasing α_g. So, reflectivity of -33 dB at normal incidence degrades to -25 dB for $\alpha_g = 40°$. [H.7: Hemming, Fig. 3-7] Room design is important. Suppose the room is L long and $W \times W$ in cross section, where the source and test antennas are near the ends of the L dimension and spaced R apart. To keep incidence as close to normal as possible, the room should be as wide as possible (i.e., W large). At the same time, R (and thus L) should be large to keep the test antenna in the far field of the source antenna. The compromise in these competing goals is to have $\alpha_g \geq 20°$, leading to the following relation [H.7: IEEE Std. 149-1979, p. 30]:

$$R \leq 2.75\,W \quad \textit{Anechoic chamber room dimension design formula} \qquad (13\text{-}17)$$

where

   $R =$ Separation of source and test antenna
   $W =$ Width of chamber that is also W high

Flat ferrite material is sometimes used below 1 GHz, mainly because it is thinner than pyramidal foam.

Rotation positioners are required in most of antenna ranges. Often, a simple *azimuth positioner* (or "turntable") is sufficient. An *elevation-over-azimuth positioner* as illustrated in Fig. 13-8a permits alignment with the source antenna placed at any height. Frequently, a *model tower* is used to position the test antenna over the axis of the rotation positioner; see Fig. 13-8b. Included with the model tower is a *roll positioner* that rotates the test antenna about its own axis for controlling the pattern cut. The source tower also often has a roll positioner for proper orientation of the source antenna polarization.

(*a*) Azimuth-over-elevation.   (*b*) Model tower.
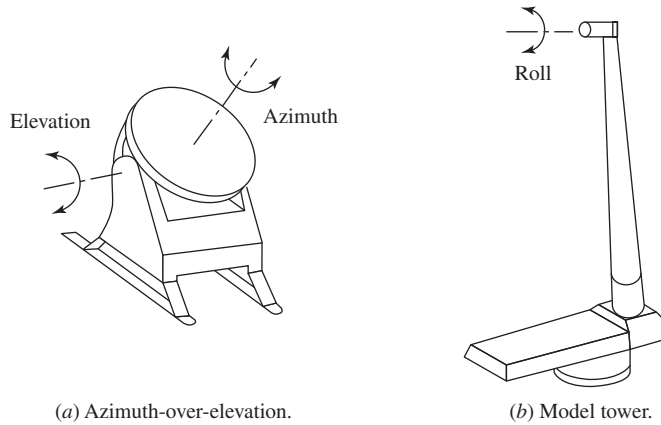
Figure 13-8 Antenna positioners for antenna testing.
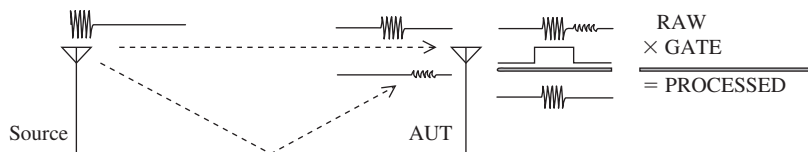
Figure 13-9 Time-domain antenna measurement showing the direct-path pulsed signal and the indirect-path delayed signal. [6] (Reprinted with permission.)

The rotational axis should pass through the phase center of the test antenna, which is the apparent origin of the spherical wave radiated from the antenna.

The instrumentation used with an antenna range varies from a simple signal source together with a relative power indicating subsystem to complete commercial systems with automatic data collection and display features. The signal source should be stable in power level and frequency. The receiving system should have a linear dynamic range of at least 40 dB. Both amplitude-only and amplitude-phase receiving systems are available. The transmitter sends a fixed frequency signal. For the receiver to have wide dynamic range, its bandwidth must be narrowed. The receiver must track the frequency of the drifting transmitted signal or an ultra-stable transmitter is needed. This approach is used for widely separated source and test antenna locations. But, most instrumentation systems use a vector network analyzer (VNA) that provides a source signal and a receiver synchronized to it. A limitation is that there must be cable runs from the VNA to the source and test antenna. The advantages of the VNA are that it steps frequency in increments under user control and provides linear response over wide dynamic range, as much as 100 dB.

The foregoing discussions have been for traditional frequency-domain techniques, but antenna measurements can also be made in the time domain. In *time-domain antenna measurement* the source transmits pulses. The indirect waves created by reflections from the ground (or floor) and objects arrive delayed in time from the direct wave as illustrated in Fig. 13-9. Using time gating, the contribution of the indirect waves is greatly reduced. The pulses can be generated with a VNA or synthetically using Fourier transforming to go back and forth between the frequency and time domains. The steps used in synthetic time-domain processing illustrated in Fig. 13-9 are: (1) The transmitter is swept in frequency; (2) The received signal is transformed to the time domain (RAW signal), time gating is applied (GATE), and an inverse transform returns the signal to the frequency domain with
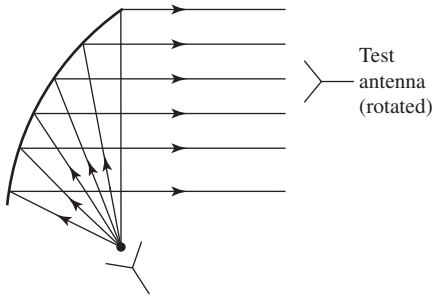
Figure 13-10 The compact range.

the indirect waves gated out (PROCESSED). The outdoor range at Virginia Tech where this is used has an 18-m direct path and a 19-m indirect path via reflection, giving a 1-m path difference corresponding to a 3.3 ns delay. [6]

A *compact range*, so-named because it can fit into a smaller room than a far-field range, is based on an offset reflector, as shown in Fig. 13-10. The room need not be absorber lined, but better performance is obtained with absorber. As noted in Sec. 9.6.1, the fields from the feed at the focal point reflect off the parabolic-shaped reflector and become collimated, providing a flat phase condition parallel to the aperture plane (perpendicular to the page in Fig. 13-10). There is, however, an amplitude taper in the aperture. Increasing the focal length-to-diameter ratio reduces the taper; see (9-208). An offset reflector is used to get the feed out of the way so that the aperture fields are not disturbed. The antenna under test is placed in the center of the collimated beam and rotated to generate pattern data. The region around the test antenna where the amplitude and phase variations are small is referred to as the *quiet zone*. The diameter of the reflector should be about 20λ larger than the diameter of the quiet zone. [H.9.4: Macnamara, p. 284] The obvious disadvantage of the compact range is the large main reflector. The upper operating frequency of the range is limited by the smoothness of the reflector. While the far-field range suffers from a phase error problem, the challenge with a compact range is amplitude error. There are two techniques to reduce the amplitude errors in the quiet zone. First, the edges of the reflector can be serrated to reduce edge diffraction effects in the quiet zone. Second, a dual-reflector can be used (see Sec. 9.6.4) in which both the main reflector and the subreflector are shaped. The subreflector is under-illuminated to avoid its spillover from reaching the quiet zone and is shaped to direct rays to the edges of the main reflector. The main reflector is shaped to correct for phase errors introduced by the shaped subreflector.

The *near-field range* is the smallest range type. In the planar near-field range shown in Fig. 13-11, the test antenna acts as a transmitter and the amplitude and phase are sampled by a receiving probe antenna that is moved over a plane in front of the test antenna. The spatial samples, spaced about a half-wavelength apart, are processed using a fast Fourier transform algorithm to obtain far-field patterns analogous to the radiation integral for finding a far-field pattern from an aperture distribution. Probe effect corrections are included in the processing. The near-field range requires accurate positioning of the probe and accurate amplitude and phase samples. A unique feature of the near-field range is that aperture distribution data are available and are useful in antenna design and evaluation. For example, the excitations of array elements can be examined and adjusted. Usually the test room is absorber-lined, or at least the walls behind the scanner are. The planar system does not require a positioner for the test antenna, but the accurate pattern data is possible only in forward hemisphere. For full pattern coverage, the test antenna is rotated in one or two angles using a cylindrical or spherical scanner, respectively. The spherical near-field scanner is popular for low-directivity antennas, such as handset antennas. Complete near-field measurement systems are commercially available. A variation on near-field scanning is to move the probe antenna around a test antenna that is stationary or on an azimuth
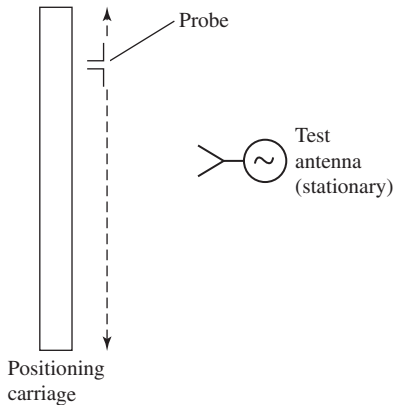
Figure 13-11 The planar near-field range.

rotation table. Commercial systems of various sizes are available for small wireless test articles and for use with antennas mounted on vehicles. Data collection speed is increased with scanning systems by using a dual-polarized probe. For more details on near-field scanning, see the books in Appendix H.7.

The operating principle of the *ground reflection range* is completely different from that of free space ranges. The source and test antenna heights are small and the ground between the towers is constructed to be flat and reflective, which causes the indirect ray to arrive with an amplitude close to that of the direct ray. The indirect ray path distance is not greatly different from that of the direct ray. This gives a slowly varying phase over the test region, which in turn gives a slowly varying interference pattern and a relatively constant field illumination over the test zone. A low test tower height is convenient for large test objects such as antennas on full-scale aircraft. The ground reflection range is sometimes used to test radiation emissions from equipment and systems, and then it is referred to as an *open area test site*. [H.9.4: Macnamara, p. 295]

Finally, we mention that antenna testing is often performed using a *scale model* to scale the test article size so that it can be accommodated in the available facilities. As an example, suppose the design antenna is an array operating at 600 MHz $(\lambda_d = 50\,\text{cm})$ that is 4 m long $(L_d = 5\,\lambda_d)$. It is convenient to test at 6 GHz $(\lambda_t = 5\,\text{cm})$, so the scale model antenna would need to be scaled by the scale factor that is the ratio of the test wavelength to the design wavelength: $s = \lambda_t/\lambda_d = 5/50 = 0.1$. The scale model antenna size is then 40 cm long (4 m × 0.1), which is easily accommodated in most ranges. When the antenna is mounted on a vehicle, the entire vehicle needs to be scaled. The details of the scale model that are a significant fraction of a wavelength at the testing frequency need to be reproduced. The model usually can be made using the same materials as the full-scale model except, in a few cases, for materials that have properties that are not constant with frequency (notably, ε and μ).

13.3 GAIN MEASUREMENT

The pattern measurement techniques discussed in the previous section are relative measurements giving the angular variation of the test antenna's radiation. Also needed is the absolute level of radiation. The parameter of gain aids in quantifying absolute power level but is usually a more difficult measurement than pattern measurement. The definitions and basic principles of gain are found in Sec. 2.5. To review, gain is found from directivity using (2-155), $G = e_r D$, where e_r is the radiation efficiency. When gain measurements include an unavoidable impedance mismatch loss, it is referred to as realized gain (see Sec. 4.4.2). Realized gain is related to gain as $G_r = qG$ where q is impedance mismatch loss and is found using (4-39). Gain alone can be found by

estimating impedance mismatch loss and removing it from the realized gain value. If the test antenna and source antenna are not of the same polarization and aligned properly, there will also be polarization mismatch loss p, which can be found using (4-45) and removed from the measured gain value also. In this section, impedance and polarization matches are assumed unless otherwise noted.

13.3.1   Gain Measurement of Linearly Polarized Antennas

Gain can be found using directivity in the following way. Directivity D is computed using multiple radiation patterns normalized to the same reference level, performing a finite summation to approximate the beam solid angle integral in (2-142), and directivity calculated using $D = 4\pi/\Omega_A$ from (2-144). Then gain is found using $G = e_r D$ and estimating the radiation efficiency e_r, which is unity for many antennas. Automated antenna measurement systems often provide directivity estimated in this fashion.

The most popular gain measurement method is the *gain comparison* (or *gain substitution*, or *gain transfer*) *method*, which is a measurement made relative to an antenna of known gain called a *standard gain antenna*. The gain comparison method measurement uses the arrangement in Fig. 13-12 with a transmitter of fixed input power, P_t, connected to a suitable source antenna whose pattern peak is centered on the test antenna location. The received power is measured using the test antenna, P_T, and for the standard gain antenna, P_S, as illustrated in Fig. 13-12. The source and receive antennas are linearly polarized and aligned in polarization (for example, both horizontally polarized). Often the antennas at each end are on positioners capable of being adjusted in angle to peak the output. The pattern peak of the standard and test antennas must be aimed in the direction of the source antenna. The gain of the test antenna is then easily computed from that of the standard gain antenna multiplied by the ratio of the received powers:

$$G_T = \frac{P_T}{P_S} G_S \tag{13-18}$$

which is expressed in dB as

$$G_T(\text{dB}) = P_T(\text{dBm}) - P_S(\text{dBm}) + G_S(\text{dB}) \tag{13-19}$$

where

$\quad G_T(\text{dB})$ = gain of the test antenna in dB—calculated
$\quad G_S(\text{dB})$ = gain of the standard gain antenna in dB—known
$\quad P_T(\text{dBm})$ = power received by the test antenna in dBm—measured
$\quad P_S(\text{dBm})$ = power received by the standard gain antenna in dBm—measured
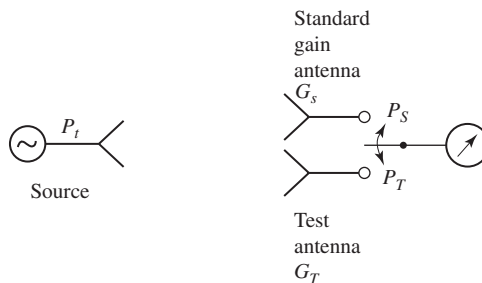
Figure 13-12 Measurement of the gain of a test antenna G_T using the gain comparison method based on the known gain of a standard gain antenna G_S and $G_T = (P_T/P_S)G_S$. (Reproduced by permission from [H.11.1: Stutzman]. © 1993 by Artech House, Inc., Norwood, MA.)

This is an intuitive result which simply says that the gain of the test antenna is larger than the gain of the standard by the increase in received power from the test antenna over that from the gain standard. This result is also easily derived using the Friis transmission formula of (4-34) for both the test and standard antenna cases and subtracting; the terms involving distance R, frequency f, and transmit power P_t are constants and drop out, leaving (13-19). Of course, if the powers are equal $(P_T = P_S)$ the gain of the test antenna is identical to that of the standard gain antenna.

EXAMPLE 13-1   *Gain Measurement by Gain Comparison*

Suppose that a standard gain antenna has a gain of 63, or 18 dB. Following the measurement technique of Fig. 13-12, the measured powers are $P_S = 3.16$ mW or 5 dBm (5 dB above a milliwatt), and $P_T = 31.6$ mW or 15 dBm. The gain of the test antenna is then $G_T = (31.6/3.16)\,63 = 630$, or in terms of decibels

$$G_T(\text{dB}) = P_T(\text{dBm}) - P_S(\text{dBm}) + G_S(\text{dB}) = 15 - 5 + 18 = 28 \text{ dB} \qquad (13\text{-}20)$$

It is obvious from (13-19) that accurate gain measurement requires accurate power measurement. With modern receivers this is often possible. An approach that does not rely on receiver linearity is the RF substitution method in which a precision attenuator is used to establish the level change. That is, the attenuator is adjusted to bring the receiver to the same reading in both cases; then the difference in the corresponding attenuator settings equals $P_T(\text{dBm}) - P_S(\text{dBm})$ that is used in (13-19). Accuracy also depends directly on knowledge of the gain of the standard gain antenna. Popular standard gain antennas are the half-wave dipole for UHF frequencies and below and the pyramidal horn for UHF frequencies and above. The gain of the dipole is quite accurately assumed to be 2.15 dB (see Table 3-2). It is constructed or adjusted to the half-wave resonant length at the desired frequency and fed with balun. Commercially available standard gain horns are provided with a calibrated gain versus frequency curve like the one in Fig. 9-20. If the gain of a horn is not known, it can be calculated using the techniques in Sec. 9.4.

Note that the term gain is synonymous with *absolute gain* or *peak gain*. Gain and pattern data can be merged into a *gain pattern* by multiplying gain by the normalized pattern so that gain as a function of angle can be displayed:

$$G(\theta, \phi) = G\,P(\theta, \phi) = G\big|F(\theta, \phi)\big|^2 \qquad (13\text{-}21)$$

This equation can be expressed in dB by taking 10 log of both sides. The unit of gain in dB is more completely written as dBi to indicate that the gain is referenced to an isotropic antenna.

Gain can also be measured without a standard gain antenna using the *three-antenna method*. The power received is measured for the three possible combinations of transmit and receive antennas, and the Friis transmission formula is used to solve for all three antenna gains. See Prob. 13.3-1.

13.3.2 Gain Measurement of Circularly Polarized Antennas

If a good-quality circularly polarized (CP) source of known gain is available, the gain comparison method of Fig. 13-12 can be used. However, a gain standard should have low cross-polarization, and making a CP antenna with low cross-polarization is more difficult than for an LP antenna. Thus, the gain of CP and elliptically polarized antennas are usually measured using two orthogonal linearly polarized (LP) antennas or one LP antenna used in two orthogonal orientations. The procedure is to use the gain comparison method of (13-19) twice, once for each orthogonal LP source antenna orientation. Suppose the gains

are measured for vertical and horizontal LP source antenna orientations, G_{Tv} and G_{Th}. These measured partial gains are combined to give the total gain [H.11.1: Stutzman, p. 94]:

$$G_T(\text{dB}) = 10\log(G_{Tv} + G_{Th}) \qquad [\text{dBic}] \qquad (13\text{-}22)$$

This procedure is referred to as the *partial gain method*. Any pair of perpendicular orientations can be used because the power in an elliptically polarized wave is contained in the sum of two orthogonal components. As a side note, we observe that a CP antenna performs this sum instantaneously. Therefore, the gain in (13-22) is relative to an ideal CP antenna. The unit dBic indicates gain relative to an isotropic, perfect CP antenna. The accuracy of gain measurement depends on the purity of the source antenna. An LP standard gain antenna usually has an axial ratio of 40 dB or better and does not contribute significantly to gain error.

EXAMPLE 13-2   *Calculation of Gain Using the Partial Gain Method*

Fig. 13-13 gives two patterns measured with an LP source antenna and a nominally CP test antenna, which is a cavity-backed spiral antenna operating at 1054 MHz. Also shown is the pattern of a standard gain horn, which has a gain at 1054 MHz of 14.15 dB based on the manufacturer's gain curve. The receiver gain setting and the source power were constant during these measurements. The peak gains for vertical and horizontal polarizations are

$$G_{Tv}(\text{dB}) = 14.15 - 16.1 = -1.95\,\text{dB}, \quad G_{Th}(\text{dB}) = 14.15 - 13.25 = 0.9\,\text{dB} \qquad (13\text{-}23)$$

because the vertical and horizontal LP pattern peaks are 13.25 and 16.1 dB below the standard gain horn pattern peak, respectively. Then

$$G_{Tv} = 10^{-1.95/10} = 0.64, \quad G_{Th} = 10^{0.9/10} = 1.23 \qquad (13\text{-}24)$$

and (13-22) gives

$$G_T(\text{dB}) = 10\log(0.64 + 1.23) = 2.71\,\text{dBic} \qquad (13\text{-}25)$$
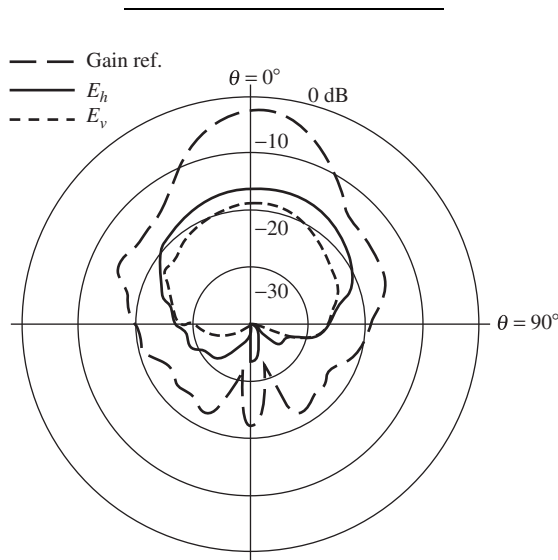
Figure 13-13 Illustration of measurement of the gain of a CP antenna using the partial gain method; see Example 13-2. The patterns are for an LP standard gain horn (long dashed curve) and the nominally CP antenna with the source vertically (solid curve) and horizontally (short dashed curve) polarized. (Reproduced by permission from [H.11.1: Stutzman]. © 1993 by Artech House, Inc., Norwood, MA.)

13.3.3   Radiation Efficiency Measurement

Low-directivity antennas have broad patterns. In fact, in many applications a very broad pattern is needed. For example, mobile terminal antennas must have low directivity in order to respond well to an incoming signal as the terminal's orientation changes. Also, any electrically small antenna will have low directivity with a pattern shaped somewhat like a distorted donut. From Table 3-2, the directivity of a donut pattern is 1.76 dB, and even a half-wave dipole has not much more directivity with 2.15 dB. Thus in practice, the directivity of electrically small antennas and half-wave dipole type antennas will range from around 0 to 2 dB. But radiation efficiency (or simply efficiency) e_r must also be determined to calculate gain using $G = e_r D$.

The Wheeler cap method is the most popular technique for efficiency measurement. The Wheeler cap method uses a conducting hemispherical cap centered on the test antenna which is mounted on a ground plane. [7] The radius of the cap, a, should satisfy the radiansphere condition of $\beta a < 1$, or $a < 0.159\lambda$, from (11-23) and as shown in Fig. 11-15. The size and shape of the Wheeler cap need not closely adhere to these conditions. It can be box shaped and have a length up to $0.5\ \lambda$. [H.7: Arai, p. 56] The two steps in the method are to measure the input resistance with and without the cap. The radiation efficiency from (2-178) is

$$e_r = \frac{R_r}{R_r + R_o} \tag{13-26}$$

where R_r is the radiation resistance and R_o is the ohmic loss resistance. Without the Wheeler cap, the antenna is essentially in free space and the input resistance is $R_{\text{Afs}} = R_r + R_o$; see Fig. 2-15$b$. With the cap on, radiation loss is prevented, leaving only ohmic loss, and the input impedance is $R_{\text{Acap}} = R_o$. So the efficiency is found from these two measurements of R_{Afs} and R_{Acap} as

$$e_r = \frac{R_{\text{Afs}} - R_{\text{Acap}}}{R_{\text{Afs}}} = 1 - \frac{R_{\text{Acap}}}{R_{\text{Afs}}} \tag{13-27}$$

This result can be expressed in terms of reflection coefficients, that is, $\left|S_{11}\right|^2$; see Prob. 13.3-2. The formulations assume no reactance is present. Methods are available where this assumption is not needed. [8]

13.3.4   Gain Measurement of Large Antennas

Electrically large antennas many wavelengths in size often cannot be measured in conventional testing facilities. For example, the Deep Space Network 70-m reflector antenna at Goldstone, California, operating at 2.3 GHz requires a measurement distance from (13-16) of 75 km! An option is to use a source flown on an airplane. Sometimes beacon signals are available on satellites, but the signal can be weak. Here we discuss the *radiometric method*, which uses a radiometric receiver with the test antenna to measure noise temperature.

A discrete "radio star" is used as a source of noise. The noise is incident on the test antenna, operating the receive mode, with power spectral flux density of S, which is the power per unit area per unit bandwidth, $W/m^2/Hz$.[3] An antenna of effective aperture A_e produces output power per Hertz from the noise source of $SA_e/2$, where the factor of ½ is needed because of the polarization mismatch loss between the unpolarized noise and

---

[3] The symbol S is conventionally used for both power density as in Sec. 2.3 and power spectral density as in this section.

polarized antenna. The received power can also be expressed as kT_A from (4-15). Equating gives:

$$\frac{SA_e}{2} = kT_A \qquad [W/Hz] \tag{13-28}$$

Using the expression for gain in terms of effective aperture $G = 4\pi A_e/\lambda^2$ from (4-27), in the above:

$$G = \frac{8\pi kT_A}{S\lambda^2} \tag{13-29}$$

where $k = 1.38 \times 10^{-23}\,\mathrm{J\,K^{-1}}$ is Boltzmann's constant and T_A is the measured antenna noise temperature. The power spectral flux density S of the source must be known and the values for intense discrete radio sources in the sky are available in [H.7: IEEE Standard 149]. The obvious disadvantage of this method is that a radiometer is required to measure the noise temperature. A radiometer can also be used to measure antenna efficiency. [9]

13.3.5 Summary of Gain Determination Methods

Accurate determination of gain is usually very important. Thus, the multiple methods available should be used to make sure an error has not been made in the measurement of gain. Gain estimation methods are valuable. The half-power beamwidths measured during the pattern measurement phase can be used in $G \approx 26{,}000/\mathrm{HP}_{E^\circ}\mathrm{HP}_{H^\circ}$ from (4-57). A second way to estimate gain is using $G = \varepsilon_{\mathrm{ap}}4\pi A_p/\lambda^2$ from (9-77), where the aperture efficiency $\varepsilon_{\mathrm{ap}}$ is estimated based on the antenna type, and the physical aperture area A_p is easily obtained. The most accurate estimation method is pattern integration over several pattern cuts to find directivity and then to estimate efficiency to compute gain with $G = e_r D$. Of course, the main gain evaluation method is direct measurement using the gain comparison or three-antenna method. Electrically large antennas are measured using radiometric methods. The efficiency of electrically small antennas is measured using the Wheeler cap method or the radiometric method.

13.4 POLARIZATION MEASUREMENT

Quite often, the polarization of an antenna can be inferred from the geometry of the active portion of the antenna. For example, the ideal dipole in Fig. 2-4 is vertical linearly polarized since the radiating element is oriented vertically. For gain and co-polarized pattern measurements, the test antenna should be illuminated with a wave of the expected polarization of the antenna: in this case, a vertical linearly polarized wave. Real antennas always have a certain amount of power in the polarization orthogonal to the intended polarization. For the practical realization of the ideal dipole, there will be a small amount of horizontal linear polarization. Such cross-polarization arises from horizontal currents flowing on the antenna or nearby structures. Thus, a complete antenna measurement set includes characterization of the polarization properties of the test antenna. This is often accomplished by making pattern cuts in the E- and H-planes of the test antenna with it both co-polarized and cross-polarized to the source antenna. This is illustrated in Fig. 13-14 for the case of a nominally LP test antenna and an LP source antenna. Of course, the cross-polarized patterns will be much lower in level than the co-polarized patterns, and will be zero for a perfect LP test antenna. Co- and cross-polarized patterns for reflector antennas are discussed in Sec. 9.6.5.

There are three measurement techniques used to characterize an antenna that is elliptically polarized but has an axial ratio that is not large (i.e., the polarization state is not close to pure LP). These methods are discussed in the remainder of this section; also see [H.11.1: Stutzman].

Source antenna (stationary)	Antenna under test (rotated)	Measured pattern

(a)   *E*-plane co-polarized

(b)   *E*-plane cross-polarized

(c)   *H*-plane co-polarized

(d)   *H*-plane cross-polarized
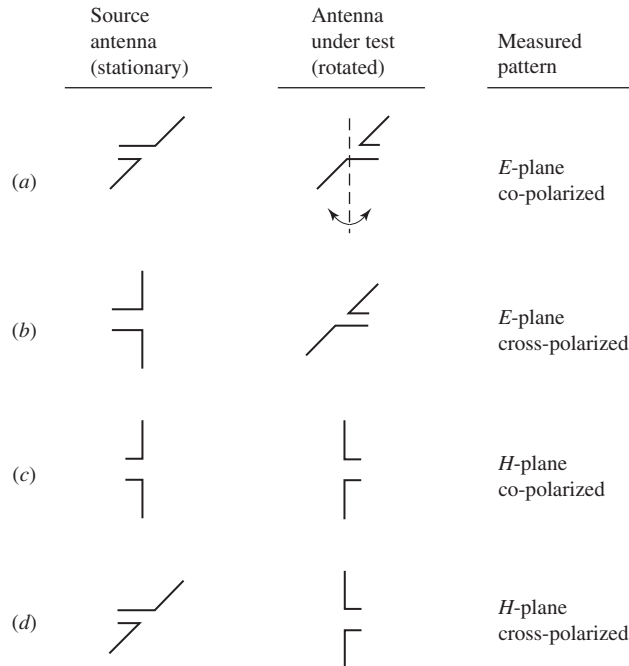
Figure 13-14 Illustration of copolarized and cross-polarized pattern measurement. The source antenna is LP and the test antenna operating in the receiving mode is nominally LP and is rotated about its axis. (Reproduced by permission from [H.11.1: Stutzman]. © 1993 by Artech House, Inc., Norwood, MA.)

13.4.1   The Polarization Pattern Method

A *polarization pattern* is the amplitude response of an antenna as it is rotated about its roll axis. It can be measured at any fixed pattern rotation angle. The resulting pattern shown in Fig. 13-15 is a polar plot of the response of the test antenna as a function of the relative angle α between the illuminating LP wave orientation and a reference orientation of the antenna. Either the LP source antenna is rolled while the test antenna is stationary or vice versa. It is easier to explain the polarization pattern method with the test antenna operated as an elliptically polarized transmitting antenna and the receiving antenna as a linearly polarized probe. Reciprocity permits us to do this. The tip of the instantaneous electric field vector from the test antenna lies on the polarization ellipse and rotates at the frequency of the wave; that is, the electric vector completes f rotations around the ellipse per second. The output voltage of the LP probe is proportional to the peak projection of the electric field onto the LP orientation line at angle α. This is the distance *OP* in Fig. 13-15 projected from the tangent point *T* on the ellipse. The locus of points *P* as the LP probe is rotated is fatter than the ellipse, which is also shown in Fig. 2-20. Of course, for a CP test antenna both curves in Fig. 13-15 are circular.

Note that the maximum and minimum of the polarization pattern are identical to the corresponding maximum and minimum of the polarization ellipse when scaled to the same size. Although the measured polarization pattern does not give the polarization ellipse, it does produce the axial ratio magnitude of the antenna polarization. It is also obvious from Fig. 13-15 that the tilt angle of the ellipse is determined as well. *The polarization pattern gives the axial ratio magnitude |AR| and tilt angle τ of the polarization ellipse, but not the sense*. The sense can be determined by additional measurements. For example, two nominally CP antennas that are identical except for sense can be used as receiving
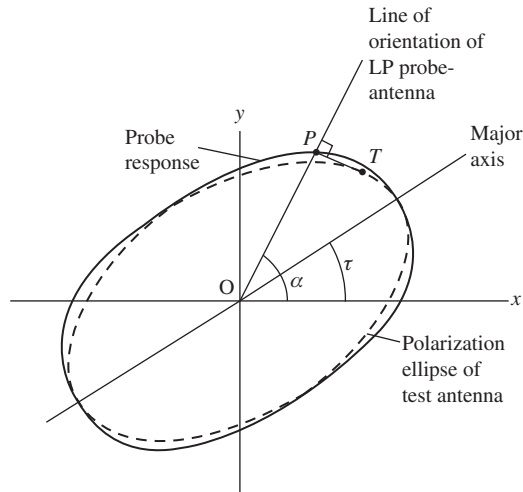
Figure 13-15 Polarization pattern (solid curve) of an elliptically polarized test antenna. It is the response of an LP receiving probe with orientation angle α to a transmitting test antenna with the polarization ellipse shown (dashed curve). (Reproduced by permission from [H.11.1: Stutzman]. © 1993 by Artech House, Inc., Norwood, MA.)

antennas with the test antenna transmitting. The sense of the antenna with the greatest output is then the sense of the test antenna.

The polarization pattern method in many cases is a practical way to measure antenna polarization. If the test antenna is nearly circularly polarized, the axial ratio is near unity and measured results are insensitive to the purity of the LP probe. If the test antenna is exactly circular, tilt angle is irrelevant. In the case of a test antenna that is nearly linearly polarized, axial ratio measurement accuracy depends on the quality of the LP probe, which must have an axial ratio much greater than that of the test antenna.

13.4.2   The Spinning Linear Method

The *spinning linear* (or *rotating source) method* provides a rapid measurement technique for determining the axial ratio magnitude as a function of pattern angle. The test antenna is rotated as in a conventional pattern measurement while an LP probe antenna (usually transmitting) is spun. The spin rate of the LP antenna should be such that the test antenna pattern does not change appreciably during one-half revolution of the LP antenna while the test antenna rotates slowly. An example pattern is shown in Fig. 13-16, which is a pattern of a helix antenna. Superimposed on the antenna pattern are rapid variations representing twice the rotation rate of the probe antenna. For logarithmic (dB) patterns as in Fig. 13-16, the axial ratio is the difference between adjacent maxima and minima at each angle. For example, at a pattern angle of 30° counterclockwise from the main beam axis, the maximum and minimum pattern envelopes are about −8 and −10 dB, corresponding to a 2-dB axial ratio.

Sense cannot be obtained using the spinning linear method. Tilt angle could, in theory, be obtained if probe orientation information were known accurately at the pattern points, but this is usually not done in practice.

13.4.3   The Dual-Linear Pattern Method

A method related to the spinning linear method is the *dual-linear pattern method*. In this method, two patterns are measured for orthogonal orientations of the LP probe source
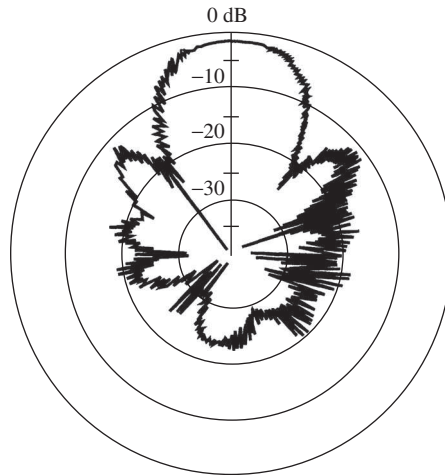
Figure 13-16 Axial ratio measurement as a function of pattern angle using the spinning linear method. The axial ratio is the difference in decibels between adjacent peaks and nulls. The test antenna is a helix antenna operating at X-band and the source is a rotating LP antenna. (Reproduced by permission from [H.11.1: Stutzman]. © 1993 by Artech House, Inc., Norwood, MA.)
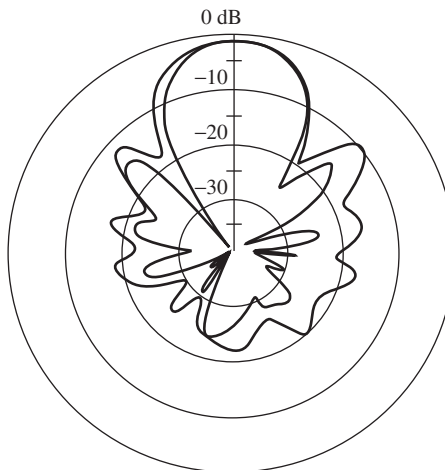
Figure 13-17 Axial ratio measurement as a function of pattern angle using the dual-linear pattern method. The axial ratio is the decibel difference between the two patterns that represent planes containing the major and minor axes of the test antenna polarization ellipse. The test antenna is identical to that in Fig. 13-16. (Reproduced by permission from [H.11.1: Stutzman]. © 1993 by Artech House, Inc., Norwood, MA.)

antenna so that they align with the major and minor axes of the test antenna polarization ellipse. Fig. 13-17 illustrates the resulting patterns for the same sample antenna as in Fig. 13-16 for the spinning linear method. For the same example pattern point at $30°$ counterclockwise from the beam peak, the two linear pattern values are -8 and -10 dB, again giving a 2-dB axial ratio value. Of course, the gains and other equipment settings must remain constant during the entire measurement period.

13.5   FIELD INTENSITY MEASUREMENT

A very small receiving antenna can be used as a *field probe*. Probes are used when it is necessary to measure the spatial amplitude distribution of electromagnetic fields. They are used in antenna measurement facilities such as near-field ranges. The probe must be small relative to the structure whose fields are being measured in order to minimize the disturbance of the fields introduced by the probe itself. The electrically small dipole, in any of its practical forms discussed in Sec. 3.1, is used to probe electric fields. The small loop is used as a magnetic field probe. In this section, we show how to calculate field intensity from probe measurements along with other aspects of field measurements.

An antenna pattern can be measured by moving a receiving antenna around a transmitting test antenna at a fixed distance from it in the far field; this is the relative field strength variation. Absolute field level is measured with a receiving antenna at a specified distance from a transmitting device. Such measurements are made for compliance to regulations and are necessary at low-frequency bands because the effects of terrain and the real Earth surface are difficult to calculate. If the gain of the measuring antenna is known (it usually is) and the voltage developed across its terminals is measured, the field strength incident upon the measuring antenna can be calculated, as we now discuss.

The same model as in Fig. 4-1b is used to derive field intensity. The power delivered to the terminating load is

$$P_D = \frac{1}{2}\frac{|V_A|^2}{R_T} = \frac{V_{A,\text{rms}}^2}{R_T} \tag{13-30}$$

where $V_{A,\text{rms}} = |V_A|/\sqrt{2}$, since V_A is a peak quantity. The field form of the delivered power expression from (4-35) and (4-26) is

$$P_D = pq\,S_{\text{av}}\,A_e = pq\frac{(E_{\text{rms}}^i)^2}{\eta}A_e \tag{13-31}$$

Equating these two relations yields

$$(E_{\text{rms}}^i)^2 = \eta\frac{V_{A,\text{rms}}^2}{pqR_L}\frac{1}{A_e} = \eta\frac{V_{A,\text{rms}}^2}{pqR_L}\frac{4\pi}{G\lambda^2} \tag{13-32}$$

where (4-27) was used for A_e. Converting wavelength to frequency using $\lambda = c/f$ and expressing the relation in decibels by taking 10 log of both sides gives

$$\begin{aligned} E_{\text{rms}}^i(\text{dB}\mu V/m) = {}& V_{A,\text{rms}}(\text{dB}\mu V) + 20\log f(\text{MHz}) - G(\text{dB}) \\ & - 10\log R_L - 10\log p - 10\log q - 12.8 \end{aligned} \tag{13-33}$$

This expression permits easy calculation of electric field intensity E_{rms}^i in decibels relative to 1 $\mu V/m$, using the voltage $V_{A,\text{rms}}$ in decibels relative to 1 μV, measured at the terminals of a probe antenna with gain G. Gain loss due to mispointing can also be included. For example, suppose the probe antenna has 6-dB gain and is pointed so that the incoming wave arrives from a direction on the receiving antenna pattern that is 2 dB below its maximum. Then 4-dB gain is used in (13-33) rather than the peak gain of 6 dB.

EXAMPLE 13-3   *Sensitivity of an FM Receiver*

As an example, suppose the antenna and transmission line input impedances are both 300 Ω. Then (13-33) becomes

$$E_{\text{rms}}^i(\text{dB}\mu V/m) = 20\log f(\text{MHz}) - G(\text{dB}) + V_{A,\text{rms}}(\text{dB}\mu V) - 37.57 \tag{13-34}$$

To be specific, consider a typical FM broadcast receiver with a sensitivity of $1\,\mu V$; that is, minimum satisfactory performance is produced when the value of $V_{A,\mathrm{rms}}$ is $1\,\mu V$, or $0\,\mathrm{dB}\mu V$. The most popular antenna for FM receivers is the half-wave folded dipole (see Sec 6.2) that has a real impedance of about $300\,\Omega$ and a gain of $2.15\,\mathrm{dB}$. At a frequency of $100\,\mathrm{MHz}$, the incident field intensity required for minimum satisfactory performance from (13-34) is $0.28\,\mathrm{dB}\,\mu V/m$, or $1.03\,\mu V/m$.

Regulatory agencies often specify the maximum field strength radiated from a device at a fixed distance away. For example, the FCC places a maximum value on the electric field strength radiated from ISM devices at a distance of $3\,\mathrm{m}$ and places limits on the electric field strength from AM broadcast radio stations $1\,\mathrm{km}$ away from the antenna as well as radiated power. Commercially available antennas for electromagnetic compliance are calibrated using manufacturer provided values of antenna factor over the operating range of the antenna.

At frequencies below $1\,\mathrm{GHz}$, antenna measurements are made by illuminating the test antenna with a known field intensity and measuring the terminal voltage. Antenna factor is used to quantify this measurement. Antenna factor was introduced in Sec. 4.2 and is defined as the ratio of the field intensity illuminating the antenna to the received voltage across the antenna terminals:

$$K = \frac{E^i}{V_A} \quad [\mathrm{m}^{-1}] \tag{13-35}$$

This is an electric field antenna factor; a corresponding one involving magnetic field intensity is also in use. Antenna factor is often used to determine receiver sensitivity. Then, (13-35) in decibel form using (13-33) becomes

$$E^i_{\mathrm{rms}}(\mathrm{dB}\mu V/m) = \text{receiver sensitivity} = V_{A,\mathrm{rms}}(\mathrm{dB}\mu V) + K(\mathrm{dB/m}) \tag{13-36a}$$

where

$$K(\mathrm{dB/m}) = 20\log\left[f(MHz)\right] - G(\mathrm{dB}) - 10\log R_L$$
$$- 10\log p - 10\log q - 12.8 \tag{13-36b}$$

It is common to specialize this definition to $R_L = 50\,\Omega$, since that is the normal receiver input impedance. Antenna factor includes impedance mismatch effects and antenna gain. The polarizations of the wave and antenna are usually assumed to be matched (i.e., $q = 1$), which is the customary measurement situation.

EXAMPLE 13-4   *Sensitivity of an FM Receiver Using Antenna Factor*

We repeat Example 13-3 using antenna factor. Substituting $R_A = Z_o = 300\,\Omega$, $G = 1.64$ and $\lambda = 3\,\mathrm{m}$ in (13-36b) gives

$$K = 20\log(100) - 2.15 - 10\log(300) - 0 - 0 - 12.8 = 0.28\,\mathrm{dB/m} = 1.03\,\mathrm{m}^{-1} \tag{13-37}$$

Then for $1\,\mu V$ sensitivity, (13-36a) gives

$$E^i_{\mathrm{rms}} = 0\,\mathrm{dB}\,\mu V + 0.28\,\mathrm{dB/m} = 0.28\,\mathrm{dB}\,\mu V/m \tag{13-38}$$

which is the result we obtained in Example 13-3.

13.6 MOBILE RADIO ANTENNA MEASUREMENTS

Physically small antennas and handset antennas for mobile communications are difficult to measure accurately. The cable connecting to the antenna under test can carry unbalanced currents that radiate, leading to highly inaccurate measured values. A standard method to avoid this problem is to make a monopole version of the antenna mounted on a large ground plane, and then the cable below the ground plane will have little effect. However, this technique is not applicable to handset antennas. Ferrite beads on the cable help but often do not solve the problem of spurious radiation. An alternative approach is to measure the antenna in the transmit mode using a battery-operated transmitter embedded in the handset. This eliminates the cable connection and its associated problems. In this section, we present techniques for measuring a small antenna and a full terminal, or more generally the *device under test* (DUT). Also, the associated figure-of-merit quantities used in the mobile communications industry are explained.

A quantity that includes the gain, radiation pattern, polarization, and multipath due to the propagation environment is *mean effective gain* (MEG). MEG is measured by moving a mobile terminal over a prescribed course that has multipath effects and averaging the signal received from a base station. The measured MEG represents the interaction of the antenna and propagation along the course. Comparing to the same measurement with a reference antenna, like a half-wave dipole, moved over the same course gives a relative measure of the test antenna performance. Because of the varying propagation conditions, measured MEG is not generally used. Instead, a theoretical MEG is computed by integrating the gain pattern over the sphere surrounding it weighted by an assumed angular probability density distribution for the incoming waves. For more details on MEG see [H.9.1: *Mobile Sys. Hdbk.*].

Full mobile equipment testing avoids having to isolate the antenna effects, and at the same time provides an end-to-end test. In the past, this was done by a cable connection from a source to a receiving terminal. For more realistic conditions an *over-the-air test system* (OATS) is used to provide end-to-end wireless testing. There are two primary measured quantities. Total radiated power (TRP) is used for the DUT transmitting and total isotropic sensitivity (TIS) for receiving. TRP and TIS were introduced in Sec. 11.6.4, and here we provide their quantitative relationships.

TRP is tested with the DUT, usually a handset, transmitting at rated power output. The received power at a fixed distance in an OATS is measured in both polarizations. Samples are collected over the sphere surrounding the DUT, commonly in 15° increments in azimuth and elevation. The data samples are then summed approximating the integral:

$$\text{TRP} = \frac{1}{4\pi} \iint [P_t \, G_\theta(\theta, \phi) + P_t \, G_\phi(\theta, \phi)] \, d\Omega \qquad (13\text{-}39)$$

For an antenna of gain G_t and normalized power pattern $|F(\theta, \phi)|^2$, the preceding result reduces to $P_t G_t \left[\iint |F(\theta, \phi)|^2 \, d\Omega \right] / 4\pi$, which using (2-142) equals $P_t G_t \Omega_A / 4\pi$. Noting that $P_t G_t = \text{EIRP}$ and $\Omega_A / 4\pi = D$ gives

$$\text{TRP} = \frac{\text{EIRP}}{D} \qquad (13\text{-}40)$$

With a handset as the DUT transmitting 24 dBm, a typical TRP value is 27 dBm. [10]

TIS is measured with the DUT terminal receiving a signal from a transmitter a fixed distance away. Samples are collected over the sphere surrounding the DUT, commonly in 30° increments in azimuth and elevation. At each point, the transmit power is adjusted until the DUT performance is just acceptable. For a handset this is the threshold bit error rate. The recorded power levels in each polarization, $P_{\text{is}\theta}$ and $P_{\text{is}\phi}$, are then summed approximating the integral:

$$\text{TIS} = \frac{4\pi}{\iint \left[\frac{1}{P_{\text{is}\theta}(\theta, \phi)} + \frac{1}{P_{\text{is}\phi}(\theta, \phi)} \right] d\Omega} \tag{13-41}$$

Note that if the powers measured have the same value of P_{is}, $\text{TIS} = 4\pi / \iint (1/P_{\text{is}}) d\Omega = P_{\text{is}}$, as expected. A typical TIS value for a handset is about -105 dBm. [10]

TIS and TRP are usually measured in an anechoic chamber. Drive tests are also used where measurements are made on a device as it is moved in a vehicle, carried by an operator, or moved on a mechanical track. However, a large number of drive tests are needed to get an average result. A faster and less expensive approach is to perform measurements in a *reverberation chamber*, which is a metal-box structure with a mode-stirring device similar to that used in a microwave oven. Reverberation chambers have been used for interference and compatibility testing of electronic products. More recently, the reverberation chamber is being used with wireless communication terminals to create a multipath environment as encountered with indoor and urban propagation conditions. [11]

13.7   RULES FOR EXPERIMENTAL INVESTIGATIONS

The order in which the experimental procedures described in this chapter are performed is important. The preferred order for most antennas is:

1. Measure impedance match over the intended operating band
2. Measure the radiation patterns
3. Measure gain
4. Measure the polarization state

Of course, the order can vary with the antenna tested and the measurement facility used. However, impedance measurement is a necessary first step. If the antenna does not provide the required input impedance, the subsequent tests will be meaningless. And the design cycle needs to iterate until the impedance is within acceptable limits. It is best to measure the full impedance (i.e., resistance and reactance) swept over frequency (usually displayed on a Smith chart), which is useful in understanding the operation of the antenna. But, specifications are most often in terms of VSWR or return loss. If possible, the impedance measurement should be made with the test article in an anechoic chamber to eliminate effects of surroundings. But often the test is made by placing the device in an absorber-lined box, or aiming the test antenna toward a room corner with absorber on the walls. For small devices and antennas, it is possible to make remote impedance measurements without any connecting cable. [12] For more on small antenna measurement see [H.8.11: Hirasawa and Haneishi, Chap. 9].

Once an acceptable impedance match is achieved, the radiation patterns can be measured, followed by gain and polarization using the procedures in this chapter. It may be convenient to combine measurement steps. The patterns and the gain can be measured simultaneously if the reference level of the pattern includes the gain. Also, polarization characterization can be combined with the pattern measurement, as shown in Figs. 13-16 and 13-17. With modern antenna measurement instruments, the operator can choose from a range of data display formats and types after data collection is complete.

Gain measurement is important because the impedance match can be good and the patterns can look normal but with the gain being extremely low. This can happen for an antenna that is very lossy. Remember that a dummy load has a VSWR of unity. To reemphasize, *just because the* VSWR *is low does not mean the antenna is operating well.*

REFERENCES

1. C. T. Tai, "A Critical Study of the Circuit Relations of Two Distant Antennas," *IEEE Trans. on Ant. & Prop. Mag.*, Vol. 44, pp. 32–37, Dec. 2002.
2. T. Milligan, "Exact Derivation of Aperture Errors in Antenna Measurements," *IEEE Trans. on Ant. & Prop. Mag.*, Vol. 38, pp. 56–60, Aug. 1996.
3. P. Hacker and H. Schrank, "Range Distance Requirements for Measuring Low and Ultra Low Sidelobe Antenna Patterns," *IEEE Trans. on Ant. & Prop.*, Vol. AP-30, pp. 956–966, Sept. 1982.
4. R. C. Hansen, "Measurement Distance Effects on Low Sidelobe Patterns," *IEEE Trans. on Ant. & Prop.*, Vol. AP-32, pp. 591–594, June 1994.
5. D. Hundt and W. Stutzman, "Pattern Measurement Distance for Broad Beam Antennas," *J. Electromagnetic Waves & App.*, Vol. 8, pp. 221–235, Feb. 1994.
6. A. Predoehl, "Time Domain Antenna Pattern Measurements," M.S. thesis, Virginia Tech, 1996.
7. H. A. Wheeler, "The Radiansphere around a Small Antenna," *Proc. IRE*, Vol. 47, pp. 1325–1331, Aug. 1959.
8. R. Johnston and J. McRory, "An Improved Small Antenna Radiation-Efficiency Measurement Method," *IEEE Trans. on Ant. & Prop. Mag.*, Vol. 40, pp. 40–47, Oct. 1998.
9. D. Pozar and B. Kaufman, "Comparison of Three Methods for the Measurement of Printed Antenna Efficiency," *IEEE Trans. on Ant. and Prop.*, Vol. 36, pp. 136–139, Jan. 1988.
10. M. Rumney, "The MIMO Antenna: Unseen, Unloved, Untested," *Microwave J.*, Vol. 53, pp. 22–34, Aug. 2010.
11. B. Furht and S. Ahson, Eds., *Long Term Evolution: 3GPP LTE Radio and Cellular Technology*, CRC Press, 2009. Chap. 12.
12. T. Yang and W. A. Davis, "Remote Measurements of Input Impedance and Radiation-Q of Handheld Antennas," URSI Radio Science Meeting (Ottawa, CA), July 2007.

PROBLEMS

13.1-1 Let sources $\mathbf{J}_a, \mathbf{M}_a, \mathbf{J}_b$ and $\mathbf{M}_b$ all be of the same frequency in a linear medium. The following steps lead to the Lorentz reciprocity theorem:
(a) Maxwell's equations for sources *a* are

$$\nabla \times \mathbf{E}_a = -j\omega\mu\,\mathbf{H}_a - \mathbf{M}_a$$

$$\nabla \times \mathbf{H}_a = j\omega\varepsilon\,\mathbf{E}_a + \mathbf{J}_a$$

Similar equations can be written for sources *b*. Manipulate these four equations and use the vector identity (C-19) to show that

$$\nabla \cdot (\mathbf{E}_a \times \mathbf{H}_b - \mathbf{E}_b \times \mathbf{H}_a) = \mathbf{E}_b \cdot \mathbf{J}_a + \mathbf{H}_a \cdot \mathbf{M}_b - \mathbf{H}_b \cdot \mathbf{M}_a - \mathbf{E}_a \cdot \mathbf{J}_b$$

(b) Integrate the above equation over a volume *v* enclosing all sources, employ the divergence theorem (C-23) for the left-hand side, and let the volume extend to infinity. Then the fields arriving at the surface of the volume behave like spherical waves, and the TEM wave relationships can be employed to show that the left-hand side is zero, leading to a proof of (13-1).

13.1-2 Use the reciprocity theorem form of (13-2) to show that the distant field of *any* finite electric current distribution in free space can have no radial component.

13.1-3 Since any two-port network can be reduced to an equivalent T-section, the general antenna system of Fig. 13-2*a* can be modeled as shown in the figure. First, excite terminals *a* with a current source I_a and find the open circuit output voltage $V_b|_{I_b=0}$. Then, excite terminals *b* with a current source I_b and find the open circuit output voltage $V_a|_{I_a=0}$. From these relationships, find Z_{ba} and Z_{ab}; they will, of course, be equal.
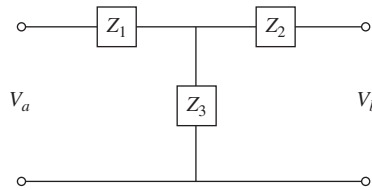
13.1-4 Write the voltage equations for the network representation of Prob. 13.1-3 and compare to (13-6) show that the T-section impedances are $Z_1 = Z_{aa} - Z_m$, $Z_2 = Z_{bb} - Z_m$, and $Z_3 = Z_m$.

13.1-5 If antennas a and b are identical, how is the network of Probs. 13.1-3 and 13.1-4 simplified?

13.1-6 Reciprocity can also be shown with voltage generators and short circuit currents:

(a) Drive terminals a of the network in Prob. 13.1-3 with a voltage generator V_a and short circuit terminals b. Find the expression of $V_a/I_b|_{V_b=0}$ in terms of Z_1, Z_2, and Z_3. Then drive terminals b with voltage source V_b while short circuiting terminals a. Find $V_b/I_a|_{V_a=0}$. It should equal $V_a/I_b|_{V_b=0}$.

(b) Find the same transfer impedance expressions in terms of Z_{aa}, Z_{bb}, Z_{ab}, and Z_{ba} from (13-6). Show that they are equal if $Z_{ab} = Z_{ba}$.

(c) Using $Z_1 = Z_{aa} - Z_m$, $Z_2 = Z_{bb} - Z_m$, and $Z_3 = Z_m$ from Prob. 13.1-4 show that the transfer impedance expressions of (a) are the same as those of (b).

13.1-7 Using the model of Fig. 13-2a, excite antenna a with voltage V_a and prove that the power received in antenna b, which is terminated in load R_L, is proportional to $|Z_m|^2$ for antenna b in the far field of a.

13.2-1 *Anechoic chamber design.* An anechoic chamber with a separation distance between the source and test antenna of 7 m is to be used to measure the pattern of a 2-dB gain antenna. Measurements are to be made from 1 to 12 GHz and a receiver dynamic range of 45 dB is required for the pattern. The receiver has a sensitivity of -110 dBm at 1 GHz and -95 dBm at 12 GHz. The transmitter power is 10 dBm. Find the gain of the source antenna (constant over the band) in order to keep the received signal power above the receiver sensitivity by an amount equal to the dynamic range.

13.3-1 *The three antenna method of gain measurement.* (a) Derive expressions for the gains of three antennas denoted 1, 2, and 3 with unknown gains G_1, G_2, and G_3 in terms of the powers measured with all three combinations of antennas used for transmit and receive: P_{12}, P_{23}, and P_{13}, as well as the transmit power P_t and the free space loss L_{fs}. Start by writing the Friis transmission equations for each case in dB form. (b) Reduce the equations to the gain comparison method result of (13-19) by using antennas with $G_1 = G_t, G_2 = G_T$, and $G_3 = G_S$.

13.3-2 Derive a radiation efficiency expression for the Wheeler cap method using measured values of $|S_{11fs}|^2$ for free space and $|S_{11cap}|^2$ with the cap on. Start with the power form of the efficiency equation in (2-177).

13.3-3 Receiving terminals are often quantified with a figure-of-merit called "*G*-over-*T*," which is the antenna gain divided by the system noise temperature expressed in dB:

$$\frac{G}{T}\Big|_{dB} = G(dB) - 10 \log\left[T_{sys}(K^\circ)\right]$$

Calculate the G/T for the direct broadcast satellite terminal in Examples 4-3 and 4-4.

13.5-1 Derive (13-33) from (13-32).

13.5-2 A voltage of 200 μV (peak) is required at the input of an FM broadcast receiver for acceptable performance. The receiver input impedance is 300 Ω (real). The antenna is a linearly polarized folded dipole with an input impedance of $300 + j0\Omega$ and has negligible loss. The antenna has a gain of 1.64 and is oriented for maximum received signal. The connecting transmission line is a 300-Ω twin lead. (a) What are the radiation and impedance efficiencies e_r and q? (b) If the radio station transmitting antenna is circularly polarized, find the minimum peak electric field in μV/m incident on the receiver required for reception at 100 MHz.

13.5-3 (a) Derive an expression for the antenna factor using (13-32) in (13-35). (b) Use this relation to derive (13-36*b*). (c) Show that (a) reduces to the following popular formula for a matched, 50-Ω system:

$$K = \frac{9.73}{\lambda\sqrt{G}} \text{ m}^{-1}$$

13.5-4 Evaluate the antenna factor of a matched antenna operating at 30 MHz with a gain of 3 dB and terminated with a 50-Ω resistor.

13.5-5 An AM broadcast tower is a 75-m-tall monopole and transmits 50 kW at 1000 kHz. Make reasonable assumptions needed to calculate the electric field 1 km away and compare to the FCC minimum of 141 mV/m.

Chapter **14**

CEM for Antennas: The Method of Moments

14.1 GENERAL INTRODUCTION TO CEM

Computational electromagnetics (CEM) is broadly defined as the discipline that intrinsically and routinely involves the use of a digital computer to obtain numerical results for electromagnetic problems. [1] It is a third tool available to electromagnetics engineers, the other two being mathematical analysis (which we employed in the early chapters of this book) and experimental observation (Chap. 13). It is not uncommon to verify analysis results and CEM results with experimental results, nor is it uncommon to employ analysis and/or CEM to understand experimental results. As CEM becomes more mature, it is increasingly being used to design experiments and provides a cost-effective alternative to solving electromagnetic problems when experimental results cannot be readily obtained.

CEM has developed a number of different techniques in recent decades, with a variety of commercial software based on them for the simulation of electromagnetic problems. It is very advantageous to understand the principles underlying the method being used in order to know how to construct an efficient simulation model for a particular application and how to set parameters in the software for the best accuracy and computation efficiency. In the remaining chapters of this book, we will examine in detail a representative sampling of CEM methods.

There are various ways to classify the assortment of techniques in CEM. First we will divide CEM into two major categories: full-wave methods based upon "first principles" (formerly called numerical methods) and high-frequency methods, as shown in Fig. 14-1. Full-wave methods involve the discretization of the computational domain, which includes the antenna or scatterer and in certain methods also the surrounding space, followed by the generation of a matrix equation of the form $[A][x] = [B]$ and its (symbolic) solution $[x] = [A]^{-1}[B]$ in the frequency domain (Chap. 14), or a marching onward in space and time scheme in the time domain (Chap. 15). These techniques are applied most easily to objects whose size in three dimensions is within an order of magnitude or two of the wavelength. On the other hand, high-frequency methods, based upon asymptotic expansions or assumptions about the current, are best suited to objects that are many wavelengths in extent (see Fig. 14-2). These methods are considered in Chap. 16.

Under the category of full-wave methods, techniques can be classified as partial differential equation (PDE) based or integral equation (IE) based, since Maxwell's equations can be expressed in either differential or integral form, respectively. PDE techniques divide the antenna or scatterer and the space around it into small segments, patches, or cells which interact with nearby segments, patches, or cells. Since surrounding space
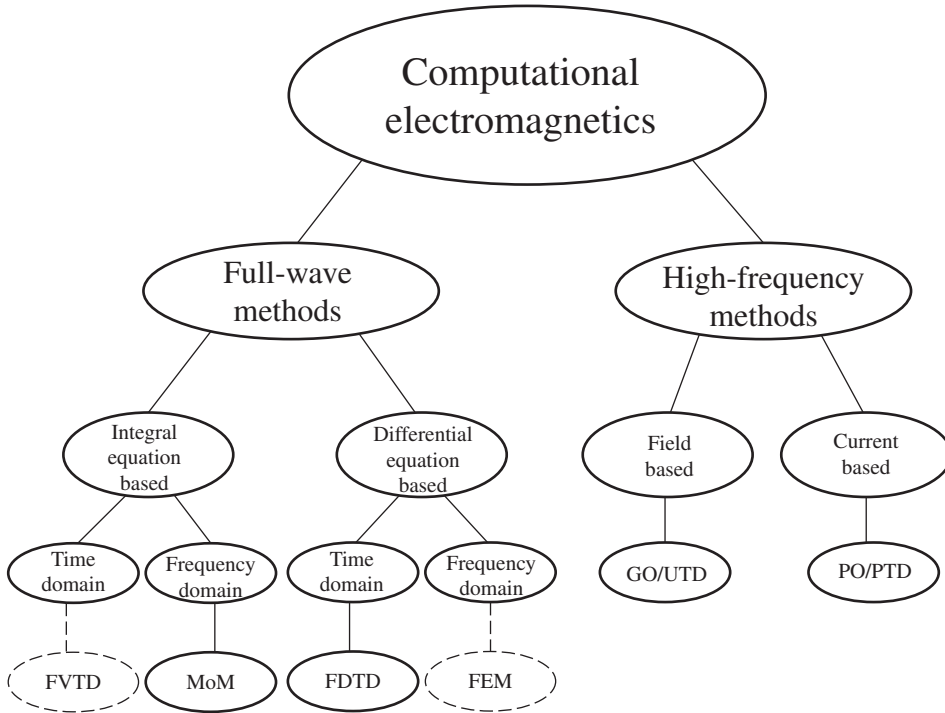
Figure 14-1 Categories within computational electromagnetics.
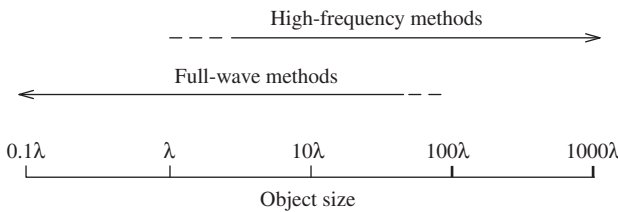
Figure 14-2 Regions of applicability for the major categories of Fig. 14-1.

cannot be subdivided out to infinity, an absorbing boundary condition (ABC) must be established around the object or objects to simulate infinity in radiation problems. In contrast to this, IE techniques subdivide only the object or objects in the problem space but not the space itself. Each subdivision interacts with all other subdivisions. The equivalence theorem is used to produce equivalent currents radiating in free space out to infinity. There are three types of integral equations: electric field integral equation (EFIE), magnetic field integral equation (MFIE), and combined field integral equation (CFIE), the latter being needed to overcome resonance phenomena.

Under the IE and PDE categories, there is a distinction between time domain and frequency domain techniques. Because frequency domain techniques solve Maxwell's equations one frequency at a time, they are usually the natural choice for problems with narrow bandwidths, in contrast to time domain techniques, which often calculate the system impulse response, yielding, via Fourier transformation, data at multiple frequencies. This makes time domain techniques relatively efficient for many wide bandwidth problems.

When a problem is large in terms of the wavelength, full-wave methods become less attractive because calculating the unknown currents or fields requires solving a very large system of linear system equations or marching fields through a very large number of time steps. This can put excessive demands upon computer resources. For such problems, high-frequency methods provide a reasonable alternative (see Fig. 14-2). High-frequency methods can be either field based or current based (see Fig. 14-1). Field-based methods use ray optics to account for the reflection, refraction, and diffraction of electromagnetic fields. Current-based methods start with a basic assumption about the relationship between the current and the surface fields and then integrate over the current, and possibly its refinements, to obtain the radiated field. In high-frequency methods, there is no need to solve for the unknown currents or fields on the antenna or scatterer as is required in full-wave methods, and therefore computer resources are often less of an issue. On the other hand, we note that some high-frequency current-based simulations require tens of thousands of facets on the radiating object because of its electrical size, resulting in a very large problem to be simulated.

There are many techniques for modeling radiation problems. The following sections of this book in Chaps. 14, 15, and 16, will examine four representative techniques in detail. These are:

1. Method of moments (MoM)
2. Finite difference time domain (FDTD)
3. Geometrical optics (GO) with the uniform theory of diffraction (UTD)
4. Physical optics (PO) with the physical theory of diffraction (PTD)

These are not the only techniques in use today, but they probably represent the ones most commonly used, other than the popular finite element method (FEM). There are many FEM books, several of which are directed toward antennas [H.10.5: Volakis; H.10.5: Jin]. Other popular CEM techniques are the finite volume time domain technique (FVTD), the finite integral technique (FIT), and the transmission line matrix (TLM) method, all of which have similarities to FDTD.

Another category of methods are hybrid methods, which combine several techniques in order to capitalize on the strengths of each while avoiding the weaknesses of each. Some of these are MoM-UTD, MoM-PO, FEM-MoM, FEM-PO, and FEM-FDTD. One MoM-UTD method is discussed in Chap. 16.

Clearly there is a rich assortment of techniques available for solving a wide range of antenna problems. An attempt to summarize the usual applicability of some of the more common techniques is presented in Table 14-1. Note that FDTD has similarities to FVTD, FIT, and TLM. Obviously no one technique is best for all problems.

Table 14-1 Representative Applicability of Some CEM Techniques

	Large Apertures	Wire Ant.	Printed Ant.	Complex Media	Resonance Phenomena
MoM	3[b]	1	2	2	1[e]
FDTD[a,c]	3	2	1	1	1
FEM[a]	3	2	1	1	2
GO/UTD	1	4	4	3[d]	4
PO/PTD	1	4	4	3[d]	4

[a] ABC required, free space discretized   1 – very good
[b] Fast solver required   2 – good
[c] Similar to FVTD, FIT, TLM   3 – sometimes applicable
[d] Via indirect means   4 – not applicable
[e] CFIE required

The next section begins the discussion of the method of moments. The focus is on antennas composed of wires or metallic tubes. The early use of the moment method in electromagnetic problems dealt mostly with wire geometries. The wire geometry is a one-dimensional problem. In recent years MoM has found application to the more complicated problem of surfaces of arbitrary shape discretized into surface patches, a two-dimensional MoM problem. The chapter will conclude with a brief discussion of surface patches. Formulations for the discretization of dielectric volumes, a three-dimensional MoM problem, have also become available.

14.2   INTRODUCTION TO THE METHOD OF MOMENTS[1]

Thus far, we have studied a variety of antenna configurations, but for the most part we have assumed either that the current distribution was uniform (e.g., the ideal dipole) or sinusoidal. It was then a relatively straightforward procedure to obtain the near- and far-zone fields created by the current.

In the chapter, we eliminate the need for assuming the form of the current distribution. Naturally, this greatly expands the number of antenna configurations that can be investigated. Indeed, we are then able to study, for example, wire antennas of almost arbitrary configuration. The methods we use to do this are, therefore, very general methods capable of yielding answers whose accuracy is within the limit of experimental error. The potential price for using such powerful methods lies in the effort required to write the necessary software, the time required for computer execution, and the effort required for validation. Fortunately, cost-effective software is readily available and it is not necessary to write software from "scratch." However, a reasonable understanding of the principles on which the electromagnetic software is based is necessary in order to avoid its misuse and the misinterpretation of results.

Consider the wire antenna along the z-axis in Fig. 14-3. A generic form for an integral equation describing such as antenna is

$$-\int I(z')K(z,\ z')\ dz' = E'(z) \tag{14-1}$$

The kernel $K(z,\ z')$ depends on the specific integral equation formulation used; the popular Pocklington form is presented in the next section.

Electromagnetic radiation problems can always be expressed as an integral equation of the general form in (14-1) with an inhomogeneous source term on the right and the unknown within the integral. However, it was not until the availability of modern high-speed digital computers in the mid-1960s that it was feasible to solve most such equations. Since that time, many MoM procedures and codes have been developed [1–3].

MoM is a solution procedure for approximating an integral equation, such as that in (14-1), with a system of simultaneous linear algebraic equations in terms of the unknown current $I(z')$. [H.10.2: Harrington] Then, as we have seen in the previous chapters, once the current is known, it is a fairly straightforward procedure to determine the radiation pattern and impedance.

In this chapter, we set forth the basic principles involved in solving integral equations via MoM and demonstrate the procedure with several examples. The serious student is encouraged to use these basic principles to write a simple computer code, such as that suggested by Prob. 14.5-1.

---

[1] In disciplines other than electromagnetics, the method of moments is known as part of the boundary element method (BEM).
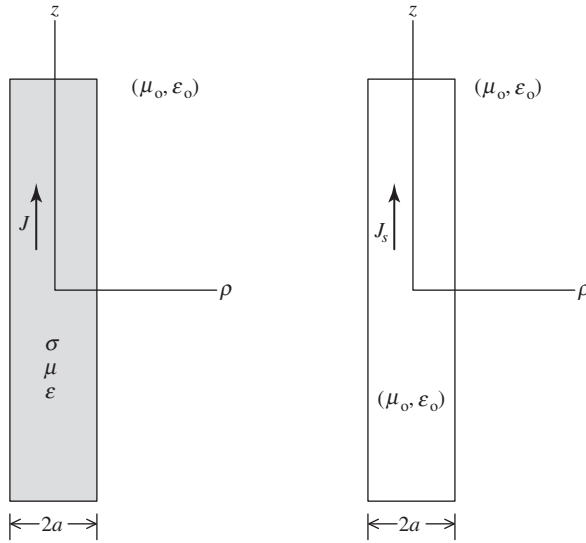
(*a*) Highly conducting wire with current density *J*.

(*b*) Surface equivalence model with equivalent surface current density J_s in free space.

Figure 14-3 Highly conducting thin wire along *z*-axis.

14.3   POCKLINGTON'S INTEGRAL EQUATION

One of the common integral equation that arises in the treatment of wire antennas is that derived by Pocklington in 1897. It enabled him to show that the current distribution on thin wires is approximately sinusoidal and propagates with nearly the speed of light. To derive this equation, consider the wire of conductivity σ in Fig. 14-3a surrounded by free space (μ_o, ε_o). Assume the conductivity of the wire is high (e.g., copper) such that the current is largely confined to the surface of the wire. The equivalence model for the wire becomes that in Fig. 14-3b (see Prob. 14.3-1), where current on the material wire is replaced by the equivalent surface current in free space (i.e., the wire material is removed). This step is necessary so that the vector potential, which employs the free-space Green's function, can be used.

When the wire radius is much less than the wavelength, we may assume only *z*-directed currents are present. From the Lorentz gauge condition in (2-44),

$$\frac{\partial A_z}{\partial z} = -j\omega\varepsilon_o\mu_o\Phi \tag{14-2}$$

where Φ is the scalar potential and A_z is the *z*-component of the magnetic vector potential. If we use (2-39), the vector electric field arising from potentials is

$$\mathbf{E} = -j\omega\mathbf{A} - \nabla\Phi \tag{14-3}$$

which for the situation in Fig. 14-3 reduces to the scalar equation

$$E_z = -j\omega A_z - \frac{\partial\Phi}{\partial z} \tag{14-4}$$

Taking the derivative of (14-2) and substituting into (14-4) gives

$$E_z = \frac{1}{j\omega\mu_o\varepsilon_o}\left(\frac{\partial^2 A_z}{\partial z^2} + \beta^2 A_z\right) \tag{14-5}$$

If we consider a z-directed volume current element $J\ dv'$,

$$dE_z = \frac{1}{j\omega\varepsilon_o}\left[\frac{\partial^2 \psi(z, z')}{\partial z^2} + \beta^2 \psi(z, z')\right]J\ dv' \tag{14-6}$$

where $\psi(z, z')$ is the free-space Green's function given in (2-59) as

$$\psi(z, z') = \frac{e^{-j\beta R}}{4\pi R} \tag{14-7}$$

and R is the distance between the observation point (x, y, z) and the source point (x', y', z') or

$$R = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2} \tag{14-8}$$

The total contribution to the electric field is the integral over the wire volume:

$$E_z = \frac{1}{j\omega\varepsilon_o}\iiint\left[\frac{\partial^2 \psi(z, z')}{\partial z^2} + \beta^2 \psi(z, z')\right]J\ dv' \tag{14-9}$$

We only need consider a volume distribution of current density if the wire is not of sufficiently high conductivity. If we assume the conductivity to be infinite, then the current is confined to the wire surface and (14-9) reduces to

$$E_z = \frac{1}{j\omega\varepsilon_o}\oint_c\int_{-L/2}^{L/2}\left[\frac{\partial^2 \psi(z, z')}{\partial z^2} + \beta^2 \psi(z, z')\right]J_s\ dz'\ d\phi' \tag{14-10}$$

where c is the cross-sectional curve of the wire surface as shown in Fig. 14-4a. For wires of good conducting material, the assumption of a surface current is approximately true and leads to no complications. If one observes the surface current distribution from a point on the wire axis as in Fig. 14-4b, then

$$R = \sqrt{(z - z')^2 + a^2} \tag{14-11}$$

For $a \ll \lambda$, the current distribution is nearly uniform with respect to ϕ', and (14-10) reduces to a line integral of (total) current. Thus,

$$E_z = \frac{1}{j\omega\varepsilon_o}\int_{-L/2}^{L/2}\left[\frac{\partial^2 \psi(z, z')}{\partial z^2} + \beta^2 \psi(z, z')\right]I(z')\ dz' \tag{14-12}$$

Note that the equivalent filamentary line source $I(z')$ is located a radial distance a from the observation point as in Figs. 14-4c and 14-4d and that we have not assumed the wire to be infinitely thin as was the case for dipoles studied in Chaps. 2 and 6.

In accordance with the surface equivalence principle of Sec. 9.1, we can denote the quantity E_z in (14-12) as the scattered field E_z^s. That is, E_z^s is the field radiated in free space by the equivalent current $I(z')$. The other field present is the incident or impressed field E_z^i. At the surface of the perfectly conducting wire and also interior to the wire,
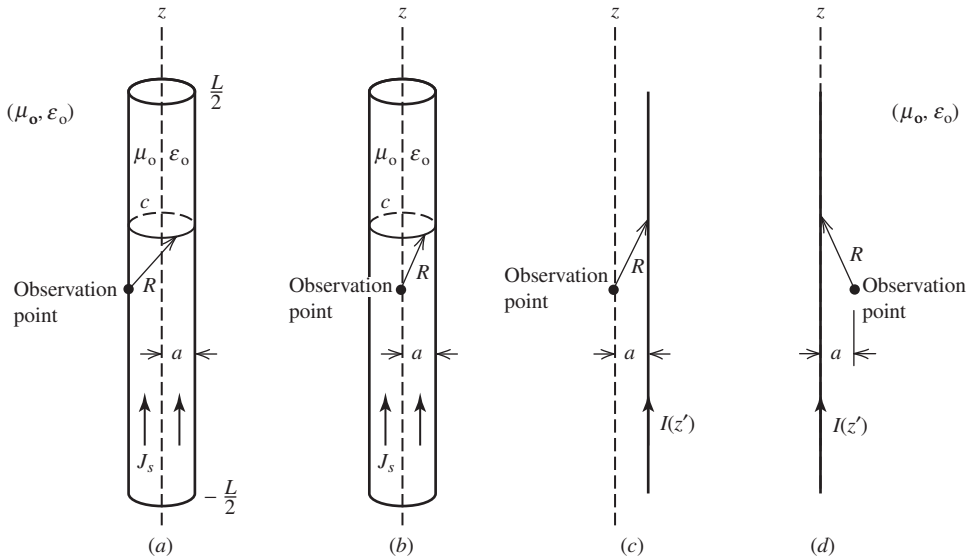
Figure 14-4 Simulation models for a thin wire. (*a*) Wire with equivalent surface current density J_s and observation point on the surface. (*b*) Wire with surface current density J_s and observation point on the wire axis. (*c*) Equivalent filamentary line source for the situation in (*b*). (*d*) Alternate representation of (*c*).

the sum of the tangential components of the scattered field and the incident field must be zero. Hence, $-E_z^s = +E_z^i$, and using (14-12) we write

$$\frac{-1}{j\omega\varepsilon_o} \int_{-L/2}^{L/2} I(z') \left[\frac{\partial^2 \psi(z, z')}{\partial z^2} + \beta^2 \psi(z, z') \right] dz' = E_z^i(z) \tag{14-13}$$

which is the type of integral equation derived by Pocklington and is of the general form used in (14-1).

Eq. (14-13) is an integral equation of the first kind because the unknown $I(z')$ appears only inside the integral. It is known as an integral equation because a boundary condition is incorporated therein. This is in contrast to (14-12) that is merely an expression for the so-called scattered field, which we can think of as that field radiated by a current independent of how the current was established (e.g., an impressed source on an antenna or incident plane wave).

Before we leave this section, it is worthwhile to summarize the important approximations that were used based on the assumption that $a \ll \lambda$.

1. Circumferential current on the wire is negligible.
2. Enforcement of the boundary condition on the surface of the wire (Fig. 14-4*a*) was performed on the axis of the wire (Fig. 14-4*b*), and the surface current then "collapsed" into a filament (Fig. 14-4*c*). By using reciprocity, the current filament was placed on the axis of the wire and the observation point placed a distance "*a*" away from the filament (Fig. 14-4*d*).
3. The distance R given by (14-11) leads to the widely used thin wire kernel or reduced kernel. R can never be zero; hence, the kernel is never singular. However, it is nearly singular and care must sometimes be taken during integrations when $R \approx a$.

In the following section, we illustrate how an integral equation such as (14-13) is solved numerically and point out how the procedure is analogous to Kirchhoff's network equation as noted by [H.3: Schelkunoff] many years ago.

14.4 INTEGRAL EQUATIONS AND KIRCHHOFF'S NETWORK EQUATIONS

One purpose of this section is to show the resemblance between integral equations of the type given in (14-13) and Kirchhoff's network equations:

$$\sum_{n=1}^{N} Z_{mn} I_n = V_m, \quad m = 1, 2, 3, \ldots N \tag{14-14}$$

Thus, we will solve the integral equation numerically by writing N equations in N unknowns just as we would do if we were solving an N mesh or N node circuit problem.

For convenience, let us write (14-13) in the form

$$-\int_{-L/2}^{L/2} I(z')K(z, z')\,dz' = E_z^i(z) \tag{14-15}$$

The first step in solving (14-15) is to approximate the unknown current by a series of known *expansion functions* F_n such that

$$I(z') = \sum_{n=1}^{N} I_n F_n(z') \tag{14-16}$$

where the I_n's are complex expansion coefficients and are unknown. To keep the discussion as simple as possible, we assume the expansion functions are a set of orthogonal pulse functions given by

$$F_n(z') = \begin{cases} 1 & \text{for } z' \text{ in } \Delta z'_n \\ 0 & \text{otherwise} \end{cases} \tag{14-17}$$

The expansion in terms of pulse functions is a "stairstep" approximation to the current distribution on the wire, where the wire is divided into N segments of length $\Delta z'_n$. See Fig. 14-5.

Substituting (14-16) into (14-15) gives

$$-\int_{-L/2}^{L/2} \sum_{n=1}^{N} I_n F_n(z')K(z_m, z')\ dz' \approx E_z^i(z_m) \tag{14-18}$$

where the subscript m on z_m indicates that the integral equation is being enforced at segment m. Note that the left side is only approximately equal to the right side because we have replaced the actual current distribution with approximate distribution. Using (14-17) in (14-18) enables us to write

Figure 14-5 "Staircase" approximation to an actual current distribution.

$$-\sum_{n=1}^{N} I_n \int_{\Delta z'_n} K(z_m, z') \ dz' \approx E_z^i(z_m) \tag{14-19}$$

For convenience, we let

$$f(z_m, z'_n) = -\int_{\Delta z'_n} K(z_m, z') \ dz' \tag{14-20}$$

Then (14-16) and (14-17) in (14-15) yield

$$-\int_{-L/2}^{L/2} I(z')K(z_m, z')dz' \approx I_1 f(z_m, z'_1) + I_2 f(z_m, z'_2) + \cdots + I_n f(z_m, z'_n)$$
$$+ \cdots + I_N f(z_m, z'_N) \approx E_z^i(z_m) \tag{14-21}$$

as illustrated in Fig. 14-5. A physical interpretation of this equation is as follows. The wire has been divided up into N segments, each of length $\Delta z'_n = \Delta z'$, with the current being an unknown constant over each segment. At the center of the mth segment, the sum of the scattered fields from all N segments is set equal to the incident field at the point z_m. The incident field is a known field arising from either a source located on the wire (transmitting case) or from a source located at a large distance (receiving case or radar scattering case). As we might surmise, if a more accurate representation of $I(z')$ is required, then shorter segments (and a larger N) can be used. More will be said about this later.

Eq. (14-21) leads to

$$\sum_{n=1}^{N} Z_{mn} I_n = V_m \tag{14-22}$$

where in this example situation

$$Z_{mn} = f(z_m, z'_n) \tag{14-23}$$

and

$$V_m = E_z^i(z_m) \tag{14-24}$$

Note that we have achieved our goal of reducing the electromagnetic problem to (14-22), which is identical to the network formulation of (14-14). It should be mentioned, however, that in network problems Z_{mn} is known at the start, whereas in electromagnetic problems it is necessary to calculate Z_{mn} as we have shown in this elementary example.

So far, we have only generated one equation in N unknowns. We need $N-1$ additional independent equations to solve for the N unknowns. To obtain these additional equations, we choose a different z_m for each equation. That is, we enforce the integral equation at N points on the axis of the wire. The process of doing this is called *point-matching*. It is a special case of the more general method of moments.

Point-matching at N points results in the following system of equations:

$$
\begin{aligned}
I_1 f(z_1, z'_1) + I_2 f(z_1, z'_2) + \cdots + I_N f(z_1, z'_N) &= E_z^i(z_1) \\
I_1 f(z_2, z'_1) + I_2 f(z_2, z'_2) + \cdots + I_N f(z_2, z'_N) &= E_z^i(z_2) \\
\vdots \qquad\qquad \vdots \qquad\qquad\qquad \vdots \qquad\qquad \vdots & \\
I_1 f(z_N, z'_1) + I_2 f(z_N, z'_2) + \cdots + I_N f(z_N, z'_N) &= E_z^i(z_N)
\end{aligned} \tag{14-25}
$$

which can be written in the matrix form as

$$
\begin{bmatrix}
f(z_1, z_1') & f(z_1, z_2') & \cdots & f(z_1, z_N') \\
f(z_2, z_1') & f(z_2, z_2') & \cdots & f(z_2, z_N') \\
\vdots & & \vdots & \vdots \\
f(z_N, z_1') & f(z_N, z_2') & \cdots & f(z_N, z_N')
\end{bmatrix}
\begin{bmatrix}
I_1 \\ I_2 \\ \vdots \\ I_N
\end{bmatrix}
=
\begin{bmatrix}
E_z^i(z_1) \\ E_z^i(z_2) \\ \vdots \\ E_z^i(z_N)
\end{bmatrix}
\tag{14-26}
$$

or in the compact notation as

$$
[Z_{mn}][I_n] = [V_m] \tag{14-27}
$$

where Z_{mn} and V_m are given by (14-23) and (14-24), respectively. We refer to the first index (m) as the match point index because it is associated with the observation point at which the mth equation is valid. The second index is the source point index since it is associated with the field from the nth segment or nth source. Because of the analogy to the network equations, the matrices $[Z_{mn}]$, $[I_n]$, and $[V_m]$ are referred to as *generalized* impedance, current, and voltage matrices, respectively. But this is only an analogy and thus the units of $[Z_{mn}]$, $[I_n]$, and $[V_m]$ need not necessarily be ohms, amperes, and volts, respectively. The analogy is not restricted to collinear segments as in the example treated here, but applies to arbitrary configurations of wires as well.

We can write the solution to (14-27) symbolically as

$$
[I_n] = [Z_{mn}]^{-1}[V_m] \tag{14-28}
$$

In practice, the explicit inverse $[Z_{mn}]^{-1}$ is not usually evaluated, but instead the system of equations is solved by one of several fairly standard matrix algorithms. Once $[I_n]$ is found, the approximate current distribution of (14-16) is known in discrete form and we can then proceed to determine impedance and radiation patterns or the radar cross section.

To summarize this section, we have obtained an elementary numerical solution to an integral equation of the form in (14-15). This was done by successively enforcing the integral equation at N different points, as illustrated in (14-25). For mathematical convenience and simplicity, the locations of the points were chosen to be at the center of the N equal-length segments into which the wire was divided. Strictly speaking, in order for the equations in (14-25) to be exact equalities, N must approach infinity. However, in practice we can obtain accurate solutions for the current distribution by allowing N to be sufficiently large, as will be demonstrated in the next section.

14.5 SOURCE MODELING

Three source models are commonly used in the MoM. For transmitting antennas, the delta gap source and frill source produce the required incident field. For a receiving antenna or scatterer, the incident field is usually a plane wave. We examine all three in this section.

No doubt, the most, used generator model in wire antenna theory is the *delta gap model*, shown in Fig. 14-6, which is often referred to as a *slice generator*. Although such
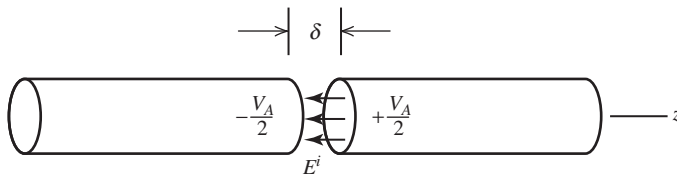
Figure 14-6 The delta gap source model with impressed field $E^i = V_A/\delta$.

sources do not exist in practice, they do permit surprisingly good calculations to be made. The source arises from the assumption that a voltage is placed across the gap, giving rise to an impressed electric field $E^i = V/\delta$ confined entirely to the gap (i.e., no fringing). With reference to Fig. 14-6, the voltage across the gap is determined by the line integral of the electric field across the gap. The result is $V_A = +E^i\delta$. The voltage V_A applied across the gap is not to be confused with the elements V_m in the generalized voltage matrix $[V_m]$. For a delta gap source model, $V_m = E^i = V_A/\delta$ when the point-matching technique is used.

A second generator model, which has practical significance, is the so-called *frill generator*. Consider Fig. 14-7a that shows a coaxial line feeding a monopole on a ground plane. Assuming a purely dominant mode distribution (TEM) in the coaxial aperture and image theory, we can replace the ground plane and the coaxial aperture with a frill of magnetic current as shown in Fig. 14-7b. Since the assumed form of the electric field in the aperture is

$$E_{\rho'}(\rho') = \frac{1}{2\rho' \ln(b/a)} \tag{14-29}$$

the corresponding magnetic current distribution from $\mathbf{M} = 2\hat{\mathbf{n}} \times \mathbf{E}$ is

$$M_{\phi'} = 2E_{\rho'} = \frac{-1}{\rho' \ln(b/a)} \tag{14-30}$$

from which it can be shown that the electric field on the axis of the monopole is [4; H.10.2: Mittra, Ed., Chap. 2]

$$E_z^i(0, z) = \frac{1}{2\ln(b/a)} \left(\frac{e^{-j\beta R_1}}{R_1} - \frac{e^{-j\beta R_2}}{R_2} \right) \tag{14-31}$$

(a) Coaxial line feeding a monopole
through a ground plane.

(b) Mathematical model of
Fig. 14-7a.
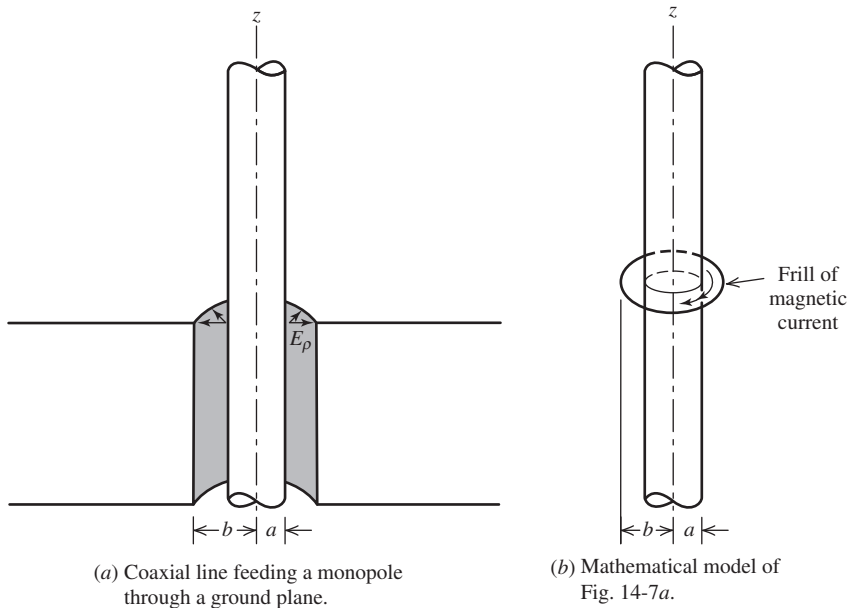
Figure 14-7 Magnetic frill source.

where

$$R_1 = \sqrt{z^2 + a^2} \tag{14-32a}$$

$$R_2 = \sqrt{z^2 + b^2} \tag{14-32b}$$

if the frill center is at the coordinate origin.

The third source to consider is that of an incident plane wave. To obtain the elements in the generalized voltage matrix in this case, we need the tangential component of the incident field at the match points along the axis of the wire dipole. For our z-directed dipole of Fig. 14-4, this would be

$$E^i_{\tan} = \hat{\mathbf{z}} \cdot \mathbf{E} e^{j\beta z \cos\theta} \tag{14-33}$$

For example, for a unit amplitude plane wave normally incident on the z-directed dipole, the elements of the generalized voltage matrix are all $(1 + j0)$.

EXAMPLE 14-1   *Point-Matching on a Short Dipole*

The purpose of this example is to illustrate the application of (14-26). An objective is to use MoM to determine the input impedance Z_A of a short dipole with a length of 0.1λ and a radius of 0.005λ. For convenience of illustration, choose $N = 5$. With reference to Fig. 14-8, the elements of $[Z_{mn}]$ are calculated to be

$$[Z_{mn}] = 10^2 \begin{bmatrix} 679.5 \angle{-89.99°} & 292.6 \angle{89.97°} & 33.03 \angle{89.73°} & 9.75 \angle{89.09°} & 4.24 \angle{87.92°} \\ 292.6 \angle{89.97°} & 679.5 \angle{-89.99°} & 292.6 \angle{89.97°} & 33.03 \angle{89.73°} & 9.75 \angle{89.09°} \\ 33.03 \angle{89.73°} & 292.6 \angle{89.97°} & 679.5 \angle{-89.99°} & 292.6 \angle{89.97°} & 33.03 \angle{89.73°} \\ 9.75 \angle{89.09°} & 33.03 \angle{89.73°} & 292.6 \angle{89.97°} & 679.5 \angle{-89.99°} & 292.6 \angle{89.97°} \\ 4.24 \angle{87.92°} & 9.75 \angle{89.09°} & 33.03 \angle{89.73°} & 292.6 \angle{89.97°} & 679.5 \angle{-89.99°} \end{bmatrix}$$

For a 1-V excitation at the center of the short dipole (i.e., segment 3), the following voltage matrix $[V_m]$ is obtained using the frill source discussed in Sec. 14.5 with $b/a = 2.3$, and upon solving (14-28), the following current matrix $[I_n]$ is also obtained:

$$[V_m] = \begin{bmatrix} 0.484 \angle{-0.31°} \\ 3.128 \angle{-0.04°} \\ 67.938 \angle{-0.002°} \\ 3.128 \angle{-0.04°} \\ 0.484 \angle{-0.31°} \end{bmatrix}, \quad [I_n] = 10^{-3} \begin{bmatrix} 0.78 \angle{89.54°} \\ 1.48 \angle{89.64°} \\ 2.35 \angle{89.75°} \\ 1.48 \angle{89.64°} \\ 0.78 \angle{89.54°} \end{bmatrix}$$

On the other hand, if a 1-V delta gap excitation is used, $V_3 = 1/\Delta z = 1/0.02 = 50 \angle{0°}$, and the resulting voltage and current matrices are

$$[V_m] = \begin{bmatrix} 0 \angle{0°} \\ 0 \angle{0°} \\ 50.0 \angle{0°} \\ 0 \angle{0°} \\ 0 \angle{0°} \end{bmatrix}, \quad [I_n] = 10^{-3} \begin{bmatrix} 0.52 \angle{89.54°} \\ 0.98 \angle{89.64°} \\ 1.63 \angle{89.76°} \\ 0.98 \angle{89.64°} \\ 0.52 \angle{89.54°} \end{bmatrix}$$
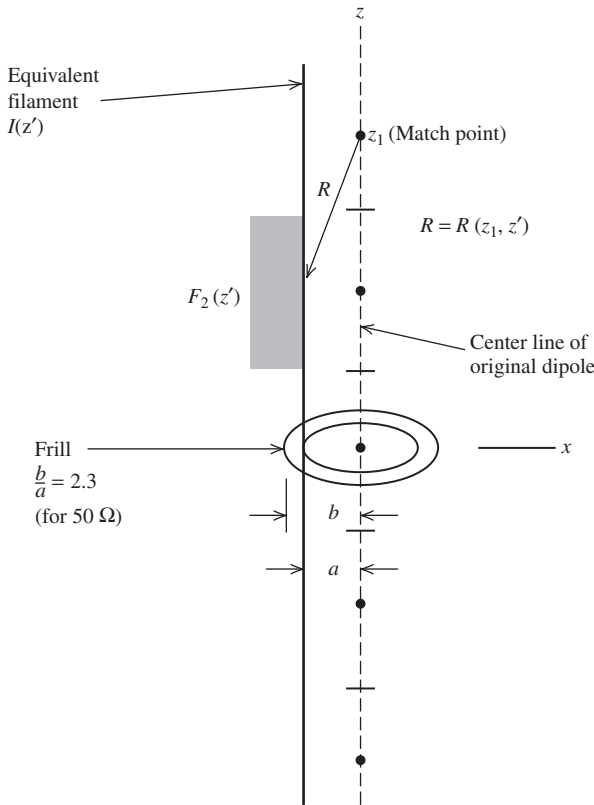
$$Z_{12} = -\int\limits_{\Delta z_2'} F_2 \, (z') \, K \, (z_1, z') \, dz'$$

Figure 14-8 Calculation of Z_{12} for the short dipole in Example 14-1.

$$K(z_m, z') = \frac{1}{4 \, \pi \, j \omega \varepsilon_o} \, \frac{e^{-j\beta R}}{R^5} \, [(1 + j\beta R) \, (2R^2 - 3a^2) + \beta^2 a^2 R^2]$$

Note that the current distribution decreases from the center toward the ends as expected. The input impedance for the frill may be found from $Z_A = V_A/I_3 = 1.0/(2.35 \times 10^{-3} \, \angle 89.75°) = 425.53 \quad \angle -89.75° = 1.857 - j425.53 \, \Omega$. Comparing with the thinner dipole in Figs. 6-5 and 6-6, we see that the input impedance of a $0.1 \, \lambda$ long dipole also has a very small real part and a large negative reactive part. Further, the real part of $1.857 \, \Omega$ compares fairly well with the approximate formula $20\pi^2(L/\lambda)^2 = 1.974 \, \Omega$ even though only five segments were used here.

In the preceding example, a short dipole was represented by only five segments for the purposes of numerical illustration. To illustrate the behavior of a pulse point-matching solution to Pocklington's equation for a resonant size dipole as the number of segments is varied, consider Fig. 14-9. Fig. 14-9 shows the input impedance of a dipole of length $L = 0.47 \, \lambda$ as the number of segments varies from 10 to 120. Both the frill source and delta gap are used. For both sources, it is apparent that for N sufficiently large, the solution has converged to a final or reasonably stable result. In many instances, N cannot be made arbitrarily large without encountering a numerically unstable result. For example, the reactance of the delta gap source exhibits divergence for large N in Fig. 14-9b, but this should not be viewed as a general behavior of the delta gap since it

(a) Input resistance.

(b) Input reactance.
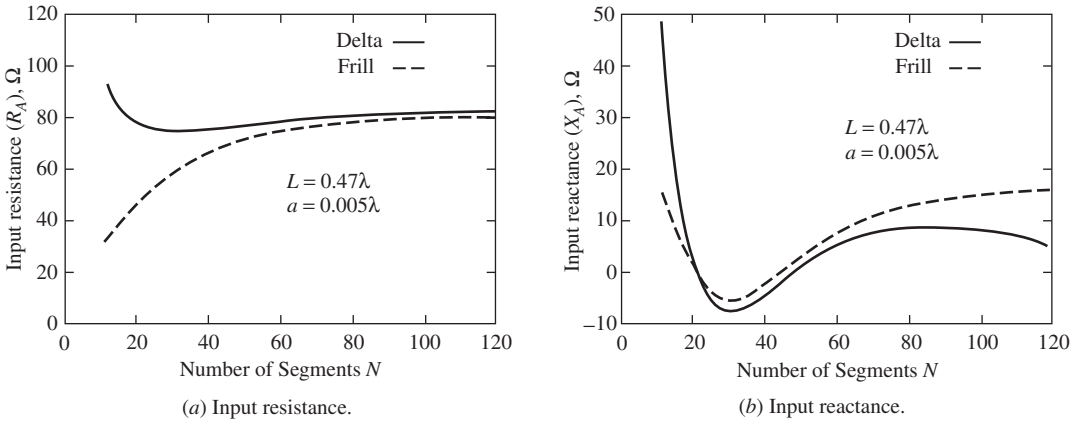
Figure 14-9 Curves showing convergence of input impedance as the number of pulse functions is increased for two different sources: the delta gap and the frill.
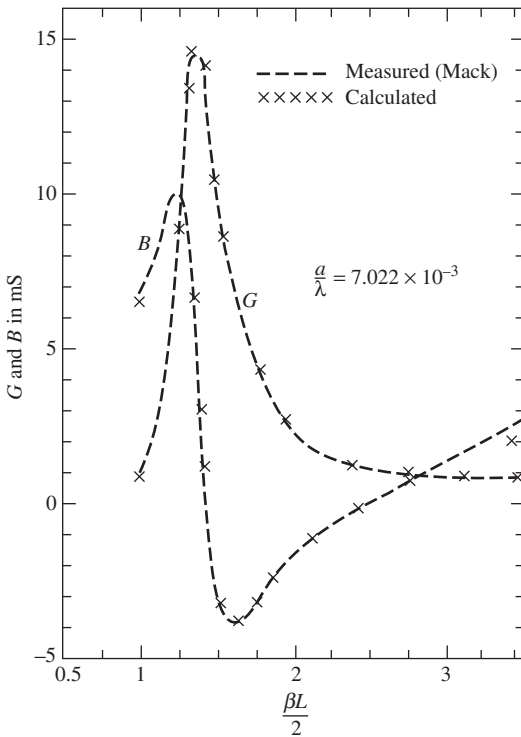
Figure 14-10 Comparison of measured dipole admittance with data calculated using pulse functions ($N = 100$) and a frill source.

does not necessarily occur for other MoM formulations (e.g., Fig. 14-13b) or for (14-13) if the order of differentiation and integration is interchanged.

A curve, such as those in Fig. 14-9, is well worth the effort since it clearly shows the convergence behavior of a solution. A comparison with experimental data is shown in Fig. 14-10.

To summarize this section, an elementary numerical solution to an integral equation of the form given in (14-15) was obtained by successively enforcing the integral equation at N different points as illustrated in (14-25). For mathematical convenience and simplicity, the segments were of equal length, and the match points were located at the center of each segment. Strictly speaking, in order for the equations (14-25) to be exact equalities, N must approach infinity. However, in practice, accurate solutions can be obtained by allowing, N to be sufficiently large. In this regard, a convergence curve, such as that in Fig. 14-9, can be invaluable.

14.6   WEIGHTED RESIDUALS AND THE METHOD OF MOMENTS

Our objective in this section is to derive a moment method procedure more general than the point-matching method of the previous section. This is accomplished by using an approach known as the method of weighted residuals [5].

Consider the straight wire example of the previous section. Define the *residual R* to be the sum of the tangential components of the scattered and incident fields:

$$R = E^s_{\text{tan}} + E^i_{\text{tan}} \tag{14-34}$$

Clearly, we wish the residual to be zero and thereby satisfy the boundary condition. In our example of Sec. 14.4, with pulse expansion functions the residual is found from (14-19) to be

$$R(z) = -\sum_{n=1}^{N} I_n f(z, \ z'_n) + E^i_z(z) \tag{14-35}$$

Stated in terms of the electric field boundary condition, the residual is the sum of the tangential components of the scattered and incident fields at the wire surface. Eq. (14-35), when evaluated for $z = z_m$, gives the residual at the mth match point, where, of course, the residual must be zero since the solution for the I_n's was obtained subject to the electric field boundary condition at the N matching points. However, at points other than the match points, the total tangential electric field will not generally be zero as Fig. 14-11 indicates. Therefore, the residual for $z \neq z_m$, $m = 1, 2, 3 \ldots N$, will not be zero, either. Physically, we can view the point-matching procedure as a relaxation of the boundary condition such that it is only satisfied at specified points. In between those points, one can only hope that the boundary condition is not so badly violated that the solution is rendered useless. Thus, it is not surprising that as N is increased (within limits) the solution tends to improve as we saw in Fig. 14-9.

In the method of weighted residuals the I_n's are found such that the residual is forced to zero in a weighted average sense. So, in the wire problem of Fig. 14-3 the weighted integrals of the residual are set to zero as follows.

$$\int W_m(z) R(z) \ dz = 0, \quad m = 1, 2, 3, \ldots, N \tag{14-36}$$
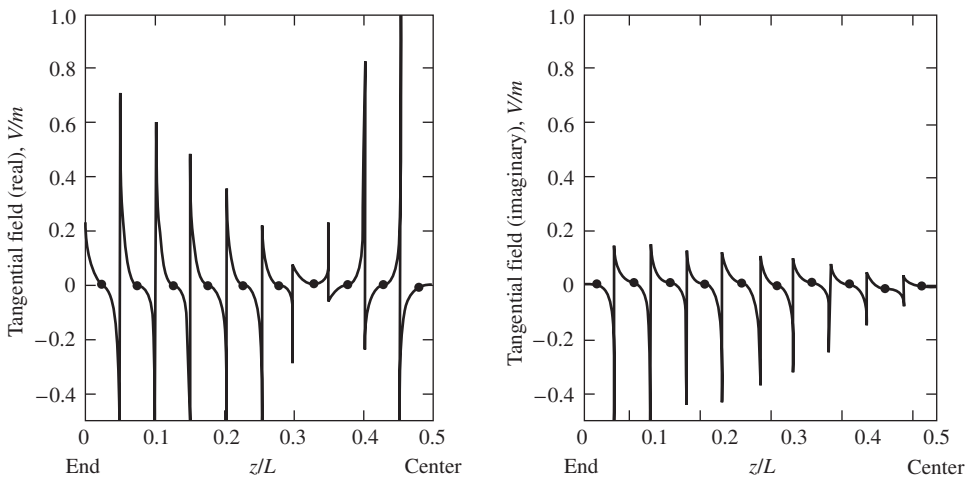
Figure 14-11 Normalized tangential electric field along one-half of a center-fed dipole with pulse expansion functions and delta weighting functions (courtesy of E. K. Miller). Dots indicate match point locations.

where $W_m(z)$ is called a weighting or testing function. Substituting (14-35) into (14-36) gives

$$-\int_{-L/2}^{L/2} W_m(z) \sum_{n=1}^{N} I_n f(z, z_n') \; dz + \int_{-L/2}^{L/2} W_m(z) E_z^i(z) \; dz = 0, \quad m = 1, 2, 3, \ldots, N$$

$$(14\text{-}37)$$

If the weighting functions are the Dirac delta functions

$$W_m(z) = \delta(z - z_m) \qquad (14\text{-}38a)$$

then (14-37) reduces to (14-21). If the weighting functions are the pulse functions

$$W_m(z) = \begin{cases} 1 & \text{for } z \text{ in } \Delta z_m \\ 0 & \text{otherwise} \end{cases} \qquad (14\text{-}38b)$$

then (14-37) becomes

$$-\sum_{n=1}^{N} I_n \int_{\Delta z_m} f(z, z_n') \; dz + \int_{\Delta z_m} E_z^i(z) \; dz = 0, \quad m = 1, 2, 3, \ldots, N \qquad (14\text{-}39)$$

It follows that

$$Z_{mn} = \int_{\Delta z_m} f(z, z_n') \; dz \qquad (14\text{-}40a)$$

and

$$V_m = \int_{\Delta z_m} E_z^i(z) \; dz \qquad (14\text{-}40b)$$

The current obtained from solving (14-40) will not necessarily be such that the sum of the scattered and incident fields (i.e., the residual) is zero everywhere along the surface of the wire, but the average over the wire will tend to be zero, presumably giving a more accurate current distribution for a given N than when the weight functions are delta functions. Actually, this may or may not be the case depending on the particular choice of expansion functions for the current and weighting (or testing) functions.

The question of how one chooses the expansion functions and weighting functions is certainly a valid one. It is, however, a question without a concise answer. But, as a rule of thumb, it is desirable to choose expansion functions that closely resemble the anticipated form of the current on the wire and to use the same functions for the weighting functions as used for the expansion functions. There are exceptions to these rules including the pulse point-matching solution of Sec. 14.4. When the expansion function and the weight function are the same, the procedure is often referred to as *Galerkin's method*, which is closely related to variational methods.

EXAMPLE 14-2   *Galerkin's Method on a Short Dipole*

The purpose of this example is to repeat Example 14-1 using pulse functions for weight functions instead of delta functions. With reference to Fig. 14-12, the impedance matrix $[Z_{mn}]$ for this pulse-pulse Galerkin solution based on (14-40) is calculated to be

$$Z_{12} = \frac{-1}{4\pi j\omega\varepsilon_o} \int\limits_{\Delta z_1} F_1(z) \int\limits_{\Delta z'_2} F_2(z') \, K(z, z') \, dz' \, dz$$
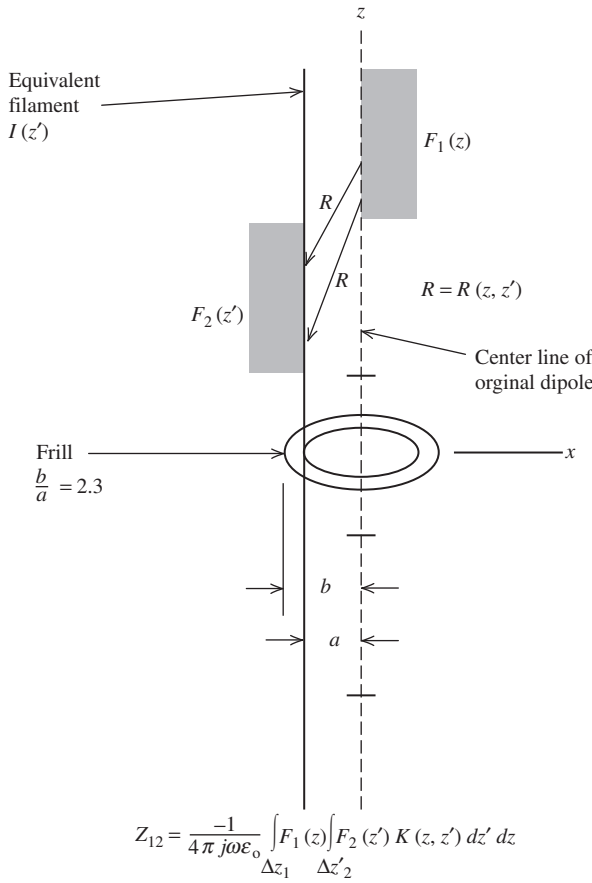
Figure 14-12 Calculation of Z_{12} for the short dipole in Example 14-2.

$$[Z_{mn}] = 10^2 \begin{bmatrix} 14.4 \angle -89.99° & 6.14 \angle 89.97° & 0.759 \angle 89.76° & 0.206 \angle 89.14° & 0.087 \angle 87.98° \\ 6.14 \angle 89.97° & 14.4 \angle -89.99° & 6.14 \angle 89.97° & 0.759 \angle 89.76° & 0.206 \angle 89.14° \\ 0.759 \angle 89.76° & 6.14 \angle 89.97° & 14.4 \angle -89.99° & 6.14 \angle 89.97° & 0.759 \angle 89.76° \\ 0.206 \angle 89.14° & 0.759 \angle 89.76° & 6.14 \angle 89.97° & 14.4 \angle -89.99° & 6.14 \angle 89.97° \\ 0.087 \angle 87.98° & 0.206 \angle 89.14° & 0.759 \angle 89.76° & 6.14 \angle 89.97° & 14.4 \angle -89.99° \end{bmatrix}$$

The voltage matrix $[V_m]$, using a 1-V frill source, and solution matrix $[I_n]$ are, respectively,

$$[V_m] = \begin{bmatrix} 0.011 \angle -0.280° \\ 0.089 \angle -0.034° \\ 0.791 \angle -0.003° \\ 0.089 \angle -0.034° \\ 0.011 \angle -0.280° \end{bmatrix}, \quad [I_n] = 10^{-3} \begin{bmatrix} 0.49 \angle 89.57° \\ 0.91 \angle 89.66° \\ 1.38 \angle 89.75° \\ 0.91 \angle 89.66° \\ 0.49 \angle 89.57° \end{bmatrix}$$

On the other hand, if a 1-V delta gap excitation is used,

$$[V_m] = \begin{bmatrix} 0 & \angle 0° \\ 0 & \angle 0° \\ 1 & \angle 0° \\ 0 & \angle 0° \\ 0 & \angle 0° \end{bmatrix}, \quad [I_n] = 10^{-3} \begin{bmatrix} 0.49 & \angle 89.57° \\ 0.91 & \angle 89.67° \\ 1.52 & \angle 89.78° \\ 0.91 & \angle 89.67° \\ 0.49 & \angle 89.57° \end{bmatrix}$$

Note that all these five matrices are different from those in Example 14-1. Of course, we would expect $[Z_{mn}]$ and $[V_m]$ to be different because they are computed by a different process. The reason $[I_n]$ is different is solely attributable to the fact that N is only 5. As in Example 14-1, a larger value of N is required in order to obtain a converged result. The input impedance based on the above current using the frill source is $Z_A = 3.162 - j724 \ \Omega$, whereas for the delta gap source $Z_A = 2.526 - j658 \ \Omega$. The impedance based on $N = 25$ is $Z_A = 2.35 - j556 \ \Omega$.

Fig. 14-13 shows the convergence of the input impedance for a dipole of length 0.47λ using pulse expansion functions and pulse weighting functions in Pocklington's equation. Comparing with Fig. 14-9, we see that the convergence is more rapid with pulse weights than delta weights, and the pulse-pulse formulation is less sensitive to the kind of source (i.e., frill or delta gap) than the pulse-delta formulation. In many formulations, as in the one here, the averaging process provided by nondelta weights tends to improve the rate of convergence and stability of the solution.

Next, we relate the quantities in the weighted residual integral to Kirchhoff's network equations, just as was done in Sec. 14.4. In doing so, let us generalize somewhat and consider a wire as shown in Fig. 14-14. In this case, the residual may be written as

$$\mathbf{R}(\ell) = \mathbf{E}_{\text{tan}}^s(\ell) + \mathbf{E}_{\text{tan}}^i(\ell) = \sum_{n=1}^{N} I_n \mathbf{E}_n^s(\ell) + \mathbf{E}_{\text{tan}}^i(\ell) \tag{14-41}$$

and the weighted residual as

$$\int_{\substack{\text{along wire}}} \mathbf{W}_m(\ell) \cdot \mathbf{R}(\ell) \, d\ell = 0 \tag{14-42}$$

(a) Input resistance.

(b) Input reactance.
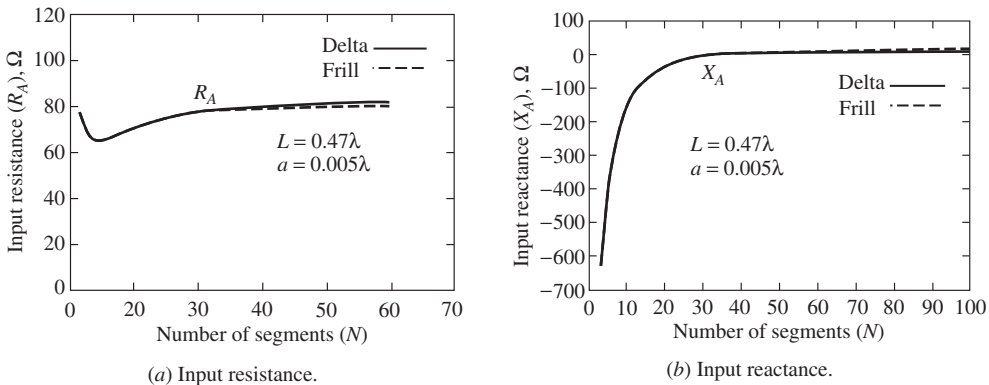
Figure 14-13 Curves showing convergence of input impedance as the number of segments (pulse Galerkin) is increased for two different sources: the delta gap and the frill.
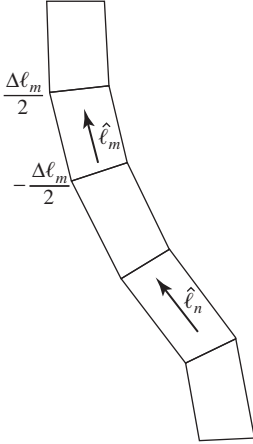
Figure 14-14 Segmented curved wire.

so that

$$\sum_{n=1}^{N} I_n \int_{-\Delta\ell_m/2}^{\Delta\ell_m/2} \mathbf{W}_m(\ell) \cdot \mathbf{E}_n^s(\ell) \ d\ell + \int_{-\Delta\ell_m/2}^{\Delta\ell_m/2} \mathbf{W}_m(\ell) \cdot \mathbf{E}^i(\ell) \ d\ell = 0, \quad m = 1, 2, 3, \ldots, N$$

$$(14\text{-}43)$$

This equation can be viewed in the form of (14-14) and if the scattered field from the nth expansion function of the current is denoted as $\mathbf{E}_n^s(\ell)$, then the general mnth element in the generalized impedance matrix is

$$Z_{\mathrm{mn}} = -\int_{-\ell_m/2}^{\ell_m/2} \mathbf{W}_m(\ell) \cdot \mathbf{E}_n^s(\ell) \ d\ell \qquad (14\text{-}44)$$

and for the mth generalized voltage matrix element,

$$V_m = \int_{\ell_m/2}^{\ell_m/2} \mathbf{W}_m(\ell) \cdot \mathbf{E}^i(\ell) \ d\ell \qquad (14\text{-}45)$$

where $\mathbf{W}_m(\ell)$ is the mth testing function taken to be located interior to the wire as suggested in Fig. 14-12. Strictly speaking, the test function should be located at the wire surface (see Fig. 14-4a), in which case (14-44) and (14-45) would be double integrals over the surface. In placing the testing function on the axis, we are in a sense modifying the electric field boundary condition for the sake of mathematical simplification. In doing this, experience has shown that we are restricted to wires for which the radius is less than about 0.01λ. This is sufficient for most wire antenna or scattering problems. For thicker wires, a more exact formulation is available [6].

The process of expanding the unknown current $I(\ell')$ in a series of expansion functions and then generating N equations in N unknowns using the weighted residual integral of (14-42) is more commonly referred to in the electromagnetics literature as the method of moments. MoM is, as we have seen in this section, equivalent to the method of weighted residuals. If the testing or weighting functions are delta functions, then the specific MoM procedure is known as point-matching. This is also known as *collocation*. This was the procedure used to obtain the system of equations in (14-25). If both the test

function and expansion function are the same, then the specific MoM procedure is known as Galerkin's method. A pulse-pulse Galerkin formulation was used in Example 14-2. There are functions other than the pulse function that have been shown to be useful. Some of these are discussed in Sec. 14.8. In the next section, we discuss two other approaches to the MoM: the reaction concept and the linear algebra formulation.

14.7   TWO ALTERNATIVE APPROACHES TO THE METHOD OF MOMENTS

In the previous sections of this chapter, MoM has been developed using an approach that makes takes advantage of concepts a student is likely to have previously experienced (e.g., Kirchhoff's network equations and the use of the "staircase approximation" to an integral in Sec. 14.4). Two other approaches to MoM are found in the literature. One has a physical interpretation (i.e., reaction) and the other is entirely mathematical (i.e., the linear algebra approach). This section will consider both of these approaches.

14.7.1   Reaction

In 1954, Rumsey introduced a physical observable (e.g., mass, length, charge, etc.) called *reaction* that permitted a general approach to boundary value problems in electromagnetic theory. [7] His approach resulted in the formulation of the reaction integral equation. Eq. (14-43) is really a special form of the reaction integral equation that applies to wire geometries. A rigorous derivation of the reaction integral equation can be derived using only principles of electromagnetic theory. The derivation is somewhat difficult to follow and so we will use inductive reasoning here, having derived (14-43) in the previous section by the relatively straightforward weighted residual approach.

Reaction is basically "a measure of the coupling" between one source and another. Thus, if we view the test function (weight function) as a *test source*, then the impedance matrix elements given by (14-44) may be taken as a calculation of the coupling between the mth test source and the scattered field from the nth expansion function or *actual source*. Similarly, the mth voltage matrix element in (14-45) can be interpreted as the coupling between the mth test source and the incident field. When referring to (14-45), for instance, we can say that we are "reacting" the mth test source with the incident field, or in the case of (14-44) that we are "reacting" the electric field from the nth actual source with the current on the mth test source.

We obtained (14-43) for a wire. The method of moments or the method of weighted residuals applies to geometries other than just wire geometries as indicated in Fig. 14-15a. Consider the equivalent situation in Fig. 14-15b. Let $(\mathbf{J}_m, \mathbf{M}_m)$ be the surface current densities of a test source and $(\mathbf{E}_m, \mathbf{H}_m)$ be the fields from the test source. The currents on the conducting body are both replaced by equivalent surface currents $(\mathbf{J}^s, \mathbf{M}^s)$ radiating the fields $(\mathbf{E}^s, \mathbf{H}^s)$ in free space. The generalization of (14-43) then becomes

$$\iint_S (\mathbf{J}_m \cdot \mathbf{E}^s - \mathbf{M}_m \cdot \mathbf{H}^s) \; ds + \iint_S (\mathbf{J}_m \cdot \mathbf{E}^i - \mathbf{M}_m \cdot \mathbf{H}^i) \; ds = 0, \quad m = 1, 2, 3, \ldots, N$$

(14-46)

The minus sign associated with $\mathbf{M}_m$ can be justified by referring to the reciprocity field theorems in Sec. 13.1. The physical interpretation of (14-46) is that we wish to have zero reaction (i.e., zero coupling) between the test source and the sum of the incident and scattered fields. Clearly, this is equivalent to the condition stated by (14-36). Nevertheless, the alternative physical interpretation offered by (14-46) and the reaction concept is a useful one and the student will find it used in the literature.
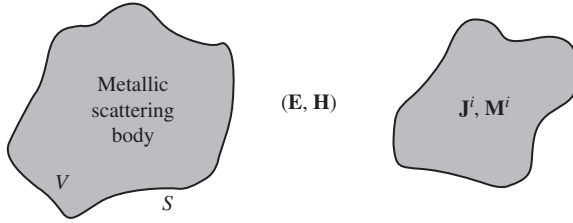
(*a*) Source current densities $\mathbf{J}^i$ and $\mathbf{M}^i$ acting in the presence of a
metallic scattering body bounded by surface *S* create
fields $(\mathbf{E}, \mathbf{H})$ exterior to *S*.
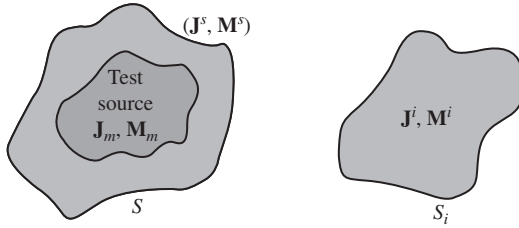
(*b*) Test source interior to surface *S* with equivalent currents
$(\mathbf{J}^s, \mathbf{M}^s)$ in free space.

Figure 14-15 Sources used in
reaction concept.

If we denote the fields from the *n*th expansion function of the actual current by
$(\mathbf{E}_n^s, \mathbf{H}_n^s)$, the sum of the *N* fields being $(\mathbf{E}^s, \mathbf{H}^s)$, then we may write for the general *mn*th
element in the generalized impedance matrix:

$$Z_{\mathrm{mn}} = -\iint\limits_S (\mathbf{J}_m \cdot \mathbf{E}_n^s - \mathbf{M}_m \cdot \mathbf{H}_n^s) \ ds \qquad (14\text{-}47)$$

Similarly, we may write the general mth element in the voltage matrix:

$$V_m = \iint\limits_S (\mathbf{J}_m \cdot \mathbf{E}^i - \mathbf{M}_m \cdot \mathbf{H}^i) \ ds \qquad (14\text{-}48)$$

The incident field $(\mathbf{E}^i, \mathbf{H}^i)$, which originates from the impressed currents $\mathbf{J}^i$ in
Fig. 14-15*a*, may be the field from a source located on *S* (antenna transmitting situation)
or be from a source located at a great distance from *S* (antenna receiving and radar
scattering situation). The general relationships in (14-47) and (14-48) will be useful later
for both wire and nonwire geometries.

14.7.2   Linear Algebra Formulation of MoM

Another way of approaching the general formulation of MoM is through the use of linear
algebra. This is an approach commonly found in the literature [3]. Consider a general
metallic body with a surface current density $\mathbf{J}$ on it. For simplicity, assume there are no
magnetic currents. The extension of what follows to the case where both $\mathbf{J}$ and $\mathbf{M}$ are
present can be deduced easily from the previous section on reaction.

The development of the linear algebra approach begins by requiring that the total
tangential electric field be zero everywhere on the surface of the body, or that

$$0 = \mathbf{E}_{\mathrm{tan}}^s + \mathbf{E}_{\mathrm{tan}}^i \qquad (14\text{-}49)$$

where $\mathbf{E}^s_{\text{tan}}$ is the scattered electric field radiated by the current density $\mathbf{J}$ and $\mathbf{E}^i_{\text{tan}}$ is the tangential component of the incident electric field due to a source located anywhere on or outside the body. We will drop the subscript *tan* since it will be understood that the tangential electric field boundary condition is being used.

Rewriting (14-49) in the form

$$-\mathbf{E}^s = \mathbf{E}^i \tag{14-50}$$

and defining the operator

$$L_{\text{op}}(\mathbf{J}) = -\mathbf{E}^s \tag{14-51}$$

we can use the concept of linear vector spaces and operators to write the operator equation

$$L_{\text{op}}(\mathbf{J}) = \mathbf{E}^i \tag{14-52}$$

Where L_{op} is an operator that must be determined for the problem of interest, $\mathbf{E}^i$ is a known excitation function or source, and $\mathbf{J}$ is the unknown response function to be determined. In the problems that are considered in this chapter, L_{op} is an integral operator operating on the current $\mathbf{J}$.

For a given problem, we must determine the domain of definition of the operator or, in other words, the space of functions on which it operates and also the range of the operator or the functions resulting from the operation. In reality, the operator performs a mapping from some subset containing $\mathbf{J}$ to one containing $\mathbf{E}^i$. If the solution is to be unique, this mapping must be one to one.

Next, expand the response (solution) function $\mathbf{J}$ in a series of basis functions $F_1, F_2, F_3, \ldots$ on a surface S and defined in the domain of L_{op}. That is,

$$\mathbf{J} = \sum_n I_n \mathbf{F}_n \tag{14-53}$$

where the coefficients I_n are, in general, complex. The I_n's are the unknown coefficients that are to be determined. Substituting (14-53) into (14-52) yields

$$L_{\text{op}}\left(\sum_n I_n \mathbf{F}_n\right) = \mathbf{E}^i \tag{14-54}$$

Using the linearity of L_{op}, we get

$$\sum_n I_n L_{\text{op}}(\mathbf{F}_n) = \mathbf{E}^i \tag{14-55}$$

The next step in the solution outlined above is to define a set of weighting functions $\mathbf{W}_1, \mathbf{W}_2, \ldots$ in the domain of L_{op} and then form the inner product:

$$\sum_n I_n \langle \mathbf{W}_m,\ L_{\text{op}}(\mathbf{F}_n)\rangle = \langle \mathbf{W}_m,\ \mathbf{E}^i\rangle \tag{14-56}$$

Note that if $\mathbf{W}_m$ is a delta function, (14-56) becomes the point-matching case, and if $\mathbf{W}_m = \mathbf{F}_m$ then (14-56) is a Galerkin formulation (e.g., Sec. 14.6). For the Galerkin formulation, write

$$\sum_n I_n \langle \mathbf{F}_m,\ L_{\text{op}}(\mathbf{F}_n)\rangle = \langle \mathbf{F}_m,\ \mathbf{E}^i\rangle \tag{14-57}$$

and the inner products appear as the reaction quantity mentioned earlier. Note that the basis functions $\mathbf{F}_n$ and the weight functions $\mathbf{F}_m$ represent currents. The inner product

$\langle \mathbf{F}, \mathbf{E} \rangle$ is a scalar quantity obtained by integrating $\mathbf{F} \cdot \mathbf{E}$ over the surface under consideration. This particular inner product is called reaction. The inner product is defined such that the following conditions are satisfied:

$$\langle \mathbf{F}, \mathbf{E} \rangle = \langle \mathbf{E}, \mathbf{F} \rangle \tag{14-58a}$$

$$\langle \alpha \mathbf{F} + \beta \mathbf{F}, \mathbf{E} \rangle = \alpha \langle \mathbf{F}, \mathbf{E} \rangle + \beta \langle \mathbf{F}, \mathbf{E} \rangle \tag{14-58b}$$

if

$$\langle \mathbf{F}^*, \mathbf{F} \rangle > 0, \quad \text{then } \mathbf{F} \neq 0 \tag{14-58c}$$

if

$$\langle \mathbf{F}^*, \mathbf{F} \rangle = 0, \quad \text{then } \mathbf{F} = 0 \tag{14-58d}$$

where α and β are scalars and * denotes complex conjugation.

The third step is to calculate the various inner products given in (14-56) and thereby form the matrix equation

$$
\begin{bmatrix}
\langle \mathbf{F}_1, L_{\text{op}}(\mathbf{F}_1) \rangle & \langle \mathbf{F}_1 \, L_{\text{op}}(\mathbf{F}_2) \rangle \dots \\
\langle \mathbf{F}_2, L_{\text{op}}(\mathbf{F}_1) \rangle & \cdot \\
\cdot & \cdot \\
\cdot & \cdot \\
\cdot & \cdot \\
\cdot & \cdot
\end{bmatrix}
\begin{bmatrix}
I_1 \\
I_2 \\
\cdot \\
\cdot \\
\cdot \\
I_N
\end{bmatrix}
=
\begin{bmatrix}
\langle \mathbf{F}_1, \mathbf{E}^i \rangle \\
\langle \mathbf{F}_2, \mathbf{E}^i \rangle \\
\cdot \\
\cdot \\
\cdot \\
\langle \mathbf{F}_N, \mathbf{E}^i \rangle
\end{bmatrix}
\tag{14-59}
$$

or, in more compact notation,

$$[Z_{\text{mn}}][I_n] = [V_m] \tag{14-60}$$

The procedure for obtaining a MoM solution in terms of linear algebra can be summarized in the following way:

1. Expand the unknown in a series of basis functions $\mathbf{F}_n$, spanning $\mathbf{J}$ in the domain of L_{op}.
2. Determine a suitable inner product and define a set of weighting functions.
3. Take the inner products and thereby form the matrix equation.
4. Solve the matrix equation for the unknown.

The first two steps are examined in more detail in Sec. 14.8.1.

EXAMPLE 14-3   *Linear Algebra Interpretation of Sec. 14.6*

The purpose of this example is to interpret the formulation given in (14-40) in terms of linear algebra as discussed in this section. From (14-40a),

$$Z_{\text{mn}} = \int_{\Delta z_m} F_m(z) f(z, z_n') \, dz$$

where

$$f(z, z_n') = -\int_{\Delta z_n'} F_n(z') K(z, z') \, dz'$$

Thus,

$$Z_{\mathrm{mn}} = \underbrace{\int_{\Delta z_m} F_m(z) \underbrace{\left[-\int_{\Delta z_n'} F_n(z') K(z, z') dz' \right]}_{L_{\mathrm{op}}(\mathbf{F}_n)} dz}_{\langle \mathbf{F}_m, L_{\mathrm{op}}(\mathbf{F}_n) \rangle}$$

where the integral operator is given by

$$L_{\mathrm{op}} = -\int_{\Delta z_n'} K(z, z') \, dz'$$

From (14-40b), write

$$V_m = \underbrace{\int_{\Delta z_m} F_m(z) E_z^i(z) \, dz}_{\langle \mathbf{F}_m, \mathbf{E}_z^i \rangle}$$

Note that the inner products actually contain a dot product of two vectors as in (14-44) and (14-45), but the integral expressions above only contain scalars since the dot products have effectively already been performed.

14.8 FORMULATION AND COMPUTATIONAL CONSIDERATIONS

The development and use of a computer model of an electromagnetic problem can be divided into the following four steps:

1. Development of the *mathematical formulation* based on the physics of the problem, the object size in terms of λ, and mathematical principles (e.g., MoM, FD-TD, etc.)
2. *Coding* the mathematical formulation into a computer algorithm
3. *Validation* of the computer code
4. *Computation* to solve analysis and design problems

The following six subsections address issues in MoM affecting item 1 of the preceding list. The last subsection addresses the important issue of validation.

14.8.1 Other Expansion and Weighting Functions

In Secs. 14.5 and 14.6, the pulse function was used as the expansion function and either the pulse function or delta function was used as the weighting function. The advantage of these functions lies in the simplicity they provide to the mathematical formulation and, hence, the coding. However, there are other functions commonly used. These include:
Triangle functions (piecewise linear):

$$J(z) = \begin{cases} \dfrac{I_n(z_{n+1} - z) + I_{n+1}(z - z_n)}{\Delta z_n} & \text{for } z \text{ in } \Delta z_n \\[2mm] 0 & \text{otherwise} \end{cases} \qquad (14\text{-}61)$$

Piecewise sinusoidal:

$$J(z) = \begin{cases} \dfrac{I_n \ \sin \beta(z_{n+1} - z) + I_{n+1} \sin \beta(z - z_n)}{\sin \beta \, \Delta z_n} & \text{for } z \ \text{in} \ \Delta z_n \\ 0 \quad \text{otherwise} \end{cases} \qquad (14\text{-}62)$$

Sinusoidal interpolation:

$$J(z) = \begin{cases} A_n + B_n \ \sin \beta(z - z_n) + C_n \ \cos \beta(z - z_n) & \text{for } z \ \text{in} \ \Delta z_n \\ 0 \quad \text{otherwise} \end{cases} \qquad (14\text{-}63)$$

where $\Delta z_n = z_{n+1} - z_n$.

The triangle functions were introduced in much of Harrington's early work and were used both as expansion and weighting functions (a Galerkin formulation). The triangle Galerkin formulation is also used in the *MININEC Professional* thin wire code [8].

Piecewise sinusoidal functions were first used by Richmond in a Galerkin formulation developed with the reaction integral equation [9]. These functions are computationally very efficient for wire geometries in free space, in part because the actual current distributions are nearly sinusoidal. A convergence curve for a dipole impedance is shown in Fig. 14-16. The rapid convergence is evident. The piecewise sinusoidal function is illustrated in Fig. 14-17a. Two segments are required to define the function. When several functions are used, as for the dipole suggested by Fig. 14-17b, each function overlaps with adjacent functions. The junction of two (or more) segments is called a node. In Fig. 14-17b, there are five nodes and there are five functions spanning six segments. Seven points (i.e., five nodes plus two end-points) are required to define the six segments. On the other hand, if there are no endpoints as in a loop, the number of nodes, the number of segments, and the number of overlapping functions are all the same.

The sinusoidal interpolation function along with delta weighting functions is used in the *Numerical Electromagnetics Code* or *NEC* code as it is widely known [10].

The choice of functions has been the subject of research in past years. Some discussion of this may be found in [11] and [H.10.2: Moore]. The choice of functions is also influenced by a consideration of how to treat junctions of more than two wires. In the case of the (nonoverlapping) pulse function, no special consideration is required. In fact, Kirchhoff's current law will automatically be satisfied at a multiwire junction as a consequence of Maxwell's equations being satisfied. In the case of overlapping functions (e.g., triangle, piecewise sinusoid) at a junction of N wires, there are $N - 1$ independent currents (the Nth being determined by Kirchhoff's current law); therefore, only $N - 1$ functions are needed that go across the N-wire junction (see Prob. 14.13-3).
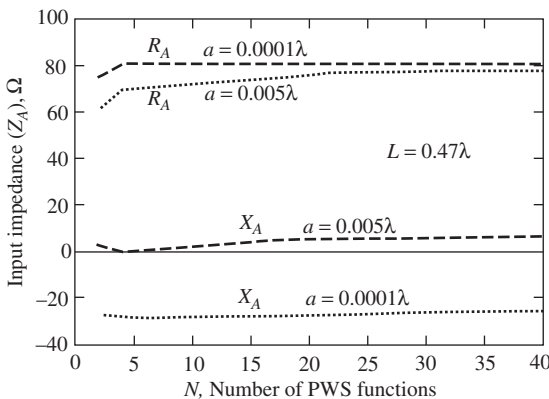
Figure 14-16 Input impedance convergence of a piecewise sinusoidal (PWS) Galerkin code [9] for two different wire radii.

(a) Piecewise sinusoidal expansion function.

(b) Set of overlapping expansion functions on an open-ended wire.
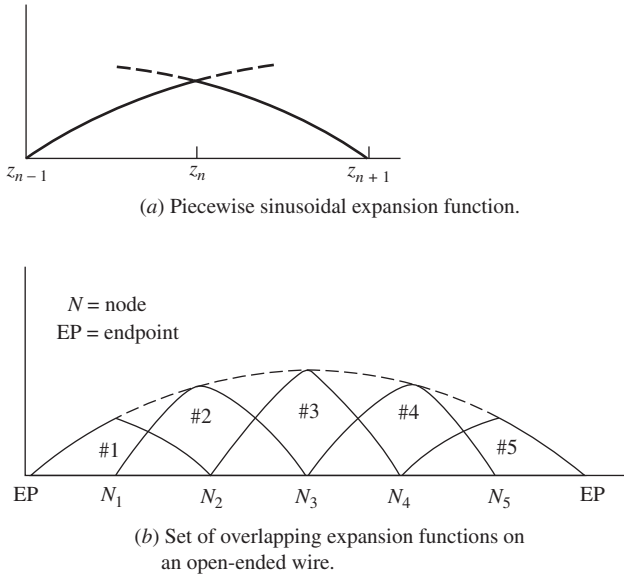
Figure 14-17 Illustration of overlapping expansion functions using the piecewise sinusoid.

14.8.2 Other Electric Field Integral Equations for Wires

One form of an electric field integral equation (EFIE) is the Pocklington form in Sec. 14.3. Another form is the potential form used by Harrington [3] in his pioneering work. For z-directed wires, the potential form is

$$E^i = \int_{-L/2}^{L/2} \left[j\omega\mu_o I(z') - \frac{1}{j\omega\varepsilon_o} \frac{\partial I(z')}{\partial z'} \frac{\partial}{\partial z} \right] \frac{e^{-j\beta R}}{4\pi R} \, dz' \tag{14-64}$$

which is derived by using both the vector and scalar potentials (see Probs. 14.3-3 and 14.8-1). This form can provide more rapid convergence than the Pocklington form, which implies that convergence rates are dependent on the characteristics of the kernel as well as the choice of expansion and weighting functions.
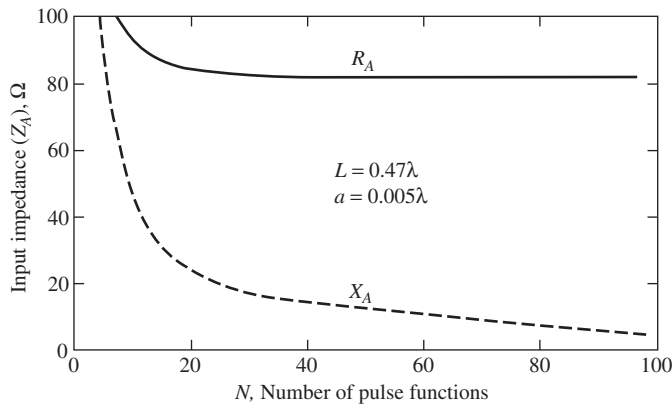
Figure 14-18 Convergence of Hallen's equation using pulse expansion functions and point-matching.

A quite different appearing equation for wire antenna work is that due to Hallen (see Prob. 14.8-2). Hallen's integral equation for a z-directed wire antenna is

$$\int_{-L/2}^{L/2} I(z') \frac{e^{-j\beta R}}{4\pi R} dz' = -\frac{j}{\eta} \left(C_1 \cos \beta z + \frac{V_A}{2} \sin \beta |z| \right) \qquad (14\text{-}65)$$

where V_A is the terminal voltage.

Hallen's equation has a simple kernel and is generally simpler to code than the either the Pocklington or potential forms. For N unknowns, there must be $N+1$ equations because the constant C_1 is an unknown as well [12]. Example 14-4 shows a sample impedance matrix, whereas Fig. 14-18 shows the rate of convergence for pulse expansion functions and delta weights.

EXAMPLE 14-4   *Subdomain Solution of Hallen's Equation*

The purpose of this example is to implement a pulse expansion function, point-matching solution to Hallen's equation for a z-directed wire dipole. The solution can be expressed as

$$\sum_{n=1}^{N} I_n \; Z_{mn} + C_1' \cos \beta z_m = V_m, \quad m = 1, 2, \ldots, N+1$$

where

$$Z_{mn} = \int_{-\Delta z_n'/2}^{\Delta z_n'/2} \frac{e^{-j\beta R}}{R} dz', \quad V_m = -j\frac{2\pi V_A}{\eta} \sin \beta |z|, \quad C_1' = j\frac{4\pi}{\eta} C_1$$

and R is given by (l4-11). Treating C_1' as an unknown rather than C_1 significantly improves the condition number of the matrix $[Z_{mn}']$ shown below. The Z_{mn} for $m \neq n$ are easily computed by numerical integration. For $m = n$, special care may be necessary for very small wire radii. The match points are chosen at the center of each pulse function, but an odd number of functions are required so that there is a match at the feed location. Further, it is necessary to have a match point at one end of the dipole (see Prob. 14.8-2). This is accomplished conceptually by employing a "phantom pulse" extending $\Delta z/2$ beyond one end of the dipole such that there is a match point at the end of the dipole.

For a dipole 0.1λ in length, radius 0.005λ, and three pulse functions, the following modified (because of the C_1' term) impedance matrix is written as

$$\begin{bmatrix} [Z_{mn}] & \vdots & Z_{1,N+1}' \\ ----\!-\!| & & \vdots \\ Z_{N+1,1}' & \!\!\cdots\cdots & Z_{N+1,N+1}' \end{bmatrix}$$

where the block $[Z_{mn}]$ is toeplitz, the $N+1$ column is given by

$$Z_{1,N+1}' = \cos(\beta z_i)$$

and the remaining elements $Z_{N+1,i}'$ are found from the numerical evaluation of the integral (above) in this example (see Fig. 14-19). Thus, for the short dipole numerical example being used here, the modified impedance matrix is

$$[Z_{mn}'] = \begin{bmatrix} 3.83 \angle{-3.13°} & 1.08 \angle{-11.1°} & 0.51 \angle{-23.6°} & 0.98 \angle{0°} \\ 1.08 \angle{-11.1°} & 3.83 \angle{-3.13°} & 1.08 \angle{-11.1°} & 1.00 \angle{0°} \\ 0.51 \angle{-23.6°} & 1.08 \angle{-11.1°} & 3.83 \angle{-3.13°} & 0.98 \angle{0°} \\ 0.40 \angle{-29.7°} & 0.69 \angle{-17.4°} & 2.59 \angle{-4.62°} & 0.95 \angle{0°} \end{bmatrix}$$

Equivalent
filament
$I(z')$

$$Z_{43} = \int\limits_{-0.05}^{\frac{-0.05}{3}} \frac{e^{-j\beta R}}{R}\, dz'$$

$$R = \sqrt{a^2 + (z_4 - z')^2}$$

$F_3(z')$

z_1

Center line of
original dipole

z_2

Voltage
source
location

z_3

R

z_4 (match point)
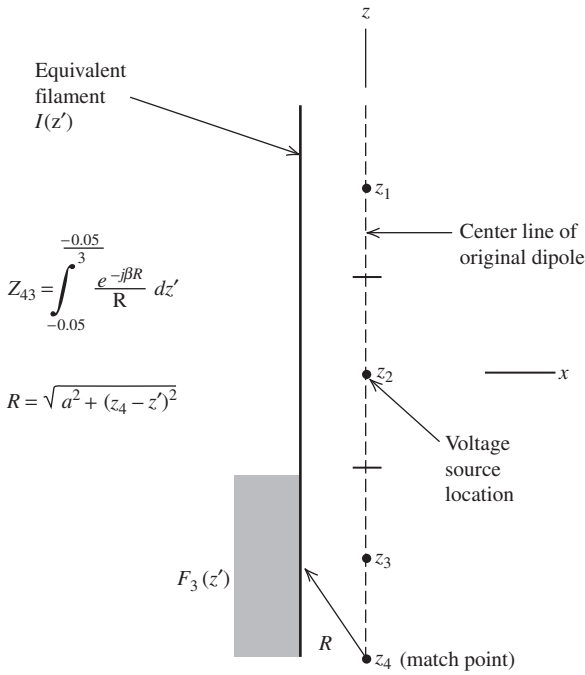
Figure 14-19 Calculation of
Z_{43} for the short dipole in
Example 14-4.

The voltage matrix is

$$[V_m] = 10^{-3} \begin{bmatrix} 3.46\ \angle -90° \\ 0\ \angle -90° \\ 3.46\ \angle -90° \\ 5.15\ \angle -90° \end{bmatrix}$$

Solving for the current matrix gives

$$[I_n] = 10^{-3} \begin{bmatrix} 0.84\ \angle 89.45° \\ 1.98\ \angle 89.69° \\ 0.84\ \angle 89.45° \\ 9.41\ \angle -95.02° \end{bmatrix}$$

Note that I_4 is the value for the coefficient C_1' for which there is no further use. The input impedance is $1.0/(1.98 \times 10^{-3} \angle 89.69°) = 2.75 - j504$ Ω. If five pulses were used instead of just three, the input impedance would be $2.1 - j489$ Ω, which more closely agrees with the approximate value of 1.974 Ω for the real part obtained from $20\pi^2 (L/\lambda)^2$.

14.8.3   Computer Time Considerations

Historically, there have been two computer limitations to the use of MoM: the amount of computer storage necessary for the N^2 impedance matrix elements and the amount of time to compute those N^2 elements and solve the resulting system of equations. Computer technology has significantly decreased the impact of the first limitation. Iterative methods have increased the speed of solution of a system of dense simultaneous equations, leaving the time required to compute the N^2 elements as a major limitation on

the electrical size of an object that can be reasonably accommodated by MoM (see Fig. 14-2). Let us examine where the computation time is spent.

A square impedance matrix of N^2 elements is said to be of order N. Let N_i be the number of different source or incident fields (i.e., radar cross section is a function of incidence angle) associated with a given impedance matrix and let N_a be the number of observation points at which the field is to be computed from the current solution; then the time t for execution will be approximately given by [6]

$$t \simeq AN^2 + B_3N^3 + CN^2N_i + DNN_iN_a \qquad (14\text{-}66)$$

where the algorithm and computer-dependent factors $A, B, C,$ and D are

A = time required to compute a typical impedance matrix element
B_3N^3 = time required to solve $[Z_{mn}][I_n] = [V_m]$ for $[I_n]$ by matrix inversion $[I_n] = [Z_{mn}]^{-1}[V_m]$ for a system of order N
CN^2N_i = time required to perform the operation $[Z_{mn}]^{-1}[V_m]$ or its equivalent for each *new* $[V_m]$
DNN_iN_a = time for computing the far field from $[I_n]$

The second term in (14-66) dominates. However, it is unlikely that we would solve a large system of equations by finding the inverse. Instead, usually an algorithm such as Gauss-Jordan or Crout is used, in which case $B_3N^3 \rightarrow B_2N^2$ and we have

$$t \simeq AN^2 + B_2N^2 + CN^2N_i + DNN_iN_a \qquad (14\text{-}67)$$

which is a significant reduction in the solution time required for large N. If iterative methods are used, more favorable reductions are possible using advanced solver algorithms such as the fast multipole method (FMM) discussed next.

Solving a MoM matrix using direct matrix inversion limits the technique to low-frequency applications because the problem size (N) is typically proportional to frequency squared using surface patch discretization. [13] This translates to computational time being proportional to frequency to the sixth power. For example, if a problem can be solved at 1 GHz, to solve the same problem at 10 GHz will require at least 10^4 times the computer resources and take 10^6 times longer to solve. Even if computer speed doubles every 18 months, the increase in computational speed only produces a minuscule decrease in solve time. The use of iterative solvers reduces the computational time to N^2. Even so the computational time is proportional to the frequency to the fourth power. It is evident that for the MoM to be of practical use, parallel computing technology needs to be coupled with advanced solver algorithms that further reduce the computational time scaling factor from N^2 to N. Several numerical techniques [14–18] were developed in the early 1990s to address this need to further reduce the solution time for iterative methods. The discussion of each technique is beyond the scope of this book; the fast multipole method will be briefly described here as it has received the most intense research compared to other fast solver techniques.

The fast multipole method (FMM) [14–16, 19–20] was developed to further reduce the solution time for iterative methods by speeding up the matrix-vector multiply step during the iterative solve. In the FMM procedure, the basis functions of an object are divided into groups. For nearby groups, the impedance matrix elements are calculated by direct numerical integration of the radiation integral similar to how the near diagonal terms of the impedance matrix are calculated for direct solve. For source and test elements in two well-separated groups, the Green's function is replaced by a multipole expansion using the addition theorem. This allows the source point and test point to be decoupled but related through the center of their respective groups. The interaction between the source and test elements is calculated by "aggregating" the radiations from source elements to the center of the source group, "translating" the interaction from the center of the source

group to the center of the test group, then "disaggregating" the received fields from the center of the test group to the test element. Applying FMM to an iterative solver can reduce the solve time from $O(N^2)$ to $O(N^{1.5})$ for the single-level FMM or to $O(N \log N)$ for the multilevel FMM. [21, 22] While the scaling factor is reduced, the constant associated with it is much larger than that for direct solve and iterative solve.

The steps involved in the FMM implementation are much more complicated than direct solvers. In addition, the accuracy of the FMM solution depends on proper setting of the convergence tolerance for the iteration steps and the number of terms included in the multipole expansion. Fast direct solvers, such as the adaptive cross approximation (ACA) [23], that minimize the burden of implementation and application have become an area of active research. Fast solver and parallel computing technologies have made the MoM practical for many engineering problems.

When the second term in (14-66) is on the order of N^2 or less, then usually the first term in (14-66), which is associated with the time required to calculate the matrix elements, becomes the dominating factor. In the following subsections, we examine briefly some ways, other than FMM, for reducing the total time required for the operations indicated in (14-66).

14.8.4   Toeplitz Matrices

Certain types of problems produce impedance matrices where there is a systematic repetition in the matrix elements. Often, this repetition can be used to decrease the impact of both the first and second terms in (14-66). Consider the straight wire in Fig. 14-3. If the segments are of equal length, all the values of the N^2 matrix elements are contained in any one row of $[Z_{mn}]$, say, the first one. All other rows are merely a rearranged version of the first. The remaining elements can be obtained by the rearrangement algorithm:

$$Z_{mn} = Z_{1,|m-n|+1}, \quad m \geq 2, \quad n \geq 1 \tag{14-68}$$

Such a matrix is said to be a *toeplitz matrix*. Computer programs exist for solving toeplitz matrices that are considerably more efficient than those for solving a non- toeplitz matrix. For a toeplitz matrix, the first two terms in (14-66) become AN and $BN^{5/3}$, respectively, and the execution time in (14-66) is reduced as in

$$t \simeq AN + B_1 N^{5/3} + CN^2 N_i + DNN_i N_a \tag{14-69}$$

for which there is a significant improvement in the first term as well as the second.

Toeplitz matrices can arise in the treatment of certain wire geometries. These are the straight wire (see Examples 14-1 and 14-2), the circular loop, and the helix. A toeplitz matrix can also arise in the treatment of geometries other than the wire, but these are outside the scope of this chapter.

14.8.5   Block Toeplitz Matrices

Consider the linear array of parallel dipoles in Fig. 14-20. The impedance matrix that characterizes the array will be toeplitz by blocks or submatrices when the array elements are of the same length and equally spaced. Thus, if the impedance matrix for the array $[Z]_{array}$ is written in terms of submatrices $[S]$ as

$$[Z]_{array} = \begin{bmatrix} [S]_{11} & [S]_{12} & \cdots & [S]_{1J} \\ [S]_{21} & [S]_{22} & \cdots & [S]_{2J} \\ \vdots & & \ddots & \vdots \\ [S]_{J1} & \cdots & & [S]_{JJ} \end{bmatrix} = \begin{bmatrix} [S]_{11} & [S]_{12} & \cdots & [S]_{1J} \\ [S]_{12} & [S]_{11} & \cdots & [S]_{1(J-1)} \\ \vdots & & \ddots & \vdots \\ [S]_{1J} & [S]_{1(J-1)} & \cdots & [S]_{11} \end{bmatrix} \tag{14-70}$$
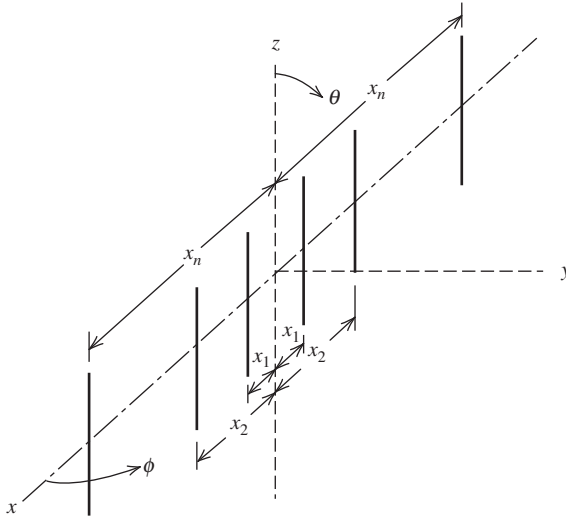
Figure 14-20 Linear array of parallel dipoles.

where $[S]_{ij} = [Z_{mn}]$, the entire impedance matrix is toeplitz by blocks. Thus, if one row of submatrices is known, the remaining submatrices may be filled by the algorithm

$$[S]_{ij} = [S]_{1,|i-j|+1}, \quad i \geq 2, \quad j \geq 1 \tag{14-71}$$

Consequently, the first term in (14-66) is of order N^2/J, where J is the number of independent submatrices. The second term in (14-66) will be of order $N^{9/5}$.

If the submatrices are themselves toeplitz, as they would be if all segments are of identical length and radius, then the matrix fill time is reduced even further. Computer programs exist for solving block toeplitz matrices. The potential savings in execution time for a problem that is block toeplitz over the same size nonblock toeplitz problem can be considerable.

14.8.6 Compressed Matrices

In certain problems, there will be a repetition in the values within $[I_n]$ due to the symmetry of the problem. If this can be recognized in advance, it can be used to advantage to compress the matrix from order N to order N/L, where L is the degree of symmetry.

Consider the following simple but very common example of symmetry suggested by Fig. 14-7*a*. Here, the monopole and its image will have a symmetrical current about the feed point. Suppose $I_n = I_{n+N/2}$; then we can write

$$\sum_{n=1}^{N/2} (Z_{mn} + Z_{m,n+N/2})I_n = V_m, \quad m = 1, 2, 3, \ldots, N/2 \tag{14-72}$$

The solution of this compressed system of $N/2$ equations will yield the $N/2$ independent I_n's. From (14-67), we see that the solution time for the system is $B_2(N/2)^2$, or a reduction in time by a factor of 4 for this portion of the computing process. For higher degrees of symmetry, the savings in time would be even more considerable. For some large problems, it may be necessary to compress the matrix for another reason, namely storage requirements. It is possible that an impedance matrix may be so large that it cannot be stored in readily available core memory and that through symmetry it may be compressed to a reasonable size.

So, it is the execution time and computer storage that tend to limit the electrical size of problems that may be reasonably treated with MoM. In some of the following sections, we will approach certain situations in such a way that the impact of these two limitations is minimized.

14.8.7 Validation

A computer code must reliably produce accurate results if it is to be useful. Errors in the code are most likely to occur because of an encoding mistake, but can also occur due to an oversight in the mathematical formulation or cumulative numerical errors caused by large numbers of numerical operations and/or inadequate numerical precision. The investigation of the possible existence of errors such as these, as well as others, is the process of validation. We examine two kinds of validation: external validation and internal validation.

External validation usually means that the output of a code is compared against either experimental measurements or the output of an independent code (perhaps thereby validating both codes). Near-field quantities, such as antenna input impedance, are more critical gages than far-field quantities. Consequently, earlier figures in this chapter have used the input impedance as a check against both experimental data and results from other independent codes.

Internal validation usually implies that additional coding and/or computational effort is involved. Some examples of internal checks are:

1. A convergence check to establish that a limiting value is smoothly approached as N is increased
2. A power balance check where the power supplied by the incident field is equal to the radiated plus dissipated powers
3. A reciprocity check where source and observation points are interchanged
4. Boundary condition check where the applicable boundary condition (e.g., total tangential E is zero) is met

Of these four checks, only the last is both necessary and sufficient. The others are necessary but not sufficient checks.

In the case of a code that is written to handle a wide variety of geometries, frequencies, and so on, representative sample of such situations must be validated much like a new aircraft must be test flown in a variety of different configurations, speeds, weather conditions, and so on, before it can be certified as safe for general use.

14.9 CALCULATION OF ANTENNA AND SCATTERER CHARACTERISTICS

Thus far, our discussion of MoM has been mainly concerned with acquiring a knowledge of some unknown current distribution. Now consider how we can obtain other information as well. But first, we should make one further remark about the currents derived from the solution of the matrix equation.

If pulse functions are used as the expansion functions in the point-matching technique, a knowledge of the current coefficients I_n means that the current distribution at the match points is known "precisely," if we assume, of course, that the solution has converged. In between the match points, we do not know the current, but since the distance between the match points is small in terms of the wavelength, one can simply fit a curve through the current values at the match points to obtain a good approximation of the current distribution along the wire.

In the case of overlapping functions, such as the piecewise sinusoidal or the triangle (see the previous section), a knowledge of the coefficients I_n again only means that the

current is known at the junctions of the segments. Along the segments, we use the overlapping functions themselves to approximate the current distribution between segment junctions.

After we have determined the current distribution, the input or terminal current can be found by evaluating the current distribution at the antenna terminal location. In turn then, the input impedance may be calculated by dividing the terminal voltage by the terminal current. The calculation of accurate impedance data is a task that is somewhat sensitive to the model used for the feed point. Two such models were discussed in Sec. 14.5.

Distributive loading, which arises when a wire is not perfectly conducting, may affect the current distribution in certain situations. For simplicity, consider a wire whose axis is parallel to the z-axis. When the wire has finite conductivity, we can relate the tangential electric field at the surface of the wire to the equivalent electric surface current density by the use of the surface impedance Z_s, which is defined as the ratio of the tangential electric field strength at the surface of a conductor to the current density that flows as a result of that tangential electric field. Thus,

$$\mathbf{E} = Z_s \mathbf{J}_s \tag{14-73}$$

Using $\mathbf{M}_s = \mathbf{E} \times \hat{\mathbf{n}}$ and the relationship $\mathbf{J}_s = \hat{\mathbf{z}} I(z)/2\pi a$, we can write

$$\mathbf{M}_s = Z_s \mathbf{J}_s \times \hat{\rho} = \frac{\hat{\phi} Z_s I(z)}{2\pi a} \tag{14-74}$$

Writing the reaction integral equation from (14-46) and reciprocity as

$$-\int\int_S (\mathbf{E}_m \cdot \mathbf{J}^s - \mathbf{H}_m \cdot \mathbf{M}^s)\, ds = V_m \tag{14-75}$$

and substituting (14-74) lead to

$$-\int I(z)[\hat{\mathbf{z}} \cdot \mathbf{E}_m - Z_s \hat{\phi} \cdot \mathbf{H}_m)\, dz = V_m \tag{14-76}$$

Using (14-16) in (14-76), we can write the generalized impedance matrix element Z'_{mn}, modified for finite conducting wires, as

$$Z'_{mn} = \int_{\Delta z_n} F_n(z)\hat{\mathbf{z}} \cdot \mathbf{E}_m\, dz - Z_s \int_{\Delta z_n} F_n(z)\hat{\phi} \cdot \mathbf{H}_m\, dz \tag{14-77}$$

From Ampere's law, a suitable approximation for $\mathbf{H}_m$ is

$$\hat{\phi} \cdot \mathbf{H}_m = \frac{F_m(z)}{2\pi a} \tag{14-78}$$

and thus (14-75) can be written as

$$Z'_{mn} = Z_{mn} - \frac{Z_s}{2\pi a} \int_{(m,n)} F_n(z) F_m(z)\, dz \tag{14-79}$$

where region (m, n) is the wire surface shared by testing or weighting function m and expansion or basis function n. In the case of overlapping expansion functions, this region covers two intersecting segments if m and n are equal. When $m \neq n$, the shared region covers, at most, one wire segment. This means that distributive loading is accounted for by a modification of only the appropriate main diagonal elements, and those elements adjacent to the modified main diagonal elements if overlapping functions are being used.

In the case of nonoverlapping functions, such as the pulse, only the main diagonal elements are modified.

The effect of either lumped loading (considered in Sec. 14.10) or distributive loading is to alter the current distribution on the wire antenna or scatterer. If we, know the current distribution, the far field can be obtained by the classical methods used previously in this book (e.g., Sec. 2.4). To illustrate, consider again z-directed segments with pulse expansion functions of the current. Then from (8-41), we can write

$$\mathbf{E} = \hat{\boldsymbol{\theta}} j\omega\mu \frac{e^{-j\beta r}}{4\pi r} \Delta z \sin\theta \sum_{n=0}^{N-1} I_n e^{j\beta z'_n \cos\theta} \tag{14-80}$$

where z'_n is the center of each short segment.

Once the far field is known, the gain may be determined from the general relationship

$$G(\theta, \phi) = \frac{[|E_\theta|^2 + |E_\phi|^2]r^2}{30|I_A|^2 R_A} \tag{14-81}$$

where R_A is the real part of the antenna input impedance. The directivity is obtained by replacing R_A with R_r, the radiation resistance.

The radar cross section may be found as (similar to (4-67))

$$\sigma = \lim_{r \to \infty} 4\pi r^2 \frac{|E^s|^2}{|E^i|^2} \tag{14-82}$$

where E^s can be determined, for example, from (14-12). The radar cross section for a short circuited dipole scatterer is shown in Fig. 14-21.

The radiation efficiency is calculated using (2-178) as

$$e_r = \frac{R_r}{R_A} = \frac{R_r}{R_r + R_o} \tag{14-83}$$

R_o is the loss resistance due to dissipative loading, either distributed or lumped (see Sec. 14.10.1). Alternatively, we could determine the radiated power by integrating the power density in the far field as we did in Chap. 2. However, the above method (14-83) is computationally more efficient.
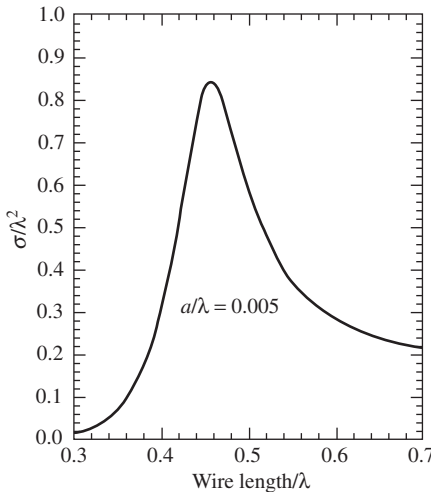
Figure 14-21 Monostatic radar cross section of a straight wire at normal incidence as a function of wire length.

From the discussion in this section, we can see, among other things, that refinements such as losses due to finite conductivity (distributive loading) or the effects of lumped loading can be included into a moment method solution in a fairly straightforward manner. In the next section, we investigate lumped loading further.

14.10   THE WIRE ANTENNA OR SCATTERER AS AN *N*-PORT NETWORK

In Sec. 14.4, we saw the resemblance between the simultaneous linear equation approximation of an integral equation and Kirchhoff 's network equations. It follows that we may view the junction of two or more segments as a port in the usual circuit sense as indicated by Fig.14-22*a*. At each port, we may place either series or parallel elements that are either passive or active. Series connections are treated on an impedance basis, whereas parallel connections are handled on an admittance basis. This section considers both types.

14.10.1   Series Connections

We already have considered a single generator placed at the junction of two wire segments (e.g., Sec. 14.5). The generator was in series with the implied port terminals located at the ends of the two adjacent segments. We could, of course, place as many generators on the wire as there are segment junctions. Thus, for an *N*-segment dipole, there would be $N - 1$ ports. If there is no generator or passive element across the port, the port is understood to be short-circuited.

Previously in Sec. 14.9, we saw how distributed loading was accounted for in the moment method by modifying certain elements in the impedance matrix. Here, let us

$$V_m = V_m^g - I_m Z_m$$

(*b*) Equivalent circuit for the *m*th port.
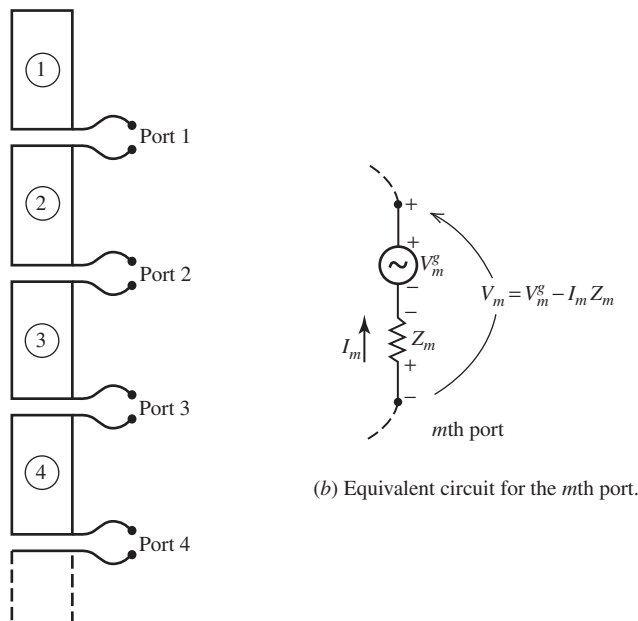
(*a*) $N - 1$ port terminal pairs.

Figure 14-22 *N* segment wire showing ports.

examine how lumped loading may be handled. If a load Z_m is inserted into a wire antenna at segment junction m having a current I_m, the total voltage acting at that point is

$$V_m = V_m^g - I_m Z_m \qquad (14\text{-}84)$$

where V_m^g represents a voltage generator that may be located at point m in series with Z_m as indicated in Fig. 14-22b. In many cases, V_m^g will be zero. Considering the mth equation in a system of N linear equations, we can write

$$\sum_{n=1}^{N} Z_{mn} I_n = V_m^g - I_m Z_m \qquad (14\text{-}85)$$

or

$$\sum_{n=1}^{N} Z'_{mn} I_n = V_m^g \qquad (14\text{-}86)$$

where

$$Z'_{mm} = Z_{mm} + Z_m \qquad (14\text{-}87)$$

Except for the diagonal elements, the new impedance matrix is the same as the original, or $Z'_{mn} = Z_{mn}$, $m \neq n$. Thus, the effect of lumped loading may be accounted for by simply adding the load impedances Z_m to the corresponding diagonal elements in the impedance matrix. The effects of lumped loading can be substantial. For example, resistive loading can be used to achieve increased bandwidth, but at the expense of lower efficiency. Or, lumped loading can be used for impedance matching as in Fig. 6-45.

14.10.2 Parallel Connections

In the previous subsection, we saw how circuit elements, when connected in series at a given port, resulted in modification of certain entries in the open-circuit moment method impedance $[Z_{mn}]$. If, however, we connect one port in parallel with another as in a log-periodic antenna, then it is necessary to work with the short-circuit moment method admittance matrix $[Y_{mn}]$.

Consider Fig. 14-23 that shows a log-periodic dipole antenna (LPDA). The LPDA is viewed as the parallel connection of two N-port networks. One N-port represents the mutual coupling between N dipole antennas. The other represents the transmission line that interconnects the dipoles. Therefore, there is one *network* port for each of the dipoles in the system.

The approach is shown schematically in Fig. 14-24. The N-port labeled "antenna elements" includes the self- and mutual impedances between N unconnected dipole antennas located arbitrarily in space. The "transmission line" N-port represents the transmission line connecting the dipole antennas. Included in this network is the effect of reversing the polarity between successive dipoles. Note that there is a current source I_s on the LPDA. If there are N_e dipole elements on the antenna, then there are voltage sources applied on ports $1, 2, 3, \ldots, N_e$. Also, there is a terminating admittance on the LPDA antenna Y_t that exists at port 1. We do not know the numerical values of the applied voltage sources. Thus, they must be found before we can solve for the currents on the LPDA.

Let $[Y_A]$ and $[Z_A]$ be the short-circuit admittance matrix and open-circuit impedance matrix, respectively, for the "antenna elements" network, where $[Y_A] = [Z_A]^{-1}$. We note that $[Z_A]$ is *not* the moment method impedance matrix. An element of $[Z_A]$, say, $[Z_A]_{ij}$, represents the voltage induced on dipole i in the LPDA by a unit current on dipole j with all other dipoles open-circuited. Thus,
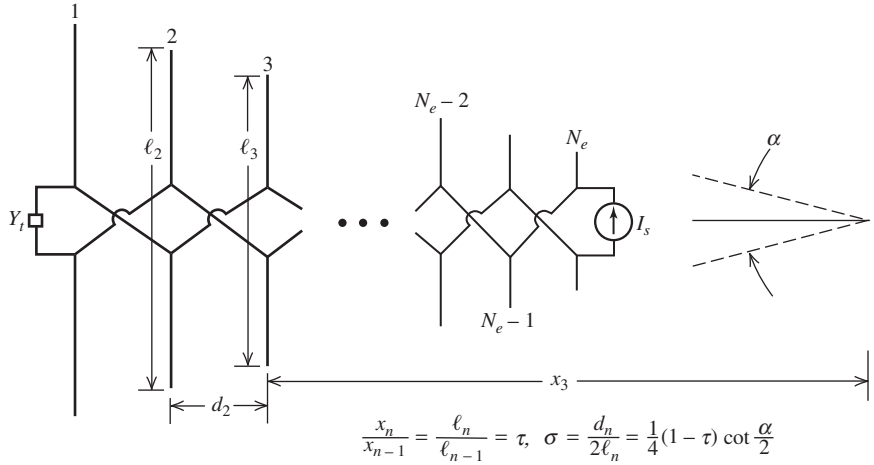
Figure 14-23 Log-periodic dipole antenna.
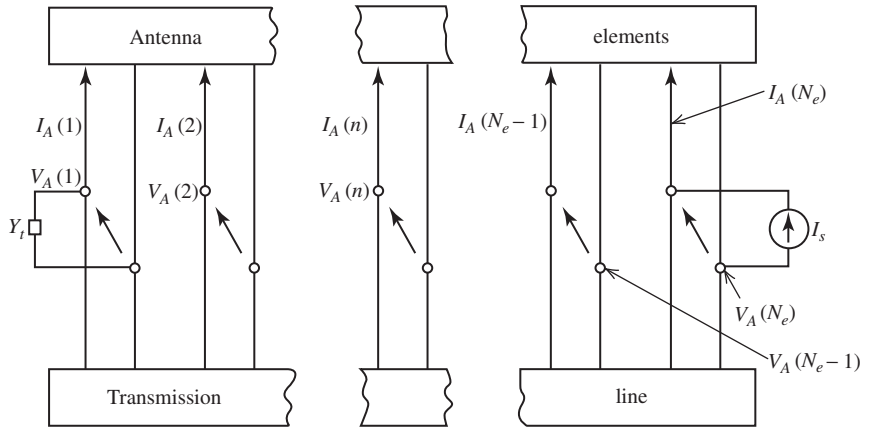
Figure 14-24 Schematic representation of the LPDA network of Fig. 14-23.

$$[Z_A]_{ij} = \frac{V_i}{I_j} \tag{14-88}$$

Let $[Y_T]$ be the short-circuit admittance matrix for the "transmission line" network. Let $[I_A]$ and $[V_A]$ be the column matrices representing the voltage and current at each port of the "antenna elements" network. Since the two networks are in parallel, the total current can be written as

$$[I_s] = [[Y_A] + [Y_T]][V_A] \tag{14-89}$$

where $[I_s]$ represents the applied current sources. The $[I_s]$ matrix contains all zero elements, except at the port where there is a current source I_s. The current source, of course, represents the excitation of the LPDA antenna. Note in (14-89) that we know the entry in $[I_s]$ but not the entries in $[V_A]$. These must be found so that the moment method column matrix $[V_m]$ can be constructed and the usual equation $[I_n] = [Z_{mn}]^{-1}[V_m] = [Y_{mn}][V_m]$ solved for the current distributions in the antenna dipole elements. But before we can solve (14-89) for $[V_A]$ and construct $[V_m]$, we must know $[Y_A]$ and $[Y_T]$.

To obtain the elements of $[Y_A]$, we proceed as follows. Consider an LPDA with N_e dipoles and M expansion functions on each dipole. The moment method impedance matrix will be of order $N_e \times M$. To obtain the moment method admittance matrix $[Y_{mn}]$, we note that $[Y_{mn}] = [Z_{mn}]^{-1}$ and

$$[I] = [Z_{mn}]^{-1}[V] = [Y_{mn}][V] \tag{14-90}$$

or

$$I_m = \sum_{n=1}^{M \times N_e} Y_{mn} V_n, \quad m = 1, 2, \ldots, M \times N_e \tag{14-91}$$

To obtain $[Y_A]$, we note that most of the V_n's will be zero since voltages are only applied by the transmission line on the center ports of each dipole in the LPDA. Suppose we rearrange the system of equations in (14-91) such that the first N_e entries in $[V]$ as well as $[I]$ correspond to the center ports of the dipoles in the LPDA. Then the currents at those ports containing a generator (i.e., antenna element ports) are related to the voltages at those ports by

$$I_j = \sum_{i=1}^{N_e} Y_{ji} V_i, \quad j = 1, 2, \ldots, N_e \tag{14-92}$$

or

$$[I_A] = [Y_A][V_A] \tag{14-93}$$

where all the V_i's in $[V_A]$ will be nonzero. (See Prob. 14.10-1.) In finding $[Y_A]$ in this manner, we have done so without making approximations other than those appropriate to the moment method itself. Indeed, all mutual couplings are included and we are not limited to LPDAs of less than 2:1 bandwidth as in the treatments by Carrel [24] and Kyle [25].

To obtain the transmission line admittance matrix $[Y_T]$ in (14-89), we first recognize that $[Y_T]$ is the transmission line admittance matrix for a simple terminated transmission line with a port at the position where each dipole is connected. Since $[Y_T]$ is the short-circuit admittance matrix, a given element $(Y_T)_{ji}$ represents the current induced across port j (which is shorted) by a unit voltage at port i, with all other ports shorted. Thus, $(Y_T)_{ji}$ is nonzero only for $i - 1 \leq j \leq i + 1$.

It is possible to write the transmission line admittance matrix $[Y_T]$ in a straightforward fashion [25]. For a single LPDA, it is

$$[Y_T] = \begin{bmatrix} (Y_t - jY_o \cot \beta d_1) & -jY_o \csc \beta d_1 & 0 & \cdots & 0 \\ -jY_o \csc \beta d_1 & -jY_o(\cot \beta d_1 + \cot \beta d_2) & -jY_o \csc \beta d_2 & \cdots & 0 \\ 0 & -jY_o \csc \beta d_2 & -jY_o(\cot \beta d_2 + \cot \beta d_3) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & -jY_o \csc \beta d_{N_e-1} & -jY_o \cot \beta d_{N_e-1} \end{bmatrix}$$

$$\tag{14-94}$$

where Y_o is the transmission line characteristic admittance and β the propagation constant of the transmission line. (See Prob. 14.10-2.)

With the proper elements of both $[Y_A]$ and $[Y_T]$ in hand, the voltages $[V_A]$ acting at the driven port of each dipole are

$$[V_A] = [[Y_A] + [Y_T]]^{-1}[I_s] \tag{14-95}$$

where $[I_s]$ has one nonzero entry. With these voltages at each dipole, the moment method voltage matrix $[V_m]$ can be filled and the current distribution on each dipole in the LPDA obtained from

$$[I_n] = [Z_{mn}]^{-1}[V_m] \tag{14-96}$$

where the elements of $[V_A]$ are the nonzero elements of $[V_m]$ and the elements of $[I_n]$ are the complex coefficients associated with the expansion functions on the various dipole elements.

It is worthwhile to summarize the above procedure for analyzing the LPDA. First, the open-circuit impedance matrix $[Z_{mn}]$ was formed in the usual manner. By taking the inverse of $[Z_{mn}]$, the short-circuit admittance matrix was obtained. Next, the antenna elements admittance matrix $[Y_A]$ was formed from $[Y_{mn}]$ as in (14-91) and (14-92). Then $[Y_A]$ was added to the transmisstion line admittance matrix $[Y_T]$, Then the current generator shown in Figs. 14-23 and 14-24 was used in (14-95) to obtain the voltage $[V_A]$ acting at each dipole port. These voltages were then used to obtain the moments method voltage matrix $[V_m]$. Solution for the currents $[I_n]$ on each dipole in the LPDA followed according to (14-96). Patterns obtained using this procedure are given in Fig. 7-43 and agree with those in [H.8.3: Rumsey].

14.11 ANTENNA ARRAYS

The use of moment methods in the analysis and design of arrays of wire antennas (or scatterers) has significant advantages over the more classical methods used in treating arrays in that mutual coupling between array elements is taken completely into account (e.g., see the LPDA treatment in Sec. 14.10.2). Furthermore, no unrealistic assumptions need be made regarding the current distributions on the wires, and the array elements can be excited at any point(s) or be loaded at any point(s) along their lengths. Thus, the type of wire element array problem that can be considered is rather general. In this section, we examine several array configurations of parallel dipoles and illustrate some typical mutual coupling effects.

14.11.1 The Linear Array

Consider the linear array of parallel wire elements shown in Fig. 14-20. The elements need not be of the same length and radius or be equally spaced in order to be treated by MoM. Clearly, they could be quite arbitrarily configured and, in fact, need not even be parallel. However, in this subsection we wish to illustrate the effects of mutual coupling in a typical linear dipole array by comparing MoM results (using a voltage generator with an internal impedance of $72\,\Omega$) with results suggested by the methods of Chap. 8 (i.e., current generator excitation). For this purpose, without loss of generality, we consider a linear array of 12 equally spaced $(d = \lambda/2)$, parallel, center-fed, half-wave dipoles phased for a beam maximum $45°$ off broadside. Each dipole is divided into six segments and a 1-V generator is placed in series with a $72\,\Omega$ resistance at the center port of each dipole, the piecewise sinusoidal current amplitudes obtained using (l4-62) and the methods of Sec. 14.7.1 are given in Table 14-2. We note that neither the feed point currents nor the input impedances [see (8-93)] are identical across the-array. This is due to mutual coupling. Since the main beam is at $\phi_o = 45°$, there is no symmetry in the currents about the array center as there would be if the array were phased for broadside radiation.

The normalized patterns are shown in Figs. 14-25a and 14-25b along with the normalized pattern for uniform current excitation. In spite of the differences noted in Table 14-2, there is little difference seen in the three normalized patterns shown in Figs. 14-25a and 14-25b. There is, of course, some small difference in the directivity

Table 14-2 Normalized Terminal Currents for a Linear Array of 12 Half-Wavelength Spaced, Parallel, Half-Wave Dipoles; $a = 0.0001\lambda$

Element Number	Zero Generator Impedance		72-Ω Generator Impedance	
	$|I_A|$	$|Z_A|$	$|I_A|$	$|Z_A|^a$
1	0.689	107.1	0.746	111.9
2	0.698	105.9	0.760	108.6
3	0.728	101.5	0.799	99.6
4	0.753	98.2	0.829	93.5
5	0.768	96.3	0.847	89.9
6	0.777	95.2	0.856	88.2
7	0.781	94.7	0.854	88.6
8	0.775	95.4	0.837	91.8
9	0.753	98.2	0.806	98.1
10	0.713	103.7	0.777	104.4
11	0.689	107.3	0.802	98.8
12	1.000	74.0	1.000	65.1

[a]Exclusive of 72-Ω generator impedance.

in the two cases. It is possible to synthesize (see Chap. 10) the excitation voltages such that maximum gain is achieved. If this were done, the resulting currents at the fed ports would be of unit magnitude, whereas the voltages needed to establish these unit magnitude currents would generally be of nonunit magnitude.

14.11.2   The Circular Array

Consider the circular array in Fig. 14-26 that is also known as a ring array [H.8.2: Ma, Chap. 3]. Such arrays have been used in radio direction finding, radar, sonar, and in other systems applications. Usually, circular arrays are composed of identical, equally spaced elements as indicated in Fig. 14-26, and each dipole is excited at its center. If we temporarily replace each dipole with a point source at the excited dipole ports, we can write for the array factor (see Sec. 8.1):

$$\text{AF}(\theta, \phi) = \sum_{n=1}^{N} I_n e^{j\alpha_n} e^{j[\beta\rho_n' \sin\theta \cos(\phi - \phi_n)]} \tag{14-97}$$

where I_n is the current excitation of the nth element located at $\phi = \phi_n$, α_n is the associated phase excitation relative to the array center located at the coordinate origin, and ρ_n' is the radial distance of each element center from the origin (all of which equal b for the circular array case). For the usual case of cophasal excitation,

$$\alpha_n = -\beta\rho_n' \sin\theta_o \cos(\phi_o - \phi_n) \tag{14-98}$$

where (θ_o, ϕ_o) is the desired position of the main beam maximum.

For the half-wave dipoles, the element pattern is given approximately by (3-4). Thus, the complete pattern for the circular array of half-wave dipoles with an assumed sinusoidal current distribution can be written as

$$F(\theta, \phi) = \frac{\cos\left(\dfrac{\pi}{2}\cos\theta\right)}{\sin\theta} \frac{\displaystyle\sum_{n=1}^{N} I_n e^{j\alpha_n} e^{j[\beta\rho_n' \sin\theta \cos(\phi - \phi_n)]}}{\displaystyle\sum_{n=1}^{N} I_n} \tag{14-99}$$

— Current generator

— — — Voltage generator (unloaded)

(*a*)

— Current generator

— — — Voltage generator (loaded)

(*b*)
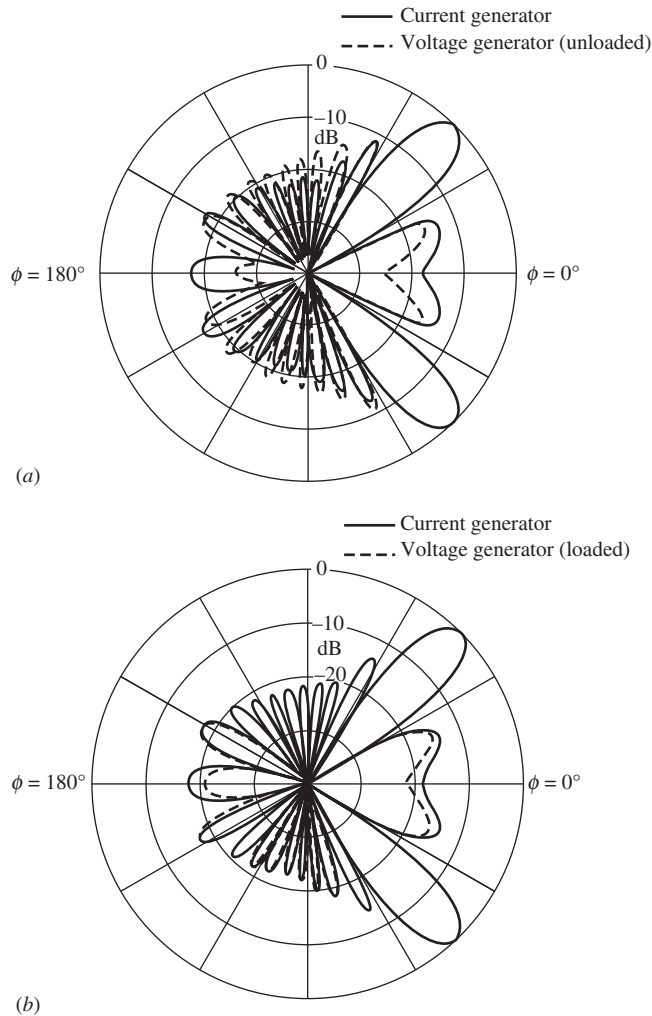
Figure 14-25 Linear array patterns with main beam steered to $\phi_o = 45°$ and ideal current generators (solid curve) compared to patterns from an array with voltage generators (see Table 14-2). (*a*) Linear array pattern for unloaded voltage generator excitations (dashed curve). (*b*) Linear array pattern for 72-Ω loaded voltage generator excitations (dashed curve).

where the assumption is made that (8-100) applies rather than (8-104).

The analysis of the circular array in (14-97) to (14-99) is, of course, based on known currents on the array elements. In practice, we usually apply voltages rather than currents to the array element ports. To determine the currents established by the voltages, we can use MoM, thereby including all mutual effects.

We will consider the circular array in Fig. 14-26 to be composed of identical, equally spaced dipoles. Thus, certain simplifications in the moment method formulation are possible. With the excitation at the centers of all dipoles, it is clear that the current distributions will have even symmetry about the $z = 0$ plane. This image symmetry can be used to compress the size of the impedance matrix $[Z_{mn}]$ of each dipole as discussed in Sec. 14.8.6. (This could also have been done for the linear array in the previous section.) In addition to this, the impedance matrix for the circular array will take the submatrix form
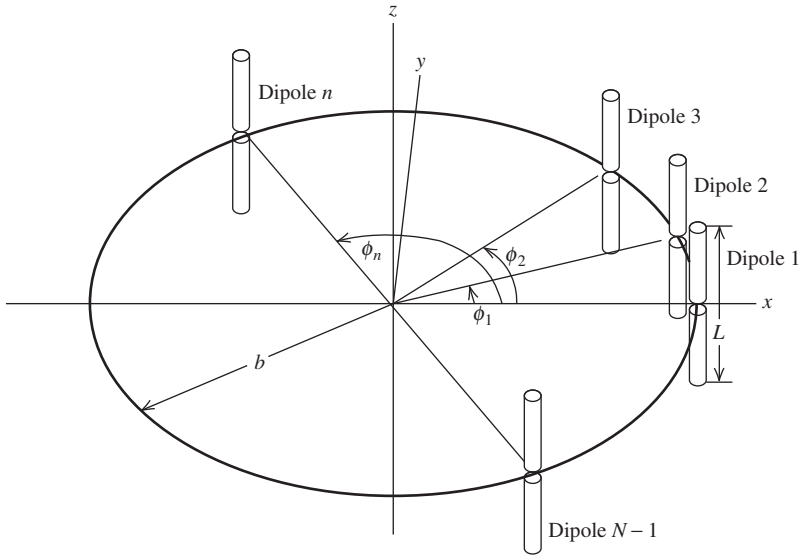
Figure 14-26 Circular array of dipoles.

$$[Z]_{\text{array}} = \begin{bmatrix} [S]_{11} & [S]_{12} & \cdots & [S]_{1N} \\ [S]_{1N} & [S]_{11} & \cdots & [S]_{1(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ [S]_{12} & [S]_{13} & \cdots & [S]_{11} \end{bmatrix} \quad (14\text{-}100)$$

where $[S]_{ij} = [Z_{mn}]$, and each $[Z_{mn}]$ may be compressed as described in Sec. 14.8.6. The matrix in (14-100) is not only toeplitz, but also goes by the name "block circulant." It can be shown that the inverse of a block circulant matrix is also block circulant. Thus, $[Y]_{\text{array}}$ would be block circulant. In a block circulant matrix, successive rows of blocks repeat the previous row but begin with the last block of the previous row.

If we use 1-V voltage generators in series with a 72-Ω impedance at the center of each dipole in a 12-element circular array with $\lambda/2$ spacing, the currents given in Table 14-3

Table 14-3 Normalized Terminal Currents for a Circular Array of 12 Half-Wavelength Spaced, Parallel Half-Wave Dipoles (72-Ω loaded voltage generators)

Element Number	$|I_A|$
1	0.735
2	0.566
3	0.628
4	0.517
5	0.547
6	0.791
7	1.000
8	0.791
9	0.547
10	0.517
11	0.628
12	0.566
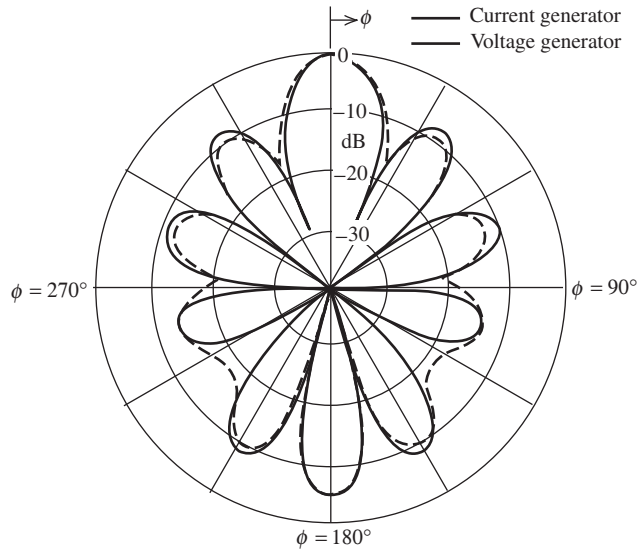
Figure 14-27 Patterns of the circular array of Fig. 14-26 with 12 elements for uniform current excitation (solid curve) and for 72-Ω loaded voltage generators with currents of Table 14-3 (dashed curve).

resulted using phases from (14-98). The almost 2:1 variation in current magnitude is the result of mutual coupling. The corresponding pattern in the azimuthal plane is shown in Fig. 14-27. For purposes of comparison, also shown is the pattern for uniform (current) excitation calculated using (14-97). The difference between the two types of patterns is more noticeable here than in Fig. 14-27 for the linear array. Although the pattern with the voltage generator obtained using the moment method is the more realistic of the two, an advantage of the moment method is that it does yield the input impedance of the elements for any scan angle, thereby providing information for the design of the feed network (see Sec. 8.9).

14.11.3   Two-Dimensional Planar Array of Dipoles

Consider a two-dimensional array of parallel dipoles located in the *xz*-plane as shown in Fig. 14-28. Our purpose here is to use MoM to show how the input impedance of an element in the array varies with scan angle.

Fig. 14-28 shows the input impedance variation of the center element in a 7×9 array (i.e., seven collinear elements in an *E*-plane row by nine parallel elements in an *H*-plane row). Three scanning conditions are illustrated: *H*-plane, *E*-plane, and the 45° plane between the *E*- and *H*-planes. It is clear from Fig. 14-28 that the input impedance does vary considerably with scan angle and the variation depends on the plane of scan. Clearly, this variation poses a challenging design problem for the engineer responsible for designing the array feed and matching network.

Note that as the array is scanned in the *E*-plane (*zy*-plane) to 90° (i.e., "endfire"), the real part of the input impedance is tending toward zero, which in turn means the element is tending not to radiate! Indeed, although the other elements in the array will not have *exactly* the same behavior, most of them (except the edge elements) will behave similarly and the entire array will tend not to radiate! This phenomenon is known as *Wood's anomaly*, or the blindspot phenomenon, and it would seem inappropriate not to mention it in a book on antenna principles. Wood's anomaly is more likely to occur in large arrays
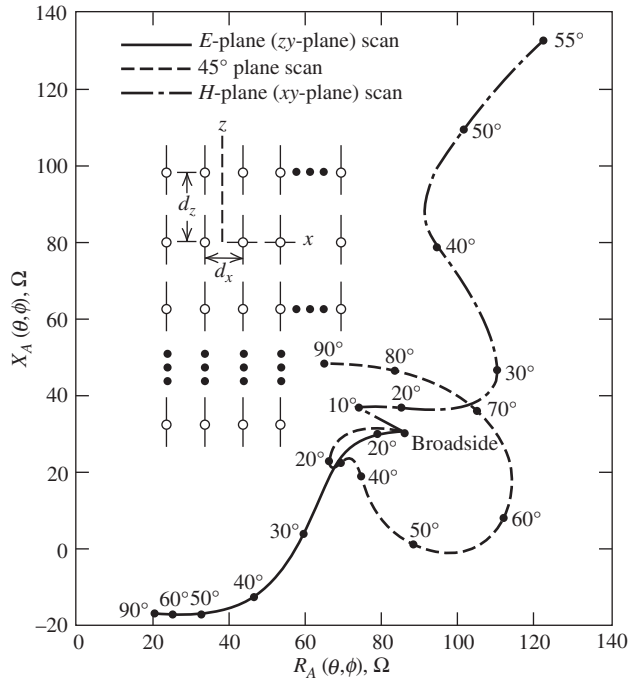
Figure 14-28 Input impedance variation of a central element in a 7×9 dipole array as a function of scan angle for three planes of scan ($d_x = d_z = 0.5\lambda$).

than in the relatively small one considered here. Fortunately, Wood's anomaly can be avoided in some arrays by suitable choices in the array design parameters as well as in the element design itself. (See Sec. 8.9.)

An explanation for Wood's anomaly lies in an understanding of surface waves. In the E-plane, as the array is scanned farther and farther away from broadside, there is a larger and larger component of the electric field normal to the array plane. The metallic elements in the dipole array simulate a metallic surface for the currents that flow in the direction established by the E-plane. Thus, a surface wave is excited in the E-plane direction that can consume the output power of the array if the array is large. This surface wave phenomenon is similar to the creeping wave discussed in Sec. 16.11.

In the H-plane, no such surface wave is possible because at large scan angles, the electric field is trying to propagate tangential to the simulated metallic surface and tends to be shorted out by it. Similarly, the creeping wave in Sec. 16.11 tends not to propagate when the electric field is parallel to a metallic surface.

The moment method analysis of a two-dimensional array such as that in Fig. 14-28 is aided by the block toeplitz nature of the problem. Much larger arrays than that in Fig. 14-28 can be analyzed and designed with the aid of MoM even if the number of unknowns is in the tens of thousands. However, for arrays of such size, other methods of analysis are available [H.8.2: Amitay].

14.11.4　Summary

In this section, we illustrated, through the use of several examples, the application of the moment method to antenna arrays. The examples show us several things. First, the moment method takes into account all mutual couplings and makes it unnecessary to assume the current distribution on the elements in the array or to assume that each

element has the same pattern. Second, the moment method directly provides accurate information concerning the input impedance of various elements under any scan condition. Third, the assumption of a sinusoidal current distribution on a thin half-wave dipole in an array environment is a pretty good one and, therefore, the classical methods of dipole array analysis based on this assumption are quite accurate. It is for elements other than the dipole that the moment method has an obvious additional advantage.

14.12 RADAR CROSS SECTION OF ANTENNAS

The study of antenna scattering is a combination of two electromagnetic disciplines: antennas and scattering. Usually, antenna analysis considers the antenna to be a transmitter, whereas part of the study of antenna scattering requires the antenna to be viewed as a receiver. Even in the receiving case, if we are just interested in the power delivered to a load, we can conveniently use antenna transmitting properties and reciprocity. But if we are also interested in how an antenna scatters energy into surrounding space, then a detailed knowledge of the induced currents on all parts of the antenna structure is required. In general, this is a difficult task, but one that is tractable using MoM.

To begin our discussion, consider the Thevenin equivalent circuit of Fig. 14-29 for an antenna as a function of its load impedance. In this circuit, $Z_A = R_A + jX_A$ is the antenna impedance, $Z_L = R_L + jX_L$ is the load impedance, and V_A is the open-circuit voltage induced at the antenna terminals. V_A can be related to the incident electric field $\mathbf{E}^i$ as in [26] by

$$V_A = -\mathbf{h}_A^r \cdot \mathbf{E}^i \tag{14-101}$$

where $\mathbf{h}_A^r$ is the antenna vector effective length upon receiving, evaluated in the direction of reception. (Note that (14-101) from [26] differs from (4-2).) By reciprocity, this antenna vector effective length is equal to that of the antenna upon transmitting $\mathbf{h}_A^t$, evaluated in the same direction and defined such that the far field radiated by the antenna under unit current excitation is

$$\mathbf{E}^t = -j\frac{\eta}{2\lambda}\mathbf{h}_A^t\frac{e^{-j\beta r}}{r} \tag{14-102}$$

where r is the radial distance from the antenna to the observation point.

The signs associated with (14-101) and (14-102) are such that the positive terminal during reception is the terminal into which positive sense current enters during transmission. The receiving current is then given by

$$I_A(Z_L) = -\frac{V_A}{Z_A + Z_L} \tag{14-103}$$

With this terminology defined, we now state that the field scattered by an antenna as a function of its load impedance is given by
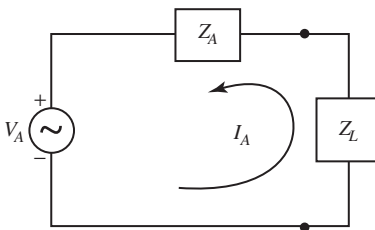
Figure 14-29 Antenna Thevenin equivalent circuit used by Green [26].

$$\mathbf{E}^s(Z_L) = \mathbf{E}^s(0) - \frac{Z_L I(0)}{Z_L + Z_A} \mathbf{E}^t \tag{14-104}$$

where $\mathbf{E}^s(0)$ and $I(0)$ are, respectively, the scattered field and terminal current under short-circuit conditions, $Z_L = 0$. The frequency, polarization, and directions of incidence and reception are assumed to be fixed. This *basic equation of antenna scattering* was derived by King and Harrison [27] using compensation and superposition theorems in circuit analysis, Stevenson [28] using field theory, and Hu [29] using circuit and field theories combined. Aharoni [H.3: Aharoni] provides a textbook discussion. A derivation from the field point of view is also offered by Green [26]. Derivations employing the scattering matrix are given by Collin [H.3: Collin] and Hansen [30]

Eq. (14-104) is not quite in the form we need for our investigation of antenna scattering. To obtain the form used by Green [26], set $Z_L = Z_A^*$, complex conjugate of the antenna impedance, and solve for $\mathbf{E}^s(0)$:

$$\mathbf{E}^s(0) = \mathbf{E}^s(Z_A^*) + \frac{Z_A^* I(0)}{2R_A} \mathbf{E}^t \tag{14-105}$$

The use of (14-103), first with $Z_L = 0$ and then with $Z_L = Z_A^*$, leads to

$$I(0) = \frac{2R_A}{Z_A} I(Z_A^*) \tag{14-106}$$

so that (14-105) becomes

$$\mathbf{E}^s(0) = \mathbf{E}^s(Z_A^*) + \frac{Z_A^*}{Z_A} I(Z_A^*) \mathbf{E}^t \tag{14-107}$$

These last two equations, substituted into (14-104), yield the fundamental equation due to Green:

$$\mathbf{E}^s(Z_L) = \mathbf{E}^s(Z_A^*) + \frac{1}{Z_A} \left[Z_A^* - \frac{2R_A Z_L}{Z_A + Z_L} \right] I(Z_A^*) \mathbf{E}^t \tag{14-108}$$

or

$$\mathbf{E}^s(Z_L) = \mathbf{E}^s(Z_A^*) - [I(Z_A^*) \mathbf{E}^t] \Gamma_m \tag{14-109}$$

where the quantity

$$\Gamma_m \equiv \frac{Z_L - Z_A^*}{Z_L + Z_A} = \frac{\frac{Z_L + jX_A}{R_A} - 1}{\frac{Z_L + jX_A}{R_A} + 1} \tag{14-110}$$

is a modified voltage reflection coefficient in contrast with the usual definition of reflection coefficient.

The quantity $[I(Z_A^*) \mathbf{E}^t] \Gamma_m$ in (14-109) is called the *antenna mode* component of the scattered field because it is completely determined by the radiation properties of the antenna. It vanishes when the antenna is conjugate-matched. This term is related to the energy absorbed in the load of a lossless antenna as well as the energy reradiated by the antenna due to load mismatch. The pattern of the energy scattered in the antenna mode is exactly that of the antenna radiation pattern. The other quantity on the right-hand side of (14-109), $E^s(Z_A^*)$, is called the *structural scattering* or *residual scattering*

(*a*) Plane wave excitation.

(*b*) Antenna mode excitation.

(*c*) Structural mode excitation.
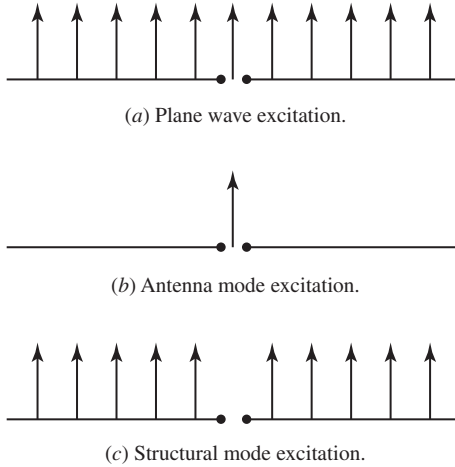
Figure 14-30 Conceptualization of antenna mode and structural mode scattering.

component. It arises from the currents induced on the antenna surface by the incident wave even when the antenna has been conjugate-matched.

Hansen [30] gives an alternative formulation to that of Green in (14-109) in terms of a conventional load match. However, his formulation does not appear to be generally superior to Green's.

A conceptualization of antenna mode and residual mode scattering is possible from the point of view of point-matching [31]. Fig. 14-30 shows the excitation of a dipole by a plane wave [31]. The arrows depict the amplitude of the plane wave at the match points. The (*b*) part of the figure shows the usual antenna excitation that gives rise to the antenna mode term, whereas the (*c*) part shows the difference between the plane wave excitation and the antenna mode excitation that is the residual (structural) mode excitation.

A primary reason for being interested in antenna scattering is to determine the radar cross section of an antenna. In the case of antenna mode scattering, it is possible to derive the following simple expression for the antenna mode component of the radar cross section:

$$\sigma_{\text{ant}} = \Gamma^2 G^2(\theta, \boldsymbol{\phi})\frac{\lambda^2}{4\pi} \tag{14-111}$$

However, no such simple expression is possible for the structural mode radar cross section.

Fig. 14-31 shows the broadside monostatic RCS of a dipole 0.5 m in length terminated in a 72-Ω load as the frequency is varied from 100 to 2000 MHz [31]. We observe peaks in the curve at the first three dipole resonances near 300, 900, and 1500 MHz as expected. Near 300 MHz, the RCS is due to the residual mode since the dipole is conjugate-matched at its first resonance. Above about 1000 MHz, the RCS is also dominated by the residual scattering, but this is because the antenna mode radiation pattern is weak or zero in the broadside direction.

Fig. 14-32 shows the bistatic scattered field for the residual mode and antenna mode when a plane wave is incident at angles of 90, 60, and 30° from the axis of a 1.723λ dipole with a load of 72 Ω. In comparing the curves in Figs. 14-32*a* and 14-32*b*, we first observe that the antenna mode bistatic electric field curves are all symmetric about $\theta_s = 90°$, whereas the structural mode curves for $\theta_s = 60°$ and 30° are not. This is a further illustration of the antenna-like behavior of the antenna mode. The very weak
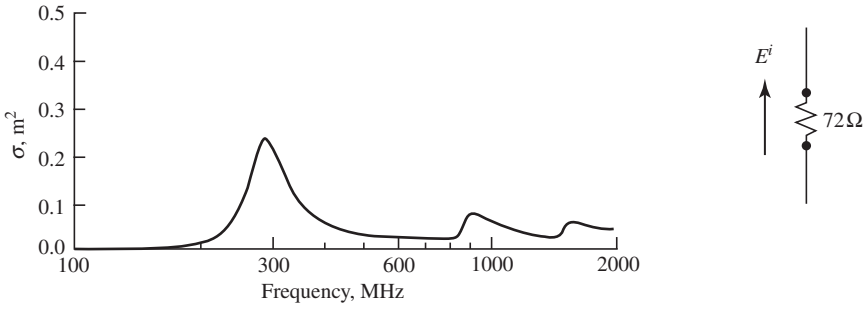
Figure 14-31 Monostatic broadside RCS vs. frequency for a 0.5-meter-long dipole terminated in a 72-Ω load.
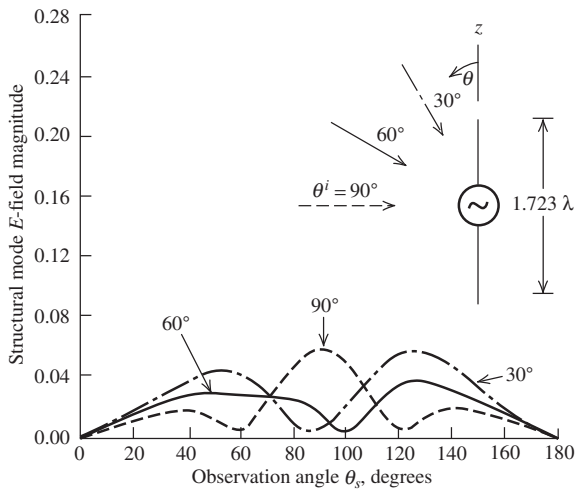
(*a*) Structural mode electric field bistatic scattering.
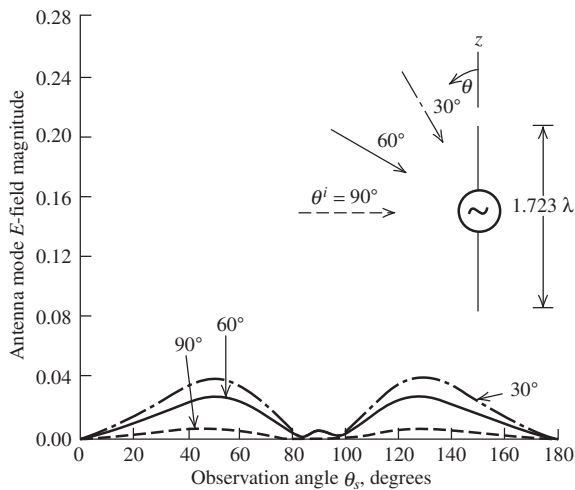
(*b*) Antenna mode electric field bistatic scattering.

Figure 14-32 Component parts of antenna scattering.
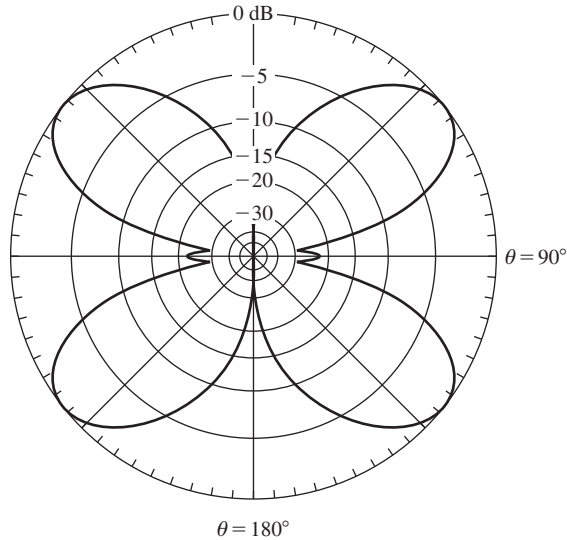
(c) Transmitting E-plane pattern of the 1.723 λ dipole.

Figure 14-32 (Continued)

response of the antenna mode scattering for 90° incidence is due to the small lobe in the antenna transmitting pattern visible in Fig. 14-32c at $\theta = 90°$.

To illustrate the structure-like behavior of the structural mode, note the large bistatic response around for the $\theta_s = 125°$ for the $\theta_i = 30°$ incidence case in Fig. 14-32a. This is due to the dipole structure acting like a moderately long reflecting surface. If the length of the dipole were to increase, the angle of maximum response would increase toward 150° as expected from the law of reflection.

It is worthwhile to comment on how the MoM is used to obtain the data in Fig. 14-32a. First, at each frequency the antenna input impedance is calculated for the dipole so that $E^s(Z_A^*)$ is known. Then the left-hand side of (14-109) is calculated for each of the three incidence angles. Next, $I(Z_A^*)$ is found so that the antenna mode scattering may be calculated. In turn, subtracting the antenna mode scattering from the total scattered field on the left-hand side of (14-109) yields the structural scattering. No other CEM technique would be as helpful as MoM for these kinds of calculations.

Fig. 14-33 shows the **total** power scattered, P_S, and the total power absorbed, P_A, by a 0.5-m-long dipole with a 72-Ω load and a plane wave normally incident. Generally, the total power scattered is greater than the total power absorbed, the two being equal at first resonance where the dipole is conjugate-matched. Below first resonance, the dipole is not conjugate-matched and it is possible for the total power absorbed to exceed the total power scattered. The curves in Fig. 14-33 were calculated using the equivalent circuit in Fig. 14-29 by assuming that Z_A accounts for the total power scattered and Z_L for the total power absorbed by the dipole. That Z_A accounts for the total power scattered by the antenna is *not* generally true, but is approximately true when the power scattered by the open-circuited antenna is much smaller than that scattered by the terminated antenna [32] (below about 400 MHz in Fig. 14-31). Numerically integrating the scattered fields over a far-zone sphere enclosing the dipole to correctly calculate the total power scattered replicates the P_A curve up to a dipole length of about 0.6λ and provides at best only rough quantitative agreement above that length. It is sometimes thought that a matched lossless antenna cannot absorb more power than it scatters. However, this is not true theoretically [33]. In practice, it is unlikely that a matched lossless antenna will absorb more power
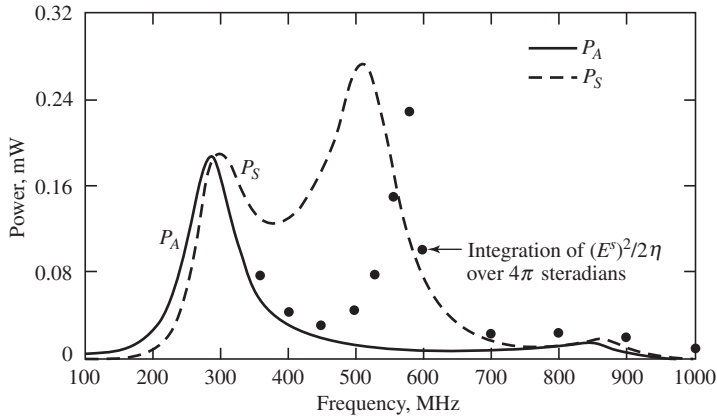
Figure 14-33 Total power absorbed and total power scattered by a 0.5-m-long dipole with a 72-Ω load.

than it scatters. In Fig. 14-33, the powers are equal just below 300 MHz where the dipole is matched. Thus, the dipole exhibits the characteristics of the so called "minimum scattering antenna." A minimum scattering antenna is one that:

1. Scatters as much power as it absorbs for conjugate match
2. Has identical transmitting and receiving patterns
3. Radiates the same fields in the forward and backward directions
4. Is "invisible" when its terminals are open circuited

The half-wave dipole, when open circuited, will consist of two collinear quarter-wave dipoles. A quarter-wave dipole is a very poor, nearly "invisible," scatterer (see Fig. 14-33 around 150 MHz). Small lossy dipoles are even poorer scatterers. [34]

The name minimum scattering antenna implies that an antenna is in a class of antennas with the lowest possible scattering. However, this is not necessarily the case, and the widespread use of the name is unfortunate for that reason [H.8.2: Munk, pg. 35]. Note that total scattering is different from backscattering (e.g., monostatic RCS). An antenna with the potential for zero backscattering is an array of conjugate-matched dipole elements a quarter-wavelength in front of a finite ground plane of a size similar to that of the array. This array basically absorbs all the energy incident upon it [35] [H.8.2: Munk, pg. 33], yet it does not meet the preceding requirements for a minimum scattering antenna!

Generally speaking, a conjugate match usually results in only a 6 dB reduction in the RCS compared to the short-circuit case. For example, the RCS of a half-wave dipole short-circuited at its terminals is $0.86\lambda^2$. If the half-wave dipole is conjugate-matched, its RCS is $0.215\lambda^2$. This is a 6 dB reduction similar to that in Fig. 14-31 just below 300 MHz, the so called "6 dB rule."

In this section, scattering by an antenna has been examined for the purpose of understanding the radar signature of an antenna (see also Example 16-5, Sec. 16.14). However, this is not the only reason to be interested in antenna scattering. For example, if an antenna did not scatter, there would be no mutual coupling in an antenna (receiving) array. Further, antenna scattering is a phenomenon central to many RFID systems.

14.13   MODELING OF SOLID SURFACES

There are two principal ways in which MoM can be used to model either two-dimensional or three-dimensional bodies (e.g., infinite cylinder or finite cylinder, respectively). A simple but older way is to model a solid surface with a grid of wires, the so-called
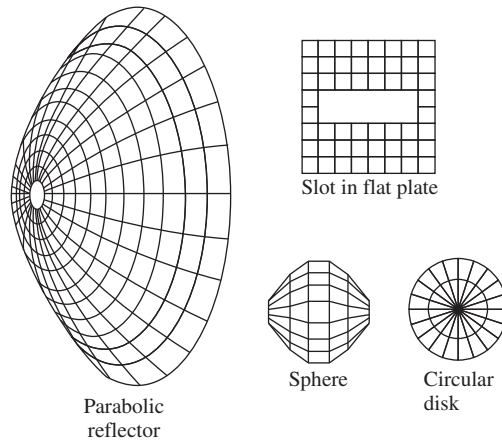
Figure 14-34 Examples of wire-grid modeling.

wire-grid model. Examples are shown in Fig. 14-34. A more modern way of modeling solid surfaces is to use either the electric-field integral equation (EFIE) or the magnetic-field integral equation (MFIE), in which the surfaces are subdivided into patches having a continuous metallic surface. In this section, we briefly examine both approaches.

14.13.1   Wire-Grid Model

In this subsection, we demonstrate the application of the wire segment procedure to model not just a wire antenna or wire scatterer, but also to model the metallic environment near the antenna. We can accomplish this by using a wire-grid or wire mesh to stimulate an actual continuous metallic surface. The idea of using a wire mesh to simulate a continuous metallic surface precedes, of course, the time when the moment method came into widespread use. There are many practical situations where the effect of a continuous metallic surface is required, but the weight and/or wind resistance offered by a continuous surface is too large (e.g., a reflector surface).

The successful substitution of a wire grid for a continuous metallic surface (in reality or in a model) depends on the fact that as the grid size becomes smaller relative to the wavelength, the grid supports a current distribution that approximates of the actual current, however, and as such it can be expected to reasonably predict the far fields but possibly not the near fields. This is due to the fact that the grid supports an evanescent reactive field on both sides of its surface [36]. An actual continuous conducting surface is not capable of supporting such a field.

The accuracy with which a wire-grid model simulates an actual surface depends on the computer code (i.e., expansion and weighting functions) used, the radius of the wire segments used, as well as the grid size. For example, with pulse expansion functions and point-matching, it has been found that a grid spacing of about 0.1 to 0.2λ yields good results [37]. With the piecewise sinusoidal Galerkin method, it has been found that the grid size should not exceed $\lambda/4$ and a suitable wire radius is $a = w/25$, where w denotes the width or length (whichever is greater) of the apertures [37].

Let us now consider the situation where a monopole is axially mounted on the base of a cone as shown in Fig. 14-35. A wire-grid representation can be used in which the cone or frustum is represented by a number of "generating lines" consisting of a number of wires joined end to end, as shown in Fig. 14-36a. Except for the base, no wires need to be provided in planes normal to the z-axis because of excitation symmetry.
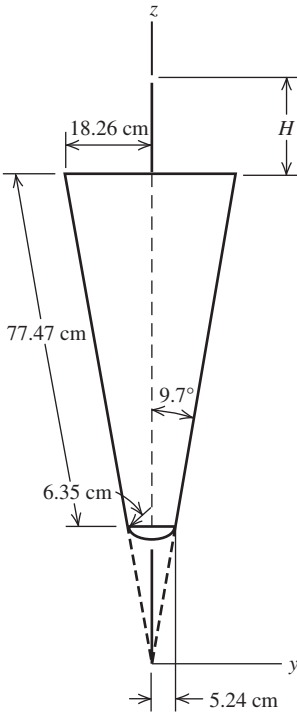
Figure 14-35 Dimensions of experimental cone model. The monopole is a quarter-wavelength at each frequency.

An interesting simplification (see Sec. 14.8.6) can be obtained from the symmetry of the configuration in the case where all generating lines have the same number of segments, each segment being identical (except for the orientation on the ϕ-coordinate) to the corresponding one on each other generating line. The currents on such corresponding segments should be equal in magnitude and phase, since $I(z)$ is independent of ϕ. Let the segments be numbered in a consecutive way, starting with the line at $\phi = 0$ and proceeding in counterclockwise direction along the other lines. Let M be the number of segments on each line and L the number of generating lines. Thus, one can write

$$\sum_{j=1}^{L \cdot M} Z_{kj} I_j = E_k^j, \quad k = 1, 2, \ldots, L \cdot M \tag{14-112}$$

Since the currents on corresponding segments are equal,

$$I_j = I_{(j+M)} = I_{(j+2M)} = I_{(j+(L-1)M)} \tag{14-113}$$

and (14-112) can be written as (see Sec. 14.8.6)

$$\sum_{j=1}^{M} I_j \left(\sum_{n=0}^{L-1} Z_{k(j+nM)} \right) = E_k^i, \quad k = 1, 2, \ldots, M \tag{14-114}$$

The advantage of (14-114) is that it permits us to reduce the number of unknown currents to M, while the actual number of wire segments is $L \cdot M$, where L is arbitrary. As a result, there is little limitation other than computer running time to the number of generating lines (and thus to the total number of segments represented). For the patterns calculated here, L was chosen to be 10, M to be 170, and pulse basis functions were used.

(a) Model of monopole and cone.

(b) Cone generating line showing
distribution of segments and
boundary matching points.
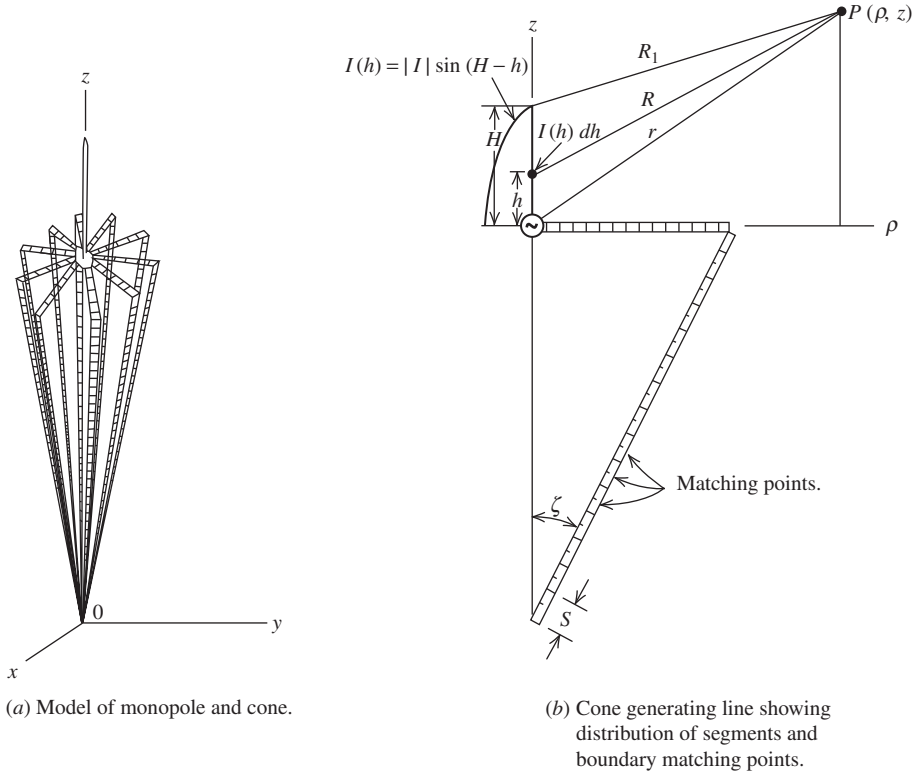
Figure 14-36 Wire-grid model of cone in Fig. 14-35.

The left-hand side of (14-114) represents $-\mathbf{E}^s$ for the cone problem under consideration here. It remains to determine $\mathbf{E}^i$. For the monopole, consider the geometry depicted in Fig. 14-36b. If we start with the vector potential, the following near-field expressions for the monopole configuration of Fig. 14-36b may be derived as

$$E_z^i = -j29.975 \ |I| \left[\frac{e^{-j\beta R_1}}{R_1} - \frac{e^{-j\beta r}}{r}\cos\beta H - j\frac{z}{r^2}e^{-j\beta r}\sin\beta H - \frac{z}{\beta r^3}e^{-j\beta r}\sin\beta H \right] \quad (14\text{-}115)$$

and

$$E_\rho^i = \frac{j29.975 \ |I|}{\rho} \left[(z-H)\frac{e^{-j\beta R_1}}{R_1} - \frac{z}{r}e^{-j\beta r}\cos\beta H - \frac{jz^2}{r^2}e^{-j\beta r}\sin\beta H \right.$$
$$\left. + \frac{\rho^2}{\beta r^3}e^{-j\beta r}\sin\beta H \right] \quad (14\text{-}116)$$

The right-hand side of (14-114), which constitutes the generalized voltage matrix, requires that the tangential components of the incident field be determined at each match point on one of the generating lines of the cone. The equations above for E_z^i and E_ρ^i are used to do this.

Solving for the current on the cone makes it possible to calculate the far-field pattern of the cone-monopole structure by superimposing the fields of the cone and those of the monopole. A necessary but not sufficient check on the validity of the moment method solution in this problem requires that the currents at the junctions of the monopole with

the wire-grid representation of the cone satisfy Kirchhoff's current law. For the formulation in (14-114) to (14-116) with $L = 10$, the current on each of the 10 wire-grid lines was found to be 0.1 A when the monopole base current $|I| = 1.0\,A$. That these 11 currents satisfy Kirchhoff's current law at their common junction is a direct consequence of Maxwell's equations since Kirchhoff's current law was not explicitly built into the system of equations [i.e., a constraining equation was not one of the equations in (14-114)].

To experimentally test the validity of the wire-grid representation of a metallic surface, an actual wire-grid cone was built around a Styrofoam core in a configuration similar to that shown in Fig. 14-36a. An experimental comparison of the solid cone and its wire-grid counterpart is shown in Fig. 14-37a. Some representative results showing both the results calculated for the wire-grid cone and measurements for the solid surface cone are illustrated in Fig. 14-37b. The results are generally quite good.

Other variations of the formulation given here are possible, of course. For example, instead of assuming the current distribution on the monopole, it can be treated as an unknown as are the currents on the metallic body. This could be done in a number of ways. The monopole terminal current value could be constrained to a particular value. This would take into account the interaction between the cone and monopole, but would not conveniently provide for the calculation of impedance. Alternatively, one could use a voltage generator at the base of the monopole such as the magnetic frill current discussed previously. Calculation of the currents on the cone and monopole would account for the cone-monopole interaction and also yield directly the monopole impedance. Note that in either case, the previously described symmetry for the cone due to the symmetrical excitation could still be used to advantage.

(a) Experimental comparison only at 400 MHz with a $\lambda/4$ monopole.

(b) Patterns at 300 MHz using a $\lambda/4$ monopole; wire grid calculation and solid cone measurement.
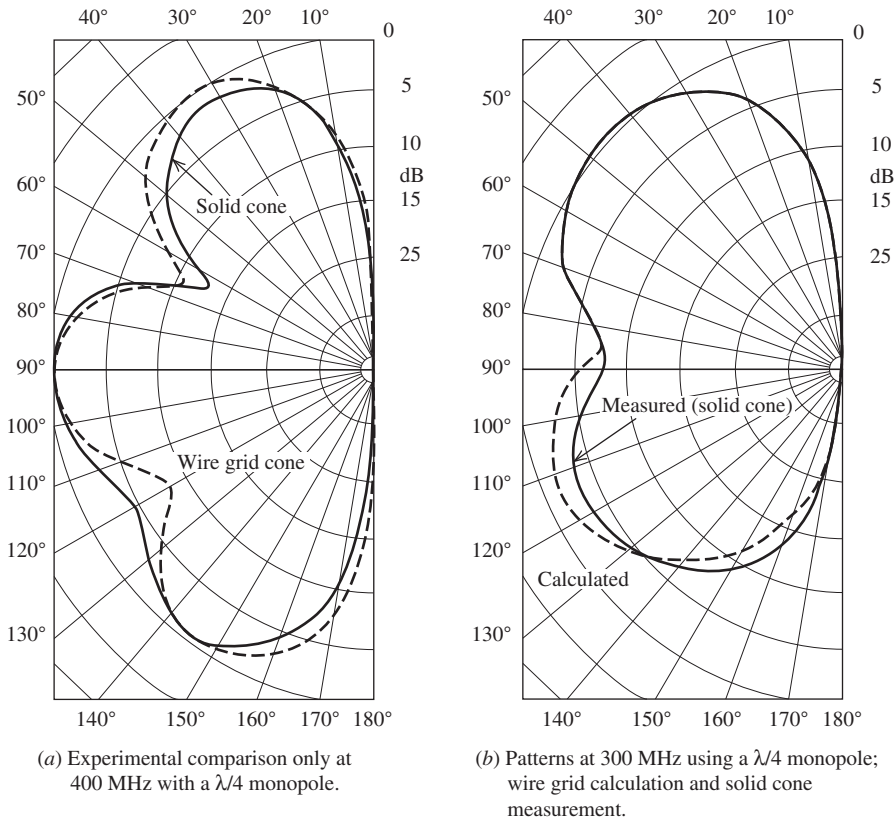
Figure 14-37 Far-field patterns of a monopole on the base of the cone in Fig. 14-36.

The accuracy of wire-grid models can be improved if the grid is reactively loaded with lumped loads [36]. The motivation for doing this to eliminate the effects of the evanescent reactive field that is in proximity to the wire grid. Not only does this increase the accuracy of the model, but it also permits larger grid sizes to be used. Nevertheless, even without this loading, the wire-grid model is a convenient and relatively straightforward tool for engineering calculations.

14.13.2   Continuous Surface Model

Continuous surface models employ surface "patches" with overlapping basis functions (see Fig. 14-17 for an example of overlapping basis functions on a wire), in either an EFIE formulation or one based upon the MFIE. The EFIE has the advantage of being applicable to both open and closed bodies, whereas the MFIE is applicable only to closed surfaces because the MFIE is based on a boundary condition which is not valid for infinitesimally thin geometries. Of the two, the EFIE is generally more difficult to apply due to the kernel of the integral containing a singularity. In closed body problems where the interior dimensions of the body correspond to a waveguide type mode, it is possible to obtain erroneous currents. This difficulty can be overcome by the use of a combined field integral equation formulation, called CFIE, that combines the EFIE and MFIE.

The patch shape of choice for an arbitrarily shaped surface is the planar triangular patch used in Fig. 14-38. Planar triangular patches are capable of accurately conforming to any geometrical surface or boundary within the desired tolerance, are easily specified for computer input, and allow for a varying patch density to accommodate small geometry features and sharp variations in anticipated current density.

If the basis functions representing the surface current are not constructed such that their normal components are continuous across the patch edges, then the continuity equation in (2-11) requires the presence of point or line charges at the edges. These fictitious charges, if present, can cause erroneous solutions in some cases and are to be avoided for that reason. Thus, the basis function of choice for triangular patches is the RWG basis function [39, 40], which avoids difficulties at the patch edges.

The RWG basis function is a subdomain vector basis function defined to be nonzero over two adjacent triangular patches that have a common (interior) edge, the nth edge, between them, as distinguished from a boundary (exterior) edge of the non-closed object. Fig. 14-39 shows two such triangles, T_n^+ and T_n^-, corresponding to the nth edge of an
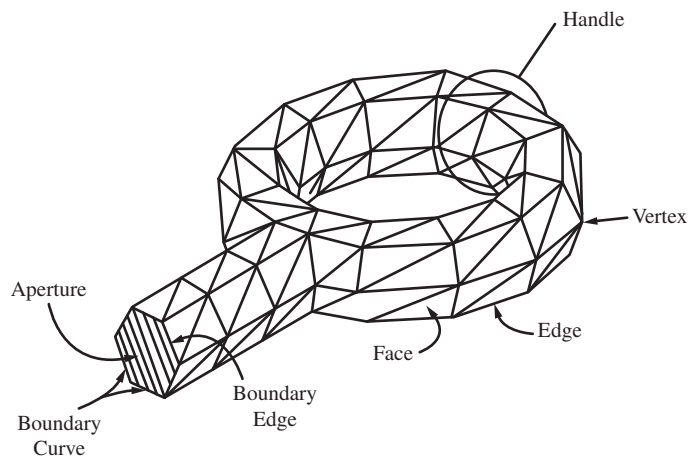
Figure 14-38 Arbitrary surface modeled by triangular patches.
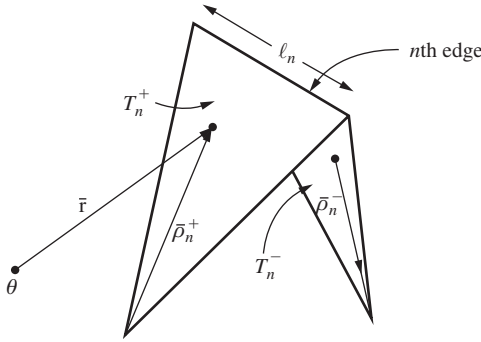
Figure 14-39 Local coordinates associated with an edge.
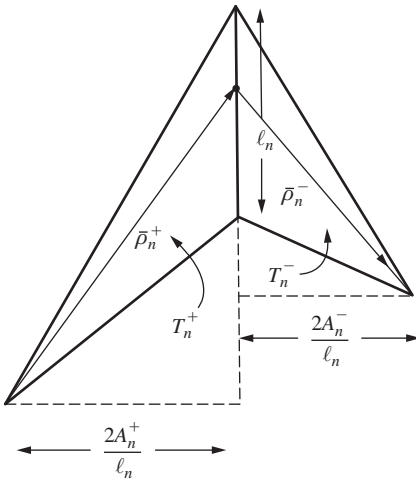
Figure 14-40 Geometry showing normal component of a basis function at an edge.

RWG basis function on a triangulated surface model. The vector basis function associated with the nth edge is defined as

$$
\boldsymbol{f}_n(\boldsymbol{r}) = \begin{cases} \dfrac{l_n}{2A_n^+} \boldsymbol{\rho}_n^+(\boldsymbol{r}), & \text{if } \boldsymbol{r} \text{ in } T_n^+ \\[2ex] \dfrac{l_n}{2A_n^-} \boldsymbol{\rho}_n^-(\boldsymbol{r}), & \text{if } \boldsymbol{r} \text{ in } T_n^- \\[2ex] 0, & \text{otherwise,} \end{cases} \tag{14-117}
$$

where l_n is the length of the edge and A_n^+ is the area of triangle T_n^+, and the position vectors $\boldsymbol{\rho}_n^\pm$ originate at the free vertex of $T_n^\pm$. The subscripts refer to edges and the superscripts refer to faces. These vector basis functions have a current representation free of line or point charges at the boundary of the object and also at all other sides of T_n^+ and T_n^- [39, 40]. The basis functions $\boldsymbol{f}_n$ provide for an adequate approximation to the surface current. Note that the $\boldsymbol{\rho}_n^+$ and $\boldsymbol{\rho}_n^-$ are not unit vectors and therefore their magnitudes produce the "rooftop" nature of the basis function shown in Fig. 14-40. Other examples of overlapping basis functions may be found in Sec. 14.8.1. The component of the current $\boldsymbol{f}_n$ normal to the nth edge is constructed to be continuous across the edge, as may be seen from Fig. 14-40 by normalizing the normal component of $\boldsymbol{\rho}_n^\pm$ along edge n, which is just

the height of triangle $T_n^{\pm}$ with edge n as the base and the height expressed as $(2A_n^{\pm})/l_n$. The RWG basis function in an EFIE formulation is capable of representing rapidly varying current densities such as the edge condition on a flat plate as shown in Fig. 14-41 for a 1.0λ square flat plate [39, 40].

The scope of this text does not permit a detailed discussion of MoM modeling with RWG vector basis functions, but as an example of the use of the RWG vector basis function, the model size, and the model complexity, consider the monopole on a large finite ground plane 30.48 cm by 30.48 cm shown in Fig. 14-42. The monopole on the finite ground plane is one of 12 antenna benchmarks established by the Electromagnetic Code Consortium (EMCC).

The ground plane in Fig. 14-42 is 3.2 mm thick. The monopole is off-center, located 5.33 cm from the nearest edge and 11.43 cm from the next nearest edge. The monopole is 1.47 cm tall and has a first resonance at about 4.5 GHz, where $\lambda/4$ is 1.65 cm. The RWG (surface) basis functions are used on the ground plane and the curved surface of the monopole itself (see Fig. 14-43). The dielectric material, $\varepsilon = 2.0$, $\tan\delta = 0.0002$, in the coax at the base of the monopole is also modeled by RWG basis functions.
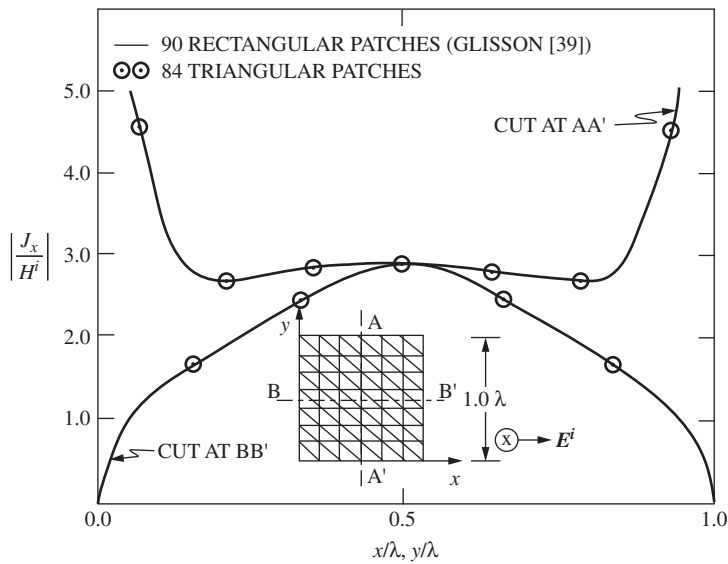
Figure 14-41 Distribution of current on a 1.0λ square flat plate for normal incidence [39].
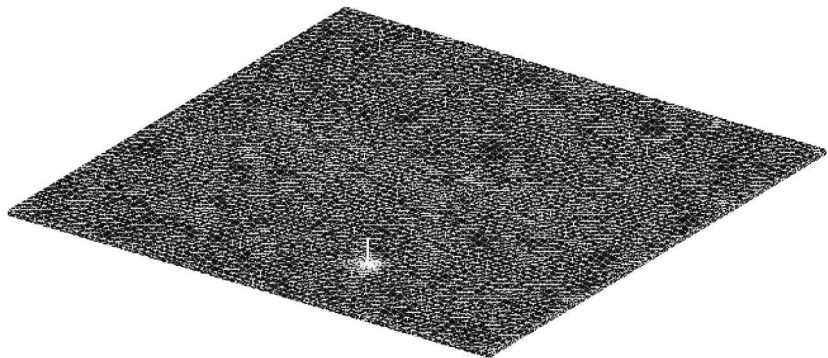
Figure 14-42 Non-uniform mesh of the monopole positioned asymmetrically on the finite ground plane. (Courtesy of BerrieHill Research Corporation.)
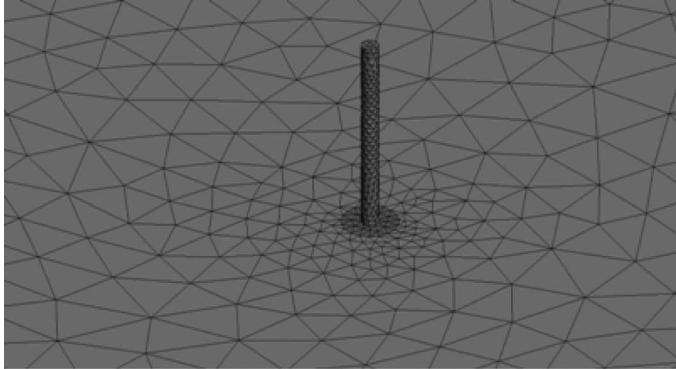
Figure 14-43 Close-up view of the mesh near and on the monopole. (Courtesy of BerrieHill Research Corporation.)

A non-uniform mesh is generated with higher density near the monopole to accurately simulate the fast-varying surface currents there. As a rule of thumb for the numerical convergence, the edge lengths of each triangle are restricted to less than $\lambda/10$ at the (effective) wavelength of interest. In addition, small geometric features whose dimensions are less than a wavelength require a large amount of small mesh elements for accurate geometric representations. Since the first resonance is at approximately 4.5 GHz, most triangles have edges less than 0.5 cm.

For the geometry in Figs. 14-42 and 14-43, there are 20,176 triangles, 20,126 metallic, and 50 dielectric. For a closed surface, the number of edges is 1.5 times the number of triangles, since there are three edges per triangle and each edge is counted twice. A general surface has junction edges which are shared by more than two triangles or exterior edges which is shared by only one triangle, that is, boundary edges of an infinitely thin plate. If an edge belongs to a PEC triangle, there is only one unknown (J) associated with it. If an edge belongs to a dielectric triangle, there are two unknowns (J and M) per edge. Because the RWG basis functions are edge-based, the number of unknown coefficients are associated with the number of edges.

The resulting EFIE matrix for the monopole on the large ground plane was 32,524 by 32,524. The code used was by FEKO [41], a commercially available code. The computing hardware used to generate very accurate input impedance data consisted of 4 physical CPU's operating at 2.4 GHz with 128 GB of RAM. Each CPU has 4 processors for a total of 16. The sum of CPU times for all processors is 6.725 hours per frequency corresponding to a run time of 0.42 hours or about 25 minutes. The complex-valued matrix equation is solved directly by standard LU factorization.

To gain a perspective on the electrical size of the simulation, at the monopole first resonance the wavelength is about 6.67 cm. Thus the ground plane is about 4.57λ by 4.57λ or 20.9 square wavelengths. This accounts for most of the unknowns, the remainder being associated with the metallic monopole and the dielectric in the coax at the base of the monopole.

The monopole itself can also be modeled as a 1D wire or a 2D planar strip to reduce the number of unknowns. However, for accurate calculation of the input impedance of the installed antenna over a wide band of frequencies, both the circular monopole and coax were included in the detailed computational model used here.

As a second example of the use of RWG basis functions, consider the cavity-backed two-arm Archimedian spiral shown in Fig. 14-44. The spiral arms, 0.09 cm wide, are tapered at the end and are on a thin dielectric substrate of radius 7.37 cm and thickness 0.15 cm, with $\varepsilon_r = 3.38 - j0.008$. The inner radius of the spiral is $r_1 = 0.198$ cm and outer radius $r_2 = 4.75$ cm. The number of spiral turns is 12. There are two layers of lossy foam in the cavity. The top layer of foam is 1.87 cm thick with $\varepsilon_r = 1.09 - j0.04$, and the bottom layer
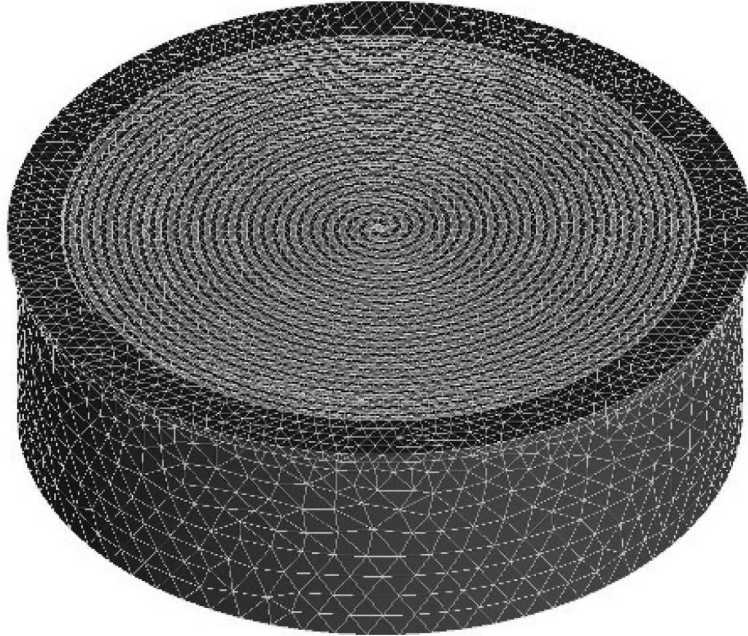
Figure 14-44 Cavity-backed two-arm spiral antenna with two layers of lossy foam inside. (Courtesy of BerrieHill Research Corporation.)

of foam is 1.47 cm thick with $\varepsilon_r = 1.1 - j0.09$. The lowest operating frequency of the cavity-backed spiral is about 1 GHz, and the highest operating frequency is about 24 GHz.

In the MoM simulation model there are 16,208 triangles, 6751 metallic and 9457 dielectric. This corresponds to 16,307 PEC edges, 16,357 dielectric (electric) edges and 12,216 dielectric (magnetic) edges. This results in a 44,880 by 44,880 matrix to be solved for 44,880 unknowns. The run time using the FEKO code with 16 processors was 0.75 hours or 45 minutes per frequency to generate the current distribution on the spiral and far field patterns for both polarizations.

To put this problem in electrical perspective, at the lowest frequency of operation, 1 GHz, the wavelength is 30 cm. Thus, the maximum circumference of the spiral arms is about 30 cm and the diameter is about 10 cm. Modeling the narrow spiral arms of 12 turns each and its substrate requires the bulk of the triangles, edges, and unknowns. The high curvature of the spiral arms near the center of the antenna can only be accurately captured with small triangles. Hence, an appropriate discretization for electromagnetic simulations must satisfy the convergence requirement (e.g., edge length $< \lambda/10$, where λ is the effective wavelength) and resolve the important geometric details such as the surface curvatures.

14.14   SUMMARY

In this chapter, we presented a very useful and powerful technique, the method of moments, for the analysis of certain types of antennas (e.g., wire antennas) and arrays of antennas (e.g., Sec. 14.10 through 14.12). Although the method has been applied primarily to z-directed wires, we have seen that it applies equally well to arbitrary configurations of wires, for example, Sec. 14.11, as well as to solid surfaces, for example, Sec. 14.13.2. Furthermore the method of moments has been used to generate some of the data presented in Chaps. 6 and 7.

The method of moments is often thought of as a low-frequency technique because it generally cannot be applied to bodies that are arbitrarily large in terms of the wavelength (e.g., Sec. 14.8). In contrast to this, in Chapter 16 we will study high-frequency techniques that apply best to bodies which are arbitrarily large in terms of the wavelength. (See Figs. 14-1 and 14-2.)

REFERENCES

1. E. K. Miller, "A Selective Survey of Computational Electromagnetics," *IEEE Trans. on Ant. & Prop.*, Vol. 36, pp. 1281–1305, Sept. 1988.
2. J. H. Richmond, "Digital Computer Solutions of the Rigorous Equations for Scattering Problems," *Proc. IEEE*, Vol. 53, Aug. 1965.
3. R. F. Harrington, "Matrix Methods for Field Problems," *Proc. IEEE*, Vol. 55, Feb. 1967.
4. L. L. Tsai, "A Numerical Solution for the Near and Far Fields of an Annular Ring of Magnetic Current," *IEEE Trans. on Ant. & Prop.*, Vol. AP-20, Sept. 1972.
5. B. A. Finlayson, *The Method of Weighted Residuals and Variational Principles*, Academic Press, New York, 1972.
6. E. K. Miller and F. J. Deadrick, Some Computational Aspects of Thin Wire Modeling," in *Numerical and Asymptotic Techniques in Electromagnetics*, Springer-Verlag, New York, 1975 Chap. 4.
7. V. H. Rumsey, "The Reaction Concept in Electromagnetic Theory," *Phys. Rev.*, Series 2, Vol. 94, June 1954.
8. Mininec Professional Software, EM Scientific, Inc., Carson City, NV, 1995.
9. J. H. Richmond, "Computer Program for Thin-Wire Structures in a Homogeneous Conducting Medium," National Technical Information Service, Springfield, VA, NASA Contractor Rep. CR-2399, July 1973.
10. G. J. Burke and A. J. Poggio, "Numerical Electromagnetics Code (NEC)—Method of Moments," Tech. Doc. No. 116, prepared for Naval Oceans Systems Center, San Diego, CA, NOSC/TD 116, revised Jan. 1980.
11. C. A. Klein and R. Mittra, "The Effect of Different Testing Functions in the Moment Method Solution of Thin Wire Antenna Problems," *IEEE Trans. on Ant. & Prop.*, Vol. AP-23, p. 258, March 1975.
12. R. Bancroft, *Understanding Electromagnetic Scattering Using the Moment Method*, Artech House, Boston 1996.
13. S. M. Rao, D. R. Wilton, and A. W. Glisson, "Electromagnetic Scattering by Surfaces of Arbitrary Shapes," *IEEE Trans. on Ant. & Prop.*, Vol. 30, pp. 409–418, May 1982.
14. V. Rokhlin, "Rapid Solution of Integral Equations of Scattering Theory in Two Dimensions," *J. Comput. Phys.*, Vol. 86, No. 2, pp. 414–439, 1990.
15. L. Greengard and V. Rokhlin, "A Fast Algorithm for Particle Simulations," *J. Comput. Phys.*, Vol. 73, pp. 325–348, 1987.
16. R. Coifman, V. Rokhlin, and S. Wandzura, "Fast Multipole Method for the Wave Equation: A Pedestrian Prescription, *IEEE Ant. & Prop. Magazine*, Vol. 35, No. 4, pp. 7–12, 1993.
17. E. Bleszynski, M. Bleszynski and T. Jaroszewicz, "AIM: Adaptive Integral Method for Solving Large-Scale Electromagnetic Scattering and Radiation Problems, *Radio Sci.*, Vol. 31, No. 5, pp. 1225–1251, 1996.
18. F. X. Canning, "Solution of IML Form of Moment Method Problems in Five Iterations," *Radio Sci.*, Vol. 30, No. 5, pp. 1371–1384, Sept./Oct. 1995.
19. M. A. Epton, B. Dembart, "Multipole Translation Theory for the Three-Dimensional Laplace and Helmholtz Equations," *SIAM Journal on Scientific Computing*, Vol. 16, Issue 4, July 1995.
20. B. Dembart and E. Yip, "The Accuracy of Fast Multipole Methods for Maxwell's Equations," *IEEE Computational Science & Engineering*, Vol. 5, Issue 3, July 1998.
21. J. M. Song and W. C. Chew, "Multilevel Fast Multipole Algorithm for Solving Combined Field Integral Equation of Electromagnetic Scattering," *Microwave and Optical Technology Letters*, Vol. 10, 1995.
22. J. M. Song, C. C. Lu, and W. C. Chew, "Multilevel Fast Multipole Algorithm for Electromagnetic Scattering by Large Complex Objects," *IEEE Trans. on Ant. & Prop.*, Vol. 45, No. 10, October 1997.
23. J. Shaeffer, "Direct Solve of Electrically Large Integral Equations for Problem Sizes to 1 M Unknowns," *IEEE Trans. on Ant. & Prop.*, Vol. 56, No. 8, August 2008.

24. R. Carrel, "Analysis and Design of the Log-Periodic Dipole Antenna," Antenna Lab., University of Illinois Tech. Rep. No. 52, April 1961.

25. R. H. Kyle, "Mutual Coupling between Log-Periodic Antennas," *IEEE Trans. on Ant. & Prop.*, Vol. AP-18, pp. 15–22, Jan. 1970.

26. R. B. Green, "The General Theory of Antenna Scattering," Rep. No. 1223–17, ElectroScience Lab., Columbus, OH, Nov. 1963; Ph.D. dissertation, Dept. Elec. Engr., The Ohio State Univ., 1963.

27. R. W. P. King and C. W. Harrison, Jr., "The Receiving Antenna," *Proc. IRE*, Vol. 32, pp. 18–49, Jan. 1944.

28. A. F. Stevenson, "Relations between the Transmitting and Receiving Properties of Antennas," *Quart. Appl. Math.*, pp. 369–384, Jan. 1948.

29. Y. Y. Hu, "Back-scattering Cross-section of a Center-Loaded Antenna," *IRE Trans. on Ant. & Prop.*, Vol. AP-6, No. 1, pp. 140–148, Jan. 1958.

30. R. C. Hansen, "Relationships between Antennas as Scatterers and as Radiators," *Proc. IEEE*, Vol. 77, pp. 659–662, May 1989.

31. G. A. Thiele and D. D. Richwine, "Antenna Mode and Residual Mode Scattering by a Dipole Antenna," paper presented at 1989 URSI EM Theory Symposium, Stockholm.

32. R. F. Harrington, "Electromagnetic Scattering by Antennas," *IEEE Trans. on Ant. & Prop.*, Vol. AP-11, No. 5, pp. 595–596, Sept. 1963.

33. M. G. Andresen, "Airborne Jamming Antenna Study—Scattering and Absorption by a Receiving Antenna," Stanford Res. Institute, April 1960, DTIC 236342.

34. R. J. Garbacz and G. A. Thiele, "Scattering and Absorption by Very Thin, Short Metallic Filaments," *Radio Science*, Vol. 3, pp. 1045–1050, Nov. 1968.

35. D. C. Jenn, "Radar and Laser Cross Section Engineering," AIAA Education Series, American Institute of Aeronautics and Astronautics, Inc., Wash., D.C., 1995, pp. 251–253.

36. K. S. H. Lee, L. Marin, and J. P. Castillo, "Limitations of Wire-Grid Modeling of a Closed Surface," *IEEE Trans. Electromag. Compatibility*, Vol. EMC-18, Aug. 1976.

37. J. H. Richmond, "A Wire-grid Model for Scattering by Conducting Bodies," *IEEE Trans. on Ant. & Prop.*, Vol. AP-14, Nov. 1966.

38. M. G. Andresen, "Comments on Scattering by Conducting Rectangular Cylinders," *IEEE Trans. on Ant. & Prop.*, Vol. AP-12, pp. 235–236, March 1964.

39. S. M. Rao, D. R. Wilton, and A. W. Glisson, "Electromagnetic scattering by surfaces of arbitrary shape," *IEEE Trans. on Ant. & Prop.*, Vol. 30, pp. 409–418, May 1982.

40. D. R. Wilton, S. S. H. Rao, and A. W. Glisson, "Electromagnetic Scattering by Surfaces of Arbitrary Shape," Report RADC TR-79-325 (AD A084347), March 1980.

41. FEKO code: www.feko.info/product-detail/numerical_methods.

PROBLEMS

14.3-1 (a) Use the equivalence principle to show that the current flowing on the highly conducting wire in Fig. 14-3*a* may be replaced by the equivalent currents radiating in free space as in Fig. 14-3*b* (i.e., the wire material is replaced by free space with zero fields and zero sources inside the original wire volume) and that the equivalent currents (Fig. 14-3*b*) are the same as the currents in the original problem.

   (b) Why is this important?

14.3-2 Show that the left-hand side of (14-13) may be expressed as

$$-E_z^s = \frac{-1}{4\pi j\omega\varepsilon_o} \int_{-L/2}^{L/2} I(z') \frac{e^{-j\beta R}}{R^5} \left[1 + j\beta R)(2R^2 - 3a^2) + \beta^2 a^2 R^2\right] dz'$$

14.3-3 Through integration by parts, show that the left-hand side of (14-13) may be written as

$$E_z^s = +\int_{-L/2}^{L/2} \left[j\omega\mu_o I(z') - \frac{1}{j\omega\varepsilon_o} \frac{\partial I(z')}{\partial(z')} \frac{\partial}{\partial z} \right] \frac{e^{-j\beta R}}{4\pi R} dz'$$

This equation maybe derived by using both the vector and scalar potentials [3].

14.4-1 (a) What are the units of the generalized voltage, current, and impedance matrix elements in (14-26)?

(b) If both sides of (14-26) are multiplied by the segment length, Δz, what are the units of the matrix elements in (14-26)?

14.4-2 Moment methods involve solving systems of linear equations. A computer technique for solving such problems is needed. This could be through a high-level language (MathCAD, MatLab, Mathematica, ...), a subroutine you write, or a canned routine such as IMSL.

(a) For the problem

$$V = ZI \quad \text{where } Z = \begin{bmatrix} 1 & j \\ j & 1 \end{bmatrix}$$

obtain the solution by hand using $I = Z^{-1}V$ if

$$V = \begin{bmatrix} 2 & + & j \\ 1 & + & 2j \end{bmatrix}$$

(b) Solve the problem in (a) using your chosen computer approach. Include details of the approach. Compare to your hand solution.

14.4-3 An important ingredient in moment methods is to be able to numerically integrate complex-valued functions. Select a computer approach and document it.

(a) Integrate the following analytically:

$$\int_0^\pi e^{jx} dx$$

(b) Use your computer approach of choice to evaluate the integral in (a) and compare to results you obtained in (a).

14.5-1 In order to obtain some feeling for MoM, it is recommended that the student write a computer program to solve the following problem. Consider a straight dipole of length L (or monopole of length $L/2$) and radius a. Divide the dipole in N segments of equal length, each containing a pulse expansion function.

(a) Use point-matching and the equation in Prob. 14.3-2 for the scattered field to compute the elements in the first row of the impedance matrix $[Z_{mn}]$ as given in (14-26), noting that these are the only independence matrix elements since the matrix is toeplitz (see Sec. 14.8.4). Note that the integrand tends toward singularity when $R = a$, but even so one may numerically integrate through this region if reasonable care is taken.

(b) Confirm the matrices in Example 14-1. Next, duplicate the curves in Fig. 14-9. (This exercise continues in Probs. 14.6-1 and 14.9-1.)

14.5-2 Starting with the electric vector potential and (14-30), derive (14-31).

14.6-1 (a) Extend the computer code of Prob. 14.5-1 to use pulse weighting functions in (14-40).

(b) Confirm the matrices in Example 14-2 and duplicate the curves in Fig. 14-13. (This exercise continues in Prob. 14.9-1.)

14.7-1 Show that (14-47) to (14-48) follow from (14-44) and (14-45), respectively.

14.7-2 Compare (14-57) to (14-37).

14.7-3 From (14-59), derive (14-26) if delta weighting functions are used.

14.8-1 Use both vector and scalar potentials to derive (14-64).

14.8-2 Another equation for the treatment of wire antennas is Hallen's integral equation:

$$\int_{-L/2}^{L/2} I(z') \frac{e^{-j\beta R}}{4\pi R} dz' = -\frac{j}{\eta}(C_1 \cos \beta z + C_2 \sin \beta |z|)$$

where C_1 and C_2 are constants. The constants C_2 may be evaluated as $V_A/2$, where V_A is the terminal voltage of the antenna. Derive Hallen's equation for the dipole by writing a solution to the wave equation for A_z that is proportional to the right-hand side of the preceding equation

and then equating this result to the integral form of the vector potential for A_z due to a perfectly conducting thin wire dipole.

14.8-3 In Secs. 14.4 and 14.6, we used pulse functions in the moment method. Expansion functions such as the pulse function, piecewise sinusoidal function, etc., are often called subdomain expansion functions because each expansion is generally nonzero on only a relatively small part of the radiating body. (The concept of domain relates to Sec. 14.7.2.)

There is another type of expansion function called entire-domain expansion functions. In this case, the function is generally nonzero over the entire radiating body and the concept of segments is not used. For example, if one were to treat the dipole with an entire-domain expansion function (i.e., a Fourier series), one could write for the current:

$$I(z') = \sum_{n=1}^{N} I_n F_n(z')$$

where

$$F_n(z') = \cos(2n - 1)\frac{\pi z'}{L}, \quad -\frac{L}{2} \le z' \le \frac{L}{2}$$

[Note that each term in $F_n(z')$ goes to zero at the ends of the dipole.]
  (a) Sketch the first two terms in the series for $F_n(z')$.
  (b) If there are N terms and N match points (i.e., a point-matching solution), write an expression for Z_{mn}, using the notation in Secs. 14.4 and 14.6.
  (c) Give a physical interpretation of Z_{25} (i.e., complete a statement similar to the following: Z_{25} represents the field from _____ at _____).

14.8-4 Verify (14-70) using the algorithm in (14-71).

14.9-1 (a) After successfully completing Probs. 14.5-1 and 14.6-1, use (14-80) to compute the far-field radiation pattern of dipole antennas with lengths of 0.1, 0.5, 1.0, 1.25, and 1.5λ. Justify the value of N that you use in each case.

(b) For the dipole lengths in part (a), plot the current distributions on the dipoles in magnitude and phase. Compare with the assumed sinusoidal distribution used in Sec. 6.1.

(c) Using (14-81), compute the gain at broadside for the dipole lengths in part (a).

(d) Consider a plane wave to be incident on a dipole short-circuited at its terminals. Use the relationship

$$E_z^i(z_m) = e^{j\beta z \cos \theta^i}$$

to compute $[V_m]$ for $\theta^i = 90°$ (i.e., the broadside case) and then compute the radar cross section, as in (14-82), when $L = \lambda/2$. Compare with Fig. 14-21 and verify several more points in Fig. 14-21.

14.9-2 Derive (14-80) by considering the dipole of length L to be compared of N ideal collinear dipoles of length L/N.

14.9-3 Derive (14-81), starting with the Poynting theorem.

14.9-4 (a) Using an available MoM code, compute, plot, and label the patterns (three planes) in polar, linear form and the current distribution for a one-wavelength loop. Compare gain to that expected.

(b) Find the input impedance for a loop of radius $a = 0.001$ wavelength for perimeter $= 1$, 1.5, and 2 wavelengths. Compare to Fig. 6-68.

14.10-1 Consider the LPDA in Fig. 14-23 to have only two dipoles. Assume each dipole is composed of three piecewise sinusoidal expansion functions numbered consecutively with the first three piecewise sinusoids on one dipole and the remaining three on the other. Thus, $[Z_{mn}]$ will be of order six, with the second and fifth piecewise sinusoids being at the centers of their respective dipoles. In accordance with (14-91) and (14-93), show that the elements of $[Y_{mn}]$ which form $[Y_A]$ are Y_{22}, Y_{25}, Y_{52}, Y_{55}.

14.10-2 (a) Show that the admittance matrix for one section of the transmission line in Fig. 14-23 without dipoles attached is

$$[Y] = \begin{bmatrix} -jY_o \cot \beta d & +jY_o \csc \beta d \\ +jY_o \csc \beta d & -jY_o \cot \beta d \end{bmatrix}$$

where d is the length of one section of transmission line with propagation constant β and characteristic admittance Y_o.

(b) Show that connecting N-1 of these sections using the scheme in Fig. 14-23 results in (14-94).

14.10-3 Extend the LPDA analysis in Sec. 14.10.2 to an array of M LPDA antennas as in [25].

14.10-4 In Sec. 14.10.2, we obtained a solution to the LPDA. One of the important points in that solution is the determination of $[Y_A]$ in the manner indicated in Eqs. (14-90) to (14-93). Had we wished to then find $[Z_A]$, we could have obtained it from $[Z_A] = [Y_A]^{-1}$. Denote this method A. Suppose instead we find $[Z_A]$ by considering the two dipole mutual impedance problem as Carrel [24] and Kyle [25] did. For example, $[Z_A]_{mn}$ is obtained by temporarily removing all dipoles except m and n from the system and then calculating $[Z_A]_{mn}$. Denote this method B.

(a) Will $[Z_A]$ obtained by method A be the same as that obtained by method B? Why?
(b) The following question refers to the concepts implied by part (a). When we calculate a moment method impedance matrix $[Z_{mn}]$, in what way does that calculation process relate to method B above?

14.11-1 Show that (14-100) is valid.

14.12-1 Assuming that Z_A in Fig. 14-29 can account for the power scattered by a dipole when $0 < L < 0.6\lambda$, and the load is $72\,\Omega$ purely resistive, find a relationship between R_A and R_L when:

(a) More power is scattered (totally) by the dipole than is absorbed by the load.
(b) The powers scattered and absorbed are equal.
(c) The power absorbed by the load is greater than that scattered.
(d) When the scattered and absorbed powers are equal at first resonance, what is the contribution by the antenna mode to the total RCS of the dipole?

14.12-2 Derive (14-111) in either of the following ways:

(a) With a plane wave incident, define an absorption aperture of an antenna as $G\lambda^2/4\pi$, find the absorbed and scattered powers, and apply the definition of radar cross section.
(b) Start with the antenna mode term in (14-109) and use the relationship that the maximum effective aperture for reradiation is $A_{re} = (h_A')^2 \eta/(4R_A)$, when $Z_A = Z_L$, and obtain an expression for the field scattered by the antenna mode, $E_{ant}^s = E^i A_{re}(-j/\lambda)\Gamma(e^{-j\beta r}/r)$, before applying the radar cross-section definition.

14.12-3 In [32] an expression for the RCS of small antennas, such as the dipole, is presented that is valid when the scattered field of the open-circuited antenna is small compared to that of the terminated antenna:

$$\sigma/\lambda^2 \approx (1/\pi)|GR_A/(Z_A + Z_L)|^2$$

(a) Use this expression to compute the RCS of the dipole used in Fig. 14-31.
(b) Also use this expression to compute the RSC of the dipole at first antiresonance (i.e., about 600 MHz). Make an idealized assumption about the input impedance at first antiresonance. Compare your result to Fig. 14-31. Explain the difference.

14.12-4 A sheet of very thin conducting material (thickness, $t \lll \lambda$) has a plane wave normally incident upon it. The resistance R of the thin material is $1/\sigma t$ where σ is the conductivity of the material, and R is measured in ohms/square.

(a) Find the optimum resistance of the thin material such that the sheet absorbs the maximum amount of power from the wave and that this amount is 50% of the power incident.
(b) Comment on the similarity of the maximum power absorbed in (a) to that absorbed by a matched resonant antenna at first resonance (e.g., a $\lambda/2$ dipole).
(c) Do the answers to (a) and (b) imply that any antenna may not absorb more power than it scatters [26,33]?

14.13-1 Derive (14-114) from (14-112).

14.13-2 Sketch a wire-grid model for a square plate $1\lambda \times 1\lambda$. If pulse expansion functions are to be used, how many unknowns will your model have?

14.13-3 Sketch a wire-grid model for a quarter-wavelength monopole at the center of a circular ground plane of $\lambda/4$ radius. If pulse expansion functions are used, how many unknowns will your models have? If piecewise sinusoidal functions are used, how many unknowns will your model have?

14.13-4 A form for the magnetic field integral equation (MFIE) is

$$\mathbf{J_s} = 2\,\hat{\mathbf{n}} \times \mathbf{H^i} - 2\,\hat{\mathbf{n}} \times \iint\limits_{S} \mathbf{J_s} \times \nabla\psi \, \mathbf{dS'}$$

Starting with the tangential H boundary condition, derive the MFIE.

Chapter 15

CEM for Antennas: Finite Difference Time Domain Method

The computational approach of the previous chapter involved setting up and solving frequency domain integral equations for the phasor electric and/or magnetic currents induced on the surfaces of antennas or scatterers. From a computing perspective, this method of moments (MoM) procedure involves setting up and solving dense (i.e., few zero elements), complex-valued systems of linear equations. These systems can involve tens of thousands of equations in the treatment of problems of even moderate electrical size, as we saw at end of the previous chapter.

As powerful as MoM is, it is inadequate for some important engineering problems, particularly those involving pulsed excitations and various transient phenomena. These kinds of problems require data to be computed over a range of frequencies. This suggests the need for a solution technique in the time domain, since all of the required frequency domain data can be generated from one-time domain solution via Fourier transformation. There is an approach that directly solves Maxwell's curl equations at points on space grids in the time domain. It is the finite difference time domain (FDTD) method (see Fig. 14-1). There are at least four reasons for the development of interest in such partial differential equation (PDE) solutions of Maxwell's equations:

1. PDE solutions are robust.
2. Time domain PDE methods usually have no matrices (frequency domain PDE methods usually have sparse matrices).
3. Complex-valued material properties are readily accommodated.
4. Computer resources are adequate to allow widespread usage of PDE methods.

The finite difference time domain (FDTD) technique to be discussed in the following sections offers many advantages as an electromagnetic modeling, simulation, and analysis tool. Its capabilities include:

- Broadband response predictions with a single excitation
- Arbitrary three-dimensional (3D) model geometries
- Interaction with an object of any conductivity from that of a perfect conductor, to that of low or zero conductivity
- Frequency-dependent constitutive parameters for modeling most materials:

Lossy dielectrics
Magnetic materials
Unconventional materials that can be anisotropic and/or nonlinear
- Any type of response such as:
  Scattered fields
  Antenna patterns
  Radar cross section (RCS)
  Surface response fields
  Currents, power densities, charge distributions
  Penetration/interior coupling

The basis of the FDTD algorithm is the two Maxwell curl equations in derivative form in the time domain. These equations are expressed in linearized form by means of central finite differencing. Only nearest-neighbor interactions need be considered as the fields are advanced temporally in discrete time steps over spatial cells usually of rectangular shape as indicated in Fig. 15-1 (other cell shapes are possible, as well as two-dimensional and one-dimensional treatments).

Although FDTD is well suited to computing responses to a continuous wave or single-frequency excitation, it is particularly well suited to computing transient responses. This is especially the case when complex geometries or difficult environments, such as an antenna that is buried in the Earth or dielectrically clad, are considered. Also, interior coupling into metallic enclosures is a situation where FDTD is a method of choice.

For problems where the modeled region must extend to infinity, absorbing boundary conditions (ABCs) are employed at the outer-grid truncation planes (grid boundary), which ideally permit all outgoing numerical waves to exit the region with negligible reflection at the grid truncation. Phenomena such as the induction of surface currents, scattering and multiple scattering, aperture penetration and cavity excitation are modeled time step by time step by the action of the numerical analog to the curl equations. The self-consistency of these modeled phenomena is generally assured if their spatial and temporal variations are well resolved by the space and time sampling process. A self-consistent model will account for the mutual coupling of all the electrically small-volume cells constituting the structure and its near field, even if the structure spans tens of wavelengths in three dimensions and there are hundreds of millions of space cells. In contrast, MoM provides a self-consistent solution of Maxwell's equations, which includes all mutual coupling, by solving a system of simultaneous equations.

In the remainder of this chapter, we develop FDTD for application to two classes of problems: antennas and scattering, with emphasis on the former. The theoretical
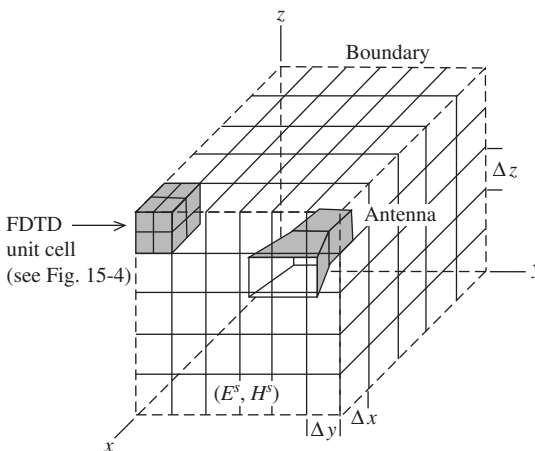
Figure 15-1 Embedding of an antenna structure in an FDTD space lattice.

development is for the three-dimensional objects in or near isotropic media, but examples are mostly specialized to two-dimensional and one-dimensional problems. The material that follows is intended as an introduction to FDTD. For more extensive information, refer to larger works on the subject [H.10.3: Taflove; H.10.3: Kunz]. The presentation here is influenced by [1] and [H.10.3: Taflove, 1st ed; H.10.3: Kunz], particularly [1] and [H.10.3: Taflove, 1st ed].

In the following section, we examine the form of Maxwell's equations we need to solve in one, two, and three dimensions before developing the rectangular coordinate system finite difference representations for those equations in Sec. 15.2. The finite difference equations are the equations used in the Yee algorithm of the FDTD technique and are subject to the constraints discussed in Sec. 15.3. Implementation of the finite difference equations is discussed in Sec. 15.4. Discussion of the absorbing boundary conditions follows in Sec. 15.5. Some sources used in FDTD are presented in Sec. l5.6, whereas transformation from the near field to the far field is contained in Sec. 15.7. Sec. 15.8 and 15.9 provide two-dimensional and three-dimensional examples, respectively.

15.1   MAXWELL'S EQUATIONS FOR THE FDTD METHOD

Before developing the basis for the FDTD method in the next section, we need to examine Maxwell's time domain curl equations in one, two, and three dimensions and put them in a form convenient for the FDTD method. Consider a region of space that is source-free but may have lossy magnetic and/or lossy electric materials that convert energy in the electromagnetic field to heat. We define an equivalent magnetic current density $\mathbf{M}$ to account for the magnetic loss mechanisms:

$$\mathbf{M} = \rho' \mathbf{H} \tag{15-1}$$

and an equivalent electric current density $\mathbf{J}$ to account for the electric loss mechanisms:

$$\mathbf{J} = \sigma \mathbf{E} \tag{15-2}$$

Here, ρ' is an equivalent magnetic resistivity in ohms per meter and σ the electric conductivity in siemens per meter. Thus, we can write

$$\frac{\partial \mathbf{H}}{\partial t} = -\frac{1}{\mu} \boldsymbol{\nabla} \times \mathbf{E} - \frac{\rho'}{\mu} \mathbf{H} \tag{15-3}$$

$$\frac{\partial \mathbf{E}}{\partial t} = \frac{1}{\varepsilon} \boldsymbol{\nabla} \times \mathbf{H} - \frac{\sigma}{\varepsilon} \mathbf{E} \tag{15-4}$$

15.1.1   Three-Dimensional Formulation

Writing out the vector components of the two curl equations above yields the following system of six coupled scalar equations in the three-dimensional rectangular coordinate system[1]:

$$\frac{\partial H_x}{\partial t} = \frac{1}{\mu} \left(\frac{\partial E_y}{\partial z} - \frac{\partial E_z}{\partial y} - \rho' H_x \right) \tag{15-5a}$$

---

[1] For clarity and notational convenience, script will not be used in this chapter for the time-varying field quantities as in Chap. 2.

$$\frac{\partial H_y}{\partial t} = \frac{1}{\mu}\left(\frac{\partial E_z}{\partial x} - \frac{\partial E_x}{\partial z} - \rho' H_y\right) \qquad (15\text{-}5b)$$

$$\frac{\partial H_z}{\partial t} = \frac{1}{\mu}\left(\frac{\partial E_x}{\partial y} - \frac{\partial E_y}{\partial x} - \rho' H_z\right) \qquad (15\text{-}5c)$$

$$\frac{\partial E_x}{\partial t} = \frac{1}{\varepsilon}\left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} - \sigma E_x\right) \qquad (15\text{-}6a)$$

$$\frac{\partial E_y}{\partial t} = \frac{1}{\varepsilon}\left(\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} - \sigma E_y\right) \qquad (15\text{-}6b)$$

$$\frac{\partial E_z}{\partial t} = \frac{1}{\varepsilon}\left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} - \sigma E_z\right) \qquad (15\text{-}6c)$$

This system of six coupled partial differential equations forms the basis of the FDTD numerical algorithm to be developed in the next section. Before proceeding with the full three-dimensional FDTD algorithm, it is useful to consider reductions to the two-dimensional and one-dimensional cases, which can yield engineering information without the computational effort required for the three-dimensional case.

15.1.2   Two-Dimensional Formulation

In the two-dimensional problem, there is no variation with respect to one of the coordinates in either the problem geometry or excitation. Here, we assume no variation with respect to z, which means that all partial derivatives of the fields with respect to z equal zero, and that the structure being modeled extends to infinity in the z-direction with no change in its geometry.

Consider grouping the previous six equations, with all partial derivatives with respect to z equal to zero, into two sets, one of which only involves magnetic field components transverse to the problem geometry axis (i.e., the z-axis) and the other in which there are only electric field components transverse to the z-axis. The first set is called the two-dimensional transverse magnetic (TM) mode and is

$$\frac{\partial H_x}{\partial t} = \frac{1}{\mu}\left(-\frac{\partial E_z}{\partial y} - \rho' H_x\right) \qquad (15\text{-}7a)$$

$$\frac{\partial H_y}{\partial t} = \frac{1}{\mu}\left(\frac{\partial E_z}{\partial x} - \rho' H_y\right) \quad \textit{two-dimensional TM mode} \qquad (15\text{-}7b)$$

$$\frac{\partial E_z}{\partial t} = \frac{1}{\varepsilon}\left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} - \sigma E_z\right) \qquad (15\text{-}7c)$$

The second set is then the two-dimensional transverse electric (TE) mode and is

$$\frac{\partial E_x}{\partial t} = \frac{1}{\varepsilon}\left(\frac{\partial H_z}{\partial y} - \sigma E_x\right) \qquad (15\text{-}8a)$$

$$\frac{\partial E_y}{\partial t} = \frac{1}{\varepsilon}\left(-\frac{\partial H_z}{\partial x} - \sigma E_y\right) \quad \textit{two-dimensional TE mode} \qquad (15\text{-}8b)$$

$$\frac{\partial H_z}{\partial t} = \frac{1}{\mu}\left(\frac{\partial E_x}{\partial y} - \frac{\partial E_y}{\partial x} - \rho' H_z\right) \tag{15-8c}$$

We observe that the TM and TE modes are decoupled; that is, they contain no common field vector components. In fact, these modes are completely independent for structures comprised of isotropic materials. That is, the modes can exist simultaneously with no mutual interactions. Problems having both TM and TE excitation can be solved as a superposition of these two separate problems.

Physical phenomena associated with the TM and TE cases can be quite different. To see this, one can look at, for example, Fig. 16-34 which shows the considerable difference in echo widths for the TE (perpendicular case) and TM (parallel case) excitation of an infinite circular cylinder.

15.1.3   One-Dimensional Formulation

Next, assume that there is no variation with respect to two coordinates in either the problem geometry or the excitation. In this instance, assume no variation with respect to either y or z, which means that all partial derivatives with respect to either y or z equal zero. This implies that the problem is one-dimensional in nature with propagation in the x-direction, but with space infinite in the y- and z-directions. Thus, while propagating in the x-direction, a wave could encounter infinite sheets of material having thickness in the x-dimension.

The one-dimensional problem is formulated by reducing either the two-dimensional TM mode or two-dimensional TE mode and ultimately obtaining almost the same result. Reducing the two-dimensional TM mode gives:

$$\frac{\partial H_x}{\partial t} = \frac{1}{\mu}\left(-\rho' H_x\right) \tag{15-9a}$$

$$\frac{\partial H_y}{\partial t} = \frac{1}{\mu}\left(\frac{\partial E_z}{\partial x} - \rho' H_y\right) \tag{15-9b}$$

$$\frac{\partial E_z}{\partial t} = \frac{1}{\varepsilon}\left(\frac{\partial H_y}{\partial x} - \sigma E_z\right) \tag{15-9c}$$

The first of these three equations can be shown to vanish by reasoning that if the fields are all zero prior to some time, say, $t = 0$, when a source is turned on, then the time derivative of H_x is zero. This, in turn, implies that H_x remains at zero. We now have a set of just two equations involving only H_y and E_z. Designate this set the TM mode in one dimension:

$$\frac{\partial H_y}{\partial t} = \frac{1}{\mu}\left(\frac{\partial E_z}{\partial x} - \rho' H_y\right) \tag{15-10a}$$

<center>*one-dimensional TM mode*</center>

$$\frac{\partial E_z}{\partial t} = \frac{1}{\varepsilon}\left(\frac{\partial H_y}{\partial x} - \sigma E_z\right) \tag{15-10b}$$

In a similar way, we can reduce the two-dimensional TE mode to a set of two equations involving only E_y and H_z. Designate this set the TE mode in one dimension:

$$\frac{\partial E_y}{\partial t} = \frac{1}{\varepsilon}\left(-\frac{\partial H_z}{\partial x} - \sigma E_y\right)$$

(15-11a)

one-dimensional TE mode

$$\frac{\partial H_z}{\partial t} = \frac{1}{\mu}\left(-\frac{\partial E_y}{\partial x} - \rho' H_z\right)$$

(15-11b)

The only practical difference between the one-dimensional TM and TE modes is that they represent plane waves of orthogonal polarizations. This renders the TM and TE labels in the one-dimensional case uncommon since we would ordinarily identify the plane wave polarization in some other way.

From either one-dimensional set, we can easily derive the one-dimensional scalar wave equation for a component of E and that for a component of H, both of which only have for solutions plane waves traveling in the $\pm x$-direction at a speed given by $1/\sqrt{\mu\varepsilon}$. That is, in the one-dimensional case we have transverse electromagnetic (TEM) plane waves traveling at a speed determined by the constitutive parameters of the medium.

In the next section, we examine numerical solutions to the one-dimensional, two-dimensional, and three-dimensional equations developed here.

15.2   FINITE DIFFERENCES AND THE YEE ALGORITHM

In this section, we develop the Yee algorithm used in the FDTD method. The Yee algorithm is based on finite difference approximations of the space derivatives and time derivatives in Maxwell's curl equations as shown later in this section. To begin our development, consider (15-10a) in the lossless case:

$$\frac{\partial H_y}{\partial t} = \frac{1}{\mu}\frac{\partial E_z}{\partial x}$$

(15-12)

Employing the classical definition of a derivative, we can write

$$\lim_{\Delta t \to 0}\frac{\Delta H_y}{\Delta t} = \frac{1}{\mu}\lim_{\Delta x \to 0}\frac{\Delta E_z}{\Delta x}$$

(15-13)

In Fig. 15-2 we illustrate (15-13) and note that in the limit a continuous and exact solution to (15-13) is obtained at the point (x, t). It is important to note that at this point, space and time derivatives are being equated, and not the actual values of the fields. In other words, it is apparent that Maxwell's equations do not directly yield electric and magnetic field values, but rather relate the rate of change between electric and magnetic field values.
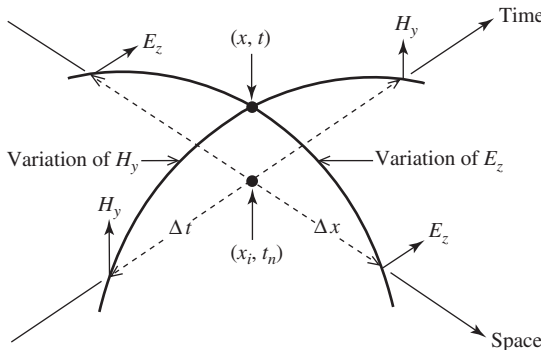
Figure 15-2 Space-time graphical interpretation of a one-dimensional component of Maxwell's equations and its discretization [1]. (Reprinted with permission of Eric Thiele.)

Thus the following strategy may suggest itself. Discretize space and time around the point (x, t) in such a way that Maxwell's equations hold true. That is, apply central differences to relate the derivatives of the neighboring discrete fields. Then, for example, (15-13) can be expressed as

$$\frac{H_y\left(t_n + \frac{\Delta t}{2}\right) - H_y\left(t_n - \frac{\Delta t}{2}\right)}{\Delta t}\Bigg|_{x_i} = \frac{1}{\mu} \frac{E_z\left(x_i + \frac{\Delta x}{2}\right) - E_z\left(x_i - \frac{\Delta x}{2}\right)}{\Delta x}\Bigg|_{t_n} \qquad (15\text{-}14)$$

This gives us the relationship between the derivatives at (x_i, t_n) which closely approximates the relationship at (x, t). However, as we shall see later, if (15-14) is solved for the field quantity at the most advanced point in time (i.e., $t_n + \Delta t/2$), then an estimate of the magnetic field value at the spatial point x_i at time $(t_n + \Delta t/2)$ can be obtained.

We could obtain (15-14) from (15-12) in a more formal way by expanding $H_y(x_i, t_n)$ in a Taylor series about the temporal point t_n to temporal point $t_n + \frac{\Delta t}{2}$, keeping space fixed at the point x_i. This yields an expansion for $H_y\left(t_n + \frac{\Delta t}{2}\right)\big|_{x_i}$. Similarly, we could obtain an explain for $H_y\left(t_n - \frac{\Delta t}{2}\right)\big|_{x_i}$. Taking the difference would give the left-hand side of (15-14) plus remainder terms on the order of $(\Delta t)^2$. Likewise, expansions of E_z about x_i in both directions with time fixed lead to the right-hand side of (15-14) plus remainder terms on the order of $(\Delta x)^2$. In this way, we formally obtain second-order-accurate central difference approximations to the first partial derivatives in time and space.

Continuing our FDTD solution to (15-12), we solve (15-14) for, $H_y\left(t_n + \frac{\Delta t}{2}\right)$ and obtain [after dropping the z and y subscripts on the field components and taking (15-14) to be an equality]

$$H\left(t_n + \frac{\Delta t}{2}\right)\Bigg|_{x_i} = H\left(t_n - \frac{\Delta t}{2}\right)\Bigg|_{x_i} + \frac{\Delta t}{\mu \Delta x}\left[E\left(x_i + \frac{\Delta x}{2}\right) - E\left(x_i - \frac{\Delta x}{2}\right)\right]_{t_n} \qquad (15\text{-}15)$$

For convenience, we notationally adopt a subscript i for the space position and a superscript n for the time observation point. If we use this shorthand notation, (15-15) can be written compactly as

$$H_i^{n+1/2} = H_i^{n-1/2} + \frac{\Delta t}{\mu \Delta x}\left[E_{i+1/2}^n - E_{i-1/2}^n\right] \qquad (15\text{-}16)$$

which implies that we can solve for $H_i^{n+1/2}$ knowing the value for H at the same spatial point but at Δt earlier in time and knowing E at spatial points $\pm \Delta x/2$ removed from x_i and $\Delta t/2$ earlier in time. This is illustrated in Fig. 15-3 that shows on a time-space diagram the three quantities linked (by the various dashed lines) to the calculation of $H_i^{n+1/2}$.

How do we obtain E at spatial points $x_i \pm \Delta x/2$? The answer, of course, is to start with the lossless version of (15-10b) and obtain for E_z the central difference approximation

$$E_{i+1/2}^{n+1} = E_{i+1/2}^n + \frac{\Delta t}{\varepsilon \Delta x}\left[H_{i+1}^{n+1/2} - H_i^{n+1/2}\right] \qquad (15\text{-}17)$$

by differencing about the temporal point $(n + 1/2)$ and the spatial point $(i + 1/2)$. This equation says that $E_{i+1/2}^{n+1}$ can be calculated with values of H and E at previous instants of time at adjacent spatial locations as Fig. 15-3 suggests. Clearly, we have the basis for a method that can march a field behavior forward in space and time through the use of
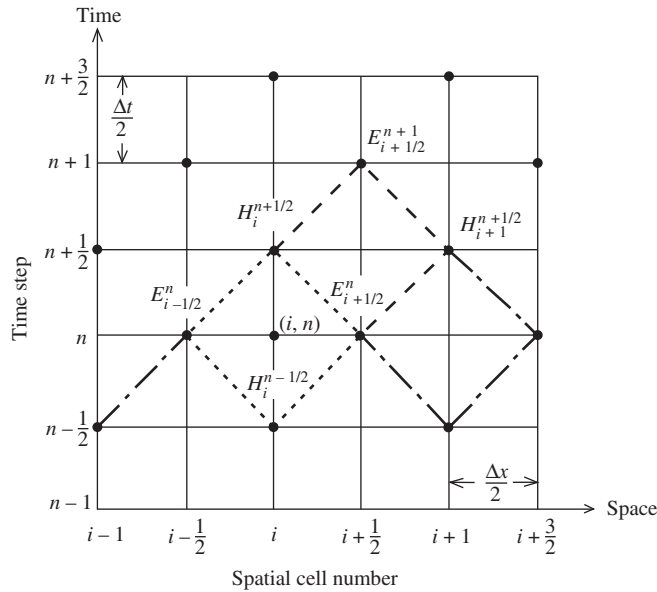
Figure 15-3 Calculation of $H_i^{n+1/2}$ and $E_{i+1/2}^{n+1}$ from its nearest neighbors in space and time [1]. (Reprinted with permission of Eric Thiele.)

difference equations like (15-16) and (15-17), more commonly called *update equations* (because they update the fields in the cells as time moves forward).

In 1966, K. S. Yee [2] originated a set of finite difference equations for the lossless three-dimensional time-dependent Maxwell's curl equations of (15-5) and (15-6) similar to that above for the lossless one-dimensional case. Yee's algorithm, introduced later in this section for the three-dimensional case, is one of great usefulness since its fundamental basis is so robust. The Yee algorithm is robust for the following reasons. First, it solves for *both* electric and magnetic fields in time and space using the coupled Maxwell's curl equations rather than solving for the electric field alone (or the magnetic field alone) as with the wave equation.

Second, the Yee algorithm interleaves its E- and H-field vector components in three-dimensional space (see Fig. 15-4), so that every E-field vector component is surrounded by H-field components, and every H-field vector component is surrounded by E-field components as suggested by Fig. 15-4. The spatial arrangement in Fig. 15-4 is not arbitrary since it must be consistent with the laws of Ampere and Faraday [1].

Third, the Yee algorithm centers its E- and H-field vector components in time in what is commonly termed a *leapfrog* arrangement. All the E-field computations in the three-dimensional space of interest are computed for a particular time point using the most recently computed H-field data stored in the computer memory (as Fig. 15-3 suggests for the one-dimensional case). Then, all the H-field computations in the three-dimensional space are computed using the E-field data just computed and stored in memory. This leapfrog arrangement is then repeated with the recomputation of the E-field based on the newly obtained H-field. This process continues for a finite number of time steps until some desired late time response is achieved (e.g., steady state).

Fourth, no matrices are involved and no large systems of simultaneous equations need to be solved as in the method of moments.

Eq. (15-17) above involved two variables: one in space and one in time. In the most general case, we have four degrees of freedom (three in space and one in time), and must choose notation carefully. Extending the notation from (15-17), denote a space point in a uniform, rectangular three-dimensional lattice as
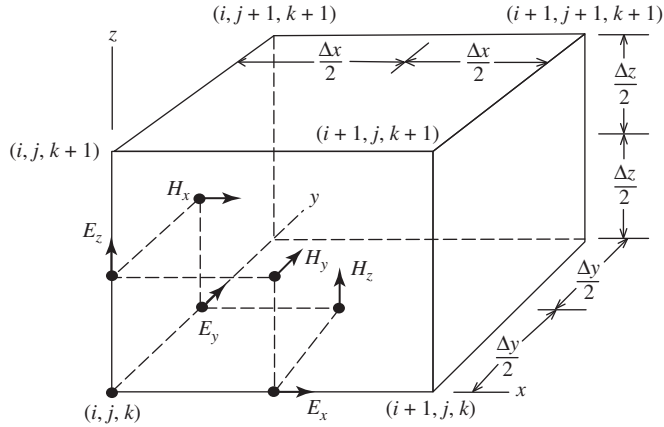
Figure 15-4 Position of the electric and magnetic field vector components in a cubic unit cell of the Yee space lattice of dimension Δx by Δy by Δz [1]. (Reprinted with permission of Eric Thiele.)

$$(i,\ j,\ k) = (i\Delta x,\ j\Delta y,\ k\Delta z) \tag{15-18}$$

Here, Δx, Δy, and Δz are, respectively, the lattice space increments in the x-, y-, and z-coordinate directions, and i, j, and k are integers. Further, we denote any field component u as a function of space and time evaluated at a discrete point in the space lattice and at a discrete point in time as

$$u(i\Delta x,\ j\Delta y,\ k\Delta z,\ n\Delta t) = u_{i,j,k}^{n} \tag{15-19}$$

Here, Δt is the time increment, assumed uniform over the observation interval, and n is an integer. Carrying this notation to derivatives, we find, for example, Yee's expression for the first space derivative of u in the x-direction, evaluated at the fixed time $t_n = n\Delta t$, to be

$$\frac{\partial}{\partial x} u(i\Delta x,\ j\Delta y,\ k\Delta z,\ n\Delta t) = \frac{u_{i+1/2,j,k}^{n} - u_{i-1/2,j,k}^{n}}{\Delta x} + O\left[(\Delta x)^2\right] \tag{15-20a}$$

We note that the $\pm\frac{1}{2}$ increment in the i subscript (x-coordinate) of u denotes a space finite difference over $\pm\Delta x/2$ as in (15-17). The remainder term $O[(\Delta x)^2]$ is a result of the Taylor series expansions that formally lead to the second-order-accurate finite difference representation of the derivatives. The numerical approximation analogous to (15-20a) for $\partial u/\partial y$ or $\partial u/\partial z$ can be written simply by incrementing the j or k subscript of u by $\pm\frac{1}{2}\Delta y$ or $\pm\frac{1}{2}\Delta z$, respectively.

Yee's expression for the first time derivative of u, evaluated at the fixed space point (i,j,k), follows by analogy:

$$\frac{\partial}{\partial t} u(i\Delta x, j\Delta y, k\Delta z, n\Delta t) = \frac{u_{i,j,k}^{n+1/2} - u_{i,j,k}^{n-1/2}}{\Delta t} + O\left[(\Delta t)^2\right] \tag{15-20b}$$

Here, the $\pm\frac{1}{2}$ increment in the n superscript (time value) of u denotes a time finite difference over $\pm\Delta t/2$ as in the left-hand side of (15-14).

We now apply the above ideas and notation to achieve a numerical approximation of Maxwell's curl equations in three dimensions, given by (15-16) and (15-17). For example, consider (15-5a), repeated here for convenience:

$$\frac{\partial H_x}{\partial t} = \frac{1}{\mu}\left(\frac{\partial E_y}{\partial z} - \frac{\partial E_z}{\partial y} - \rho' H_x\right) \tag{15-21}$$

Substituting for the time and space derivatives at time step n and assuming the space lattice point (i, j, k), we have initially[2]

$$\frac{H_x\big|_{i,j,k}^{n+1/2} - H_x\big|_{i,j,k}^{n-1/2}}{\Delta t} = \frac{1}{\mu_{i,j,k}}\left(\begin{array}{c} \dfrac{E_y\big|_{i,j,k+1/2}^{n} - E_y\big|_{i,j,k-1/2}^{n}}{\Delta z} \\[12pt] -\dfrac{E_z\big|_{i,j+1/2,k}^{n} - E_z\big|_{i,j-1/2,k}^{n}}{\Delta y} \\[12pt] -\rho'_{i,j,k}\cdot H_x\big|_{i,j,k}^{n} \end{array}\right) \tag{15-22}$$

Note that all field quantities on the right-hand side (in a compact format) are evaluated at time step n, including the magnetic field term H_x, appearing due to the magnetic loss ρ'. Since H_x at time step n is not assumed to be stored in computer memory (only the previous values of H_x at time step $n - \frac{1}{2}$ are assumed to be in memory), we need some way to estimate this term. A very good approach is to apply a "semi-implicit" approximation to the H_x term on the right-hand side:

$$H_x\big|_{i,j,k}^{n} = \frac{H_x\big|_{i,j,k}^{n+1/2} + H_x\big|_{i,j,k}^{n-1/2}}{2} \tag{15-23}$$

H_x at time step n is assumed to be simply the arithmetic average of the stored value of H_x at the time step $n - \frac{1}{2}$ and the yet to be computed new value of H_x at time step $n + \frac{1}{2}$ Substituting into (15-22) after multiplying both sides by Δt, we obtain

$$H_x\big|_{i,j,k}^{n+1/2} - H_x\big|_{i,j,k}^{n-1/2} = \frac{\Delta t}{\mu_{i,j,k}}\left(\begin{array}{c} \dfrac{E_y\big|_{i,j,k+1/2}^{n} - E_y\big|_{i,j,k-1/2}^{n}}{\Delta z} \\[12pt] -\dfrac{E_z\big|_{i,j+1/2,k}^{n} - E_z\big|_{i,j-1/2,k}^{n}}{\Delta y} \\[12pt] -\rho'_{i,j,k}\cdot\left(\dfrac{H_x\big|_{i,j,k}^{n+1/2} + H_x\big|_{i,j,k}^{n-1/2}}{2}\right) \end{array}\right) \tag{15-24}$$

We note that $H_x\big|_{i,j,k}^{n+1/2}$ and $H_x\big|_{i,j,k}^{n-1/2}$ appear on both sides of (15-24). Collecting all terms of these two types and isolating $H_x\big|_{i,j,k}^{n+1/2}$ on the left-hand side yield

---

[2] For ease of presentation in some of the equations that follow, the terms inside the parentheses are "stacked" vertically rather than written horizontally.

$$\left(1 + \frac{\Delta t}{\mu_{i,j,k}} \cdot \frac{\rho'_{i,j,k}}{2}\right) \cdot H_x\Big|_{i,j,k}^{n+1/2} = \left(1 - \frac{\Delta t}{\mu_{i,j,k}} \cdot \frac{\rho'_{i,j,k}}{2}\right) \cdot H_x\Big|_{i,j,k}^{n-1/2}$$

$$+ \frac{\Delta t}{\mu_{i,j,k}} \left(\frac{E_y\Big|_{i,j,k+1/2}^{n} - E_y\Big|_{i,j,k-1/2}^{n}}{\Delta z} - \frac{E_z\Big|_{i,j+1/2,k}^{n} - E_z\Big|_{i,j-1/2,k}^{n}}{\Delta y} \right) \qquad (15\text{-}25)$$

Finally, dividing both sides by $\left(1 + \frac{\Delta t}{\mu_{i,j,k}} \cdot \frac{\rho'_{i,j,k}}{2}\right)$ yields an explicit expression for $H_x\Big|_{i,j,k}^{n+1/2}$:

$$H_x\Big|_{i,j,k}^{n+1/2} = \left(\frac{1 - \dfrac{\rho'_{i,j,k}\,\Delta t}{2\mu_{i,j,k}}}{1 + \dfrac{\rho'_{i,j,k}\,\Delta t}{2\mu_{i,j,k}}} \right) \cdot H_x\Big|_{i,j,k}^{n-1/2}$$

$$+ \left(\frac{\dfrac{\Delta t}{\mu_{i,j,k}}}{1 + \dfrac{\rho'_{i,j,k}\,\Delta t}{2\mu_{i,j,k}}} \right) \cdot \left(\frac{E_y\Big|_{i,j,k+1/2}^{n} - E_y\Big|_{i,j,k-1/2}^{n}}{\Delta z} - \frac{E_z\Big|_{i,j+1/2,k}^{n} - E_z\Big|_{i,j-1/2,k}^{n}}{\Delta y} \right) \qquad (15\text{-}26)$$

In a similar manner, we can derive finite difference expressions based on Yee's algorithm for the H_y and H_z field components in the curl equations.

By analogy, we can derive finite difference expressions based on Yee's algorithm for the E_x, E_y, and E_z field components given by (15-6). Here, $\sigma E^{n+1/2}$ represents the loss term on the right-hand side of each equation that is estimated using a semi-implicit procedure similar to that of (15-23). This results in a set of three equations for E_x, E_y, and E_z. For example, the result for E_z, also at space lattice point (i, j, k), is

$$E_z\Big|_{i,j,k}^{n+1} = \left(\frac{1 - \dfrac{\sigma_{i,j,k}\,\Delta t}{2\varepsilon_{i,j,k}}}{1 + \dfrac{\sigma_{i,j,k}\,\Delta t}{2\varepsilon_{i,j,k}}} \right) \cdot E_z\Big|_{i,j,k}^{n}$$

$$+ \left(\frac{\dfrac{\Delta t}{\varepsilon_{i,j,k}}}{1 + \dfrac{\sigma_{i,j,k}\,\Delta t}{2\sigma_{i,j,k}}} \right) \cdot \left(\frac{H_y\Big|_{i+1/2,j,k}^{n+1/2} - H_y\Big|_{i-1/2,j,k}^{n+1/2}}{\Delta x} - \frac{H_x\Big|_{i,j+1/2,k}^{n+1/2} - H_x\Big|_{i,j-1/2,k}^{n+1/2}}{\Delta y} \right) \qquad (15\text{-}27)$$

With the above expressions for $H^{n+1/2}$ and E^{n+1}, the new value of a field vector component at any space lattice point depends only on its previous value and the previous values of the components of the other field vectors at adjacent points.

To implement a solution like (15-26) and (15-27) for a region having a continuous variation of isotropic material properties with spatial position, it is desirable to define and store the following constant coefficients for each field vector component before time

stepping is commenced. For a cubic lattice where $\Delta x = \Delta y = \Delta z = \Delta s$, we have for the electric field algorithm coefficients at point $(i,\ j,\ k)$

$$C_a\big|_{i,j,k} = \frac{1 - \dfrac{\sigma_{i,j,k}}{2\varepsilon_{i,j,k}}\Delta t}{1 + \dfrac{\sigma_{i,j,k}}{2\varepsilon_{i,j,k}}\Delta t} \tag{15-28a}$$

$$C_b\big|_{i,j,k} = \frac{\dfrac{\Delta t}{\varepsilon_{i,j,k}\,\Delta s}}{1 + \dfrac{\sigma_{i,j,k}}{2\varepsilon_{i,j,k}}\Delta t} \tag{15-28b}$$

And for the magnetic field algorithm coefficients at point $(i,\ j,\ k)$, we have

$$D_a\big|_{i,j,k} = \frac{1 - \dfrac{\rho'_{i,j,k}}{2\mu_{i,j,k}}\Delta t}{1 + \dfrac{\rho'_{i,j,k}}{2\mu_{i,j,k}}\Delta t} \tag{15-29a}$$

$$D_b\big|_{i,j,k} = \frac{\dfrac{\Delta t}{\mu_{i,j,k}\,\Delta s}}{1 + \dfrac{\rho'_{i,j,k}}{2\mu_{i,j,k}}\Delta t} \tag{15-29b}$$

Note that the lattice increment Δs is contained in C_b and D_b.

The complete set of finite difference equations suggested by (15-26) and (15-27) can now be written to conform to the spatial arrangement in Fig. 15-4 by adjusting the spatial indices appropriately. For example, to the spatial indices in (15-26), we add $\frac{1}{2}$ to both j and k to obtain the following equation for H_x and in (15-27) add $\frac{1}{2}$ to the index k to obtain the following equation for E_z. Thus, the complete set of six equations can be written as

$$H_x\big|_{i,j+1/2,k+1/2}^{n+1/2} = D_{aHX}\big|_{i,j+1/2,k+1/2} \cdot H_x\big|_{i,j+1/2,k+1/2}^{n-1/2} + D_{bHX}\big|_{i,j+1/2,k+1/2}$$
$$\cdot \left(E_y\big|_{i,j+1/2,k+1}^{n} - E_y\big|_{i,j+1/2,k}^{n} + E_z\big|_{i,j,k+1/2}^{n} - E_z\big|_{i,j+1,k+1/2}^{n}\right) \tag{15-30a}$$

$$H_y\big|_{i+1/2,j,k+1/2}^{n+1/2} = D_{aHY}\big|_{i+1/2,j,k+1/2} \cdot H_y\big|_{i+1/2,j,k+1/2}^{n-1/2} + D_{bHY}\big|_{i+1/2,j,k+1/2}$$
$$\cdot \left(E_z\big|_{i+1,j,k+1/2}^{n} - E_z\big|_{i,j,k+1/2}^{n} + E_x\big|_{i+1/2,j,k}^{n} - E_x\big|_{i+1/2,j,k+1}^{n}\right) \tag{15-30b}$$

$$H_z\big|_{i+1/2,j+1/2,k}^{n+1/2} = D_{aHZ}\big|_{i+1/2,j+1/2,k} \cdot H_z\big|_{i+1/2,j+1/2,k}^{n-1/2} + D_{bHZ}\big|_{i+1/2,j+1/2,k}$$
$$\cdot \left(E_x\big|_{i+1/2,j+1,k}^{n} - E_x\big|_{i+1/2,j,k}^{n} + E_y\big|_{i,j+1/2,k}^{n} - E_y\big|_{i+1,j+1/2,k}^{n}\right) \tag{15-30c}$$

$$E_x\big|_{i+1/2,j,k}^{n+1} = C_{aEX}\big|_{i+1/2,j,k} \cdot E_z\big|_{i+1/2,j,k}^{n} + C_{bEX}\big|_{i+1/2,j,k}$$
$$\cdot \left(H_z\big|_{i+1/2,j+1/2,k}^{n+1/2} - H_z\big|_{i+1/2,j-1/2,k}^{n+1/2} + H_y\big|_{i+1/2,j,k-1/2}^{n+1/2} - H_y\big|_{i+1/2,j,k+1/2}^{n+1/2}\right)$$
$$\tag{15-31a}$$

$$E_y\big|_{i,j+1/2,k}^{n+1} = C_{a_{EY}}\big|_{i,j+1/2,k} \cdot E_y\big|_{i,j+1/2,k}^{n} + C_{b_{EY}}\big|_{i,j+1/2,k}$$

$$\cdot \left(H_x\big|_{i,j+1/2,k+1/2}^{n+1/2} - H_x\big|_{i,j+1/2,k-1/2}^{n+1/2} + H_z\big|_{i-1/2,j+1/2,k}^{n+1/2} - H_z\big|_{i+1/2,j+1/2,k}^{n+1/2}\right)$$

$$(15\text{-}31b)$$

$$E_z\big|_{i,j,k+1/2}^{n+1} = C_{a_{EZ}}\big|_{i,j,k+1/2} \cdot E_z\big|_{i,j,k+1/2}^{n} + C_{b_{EZ}}\big|_{i,j,k+1/2}$$

$$\cdot \left(H_y\big|_{i+1/2,j,k+1/2}^{n+1/2} - H_y\big|_{i-1/2,j,k+1/2}^{n+1/2} + H_x\big|_{i,j-1/2,k+1/2}^{n+1/2} - H_x\big|_{i,j+1/2,k+1/2}^{n+1/2}\right)$$

$$(15\text{-}31c)$$

The above six equations can be used for the three-dimensional case in Fig. 15-4 or reduced appropriately for the two-dimensional and one-dimensional cases, for example, the two-dimensional TE case contains the field components in the spatial arrangement given in Fig. 15-4 in the x-y plane, whereas the two-dimensional TM case contains the field components in the $k + \frac{1}{2}$ plane. A one-dimensional case may be obtained by appropriately reducing either of the two-dimensional cases.

Next, we develop bounds on the cell size and the time step used in the update equations and discuss the effects of dispersion.

15.3   CELL SIZE, NUMERICAL STABILITY, AND DISPERSION

Before we can implement the difference equations presented in the previous section, the cell size and time increment must be determined. In practice, the cell size is determined first. It is primarily influenced by numerical dispersion, which is the propagation of different frequencies with different velocities. Then after we have established the cell size, the time increment is determined such that numerical stability is achieved.

In view of our study of MoM earlier in this chapter, we can appreciate that over one FDTD cell dimension the electromagnetic field should not change significantly. This means that for meaningful results, the grid size should be only a fraction of the wavelength of the highest significant frequency content f_U in the excitation frequency spectrum. For example, from a study of Fourier analysis we know that for a pulse of width τ, the major portion of the frequency spectrum lies between zero and $f_U = 1/\tau$. The Nyquist sampling theorem would suggest that the cell size be less than $\lambda_U/2$ in order that the spatial variation of the fields be adequately sampled. However, our pulse has frequency content higher than f_U, numerical dispersion is present in the two-dimensional and three-dimensional cases, and our difference equations are themselves approximations, so a higher spatial sampling rate (i.e., smaller cell size) is required. Depending on the accuracy of desired results, it has been found that the cell size should be smaller than approximately $\lambda_U/10$ in the material medium (e.g., $\lambda_U/20$ if computational resources allow), primarily to minimize the effects of numerical dispersion. Fig. 15-5 shows, for the one-dimensional case, the effects of cell size on phase velocity and suggests a cell size at least as small as $\lambda_U/20$. Details of the geometry may dictate a still smaller cell size.

Now that we have established cell size, the time step Δt can be determined. Let us first consider the one-dimensional case. In one time step, any point on the wave must not travel more than one cell because during one time step the FDTD algorithm can propagate the wave only from one cell to its nearest neighbors. Any attempt to use even a slightly larger time step will quickly lead to numerical instability. We can do less than one cell in one time step, but it is not an optimum situation and will not lead to increased accuracy. Thus, the condition in the one-dimensional case is

$$\Delta t \leq \frac{\Delta x}{c} \qquad (15\text{-}32)$$

Phase velocity vs. spatial discretization
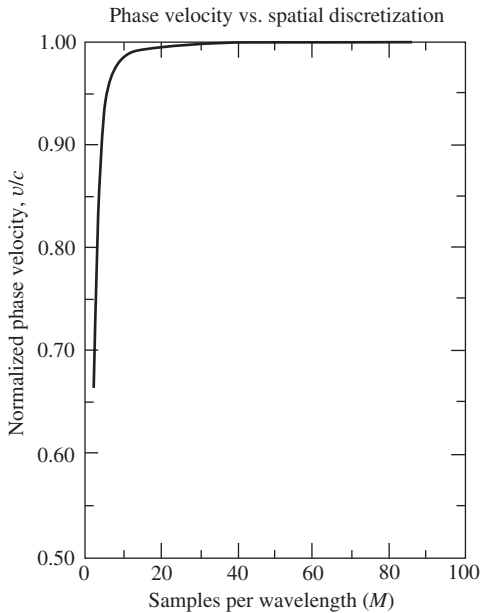
Figure 15-5 Variation of numerical phase velocity with samples (cells) per wavelength in a one-dimensional FDTD grid (From [H.10.3: Taflove, 1st ed.], © 1995. Reprinted with permission of Artech House, Inc. Norwood, MA.)

If the equality sign is used, we have what is referred to by Taflove [H.10.3: Taflove, 1st ed] as the *magic time step*, $c \, \Delta t = \Delta x$. It can be shown that the one-dimensional central difference equations produce an exact solution when the magic time step is used. This is an interesting result given that the difference equations are themselves approximations (see Fig. 15-2). Unfortunately, a similar condition does not exist in the two-dimensional and three-dimensional cases.

To guarantee numerical stability in the general case, it has been shown that

$$\Delta t \leq \frac{1}{c \sqrt{\dfrac{1}{(\Delta x)^2} + \dfrac{1}{(\Delta y)^2} + \dfrac{1}{(\Delta z)^2}}} \tag{15-33}$$

The above condition was obtained using the classical approach first suggested by Courant et al. in [5]. In this approach, a time eigenvalue problem is first solved and then a space eigenvalue problem is solved. Next, a stable range of space eigenvalues is forced to lie within the stable range of the time eigenvalues, resulting in the general relation above.

In the two-dimensional case, if $\Delta x = \Delta y = \Delta z = \Delta s$, (15-33) reduces to

$$\Delta t \leq \frac{\Delta s}{c \sqrt{2}} \tag{15-34}$$

whereas in the three-dimensional case, (15-33) reduces to

$$\Delta t \leq \frac{\Delta s}{c \sqrt{3}} \tag{15-35}$$

Examination of the above results shows that the minimum number of time steps required to travel the maximum dimension of a unit cell is equal to the dimensionality of the cell. Thus, at least two time steps are required to traverse the diagonal of a two-dimensional square cell and at least three time steps to traverse the diagonal of a three-dimensional cubic cell.

Before we leave this section, it is necessary to mention dispersion. Dispersion is the propagation of different numerical wavelengths with different velocities within the grid. Dispersion, for example, can cause the distortion of a pulse shape. In the one-dimensional case, dispersion is zero if the magic time step is used.

In the two-dimensional case, dispersion is zero if the equality is used in (15-33) and propagation is along the square cell diagonal. In any other direction of propagation, there will be dispersion. The situation is similar in the three-dimensional case. If the equality sign in (15-33) is used and propagation is along the cube diagonal, dispersion will be zero, otherwise not. Generally, numerical dispersion can be reduced, but not eliminated, by reducing the cell size.

Dispersion is illustrated in Fig. 15-6, which shows the variation of the normalized numerical phase velocity with the propagation angle in a two-dimensional FDTD grid where the inequality of (15-33) was used. The time step $c\,\Delta t = \Delta s/2$ was employed; it is an example of a time step commonly used in two-dimensional (and three-dimensional) grids to satisfy the stability criterion in (15-33) with a margin of error. The figure shows that the phase velocity is a minimum along the Cartesian axes ($\alpha = 0°$ and $\alpha = 90°$) and is a maximum at $\alpha = 45°$ (along the square cell diagonal), but is slightly less than c even there since the equality of (15-33) was not used. The general behavior in Fig. 15-6 represents a numerical anisotropy that is inherent in the Yee algorithm.

Fig. 15-7 shows the variation of the numerical phase velocity versus cell size for the same incidence angles and time step. The beneficial effect of small cell size is apparent. If too large of a cell size (i.e., too close to the Nyquist limit) is used, the wave will actually stop propagating.

Both figures imply that different frequency components of the excitation will propagate with different speeds, resulting in pulse distortion that will increase with distance. On the other hand, for a sinusoidal wave, the effect of an incorrect phase velocity would be to develop a lagging phase error that increases with propagation distance.

Now that we have bounds on the cell size and the time step, and understand the effects of dispersion, we are in a position to implement the central difference update equations. The next section examines how to do this.
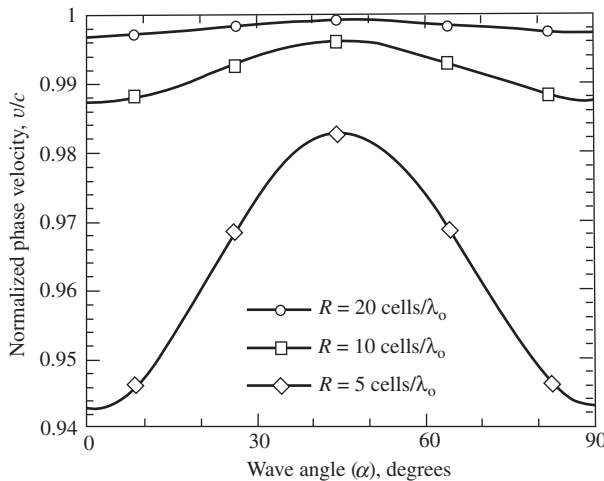
Figure 15-6 Variation of the numerical phase velocity with wave propagation angle in a two-dimensional FDTD grid for three grid resolutions. At 0 and 90°, incidence is along either Cartesian grid axis. (From [H.10.3: Taflove, 1st ed.] © 1995. Reprinted by permission of Artech House, Inc., Norwood, MA.)
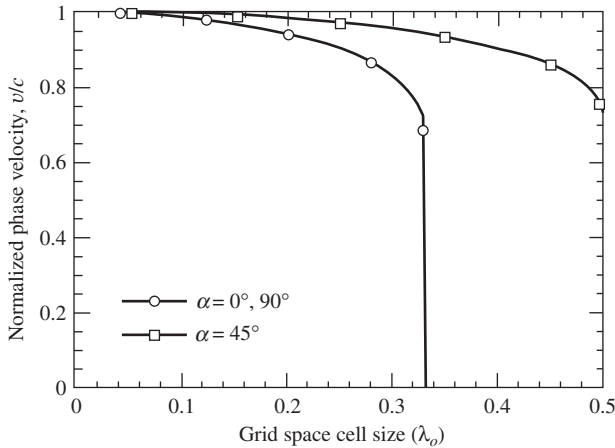
Figure 15-7 Variation of the numerical phase velocity with grid cell size in a two-dimensional FDTD grid for three wave propagation angles relative to Cartesian grid axes at 0 and 90°. (From [H.10.3: Taflove, 1st ed.] © 1995. Reprinted by permission of Artech House, Inc., Norwood, MA.)

15.4   COMPUTER ALGORITHM AND FDTD IMPLEMENTATION

In previous sections, we developed the Yee algorithm and explained some of the basic concepts in the FDTD approach. Although there are other fundamental issues to be discussed, it is helpful at this point to view the generalities of the computer architecture and how the Yee algorithm is implemented. Some specific points will be illustrated with a one-dimensional example.

The primary computational feature of an FDTD code is the time stepping process. This is a small part of the code, but the most heavily used part. Prior to time stepping, the FDTD grid must be defined as well as parameters such as cell size, time step, and the source condition. Constant multipliers that are not computed at each time step, such as the C and D coefficients in (15-28) and (15-29), should also be evaluated and stored before time stepping begins. There must be a geometrical definition of the antenna or scatterer that consists of identifying those cell locations containing material other than free space. This is done via the C_a, C_b, D_a, and D_b coefficients. In addition, desired responses must be specified so that they will be available for output after time stepping is completed (or perhaps during the time stepping if transient information is desired).

The code requirements consist of the following major steps:

Preprocessing

- Define the FDTD grid (sets the number of cells in each dimension and the cell size).
- Calculate the time step according to the Courant stability condition presented in the previous section.
- Calculate constant multipliers, including the C and D coefficients from Sec. 15.2, which serves to define the antenna or scatterer geometry in the FDTD grid.

Time stepping

- Update the source conditions (to be discussed in Sec. 15.6).
- Calculate the response of an E-field component from that of the nearest-neighbor field quantities according to the type of material present at the nearest-neighbor locations.
- Update the absorbing boundary condition (ABC), also called the outer radiation boundary condition. The purpose of the ABC (discussed in Sec. 15.5) is to absorb,

at the extremities of the FDTD grid, as much of the radiation field as possible to prevent nonphysical reflections within the FDTD grid.

- Update *H*-field components.

Postprocessing

- In software, arrays save response data such as *E*- and *H*-field components, currents, voltages, etc., at desired time steps.
- Determine the tangential electric and magnetic fields on a closed fictitious surface surrounding the antenna or scatterer and compute the corresponding scattered or radiated fields in the far zone (see Sec. 15.7).

A code structure that will implement the above requirements is suggested by the simplified flowchart shown in Fig. 15-8.

To illustrate how some of the calculations are done, a simplified one-dimensional model is used. The model is along the *x*-axis and in free space. We use the one-dimensional equations of (15-10*a*) and (15-10*b*), but the FDTD equations are taken from (15-30*b*) with $E_x = 0$ and from (15-31c) with $H_x = H_z = 0$. Note that $C_a = 1$ and $C_b = \frac{\Delta t}{\varepsilon_o \Delta x}$ and that $D_a = 1$ and $D_b = \frac{\Delta t}{\mu_o \Delta x}$ using the magic time step condition. Thus,

$$H_y\big|_{i+1/2}^{n+1/2} = H_y\big|_{i+1/2}^{n-1/2} + \frac{\Delta t}{\mu_o \Delta x}\left(E_z\big|_{i+1}^{n} - E_z\big|_{i}^{n}\right) \tag{15-36}$$
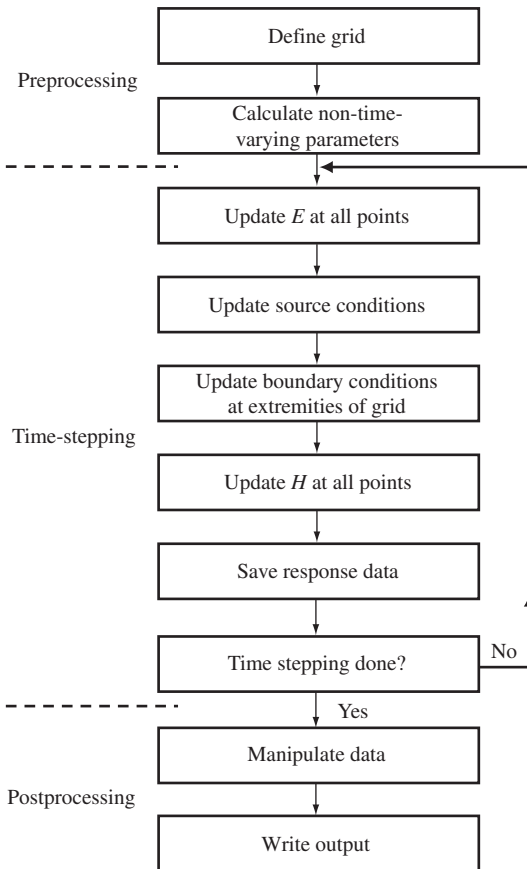
Figure 15-8 FDTD flowchart.
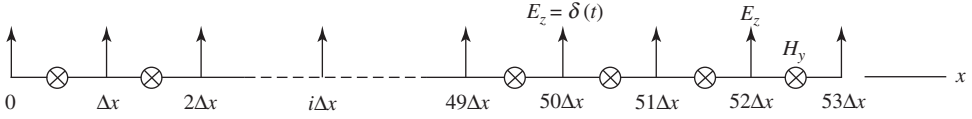
Figure 15-9 A one-dimensional 53-cell model.

$$E_z\big|_i^{n+1} = E_z\big|_i^n + \frac{\Delta t}{\varepsilon_o \Delta x}\left(H_y\big|_{i+1/2}^{n+1/2} - H_y\big|_{i-1/2}^{n+1/2}\right) \tag{15-37}$$

We note that $\frac{\Delta t}{\varepsilon_o \Delta x} = \frac{1}{\varepsilon_o c} = \eta$ and that $\frac{\Delta t}{\mu_o \Delta x} = \frac{1}{\mu_o c} = \frac{1}{\eta}$, where $c = (\mu_o \varepsilon_o)^{-1/2}$.

Next, an FDTD grid is set up along the x-axis starting with cell #1, ending with cell #53, and a source at cell 50 as indicated in Fig. 15-9. The source is a delta function with amplitude η. As an initial condition, for $n < 1$ all fields in the FDTD grid are taken to be zero. The source turns on at $n = 1$. Thus, dropping the coordinate subscripts on H and E in (15-36) and (15-37), we can write (for $i \geq 50$)

$$E_{50}^{1.0} = \eta \tag{15-38a}$$

$$H_{50.5}^{1.5} = 0 + \frac{1}{\eta}\left(E_{51}^{1.0} - E_{50}^{1.0}\right) = -1 \tag{15-38b}$$

All other $E_i^{1.0}$ and $H_{i+1/2}^{1.5}$ for $i > 50$ are zero since the delta function has not yet propagated to those other locations in the FDTD grid. A word on the notation is in order here. Eq. (15-36) shows $H_y\big|_{i+1/2}^{n+1/2}$ on the left side, whereas (15-38b) shows $H_{50.5}^{1.5}$. In a computer array, there is no location 50.5. There are only the data locations for E_{50} and H_{50}. It is important to understand that $H_{i+\Delta x/2}$ is H_i in the computer. At the next time step, we write (for $i \geq 50$)

$$E_{50}^{2.0} \equiv 0 \tag{15-39a}$$

$$E_{51}^{2.0} = E_{51}^{1.0} + \eta\left(H_{51.5}^{1.5} - H_{50.5}^{1.5}\right) = 0 + \eta(0 + 1) = \eta \tag{15-39b}$$

$$E_{52}^{2.0} = E_{52}^{1.0} + \eta\left(H_{52.5}^{1.5} - H_{51.5}^{1.5}\right) = 0 + \eta(0 - 0) = 0 \tag{15-39c}$$

$$E_{53}^{2.0} = 0 \tag{15-39d}$$

$$H_{50.5}^{2.5} = H_{50.5}^{1.5} + \frac{1}{\eta}\left(E_{51}^{2.0} - E_{50}^{2.0}\right) = -1 + \frac{1}{\eta}(\eta - 0) = 0 \tag{15-39e}$$

$$H_{51.5}^{2.5} = H_{51.5}^{1.5} + \frac{1}{\eta}\left(E_{52}^{2.0} - E_{51}^{2.0}\right) = 0 + \frac{1}{\eta}(0 - \eta) = -1 \tag{15-39f}$$

$$H_{52.5}^{2.5} = H_{52.5}^{1.5} + \frac{1}{\eta}\left(E_{53}^{2.0} - E_{52}^{2.0}\right) = 0 + \frac{1}{\eta}(0 - 0) = 0 \tag{15-39g}$$

$$H_{53.5}^{2.5} = 0 \tag{15-39h}$$

At the third time step, we find that, for $i \geq 50$,

$$E_{50}^{3.0} \equiv 0, \qquad E_{51}^{3.0} = 0 \tag{15-40a}$$

$$E_{52}^{3.0} = E_{52}^{2.0} + \eta\left(H_{52.5}^{2.5} - H_{51.5}^{2.5}\right) = 0 + \eta(0 + 1) = \eta \tag{15-40b}$$

$$E^{3.0}_{53} = 0, \qquad H^{3.5}_{50.5} = 0, \qquad H^{3.5}_{51.5} = 0 \tag{15-40c}$$

$$H^{3.5}_{52.5} = H^{2.5}_{52.5} + \frac{1}{\eta}\left(E^{3.0}_{53} - E^{3.0}_{52}\right) = 0 + \frac{1}{\eta}(0 - \eta) = -1 \tag{15-40d}$$

$$H^{3.5}_{53.5} = 0 \tag{15-40e}$$

At the fourth time step, for $i \geq 50$, we encounter the following:

$$E^{4.0}_{50} \equiv 0, \quad E^{4.0}_{51} = 0, \quad E^{4.0}_{52} = 0 \tag{15-41a}$$

$$E^{4.0}_{53} = E^{3.0}_{53} + \eta\left(H^{3.5}_{53.5} - H^{3.5}_{52.5}\right) = 0 + \eta(0 + 1) = \eta \tag{15-41b}$$

$$H^{4.5}_{50.5} = 0, \qquad H^{4.5}_{51.5} = 0, \quad H^{4.5}_{52.5} = 0 \tag{15-41c}$$

$$H^{4.5}_{53.5} = H^{3.5}_{53.5} + \frac{1}{\eta}\left(E^{4.0}_{54} - E^{4.0}_{53}\right) \tag{15-41d}$$

Here, we have a problem in that $E^{4.0}_{54}$ is undefined. If a computer software could take $E^{4.0}_{54}$ to be zero, then in the computer we would obtain the "correct" value for $E^{4.5}_{53.5}$. However, since the grid was specified to extend only to $i = 53$, we have no reason to expect this will happen. The difficulty can be overcome with an absorbing boundary condition, as discussed in the next section.

Before proceeding to the next section, we observe that in this section we applied the leapfrog time-marching finite difference algorithm equations, and that for our impulsive source, propagation at the speed of light is predicted in the positive x-direction. It is left as an exercise for the reader to show that the same equations will predict propagation in the negative x-direction, and that the right-hand rule for power flow is automatically obeyed.

15.5   ABSORBING BOUNDARY CONDITIONS

At the end of the previous section, it was seen that there was a problem in computing the fields at the edge of the FDTD grid. Without some means of absorbing the outward propagating waves at the extremities of the FDTD grid, nonphysical reflections at the edge of the grid will contaminate the fields inside the grid. Of course, we could terminate the time-stepping procedure before such a reflection reaches the observation area of interest or make the grid very, very large, but these are not computationally viable alternatives. Therefore, some special attention must be given to the problem of updating field components at the edge of the grid.

The most practical solution to updating at the edge of the grid is to employ an *absorbing boundary condition* (ABC), sometimes referred to as a radiation boundary condition (RBC). In the one-dimensional case, the required condition is simple and exact because there is a plane wave normally incident on the edges of the grid. Thus, simple propagation delay can be used. In the two-dimensional and three-dimensional cases, the problem is considerably more difficult because the wave is not likely to be normally incident on the edges of the grid and the waves are not likely to be planar as indicated in Fig. 15-10.

Numerous ABCs have been developed over the past several decades. It is beyond the scope of this book to derive them or even to present more than two of them. Thus, we will present only the Mur ABC [4] and the Berenger perfectly matched layer ABC [5].

There exist two Mur estimates for the fields on the boundary which are first-order- and second-order-accurate. Consider the E_z component located at $x = i\Delta x$, $y = j\Delta y$ for the two-dimensional case. The first-order Mur estimate of this E_z field component is [H10.2: Kunz].
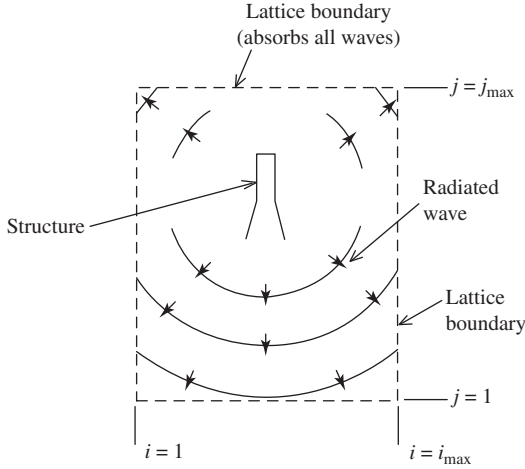
Figure 15-10 FDTD electromagnetic wave interaction region with (ideally) no reflection from the lattice boundary.

$$E_{i,j}^{n+1} = E_{i-1,j}^n + \frac{c\Delta t - \Delta x}{c\Delta t + \Delta x}\left(E_{i-1,j}^{n+1} - E_{i,j}^n\right) \tag{15-42}$$

In the one-dimensional case if $\Delta x = c\Delta t$, $E_i^{n+1} = E_{i-1}^n$, which says that the estimate at location i at the $n+1$ time step is the field from the previous location and previous time step. This is an exact result in one dimension only.

The second-order Mur estimate for E_z in the two-dimensional case, if we assume $\Delta x = \Delta y$, is [2]

$$E_{i,j}^{n+1} = -E_{i-1,j}^{n-1} + \frac{c\Delta t - \Delta x}{c\Delta t + \Delta x}\left(E_{i-1,j}^{n+1} + E_{i,j}^{n-1}\right) + \frac{2\Delta x}{c\Delta t + \Delta x}\left(E_{i-1,j}^n + E_{i,j}^n\right)$$
$$+ \frac{(c\Delta t)^2}{2(\Delta x)(c\Delta t + \Delta x)}\left(E_{i,j+1}^n - 2E_{i,j}^n + E_{i,j-1}^n + E_{i-1,j+1}^n - 2E_{i-1,j}^n + E_{i-1,j-1}^n\right)$$
$$\tag{15-43}$$

In the case of the first-order Mur estimate, we see that the current value of E_z at $x = i\Delta x$ is estimated from the previous E_z value at $x = i\Delta x$ and the value of E_z at $x = (i-1)\Delta x$ at the current time step, both at the same y-position. The second-order Mur estimate uses values from the preceding two time steps, and values at the adjacent x- and y-positions. The equations needed at the $y = j\Delta y$ surface (where the index j is not to be confused with $\sqrt{-1}$) are appropriate permutations of the positional coordinates given in (15-43) above. The second-order Mur estimate is an exact solution for waves impinging normal to a grid boundary. At the intersection of the xz- and yz-planes, some type of first-order-accurate approximation may be employed based on propagation delay as suggested by (15-42).

In 1994, Berenger [5] published a technique that lowered the reflection from the outer grid boundary by several orders of magnitude over other approaches. He called his approach the "perfectly matched layer (PML) for the absorption of electromagnetic waves" in his paper that treated the two-dimensional TE and TM cases. Ingenuously, he artificially split the fields at the boundaries into two components, creating four coupled equations rather than the usual three. This extra degree of freedom permitted Berenger to derive a nonphysical anisotropic absorbing medium, adjacent to the outer boundary (see Fig. 15-11), with a remarkable wave impedance that is independent of the angle of incidence and frequency of the outgoing waves. For the TM case, except in the interface for H_y, the applicable FDTD equations for H_y and E_z are
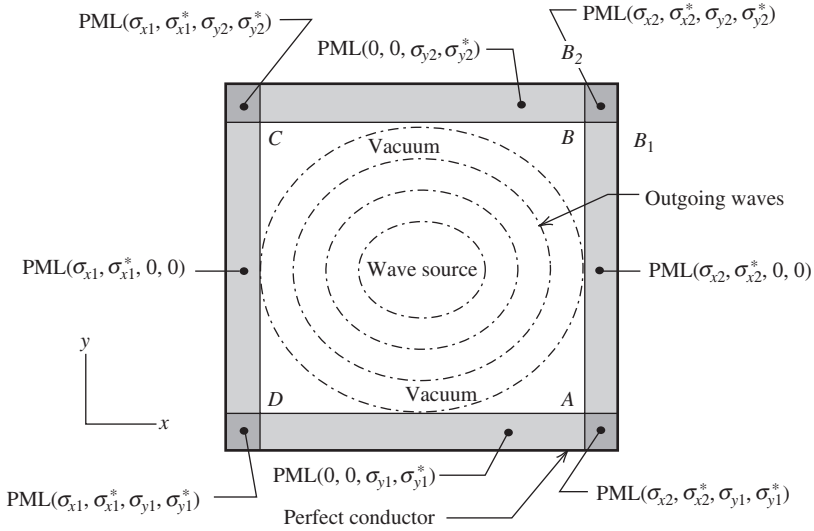
Figure 15-11 Structure of a two-dimensional FDTD grid having the Berenger PML ABC. (J. Berenger, *Comput. Phys.,* Vol. 114, pp. 185–200, 1994. Reprinted with permission *of Computational Physics.*)

$$H_y\big|_{i+1/2,j}^{n+1/2} = e^{-\sigma_x^*(i+1/2)\Delta t/\mu_0}\, H_y\big|_{i+1/2,j}^{n-1/2} - \frac{1 - e^{\sigma_x^*(i+1/2)\Delta t/\mu_0}}{\sigma_x\left(i+\dfrac{1}{2}\right)\Delta x}$$

$$\times \left[E_{zx}\big|_{i+1,j}^{n} + E_{zy}\big|_{i+1,j}^{n} - E_{zx}\big|_{i+j}^{n} - E_{zy}\big|_{i,j}^{n}\right] \tag{15-44}$$

$$E_{zx}\big|_{i,j}^{n+1} = e^{-\sigma_x(i)\Delta t/\varepsilon_0}\, E_{zx}\big|_{i,j}^{n} - \frac{1 - e^{\sigma_x(i)\Delta t/\varepsilon_0}}{\sigma(i)\Delta x} \times \left[H_y\big|_{i+1/2,j}^{n+1/2} - H_y\big|_{i-1/2,j}^{n+1/2}\right] \tag{15-45}$$

where the electric and magnetic conductivities σ_x and σ_x^* are functions of $x(i)$ in the left, right, and corner layers. In the upper and lower PML layers, σ_x and σ_x^* are equal to zero for all $x(i)$, where in fact, the medium behaves as a vacuum for the equations dependent on $x(i)$. Note that E_{zx} and E_{zy} are colocated at the same point.

For an H_y component lying on the interface, the update equation is based on the values of three adjacent E-field components: one E_z component in the regular FDTD grid, and two components, E_{zx} and E_{zy}, in the PML region. Through the application of a normal Maxwell-based FDTD update equation to this H_y component, an interface is established between the regular FDTD grid and PML grid. Here, the assumption is made that the E_{zx} and E_{zy} components are summed to yield an E_z component that is effectively a regular FDTD field component. This is seen to be valid by simply letting $\sigma_x = \sigma_y$ and reducing the four PML update equations to the usual three FDTD governing equations. Thus, the finite difference equations have to be modified. So, in the right side interface normal to x, the equation for H_y above becomes

$$H_y\big|_{il+1/2,j}^{n+1/2} = e^{-\sigma_x^*(il)\Delta t/\mu_0}\, H_y\big|_{il+1/2,j}^{n-1/2} - \frac{1 - e^{\sigma_x^*(il)\Delta t/\mu_0}}{\sigma_x^*(il)\Delta x}$$

$$\times \left[E_{zx}\big|_{il+1,j}^{n} + E_{zy}\big|_{il+1,j}^{n} - E_z\big|_{il,j}^{n}\right] \tag{15-46}$$

Table 15-1 Pulse Propagation with Perfect Absorption ($|E| = \eta$, $|H| = 1$) Spatial Cell Location ($i \geq 50$)

	50	50.5	51	51.5	52	52.5	53	53.5	
Time Step									
1	η		0		0		0		E
1.5		−1		0		0		0	H
2	0		η		0		0		E
2.5		0		−1		0		0	H
3	0		0		η		0		E
3.5		0		0		−1		0	H
4	0		0		0		η		E
4.5		0		0		0		−1	H
5	0		0		0		0		E
5.5		0		0		0		0	H

This particular update is applied at the right-side interface normal to x, where il indicates the inner boundary of the PML.

The discretized equations needed for the components H_x and E_{zy} are left as an exercise for the reader. Not surprisingly, the two-dimensional PML equations reduce to the exact one-dimensional result. This is also left as an exercise for the reader. A Maxwellian derivation of the PML technique may be found in [8].

Of the two ABCs presented here, Mur and Berenger, the former is somewhat simpler to implement but the latter offers substantially lower reflection characteristics, perhaps more than is really necessary for many applications. One general consideration with ABCs is that of determining the necessary distance from the antenna or scatterer to the outer boundary where the ABC is applied. The greater this distance, the more effective the ABC tends to be. (This is particularly true with the Mur ABC but not necessarily so with the PML.) The increased effectiveness of the ABC is due to the more plane-like nature of the outward traveling wavefront as the distance becomes large. A common criterion is a minimum of 10 cells between the antenna or scatterer and the outer boundary, with 15 to 20 being preferred for the Mur ABC and as few as 4 or 5 for the PML.

More recently, improvements on the PML have been developed. Two of these are the UPML and the CPML. The UPML is the *uniaxial* PML first applied to FDTD by Gedney. [6, 7] It is based on a Maxwellian formulation [H.10.3: Taflove, 3rd ed., p. 274] rather than on a mathematical model. As a consequence, there are been attempts to physically realize a UPML. The development of the UPML was followed by the CPML, which is a *convolutional* PML. [8] The CPML is more accurate than the classical UPML, more efficient, and better suited for application to domains with generalized materials. It has been reported that the CPML reduces errors due to reflections from the absorbing boundary by better than 20 dB relative to the UPML [H.10.3: Taflove, 3rd ed., p. 324].

Before we leave this section, let us apply the one-dimensional exact result to (15-41d). In this case, the exact result is $E_{54}^{4.0} = E_{53}^{3.0} = 0$. Thus,

$$H_{53.5}^{4.5} = 0 + \frac{1}{\eta}(0 - \eta) = -1 \tag{15-47}$$

Then, to evaluate the reflection from the edge of the grid,

$$E_{53}^{5.0} = E_{53}^{4.0} + \eta\left(H_{53.5}^{4.5} - H_{52.5}^{4.5}\right) = \eta + \eta(-1 - 0) = 0 \tag{15-48a}$$

$$H_{52.5}^{5.5} = H_{52.5}^{4.5} + \frac{1}{\eta}\left(E_{53}^{5.0} - E_{52}^{5.0}\right) = 0 + \frac{1}{\eta}(0 - 0) = 0 \tag{15-48b}$$

We see that there is perfect absorption. Table 15-1 summarizes the situation for $1 \leq n < 6$ and $i \geq 50$. Note that all points in the grid are calculated at all time steps. The grid used for Table 15-1 ends on an H-field calculation. The example could also be developed to end on an E-field calculation at E_{53} by defining the H-field only out to $H_{52.5}$. In this case, $H_{53.5}^{3.5}$ would be undefined in (15-41b) and the one-dimensional exact absorption condition would be applied to E_{53}. For simplicity, an ABC is typically applied to only E- or only H-fields. In this way, regular FDTD update equations will govern all behavior for one type of field (say, H) and special updates then need only be applied at the grid boundary to the other type of field (say, E). Note that the Mur equations are given here for the electric field.

15.6 SOURCE CONDITIONS

In this section, we introduce into the FDTD lattice several electromagnetic wave excitations appropriate for modeling engineering problems. An excitation of interest is the linearly polarized plane wave propagating in free space for use in scattering analysis, but we are also interested in waves radiated by antennas. With FDTD, we usually study antennas in the transmitting mode since it is not computationally efficient to do so in the receiving mode.

This section covers source conditions for antennas and scatterers. Following the designations of Taflove, the sources will be classified as either "hard" or "soft." A hard source forces a field quantity to a value independent of neighboring fields, which means that the update equations are not allowed to update the field(s) at the source location (e.g., a metallic monopole near a scatterer). A soft source does permit the fields to be updated at the source location(s) (e.g., a plane wave injected into the grid).

15.6.1 Source Functionality

A common source is one that generates a continuous sinusoidal wave of frequency f_o that is switched on at $n = 0$:

$$f(t) = E_o \sin(2\pi f_o n \Delta t) \qquad (15\text{-}49a)$$

A second source provides a wideband Gaussian pulse with finite dc spectral content that is centered at time step n_o and has a $1/e$ characteristic decay of n_{decay} time steps:

$$f(t) = E_o e^{-[(n-n_o)/n_{\text{decay}}]^2} \qquad (15\text{-}49b)$$

Note that (15-49b) has a nonzero value at $n = 0$, so that if a smooth transition from zero into the Gaussian pulse is required, n_o should be taken as at least $3n_{\text{decay}}$. A third source that provides a zero-dc content is a sine modulated (bandpass) Gaussian pulse with a Fourier spectrum symmetrical about f_o. The pulse is again centered at time step n_o and has a $1/e$ characteristic decay of n_{decay} time steps:

$$f(t) = E_o e^{-[(n-n_o)/n_{\text{decay}}]^2} \sin[2\pi f_o(n - n_o)\Delta t] \qquad (15\text{-}49c)$$

Each source of (15-49) radiates a numerical wave having a time waveform corresponding to the source function $f(t)$. The numerical wave propagates symmetrically in all directions from the source point at i_s. If a material structure is specified at some distance from the source point, the radiated numerical wave eventually propagates to this structure and undergoes partial transmission and partial reflection. In principle, time-stepping can be continued until all transients decay. For the source of (15-49a), this would mean the
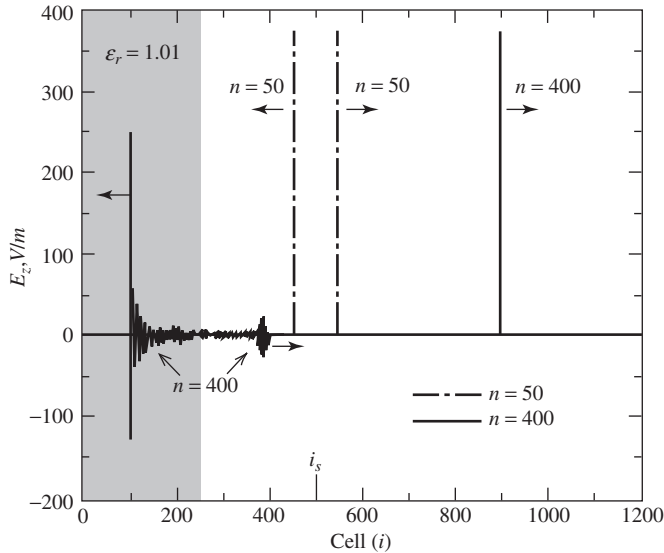
Figure 15-12 Delta wave function at $n = 50$ and 400 with $\varepsilon_r = 1.01$ for $0 \leq i \leq 250$ and $i_s = 500$. Source magnitude is 377 V/m.

attainment of the sinusoidal steady state for the transmitted and reflected fields. For the sources of (15-49b) and (15-49c), this would mean the evolution of the complete time histories of the transmitted and reflected waves. Discrete Fourier analysis of these time histories obtained in a single FDTD run can provide the magnitude and phase of the transmission and reflection coefficients over a potentially wide frequency band starting at dc.

Thus far, we have discussed three time functions used in FDTD work. The delta function used in Secs. 15.4 and 15.5 is not a generally useful time function for FDTD calculations because of its theoretically infinite bandwidth. The delta function was used in Secs. 15.4 and 15.5 because it provided a simple way of illustrating how the update equations worked, how the one-dimensional absorbing boundary condition worked, and even permitted a small number of calculations to be easily done by hand. Unfortunately, the delta function hard source will only work in one dimension, and then only when propagation is exactly one cell per time step (e.g., the magic time step). Fig. 15-12 illustrates the consequences of violating this condition. Consider Fig. 15-9 but with the grid extending from zero to $i\Delta x = 1200\Delta x$, with a delta function source of amplitude η at $i = 500$ and the region from $i = 0$ to $i = 250$ filled with a relative permittivity of only 1.01. Fig. 15-12 shows the electric field in the grid after 50 and 400 time steps. Note that the wave encounters the small change in dielectric constant at $n = 250$ and there is obvious evidence of numerical dispersion for $n = 400$ in the region $i < 400$ as a consequence of the (slight) violation of the condition one cell in one time step for $i < 250$ and $n > 250$. On the other hand, at $n = 400$, $i = 900$ the delta function propagates to the right in a dispersionless manner since no violation of the one cell in one time step condition has occurred.

15.6.2   The Hard Source

The hard source is set up simply by assigning a desired time function to specific electric or magnetic field components in the FDTD lattice as we did in the example in Sec. 15.4. For example, in a one-dimensional grid, the following hard source on E_z could be established at the grid source point i_s to generate a continuous sinusoid that is switched on at $n = 0$:

$$E_z\big|_{i_s}^{n} = f(t) = E_o \sin\left(2\pi f_o n\Delta t\right) \tag{15-50}$$

Note that the electric field at i_s is forced to have a value determined entirely by the source and it is independent of the update equation.

There are some difficulties with the hard source scenario. As time-stepping is continued to obtain either the sinusoidal steady state or the late-time pulse response, we note that the reflected numerical wave eventually returns to the source grid location i_s. Because the total electric field is specific at i_s without regard to any possible reflected waves in the grid (hence the terminology, "hard source"), the hard source causes a retro-reflection of these waves at i_s back toward the material structure of interest. In effect, it prevents the movement of reflected wave energy through its position toward infinity, and thereby may fail to properly simulate the true physical situation.

15.6.3   The Soft Source

A simple way to mitigate the reflective nature of a hard source is to allow a new value of the electric field at the source location i_s to equal the update value plus the value of an impressed electric field described by the time function $f(t)$. For our one-dimensional example, this means that

$$E_z\big|_{i_s}^n = E_z\big|_{i_s}^{n-1} + \frac{\Delta t}{\varepsilon_o \Delta x}\left(H_y\big|_{i_s+1/2}^{n-1/2} - H_y\big|_{i_s-1/2}^{n-1/2}\right) + f(t) \tag{15-51}$$

where $f(t)$ can be obtained, for example, from (15-49). The relationship in (15-51) is conceptually similar to that of the resistive voltage source in [H.10.3: Taflove, 1st ed., Chap. 6, p. 459].

Fig. 15-13 illustrates the difference between one-dimensional hard and soft sources. The FDTD model has cells from $i = 0$ to $i = 1200$ with a source at $i = 500$ and a dielectric with $\varepsilon_r = 9$ in cells 1 to 200. The time function is a Gaussian pulse as in (l5-49b). The (a) and (b) parts of the figure apply whether the source is hard or soft, whereas the (c) and (d) parts only apply to the soft source and the (e) and (f) to the hard source. For $n = 600$, 700, and 800, the differences between the hard and soft sources are apparent. Note the effect of the ABC at $i = 1200$.

15.6.4   Total-Field/Scattered-Field Formulation

The total-field/scattered-field FDTD formulation [H.10.3: Taflove, Chap. 6] (see Fig. 15-14) finds its greatest use in simulating plane wave illumination. This approach is based on the linearity of Maxwell's equations and the usual decomposition of the total electric and magnetic fields into incident fields ($\mathbf{E}^i$, $\mathbf{H}^i$) and scattered fields ($\mathbf{E}^s$, $\mathbf{H}^s$). $\mathbf{E}^i$ and $\mathbf{H}^i$ are the values of the incident fields that are *known* at all points of an FDTD grid at all time steps. These are the field values that would exist if there were no materials of any sort in the modeling as Fig. 15-15 shows for the fields on two separate grids, one being used for the incident field and the second for the total field-scattered field. The incident field only exists in the total field portion ($100 < i < 1100$) of the total field grid ($0 < i < 1200$).

In Fig. 15-15, the total field grid is free space only. At time step $n = 350$, the incident field grid is seen to be propagating a sine wave. For $i > 100$ the incident field is subtracted from the (zero) total field leaving only the scattered field (which is equal to the incident field) in the total field grid. The scattered field region of the total field grid (dotted line in Fig. 15-15) is exactly equal to zero, as one would expect with no scatterer present. At time step $n = 650$, the scattered field for $i < 100$ remains zero, while for $i > 1100$ the wave has propagated through, and is subtracted out, leaving only zero scattered field in this region as well for this example with no scatterer present.

The finite difference approximations of the Yee algorithm can be applied with equal validity to either the incident-field vector components, the scattered-field vector components, or the total-field vector components. FDTD codes can utilize this property to
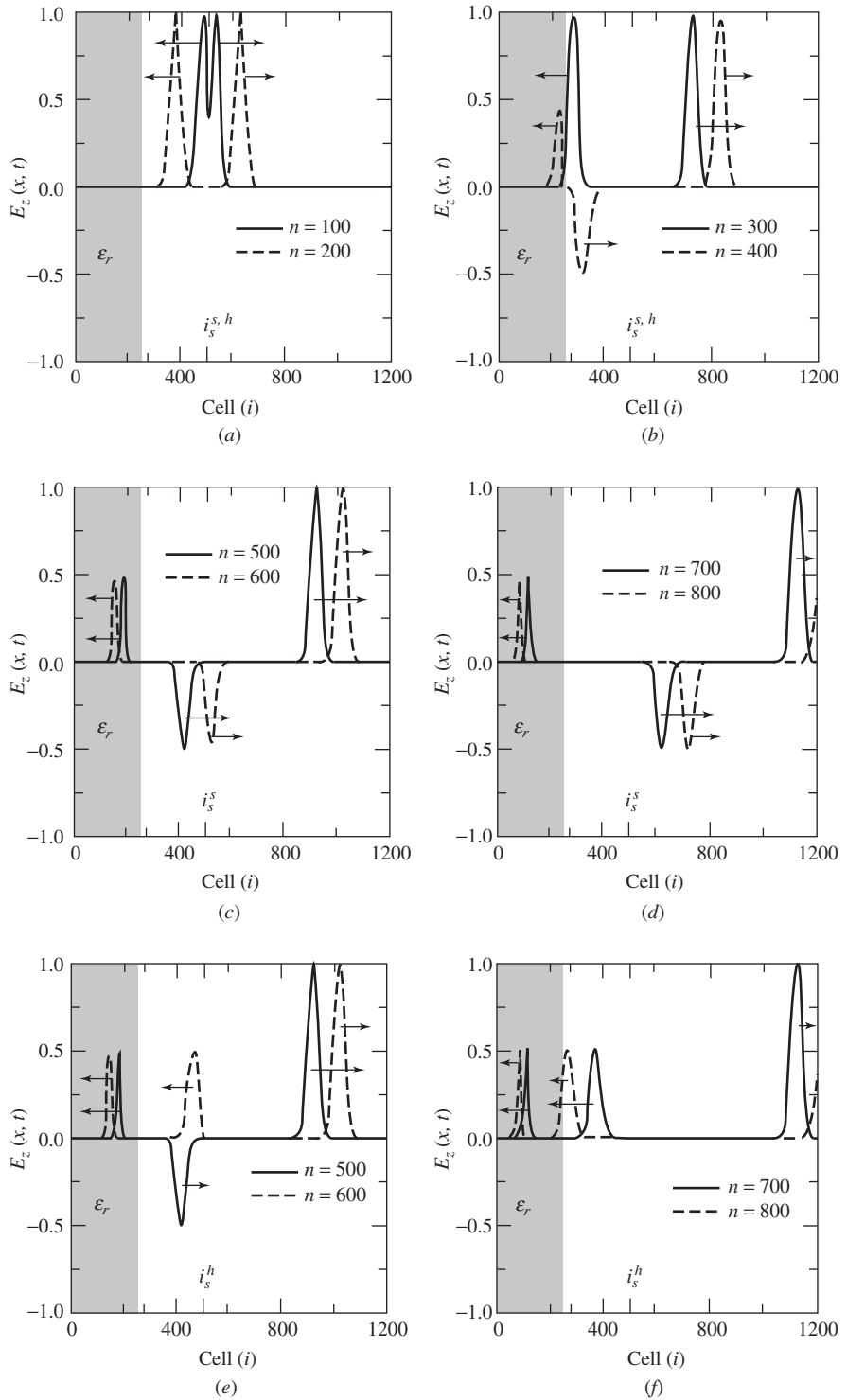
Figure 15-13 Gaussian pulse electric field at $n = 100, 200, 300, 400, 500, 600, 700$, and 800. Parts $(a-d)$ are for a soft source at $i_s = 500$. Parts (a), (b), (e), and (f) are for a hard source at $i_s = 500$. Cells from $i = 0$ to $i = 200$ contain $\varepsilon_r = 9$.
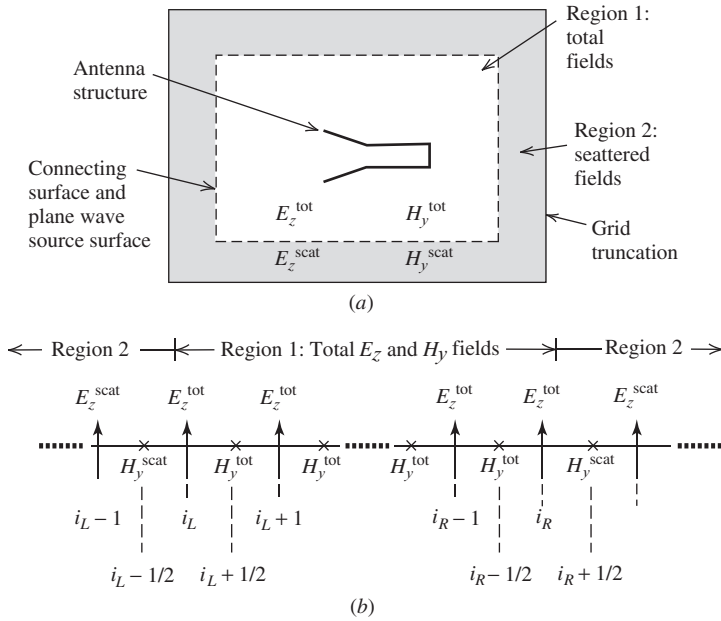
Figure 15-14 Zoning of the FDTD grid, (*a*) Total-field and scattered-field regions, connecting surface/plane wave source, and lattice truncation (ABC). (*b*) Detail of field component locations in a one-dimensional horizontal cut through the grid of (*a*).
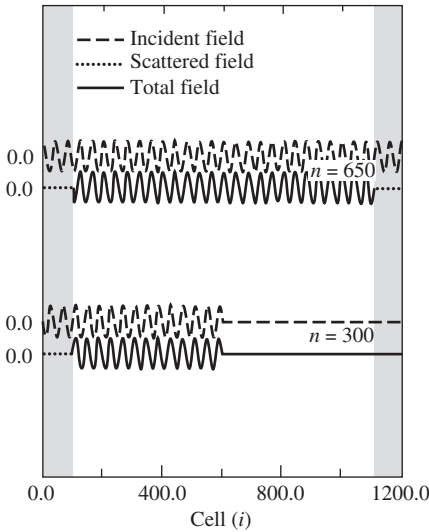
Figure 15-15 One-dimensional free-space grid from $i = 0$ to $i = 1200$. Total field region is from $i = 100$ to $i = 1100$. 40 cells per wavelength with a 10-GHz sine wave propagating from left to right. *E*-fields shown at $n = 300$, and $n = 650$. The scattered field region is shaded grey.

zone the numerical space lattice into two distinct regions separated by a nonphysical surface that serves to connect the fields in each region as shown in Fig. 15-14.

Region 1, the inner zone of the lattice, is denoted as the *total-field* region where the Yee algorithm operates on total-field vector components. The interacting structure of interest is embedded within this region.

Region 2, the outer zone of the lattice, is denoted as the *scattered-field* region where the Yee algorithm operates only on scattered-field vector components. This implies that
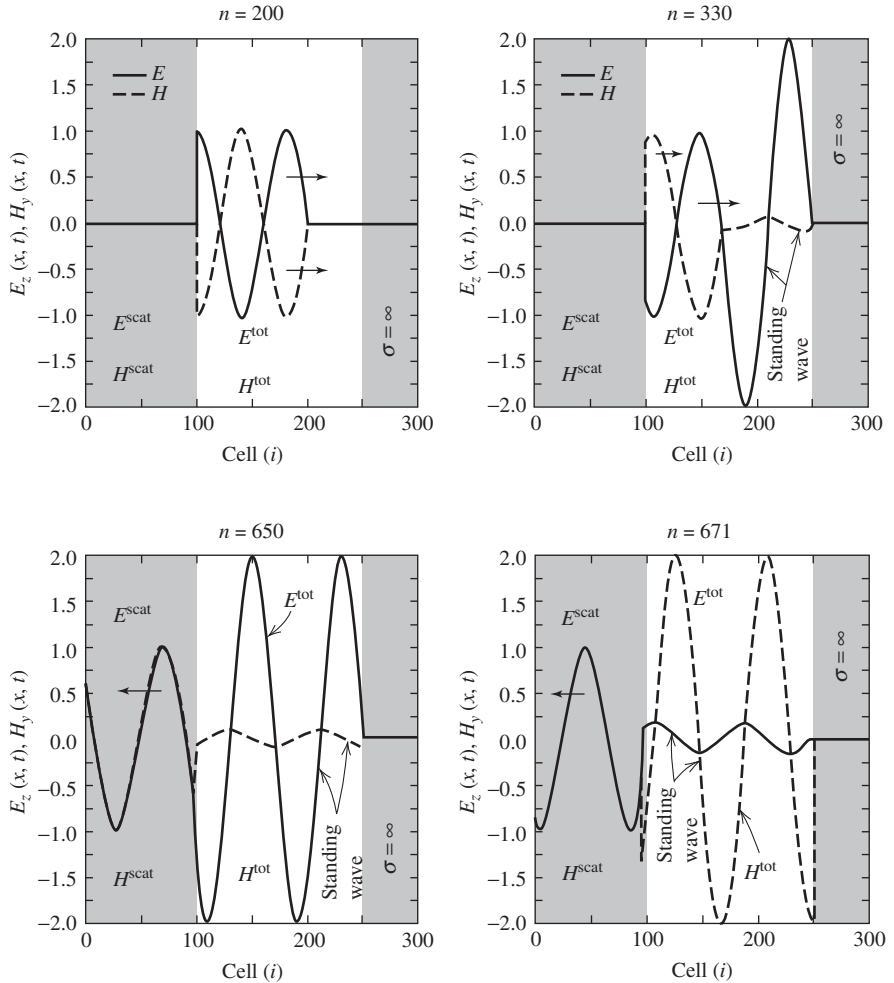
Figure 15-16 One-dimensional free-space grid from $i = 0$ to $i = 249$ with a PEC from $i = 250$ to $i = 300$. Total field region is from $i = 100$ to $i = 300$. 40 cells per wavelength with a 10-GHz sine wave incident from the left. Normalized E- and H-fields shown at $i = 200$, 330, 650, and 671.

there is no incident wave in Region 2 as illustrated by Fig. 15-16 for $i < 100$ and $n \leq 330$. For $n = 200$, the total field is the incident field (E and H considered to be 180° out of time phase), whereas for $n = 330$ reflection has occurred from the perfect electric conductor (PEC) and some of the total field is just the incident field and some is a standing wave of the incident and scattered (reflected) fields (E and H 90° out of time phase). In Fig. 15-16, for $n = 650$ and $n = 671$ there is a scattered field propagating to the left for $i < 100$ (E and H are in time phase) with a standing wave for $100 < i < 250$, with E always zero at $i = 250$ and H at times reaching a maximum value of 2 at the PEC.

To illustrate how the total-field/scattered-field formulation is implemented, consider the one-dimensional case. The nonphysical surface constituting the interface of Regions 1 and 2 contains **E** and **H** components that obviously require the formulation of various field component spatial differences in the update equations. When a spatial difference is taken across the interface plane, a problem of consistency arises. That is, on the Region 1 side of the interface, the field in the difference expression is assumed to be a total field,

whereas on the Region 2 side of the interface, the field in the difference expression is assumed to be a scattered field. It is inconsistent to perform an arithmetic difference between scattered- and total-field values.

This problem of consistency can be solved by using the values of the components of the incident-field vectors[3] $\mathbf{E}^{\text{inc}}$ and $\mathbf{H}^{\text{inc}}$, which are assumed to be known or calculable at each space lattice point. As illustrated in Fig. 15-14b, let the left interface between scattered-field and total-field zones be positioned between E_z at i_L and H_y at $i_{L-1/2}$. From this arrangement, it is clear that E_z is a total-field component. We then write

$$E_z^{\text{tot}}\big|_{i_L}^{n+1} = E_z^{\text{tot}}\big|_{i_L}^{n} + \frac{\Delta t}{\varepsilon_o \Delta x}\left(H_y^{\text{tot}}\big|_{i_L+1/2}^{n+1/2} - H_y^{\text{scat}}\big|_{i_L-1/2}^{n+1/2}\right) - \frac{\Delta t}{\varepsilon_o \Delta x} H_y^{\text{inc}}\big|_{i_L-1/2}^{n+1/2} \qquad (15\text{-}52)$$

where the right-most term corrects the problem of inconsistency since

$$-H_y^{\text{scat}}\big|_{i_L-1/2}^{n+1/2} - H_y^{\text{inc}}\big|_{i_L-1/2}^{n+1/2} = -H_y^{\text{tot}}\big|_{i_L-1/2}^{n+1/2} \qquad (15\text{-}53)$$

Similarly at grid point $(i_L - \frac{1}{2})$, we write

$$H_y^{\text{scat}}\big|_{i_L-1/2}^{n+1/2} = H_y^{\text{scat}}\big|_{i_L-1/2}^{n-1/2} + \frac{\Delta t}{\mu_o \Delta x}\left(E_z^{\text{tot}}\big|_{i_L}^{n} - E_z^{\text{scat}}\big|_{i_L-1}^{n}\right) - \frac{\Delta t}{\mu_o \Delta x} E_z^{\text{inc}}\big|_{i_L}^{n} \qquad (15\text{-}54)$$

where the right-most term corrects the problem of inconsistency since

$$E_z^{\text{scat}}\big|_{i_L}^{n} = E_z^{\text{tot}}\big|_{i_L}^{n} - E_z^{\text{inc}}\big|_{i_L}^{n} \qquad (15\text{-}55)$$

A similar procedure is carried out at the right-hand-side total-field/scattered-field interface. Let the right interface between scattered-field and total-field zones be positioned exactly at an E_z component at grid point i_R, and further assume that this E_z is a total-field component. The electric field expression analogous to (15-52) is

$$E_z^{\text{tot}}\big|_{i_R}^{n+1} = E_z^{\text{tot}}\big|_{i_R}^{n} + \frac{\Delta t}{\varepsilon_o \Delta x}\left(H_y^{\text{scat}}\big|_{i_R+1/2}^{n+1/2} - H_y^{\text{tot}}\big|_{i_R-1/2}^{n+1/2}\right) + \frac{\Delta t}{\varepsilon_o \Delta x} H_y^{\text{inc}}\big|_{i_R+1/2}^{n+1/2} \qquad (15\text{-}56)$$

The magnetic field expression analogous to (15-54) is

$$H_y^{\text{scat}}\big|_{i_R-1/2}^{n+1/2} = H_y^{\text{scat}}\big|_{i_R+1/2}^{n-1/2} + \frac{\Delta t}{\mu_o \Delta x}\left(E_z^{\text{scat}}\big|_{i_R+1}^{n} - E_z^{\text{tot}}\big|_{i_R}^{n}\right) + \frac{\Delta t}{\mu_o \Delta x} E_z^{\text{inc}}\big|_{i_R}^{n} \qquad (15\text{-}57)$$

The important effect of this procedure is to generate a plane wave at the left-hand scattered-field/total-field interface point i_L, propagate it through the total-field region to the right-hand total-field/scattered-field interface point i_R, and then cancel it out in the right-hand scattered-field region. In the absence of a scattering object in the central total-field zone, there are zero fields present in the scattered-field regions to the left and right of the center zone as is the case in Fig. 15-15.

15.6.5 Pure Scattered-Field Formulation

The pure scattered-field formulation borrows from a method popular with the frequency domain integral equation (i.e., MoM) community. Again, the concept evolves from the linearity of Maxwell's equations and the decomposition of the total electric and magnetic

---

[3] To avoid confusion with the index i, "inc" is used to denote the incident field. For further consistency, "scat" denotes the scattered field and "tot" the total field.

fields into a known incident field and an unknown scattered field. Here, however, the FDTD method is used to time step only the scattered electric and magnetic fields. That is, the FDTD grid is not segmented into total-field and scattered-field regions, but instead assumes scattered-field quantities everywhere. This is the case for (transmitting) antenna analysis where the scattered field is thought of as the radiation field. The scattered (radiation) field is, however, a near field since it is not practical to extend the grid to the far-field. To obtain a far-field radiation pattern, it is necessary to transform data in the near field to the far-field as discussed in the next section.

15.7 NEAR FIELDS AND FAR FIELDS

As we have implied earlier, it is not practical to directly calculate far-field FDTD data within the FDTD grid because for most problems, the grid space cannot be made large enough to include the far field. Thus, near-field data must be transformed into far-field data. The existence of a well-defined scattered-field region in the FDTD lattice, as described in the previous section, facilitates a near-to-far-field transformation that is discussed here. If we use the near-field data stored in a single FDTD modeling run, this transformation efficiently and accurately calculates the complete radiation pattern of an antenna or the complete far-field bistatic scattering response of an illuminated structure for a single illumination angle.

To begin developing the near-to-far-field transformation, refer to Fig. 15-17, where a rectangular virtual surface S_{ab} fully enclosing the scatterer (region B) is located in the scattered-field region (region 2 of Fig. 15-17) near the lattice boundary. The tangential components of the scattered fields $\mathbf{E}^s$ and $\mathbf{H}^s$ are first obtained at S_{ab} using FDTD. Then, as indicated in Fig. 15-17, an equivalent problem is set up that is completely valid for Region A, external to S_{ab}. The new excitation data are $\mathbf{J}_s$ and $\mathbf{M}_s$, the equivalent surface electric and magnetic currents, respectively, on S_{ab} that are obtained according to (see Sec. 9.1).

$$\mathbf{J}_s(\mathbf{r}) = \hat{\mathbf{n}} \times \mathbf{H}^s(\mathbf{r}) \tag{15-58a}$$

$$\mathbf{M}_s(\mathbf{r}) = -\hat{\mathbf{n}} \times \mathbf{E}^s(\mathbf{r}) \tag{15-58b}$$

where $\hat{\mathbf{n}}$ is the outward unit normal vector at the surface S_{ab}.

The scattered far fields are then given by the integration of the equivalent currents of (15-58a) and (15-58b). If (μ_o, ε_o) are the region A medium characteristics, then the following scattered far-field expressions for θ and ϕ polarizations are obtained:

$$E_\theta = -j\omega \left[A_\theta + \eta F_\phi \right] \tag{15-59a}$$

$$E_\phi = -j\omega \left[A_\phi - \eta F_\theta \right] \tag{15-59b}$$
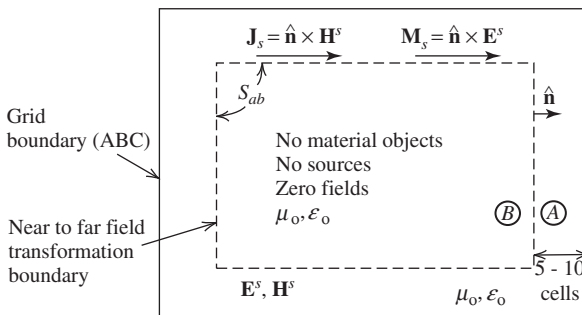
Figure 15-17 Electromagnetic equivalence to transform near fields to far fields.

where

$$A_\theta = A_x \cos\theta \cos\phi + A_y \cos\theta \sin\phi - A_z \sin\theta \qquad (15\text{-}59c)$$

$$F_\theta = F_x \cos\theta \cos\phi + F_y \cos\theta \sin\phi - F_z \sin\theta \qquad (15\text{-}59d)$$

$$A_\phi = -A_x \sin\phi + A_y \cos\phi \qquad (15\text{-}59e)$$

$$F_\phi = -F_x \sin\phi + F_y \cos\phi \qquad (15\text{-}59f)$$

and the potentials in the far-field region are given by

$$\begin{bmatrix} \mathbf{A} \\ \mathbf{F} \end{bmatrix} = \left(\frac{e^{-j\beta r}}{4\pi r}\right) \iint \begin{bmatrix} \mu_o \mathbf{J}_s \\ \varepsilon_o \mathbf{M}_s \end{bmatrix} e^{j\beta r' \cos\xi} \, dS'_{ab} \qquad (15\text{-}60a)$$

$$r' \cos\xi = (x' \cos\phi + y' \sin\phi) \sin\theta + z' \cos\theta \qquad (15\text{-}60b)$$

This approach to computing the far scattered fields is straightforward because (1) the near-field data for arbitrary antennas or scatterers can be obtained from the FDTD calculations themselves and (2) the transformation of the near-field data to the far field is independent of the nature of the scatterer that resides within the integration surface S_{ab}.

Early FDTD calculations of far-zone scattered fields used sinusoidal excitation. Because of this, the FDTD far-zone results were obtained at only one frequency per FDTD calculation run. The procedure for such single-frequency far-zone calculations is straightforward. First, the FDTD calculations are stepped through time until steady-state conditions are reached. Then the complex time-harmonic electric and magnetic currents flowing on a closed surface surrounding the object are obtained. If these time-harmonic fields or currents are stored, then during postprocessing the far-zone radiated or scattered fields can be calculated in any desired direction. This approach is particularly suited to far-zone radiation or scattering patterns at only a single frequency.

To obtain far-zone results at multiple frequencies, the approach is to use pulsed excitations for the FDTD calculations. For each frequency of interest, a running discrete Fourier transform (DFT) of the time-harmonic surface currents on a closed surface surrounding the FDTD geometry is updated at each time step. The running DFT provides the complex frequency domain currents for any number of frequencies when using pulse excitation for the FDTD calculation. This is much more efficient than using a time-harmonic excitation for each frequency of interest. It requires no more computer storage per frequency for the surface currents than when sinusoidal excitation is used and provides frequency domain far-zone fields at any far-zone angle. If far-zone results are desired at several frequencies, then the running DFT approach is the better choice.

Before leaving this section, we should mention that the near-field data itself may be of interest. Near-field data are readily obtained by selecting appropriate field values directly from the FDTD grid. Data can include instantaneous fields, phasor fields obtained via Fourier transformation of the instantaneous fields, scalar or vector-interpolated field maps. The near-field radiation pattern of an antenna is simply the spatial distribution of the FDTD computed radiated fields in the vicinity of the antenna. Near-fields provide insight into basic physical interactions such as reflection and diffraction. Near-field data can also be used to determine, for example, magnitude and phase data across an antenna aperture (as in the next section), surface current densities on an antenna, and current or field distributions in or along an antenna feed.

15.8 A TWO-DIMENSIONAL EXAMPLE: AN *E*–PLANE SECTORAL HORN ANTENNA

The previous sections presented all the basics needed to do a problem from grid layout to computation of the far field. This section considers a two-dimensional TE example

problem in detail. The problem is a two-dimensional model of an *E*-plane sectoral horn as illustrated in Fig. 9-15. The horn was chosen according to the optimum condition $B = \sqrt{2\lambda R_2}$ with $R_2 = 8\lambda$ so that $B = 4\lambda$. (See Sec. 9.4.2.) The resulting FDTD model of the horn is illustrated in Fig. 15-18. Notice that the walls of the horn are "stepped." To see why this stepping comes about, examine the two-dimensional grid in Fig. 15-19. At the walls of the horn and waveguide, the coefficients C_a and C_b are calculated with a high value of conductivity (e.g., $5.7 \times 10^7 \text{siemens/m}$). Since the cells are square, a stepped contour naturally results.

The two-dimensional grid used for this example was 260×200 cells as indicated in Fig. 15-18. The cells are $\lambda/20$ by $\lambda/20$ at the center frequency. The boundary for the near-to-far-field transformation is taken to be 12 cells inside the extremities of the grid. The time function chosen is the sine modulated Gaussian pulse of (15-49c) expressed as

$$ f(t) = 1.148 e^{-[(t-3\tau_o)/\tau_o]^2} \, \sin\left[2\pi f_c (t - 3\tau_o)\right] \tag{15-61} $$

where
$$ \tau_o = 1.0/[\pi(f_h - f_c)] $$
$$ f_h = 15 \ \text{GHz} $$
$$ f_c = 10 \ \text{GHz} $$
$$ \Delta x = \lambda/20 \text{ at the center frequency.} $$
$$ 2\Delta t = \Delta x/2.99792458 \times 10^8 $$
$$ t = n\Delta t, \quad \Delta t = 2.5 \times 10^{-12} $$

The plot of this function is shown in Fig. 15-20. A soft source with this time function is located $\lambda/4$ from the back wall of the waveguide at the center frequency.

(200, 0)   (200, 260)

(0, 0)   (0, 260)
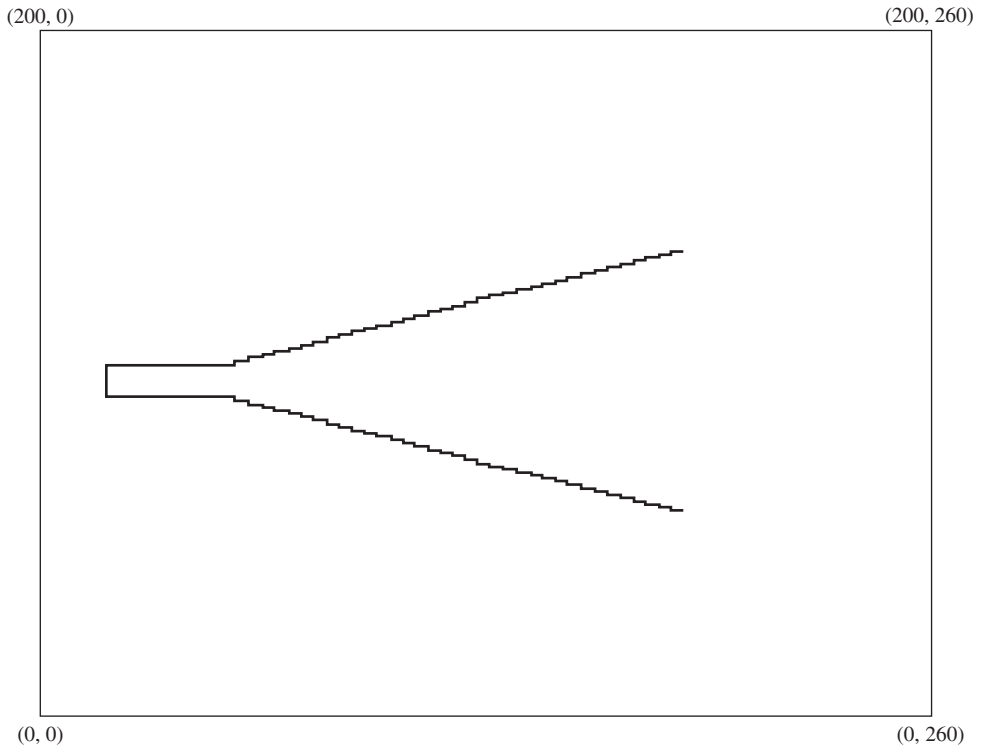
Figure 15-18 TE model of an *E*-plane sectoral horn.
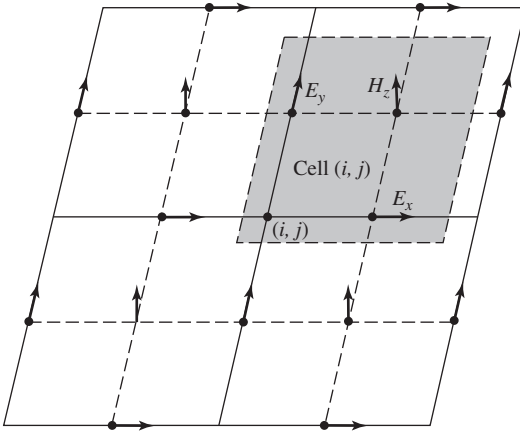
Figure 15-19 Perspective view of a two-dimensional TE grid.
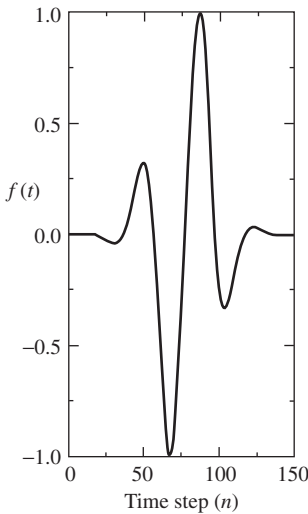
Figure 15-20 Sine modulated Gaussian waveform.

Fig. 15-21a shows the fields in the horn at the 250th time step. The pulse has not yet reached the aperture and there is evidence of some reflection at the discontinuity where the walls of the horn join with the waveguide walls. In Fig. 15-21b, where $n = 475$, the peak energy content of the pulse has passed just beyond the aperture and there is some evidence of energy reflected from the aperture traveling back toward the waveguide. Diffraction at the edges of the horn is in clear evidence. Time steps $n = 600$ and 800 are presented in the previous edition of this book on p. 528. They show double diffraction effects at $n = 600$ and 800 and diffraction by a wedge whose interior angle is greater than π (i.e., where the horn itself joins the waveguide) at $n = 800$.

In Figs. 15-21$a-b$, the darkest areas indicate highest levels of field intensity and the white areas indicate zero field intensity. Even within the pulse itself there are several instants in time when the pulse is zero (see Fig. 15-20). These times of zero field intensity are the thin white lines within the main pulse and its diffractions or reflections in Fig. 15-21. Large almost-white areas are evidence of numerical noise.

Fig. 15-22 shows the amplitude and phase distributions across the aperture at 9,10, and 11 GHz. These are obtained from a Fourier transform of the fields at the aperture. From

(0, 200)   (200, 260)

(0, 0)   (0, 260)

(*a*) At *n* = 250

(0, 200)   (200, 260)

(0, 0)   (0, 260)

(*b*) At *n* = 475
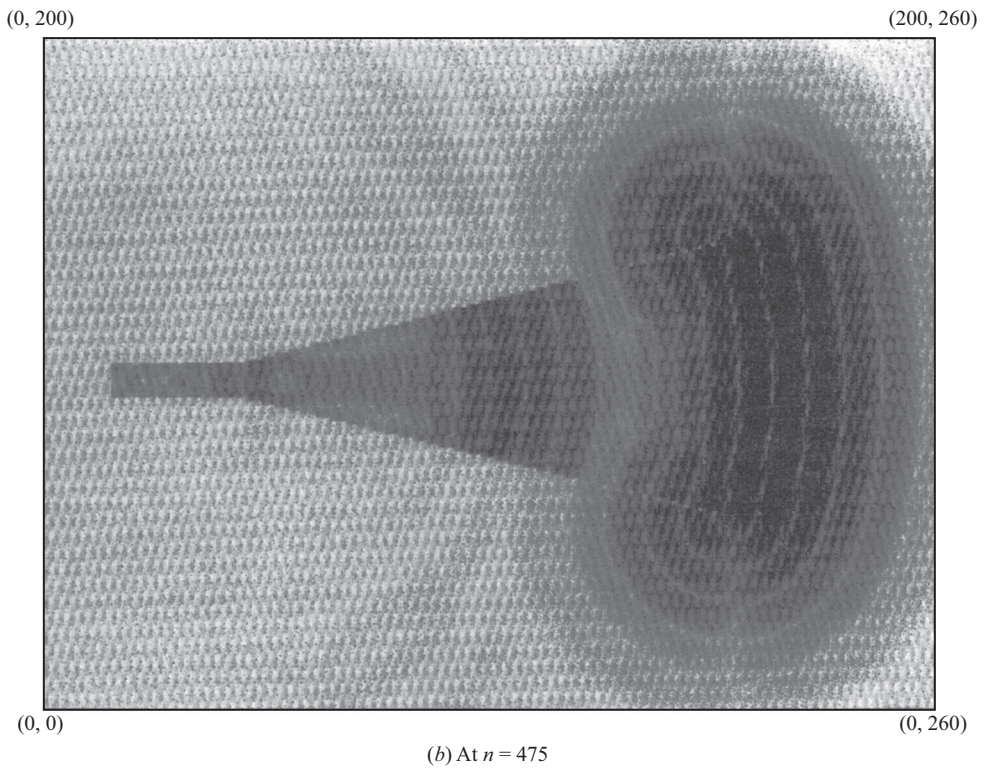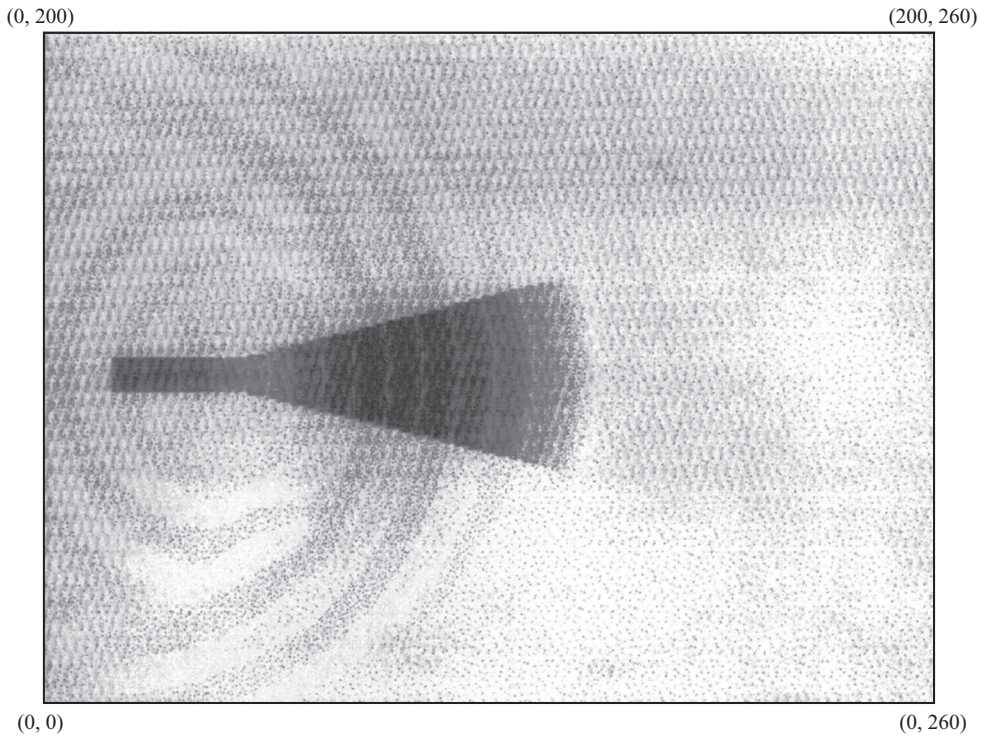
Figure 15-21 Pulse propagation at *n* = 250, 475, 600, 800.
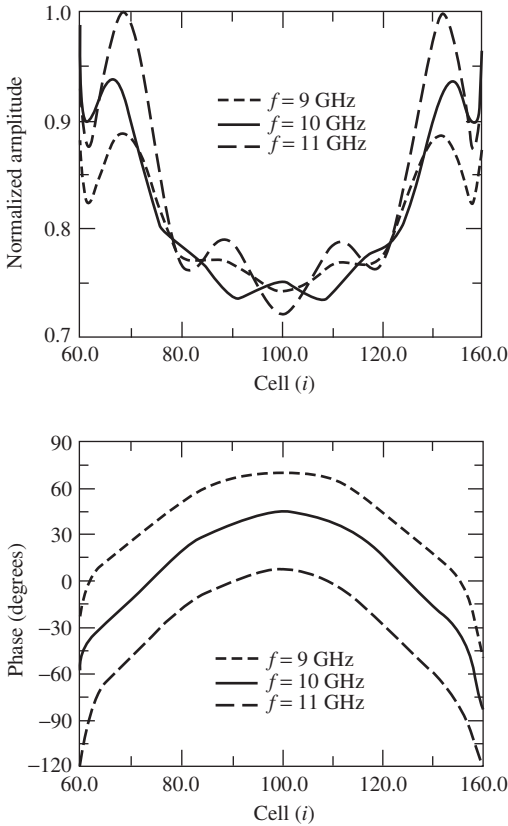
Figure 15-22 Horn aperture field distribution at $n = 800$.

our design condition for this horn, $B = \sqrt{2\lambda R_2}$, a $90°$ phase change is expected from the center of the aperture to the edge. The FDTD results are nearly in agreement with this if allowance is made for the diffraction effects near the edges of the aperture that are not included in the classical analysis of Chap. 9, but are included in the nearly exact FDTD results. From the amplitude distribution near the edges of the aperture in Fig. 15-22 as well as the more rapid change in phase there, it is apparent that the electromagnetic wave has a strong interaction with the edges of the horn.

Fig. 15-23 shows the magnitude of the electric field along the transformation boundary that partly contributes to the near-to-far-field transformation; the magnetic field accounts for the remainder. The field is evaluated at $n = 800$. Not surprisingly, the field is strongest on the side in front of the aperture and is zero on the side opposite the aperture (i.e., the back side) since the pulse has not yet had time to reach the back side. Along the sides (i.e., the top or bottom of Fig. 15-18), there is some field present, particularly near the horn aperture.

Fig. 15-24 shows the far-field pattern at 10 GHz computed from the near-to-far-field transformation using data at a boundary 12 cells inside the extremities of the grid in Fig. 15-18. The pattern is in good agreement with the classical pattern shown in Fig. 9-16 for the case s $= \frac{1}{4}$ when the E-plane scale on the abscissa is 4 $\sin\theta$ since $B/\lambda \sin\theta = 4\sin\theta$ here.

The forward-region far-field pattern was calculated at $n = 800$ because data in Fig. 15-25 show that steady state has been achieved on the front (right) face of the near-to-far-field transformation boundary. Near-field convergence guarantees far-field convergence, but the reverse is not true. One could calculate the far-field pattern after more time steps, but
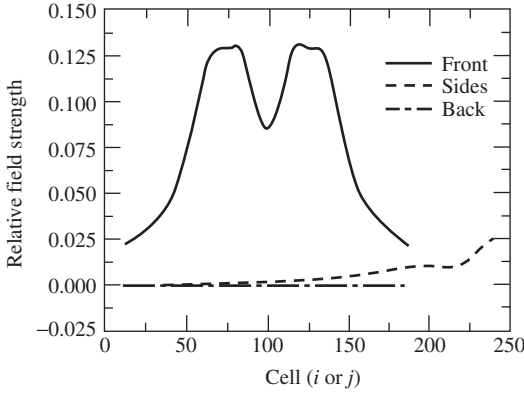
Figure 15-23 Amplitude of the 10-GHz electric field on the near-to-far-field transformation boundary at $n = 800$.
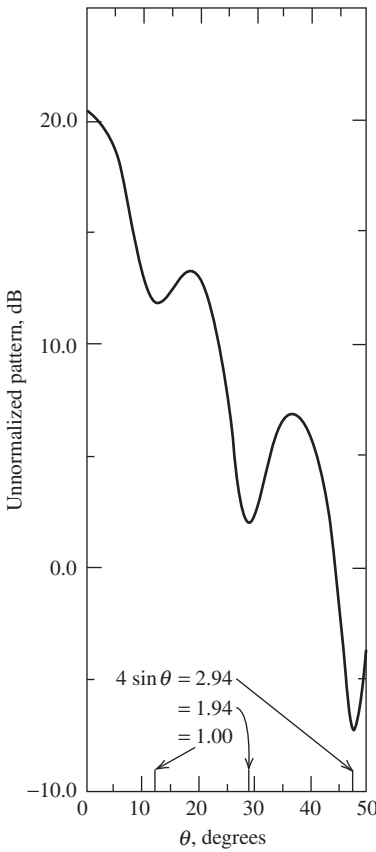
Figure 15-24 Far-field pattern at 10 GHz after 800 time steps. $B/\lambda \, \sin\theta = 4 \, \sin\theta$.

unwanted effects such as the interaction of the wave with the exterior of the waveguide may appear in the data. At, say, n $= 800$, the start of such occurs before the disturbance has had time to reach the back face of the near-to-far-field transformation boundary. In other words, it is important to march out in time far enough to achieve steady state, but not so far out in time that the desired data become contaminated with unwanted effects.
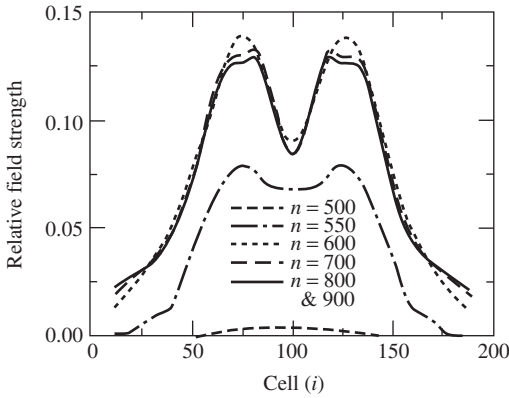
Figure 15-25 Amplitude of the 10-GHz electric field on the front part of the near-to-far-field boundary at six different time steps.
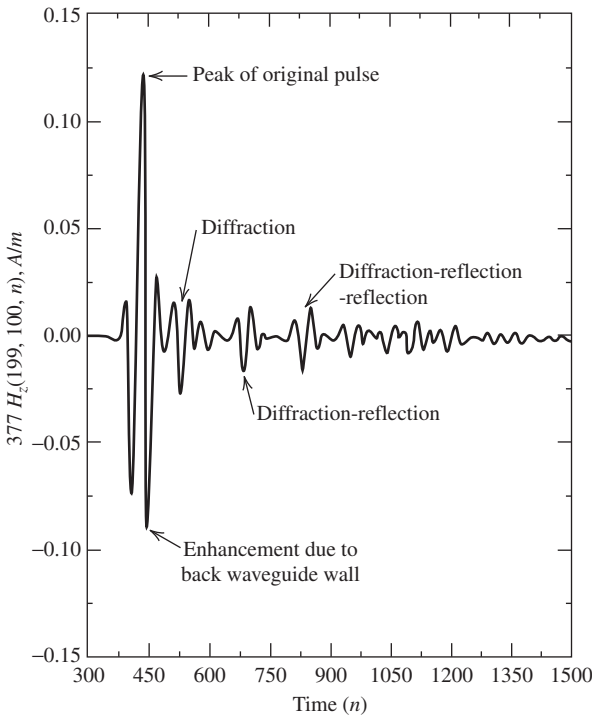
Figure 15-26 H-field at the center of the aperture center of the aperture ($i = 199$, $j = 100$) vs. n.

   FDTD has the potential to provide other data as well. Consider Fig. 15-26 that shows the magnetic field at the center of the aperture ($i = 199$, $j = 100$) from $n = 300$ to $n = 1500$. The first major feature to appear is the main pulse at approximately $n = 420$ (e.g., about 210 cells traveled with two time steps per cell). About 80 time steps later (i.e., 40 cells from an edge to the aperture center), the diffraction from the edges of the horn arrives. At approximately $n = 700$ and $n = 850$, respectively, diffraction that has reflected from the opposite wall arrives, followed by diffraction that has again reflected from the other wall. These last two effects at around $n = 700$ and $n = 825$ do not affect the main-lobe region of the far-field pattern because they are propagating in directions that are not in the main-lobe direction.

The next section will consider two three-dimensional antenna problems, although not in as much detail as was done here for the two-dimensional model of the *E*-plane sectoral horn.

15.9   ANTENNA ANALYSIS AND APPLICATIONS

Application of FDTD to antennas has occurred only recently relative to other applications such as shielding and radar cross section. One reason for this is that MoM can provide results for small, relatively simple antennas with much less computer time and memory than FDTD, since MoM finds only the currents flowing on the wire or conducting surface, whereas FDTD must calculate the fields in the entire computational region. This region should contain enough cells to allow some near-field decay between the antenna and the absorbing boundaries. If the antenna is small and geometrically simple, computing the fields in all the surrounding free-space cells makes FDTD much less efficient than MoM. For medium-sized antennas, or antennas with geometries and/or materials [9] that are not easily included in MoM formulations, or where data are needed at many frequencies, FDTD becomes a superior method.

When FDTD is applied to receiving antenna calculations, it loses one of its advantages relative to the MoM in applications that require results at multiple far-zone angles. For example, in scattering applications the MoM produces results for different plane wave incidence angles efficiently from a single impedance matrix, whereas FDTD requires a complete recalculation for each different incidence angle. However, for antenna trans-mitting problems, FDTD can produce far-zone fields in any number of different directions efficiently during one computation, as can MoM. Because FDTD also provides wide frequency band results with pulse excitation, it is extremely efficient in antenna appli-cations, since results for impedance and radiation patterns over a wide frequency band can be obtained from one FDTD computation.

15.9.1   Impedance, Efficiency, and Gain

It should not be forgotten that antenna descriptors we have become so comfortable with elsewhere in this book, such as impedance, gain, far-field patterns, and radar cross section, are frequency domain concepts. To obtain them from the FDTD calculation process, it is necessary to Fourier-transform the appropriate voltages, currents, and fields from the time domain to the frequency domain.

In order to capitalize on the advantages of FDTD (i.e., wide bandwidth data), it is common to utilize a Gaussian voltage pulse to excite an antenna. The Fourier transform of the voltage excitation pulse at the feed point is denoted as $V_A(\omega)$ and the Fourier transform of the current at the feed point is denoted $I_A(\omega)$ Then the input impedance is given by

$$Z_A(\omega) = \frac{V_A(\omega)}{I_A(\omega)} \tag{15-62}$$

To determine $\upsilon_A(t)$ and $i_A(t)$ in the FDTD grid, from which $V_A(\omega)$ and $I_A(\omega)$ are derived, we employ the line integrals of **E** and **H**, respectively.

Consider a situation in which an antenna is fed with one voltage source modeled as an electric field $E_z\big|_{i,j,k}^{n}$ with corresponding voltage $\upsilon_i(t)$ across the cell at the antenna feed gap, and this source supplies a time domain current $i_i(t)$. After all transients are dissipated and the time domain results for these two quantities are Fourier-transformed, the equivalent steady-state input power *at each frequency* is given quite simply by

$$P_{\text{in}}(\omega) = \frac{1}{2}\text{Re}\big[V_A(\omega)I_A^*(\omega)\big] \tag{15-63}$$

Dissipated power due to ohmic losses is computed as follows. Suppose that an FDTD electric field component $E_z(t)$ is in a region with conductivity σ. If we assume that the electric field is uniform within a single FDTD cell, then at each frequency the equivalent steady-state power dissipated in this region is given by

$$P_o = \frac{1}{2} \iiint \sigma |E_z(\omega)|^2 \, dv = \frac{1}{2} \sigma |E_z(\omega)|^2 \Delta x \ \Delta y \ \Delta z$$
$$= \frac{1}{2} \frac{\sigma \ \Delta x \ \Delta y}{\Delta z} |E_z(\omega) \Delta z|^2 = \frac{|V(\omega)|^2}{2R} \tag{15-64}$$

where $E_z(\omega)$ is the Fourier transform of $E_z(t)$ and R is a lumped resistance across the cell in the z-direction. Knowing P_{in} and P_o leads to a determination of radiation efficiency from (2-153).

To determine the antenna gain, the far-zone electric field in the desired direction must be determined at specified frequencies. If we use the approach given in the previous section, this can be done for pulsed far-zone fields. Since the far-zone electric field is computed so that the $1/r$ amplitude factor and the propagation delay are suppressed, the antenna gain relative to a lossless isotropic antenna in the (θ, ϕ) direction is given by

$$G(\omega, \theta, \phi) = \frac{1}{2} \frac{|E(\omega, \theta, \phi)^2|/\eta}{P_{\text{in}}/4\pi} \tag{15-65}$$

where $E(\omega, \theta, \phi)$ is the peak value of the Fourier transform of the pulsed far-zone time domain electric field radiated in the (θ, ϕ) direction.

15.9.2 The Monopole over a PEC Ground Plane

Maloney et al. [10] used FDTD to model the radiation from two simple antennas: the cylindrical monopole and the conical monopole. Here, we shall consider only the former, the cylindrical monopole of height h over a PEC ground plane. The FDTD grid used to model this antenna is shown in Fig. 15-27. The grid used a two-dimensional cylindrical-coordinate algorithm, exploiting the rotational symmetry of both the antenna and feeding coaxial line about the z-axis. In the cylindrical system, the TE mode is composed of E_ϕ, H_ρ, and H_z, whereas the TM mode has the components E_ρ, E_z, and H_ϕ. Since the coaxial line was excited with a TEM mode consisting of just E_ρ and H_ϕ, only the TM cylindrical mode was modeled.

A Gaussian pulse voltage excitation within the coaxial line $\upsilon(t) = \upsilon_0 \exp(-t^2/2\tau_p^2)$ was used at source plane $A - A'$ in combination with an exact ABC to emulate a matched source at that location. The following normalized parameters apply to the model: $b/a = 2.30$ (50-Ω coaxial line); $h/a = 65.8$; $\tau_p/\tau_a = 8.04 \times 10^{-2}$. Here τ_p is the $1/e$ width of the excitation pulse and $\tau_a = h/c =$ characteristic antenna height. In other words, τ_a represents the time required for an electromagnetic wave to travel the length of the monopole. Finally, an ABC of only first-order accuracy was used to terminate the grid at its outer boundary S_e.

The cylindrical monopole represents a two-dimensional electromagnetic problem. For example, the radiator in Fig. 15-27 is rotationally symmetric and is excited by a rotationally symmetric source

$$\mathbf{E}^i(t) = \frac{\upsilon^i(t)}{\ln(b/a)\rho} \hat{\boldsymbol{\rho}} \tag{15-66}$$
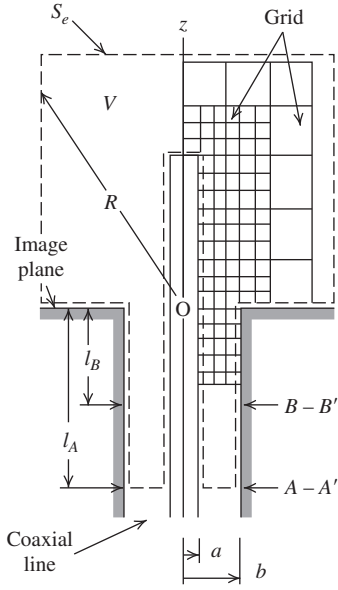
Figure 15-27 Geometry of the two-dimensional cylindrical-coordinate FDFD grid used to model the transient excitation of a coaxial-fed monopole over a PEC ground plane. (Maloney et al., *IEEE Trans. on Ant. & Prop.,* Vol. 38, pp. 1059–1068, July 1990. Reprinted with permission of IEEE.)

The applicable components of Maxwell's curl equations are

$$\frac{\partial E_\rho}{\partial z} - \frac{\partial E_z}{\partial \rho} = -\mu_o \frac{\partial H_\phi}{\partial t} \tag{15-67a}$$

$$-\frac{\partial H_\phi}{\partial z} = -\varepsilon_o \frac{\partial E_\rho}{\partial t} \tag{15-67b}$$

$$\frac{1}{\rho} \frac{\partial(\rho H_\phi)}{\partial \rho} = \varepsilon_o \frac{\partial E_z}{\partial t} \tag{15-67c}$$

After discretization of the above, we have

$$H_\phi\big|_{i,j}^{n+1/2} = H_\phi\big|_{i,j}^{n-1/2} + \frac{\Delta t}{\mu_o \Delta \rho} \left[E_z\big|_{i+1/2,j}^{n} - E_z\big|_{i-1/2,j}^{n} \right]$$
$$- \frac{\Delta t}{\mu_o \Delta z} \left[E_\rho\big|_{i,j+1/2}^{n} - E_\rho\big|_{i,j-1/2}^{n} \right] \tag{15-68a}$$

$$E_\rho\big|_{i,j-1/2}^{n+1} = E_\rho\big|_{i,j-1/2}^{n} - \frac{\Delta t}{\varepsilon_o \Delta z} \left[H_\phi\big|_{i,j}^{n+1/2} - H_\phi\big|_{i,j-1}^{n+1/2} \right] \tag{15-68b}$$

$$E_z\big|_{i+1/2,j}^{n+1} = E_z\big|_{i+1/2,j}^{n} + \frac{\Delta t}{\varepsilon_o \Delta \rho} \frac{1}{\rho_{i+1/2}} \left[\rho_{i+1} H_\phi\big|_{i+1,j}^{n+1/2} - \rho_i H_\phi\big|_{i,j}^{n+1/2} \right] \tag{15-68c}$$

Note that the grid in Fig. 15-28 is arranged so that the electric field component tangential to the surface of a perfect conductor is evaluated at the surface.

An absorbing boundary condition is used at the surface S_e; this allows the observation period to be extended beyond $t = t_o$. If we look in the opposite direction, the TEM field within the coaxial line behaves like the one-dimensional case examined in previous
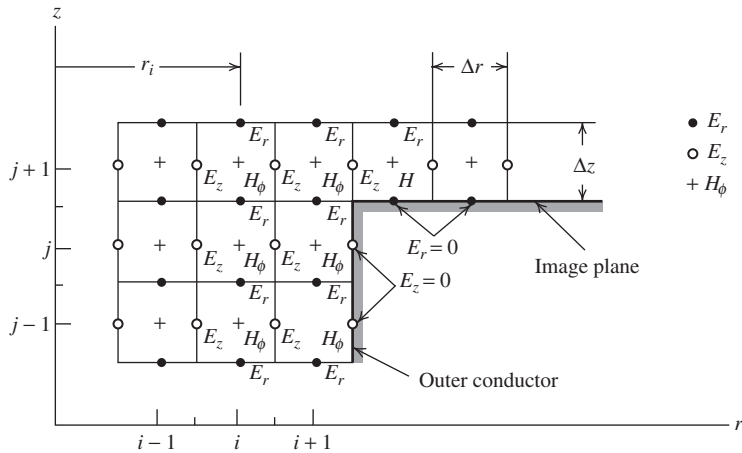
Figure 15-28 Spatial grid and field components for the two-dimensional problem with cylindrical symmetry. (Maloney et al., *IEEE Trans. on Ant. & Prop.,* pp. 1059–1068, July 1990. Reprinted with permission.)

sections. Thus, an exact absorbing boundary condition can be constructed within the coaxial line. The incident field is additively injected at a plane $z = -\ell$, and the absorbing boundary condition, placed at $z = -(\ell + \Delta z)$, exactly absorbs the field of a TEM mode propagating in the $-z$-direction. This allows the cross section at which the incident field is specified to be moved closer to the ground plane; namely, in Fig. 15-27, $B - B'(z = \ell_B)$ is used instead of $A - A'(z = -\ell_A)$. This reduces both the time required for observation and the size of the grid.

The spatial and temporal increments ($\Delta\rho$, Δz, and Δt) are chosen to satisfy the Courant-Friedrichs-Levy condition in the cylindrical system [3]:

$$c\Delta t \le \sqrt{\frac{\Delta\rho^2\,\Delta z^2}{\Delta\rho^2 + \Delta z^2}} \tag{15-69}$$

In this work, two spatial grid spacings are used: a fine spacing ($\Delta\rho_1 \approx \Delta z_1$) within the coaxial line and close to the antenna where the field is varying rapidly with spatial position, and a coarse grid [$\Delta\rho_2 = (3-5)\,\Delta\rho_1$, $\Delta z_2 = (3-5)\,\Delta z_1$] in the remainder of the space. The use of the dual grid reduces computer storage. Note that when (15-69) is satisfied for the fine grid, it is automatically satisfied for the course grid. In the example that follows,

$$c\Delta t = \frac{\min(\Delta\rho_1, \Delta z_1)}{2} \tag{15-70}$$

and the increments $\Delta\rho_1$, Δz_1 are chosen small enough to resolve the spatial variation of the field.

Fig. 15-29 is a space-time plot of the FDTD calculated surface charge density on the monopole antenna and its feeding coaxial line. At point *A* in this figure, the incident pulse has reached the antenna. An impedance mismatch between the feed-line and the antenna causes some of the energy to reflect back down the line. The remaining energy then propagates along the length of the antenna until the end of the antenna is reached at point *B*. Here, some energy radiates while the remaining energy reflects back down the antenna. This represents the (imperfect) transition from the antenna to free space.
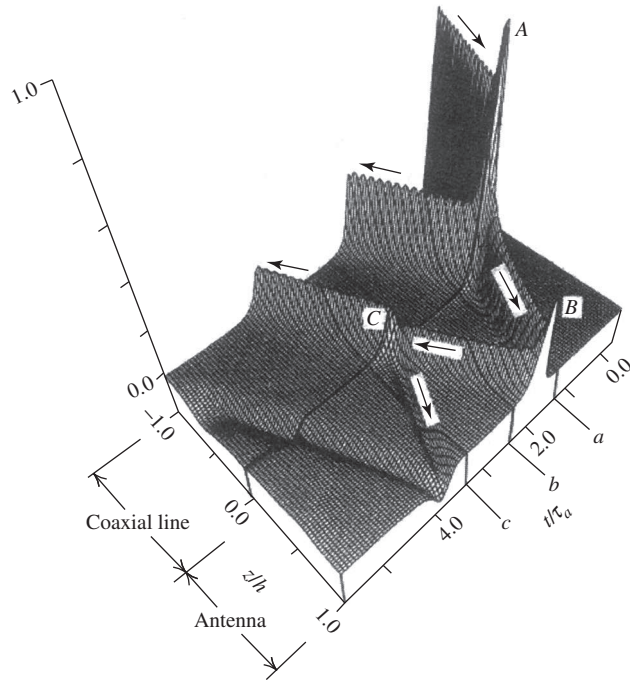
Figure 15-29 Normalized surface charge density on the cylindrical monopole antenna as a function of the normalized position z/h and the normalized time t/τ_a. (Maloney et al., *IEEE Trans. on Ant. & Prop.,* Vol. 38, pp. 1059–1068, July 1990. Reprinted with permission.)

At point C, the antenna/feedline junction causes a partial retroreflection, with some energy continuing down the coax and the remainder going back up the monopole. This process repeats itself until all transients have decayed. It is significant to note that at no time did energy enter the antenna from the coax once the incident pulse had propagated through. This verified that the ABC at the line feedpoint was working properly.

Fig. 15-30 shows the FDTD-computed radiated fields for three snapshots in time. In Fig. 15-30a, the initial outgoing wavefront W_1 was produced after the exciting pulse passed the feedline/antenna transition. Note the reflected energy traveling back down the coaxial feedline. In Fig. 15-30b, a second outgoing wavefront W_2 was produced when the energy traveling up the antenna was reflected from its top end. In Fig. 15-30c, both W_1 and W_2 have propagated away from the antenna, but a third wavefront W_3 was generated when energy retroreflected from the feed-line/antenna transition. Also in Fig. 15-30c, the wavefront W_{2R} arises from the reflection of the W_2 wavefront from the ground plane. This process repeats until the surface charge density decays to zero.

The far-zone electric field E_θ for the cylindrical monopole antenna is shown in Fig. 15-31. The surface used for these calculations was the cylindrical boundary separating the fine and coarse grids in Fig. 15–27. Each trace in this figure shows the electric field at a fixed polar angle θ as a function of the normalized time t/τ_a. Notice that the shape of each time domain trace is different for each polar angle because each trace has a different frequency content. This is due to the radiation patterns in the frequency domain being different at each frequency. Also notice that wave-fronts from the same point on the antenna are always separated by a time interval that is a multiple of $2\tau_a$, the round-trip transit time for the pulse on the antenna. For example, wavefronts W_1 and W_3, which are centered on the drive point, are separated by the time $2\tau_a$, as are wavefronts W_2 and W_4, which are centered on the end.
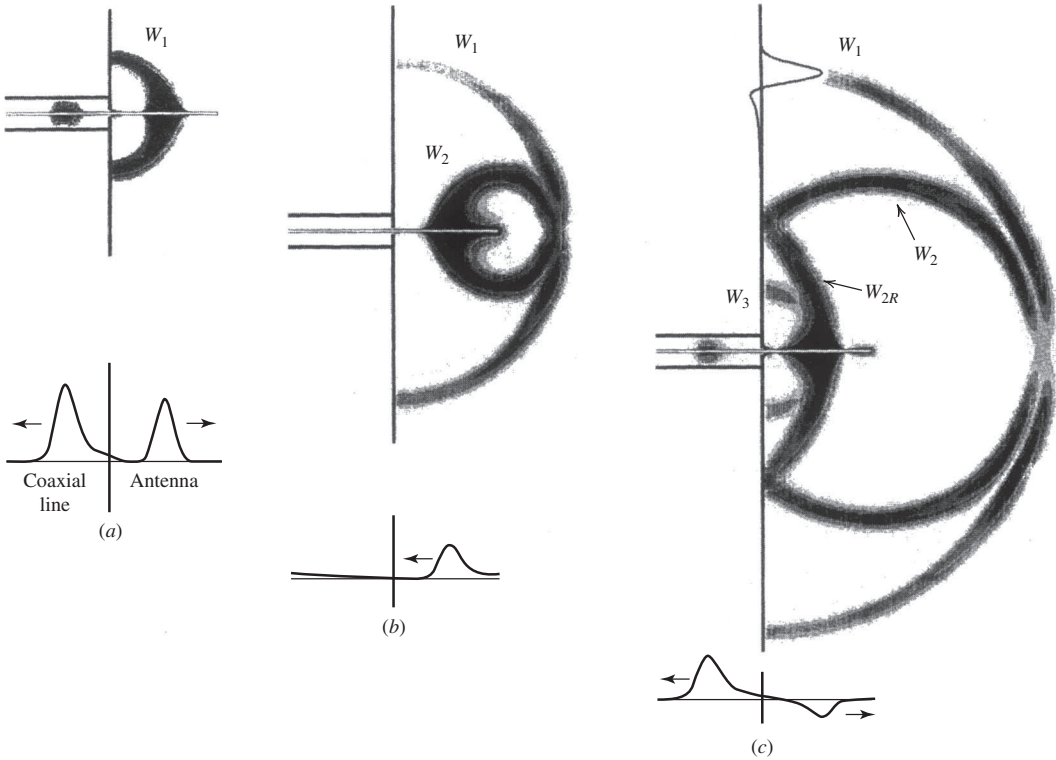
Figure 15-30 FDTD-computed radiation of a Gaussian pulse from the cylindrical monopole antenna. The gray scale plots show the magnitude of the electric field, whereas the line drawings show the surface charge density on the antenna and the feeding coaxial life. (Maloney et al., *IEEE Trans. on Ant. & Prop.,* Vol. 38, pp. 1059–1068, July 1990. Reprinted with permission.)
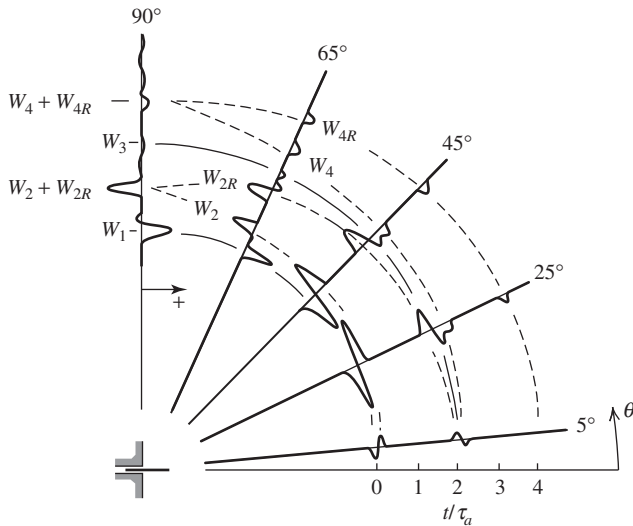
Figure 15-31 Radiation of a Gaussian pulse from a cylindrical monopole antenna. Each trace shows the far-zone electric field ε_θ at a fixed polar angle θ as a function of the normalized time t/τ_a. $b/a = 2.30$, $h/a = 65.8$, and $\tau_p/\tau_a = 8.04 \times 10^{-2}$. (Maloney et al., *IEEE Trans. on Ant. & Prop.,* pp. 1059–1068, July 1990. Reprinted with permission.)

15.9.3   Microstrip Leaky Wave Antennas

The microstrip leaky wave antenna was discussed in Sec. 11.4.3. This section discusses how the configuration in Fig. 11-8*a* was simulated on a PC with limited computer resources. The dimensions of interest for the full-width configuration in Fig. 11-8*a* where a length of 220 mm with the vias replaced by a solid wall down the centerline. The width of the microstrip was 15 mm. At 8 GHz the free-space wavelength is 37.5 mm, making the antenna 5.86 λ long. The material was Rogers Corporation 5870 Duroid with 1-ounce copper cladding 35 μm thick. The following discussion is taken from [11].

As a starting point in building the FDTD simulation model, the entire antenna was surrounded by a 20-cell-thick layer of free space, except for the space above the antenna (i.e., between the top of the antenna and the UPML), which was 40 cells thick. The substrate was extended one antenna width in both transverse directions, making the overall width of the configuration 45 mm, as shown in Fig. 15-32. The copper in the microstrip and the ground plane was a one-cell-thick layer of conductivity, $\sigma = 5.8 \times 10^7$ S/m. With cells nearly cubical in shape, the substrate would be 22 cells thick. Thus the initial simulation model would require nearly 900 million cells, far too large for a PC. By approximating the copper conductor as a perfect electric conductor (PEC) of zero thickness, larger cells could be used throughout, decreasing the total number of cells required. For example, by making the substrate only 5 cells thick, the simulation model is reduced from 900 million cells to about 50 million cells, a 94% decrease but still too large for a PC at the time this work was undertaken.

The PML used in this study was the UPML of Sec. 15.5 matched to free space, which surrounds the leaky wave antenna on all sides. Of interest to us is the propagation constant of the forward traveling wave as discussed in Sec. 11.4.3. To isolate the forward traveling wave, any reflected waves need to be eliminated in the simulation model. This can be accomplished by extending the substrate in the transverse and longitudinal directions directly into the UPML and modifying the affected UPML cells to match the substrate. This permits the UPML layer under the ground plate to be removed since it is now redundant. The UPML now absorbs the propagating waves except for the forward traveling wave. At this point the UPML is inhomogeneous in the broadside, or *y*-direction. If the electric flux density, D_y, is discontinuous between two materials, Gauss's Law states
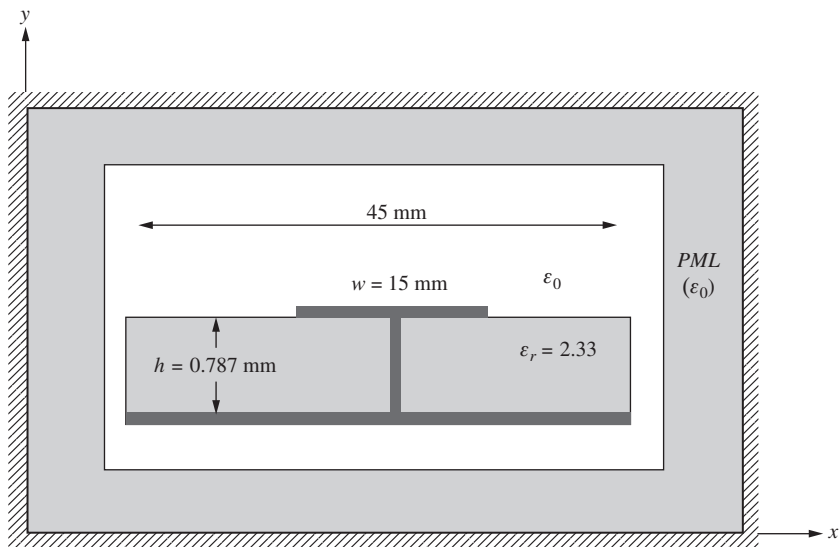
Figure 15-32 An *x-y* cross-section slice of the full-width antenna surrounded by free space (not to scale).

that a surface charge will develop at the boundary. Such a charge build-up is a potential source of instability in an FDTD simulation but can be avoided by choosing the imaginary component of UPML permittivity to be normalized to free-space permittivity instead of the material's permittivity.

Next in the development of the simulation model, free space was removed except for the region on the top side of the antenna, as shown in Fig. 15-33. Further, the thickness of the free-space layer above the antenna only needed to be 2 cells thick. The resulting simulation space is now reduced to 10 million cells.

A further reduction in the number of cells is possible if the cells used are elongated in the longitudinal direction instead of being cubic. The cells used, however, had the same dimensions in the transverse x-direction and broadside y-direction direction. It was found that the error in β_z at the center frequency of 7 GHz with respect to the transverse resonance solution was less than 1% if the ratio of the cell size was 1:1:6 (transverse:broadside: longitudinal) or less. Using a 5-cell-thick substrate with a ratio of 1:1:5 reduced the simulation space to 572, 000 cells.

MATLAB defaults to double precision unless specified otherwise. Runs in single precision used roughly 40% less memory and results were mostly consistent with double precision results. All runs were done in double precision unless the size of the simulation required more memory than was available.

The worst resolution will occur in areas with the shortest wavelength and largest cells, which occurs in the substrate outside the microstrip where the wavelength is the free-space wavelength reduced by $\sqrt{\varepsilon_r}$. When the substrate is 5 cells thick (i.e., each cell is $787/5\,\mu$m thick) and the size is 1:1:5 ($x:y:z$), the cells are $157.4 \times 157.4 \times 787\,\mu$m. When the size is 1:1:3, the cells are $157.4 \times 157.4 \times 472.2\,\mu$m. Between the microstrip and the ground plate where the z-directed guided wavelength is much greater, the resolution is hundreds of cells per guided wavelength, which is more than adequate for accurate determination of α_z and β_z.

A general rule for cell size is that the cells size be small enough to represent geometrical features adequately and that the cell size be smaller than one-tenth the shortest wavelength in the simulation space. The cell size used for the microstrip leaky wave easily meets these criteria and is further driven in this case by the discovered need to sample the fields more accurately in the critical transverse x-direction.
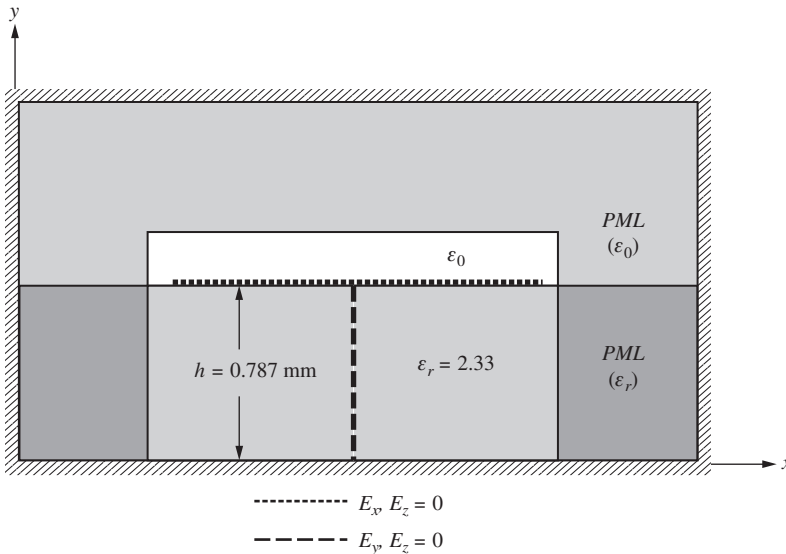
Figure 15-33 An x-y cross-section slice of the full-width antenna extending into the PML with unneeded material removed (not to scale).
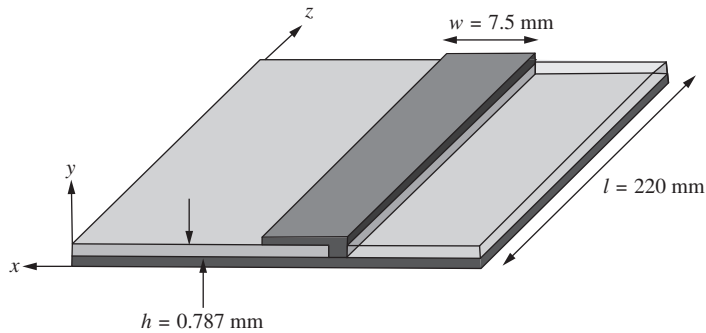
Figure 15-34 Half-width antenna without microstrip feed used in the FDTD simulation.

Finally, the unexcited half of the antenna in Fig. 11-8a was eliminated and the half-width configuration in Fig. 15-34 was used for the final simulation. The half-width antenna used approximately 45% less memory than the full-width simulation.

The source placement and size required some experimentation. It was found that at least one source cell must be more than two cells from the UPML or too much energy will be absorbed for the traveling wave to adequately develop. A single source cell centered vertically in the substrate between the microstrip and ground plate allowed a traveling wave of the largest magnitude to develop in the half-width antenna provided the source was two cells inside the open side edge of the antenna and five cells from the UPML. The source cell location is indicated in Fig. 15-33.

Results of the FDTD simulation are found in Figs. 11-12 and 11-13 and also in [12].

A purpose of this subsection is to demonstrate how a seemingly large simulation problem can be reduced in size without compromising accuracy and usefulness. As computer capability increases, larger and larger simulations become the norm. Recent advances to 64-bit architecture with corresponding increases in computer memory enable current desktop computers to overcome some of the computer limitations encountered in this study. Nevertheless, it is prudent to take advantage of opportunities to reduce the size of the FDTD simulation space.

15.10   SUMMARY

In this chapter, the finite difference time domain, or FDTD, method has been presented. FDTD is a differential-equation-based method in the time domain that employs approximations to derivatives in a solution of Maxwell's equations that "marches on in time and space." The basic features of FDTD were presented in Secs. 15.2 through l5.6. The one-dimensional case was used to illustrate fundamental principles, but the two-dimensional and three-dimensional cases were also considered (e.g., Secs. 15.8 and l5.9). Only simple media were employed, but FDTD is well suited to complex media as well, such as inhomogeneous and anisotropic media.

Both MoM (in the previous chapter) and FDTD are usually thought of as intermediate frequency techniques because they cannot easily accommodate bodies that are arbitrarily large in terms of the wavelength. In contrast to this, Chap. 16 presents high-frequency or asymptotic methods that apply best to material structures (i.e., antennas or scatterers) arbitrarily large in terms of the wavelength.

REFERENCES

1. E. T. Thiele, private communications, 1995.
2. K. S. Yee, "Numerical Solution of Initial Boundary Value Problems Involving Maxwell's Equations in Isotropic Media," *IEEE Trans. Ant. & Prop.*, Vol. 14, pp. 302–307, April 1966.

3. E. Isaacson and H. B. Keller, *Analysis of Numerical Methods*, Wiley, New York, 1967.
4. G. Mur, "Absorbing Boundary Conditions for the Finite-Difference Approximation of the Time-Domain Electromagnetic Field Equations," *IEEE Trans. Electromag. Compatibility*, Vol. 23, pp. 377–382, 1981.
5. J. P. Berenger, "A Perfectly Matched Layer for the Absorption of Electromagnetic Waves," *J. Computational Phys.*, Vol. 114, pp. 185–200, 1994.
6. S. D. Gedney, "An Anisotropic PML Absorbing Media for FDTD Simulation of Fields in Lossy Dispersive Media," *Electromagnetics*, Vol. 16, pp. 399–414, 1996.
7. S. D. Gedney, "An Anisotropic Perfectly Matched Layer Absorbing Media for the Truncation of FDTD Lattices," *IEEE Trans. Ant. & Prop.*, Vol. 44, pp.1630–1639, 1996.
8. J. A. Roden and S. D. Gedney, "Convolutional PML (CPML): An Efficient FDTD Implementation of the CFS-PML for Arbitrary Media" *Microwave Optical Tech. Lett.*, Vol. 50, pp. 334–339, 2002.
9. M. A. Jensen and Y. Rahmat-Samii, "EM Interaction of Handset Antennas and a Human in Personal Communications," *Proc. IEEE*, Vol. 83, pp. 7–17, January 1995.
10. J. G. Maloney, G. S. Smith, and W. R Scott, "Accurate Computation of the Radiation from Simple Antennas Using the Finite-Difference Time-Domain Method," *IEEE Trans. on Ant. & Prop.*, Vol. 38, pp. 1059–1068, July 1990.
11. G. M. Zelinski, "FDTD Analysis of a Leaky Traveling Wave Microstrip Antenna," Thesis, Air Force Institute of Technology, March 2005.
12. G. M. Zelinski, G. A. Thiele, M. L. Hastriter, M. J. Havrilla, and A. J. Terzuoli, "Half Width Leaky Wave Antennas," *IET Proc.-Microwave Antennas Propagation (London)*, Vol. 1, no. 2, pp 341–348, Apr. 2007.

PROBLEMS

15.1-1 Reduce (15-5) to (15-7) and (15-8).

15.1-2 Reduce (15-7) to (15-10) and reduce (15-8) to (15-11).

15.2-1 (a) Obtain (15-14) from (15-12) by expanding H_y in a Taylor series about the temporal point t_n and by expanding E_z, about the spatial point x_i.

(b) Show that the central difference approximations are second-order-accurate. That is, error $= O(\Delta z^2) + O(\Delta t^2)$.

(c) Does this tell how much error is in the solution?

15.2-2 Derive (15-17) from (15-10b).

15.2-3 Derive (15-22) from (15-21).

15.2-4 Derive (15-27) from (15-6c).

15.2-5 Derive from (15-5b) an expression for H_y similar to (15-26).

15.2-6 Derive from (l5-6b) an expression for E_y similar to (15-27).

15.2-7 Derive (15-30a) from (15-26).

15.3-1 Show that the one-dimensional central difference eqs. (15-14) and (15-17) produce an exact solution when $c\Delta t = \Delta x$.

15.3-2 Can numerical dispersion occur in a non-dispersive medium?

15.3-3 Consider a plane wave in free space:

$$H_y = H_o e^{j(\omega t - \beta x)}$$

If the plane wave is discretized in time and space

$$H_y(x_i,\ t_n) = H_o e^{j(\omega n\Delta t - \beta_{\text{num}} i\Delta x)}$$

where $t = n\Delta t$, $x = i\Delta x$, $\beta_{\text{num}} = \frac{\omega}{v}.v$ is the numerical phase velocity.

(a) Write an expression for the phase error of the discretized wave relative to the actual wave and comment on the amount of phase error as the propagation distance increases.

(b) How can the error in phase be overcome?

15.4-1 With reference to Fig. 15-9, calculate by hand E and H for the fourth time step when $i < 50$. Then, from your results, verify that there is power flow in the negative x-direction.

15.5-1 Recompute by hand Table 15-1 if there is a perfect electrical conductor in cell 53.

15.5-2 Show that the second-order Mur estimate in (15-43) reduces to an exact 1-D result.

15.6-1 Write a one-dimensional computer code and verify the results presented in Fig. 11-12 for the delta function hard source.

15.6-2 Extend the computer code in Prob. 15.6-1 to accommodate soft sources and verify Fig. 15-13.

15.6-3 With reference to Fig. 15-13, calculate by classical means the reflection and transmission coefficients at the dielectric to air interface and compare with the magnitude of the reflected and transmitted fields in the figure when $n = 400$.

15.6-4 At what time step in Fig. 15-13 does the peak value of the reflected Gaussian pulse arrive back at the source? Arrive at your answer by assuming the magic time step and then determining how many time steps are required.

15.6-5 In terms of the update equations, explain why the soft source in Fig. 15-13 allows the wave reflected from the dielectric to pass onward to the right and the hard source does not.

15.6-6 Verify (15-52) using (15-53).

15.6-7 Verify (15-54) using (15-55).

15.6-8 Explain why it is more computationally efficient to use FDTD for antennas when the antenna is transmitting than when it is receiving.

15.6-9 Extend Prob. 15.5-1 out to $n = 8$ if the source at $i = 50$ is a hard source. Repeat if the source is a soft source.

15.7-1 Using the surface equivalence theorem, show that the far-field pattern of an antenna may be computed using (15-59) and (15-60) applied to a surface surrounding the antenna.

15.8-1 In Fig. 15-22, exclude edge effects on the phase and show that the phase change from the center of the aperture to the edges is $90°$ at 10 GHz as required by the optimum condition under which this horn was designed. (Note the amplitude distribution near the edges and use that to estimate the region over which the edges are having a strong effect on both the amplitude and phase distributions.)

15.8-2 Compare Fig. 15-24 to Fig. 15-16.

15.8-3 By making measurements on Fig. 15-18 with a ruler, verify the time at which the various physical phenomena depicted in Fig. 15-26 occur.

15.8-4 Write a two-dimensional computer code (or use an existing one) and verify Fig. 15-22.

15.8-5 Make a photocopy of Figs. 15-21$a-b$ and indicate on the copy the various physical phenomena that you see there.

15.9-1 Verify (15-64).

15.9-2 Derive (15-65).

15.9-3 Derive (15-68) from (15-67).

Chapter 16

CEM for Antennas: High-Frequency Methods

Optics is a well-understood area of physics that deals with the characteristics of light wave propagation. It was Maxwell who showed before 1873 that the propagation of light could be viewed as an electromagnetic phenomenon. Since the wavelength of light waves is usually small compared to objects with which it interacts, the analytical treatment of light wave propagation is much different than that employed to analyze lower-frequency propagation where the size of a scattering surface is comparable to the wavelength.

A very useful and easily understood method for analyzing optical problems is the ray concept. The relationship between ray optics and wave propagation is apparent from the famous works of Huygens in 1690 and Fresnel in 1818, but was not formally shown until the works of Luneberg in 1944 and Kline in 1951. [1] Since that time the well-known methods of optics have found increasing use in the treatment of many electromagnetic problems in the radio frequency portion of the spectrum for situations where the wavelength is small compared to the geometrical dimensions of the scatterer or antenna. In these cases, asymptotic high-frequency methods must be employed since it is not practical to use moment methods (Chap. 14) or eigenfunction expansions. This is because the rate of convergence of both of these techniques is generally quite poor when dealing with an electrically large antenna or scatterer.

In this chapter, we will first examine the principles of geometrical optics. We will then see that in many situations geometrical optics is inadequate to completely describe the behavior of the electromagnetic field, and it is necessary to include another field called the diffracted field. The diffracted field, when added to the geometrical optics field, permits us to solve many practical radiation and scattering problems in a moderately straightforward manner that could not be solved any other way.

Geometrical optics and its extension to include diffracted fields is a field-based method (see Fig. 14-1) and does not require the calculation of currents. Later in this chapter, current-based methods will be discussed wherein currents are used to ultimately determine the field quantities of interest. These methods are physical optics and its extension to include diffraction. In many situations, a physical optics current is inadequate to produce accurate fields from a radiating object and it is necessary to include another current called the nonuniform current. The nonuniform current, when added to the physical optics current, permits an accurate representation of the fields to be obtained. Whether a field-based or a current-based method is to be used depends on the specific application, as we shall see in the sections that follow.

700

16.1   GEOMETRICAL OPTICS

Geometrical optics, or *ray optics* as it is often called, was originally developed to analyze the propagation of light where the frequency is sufficiently high that the wave nature of light need not be considered. Indeed, geometrical optics can be developed by simply considering the transport of energy from one point to another without any reference to whether the transport mechanism is particle or wave in nature.

Classical geometrical optics applies to isotropic lossless media that may or may not be homogeneous. In this chapter, we will only consider homogeneous media where the index of refraction n is assumed to be real and is given by

$$n = \frac{c}{v} \tag{16-1}$$

and is not a function of position within a given medium. Here, c is approximately 3×10^8 m/s and v is the velocity of propagation in the medium. In a homogeneous medium, energy moves along ray paths that are straight lines. Normal to these ray paths are a family of surfaces called the *eikonal* of the ray system. In applying geometrical optics, it is only necessary that we know either the eikonal of the ray system or the ray paths, since the two are uniquely related.

For a plane wave in homogeneous media, the eikonal surfaces are planes perpendicular to the ray paths as shown in Fig. 16-1a. For a spherical source, the eikonal surfaces are spherical surfaces perpendicular to the ray paths as shown in Fig. 16-1b.

The variation of the amplitude of the geometrical optics field within a ray tube is determined by the law of energy conservation since the rays are lines of energy flow. Consider two surfaces ρ_o and $\rho_o + \Delta\rho$ as shown in Fig. 16-2. Between the two surfaces, we can construct a tube of constant energy flux by using the rays. Thus, the energy through cross section $d\sigma_o$ at P_o must equal the energy flux through cross section $d\sigma$ at P. If S is the power per unit area, the condition of constant energy flow through the flux tube is

$$S_o \ d\sigma_o = S \ d\sigma \tag{16-2}$$

In the case of electromagnetic waves, the quantity S is the real part of the complex Poynting vector and we can assume that

$$S = \frac{1}{2}\sqrt{\frac{\varepsilon}{\mu}}|E|^2 \tag{16-3}$$

Substituting (16-3) into (16-2) yields

$$|E_o|^2 \ d\sigma_o = |E|^2 \ d\sigma \tag{16-4}$$
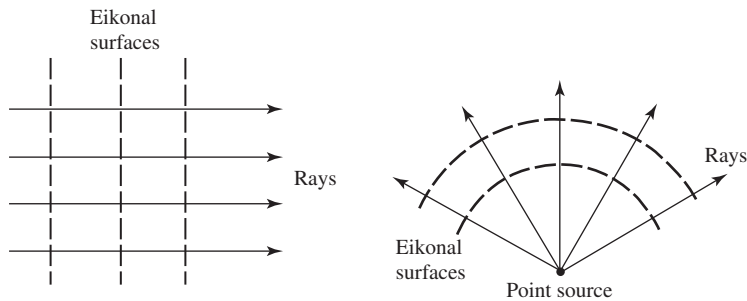
(a) Plane wave.   (b) Spherical wave.

Figure 16-1 The relationship of rays and eikonals for two types of sources.
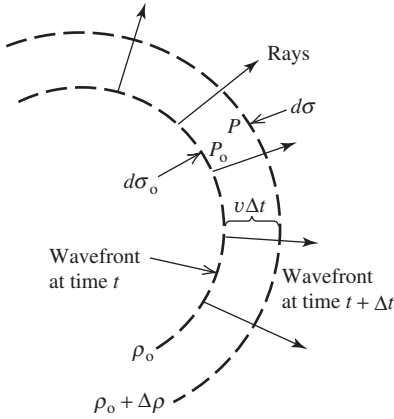
Figure 16-2 The relationship of rays and wavefronts.
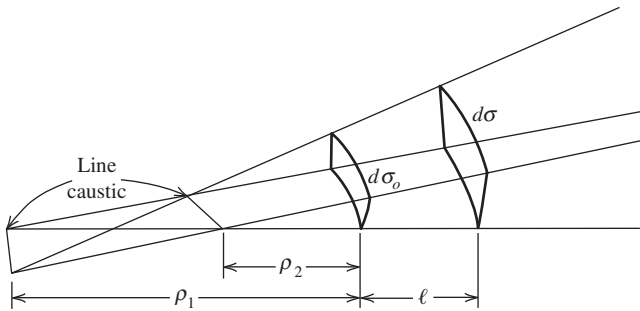
Figure 16-3 An astigmatic ray tube.

Solving for $|E|$, we obtain

$$|E| = |E_o| \sqrt{\frac{d\sigma_o}{d\sigma}} \tag{16-5}$$

Therefore, we have obtained a relationship between the amplitude of the geometrical optics field at one point in terms of the amplitude at another.

The relationship in (16-5) would be more useful if the radii of curvature of the wavefront surfaces $d\sigma$ and $d\sigma_o$ were used. Consider the astigmatic ray tube picture in Fig. 16-3. The principal radii of curvature of $d\sigma_o$ are ρ_1 and ρ_2, whereas the principal radii of curvature of $d\sigma$ are $(\rho_1 + \ell)$ and $(\rho_2 + \ell)$. we can write for the ratios

$$\frac{d\sigma_o}{\rho_1 \rho_2} = \frac{d\sigma}{(\rho_1 + \ell)(\rho_2 + \ell)} \tag{16-6}$$

and thus

$$\frac{d\sigma_o}{d\sigma} = \frac{\rho_1 \rho_2}{(\rho_1 + \ell)(\rho_2 + \ell)} \tag{16-7}$$

From (16-5), we have

$$|E| = |E_o| \sqrt{\frac{\rho_1 \rho_2}{(\rho_1 + \ell)(\rho_2 + \ell)}} \tag{16-8}$$
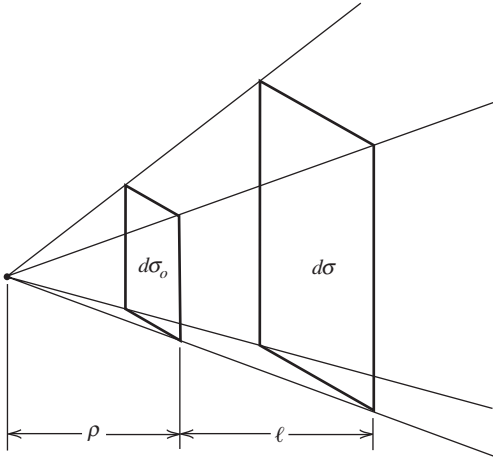
Figure 16-4 A tube of rays from a point source.

Note that the tube of rays converge to a line at $\rho_1 = 0$ and $\rho_2 = 0$ where the cross section of the ray tube goes to zero. Therefore, the amplitude of the geometrical optics field description becomes infinite there although the actual field does not. The locus of points where the ray tube cross section exhibits such behavior is called a *caustic*. Caustics may be a point, line, or surface. For example, consider a point source as shown in Fig. 16-4. We can construct a ray tube from four rays and write

$$\frac{d\sigma_o}{\rho^2} = \frac{d\sigma}{(\rho + \ell)^2} \tag{16-9}$$

Thus,

$$|E| = |E_o|\sqrt{\frac{\rho^2}{(\rho + \ell)^2}} = |E_o|\frac{\rho}{(\rho + \ell)} \tag{16-10}$$

The caustic would be located at the point source in this case.

In both (16-8) and (16-10), we note that as ℓ becomes large, we have the usual inverse distance-type field dependence found in the far zone of a three-dimensional source. Often, however, one is concerned with two-dimensional problems where one of the radii of curvature, say, ρ_2, becomes infinite. In such problems,

$$|E| = |E_o|\sqrt{\frac{\rho_1}{\rho_1 + \ell}} \tag{16-11}$$

Here, the eikonal surfaces are cylindrical and, as $\ell \to \infty$, we have an amplitude dependence of the field at large distances of the form $1/\sqrt{\ell}$. Obviously, if both ρ_1 and ρ_2 are infinite, the eikonal surfaces are planes and $|E|$ is a constant for all values of ℓ (e.g., a plane wave).

The results of (16-8), (16-10), and (16-11) are extremely important for they permit us to easily compute the amplitude of the geometrical optics field at one point in terms of its known value at another. In electromagnetic field problems, however, we must also include the phase. Phase can be introduced into (16-8) artificially. First, we take our phase reference to coincide with the amplitude reference. Thus, the electrical phase of the ray tube is given by $e^{-j\beta\ell}$ and we may write for the amplitude and phase of the field in the ray tube of Fig. 16-3

$$E = E_o e^{j\phi_o} \sqrt{\frac{\rho_1 \rho_2}{(\rho_1 + \ell)(\rho_2 + \ell)}} \, e^{-j\beta\ell} \tag{16-12}$$

or

$$E = E_o e^{i\phi_o} A(\rho_1, \rho_2, \ell) e^{-j\beta\ell} \tag{16-13}$$

where E_o is the reference amplitude at $\ell = 0$, ϕ_o is the reference phase at $\ell = 0$, $A(\rho_1, \rho_2, \ell)$ is the general spatial attenuation factor, and $e^{-j\beta\ell}$ is the spatial phase delay factor.

Note that when ℓ becomes less than $-\rho_2$, the quantity under the radical sign in $A(\rho_1, \rho_2, \ell)$ becomes negative and a phase jump of $\pi/2$ occurs when the observer passes through the caustic. Although we cannot predict the amplitude or the phase of the geometrical optics field at the caustic, we can determine the fields on either side of the caustic.

Eq. (16-12) or (16-13) permits us to approximately express the field at a point (i.e., ℓ) in terms of the value at a known point (i.e., $\ell = 0$). Rigorously, the result is only approximate, becoming more accurate as the wavelength tends toward zero. In practice, however, we will find the geometrical optics expression above to be highly accurate for engineering purposes where the assumptions of geometrical optics are valid.

To finish our initial discussion of geometrical optics, we illustrate its use by considering the problem of reflection at a curved smooth surface and the subsequent calculation of the radar cross section of a sphere. From (16-12), it is apparent that we need an expression for the radii of curvature of the wavefront in terms of the geometrical radii of curvature of the surface. Consider Fig. 16-5 that depicts a line source parallel with the axis of a convex cylinder of arbitrary cross section. From Fig. 16-5a,

$$\gamma_1 = \pi - \alpha - (\pi - \theta_o) = \theta_o - \alpha \tag{16-14}$$

The element of arc length in Fig. 16-5b is equal to $r_1^c \Delta\alpha$ and

$$r_1^c \Delta\alpha = \frac{\Delta\gamma_1 \ell_o}{\cos\theta_o} = \frac{(\Delta\theta_o - \Delta\alpha)\ell_o}{\cos\theta_o} \tag{16-15}$$

Since $\Delta\gamma_2 = \Delta\theta_o + \Delta\alpha$, we have

$$r_1^c \Delta\alpha = \frac{\rho_1 \Delta\gamma_2}{\cos\theta_o} = \rho_1 \frac{\Delta\theta_o + \Delta\alpha}{\cos\theta_o} \tag{16-16}$$

$$\Delta\gamma_1 = \Delta\theta_o - \Delta\alpha \qquad \Delta\theta_o = \theta_2 - \theta_1$$
$$\Delta\gamma_2 = \Delta\theta_o + \Delta\alpha$$
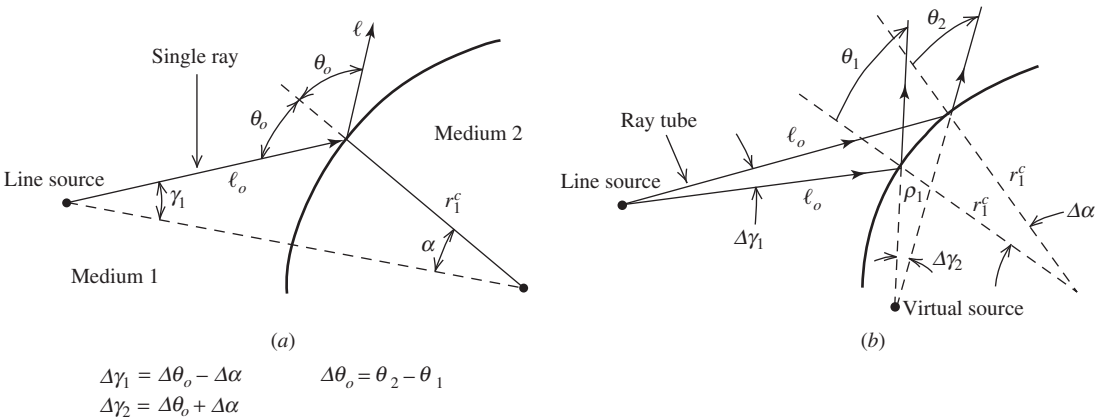
Figure 16-5 Ray geometry for reflection by a curved conducting surface.

Solving both (16-15) and (16-16) for $r_1^c \, \Delta\alpha \, \cos\theta_o$, we have, respectively,

$$r_1^c \, \Delta\alpha \cos\theta_o = \ell_o \, \Delta\theta_o - \ell_o \, \Delta\alpha \tag{16-17}$$

and

$$r_1^c \Delta\alpha \cos\theta_o = \rho_1 \, \Delta\theta_o + \rho_1 \, \Delta\alpha \tag{16-18}$$

Solving both these equations for $\Delta\alpha$ and equating the two results yields

$$\frac{\ell_o \, \Delta\theta_o}{r_1^c \, \cos\theta_o + \ell_o} = \frac{\rho_1 \, \Delta\theta_o}{r_1^c \, \cos\theta_o - \rho_1} \tag{16-19}$$

which after some manipulation gives us the desired result

$$\frac{1}{\rho_1} = \frac{1}{\ell_o} + \frac{2}{r_1^c \, \cos\theta_o} \tag{16-20}$$

This equation[1] relates a principal radii of curvature of the reflected wavefront to the geometrical radius of curvature of the surface at the point where the ray strikes the surface.

As a simple example of the application of (16-20), consider the situation shown in Fig. 16-6 where a plane wave is incident on a sphere. We wish to find the field scattered back in the direction of the transmitter and from this *back-scattered field* find the radar cross section of the sphere. Thus, the only ray we need consider is that reflected from what is called the *specular point*. In this situation, then, $\ell_o = \infty$, $\theta_o = 0°$, and $r_c = a$ in (12-20) and we have the following result:

$$\rho_1 = \frac{a}{2} = \rho_2 \tag{16-21}$$

where ρ_2 is the radius of curvature of the reflected wavefront orthogonal to ρ_1. (See Prob. 16.1-1 for an expression for ρ_2.)

If the incident field has a value of E_o at the specular point, then in the backscattered direction,

$$E^s = -E_o \frac{\rho_1}{\rho_1 + \ell} e^{-j\beta\ell} \tag{16-22}$$

Reflected wavefront

Virtual source

a

$\frac{a}{2}$

Specular point
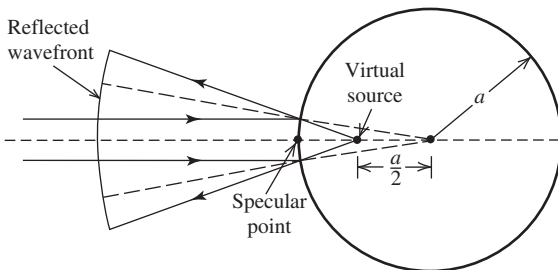
Figure 16-6 Geometrical optics scattering by a sphere.

[1] Even though this result is based on a two-dimensional configuration, the result is somewhat more general than this in that it holds true in the plane of incidence whenever the plane of incidence coincides with the principal planes of the surface.
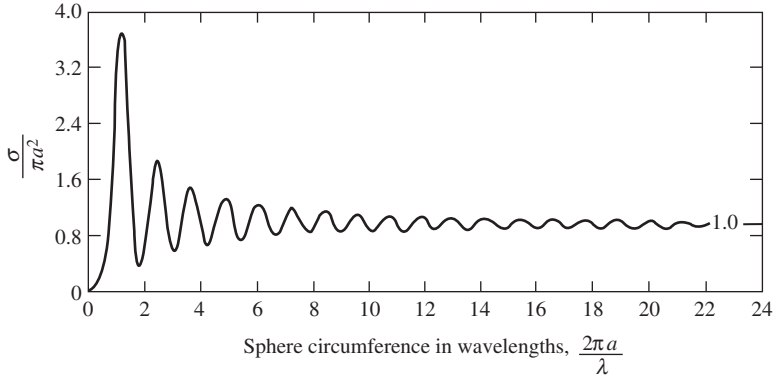
Figure 16-7 Radar cross section σ of a sphere versus the electrical size of the sphere.

with ρ_2 having the same value as ρ_1 in this example. Therefore, similar to (4-67), the radar cross section is (at high frequencies)

$$\sigma = \lim_{\ell \to \infty} 4\pi\ell^2 \left[\frac{a/2}{a/2 + \ell} \right]^2 = \pi a^2 \tag{16-23}$$

The exact value for $\sigma/\pi a^2$ is shown in Fig. 16-7. We note that as the radius of the sphere becomes larger, the more closely the geometric optics cross section approaches the exact (high frequency) result. That is what one would expect since geometrical optics assumes the wavelength is small when compared to the geometrical dimensions of the scattering surface. Furthermore, the result in (16-23) is frequency-independent, which is typical of geometrical optics calculations [2-4].

We can extrapolate from (16-22) to write a general expression for the geometrical optics field due to a plane wave reflected from a smooth surface. Let a plane of incidence be defined by the incident ray and the normal to the surface. Let $E_\parallel^i(Q_r)$ and $E_\perp^i(Q_r)$ be the components of the incident field that are parallel and perpendicular, respectively, to the plane of incidence at the point of reflection Q_r, and let $E_\parallel^r(\ell)$ and $E_\perp^r(\ell)$ be the components of the reflected field that are parallel and perpendicular to the plane of incidence, respectively. Then, in matrix form,

$$\begin{bmatrix} E_\parallel^r(\ell) \\ E_\perp^r(\ell) \end{bmatrix} = \begin{bmatrix} E_\parallel^i(Q_r) \\ E_\perp^i(Q_r) \end{bmatrix} \cdot [R] \sqrt{\frac{\rho_1 \rho_2}{(\rho_1 + \ell)(\rho_2 + \ell)}} \, e^{-j\beta\ell} \tag{16-24}$$

where R is a reflection coefficient matrix, and for perfectly conducting surfaces appears as

$$R = \begin{bmatrix} R_\parallel & 0 \\ 0 & R_\perp \end{bmatrix} \tag{16-25}$$

where $R_\parallel = +1$ and $R_\perp = -1$ and are recognized as representing the Fresnel reflection coefficients for parallel and perpendicular polarization reflection from plane perfectly conducting surfaces.[2] The Fresnel reflection coefficients imply that the incident wave is a plane wave and that the reflecting surface is also planar. We can deviate from these restrictions at high frequencies (short wavelengths) by noting that geometrical optics reflection is a local phenomenon and, therefore, the incident field need only be locally

---

[2] The signs on the entries $R_\parallel$ and $R_\perp$ depend on the reference directions used for the incident and reflected electric field vectors. The treatment here is consistent with that in Fig. 6-62, but others are in common use.

plane at the reflecting point and the surface need only be adequately approximated by a plane tangent to the surface at the point of reflection.

If we apply geometrical optics to reflection from a surface when the source point and reflection point are fixed, then the observation point is determined for us by the law of reflection. That is, we obtain information about the reflected field in one direction only (the specular direction), whereas the reflection typically spreads out over some angular region. To obtain information about the reflected field in nonspecular directions, it is necessary to first consider what the current is on the reflecting surface and then integrate that current to get the reflected field (e.g., the aperture integration of Chap. 9). In the Sec. 16.14, we will examine the physical optics method of doing this.

16.2   WEDGE DIFFRACTION THEORY

In the previous section, we introduced the ray-optical concept of geometrical optics. The theory was applied to the calculation of the backscattered field from a sphere, but no attempt was made to determine the field in the forward scattering direction, in particular, the shadowed region in Figs. 16-6 or 16-8. By simple ray tracing, it is quite apparent that geometrical optics is incapable of correctly predicting a nonzero field in the shadow region. However, geometrical optics may be extended to include a class of rays, called *diffracted rays* [5, 6], that permit the calculation of fields in the shadow region of a scatterer. Diffracted rays are produced, for example, when a ray strikes an edge, a vertex, or is incident tangentially to a curve surface as illustrated in Fig. 16-8. It is these rays that account for a nonzero field in the shadow region. In addition, they also modify the geometrical optics field in the illuminated region. It is the purpose of this section to examine in some detail one type of diffracted ray, the wedge diffracted ray of Fig. 16-8*a*.

Consider the wedge diffraction situation shown in Fig. 16-8*a*. Geometrical optics would predict a sharp discontinuity in the field at a *shadow boundary* as shown in Fig. 16-9. Since physical phenomena in nature are not perfectly discontinuous, it is apparent that geometrical optics by itself constitutes an incomplete treatment of problems such as those in Fig. 16-8. It will be shown that the wedge diffracted rays will make the total electric field continuous across the shadow boundary in Fig. 16-8*a*.

Because diffraction is a local phenomena at high frequencies, the value of the field of a diffracted ray is proportional to the field value of the incident ray at the point of diffraction multiplied by a coefficient called the *diffraction coefficient*. That is, the diffraction coefficient is determined largely by the local properties of the field and the boundary in the immediate neighborhood of the point of diffraction. Since it is only

(*a*) Wedge diffraction.   (*b*) Tip diffraction.   (*c*) Curved surface diffraction.
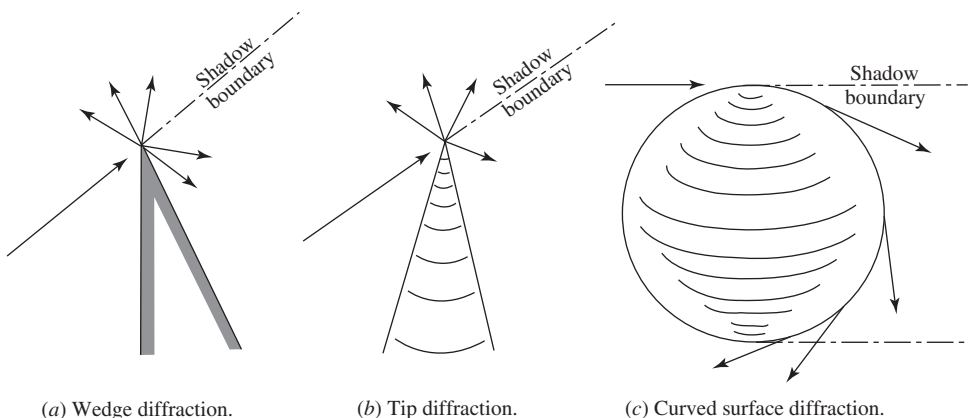
Figure 16-8 Examples of diffraction.

the local conditions near the point of diffraction that are important, the diffracted ray amplitude may be determined from the solution of the appropriate boundary value problem having these local properties. Such a problem is called a canonical problem and wedge diffraction is one such canonical problem. Wedge diffraction is perhaps the most important canonical problem in the extension of geometrical optics as originally proposed by Joseph Keller in 1953. Keller's theory is known as the *geometrical theory of diffraction,* or GTD [5–6].

Through the use of geometrical optics and the solution to a number of canonical problems, such a those in Fig. 16-8, we can construct solutions to more complex problems via the principle of superposition. Let us now consider the canonical problem of wedge diffraction. To start, we will consider scalar diffraction by an infinitely conducting and infinitesimally thin half-plane sheet as shown in Fig. 16-10. The half-plane is a wedge of zero included angle. To calculate the field in the region $z > 0$, we will use Huygens' principle in two dimensions; see Sec. 9.1. Thus, each point on the primary wave front along $z = 0$ is considered to be a new source for a secondary cylindrical wave, the envelope of these secondary cylindrical waves being the secondary wave-front. Thus,

$$E(P) = \int_{x=a}^{x=\infty} dE \tag{16-26}$$

where dE is the electric field at P due to a magnetic line source parallel to the y-axis in the $z = 0$ plane, or

$$dE = \frac{C_1}{\sqrt{\ell + \delta(x)}} \, e^{-j\beta[\ell + \delta(x)]} \, dx \tag{16-27}$$
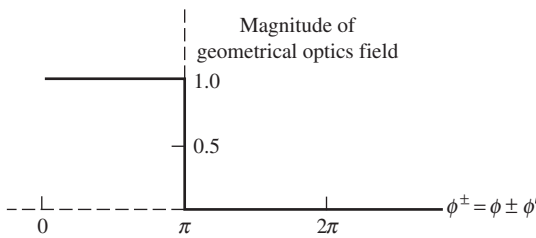
Figure 16-9 Magnitude of the geometrical optics field near either a reflected field shadow boundary $(\phi^+ = \phi + \phi' = \pi)$ or an incident field shadow boundary $(\phi^- = \phi - \phi' = \pi)$.
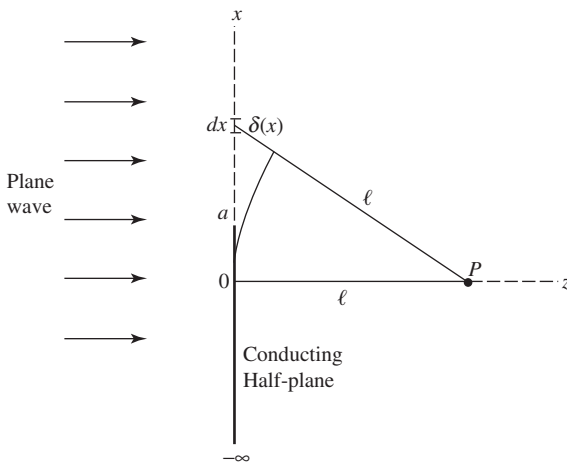
Figure 16-10 Plane wave diffraction by a conducting half-plane.

where C_1 is a constant. If $(\ell + \delta) \gg \lambda$, and $\ell \gg \delta$, we may write for the contribution to $E(P)$ from those two-dimensional Huygens' sources between $x = a$ and $x = x_o$

$$E(P) = \frac{C_1}{\sqrt{\ell}} e^{-j\beta\ell} \int_a^{x_o} e^{-j\beta\delta(x)} \, dx \qquad (16\text{-}28)$$

We will consider the contribution from x_o to ∞ later. When $\delta \ll \ell$, we can follow the same reasoning as in (2-86) to show that $\ell + \delta \approx \ell + x^2/2\ell$. Making the substitutions $\gamma^2 = 2/\lambda\ell$ and $u = \gamma x$ gives

$$E(P) = C_1 \sqrt{\lambda/2} \, e^{-j\beta\ell} \int_{\gamma a}^{\gamma x_o} e^{-j(\pi/2)u^2} du \qquad (16\text{-}29)$$

If the upper limit in (16-29) is allowed to go to infinity, the integral will be in the standard form of a Fresnel integral [H.10.4: James], [H.4: Jordan]. The Fresnel integral may be easily evaluated on a computer or from a graph known as Cornu's spiral, which is shown in Fig. 16-11a. A vector drawn from the origin to any point on the curve represents the magnitude of a Fresnel integral with lower limit zero and upper limit u_o. As u_o approaches infinity, the tip of the vector will circle the point $(\frac{1}{2}, \frac{1}{2})$ an infinite number of times, which suggests that the contribution to the value of the integral comes primarily between the limits zero and u_o provided $u_o > 1.26$. For this reason, we can argue that allowing $\gamma x_o \rightarrow \infty$ in (16-29) has little effect on the value of the integral. Thus,

$$E(P) \approx C_1 \sqrt{\lambda/2} \, e^{-j\beta\ell} \int_{\gamma a}^{\infty} e^{-j(\pi/2)u^2} \, du \qquad (16\text{-}30)$$

The value of the integral in (16-30) can be represented by a vector drawn from any point on the Cornu spiral to the point $(\frac{1}{2}, \frac{1}{2})$ (e.g., see Prob. 16.2-2).

If the lower limit in (16-30) is allowed to go to minus infinity, $E(P)$ will equal the field strength without the half-plane present [H.4: Jordan]. Thus,

$$E(P)|_{a=-\infty} = C_1 \sqrt{\lambda/2}(1 - j)e^{-j\beta\ell} = E_o e^{-j\beta\ell} \qquad (16\text{-}31)$$

Solving for C_1 and substituting into (16-30) give the value of $E(P)$ in terms of the free-space field E_o:

$$E(P) \approx \frac{E_o e^{+j(\pi/4)}}{\sqrt{2}} e^{-j\beta\ell} \int_{\gamma a}^{\infty} e^{-j(\pi/2)u^2} \, du \qquad (16\text{-}32)$$

where for this approximate analysis to hold, it is necessary that $\ell \gg \lambda$ and the point $x = a$ not be far removed from the z-axis so that the assumption $\ell \gg \delta$ holds. A more exact (and complicated) analysis of this problem is possible, but it has not been presented here for we wish simply to show how the Fresnel integral arises naturally in the study of wedge diffraction.

Eq. (16-32) and the Cornu spiral make it possible to visualize the variation of the electric field as the point a moves along the x-axis, causing the observation point to change from the lit region to the shadow region. The corresponding plot of the relative electric field in the vicinity of the shadow boundary is shown in Fig. 16-11b. We note that on the shadow boundary the value of the relative field is $\frac{1}{2}$ and in the lit region the value of the field oscillates about the value of unity. This oscillation can be interpreted as being caused by interference between the diffracted field and the direct field. Since there is no direct field in the shadow region, we observe that no such oscillation occurs. Unfortunately, it is not convenient to explicitly distinguish between the direct and diffracted field in (16-32). In many applications of diffraction theory, it is essential that we be able to

Line length is proportional to field. a_1, b_1, and c_1 correspond to positive values of a. At f_1 the direct field and diffracted fields are in-phase.

Note: The value of the relative field is 0.5 at the shadow boundary corresponding to line d_1.

A vector drawn from the origin to any point on the curve represents the magnitude of the Fresnel integral and the negative of its phase.

$u = 1.26$

(a)

(b)

Figure 16-11 (a) Use of the Cornu spiral in evaluating the Fresnel integral as a function of the parameter a. (b) Relative magnitude of the diffracted field in the vicinity of a shadow boundary. Refer to Fig. 16-10 for values of a.

mathematically distinguish between the direct and diffracted fields, as well as the reflected field that we have yet to consider.

Referring to Fig. 16-12, we can identify two shadow boundaries: the incident or direct field shadow boundary and the reflected field shadow boundary. These two shadow boundaries serve to divide space into three regions where region I contains direct and diffracted rays as well as reflected rays, region II direct and diffracted rays but no reflected rays, and region III only diffracted rays.

For a field in any one of the three regions, let us write $E(\rho, \phi)$ as consisting of a reflected field $v^r(\rho, \phi + \phi')$ and an incident field $v^i(\rho, \phi - \phi')$. Thus,

$$E(\rho, \phi) = \pm v^r(\rho, \phi + \phi') + v^i(\rho, \phi - \phi') \tag{16-33}$$

The choice of sign depends on the polarization of the incident field. If the electric field is perpendicular (parallel) to the diffracting edge, the plus (minus) sign is used. The field E at the point P must be a solution to the scalar wave equation with the appropriate boundary conditions. The boundary value problem depicted in Fig. 16-12 was first solved by Sommerfeld in 1896. We will first consider his solution. To do so, we must examine (16-33) more fully.

The first term in (16-33) gives the reflected fields, whereas the term $v^i(r, \phi - \phi')$ represents the incident field. If the ground plane were infinite in extent, the reflected field term would simply be the geometrical optics reflected field. However, in the case of the half-plane in Fig. 16-12, the reflected field will consist of two parts: namely, a geometrical optics reflected field and a diffracted field. Both parts of the reflected field will appear to originate from an image source behind the half-plane. Similarly, the incident field can be thought to consist of two parts: namely a geometrical optics incident field and a diffracted field. Thus, for the reflected field,

$$\pm v^r(\rho, \phi + \phi') = \pm[v^r_*(\rho, \phi + \phi') + v^r_B(\rho, \phi + \phi')] \tag{16-34}$$

and for the incident field,

$$v^i(\rho, \phi - \phi') = v^i_*(\rho, \phi - \phi') + v^i_B(\rho, \phi - \phi') \tag{16-35}$$

where v_* denotes the geometrical optics field and v_B the diffracted field. Thus, (16-33) may be thought of as being composed of four parts. Each of the terms on the right-hand side of (16-34) and (16-35) satisfies the wave equation individually except at the reflected field and incident field shadow boundaries, respectively. However, the sum of v^r_* and v^r_B makes v^r continuous across the reflected field shadow boundary and thus v^r satisfies the wave equation there. (Similar comments apply to v^i.) But, neither v^r nor v^i alone satisfies the boundary conditions at the wedge. However, the sum of v^r and v^i in (16-33) does satisfy the boundary conditions as well as the wave equation.

From simple geometrical considerations, we can see that for reflected geometrical optics rays, all points on a constant phase wavefront are given by

$$v^r_*(\rho, \phi + \phi') = e^{j\beta\rho\cos(\phi+\phi')}, \qquad 0 < \phi < \pi - \phi' \text{ in region I} \tag{16-36}$$

where the phase reference is taken to be at the edge of the half-plane in Fig. 16-13 since we are using a cylindrical coordinate system whose origin is on the edge of the wedge. By
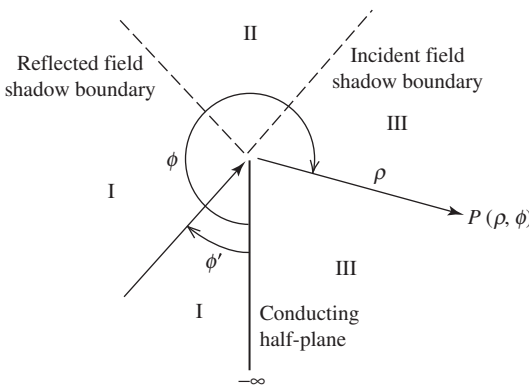
Figure 16-12 Diffraction by a conducting half-plane showing the location of shadow boundaries.

$$\frac{\rho_o}{\rho} = \cos\left[\pi - (\phi + \phi')\right]$$

$$\frac{\rho_o}{\rho} = -\cos(\phi + \phi')$$
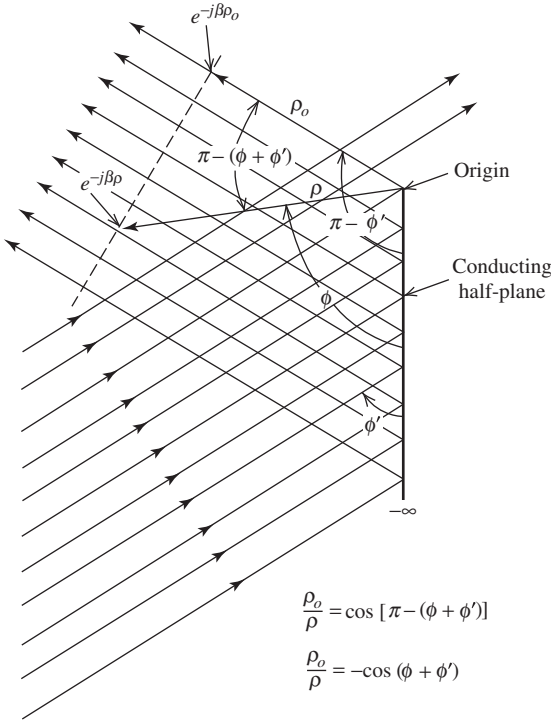
Figure 16-13 Geometry for the reflected field wavefront from a conducting half-plane.

similar considerations for direct incident rays, points on a constant phase wavefront are given by

$$v_*^i\left(\rho, \phi - \phi'\right) = e^{j\beta\rho\,\cos(\phi-\phi')}, \qquad 0 < \phi < \pi + \phi' \text{ in regions I and II} \qquad (16\text{-}37)$$

For other values of ϕ,

$$v_*^i = 0 = v_*^r \qquad (16\text{-}38)$$

In other words, v_* is zero in regions II and III for reflected rays and is zero in region III for direct or incident rays. It is the diffracted field that compensates for this discontinuity in the geometrical optics field as shown in Fig. 16-14. We see in Fig. 16-14a that for $\phi > 255°$, the total field is just the diffracted field and the total field is continuous across the incident field shadow boundary at $\phi = 255°$, where the value of the diffracted field is 0.5. For $105° < \phi < 255°$, the total field oscillates due to the interference between the incident field and the diffracted field. At $\phi = 105°$, the diffracted field again rises to 0.5 and the total field is continuous across the field shadow boundary. For $\phi = 105°$, the total field oscillates almost between 0 and 2 due mainly to the standing wave produced in region I by the incident and reflected fields and the fact that the field is observed at a constant distance $(\rho = 3\lambda)$ from the edge of the half-plane, requiring the observation point to sweep through the standing wave field. The electric field is normal to the half-plane at $\phi = 0$ and is nonzero there. Fig. 16-14b shows a time domain representation of the total electric field in the vicinity of the edge of the half-plane when a sinusoidal plane wave is incident at $\phi' = 75°$. Since the presentation is essentially a "snapshot" in time,
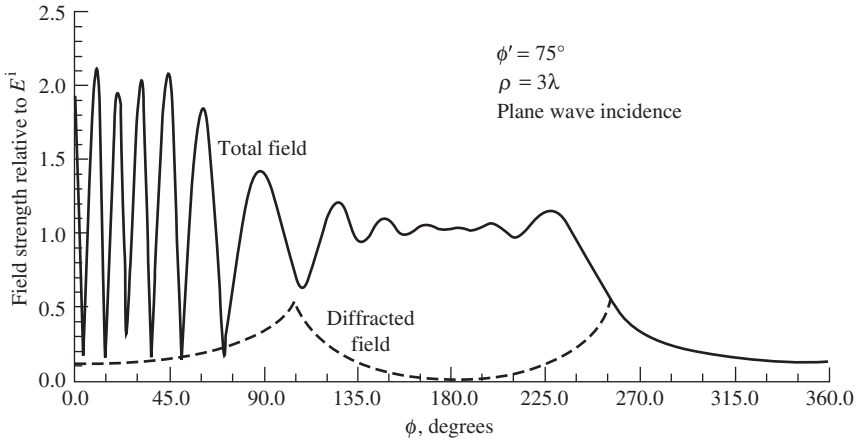
(a) Diffraction by a half-plane (frequency domain).
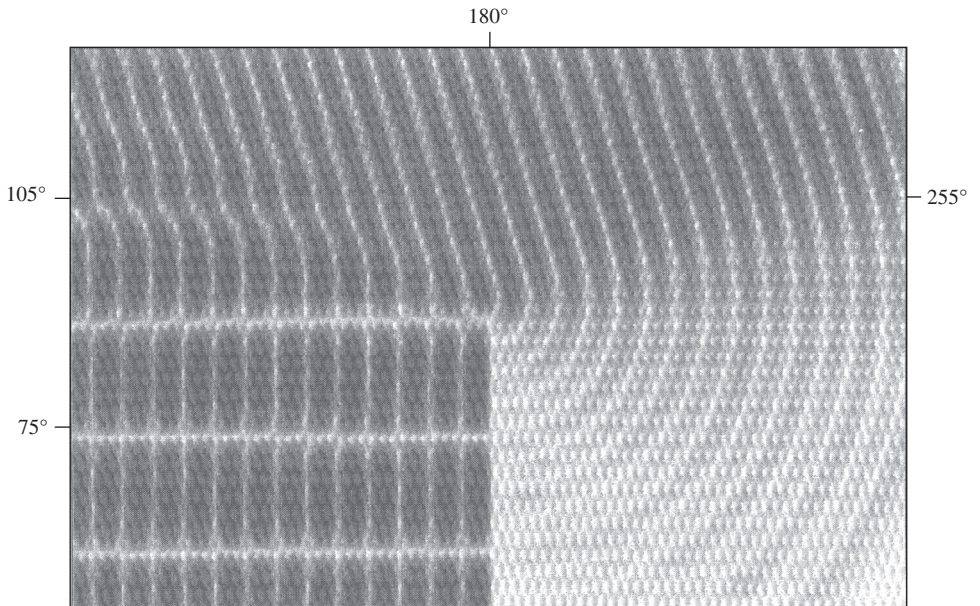
(b) FDTD calculated field distribution for a fixed instant of time in the vicinity of a half-plane. Sinusoidal plane wave incident at $\phi' = 75°$.

Figure 16-14 Diffraction by a half-plane. The incident field is polarized perpendicular to the edge of the half-plane. Refer to Figs. 16-12 and 16-13.

the almost white areas indicate zero field at an instant of time. (See Fig. 15-21 and associated text.) Note evidence of the reflection and shadow boundaries, the weak field when $\phi > 255°$, and the interference pattern when $\phi < 105°$. In the interference pattern for $\phi < 105°$ there is a standing wave in directions both normal and tangential to the half plane since $\phi' \neq 90°$.

Mathematical expressions for the diffracted field v_B have been a subject of considerable research in the past several decades in an effort to improve on the early classical work of Sommerfeld [7]. For the half-plane problem of Fig. 16-12, Sommerfeld obtained

an expression for the diffracted field due to an incident plane wave in terms of the Fresnel integral. This expression is[3]

$$v_B(\rho, \phi^{\pm}) = -e^{j(\pi/4)} \sqrt{\frac{2}{\pi\alpha}} \cdot e^{j\beta\rho \cos \phi^{\pm}} \cos \frac{\phi^{\pm}}{2} \int_{\sqrt{\alpha\beta\rho}}^{\infty} e^{j\tau^2} \, d\tau \qquad (16\text{-}39)$$

where

$$\phi^{\pm} = \phi \pm \phi' \qquad (16\text{-}40)$$

and

$$\alpha = 1 + \cos \phi^{\pm} \qquad (16\text{-}41)$$

We note that this solution is in a form somewhat similar to that of (16-32). The mathematical details of deriving the above are beyond the scope of this text.

Sommerfeld's work was more general than that of just a half-plane. He also considered the more general case of a plane wave illuminating a conducting wedge of interior angle $(2 - n)\pi$, where $0 < n \le 2$.[4] For this case, he obtained an asymptotic evaluation of a contour integral representation for the diffracted field that is given by

$$v_B(\rho, \phi^{\pm}) = \frac{e^{-j(\beta\rho+\pi/4)}}{\sqrt{2\pi\beta\rho}} \frac{(1/n) \sin(\pi/n)}{\cos(\pi/n) - \cos(\phi^{\pm}/n)} \qquad (16\text{-}42)$$

Unfortunately, this asymptotic form yields infinite fields in the immediate vicinity of the shadow boundary [8]. The region near a shadow boundary is usually referred to as a *transition region*. Eq. (16-42) is only valid outside a transition region where the condition

$$\beta\rho \left(\cos \frac{\pi}{n} - \cos \frac{\phi^{\pm}}{n} \right)^2 \gg 1 \qquad (16\text{-}43)$$

is satisfied. This condition is always met if the quantity $\beta\rho(1 + \cos \phi^{\pm})$ is large, which means that the observation point at $P(\rho, \phi, z)$ must be at a large electrical distance from the diffracting edge. Nevertheless, (16-42) is a useful one if the observation point is not near a shadow boundary and the above conditions are met.

In 1938, Pauli [9] improved on the work of Sommerfeld by obtaining a series form for Sommerfeld's contour integral solution. Pauli's result is given by

$$v_B(\rho, \phi^{\pm}) = \frac{2e^{j(\pi/4)}}{n\sqrt{\pi}} \frac{\sin(\pi/n)}{\cos(\pi/n) - \cos(\phi^{\pm}/n)} \left| \cos \frac{\phi^{\pm}}{2} \right| \cdot e^{j\beta\rho \cos \phi^{\pm}} \int_{\sqrt{\alpha\beta\rho}}^{\infty} e^{-j\tau^2} \, d\tau$$

$$+ \text{[higher-order terms]} \qquad (16\text{-}44)$$

---

[3] Note that in (16-39), we are really writing two equations, one for $v_B^r(r, \phi^+)$ and the other for $v_B^i(r, \phi^-)$. The use of the notation $\phi^{\pm}$ is for convenience and the reader should keep in mind that wherever it appears there are two separate equations implied, one associated with the reflected field and one associated with the incident field.

[4] Refer forward to Fig. 16-15. Note, the parameter n as used here is not to be confused with the index of refraction.

This expression is far more accurate, particularly near the shadow boundaries, than (16-42) while being only slightly more difficult to evaluate. It is valid for $0 < n \leq 2$. In the case of the half-plane $(n = 2)$, the higher-order terms are identically zero and Pauli's result in (16-44) reduces to that of Sommerfeld in (16-39). Pauli's expression was the first practical formulation of Sommerfeld's original solution that included a finite observation distance.

EXAMPLE 16-1   *Sample Wedge Diffraction Calculations*

The use of Eqs. (16-39), (16-42), and (16-44) is best illustrated by an example. Let us calculate the diffracted field in Fig. 16-14 for $\phi = 250°$. Assume $\lambda = 1$m. Using (16-39), we obtain

$$v_B(3, \phi^-) = (-9.146 - j9.146)(0.0436) \cdot (0.997 + j0.0717)(0.359 - j0.620)$$
$$= -0.397 + j0.0760$$

$$v_B(3, \phi^+) = (-0.418 - j0.418)(-0.954) \cdot (-0.964 + j0.264)(-0.0237 + j0.0820)$$
$$= 0.0345 - j0.0335$$

Thus, the exact Sommerfeld solution gives for the diffracted field magnitude $|-0.3625 + j0.0435| = 0.365$, which agrees with Fig. 16-14. Using (16-44), we should obtain the same result for the half-plane case since Pauli's equation reduces to Sommerfeld's. Thus,

$$v_B(3, \phi^-) = (0.798 + j0.798)(-11.46)(0.0436) \cdot (0.997 + j0.0717)(0.359 - j0.620)$$
$$= -0.397 + j0.0760$$

$$v_B(3, \phi^+) = (0.798 + j0.798)(0.524) \cdot (0.954)(-0.964 + j0.264) \cdot (-0.0237 + j0.0820)$$
$$= 0.0345 - j0.0335$$

and the diffracted field magnitude is once again 0.365. We note that since $\phi = 250°$ is near the incident field shadow boundary, $v_B(3, \phi^-)$ is the major contributor to the diffracted field and $v_B(3, \phi^+)$, which is associated with the reflected field shadow boundary, makes only a minor contribution. Both (16-39) and (16-44) would go to infinity precisely at the shadow boundary $\phi = 255°$ (or $\phi = 105°$). For this reason, we have elected to use $\phi = 250°$ in this example. Finally, let us use the asymptotic form in (16-42). Thus,

$$v_B(3, \phi^-) = (0.065 - j0.065)(-11.46)$$
$$= -0.745 + j0.745$$
$$v_B(3, \phi^+) = (0.065 - j0.065)(0.524)$$
$$= 0.034 - j0.034$$

and the magnitude of the diffracted field alone exceeds unity or that of the incident field. This result is in error because the condition in (16-43) has been violated. The result would be only 10% in error at $\rho = 10\lambda$ if $\phi = 255° \pm 12°$, at 20λ if $\phi = 255° \pm 5°$, at 30λ if $\phi = 255° \pm 4°$, and at 100λ if $\phi = 255° \pm 3°$. However, no matter how large ρ is, the asymptotic form will be singular right at the shadow boundary.

Starting in 1953, it was Keller [5, 6] who systematically developed the geometrical theory of diffraction, or GTD as it is often referred to. In his work, he has called the quantities $D(\phi^+)$ and $D(\phi^-)$ diffraction coefficients, where

$$\left[v_B^i(\rho, \phi^-) \mp v_B^r(\rho, \phi^+)\right] = [D(\phi^-) \mp D(\phi^+)]\frac{e^{-j\beta\rho}}{\sqrt{\rho}} \qquad (16\text{-}45)$$

and used the asymptotic expression of Sommerfeld in (16-42) to calculate the diffracted field due to plane wave incidence. The postulates of Keller's theory are:

1. The diffracted field propagates along ray paths that include points on the boundary surface. These ray paths obey the principle of Fermat, known also as the principle of shortest optical path.
2. Diffraction, like reflection and transmission, is a local phenomenon at high frequencies. That is, it depends only on the nature of the boundary surface and the incident field in the immediate neighborhood of the point of diffraction.
3. A diffracted wave propagates along its ray path so that:
   a. power is conserved in a tube of rays, and
   b. phase delay equals the wave number times the distance along the ray path.

A consequence of the second postulate is that the diffracted fields caused by the edge of the infinite wedge in Fig. 16-13, for example, appear to be cylindrical wave fields that originate at the wedge edge. This is consistent with the $(\rho)^{-1/2}$ dependence in (16-45).

The simple ray formulation of Keller's geometrical theory of diffraction is restricted to the calculation of fields in regions of space that exclude transition regions adjacent to shadow boundaries, caustics, and focal points. To calculate the field at such points, additions and modifications to the geometrical theory of diffraction are required. Further, if the incident field is not a plane wave, but a cylindrical or spherical wave, GTD must be modified to accept these incident fields as well. These various modifications will be considered in later sections.

16.3 THE RAY-FIXED COORDINATE SYSTEM

In the previous section, we considered the scalar diffracted field due to a plane wave normally incident (i.e., traveling in the negative ρ-direction) on a perfectly conducting infinite wedge whose edge was along the z-axis. Such a coordinate system is said to be an edge-fixed coordinate system. On the other hand, the obliquely incident and diffracted rays associated with the point Q in Fig. 16-15 are more conveniently described in terms of spherical coordinates centered at Q. Such a coordinate system is said to be ray-fixed [1]. Let the position of

Figure 16-15 Geometry for three-dimensional wedge diffraction problem.

the source of the incident ray be defined by the spherical coordinates (s', γ'_o, ϕ'), and the observation point by the coordinates (s, γ_o, ϕ) as indicated in Fig. 16-15. Note that the point Q is a unique point on the edge for a given source location and observation point.

The plane containing the incident ray and the edge of the wedge will be referred to as the plane of incidence, whereas that plane containing the diffracted ray and the edge of the wedge will be referred to as the plane of diffraction. The unit vector $\hat{s}'$ is in the direction of incidence and the unit vector $\hat{s}$ in the direction of diffraction. It is then apparent that the unit vectors $\hat{\gamma}'_o$ and $\hat{\phi}'$ are parallel and perpendicular, respectively, to the plane of incidence, and that the unit vectors $\hat{\gamma}_o$ and $\hat{\phi}$ are parallel and perpendicular, respectively, to the plane of diffraction as shown in Fig. 16-16. γ'_o and γ_o are angles less than $\pi/2$ measured from the edge to the incident and diffracted rays, respectively, whereas $\hat{\gamma}'_o$ and $\hat{\gamma}_o$ are the implied unit vectors. Further, ϕ' and ϕ are angles measured from one face of the wedge to the plane of incidence and diffraction, respectively, whereas $\hat{\phi}'$ and $\hat{\phi}$ are the implied unit vectors. Note that ϕ' and ϕ are measured from the same face of the wedge.

Let us write a symbolic expression for the diffracted field in matrix form as

$$[\mathbf{E}^d] = [D][\mathbf{E}^i]A(\rho)e^{-j\beta\rho} \tag{16-46}$$

where $[\mathbf{E}^d]$ and $[\mathbf{E}^i]$ are column matrices consisting of the scalar components of the diffracted and incident fields respectively, $[D]$ is a square matrix of the appropriate scalar diffraction coefficients, and ρ is the distance from the wedge edge to the observation point, and $A(\rho)$ is a spreading factor. Now if the edge-fixed coordinate system is used, it is clear that $[\mathbf{E}^d]$ will have, in general, three scalar components E^d_ρ, E^d_ϕ, and E^d_z, and $[D]$ will be a 3×3 matrix. It can be shown that in such a situation seven of the nine terms in $[D]$ are nonvanishing. However, when the ray-fixed coordinate system is used, there is no (radial) component of the diffracted field in the direction of the diffracted ray tube since the incident field is not allowed to have a component in the direction of the incident ray tube. It follows that there are then only two possible components of the diffracted field, E^d_γ and E^d_ϕ, and only two components of the incident field, $E^i_{\gamma'}$ and $E^i_{\phi'}$. Clearly, $[D]$ is then a 2×2 matrix. In this case, $[D]$ has nonvanishing terms on the main diagonal. Thus, for plane wave incidence in the ray-fixed system, (16-46) can be written as

$$\begin{bmatrix} E^d_\gamma(s) \\ E^d_\phi(s) \end{bmatrix} = \begin{bmatrix} -D_\parallel & 0 \\ 0 & -D_\perp \end{bmatrix} \begin{bmatrix} E^i_{\gamma'}(Q) \\ E^i_{\phi'}(Q) \end{bmatrix} A(s)e^{-j\beta s} \tag{16-47}$$
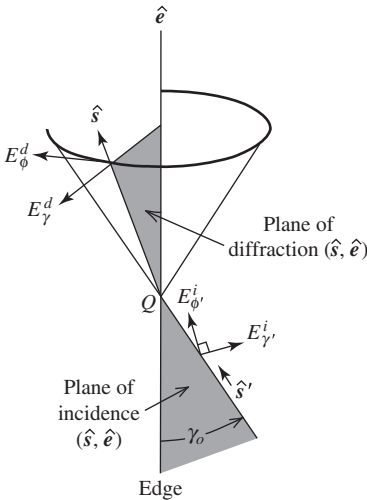
Figure 16-16 Ray-fixed coordinate system.

where the scalar diffraction coefficients $D_\parallel$ and $D_\perp$ are momentarily undefined and will be given in the following section.

It is apparent that the use of the ray-fixed coordinate system instead of the edge-fixed system reduces the diffraction matrix from a 3×3 matrix with seven nonvanishing terms to a 2×2 matrix with but two nonvanishing terms. Thus, the ray-fixed system is the natural coordinate system to be used for wedge diffraction and the importance of using it can hardly be overemphasized.

We have chosen to use the notation $D_\parallel$ in association with $E^i_{\gamma'}(Q)$ and $D_\perp$ in association with $E^i_{\phi'}(Q)$ not because $E^i_{\gamma'}$ and $E^i_{\phi'}$ are parallel and perpendicular, respectively, to the diffracting edge (which they are at normal incidence when $\gamma'_o = 90°$), but because $E^i_{\gamma'}$ and $E^i_{\phi'}$ are parallel and perpendicular, respectively, to the plane of incidence as shown in Fig. 16-16.

Since $E^i_{\gamma'}$ and $E^i_{\phi'}$ are parallel and perpendicular, respectively, to the plane of incidence, we will let $E^i_{\gamma'}$ be written as $E^i_\parallel$ and $E^i_{\phi'}$ as $E^i_\perp$. Similarly, $E^d_\gamma(s)$ and $E^d_\phi(s)$ are parallel and perpendicular, respectively, to the plane of diffraction as shown in Fig. 16-16. Thus, we let E^d_γ be written as $E^d_\parallel$ and E^d_ϕ as $E^d_\perp$. With these notational changes, (16-47) may be rewritten as

$$\begin{bmatrix} E^d_\parallel(s) \\ E^d_\perp(s) \end{bmatrix} = \begin{bmatrix} -D_\parallel & 0 \\ 0 & -D_\perp \end{bmatrix} \begin{bmatrix} E^i_\parallel(Q) \\ E^i_\perp(Q) \end{bmatrix} A(s) e^{-j\beta s} \qquad (16\text{-}48)$$

We will use this notation throughout the remainder of the chapter, keeping in mind that when the $\parallel$ and $\perp$ symbols are associated with E^i, reference to the plane of incidence is implied. When the $\parallel$ and $\perp$ symbols are associated with E^d, reference to the plane of diffraction is implied.

16.4 A UNIFORM THEORY OF WEDGE DIFFRACTION

The modern version of GTD can be divided into the two basic canonical problems of wedge diffraction and curved surface diffraction plus the lesser but more complex problems of vertex diffraction, tip diffraction, and other higher-order phenomena. In the application of wedge diffraction to antenna problems, the important features of antennas are modeled by perfectly conducting wedges. For example, the sectoral horn antenna can be modeled by two half-planes as shown in Fig. 16-18 for the purpose of analyzing the E-plane pattern [8]. In such a problem, however, it is necessary to use cylindrical wave diffraction coefficients instead of plane wave diffraction coefficients as in Sec. 16.2. The first use of cylindrical wave diffraction in the treatment of antenna problems, such as in Sec. 16.5, was by Rudduck [8], who used Pauli's formulation together with the principle of reciprocity to calculate the necessary cylindrical wave diffraction. Problems involving spherical wave diffraction are also common.

In Sec. 16.2, some early developments in the study of diffraction by a conducting wedge were presented. We saw that although some of the formulas presented are certainly useful for some engineering calculations, they are limited in their accuracy in a transition (shadow boundary) region [e.g., (16-42)], or when the observation point is near $(r < \lambda)$ the diffracting edge [e.g., (16-44)]. It would obviously be useful and convenient if there were available to us a theory of wedge diffraction having the property that it could accurately predict the diffracted field in such places as the transition regions or near the diffracting edge without the necessity for considering each type of incident field separately. Such a theory is available and is known as a *uniform theory* of wedge diffraction because it applies in all situations consistent with the postulates of the geometrical theory of diffraction given in Sec. 16.2. It is the purpose of this section to present the important results in this theory, known as UTD, which is based on the numerous works of Kouyoumjian and Pathak [10–12].

In 1967, Kouyoumjian and co-workers obtained a generalized version of Pauli's result [i.e., (16-44)] with the resultant diffraction function v_B expressed as $v_B(L, \phi^\pm)$, where L is a distance parameter more general than just the distance ρ used in Sec. 16.2, whereas $\phi^\pm$ retains the meaning used previously. In their work, the distance parameter is given by

$$
L = \begin{cases}
s \sin^2 \gamma_o' & \text{for plane waves} \\[2ex]
\dfrac{\rho' \rho}{\rho + \rho'} & \text{for cylindrical waves} \\[3ex]
\dfrac{s's \sin^2 \gamma_o'}{s + s'} & \text{for conical and spherical waves } \bullet
\end{cases}
\tag{16-49}
$$

We note immediately that L is dependent on the type of incident wave and the angle of incidence γ_o' (which equals the angle of reflection γ_o) as well as the distances involved. The distance parameter L in (16-49) can be found by imposing the condition that the total field, which is the sum of the geometrical-optics field and the diffracted field, be continuous at shadow or reflection boundaries.

When the work of Kouyoumjian and co-workers is expressed in terms of the scalar diffraction coefficients $D_\parallel$ and $D_\perp$, where

$$
D_\parallel(L, \phi, \phi') = [v_B(L, \phi^-) - v_B(L, \phi^+)] \frac{\sqrt{L} e^{j\beta L}}{\sin \gamma_o}
\tag{16-50}
$$

$$
D_\perp(L, \phi, \phi') = [v_B(L, \phi^-) + v_B(L, \phi^+)] \frac{\sqrt{L} e^{j\beta L}}{\sin \gamma_o}
\tag{16-51}
$$

We have (without proof) [10, 12]

$$
D_{\substack{\parallel \\ \perp}}(L, \phi, \phi') = \frac{-e^{-j(\pi/4)}}{2n\sqrt{2\pi\beta}\,\sin\gamma_o'}
$$

$$
\times \left[\cot\left(\frac{\pi + (\phi - \phi')}{2n} \right) F[\beta L a^+(\phi - \phi')] \right.
$$

$$
+ \cot\left(\frac{\pi - (\phi - \phi')}{2n} \right) F[\beta L a^-(\phi - \phi')]
$$

$$
\mp \left\{ \cot\left(\frac{\pi + (\phi + \phi')}{2n} \right) F[\beta L a^+(\phi + \phi')] \right.
$$

$$
\left. \left. + \cot\left(\frac{\pi - (\phi + \phi')}{2n} \right) F[\beta L a^-(\phi + \phi')] \right\} \right]
\tag{16-52}
$$

where, if the argument of F is represented by X,

$$
F(X) = 2j \left| \sqrt{X} \right| e^{jX} \int_{|\sqrt{X}|}^{\infty} e^{-j\tau^2} \, d\tau
\tag{16-53}
$$

Again, we see that a Fresnel integral appears in the expression for the diffraction coefficient. The factor $F(X)$ may be regarded as a correction factor to be used in the transition regions of the shadow and reflection boundaries. Outside of the transition
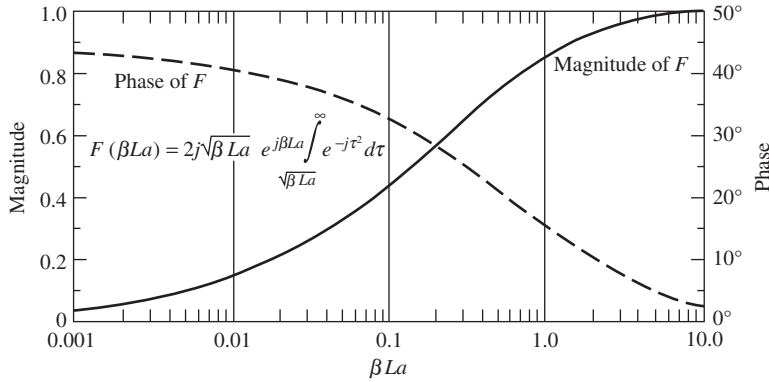
Figure 16-17 Magnitude and phase of the transition function $F(\beta La)$, where $a = a^+$ or a^-.

regions where the argument of F exceeds about 3, the magnitude of F is approximately equal to 1, as Fig. 16-17 shows. Even within a given transition region, usually only one of the four terms in (16-52) is significantly different from unity. The transition function that is significantly different from unity goes to zero at the same rate that its cotangent multiplier goes to infinity. Thus the transition function prevents the singularity in (16-42) from occurring in (16-52) (see Prob. 16.4-7).

The argument of the transition function, which is $X = \beta La^{\pm}(\phi \pm \phi')$, may be calculated for a known value of βL if $a^{\pm}$ as a function of $(\phi \pm \phi')$ is known. To determine $a^+(\phi \pm \phi')$ and $a^-(\phi \pm \phi')$, we use

$$a^{\pm}(\phi \pm \phi') = 2 \, \cos^2\left[\frac{2n\pi N^{\pm} - (\phi \pm \phi')}{2}\right] \tag{16-54}$$

in which $N^{\pm}$ are the integer that most nearly satisfy the four equations

$$2\pi n N^+ - (\phi \pm \phi') = \pi \tag{16-55}$$

and

$$2\pi n N^- - (\phi \pm \phi') = -\pi \tag{16-56}$$

We note that N^+ and N^- may each have two separate values in a given problem. For exterior wedge diffraction where $1 < n \leq 2$, $N^+ = 0$ or 1, but $N^- = -1, 0$ or 1. The factor $a^{\pm}(\phi \pm \phi')$ may be interpreted physically as a measure of the angular separation between the field point and a shadow or reflection boundary.

Now that we have all the necessary relationships to calculate $D_{\parallel}$ and $D_{\perp}$, we repeat (16-47) in the format of UTD as

$$\begin{bmatrix} E_{\parallel}^d(s) \\ E_{\perp}^d(s) \end{bmatrix} = \begin{bmatrix} -D_{\parallel} & 0 \\ 0 & -D_{\perp} \end{bmatrix} \begin{bmatrix} E_{\parallel}^i(Q) \\ E_{\perp}^i(Q) \end{bmatrix} A(s)e^{-j\beta s} \tag{16-57}$$

where the spatial attenuation factor $A(s)$ is defined as

$$A(s) = \begin{cases} \dfrac{1}{\sqrt{s}} & \text{for plane, cylindrical, and conical wave incidence} \\[2ex] \left[\dfrac{s'}{s(s'+s)}\right]^{1/2} & \text{for spherical wave incidence} \end{cases} \tag{16-58}$$

It should also be mentioned that, since diffraction concepts apply to acoustical problems, the diffraction coefficients $D_\parallel$ and $D_\perp$ in (16-57) are sometimes written D_s and D_h, respectively, which correspond to the acoustic soft and hard boundary conditions [14]. Software code based upon (16-52) for UTD is available in the previous editions of this text.

EXAMPLE 16-2   *Sample UTD Calculation*

The use of (16-49) to (16-58) is best illustrated by an example. Suppose we wish to calculate the diffracted field in Fig. 16-14 when $\phi = 250°$. We have in this case: $\phi + \phi' = 325°$; $\phi - \phi' = 175°$; $L = 3\lambda$; $\beta L = 6\pi$; $n = 2$; $\lambda = 1$m. Thus, from (16-54) to (16-56)

$$a^+(\phi + \phi') = 2 \ \cos^2(197.5°), \qquad \text{where } N^+ = 1$$
$$a^+(\phi - \phi') = 2 \ \cos^2(87.5°), \qquad \text{where } N^+ = 0$$
$$a^-(\phi + \phi') = 2 \ \cos^2(162.5°), \qquad \text{where } N^- = 0$$
$$a^-(\phi - \phi') = 2 \ \cos^2(87.5°), \qquad \text{where } N^- = 0$$

From Fig. 16-17, using the respective values of a^+ and a^- above, we obtain

$$F(6\pi \cdot 1.819) = 0.999 + j0.0146$$
$$F(6\pi \cdot 0.0038) = 0.318 + j0.216$$
$$F(6\pi \cdot 1.819) = 0.999 + j0.0146$$
$$F(6\pi \cdot 0.0038) = 0.318 + j0.216$$

Using (16-52) and (16-58), we obtain

$$D_\perp(L, \phi, \phi') = -0.628 + j\,0.0735$$
$$A(s)e^{-j\beta s} = 0.577$$

From (16-57),

$$E_\perp^d(s) = -0.363 + j0.0424$$

or

$$\left| E_\perp^d(s) \right| = 0.365$$

which agrees with Fig. 16-14. It is worth noting that when the four correction factors F above are multiplied by their associated cotangent factor, it is the fourth term above that is much larger than the others. As mentioned earlier, usually just one of the terms in (16-52) turns out to be large, even close to a shadow boundary. Eq. (16-52) will not exhibit a singular behavior at a shadow boundary as was the case in Sec. 16.2 with (16-39) and (16-42).

If the field point is not close to a shadow or reflection boundary and $\phi' \neq 0$ or $n\pi$ (grazing incidence), the scalar diffraction coefficients $D_\parallel$ and $D_\perp$ reduce to Keller's diffraction coefficients [see (16-42) and (16-45)] that may be written as

$$D_{\genfrac{}{}{0pt}{}{\parallel}{\perp}} (\phi, \phi'; \gamma_o') = \frac{e^{-j(\pi/4)} \sin(\pi/n)}{n\sqrt{2\pi\beta} \ \sin\gamma_o'} \cdot \left[\frac{1}{\cos\dfrac{\pi}{n} - \cos\dfrac{\phi - \phi'}{n}} \mp \frac{1}{\cos\dfrac{\pi}{n} - \cos\dfrac{\phi + \phi'}{n}} \right] \quad (16\text{-}59)$$

This expression is valid for all four types of incident waves given in (16-49), which is important because the diffraction coefficient should be independent of the edge illumination away from shadow and reflection boundaries. However, from Sec. 16.2, we know that (16-59) will become singular as a shadow or reflection boundary is approached.

Grazing incidence, where $\phi' = 0$ or $n\pi$, is a special case that must be considered separately. In this case, $D_\parallel \approx 0$, and the expression for E_1^d must be multiplied by a factor of $\frac{1}{2}$. If we consider grazing incidence to be the limit of oblique incidence, we can see how the need for the factor of $\frac{1}{2}$ arises, because at grazing incidence the incident and reflected fields merge. When they merge, one-half of the total field propagating along the face of the wedge toward the edge is the incident field and the other half is the reflected field. The merged field is then regarded as being the "incident" field, but it is too large by a factor of 2 and the factor of $\frac{1}{2}$ becomes necessary. That is, (16-57) requires the use of the free-space incident field and not the merged field.

The uniform theory of wedge diffraction described in this section permits us to consider diffraction problems wherein both the source and observation points are quite close to the diffracting edge (i.e., a wavelength or even less). It also permits us to consider any type of TEM incident field. A more general expression for L, valid for an arbitrary wavefront incident on the straight edge of a wedge, appears in the literature [11, 12].

Unlike the edge diffraction formulas presented in Sec. 16.2, (16-52) is valid in the transition regions of the incident field shadow boundary and the reflected field shadow boundary. Eq. (16-52) cannot be used to calculate the field at a caustic of the diffracted ray. This does not conflict with the concept of a uniform theory of wedge diffraction because geometrical optics itself is incapable of determining the field at a caustic. The field at a caustic may, however, be found through the use of a supplementary solution in the form of an integral representation of the field. The equivalent sources in the integral representation are determined from a suitable high-frequency approximation such as geometrical optics or the geometrical theory of diffraction. The calculation of the field at a caustic by such methods will be considered in Sec. 16.9.

16.5   *E*-PLANE ANALYSIS OF HORN ANTENNAS

To illustrate the application of the uniform theory of diffraction presented in the previous section, consider the *E*-plane horn antenna shown in Fig. 16-18*a*. In this section, we use
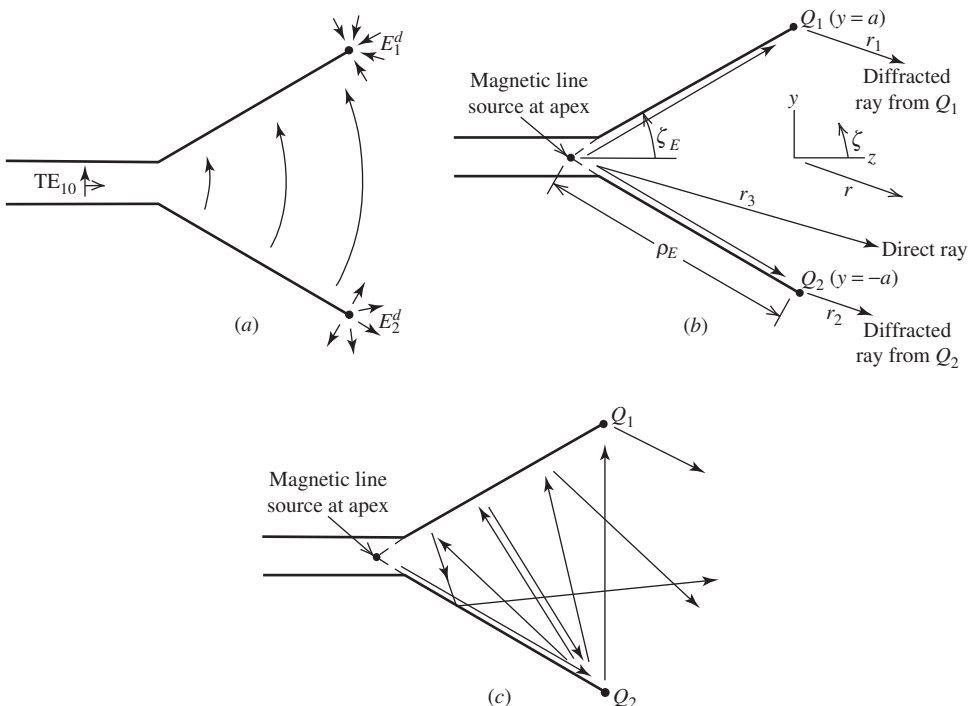
Figure 16-18 Diffraction by a horn antenna. (*a*) *E*-plane sectoral horn. (*b*) Model of *E*-plane sectoral horn. (*c*) Neglected rays.

the model shown in Fig. 16-18*b* to compute the complete *E*-plane pattern of the horn antenna. The model is simple and therefore particularly well-suited to use as a first example of the application of UTD. The model has only three sources of radiation and is two-dimensional in nature (i.e., infinite in the $\pm x$-directions), which in the *E*-plane well-represents a three-dimensional horn antenna.

The equations applicable to the analysis are as follows. Note that the angle ζ $(0 \leq \zeta \leq 2\pi)$ is used instead of θ $(0 \leq \theta \leq \pi)$ so that positions in the *yz*-plane may be defined unambiguously. In the far field, we have (in the *yz*-plane)

$$r_1 = r - a \, \sin \zeta \tag{16-60}$$

$$r_2 = r + a \, \sin \zeta \tag{16-61}$$

$$r_3 = r + \rho_E \, \cos \zeta \, \cos \zeta_E \tag{16-62}$$

where r_1 and r_2 are distances to the far-field observation point $P(r, \zeta)$ from diffracting edges 1 and 2, respectively, and r_3 is the distance from the line source to the far-field observation point as shown in Fig. 16-18*b*. Thus, the incident field along the direct ray can be expressed by

$$E^i(P) = \frac{e^{-j\beta r_3}}{\sqrt{r_3}} \approx \frac{e^{-j\beta r}}{\sqrt{r}} e^{j\beta \rho_E \cos \zeta \, \cos \zeta_E}, \qquad -\zeta_E \leq \zeta \leq \zeta_E \tag{16-63}$$

and

$$E^i(P) = 0, \qquad \zeta_E < \zeta < 2\pi - \zeta_E \tag{16-64}$$

Note that in applying UTD, we do not replace the conducting surfaces with equivalent currents radiating in free space as in the preceding chapters of this book. Instead, the conducting surfaces are retained. As a consequence, for example, $E^i(P) = 0$ when $\zeta > \zeta_E$.

The edge diffracted field at $P(r, \zeta)$ from a diffraction point Q_1 on the "top" edge may be written as

$$E_1^d(P) = \frac{1}{2} E_\perp^i(Q_1) D_\perp(L, \phi, \phi') \frac{e^{-j\beta r_1}}{\sqrt{r_1}} = \frac{1}{2} E_\perp^i D_\perp(L, \phi, \phi') \frac{e^{-j\beta r}}{\sqrt{r}} e^{j\beta a \, \sin \zeta},$$
$$-\frac{\pi}{2} \leq \zeta \leq \pi + \zeta_E \tag{16-65}$$

and

$$E_1^d(P) = 0, \qquad \pi + \zeta_E < \zeta < \frac{3\pi}{2} \tag{16-66}$$

Similarly, the diffracted field at $P(r, \zeta)$ from a diffraction point Q_2 on the "bottom" edge may be written

$$E_2^d(P) = \frac{1}{2} E_\perp^i(Q_2) D_\perp(L, \phi, \phi') \frac{e^{-j\beta r_2}}{\sqrt{r_2}} = \frac{1}{2} E_\perp^i D_\perp(L, \phi, \phi') \frac{e^{-j\beta r}}{\sqrt{r}} e^{-j\beta a \, \sin \zeta},$$
$$-\pi - \zeta_E \leq \zeta \leq \frac{\pi}{2} \tag{16-67}$$

and

$$E_2^d(P) = 0, \qquad \frac{\pi}{2} < \zeta < \pi - \zeta_E \tag{16-68}$$

where

$$E_\perp^i(Q_1) = E_\perp^i(Q_2) = \frac{e^{-j\beta\rho_E}}{\sqrt{\rho_E}} \tag{16-69}$$

Thus, the total field at an observation point $P(r, \zeta)$ may be written as the scalar sum

$$E(P) = E^i(P) + E_1^d(P) + E_2^d(P) \tag{16-70}$$

In the above equations, scalar $D_\perp$ denotes the diffraction coefficient at the point of diffraction Q_m for the case where the incident electric field is normal to the edge. The diffraction coefficient at Q_m depends on the geometry of the incident and diffracted rays at Q_m and is most accurately given by (16-49) and (16-52). Here, of course, we consider the incident field to be cylindrical and use the cylindrical wave form for the distance parameter L. $E^i(Q_m)$ is the incident field that is perpendicular to both the edge and incident ray.

At first glance, the factor of $\frac{1}{2}$ in (16-65) and (16-67) might appear to be incorrect. However, in this problem, the rays from the line source are incident at a grazing angle with the surface of the horn walls and therefore deserve special consideration. Grazing incidence, where $\phi' = 0$ or $n\pi$, requires that $D_\perp$ in (16-57) be multiplied by a factor of $\frac{1}{2}$ as discussed in the preceding section below (16-59).

Fig. 16-19 shows results calculated with the model shown in Fig. 16-18b and also experimental data. The agreement between the calculated results without using double diffractions (dashed curve) and the experimental results is seen to be very good. Note that there is a discontinuity in the calculated results when $\zeta = 90°$ (or $270°$). This discontinuity may be removed simply by including rays that diffract from Q_2 (or Q_1) and travel across the horn aperture to Q_1 (or Q_2) and are diffracted a second time as indicated in Fig. 16-18c.

Also shown in Fig. 16-18c are several other rays that have not been included in the calculated results because in this problem they provide a relatively weak numerical contribution. Strictly speaking, those rays shown in Fig. 16-18c that do not involve double diffractions should be included in the analysis. These are the two rays that experience a

Figure 16-19 Calculated and experimental E-plane patterns of an E-plane sectoral horn.

reflection after undergoing diffraction at Q_2 (or Q_1, which are not shown). Of the two doubly diffracted rays shown, here only the one from Q_2 to Q_1 is important, because it compensates for the shadowing of Q_2 when $\zeta > 90°$. There is no similar compensation needed in the case of the other doubly diffracted ray that goes from Q_2 to the "top" wall and back to Q_2.

In conclusion, we have used a simple model (i.e., Fig. 16-18b) to calculate the E-plane pattern of horn antennas with good results. Strictly speaking, we should have included some of the rays in Fig. 16-18c, but did not do so for the sake of simplicity without a loss of accuracy. It is a fundamental fact that in applying UTD (or GTD), one must be careful to identify and include all rays that arise in the problem. In the horn problem here, we were able to omit some of the rays only because they were not in or near a transition region, and because the rays in Fig. 16-18b are one or more orders of magnitude stronger than those in Fig. 16-18c.

16.6 CYLINDRICAL PARABOLIC REFLECTOR ANTENNAS

As a second example of the application of UTD, we consider the cylindrical parabolic antenna shown in Fig. 16-20. We use the aperture integration procedure given in Chap. 9 to obtain the pattern in and near the main beam, but use UTD to compute the pattern everywhere else. As in the study of the horn antenna in the previous section, the model here is two-dimensional. We consider only the diffractions that occur at the edges of the parabolic surface and ignore any higher-order rays associated with the curved surface (e.g., see Sec. 16.11). A simple model was presented in Example 5-3.

First, let us consider the equation for obtaining the main beam and first few side lobes. From Sec. 9.1, we may write for the far field E^A obtained by aperture integration

$$E^A(P) = \sqrt{\frac{j\beta}{2\pi r}} e^{-j\beta r} \cos \zeta \int_{-a}^{a} \frac{F_f(\theta_s)}{\sqrt{\rho}} e^{j\beta y' \sin \zeta} \, dy' \qquad (16\text{-}71)$$

where $F_f(\theta_s)$ is the pattern of the electric line source current I^e that serves as the feed for the cylindrical parabolic reflector antenna. (If the line source pattern is isotropic, $F_f(\theta_s) = 1$.) Eq. (16-71) is a two-dimensional specialization of the equations in Sec. 9.1.

Figure 16-20 Cylindrical parabolic antenna geometry.

Eq. (16-71) can, of course, give us the pattern for $90° \geq \zeta \geq -90°$. However, since we must perform the aperture integration anew for each value of ζ, it is more efficient in the computational sense to use (16-71) for $\zeta_o \geq \zeta \geq -\zeta_o$, where ζ_o is the angular extent of the main beam and the first side lobe or two, and then to use UTD for the remainder of the pattern. Clearly, we do not use aperture integration and UTD simultaneously in the *same* angular sector.

For the UTD model of the antenna, the following equations apply. For the singly diffracted field from Q_1, we have at the far-field observation point $P(r, \zeta)$

$$E_1^d(P) = 0, \quad \frac{3\pi}{2} - \gamma \leq \zeta \leq \frac{3\pi}{2} \tag{16-72}$$

and elsewhere

$$E_1^d(P) = E_\parallel^i(Q_1) D_\parallel(L, \phi, \phi') \frac{e^{-j\beta r_1}}{\sqrt{r_1}}$$
$$\approx E_\parallel^i(Q_1) D_\parallel(L, \phi, \phi') \frac{e^{-j\beta r}}{\sqrt{r}} e^{j\beta a \sin \zeta} \tag{16-73}$$

where (16-60) has been used in (16-73). Similarly, the diffracted field at $P(r, \zeta)$ from Q_2 may be written

$$E_2^d(P) = 0, \quad \frac{\pi}{2} \leq \zeta \leq \frac{\pi}{2} + \gamma \tag{16-74}$$

and elsewhere

$$E_2^d(P) = E_\parallel^i(Q_2) D_\parallel(L, \phi, \phi') \frac{e^{-j\beta r_2}}{\sqrt{r_2}}$$
$$\approx E_\parallel^i(Q_2) D_\parallel(L, \phi, \phi') \frac{e^{-j\beta r}}{\sqrt{r}} e^{-j\beta a \sin \zeta} \tag{16-75}$$

where (16-61) has been used in (16-75). In both (16-73) and (16-75),

$$E_\parallel^i(Q_1) = E_\parallel^i(Q_2) = \frac{e^{-j\beta \rho_o}}{\sqrt{\rho_o}} F_f(\theta_o) \tag{16-76}$$

The total field at an observation point $P(r, \zeta)$ may be written as either

$$E(P) = E^i(P) + E^A(P) \tag{16-77}$$

or

$$E(P) = E^i(P) + E_1^d(P) + E_2^d(P) \tag{16-78}$$

depending on the angle ζ as mentioned earlier.

Fig. 16-21 shows a calculated pattern for a cylindrical parabolic reflector having a 10λ aperture (i.e., $2a = 10\lambda$) and a focal length-to-diameter ratio of 0.5. The electric line source that models the feed has a pattern of $F_f(\theta_s) = \cos^2 \theta_s$ for $\theta_s \geq 90°$ and $F_f(\theta_s) = 0$ in the forward half-space where $\theta_s < 90°$. We note that the pattern has a small discontinuity at $\zeta = 90°$ (and $270°$) and this discontinuity can be removed by including double diffracted rays between Q_1 and Q_2 as was done for the horn in the previous section. We
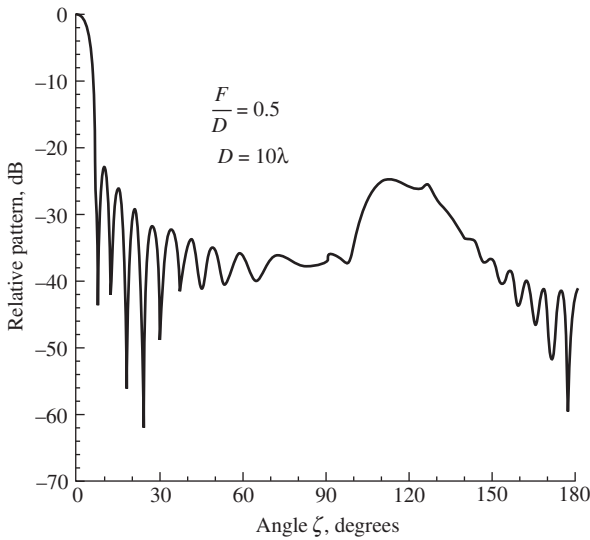
Figure 16-21 Calculated H-plane pattern of a cylindrical parabolic reflector with $D = 10\lambda$ having a focal-length-to-diameter ratio of 0.5.
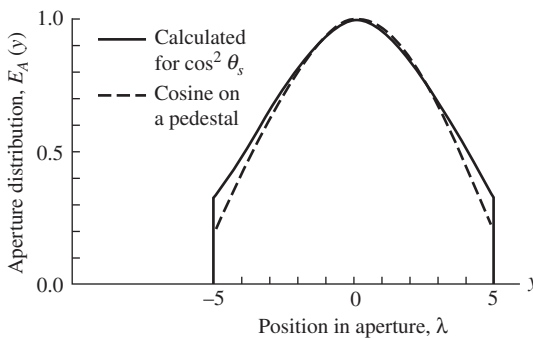
Figure 16-22 Aperture distribution for the parabola of Fig. 16-21 with a feed pattern of $\cos^2 \theta_s$.

also note that there is a small discontinuity at about $\zeta = 127°$ (and $233°$) that is a result of the shadowing of Q_2 (or Q_1 when $\zeta \approx 233°$). The relatively high level of the pattern in the vicinity of $\zeta = 120°$ is due to the spillover caused by the feed pattern.

It is interesting to note that for the chosen feed pattern of $\cos^2 \theta_s$, the aperture electric field distribution is almost that of a cosine on a pedestal with a -15-dB edge illumination as shown in Fig. 16-22. Referring to Table 5-2, we see that such a distribution should produce a pattern with a side-lobe level of -22 dB. Examination of the pattern in Fig. 16-21 shows that indeed the side-lobe level is -22 dB. Thus, the pattern in the forward half-space could be well represented by a line source, as discussed in Chap. 5, once the aperture field distribution is known.

In this section, we have examined the H-plane pattern of a cylindrical parabolic antenna (i.e., an electric line source was used to model the feed). We could also analyze the E-plane pattern when a magnetic line source is used to model the feed. This is left as an exercise for the student.

16.7   RADIATION BY A SLOT ON A FINITE GROUND PLANE

To illustrate further the application of the uniform theory of diffraction, consider the situation in Fig. 16-23 where a radiating slot is asymmetrically located along the x-axis of
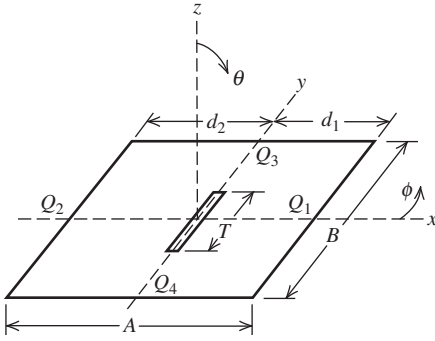
Figure 16-23 Geometry of a slot on a rectangular conducting plate.

the rectangular plate. We desire pattern information in both principal planes to determine the amount of ripple in the pattern caused by edge diffraction. In general, the edges denoted Q_1 and Q_2 will be illuminated unequally unless $d_1 = d_2$ and thus the pattern in the xz-plane will not be symmetrical about the z-axis.

The geometry of the problem to be investigated, as depicted in Fig. 16-23, is a narrow aperture (or slot) with length T on a finite ground plane of dimensions A and B. The narrow slot has an electric field polarized in the x-direction and has a cosine-distribution in the y-direction. The length of the slot is taken to be one-half wavelength at the operating frequency.

For radiation in the xz-plane above the ground plane, the problem is represented, to a first-degree approximation, by an infinitely long slot. According to UTD, there exist two edge-diffracted rays originating from edge points Q_1 and Q_2 due to the finiteness of the ground plane. Therefore, for a far-field observation point $P(r, \theta, \phi = 0)$ in the region of interest, the total field is the sum of the contributions from the direct ray and two edge-diffracted rays as shown in Fig. 16-24. Doubly diffracted rays exist but are small compared to the singly diffracted rays shown in Fig. 16-24 and are not included in the present analysis.

For radiation in the yz-plane above the ground plane, a sampling of $N + 1$ ideal sources forming a cosine distribution is performed. There exist no first-order edge-diffracted rays because the incident ray is zero in the yz-plane. A geometry of five samplings is shown in Fig. 16-25. The end sources are of zero amplitude since tangential E is zero at the ends of the slot.

First, let us consider the radiation pattern in the xz-plane. The direct ray from the narrow slot at an observation point $P(r, \theta, \phi = 0)$ is

$$\mathbf{E}^i(P) = \hat{\boldsymbol{\theta}} E_o \frac{e^{-j\beta r}}{\sqrt{r}} \tag{16-79}$$

The edge-diffracted ray from Q_1 at $P(r, \theta, \phi = 0)$ becomes

$$\mathbf{E}_1^d(P) = \hat{\boldsymbol{\theta}} \frac{1}{2} E_\perp^i(Q_1) D_\perp(L, \phi, \phi') \frac{e^{-j\beta r_1}}{\sqrt{r_1}} \tag{16-80}$$

with

$$\mathbf{E}_\perp^i(Q_1) = \hat{\mathbf{z}} E_o \frac{e^{-j\beta d_1}}{\sqrt{d_1}} = \hat{\mathbf{z}} E_\perp^i(Q_1) \tag{16-81}$$

The edge-diffracted ray from Q_2 at $P(r, \theta, \phi = 0)$ yields

$$\mathbf{E}_2^d(P) = \hat{\boldsymbol{\theta}} \frac{1}{2} E_\perp^i(Q_2) D_\perp(L, \phi, \phi') \frac{e^{-j\beta r_2}}{\sqrt{r_2}} \tag{16-82}$$
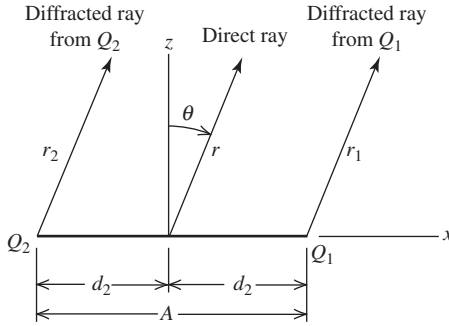
Figure 16-24 Direct and diffracted rays contributing to the xz-plane pattern.
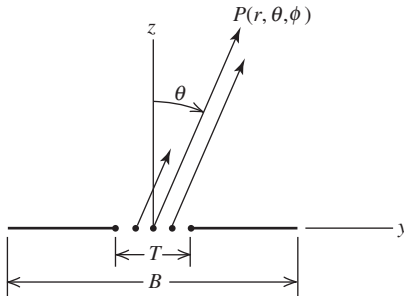
Figure 16-25 Direct rays from the weighted sources representing the slot contribution to the yz-plane pattern. See Fig. 16-23.

with

$$\mathbf{E}^i_{\perp}(Q_2) = \hat{\mathbf{z}} E_o \frac{e^{-j\beta d_2}}{\sqrt{d_2}} = \hat{\mathbf{z}} E^i_{\perp}(Q_2) \tag{16-83}$$

The total field at an observation point $P(r, \theta, \phi = 0)$ then becomes (in the symmetrical case)

$$\mathbf{E}(P) = \mathbf{E}^i(P) + \mathbf{E}^d_1(P) + \mathbf{E}^d_2(P) \tag{16-84}$$

The parameters r, r_1, d_1, r_2 and d_2, are shown in Fig. 16-24. The parameter E_o represents the magnitude of the electric field at the narrow slot in the xz-plane. $E^i_{\perp}(Q_m)$ is that component of the incident field which is perpendicular to both the edge and the incident ray. To first order, $D_{\parallel}$ is zero. However, there is a small amount of diffraction that does take place and this is called the slope diffraction (see Prob. 16.7-1). The addition of slope diffraction to the diffracted field ensures that not only is the total field continuous across a shadow boundary, but also the derivative of the total field is continuous.

For the slot problem of Fig. 16-23, the radiation in the yz-plane may be analyzed in the region above the ground plane, to a first degree approximation, by an array of dipole sources with a cosine-distributed amplitude across the array. Let the total number of dipoles in the array be $N + 1$; then the separation between dipoles is

$$s = \frac{T}{N} \tag{16-85}$$

The total field at an observation point $P(r, \theta, \phi = \pi/2)$ then becomes

$$\mathbf{E}(P) = \hat{\boldsymbol{\theta}} E_o \sin(90° - \theta) \sum_{n=-N/2}^{N/2} \cos\left(\frac{ns\pi}{T}\right) e^{jn\beta s \sin\theta} \tag{16-86}$$

In Fig. 16-25 is shown the geometry of the yz-plane with five dipoles ($N = 4$) in the array.
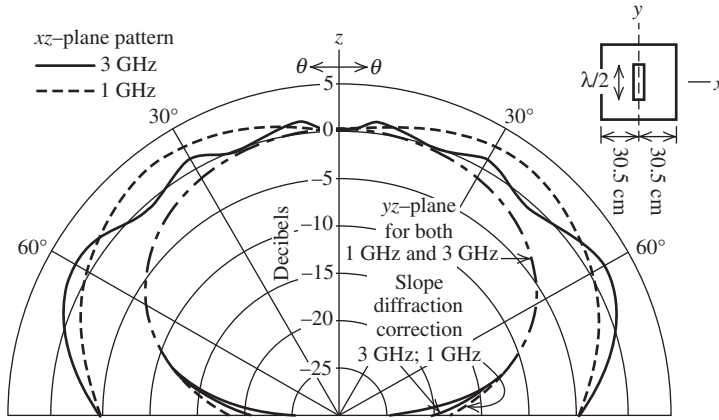
Figure 16-26 Far-field pattern in both principal planes at 1 and 3 GHz.

Fig. 16-26 shows the far-field pattern results in the xz- and yz-plane at both 1 and 3 GHz. The ground plane is 61×61 cm but the slot is taken to be one-half wavelength at each frequency and diffraction in the yz-plane has been assumed to be negligible, and under this assumption the pattern in the yz-plane is the same at each frequency as indicated in Fig. 16-26. However, due to diffraction, the two patterns in the xz-plane are different, the "ripple" in the patterns being the result of the diffracted energy. Since the slot is located symmetrically on the ground plane, we see that the pattern is symmetric about the z-axis.

For completeness, the slope diffraction contribution to the yz-plane pattern at 1 and 3 GHz is also shown in Fig. 16-26. Slope diffraction is proportional to the spatial derivative of the incident field with respect to the direction that is normal to both the incident ray and the edge at Q (see Prob. 16.7-1). In contrast, the edge diffraction we have considered thus far is proportional to the incident field at Q. In the problem considered in this section, the incident field at Q_3 and Q_4 in Fig. 16-23 is zero. However, the derivative of the incident field with respect to the normal (z in this case) is not zero at either Q_3 or Q_4. Although an experimental comparison for vertex diffraction is not shown here, such comparisons have been made with excellent results [12]. We know from experimental measurements that vertex diffraction is generally much weaker than wedge diffraction. Thus, the total far field is given to a good approximation by (16-84). In the problem considered here, vertex diffraction is weak in the xz- and yz-planes and somewhat stronger in the $\phi = 45°$ and $\phi = 135°$ planes. A UTD vertex diffraction coefficient is available in [13].

16.8 RADIATION BY A MONOPOLE ON A FINITE GROUND PLANE

As another application of the uniform theory of diffraction and also as an example of a problem with a caustic, consider the two situations depicted in Fig. 16-27. First, consider the situation in Fig. 16-27a of a $\lambda/4$ monopole on a square plate and suppose we wish to obtain the pattern in the xz-plane. For purposes of far-field calculation and conceptual simplicity, a suitable approximation to the $\lambda/4$ monopole is the ideal dipole of Chap. 2. We will consider the ideal dipole to be resting on the surface of the ground plane. Thus, following the development of the previous section for the slot, we have for the direct ray from the ideal dipole at $P(r, \theta, \phi = 0)$

$$\mathbf{E}^i(P) = \hat{\boldsymbol{\theta}} \, E_o \, \frac{e^{-j\beta r}}{r} \sin \theta \tag{16-87}$$

(*a*) Square ground plane.   (*b*) Circular ground plane.
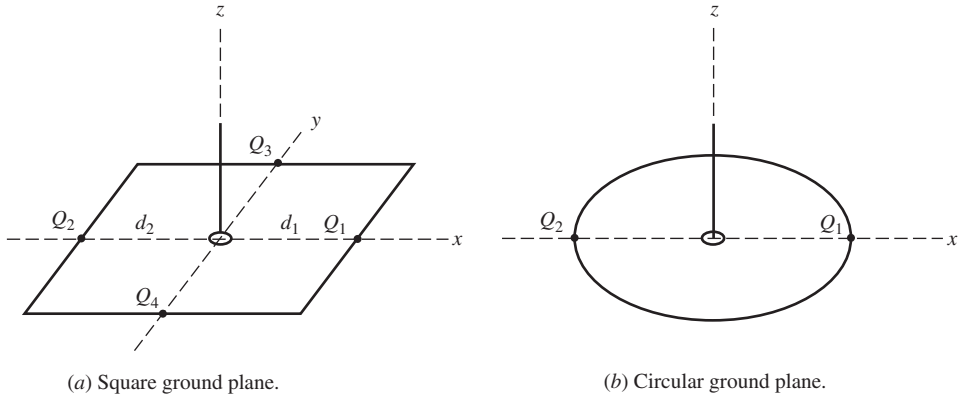
Figure 16-27 Monopole on a finite ground plane.

which now must obviously be considered a spherical wave. The edge-diffracted ray from Q_1 at $P(r, \theta, \phi = 0)$ appears to emanate from a single point and is therefore

$$\mathbf{E}_1^d(P) = \hat{\boldsymbol{\theta}} \frac{1}{2} E_\perp^i(Q_1) D_\perp(L, \phi, \phi') \sqrt{d_1}\, \frac{e^{-j\beta r_1}}{r_1} \qquad (16\text{-}88)$$

with

$$\mathbf{E}_\perp^i(Q_1) = \hat{\mathbf{z}} E_o \frac{e^{-j\beta d_1}}{d_1} = \hat{\mathbf{z}} E_\perp^i(Q_1) \qquad (16\text{-}89)$$

Similarly, the edge-diffracted ray from Q_2 at $P(r, \theta, \phi = 0)$ is

$$\mathbf{E}_2^d(P) = \hat{\boldsymbol{\theta}} \frac{1}{2} E_\perp^i(Q_2) D_\perp(L, \phi, \phi') \sqrt{d_2}\, \frac{e^{-j\beta r_2}}{r_2} \qquad (16\text{-}90)$$

with $E_\perp^i(Q_2)$ given by (16-89) since the source is located at the center of the ground plane. Diffraction from the sides containing Q_3 and Q_4 does not contribute to the far field in the *xz*-plane since the monopole is positioned at the center of the ground plane and the diffracted fields from these two sides will cancel one another. As in the case of the slot of the previous section, we neglect diffraction from the four corners or vertices of the ground plane. The corresponding pattern for a 6λ square ground plane is given in Fig. 16-28 (dashed curve).

   If we now consider the geometry of Fig. 16-27*b*, we note that in the *xz*-plane, the diffracted radiation will also appear (due to Fermat's principle) to come from two points that are called stationary points. We note also that the *z*-axis is a caustic in this problem because all rays from the circular edge of the ground plane intersect along the *z*-axis. Therefore, although we can expect to use the two stationary points to calculate the diffracted field contribution to the pattern in regions not near the caustic [14], we can likewise expect the "two-point approximation" to be increasingly in error as the observation point *P* moves nearer the caustic. Fig. 16-28 shows that indeed this is the case since the measured and two-point calculated patterns diverge as both $\theta \to 0$ and $\theta \to \pi$, which is also a caustic. The apparent difficulty in the vicinity of the caustic can be overcome, as suggested in Fig. 16-28, by the use of a fictitious equivalent edge current. As will be seen in the next section, the so-called equivalent current is not a physical current at all, but rather a mathematical artifice for predicting the correct diffracted field at or near a caustic.
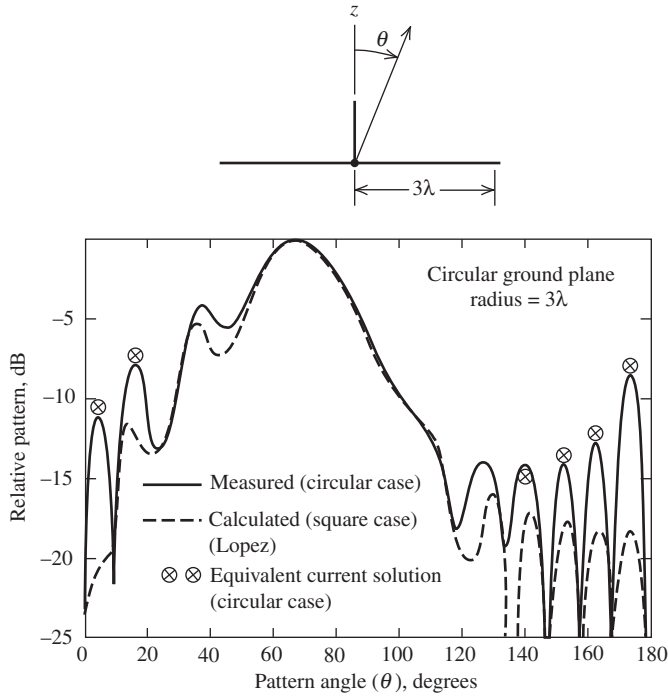
Figure 16-28 Radiation pattern of an ideal monopole on a circular ground plane 6λ in diameter. The calculated curve by Lopez [14] is for a $6\lambda \times 6\lambda$ square ground plane.

16.9 EQUIVALENT CURRENT CONCEPTS

In the previous section, we saw that, in the treatment of the circular ground plane, we could obtain the diffracted field using ordinary wedge diffraction theory if the point of observation was not near a caustic. In essence, we treated the problem as a two-dimensional one with the diffraction taken as that from an infinite two-dimensional wedge, whereas in fact we had a finite edge that was not straight, but curved.

To properly treat the diffraction by a curved edge or finite wedge (i.e., finite length of the edge), it is necessary to consider the concept of equivalent currents [15]. As we shall see, the strengths (i.e., amplitude and phase) of these so-called equivalent currents will be determined by the canonical problem of wedge diffraction.

Consider the wedge of Fig. 16-15 to be of finite extent, $-\ell/2 \leq z(Q) \leq \ell/2$. To start, we assume the currents are the same as those on an infinite wedge. Let us determine the current flowing on the edge of the infinite wedge that would produce the scattered field predicted by wedge diffraction analysis. Thus, we specify an infinite line source whose current is determined by the diffraction coefficient. If the z-directed line source is an electric current, it can be shown that the solution to the scalar wave equation is [16]

$$\psi = \frac{I^e}{4j} H_o^{(2)}(\beta\rho) \tag{16-91}$$

and therefore that the z-component of the electric field is

$$E_z = \frac{-\beta^2 I^e}{4\omega\varepsilon} H_o^{(2)}(\beta\rho) \tag{16-92}$$

where I^e denotes an electric current. If the argument of the Hankel function $H_o^{(2)}(\beta\rho)$ is large, then using the asymptotic representation of the Hankel function, we obtain

$$E_z = \eta\beta I^e \frac{e^{j(\pi/4)}}{2\sqrt{2\pi\beta\rho}} e^{-j\beta\rho} \tag{16-93}$$

We note that (16-93) represents an outward traveling wave in the cylindrical coordinate system with the proper $\rho^{-1/2}$ dependence for a two-dimensional problem. If instead the line source is a magnetic current I^m, then we have

$$H_z = -\frac{\beta}{\eta} I^m \frac{e^{j(\pi/4)}}{2\sqrt{2\pi\beta\rho}} e^{-j\beta\rho} \tag{16-94}$$

Since we are considering a two-dimensional problem, we can also apply wedge diffraction theory to obtain the diffracted field from the edge for the two orthogonal polarizations. Thus,

$$E_z = D_{\parallel}(L, \phi, \phi')E_z^i \frac{e^{-j\beta\rho}}{\sqrt{\rho}} \tag{16-95}$$

and

$$H_z = D_{\perp}(L, \phi, \phi')H_z^i \frac{e^{-j\beta\rho}}{\sqrt{\rho}} \tag{16-96}$$

where $D_{\parallel}$ and $D_{\perp}$ are given in Sec. 16.4. Usually, however, we find that the use of equivalent currents involves the calculation of diffracted fields in regions away from an incident field or a reflected field shadow boundary or their associated transition regions. Thus, the asymptotic form in (16-59) for arbitrary incidence angle γ_o is usually sufficient.

From (16-93) with (16-95) and also (16-94) with (16-96), we can solve for the electric and magnetic currents of an infinite line source that will produce the same far fields predicted using the diffraction coefficients. Thus,[5]

$$I^e = \frac{-2j}{\eta\beta} E_z^i D_{\parallel}\left(\phi, \phi'; \frac{\pi}{2}\right)\sqrt{2\pi\beta} \; e^{j(\pi/4)} \tag{16-97}$$

and

$$I^m = \frac{2j\eta}{\beta} H_z^i D_{\perp}\left(\phi, \phi'; \frac{\pi}{2}\right)\sqrt{2\pi\beta} \; e^{j(\pi/4)} \tag{16-98}$$

We note that (16-97) and (16-98) give the equivalent currents I^e and I^m, but they are numerically different for each value of ϕ and ϕ'. The fact that these currents are different for different observation points (i.e., values of ϕ) serves to emphasize the fact that these equivalent currents are not true currents, but fictitious currents that simply aid us in calculating diffracted fields.

Considering Fig. 16-15 with the ray incident normally on the edge ($\gamma_o = \pi/2$), we have, respectively, for the far-zone diffracted fields

---

[5] Note that we denote the diffraction coefficient to be a function of L, ϕ, and ϕ' to imply the Fresnel integral form of the uniform theory in (16-52) and use ϕ, ϕ', and γ_o when the asymptotic form in (16-59) is intended.

$$E_\theta^e = \frac{j\omega\mu \sin\theta}{4\pi r} e^{-j\beta r} \int_{-\ell/2}^{\ell/2} I_z^e(z')e^{j\beta z' \cos\theta} dz' \tag{16-99}$$

and

$$H_\theta^m = \frac{j\omega\mu \sin\theta}{4\pi r} e^{-j\beta r} \int_{-\ell/2}^{\ell/2} I_z^m(z')e^{j\beta z' \cos\theta} dz' \tag{16-100}$$

As in Chap. 5, we see that since the currents are constant with respect to z', (16-99) and (16-100) reduce to results in the general form of $\sin(x)/x$ with respect to the θ-coordinate.

For the case of nonnormal incidence (i.e., $\gamma_o \neq \pi/2$), we can proceed in the same manner and show that

$$I^e = \frac{-2j}{\eta\beta} E_z^i D_\parallel(\phi, \phi'; \gamma_o)\sqrt{2\pi\beta} \; e^{j(\pi/4)}e^{j\beta\ell \cos\gamma_o} \tag{16-101}$$

and

$$I^m = \frac{2j\eta}{\beta} H_z^i D_\perp(\phi, \phi'; \gamma_o)\sqrt{2\pi\beta} \; e^{j(\pi/4)}e^{j\beta\ell \cos\gamma_o} \tag{16-102}$$

which includes the phase term to account for the traveling-wave-type current due to the oblique angle of incidence. In obtaining (16-101) and (16-102), we have neglected the effects of the terminations at $z = \pm\ell/2$. If the effect of the termination could be specified, an alternative equivalent current could be composed of the currents given above plus a reflected current due to the termination. These reflection effects would be expected to be of most concern in the backscatter direction, rather than in the direction of the bistatic scattered field. Even so, as the edge becomes long in terms of the wavelength, termination effects diminish. In addition, usually the above currents find their application in the angular region near the plane normal to the edge, further diminishing any possible termination effects.

When we obtain equivalent currents, we invoke the postulate of diffraction theory that diffraction is a local phenomena. For curved edges, we stretch this postulate even further than for the straight edge. That is, we assume that each point on a curved edge acts as an incremental section of an infinite straight edge and thereby determine the equivalent current. Thus, for example, the equivalent current that would enable us to calculate the diffracted field at the caustic of the problem in Fig. 16-27b would be [12]

$$I^m = -(\hat{\boldsymbol{\phi}} \times \hat{\mathbf{s}}') \cdot \mathbf{E}^i D_\perp\left(\phi, \phi'; \frac{\pi}{2}\right)\sqrt{\frac{8\pi}{\beta}} \; e^{-j(\pi/4)} \tag{16-103}$$

where we have used the result of (16-98) and the fact that $(\hat{\boldsymbol{\phi}} \times \hat{\mathbf{s}}')$ gives us the unit vector perpendicular to the ray from the sources to the edge. The use of (16-103) gives the calculated results in Fig. 16-28, which agree with experimental measurements in the caustic region.

If, on the other hand, the source in Fig. 16-27b were a magnetic dipole, then the required equivalent current would be [14]

$$I^e = -\frac{\hat{\boldsymbol{\phi}} \cdot \mathbf{E}^i}{\eta}D_\parallel\left(\phi, \phi'; \frac{\pi}{2}\right)\sqrt{\frac{8\pi}{\beta}} \; e^{-j(\pi/4)} \tag{16-104}$$

For an arbitrary polarization of the incident wave, both electric and magnetic currents are necessary to obtain the total diffracted field. Such a situation would occur, for example, in the calculation of the fields at or near the rear axis (caustic region) of a

circular parabolic reflector antenna. At the rim of the parabolic dish, the polarization of the field incident from the feed is generally neither perpendicular nor parallel to the edge. Thus, both electric and magnetic equivalent currents at the rim would be required to obtain the total diffracted field in the rear axial region.

16.10   A MULTIPLE DIFFRACTION FORMULATION

In the previous two sections, we considered radiating elements on infinitely thin ground planes (i.e., $n = 2$). If, instead, the ground plane were "thick" such that one side could be represented by two $90°$ wedges as shown in Fig. 16-29, then it would have been necessary to consider the multiple diffractions that occur between the two closely spaced edges. In such a situation, some of the energy diffracted by one edge is, in turn, diffracted by the other, giving rise to second-order diffraction or double diffracted rays. Clearly, some of these double diffracted rays give rise to still higher-order multiple diffractions.

If, to reasonably approximate the total diffracted energy, it is necessary to include doubly diffracted rays, then it is usually simplest to include them in the same manner used to account for the first-order diffraction in the previous two sections. On the other hand, if it is necessary to account for diffractions higher than second-order, it is advantageous to use a procedure known as the *method of self-consistency* [17].

Briefly, the method of self-consistency incorporates all the diffracted rays (i.e., single as well as all higher-order multiple ones) into a *total* (or net) diffracted field from each diffracting edge. Thus, each of these total edge diffracted fields is excited by a surface ray. Between the two diffracting edges there are, therefore, two surface rays traveling in opposite directions. The amplitudes and phases of these two surface rays are treated as unknowns. To solve for the two unknowns, two equations are generated by applying single diffraction conditions at each of the two diffracting edges.

To illustrate, consider Fig. 16-29. The coefficients C_1 and C_3 are the unknown amplitudes (i.e., magnitude and phase) of the two surface rays on the surface *ab*. The coefficients C_2 and C_4 are known once C_1 and C_3 are known. Thus, we may write the following equations. At edge *a*,

$$C_1 = C_3 R_{ba} + V_1 \qquad (16\text{-}105)$$

and at edge *b*,

$$C_3 = C_1 R_{ab} + V_3 \qquad (16\text{-}106)$$

and knowing C_1 and C_3, we have

$$C_2 = T_{ba} C_3 + V_2 \qquad (16\text{-}107)$$

$$C_4 = T_{ab} C_1 + V_4 \qquad (16\text{-}108)$$
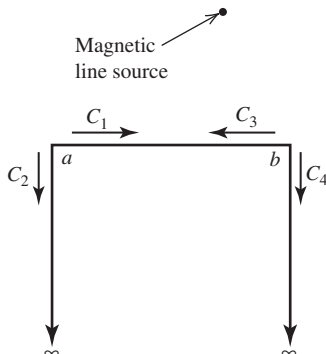
Figure 16-29 Magnetic line source exciting surface rays on a half-plane of finite thickness.

where R and T are reflection and transmission coefficients, respectively, and V is the direct source contribution to the corresponding surface ray.

Eqs. (16-105) and (16-106) may be written in matrix form as

$$\begin{bmatrix} 1 & -R_{\text{ba}} \\ -R_{\text{ab}} & 1 \end{bmatrix} \begin{bmatrix} C_1 \\ C_3 \end{bmatrix} = \begin{bmatrix} V_1 \\ V_3 \end{bmatrix} \tag{16-109}$$

or compactly as

$$[Z][C] = [V] \tag{16-110}$$

where $[Z]$ is taken to be a coupling matrix and $[V]$ is the excitation matrix. The elements of the coupling matrix specify the interactions between the two surface rays. In general, two surface rays can couple only if they travel on the same or adjacent faces of a polygon as shown in Fig. 16-30. This, in general, leads to a sparse $[Z]$ matrix.
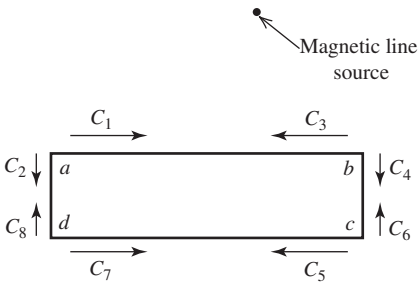
Figure 16-30 Magnetic line source exciting surface rays on an infinite four-sided polygon cylinder.

For the situation in Fig. 16-29, the reflection and transmission coefficients are

$$R_{ab} = \frac{e^{-j\beta\rho_{\text{ab}}}}{\sqrt{\rho_{\text{ab}}}} \frac{1}{2} D_\perp(L, \phi, \phi') \tag{16-111}$$

where $\phi^+ = \phi^- = 0$ and $\gamma_o = 90°$, and

$$T_{ab} = \frac{e^{-j\beta\rho_{\text{ab}}}}{\sqrt{\rho_{\text{ab}}}} \frac{1}{2} D_\perp(L, \phi, \phi') \tag{16-112}$$

where $\phi^+ = 2\pi - \pi/2$, $\phi^- = 0$, and $\gamma_o = 90°$. For T_{ab}, ϕ^+ is 2π less the interior wedge angle of $\pi/2$. In both cases, the distance parameter L used is that for cylindrical waves. For the special situation depicted in Fig. 16-29, $R_{ab} = R_{ba}$ and $T_{ab} = T_{ba}$. This is not true in general. For example, it would not be true for the situation depicted in Fig. 16-31.

For the two excitation matrix elements, we have

$$V_1 = \frac{e^{-j\beta\rho_{\text{sa}}}}{\sqrt{\rho_{\text{sa}}}} D_\perp(L, \phi, \phi') \tag{16-113}$$

(a) Curved surface.   (b) Polygon approximation.

Figure 16-31 Polygon approximation of a curved surface cylinder.

$$V_3 = \frac{e^{-j\beta\rho_{sb}}}{\sqrt{\rho_{sb}}} D_\perp(L, \phi, \phi') \qquad (16\text{-}114)$$

where ρ_{sa} is the distance from the line source to edge a and ρ_{sb} is the distance from the line source to edge b. If the line source did not directly illuminate, say, edge b, then V_3 would be zero. Here, we have considered only the TE case. A consideration of the TM case requires a knowledge of slope diffraction (see Sec. 16.7).

As stated earlier, it may be sufficient in many problems to only take into account second-order diffraction, thereby neglecting all higher-order multiply-diffracted rays. An example of a situation where the method of self-consistency greatly simplifies the amount of work required for solution is suggested by Fig. 16-31. It is possible and practical to approximate the curved surface of Fig. 16-31a with a polygon such as that in Fig. 16-31b. For an accurate approximation to the curved surfaces, the edges in Fig. 16-31b may be sufficiently close together that higher-order multiple diffractions should be taken into account. The easiest way of doing that is via the method of self-consistency. However, the self-consistent field method only works well provided an edge is not in the transition region of a diffraction from another edge. This is a possibility if adjacent edges of the polygon are closely aligned and this limits the degree to which the curved surface may be approximated.

16.11 DIFFRACTION BY CURVED SURFACES

In previous sections, we saw how a perfectly conducting wedge diffracts energy into the shadow region. Curved surfaces can also diffract energy. That is, when an incident ray strikes a smooth, convex-curved perfectly conducting surface at grazing incidence, a part of its energy is diffracted into the shadow region as illustrated by Fig. 16-32.

In Fig. 16-32, the incident plane wave undergoes diffraction at the shadow boundary at point Q_1 that is a point of tangency for the incident ray. At this point, a portion of the energy

Figure 16-32 Diffraction by a smooth convex surface.

is trapped, resulting in a wave that propagates on the surface of the scatterer, shedding energy by radiation in directions tangent to the surface (e.g., point Q_2) as it progresses around the curved surface.

This wave that propagates along the surface in the shadow region is known as a *creeping wave*. The creeping wave can be described by an attachment (diffraction) coefficient at the point of capture, a launching (diffraction) coefficient at the point of radiation, an attenuation factor to account for the rate of radiation, and a description of the path on the scatterer transversed by the creeping wave in order to account for phase delay and total attenuation. Thus, the creeping wave field $E^{cw}(s'')$ along the path s'', in the case of a two-dimensional problem, can be written as

$$E^{cw}(s'') = E^i(Q_1)D^s(Q_1)G(s'')e^{-\int \gamma(s'')ds''} \tag{16-115}$$

where
$E^{cw}(s'')$ = creeping wave field along s''
$D^s(Q_1)$ = the surface diffraction coefficient (attachment coefficient) at point A
$\gamma(s'') = \alpha(s'') + j\beta(s'')$ = creeping wave propagation factor
s'' = arc length along the creeping wave path
$G(s'')$ = the ray divergence factor determined by the geometry of the ray

As already stated, the surface ray sheds energy tangentially as it propagates along a geodesic on the curved surface, with the result that energy is continuously lost, resulting in attenuation. As in geometrical optics, we assume that energy in the flux tube between adjacent rays is conserved, which gives the two-dimensional geometrical optics spreading factor as

$$G(s'') = \sqrt{\frac{d\eta(Q_1)}{d\eta(s'')}} \tag{16-116}$$

where $d\eta(s'')$ is the transverse dimension of the surface ray tube as shown in Fig. 16-32. For a plane wave at normal incidence, $G(s'')$ is unity.

Keller and Levy [18] have given the first-order terms in the expressions for the curved surface diffraction coefficients and attenuation constants. Kouyoumjian [12, 19] gives higher-order terms and uses the notation of a soft surface for the case where E is tangential to the surface and hard surface for the case where E is normal to the surface. To be consistent with our earlier notation, we will use the perpendicular and parallel notation of earlier parts of this chapter.

At point Q_2, there will be a second surface diffraction coefficient $D^s(Q_2)$ that will account for the tangential detachment (launch) of a ray from the surface toward the observation point at a distance ℓ from point Q_2. Thus, we can write for the creeping wave field shed at Q_2 and observed at the observation point P

$$E^{cw}(P) = E^i(Q_1)D^s(Q_1)D^s(Q_2)\sqrt{\frac{d\eta(Q_1)}{d\eta(Q_2)}} \, e^{-\int_{Q_1}^{Q_2} \alpha(s'')ds''} e^{-j\beta(Q_2-Q_1)} \frac{e^{-j\beta\ell}}{\sqrt{\ell}} \tag{16-117}$$

For a circular cylinder, the diffraction coefficients and attenuation constants for the asymptotic approximation to the exact solution are given by the quantities in Table 16-1, where ρ_g is the radius of curvature along a geodesic. For a normally incident plane wave on a circular cylinder, $\rho_g = a \cdot \text{Ai}(-x)$ is the Airy function [12, 20]. The creeping wave surface field is more accurately represented by a series of modes, but only the first such mode is given in Table 16-1 since the higher-order modes are not numerically significant for the circular cylinder treated here.

Table 16-1 Diffraction and Attenuation Coefficients for a Convex Cylindrical Surface

Case	$(D^s)^2$	α	Airy Function and Its Zeros
$E_\parallel$	$\dfrac{\rho_g^{1/3}e^{-j\pi/12}}{\pi^{1/2}\,2^{5/6}\,\beta^{1/6}(\mathrm{Ai}'(-q))^2}$	$\dfrac{q}{\rho_g}e^{j\pi/6}\left(\dfrac{\beta\rho_g}{2}\right)^{1/3}$	$q = 2.33811$ $\mathrm{Ai}'(-q) = 0.7012$
$E_\perp$	$\dfrac{\rho_g^{1/3}\,e^{-j\pi/12}}{\pi^{1/2}\,2^{5/6}\,\beta^{1/6}(\overline{q})(\mathrm{Ai}(-\overline{q}))^2}$	$\dfrac{\overline{q}}{\rho_g}e^{j\pi/6}\left(\dfrac{\beta\rho_g}{2}\right)^{1/3}$	$\overline{q} = 1.01879$ $\mathrm{Ai}(-\overline{q}) = 0.5356$

EXAMPLE 16-3   *Creeping Wave on a Circular Cylinder*

Consider the two-dimensional problem of calculating the radar echo width of a right circular cylinder normal to the axis of the cylinder. The echo width σ_w is the two-dimensional counterpart to the three-dimensional echo area or radar cross section. σ_w is defined to be

$$\sigma_w = \lim_{\ell \to \infty} 2\pi\ell \frac{|\overline{E}^s|^2}{|\overline{E}^i|^2} \tag{16-118}$$

where ℓ is the range to the target.

   Let the cylinder be of radius a and centered about the z-axis as in Fig. 16-33. Assume the incident wave is given by

$$\begin{pmatrix} E_\parallel^i \\ E_\perp^i \end{pmatrix} = e^{+j\beta\ell} \tag{16-119}$$

The two-dimensional geometrical optics field reflected in the backscatter direction may be found from (16-21) and (16-24) to be

$$\begin{pmatrix} E_\parallel^r \\ E_\perp^r \end{pmatrix} = \begin{pmatrix} E_\parallel^i \\ E_\perp^i \end{pmatrix} [R]\sqrt{\frac{a}{2\ell}} \tag{16-120}$$

where $\ell \gg a$ and where R is given by (16-25). Applying the echo width definition above gives

$$\sigma_w = \lim_{\ell \to \infty} 2\pi\ell \frac{|\overline{E}^s|^2}{|\overline{E}^i|^2} = \lim_{\ell \to \infty} 2\pi\ell\left(\frac{a}{2\ell}\right) = \pi a \tag{16-121}$$

   Next, the effects of creeping waves around the cylinder are included. At the attachment point, write

$$E^{cw}(Q_1) = E^i(Q_1)D^s(Q_1) \tag{16-122}$$

and at the detachment point, write

$$E^{cw}(Q_2) = E^i(Q_1)D^s(Q_1)e^{-j\beta\pi a}e^{-\int_0^{\pi a}\alpha\,ds''} = E^i(Q_1)D^s(Q_1)\,e^{-j\beta\pi a}e^{-\alpha\,\pi a} \tag{16-123}$$

$G(s'')$ is unity since there is no transverse spreading of the rays on the surface of the cylinder.
   When $E^{cw}(Q_2)$ is multiplied by the detachment coefficient $D(Q_2)$, we have the radiated creeping wave field. By a reciprocity argument, we can see that for the circular cylinder, the launching and attachment coefficients are the same. Thus in the following expression for the radiated field in the backscatter direction, the surface diffraction coefficient is squared:
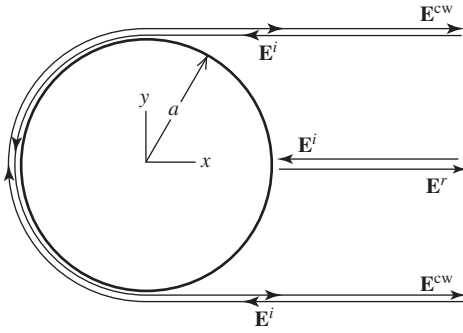
Figure 16-33 Backscatter from a circular cylinder.

$$\begin{pmatrix} E_\parallel^{cw} \\ E_\perp^{cw} \end{pmatrix} = \begin{pmatrix} E_\parallel^{i} \\ E_\perp^{i} \end{pmatrix} \cdot \begin{pmatrix} D_\parallel^{s} \\ D_\perp^{s} \end{pmatrix}^2 e^{-j\beta\pi a} e^{-\pi a \alpha_{\parallel,\perp}} \sqrt{\frac{a}{2\ell}} \tag{16-124}$$

We must consider the fact that there are attachment points at both the top and bottom of the cylinder requiring us to double the creeping wave field strength. Thus, the total back-scattered field is

$$\begin{pmatrix} E_\parallel^{s} \\ E_\perp^{s} \end{pmatrix} = \begin{pmatrix} E_\parallel^{r} \\ E_\perp^{r} \end{pmatrix} + 2 \begin{pmatrix} E_\parallel^{cw} \\ E_\perp^{cw} \end{pmatrix} \tag{16-125}$$

which for the cylinder is

$$\begin{pmatrix} E_\parallel^{s} \\ E_\perp^{s} \end{pmatrix} = \begin{pmatrix} E_\parallel^{i} \\ E_\perp^{i} \end{pmatrix} \cdot \begin{pmatrix} -\sqrt{\dfrac{a}{2}}\, e^{j2\beta a} + 2(D_\parallel^{s})^2 e^{-\pi a \alpha_\parallel}\, e^{-j\beta\pi a} \\[2mm] +\sqrt{\dfrac{a}{2}}\, e^{j2\beta a} + 2(D_\perp^{s})^2 e^{-\pi a \alpha_\perp}\, e^{-j\beta\pi a} \end{pmatrix} \frac{e^{-j\beta\ell}}{\sqrt{\ell}} \tag{16-126}$$

The echo width then becomes

$$\sigma_w = \pi a \left| \mp e^{j2\beta a} + 2(D_{\parallel,\perp}^{s})^2 \left(\frac{2}{a}\right)^{1/2} e^{-\pi a \alpha_{\parallel,\perp}}\, e^{-j\beta\pi a} \right|^2 \xrightarrow[\beta a \to \infty]{} \pi a^2 \tag{16-127}$$

Expressions for D and α can be obtained from Table 16-1 with $\rho_g = a$.

For the problem considered in the above example, the exact result for the echo width for both principal polarizations appears in Fig. 16-34a. The echo width is normalized with respect to πa just as the sphere echo area was normalized with respect to its specular contribution. Note that the creeping wave contribution to the echo width for the cylinder is quite visible for the perpendicular polarization and does not appear for the parallel polarization. This is due to the tangential electric field boundary condition that tends to short out the creeping wave contribution in the parallel case but not the perpendicular case.

The above GTD-based expression in (16-127) for the echo width of a circular cylinder cannot be expected to reproduce the curves in Fig. 16-34a, which are based on the exact eigenfunction solution when the diameter is on the order of the wavelength or less. This, of course, is because GTD is an asymptotic theory valid most when the wavelength is small compared to the scatterer. In Fig. 16-34a, the wavelength is not small compared to the scatterer. Results using (16-127) are shown in Fig. 16-34b. The differences between the results from (16-127) and the exact solution are apparent. As a point of fact, there is no theoretical reason to expect them to agree. Interestingly, most of the error in attempting to reproduce Fig. 16-34a with (16-127) arises from the geometrical optics term and not the creeping wave term.

(a) Exact eigenfunction solution.

(b) GTD solution using (16-127).

(c) GTD solution with GO correction.
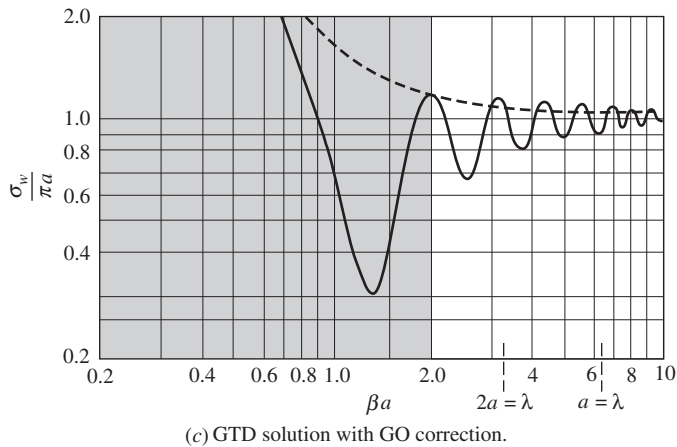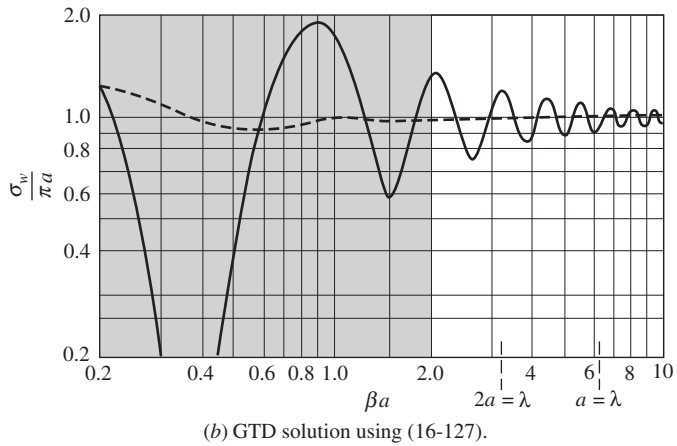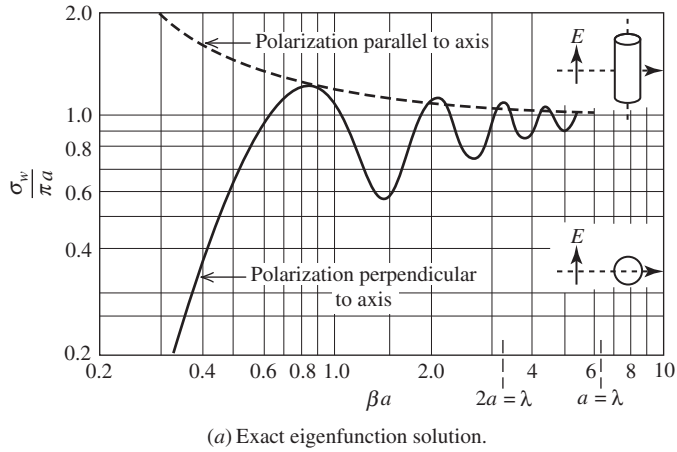
Figure 16-34 Echo width of an infinitely long circular cylinder calculated by three methods.

To improve the accuracy of the geometrical optics contribution, it is necessary to include correction terms [20] (see Prob. 16.11-1). The application of correction terms to the geometrical optics contribution is shown in Fig. 16-34c, and the improvement over Fig. 16-34b is substantial. In fact, the agreement between Fig. 16-34a and Fig. 16-34c when $\beta a > 2$ (i.e., $a > \lambda/3$) is surprisingly good for a high-frequency method in a lower
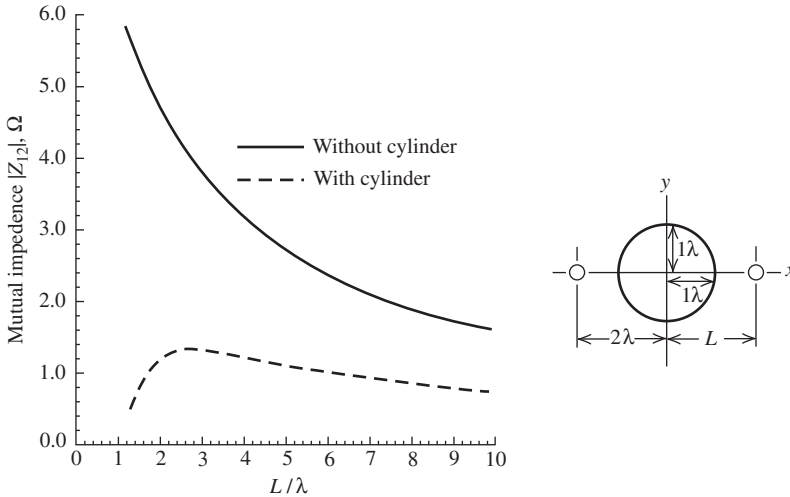
Figure 16-35 Mutual impedance Z_{12} between two $\lambda/2$ dipoles in the deep shadow region of each other compared to Z_{12} without cylinder present.

portion of the intermediate frequency region. In this region, one would normally employ the method of moments or MoM to produce results such as those in Fig. 16-34a.

The concept of a creeping wave is valuable in that it helps one visualize the physical process involved in diffraction by curved surfaces. For example, the RCS of a sphere as a function of the sphere radius is presented in Fig. 16-7. We can interpret the oscillatory feature of the curve as being caused by two creeping waves traveling around the sphere in opposite directions. A similar explanation applies to the cylinder in Fig. 16-34. Depending on the electrical size of the sphere or cylinder, these two creeping waves tend to either constructively or destructively interfere with each other, causing the RCS to oscillate about the value contributed by the specular scattering. As the sphere or cylinder becomes larger, the amount of oscillation decreases, which may be attributed to the decreasing amplitudes of the two creeping waves due to the product α times the total path length s''. In the backscatter case, as the radius becomes large, α becomes small, but the product $\alpha s''$ becomes large, causing the creeping wave to become small compared to the specular contribution as evidenced by Fig. 16-7 and (16-34).

In the case of backscatter for the cylinder or sphere, the creeping wave travels a 180° geodesic path. A 180° path is not the only path of possible interest. Consider Fig. 16-35 that shows the mutual coupling between two half-wave dipoles on opposite sides of a circular cylinder [24]. The coupling with the cylinder present can be calculated by the creeping wave formulation in this section.

16.12   APPLICATION OF UTD TO WIRELESS MOBILE PROPAGATION

Diffraction theory presented in this chapter has application in the propagation of wireless mobile signals. This section will discuss propagation mainly in urban macrocells (see Fig. 4-7). Macrocells are typically a few tens of kilometers across as discussed in Sec. 4.5.1. Macrocells are formed by base stations having antenna heights that are significantly higher than the surrounding clutter introduced by buildings and tall vegetation. On the other hand, mobile receivers are likely to be shadowed by these obstructions in the nearby environment, particularly buildings, as shown in Fig. 16-36.

There are two broad classes of macrocell propagation models: empirical and physical. Early propagation models were based upon empirical relationships derived from large amounts of measured data, but these relationships provide little understanding of the
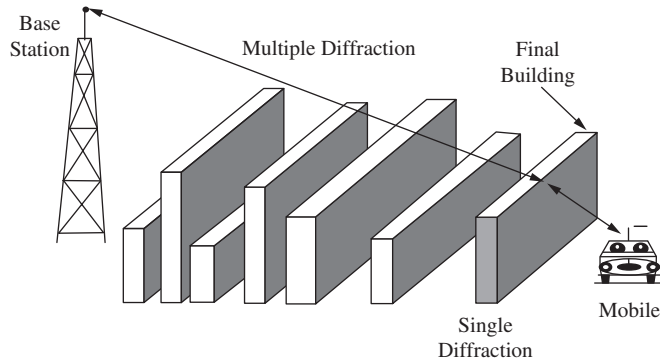
Figure 16-36 Multiple diffractions over flat building rooftops. (Reproduced by permission from [H.9.1: Fujimoto], Norwood, MA: Artech House, Inc. © 2008 by Artech House, Inc.)

physical processes involved in producing various signal levels at the mobile receiver. Subsequently, effort was made to construct physical propagation models based upon an understanding of the interaction of the signal environment with the signal transmitted by the base station. These physical models treat the path loss in a given macrocell as distance dependent on the assumption that the environment around the base station is fairly uniform. The effect of obstructions in the signal environment is usually presented as excess path loss over what the unobstructed free space path loss alone would be. Diffraction (UTD) plays a major role in determining the excess path loss in urban macrocells.

Wireless signals from a macrocell mobile system, operated in a reasonably flat terrain city environment, will experience propagation over and around multiple buildings. Diffraction over multiple buildings is stronger than diffraction around buildings and models using diffraction theory are of the most interest. In a scenario with propagation over multiple rooftops, the diffraction is largely unaffected by the shape of the rooftop except for the "last" rooftop as suggested in Fig. 16-36. Consequently, simple knife edge (i.e., thin half-plane as in Fig. 16-10) diffraction was used in early models for all but the last rooftop. Later propagation models used the UTD wedge diffraction for all rooftops. An early example of the utilization of UTD was by Luebbers [22] in a model for a single wedge wherein the part of the UTD diffraction coefficient associated with the reflection boundary is modified to account for finite conductivity. This was done by multiplying the diffraction transition functions $F[\beta La^+(\varphi+\varphi')]$ and $F[\beta La^-(\varphi+\varphi')]$ in (16-52) by the appropriate reflection coefficient as in (6-62) and (6-63). This single wedge model was later used by others [23, 24] in multiple wedge models.

The problem with the sequential application of first-order diffraction (see Sec. 16.4) to multiple edge (i.e., multiple rooftops as in Fig. 16-36) is that a rooftop will likely be in the transition region of the previously encountered rooftop rendering the field incident on all but the first rooftop non-ray optical (i.e., incident fields from multiple sources). This will introduce error into the propagation model [23, 25]. The ray-optical nature of the field can be restored either by including higher-order diffraction terms such as multiple diffractions (see Sec. 16.10) and [25 or 26] or by including slope diffraction [23, 24]. Results using multiple diffraction are shown in Figs. 16-37 and 16-38, where the calculated results are compared against measurements. The field is normalized with respect to the free-space value and therefore represents the excess path loss due to the rooftops. Values above zero dB at a receiver height around 65 wavelengths are due to the constructive interference between the direct and diffracted rays. Both the calculated results and the measurements were done for a scale model at 25 GHz in a controlled environment. The agreement between the theoretical and measured scale model results is quite good, especially in the parallel polarization case. Diffraction by straight edges is not frequency dependent (see Sec. 16.16).

Slope diffraction was used in Fig. 16-27 and also in Prob. 16.7-1 for a perfectly conducting wedge. Here we are interested in an absorbing (i.e., imperfectly conducting)
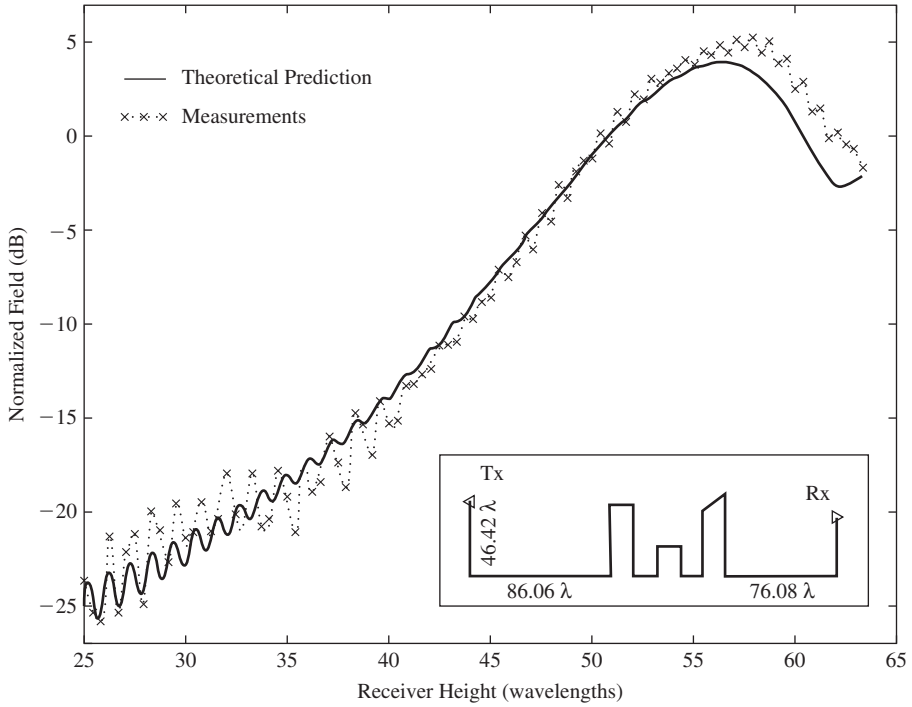
Figure 16-37 Propagation with Tx above rooftops. Parallel polarization case: mean error 1.06 dB; standard deviation 0.85 dB. (Used with permission from Erricolo [25] © 2001 IEEE.)
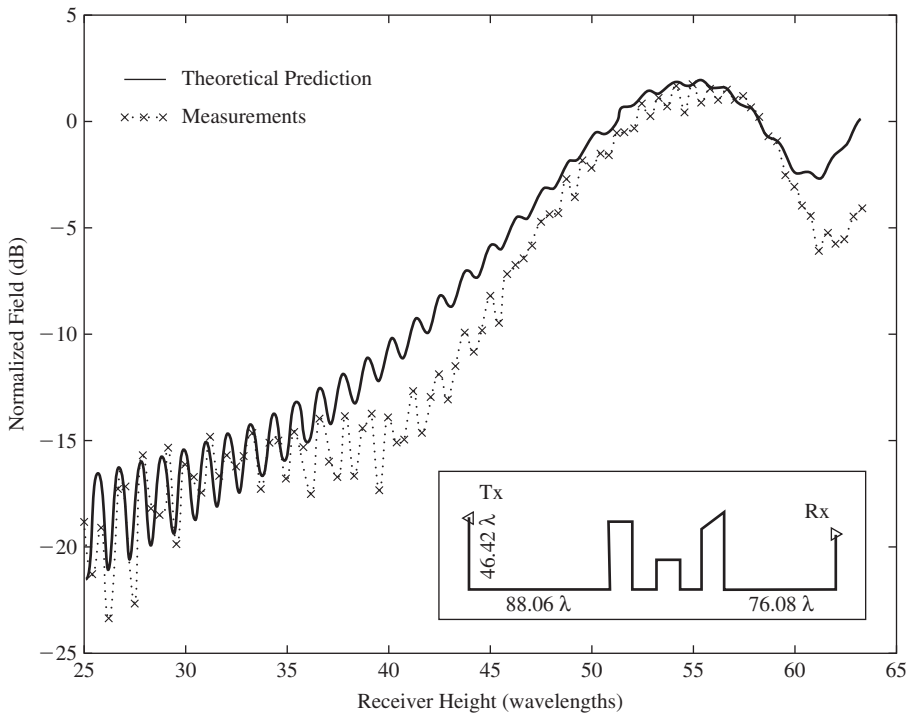
Figure 16-38 Propagation with Tx above rooftops. Perpendicular polarization case: mean error 1.78 dB; standard deviation 1.35 dB. (Used with permission from Erricolo [25] © 2001 IEEE.)

wedge or screen. We can add slope diffraction to (16-57) and write that matrix equation in compact form as [H.10.4: McNamara]

$$E^d_{\parallel,\perp}(s) = -\left[D_{\parallel,\perp} E^i_{\parallel,\perp}(Q) + \frac{1}{j\beta} \frac{\partial D_{\parallel,\perp}}{\partial \phi'} \frac{\partial E^i_{\parallel,\perp}(Q)}{\partial n} \right] A(s) e^{-j\beta s} \tag{16-128}$$

which applies to both perfectly conducting and imperfectly conducting wedges when the appropriate reflection representation is used in the diffraction coefficient.

The parameter, L, implicit in the preceding expression is determined by continuity equations in [23] which ensure the continuity of the diffracted field amplitude across the shadow boundaries. This concept was improved in [24] which allowed for independent continuity relations for each ray in an n-ray propagation. A tutorial review of wireless propagation can be found in [27], and a more detailed discussion in [H.4: Saunders].

16.13 EXTENSION OF MOMENT METHOD USING THE UTD

In Chap. 14, we saw how the MoM could be applied to many antenna and scattering problems where the antenna or scatterer was not excessively large in terms of the wavelength. In this chapter, we have seen how geometrical optics and GTD can be applied to problems that are large in terms of wavelength. The purpose of this section is to show how the class of problems solvable by moment methods can be enlarged by incorporating GTD into the moment method solution [26]. In studying this section, the student will have an opportunity to test his or her understanding of the fundamental concepts developed in Chap. 14 and the previous sections of this chapter.

Recall from Chap. 14 the elements of the generalized impedance matrix can be given in inner product notation as

$$Z_{mn} = \langle \mathbf{J}_m, \mathbf{E}_n \rangle \tag{16-129}$$

This states that the $Z_{mn}th$ element of the impedance matrix is found by reacting the mth test function (weight function) with the electric field from the nth basis function. Similarly, the mth element in the generalized voltage matrix is found by reacting the mth test function with the incident field.

In a strictly moment method formulation of a given problem, all material bodies are removed and replaced with equivalent currents radiating in free space. Thus, when one reacts the mth test function with the field from the nth basis function, it is only that field which directly arrives at the mth test function via the shortest free-space path that one needs to consider since it is the only possible field. However, suppose there exists in a given situation a portion of the structure that is not represented by equivalent currents (i.e., a material body remains as shown in Fig. 16-39). In this case, the calculation of the impedance matrix elements is more complex, but not unduly so. Let these new impedance matrix elements be denoted Z'_{mn}. In terms of (16-129), the reaction of $\mathbf{J}_m$ with $\mathbf{E}_n$ may

Material body replaced by equivalent currents in localized free space

Material structure not characterized by equivalent currents
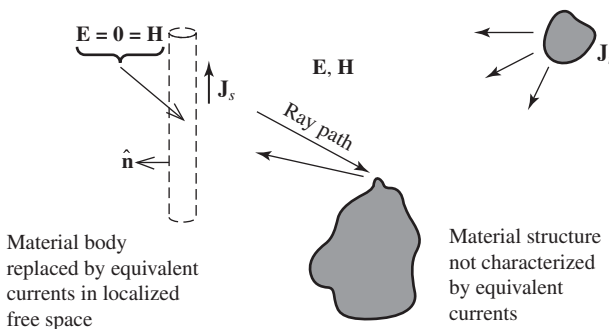
Figure 16-39 A source radiating in free space with one scatterer replaced by an equivalent current and the other remaining as a material body.

be interpreted to mean the reaction of the test source with not only the field from the true source arriving at the test source directly, but also the reaction of the test source with fields from the true source that arrive by other means, as suggested by Fig. 16-39. Therefore, one can write

$$Z'_{mn} = \langle \mathbf{J}_m, a\mathbf{E}_n + b\mathbf{E}_n \rangle \qquad (16\text{-}130)$$

where a may be set to unity and $b = b(m, n)$ is different for each m and n. The quantity $b\mathbf{E}_n$ also represents the field due to $\mathbf{J}_n$, but arriving at the mth observation point or region due to a physical process, such as a geometrical optics or diffraction mechanism, which is not accounted for in that portion of the problem formulated by the moment method. Thus,

$$Z'_{mn} = \langle \mathbf{J}_m, \mathbf{E}_n \rangle + \langle \mathbf{J}_m, b\mathbf{E}_n \rangle \qquad (16\text{-}131)$$

or

$$Z'_{mn} = Z_{mn} + Z^g_{mn} \qquad (16\text{-}132)$$

where the superscript g denotes that Z^g_{mn} is an additional term added to, in general, each impedance matrix element due to a physical process g that redirects energy from the nth basis current function to the location of the mth test source.

As implied by Fig. 16-39, there is also a modification of the usual generalized voltage matrix terms. That is,

$$V'_m = \langle \mathbf{J}_m, \mathbf{E}^i + c\mathbf{E}^i \rangle \qquad (16\text{-}133)$$

where $\mathbf{E}^i$ is the incident field arriving directly at region m and $c\mathbf{E}^i$ is that field from the source redirected to region m by a physical process g. We note that $c = c(m)$ is different for each m:

$$V'_m = \langle \mathbf{J}_m, \mathbf{E}^i \rangle + \langle \mathbf{J}_m, c\mathbf{E}^i \rangle \qquad (16\text{-}134)$$

or

$$V'_m = V_m + V^g_m \qquad (16\text{-}135)$$

As a direct consequence of the foregoing discussion, we have

$$[Z'][I'] = [V'] \qquad (16\text{-}136)$$

and its solution as

$$[I'] = [Z']^{-1}[V'] \qquad (16\text{-}137)$$

where $[I']$ is the current on, for example, an antenna operating in the presence of scattering mechanisms that may be accounted for by either geometrical optics techniques or GTD.

Initially, to combine the method of moments and GTD into a hybrid technique [28], consider the problem of a monopole near a perfectly conducting wedge as show in Fig. 16-40. If we describe the monopole on an infinite ground plane strictly by the moment method matrix representation given in (14-43), then for the monopole near the conducting wedge, we utilize (16-135), where in (16-131) the term Z^g_{mn} is obtained by considering that energy radiated by the nth basis function on the monopole that is diffracted by the wedge to the mth observation point or region.

To investigate a monopole on a circular ground plane as in Fig. 16-41, we must use the equivalent edge currents described in Sec. 16.9. Thus, we replace the edge of the disk with an equivalent magnetic current M given by
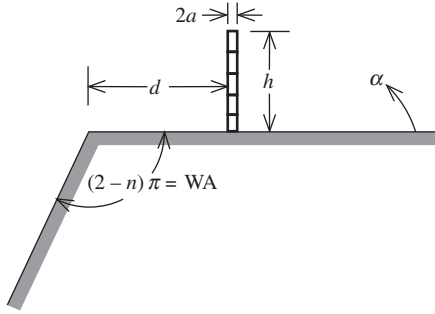
Figure 16-40 Monopole on a conducting wedge.

$$M = -2E_\theta e^{-j(\pi/4)} D_\perp (L, \phi, \phi') \sqrt{\lambda} \qquad (16\text{-}138)$$

This equivalent magnetic current is used to calculate the field at the segment at s due to the current at s' as indicated in Fig. 16-41. Note that an equivalent magnetic ring current must be calculated for each choice of s and s'.

It is useful for us to break up the equivalent magnetic ring current of Fig. 16-41 into differential elements dc' so that the observation point is in the far field of each element even though it may be in the near field of the total ring current.

The electric field in a plane perpendicular to an element dc' is given by

$$dE_z = \frac{M \, dc'}{4\pi} \left(\frac{j\omega}{cr} + \frac{1}{r^2} \right) e^{-j\beta r} \qquad (16\text{-}139)$$

where $r = (a^2 + z^2)^{1/2}$. Letting $dc' = a \, d\psi$ where ψ is the azimuth angle, taking only the z-component at the monopole, and integrating over the range $\psi = 0$ to $\psi = 2\pi$ yields

$$E_z = \frac{Ma^2}{2r} \left(\frac{j\beta}{r} + \frac{1}{r^2} \right) e^{-j\beta r} \qquad (16\text{-}140)$$

The value for E_z is the term Z^g_{mn} that is added to the impedance element obtained for a monopole on an infinite ground plane. This process gives the modified impedance element needed, Z'_{mn}, to calculate the modified currents (and hence input impedance) of a monopole on the finite circular ground plane.
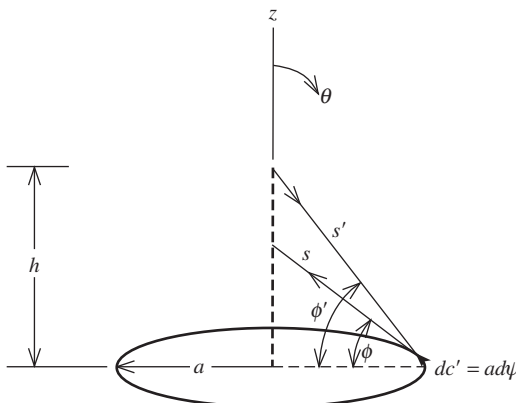
Figure 16-41 Segmented monopole encircled by a magnetic ring current for analysis of a monopole on a circular ground plane.

Figs. 16-42*a* and 16-42*b* show a comparison between calculations made with the equivalent magnetic ring current and measurements for a monopole of length 0.224λ and radius 0.003λ on a circular ground plane for varying radius. It is apparent that the correct variation is accurately predicted for both the real and imaginary parts of the input impedance. For the input resistance, the agreement between the measurements and the theory is excellent. For the input reactance, the agreement is very good, but there is a slight shift in the calculated curve when compared to the measurements. The amount of this shift is sufficiently small that it can be attributed to the usual problems associated with modeling the region in proximity to the driving point.

Next, consider the situation shown in Fig. 16-43 where a monopole of height h is a distance d_1 away from a vertical conducting step. To properly determine the Z_{mn}^g term in (16-131), it is necessary to determine all the various combinations of reflections that can occur for rays emanating from the monopole and reflecting back to it as well as the diffraction from the top edge of the step. Since the vertical wall is at a right angle to the lower horizontal surface, there will be no diffraction from the interior wedge and all the reflections can most conveniently be accounted for by imaging the monopole into the horizontal ground plane and then imaging the resulting dipole into the plane of the vertical wall.
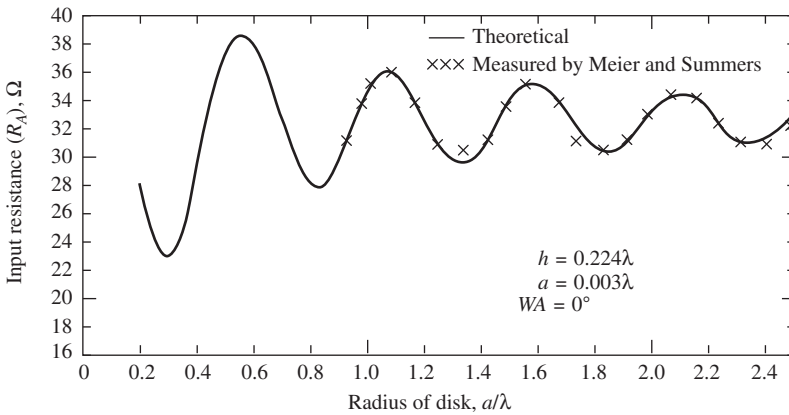
(*a*) Input resistance.
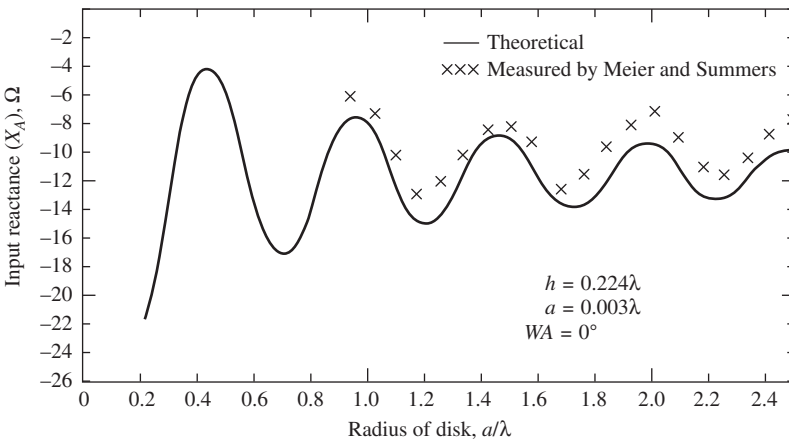
(*b*) Input reactance.

Figure 16-42 Theoretical and experimental input impedance of a monopole of radius 0.003λ at the center of a circular disk as shown in Fig. 16-41.
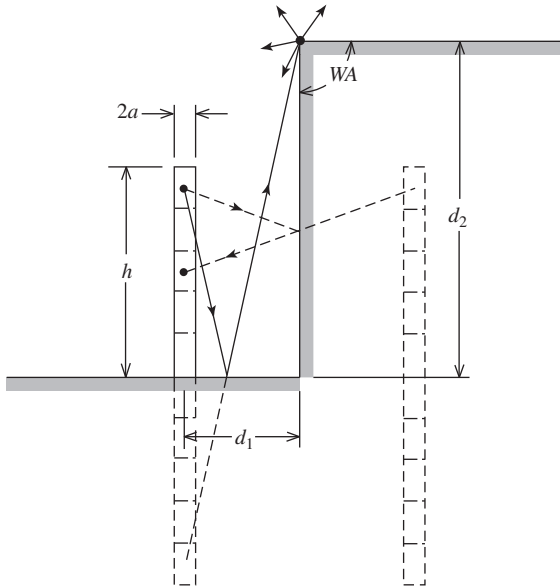
Figure 16-43 Monopole near a conducting step showing the partial use of images.

Shown in Fig. 16-43 are two example situations that depict the utilization of the images. If we consider the uppermost segment of the monopole to be the source segment, one set of rays shows the use of the image in the horizontal surface to calculate reflected-diffracted energy reaching the segments of the monopole. The other set of rays shows the use of the image in the vertical wall to calculate singly reflected energy. In the calculated results that follow, all combinations of singly reflected, doubly reflected, diffracted, diffracted-reflected, reflected-diffracted, and reflected-diffracted-reflected rays are taken into account. All rays that involve combinations of double (or higher-order) diffractions are negligible.

Fig. 16-44 shows the calculated input impedance for a quarter-wavelength monopole a quarter-wavelength away from a vertical wall whose height is $d_2 > 0.25\lambda$. As d_2 increases, the impedance oscillates about the value for the case where $d_2 = \infty$. The results of Figs. 16-44a and 16-44b shows that as the diffracting edge recedes from the vicinity of the monopole, its effect on the input impedance rapidly diminishes. Although we have not shown results for the case where the step height is less than the height of the monopole, the same method could be used to investigate such situations.

In combining moment methods with GTD, we have proceeded from the philosophical viewpoint of extending the method of moments via GTD. In so doing, we have shown that modifying the impedance matrix to account for diffraction effects (or geometrical optics effects) enables one to accurately treat a larger class of problems than could be treated by moments methods alone. An alternative interpretation of the hybrid method is also possible. That is, the procedure employed can be viewed as using UTD to obtain an approximation to the exact Green's function needed. Other hybrid methods are discussed in [29].

Although this hybrid method possesses many of the advantages inherent in both the moment method and UTD, it also has some of the limitations peculiar to each. For example, as in the usual moment method, one can treat arbitrary configurations of wire antennas (or slot antennas), taking into account lumped loading, finite conductivity, and so forth, and obtain accurate impedance data and current distributions. Naturally, one still must take the usual precaution of using a sufficient number of basis functions to assure convergence. On the other hand, as in the usual UTD problem, one must take care that the antenna is not too close to a source of diffraction (e.g., $d > 0.2\lambda$).

$d_2 = \infty$

$h = 0.25\lambda$
$a = 0.00762\lambda$
$WA = 90°$
$d_1 = 0.25\lambda$

(*a*) Input resistance.

$d_2 = \infty$

$h = 0.25\lambda$
$a = 0.00762\lambda$
$WA = 90°$
$d_1 = 0.25\lambda$
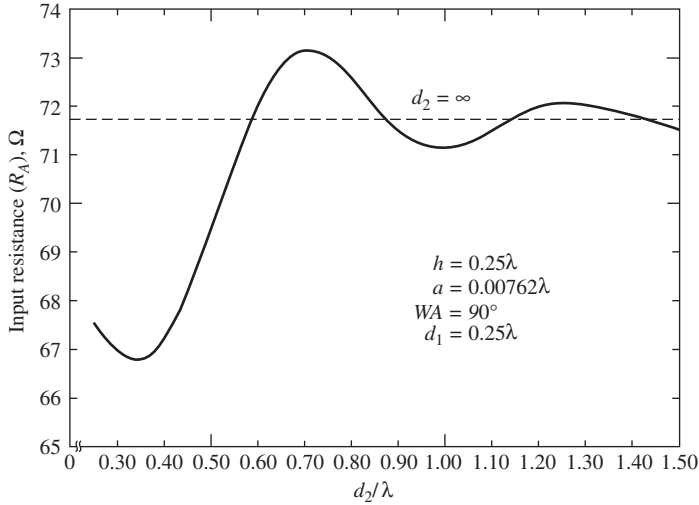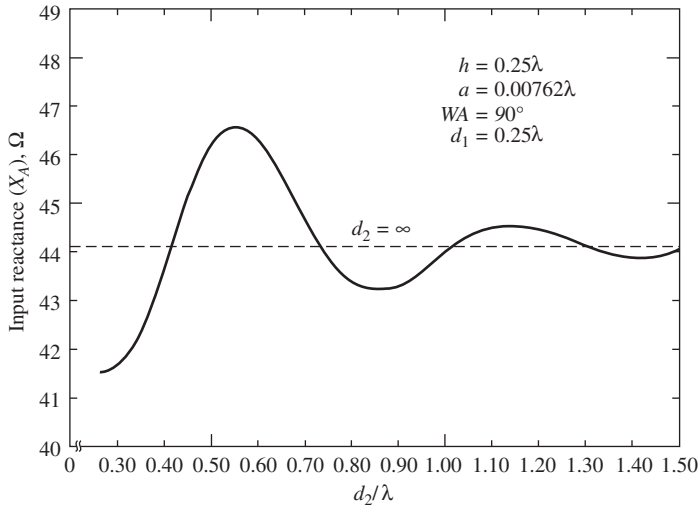
(*b*) Input reactance.

Figure 16-44 Input impedance of a quarter-wave monopole as a function of step height for the geometry of Fig. 16-43.

16.14 PHYSICAL OPTICS

In Sec. 16.1, we calculated the scattered field from the sphere by geometrical optics. Often, we can calculate these same scattered fields by physical optics. The concept of *physical optics* can be considered to be somewhat more general than geometrical optics, since the equations obtained from physical optics for the scattered field from a conducting body often reduce to the equations of geometrical optics in the high frequency limit. In fact, it is assumed in physical optics that the field at the surface of the scattering body is the geometrical optics surface field. This implies that at each point on the illuminated side of the scatterer, the scattering takes place as if there were an infinite tangent plane at that point, while over the shadowed regions of the scatterer the field at the surface is zero [2].
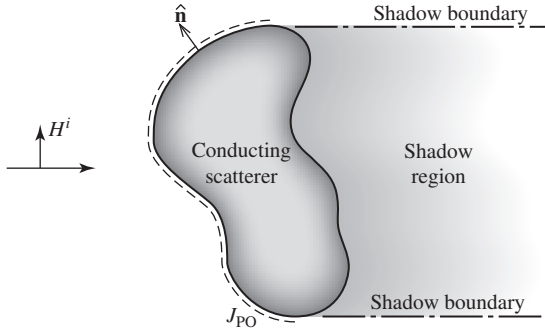
Figure 16-45 Physical optics current on a conducting scatterer.

For a perfectly conducting body, the assumed physical optics surface current is

$$\mathbf{J}_{PO} = \begin{cases} \hat{\mathbf{n}} \times \mathbf{H}_{total} & \textit{in the illuminated region} \\ 0 & \textit{in the shadowed region} \end{cases} \tag{16-141}$$

where $\hat{\mathbf{n}}$ is a unit normal vector outward from the surface of interest as shown in Fig. 16-45.

From image theory, the tangential components of $\mathbf{H}$ at a perfect conductor are just twice those from the same source when the conducting scatterer is replaced by equivalent currents in free space. Thus, the physical optics current is given by

$$\mathbf{J}_{PO} = 2(\hat{\mathbf{n}} \times \mathbf{H}^i) \tag{16-142}$$

From Chap. 2, we know that in the far field

$$\mathbf{E}^s = -j\omega\mathbf{A} \tag{16-143}$$

having neglected any radial terms. For the purposes of simplification, let ψ represent the free-space Green's function as in (2-59). Then

$$\mathbf{E}^s = -j\omega\mu\iint \mathbf{J}\psi \ ds' \tag{16-144}$$

Using the curl E equation, we can write for $\mathbf{H}^s$

$$\mathbf{H}^s = \nabla \times \iint \mathbf{J}\psi \ ds' \tag{16-145}$$

Since the del operation is in the unprimed coordinate system and the integration is in the primed coordinate system, we can write

$$\mathbf{H}^s = \iint (\nabla \times \mathbf{J} \, \psi) \ ds' \tag{16-146}$$

Since $\nabla \times \mathbf{J}\psi = \nabla\psi \times \mathbf{J} + \psi\nabla \times \mathbf{J}$ using (C-16), and the last term on the right is zero,

$$\mathbf{H}^s = \iint (\nabla\psi \times \mathbf{J}) \ ds' \tag{16-147}$$

Since $R = r - \hat{\mathbf{r}} \cdot \mathbf{r}'$, as in (2-98), we can express $\nabla\psi$ in the far field as

$$\nabla\psi = -\hat{\mathbf{r}} \frac{1 + j\beta r}{4\pi r^2} e^{-j\beta r} e^{j\beta\hat{\mathbf{r}} \cdot \mathbf{r}'} \tag{16-148}$$

giving

$$\mathbf{H}^s = e^{-j\beta r} \iint (\mathbf{J} \times \hat{\mathbf{r}}) \frac{1 + j\beta r}{4\pi r^2} e^{j\beta \hat{\mathbf{r}} \cdot \mathbf{r}'} \, ds' \qquad (16\text{-}149)$$

which is approximately equal to

$$\mathbf{H}^s \approx \frac{j\beta}{4\pi r} e^{-j\beta r} \iint (\mathbf{J} \times \hat{\mathbf{r}}) e^{j\beta \hat{\mathbf{r}} \cdot \mathbf{r}'} \, ds' \qquad (16\text{-}150)$$

since $\beta r \gg 1$ and $r \approx R$ in the denominator. It should be noted that this expression for the scattered field is frequency dependent in contrast to the geometrical optics expression that is frequency independent. It might, therefore, be intuitively inferred that physical optics provides a more accurate approximation to the scattered field. Although this may be so in certain cases, a general conclusion cannot be reached since necessary and sufficient conditions for the valid application of physical optics are not known [2]. It is fortunate for the engineer that physical optics works in many practical problems, even though in some problems prior justification of its application would be difficult to make.

Next, let us develop a general expression for the radar cross section using our expression in (16-150) for $\mathbf{H}^s$. Writing the radar cross-section definition as

$$\sigma = \lim_{r \to \infty} 4\pi r^2 \frac{|\mathbf{H}^s|^2}{|\mathbf{H}^i|^2} \qquad (16\text{-}151)$$

and inserting the physical optics current $\mathbf{J} = 2\hat{\mathbf{n}} \times \mathbf{H}^i$ gives

$$\sigma = \lim_{r \to \infty} 4\pi r^2 \left(\frac{\beta}{4\pi}\right)^2 \left|\frac{1}{\mathbf{H}^i}\right|^2 \left|\iint ((2\hat{\mathbf{n}} \times \mathbf{H}^i) \times \hat{\mathbf{r}}) \frac{e^{j\beta \hat{\mathbf{r}} \cdot \mathbf{r}'}}{r} \, ds'\right|^2 \qquad (16\text{-}152)$$

which reduces to

$$\sigma = \frac{\pi}{\lambda^2} \left|\frac{1}{\mathbf{H}^i}\right|^2 \left|\iint ((2\hat{\mathbf{n}} \times \mathbf{H}^i) \times \hat{\mathbf{r}}) e^{j\beta \hat{\mathbf{r}} \cdot \mathbf{r}'} \, ds'\right|^2 \qquad (16\text{-}153)$$

Using the vector identity (C-8), we can write

$$(\hat{\mathbf{n}} \times \mathbf{H}^i) \times \hat{\mathbf{r}} = (\hat{\mathbf{r}} \cdot \hat{\mathbf{n}})\mathbf{H}^i - (\hat{\mathbf{r}} \cdot \mathbf{H}^i)\hat{\mathbf{n}} \qquad (16\text{-}154)$$

At this point, consider the backscattered or monostatic radar cross section where (16-154) reduces to $(\hat{\mathbf{r}} \cdot \hat{\mathbf{n}})\mathbf{H}^i$ and the phase of $\mathbf{H}^i$ is $e^{j\beta \hat{\mathbf{r}} \cdot \mathbf{r}'}$ on the illuminated surface. Thus,

$$\sigma = \frac{4\pi}{\lambda^2} \left|\iint (\hat{\mathbf{r}} \cdot \hat{\mathbf{n}}) e^{j2\beta \hat{\mathbf{r}} \cdot \mathbf{r}'} \, ds'\right|^2 \qquad (16\text{-}155)$$

Next, take $\hat{\mathbf{r}} = \hat{\mathbf{z}}$ for the purpose of monostatic illustration (i.e., the radar is on $+z$ axis), giving us a final result of

$$\sigma = \frac{4\pi}{\lambda^2} \left|\iint (\hat{\mathbf{z}} \cdot \hat{\mathbf{n}}) e^{j2\beta \hat{\mathbf{z}} \cdot \mathbf{r}'} \, ds'\right|^2 \qquad (16\text{-}156)$$

where the factor of 2 in the exponent represents the phase advance of the backscattered field relative to the origin due to the two-way path.

EXAMPLE 16-4   **RCS of a Sphere by Physical Optics**

Here we apply the result in (16-156) to obtain the physical optics expression for the monostatic RCS of a sphere and compare it to the result obtained by geometrical optics. First, we note that from Fig. 16-46

$$\hat{\mathbf{z}} \cdot \hat{\mathbf{n}} = \cos \theta' = \frac{a - \ell'}{a} \qquad (16\text{-}157)$$

where ℓ' is the distance from the reference plane to the spherical surface, and that an element of surface area is $a^2 \sin \theta' \, d\theta' \, d\phi'$ on the spherical surface. Since $a - \ell' = a \cos \theta'$ and $d\ell' = a \sin \theta' \, d\theta'$, we find that on the projected area of the sphere onto the reference plane is $ds' = a^2 \sin \theta' \, d\theta' \, d\phi' = a \, d\phi' \, d\ell'$. Noting that $\hat{\mathbf{z}} \cdot \mathbf{r}' = (a - \ell')$, substituting into our general expression for monostatic RCS, and performing the ϕ integration, we obtain

$$\sigma_{sp} = \frac{4\pi}{\lambda^2} \left| 2\pi \ e^{j2\beta a} \int_0^a e^{-j2\beta \ell'} (a - \ell') \ d\ell' \right|^2 \qquad (16\text{-}158)$$

Performing the remaining integration yields

$$\sigma_{sp} = \frac{4\pi}{\lambda^2} \left| 2\pi \ \frac{e^{j2\beta a}}{j2\beta} \left[a - \frac{1 - e^{-j2\beta a}}{j2\beta} \right] \right|^2 \qquad (16\text{-}159)$$

which can be put in the form

$$\sigma_{sp} = \frac{4\pi}{\lambda^2} \left| \frac{a\lambda}{j2} \left[\left(1 + \frac{j}{2\beta a} \right) e^{j2\beta a} - \frac{j}{2\beta a} \right] \right|^2 \qquad (16\text{-}160)$$

The terms associated with the exponential $e^{j2\beta a}$ are due to the front face (i.e., specular) reflection, whereas the right-most term $(j/2\beta a)$ is due to the contribution from the artificially imposed discontinuity in the current at the $\theta = \pi/2$ location on the sphere (i.e., shadow boundary). Since this discontinuity is nonphysical, so too is the right-most term $(j/2\beta a)$ and we must disregard it. Thus,

$$\sigma = \pi a^2 \left| \frac{1}{j} \left(1 + \frac{j}{2\beta a} \right) e^{j2\beta a} \right|^2 \xrightarrow[\beta a \to \infty]{} \pi a^2 \qquad (16\text{-}161)$$

We see then that the radar cross section of the sphere obtained via physical optics reduces to the geometrical optics result of (16-23) in the high-frequency limit.

   The fact that we have had to eliminate the right-most term $(j/2\beta a)$ in (16-140) is not a peculiarity of the sphere, but is common to any problem employing physical optics where a nonphysical discontinuity in current gives rise to an erroneous contribution to the scattered field that can be numerically significant when compared to the geometrical optics contribution.
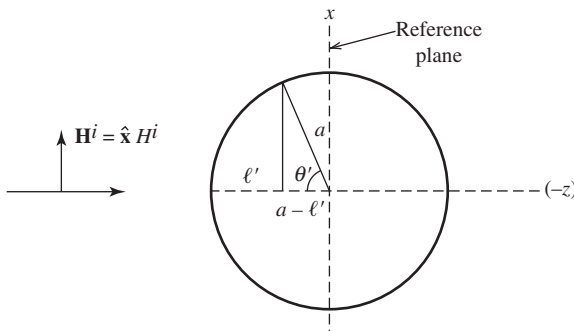
Figure 16-46 Physical optics scattering by a sphere.
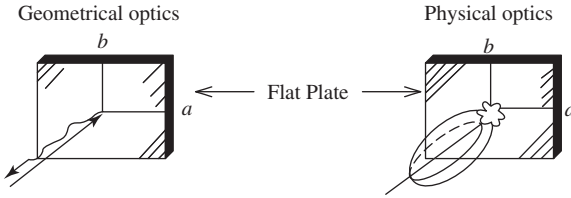
Figure 16-47 Geometrical optics compared to physical optics.

The second term in (16-161) may be taken to be the second term in a high-frequency asymptotic expansion of the scattered field. Such an expansion is in inverse powers of the frequency and is known as Luneburg–Kline expansion [1]. The Luneburg–Kline expansion satisfies the wave equation and is a formal way of showing the correspondence between optics and electromagnetics in the high-frequency limit. The leading term in the Luneburg–Kline expansion is, in fact, the geometrical optics term that is also the first term in (16-161).

Physical optics is more useful to us than just finding radar cross sections. For example, if we wish to find the far-field pattern of a parabolic reflector antenna, physical optics is one way of doing so. In fact, it is probably the easiest way of finding the radiated field on the forward axis of the reflector antenna. In directions other than on the forward axis of the reflector antenna, physical optics provides us with a nonzero estimate of the radiation pattern. This should be contrasted with geometrical optics that can only provide information in a specular direction (see Fig. 16-47), but does so in a straightforward manner. Fig. 16-47 shows a ray normally incident on a flat plate and the reflected or scattered field coming back only in one direction, whereas the figure indicates that the physical optics current produces a scattered field in all directions for the same incident field.

EXAMPLE 16-5 *RCS of a 3-D Parabolic Reflector*

As an example of the application of (16-156) to antenna scattering, consider a circularly symmetric parabolic reflector (see Fig. 9-25) antenna with the prime focus feed temporarily removed, as shown in Fig. 16-48. We wish to compute the monostatic RCS on the z-axis. To do this, we need to determine $\hat{\mathbf{n}}$, r', z', and ds' in (16-156).

The location of the origin is on the parabola center, and the equation of the parabola is $(\rho')^2 = 4Fz'$, where F is the focal length. Using this for z' in the representation of r', we have

$$\mathbf{r}' = \rho'\hat{\boldsymbol{\rho}} + z'\hat{\mathbf{z}} = \rho'\hat{\boldsymbol{\rho}} + \frac{(\rho')^2}{4F}\hat{\mathbf{z}} \tag{16-162}$$

The normal is found by setting the gradient of $4Fz' - (\rho')^2$ equal to zero, which yields

$$\hat{\mathbf{n}} = \frac{-\rho'\hat{\boldsymbol{\rho}} + 2F\hat{\mathbf{z}}}{\sqrt{4F^2 + (\rho')^2}} \tag{16-163}$$

The differential surface element on the curved surface of the parabola in Fig. 16-48 is dl' by dw' where the circular arc is $dl' = \rho' d\phi'$ and the parabolic arc is

$$dw' = \sqrt{1 + \left(\frac{dz'}{d\rho'}\right)^2}\, d\rho' = \sqrt{1 + \left(\frac{\rho'}{2F}\right)^2}\, d\rho' = \frac{\sqrt{4F^2 + (\rho')^2}}{2F}\, d\rho' \tag{16-164}$$

Thus, the differential element of curved surface area is

$$ds' = \frac{\sqrt{4F^2 + (\rho')^2}}{2F}\, \rho' d\rho' d\phi' \tag{16-165}$$
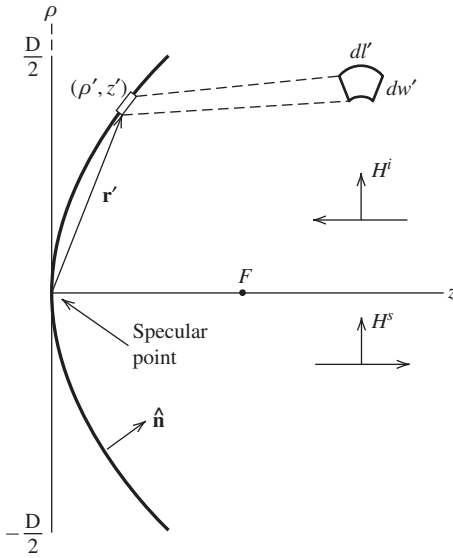
Figure 16-48 Circularly symmetric parabolic reflector. Monostatic case with axial incidence.

Next we have for $(\hat{\mathbf{z}} \cdot \hat{\mathbf{n}})ds'$ in (16-156)

$$(\hat{\mathbf{z}} \cdot \hat{\mathbf{n}})ds' = \rho' d\rho' d\phi' \tag{16-166}$$

and for the exponential term

$$2\beta\hat{\mathbf{z}} \cdot \mathbf{r}' = 2\beta\frac{(\rho')^2}{4F} = \frac{\pi}{\lambda F}(\rho')^2 \tag{16-167}$$

Then (16-156) becomes

$$\sigma = \frac{4\pi}{\lambda^2}\left|\iint e^{j\frac{\pi}{\lambda F}(\rho')^2}\rho' d\rho' d\phi'\right|^2 \tag{16-168}$$

Let

$$u = j\frac{\pi}{\lambda F}(\rho')^2 \qquad \text{and} \qquad du = j\frac{2\pi}{\lambda F}\rho' d\rho' \tag{16-169}$$

$$u_1 = 0 \quad \text{and} \quad u_2 = j\frac{\pi}{\lambda F}\left(\frac{D}{2}\right)^2 \tag{16-170}$$

After performing the trivial ϕ integration in (16-168), we have

$$\sigma = \frac{4\pi}{\lambda^2}\left|\frac{\lambda F}{j}\int_{u_1}^{u_2}e^u du\right|^2 \rightarrow 4\pi F^2 \tag{16-171}$$

Similar to the previous example for the sphere, we have identified the e^{u_2} term resulting from the integration as being associated with the non-physical abrupt termination of the physical optics current at the rim of the parabola, and have discarded that term. The simple result in (16-171) may be confirmed by a consideration of Prob. 16.1-4c which obtains the same result via geometrical optics [30, p. 618]. The result in (16-171) is the same result as for the convex case since the dot product $\hat{\mathbf{z}} \cdot \hat{\mathbf{n}}$ in the kernel in (16-156) is the same in both cases, and except for sign, the exponential term in (16-167) is the same for both cases as well. Furthermore, the results for the convex case agree with the exact solution [30, p. 593]. The results for the concave case are limited

to cases where there are no multiple reflections that contribute to the monostatic RCS. The procedure used here for the special case of monostatic axial incidence may be generalized for oblique incidence. The details are left as an exercise for the interested reader.

The result in Example 16-5 gives the major part of the structural, σ_s (or residual) component of the antenna scattering by a front-fed circular parabolic reflector antenna at boresight (i.e., $\theta = 0°$). Recall from Sec. 14.12 that here is also a contribution to the total scattering, σ_T, by the antenna mode (see (14-111) for σ_{ant}). The total cross section can be written as [31]

$$\sigma_T = \sigma_s + \sigma_{ant} + 2\sqrt{\sigma_s \sigma_{ant}} \, \cos \phi \qquad (16\text{-}172)$$

where ϕ is the relative phase between the structural scattering and antenna scattering.

Fig. 16-49 [H.8.2: Munk, p. 229] shows the bistatic scattering from a parabolic cylindrical reflector of modest size for boresight incidence ($\theta = 0°$). Fig. 16-49 presents bistatic RCS results for no feed and also for a low RCS feed. The presence of the low RCS feed produces a lower level of bistatic scattering in a wide angular region about
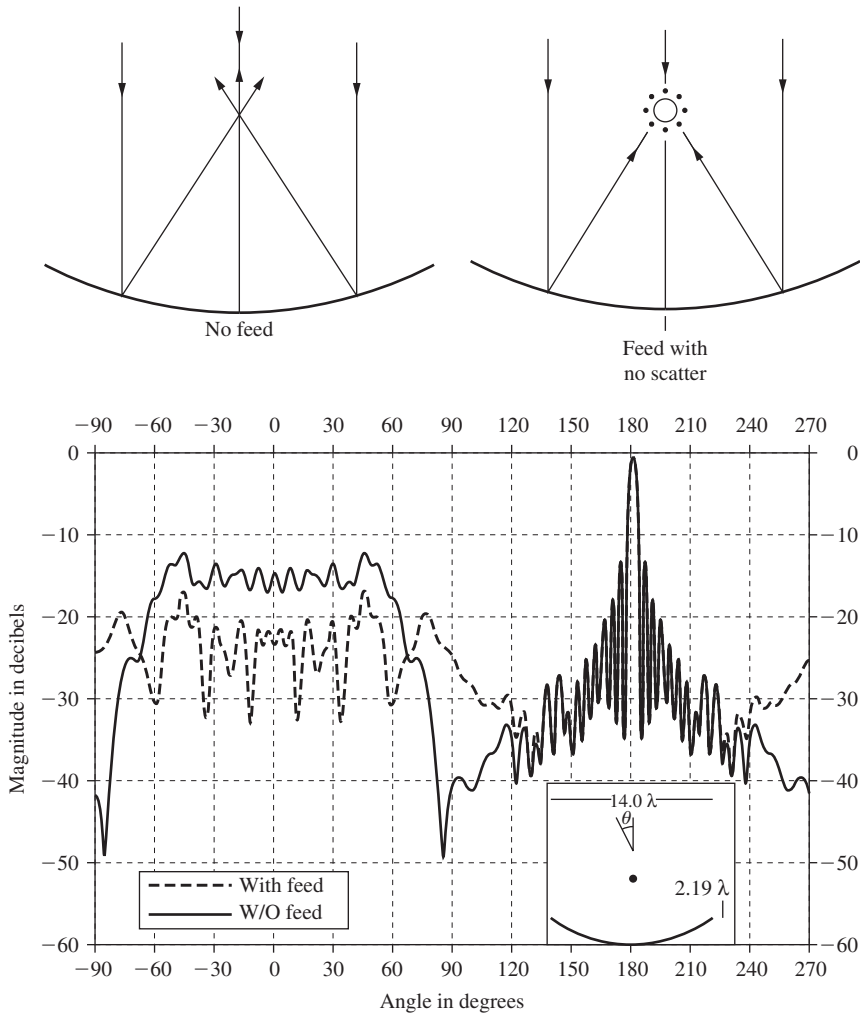
Figure 16-49 The bistatic scattered field from a parabolic cylinder for an incident plane wave along boresight; that is, $\theta = 0°$. Shown with and without a low RCS feed. Frequency is 9.0 GHz. (From [H.8.2: Munk] © 2003. Reprinted by permission of John Wiley & Sons.)

$$\beta a = 8.94$$

TE
polarization

TM
polarization

Physical
optics

θ, degrees

$\dfrac{2\pi}{\lambda}\,\sigma_w$
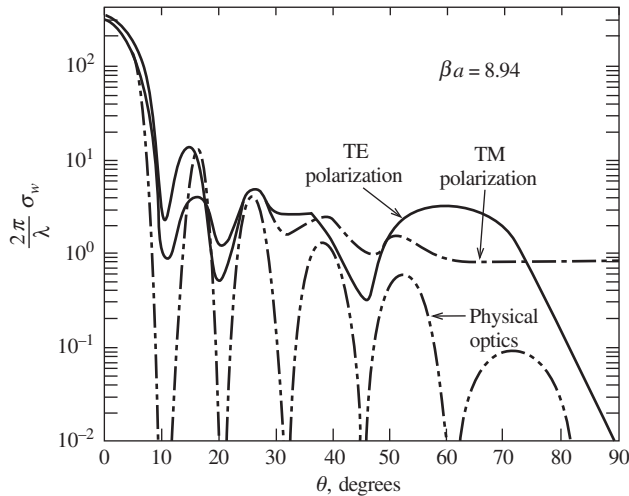
Figure 16-50 TE and TM physical optics scattering by a strip of width a.

boresight than with the feed removed. This is due to the feed absorbing power incident upon it rather than allowing that power to return to the radar.

The bistatic scattering remains strong in Fig. 16-49 as long as there is a defined normal to the parabolic surface that bisects the angle of incidence and the angle of reflection. In Fig. 16-49 this accounts for the bistatic scattering for approximately $-45° < \theta < 45°$ due the normal making an approximate angle of $\theta = 22.5°$ at the edges of the parabolic cylinder (i.e., the angle of the normal bisects the angle formed by the incident ray at $0°$ and the angle of the reflected ray at $45°$).

If the feed is not low RCS and/or is seriously mismatched at an out-of-band frequency, there can be a strong monostatic return at $\theta = 0°$ [H.8.2: Munk, p. 227] due to the antenna mode. The strong RCS at $\theta = 180°$ is due to the interesting phenomena of forward scattering. The two-dimensional H-plane results in Fig. 16-49 are representative of H-plane results for the three-dimensional parabolic reflector in Example 16-5.

In summary, physical optics is an approximate high-frequency method of considerable usefulness that can be expected to provide an accurate representation of the scattered field arising from a surface where the postulated physical optics current is reasonably close to the true current distribution. We recall from the discussion at the beginning of this section that the physical optics current will be a reasonable representation of the true current if the field at the scatterer surface is correctly given by the geometrical optics surface field.

An example of a situation where the geometrical optics surface field does not give us the true current is in the vicinity of an edge (where a plane tangent to the surface is not defined). Consider Fig. 16-50 that shows $\beta\sigma_w$ for an infinite strip when physical optics is used compared to an exact solution [2]. Both TE (E perpendicular to the edges) and TM (E parallel to the edges) are shown. As the backscatter angle θ moves away from the normal to the strip ($\theta = 0°$), the difference between physical optics and the correct solution for the two polarizations becomes increasingly larger. This difference may be eliminated by adding to the physical optics current an additional current to account for each edge. This is the subject of Sec. 16.17.

Before doing that, we will utilize physical optics to investigate the frequency dependence of some common sources of scattering.

16.15   FREQUENCY DEPENDENCE OF FIRST-ORDER SCATTERING SOURCES

Part of antenna scattering is the scattering from the structure which comprises the antenna and its immediate platform. It is insightful, therefore, to examine the frequency

dependence of the major geometrical features likely to be found as part of an antenna and/or the platform upon which the antenna is placed. To do this we will examine the monostatic RCS of a number of elementary shapes having a common geometrical feature. The dominant scattering contributors in elementary shapes will have scattering cross sections that vary with frequency (or wavelength) according to the form, wavelength times a function of a characteristic dimension, d, of the object as [32]:

$$\sigma = \lambda^n f(d/\lambda) \tag{16-173}$$

where $n = -2, -1, 0, 1, 2$ for the geometries we will consider here. The function $f(d/\lambda)$ in (16-173) is different for each shape and is not of concern in this section.

Some elementary scattering contributors can be classified in terms of their principal radii of curvature at the point where the normal to the surface is parallel to the direction of incidence. Table 16-2 provides some examples.

Fig. 16-51 shows the backscatter radar cross-section dependence of five simple shapes. These are a circular flat plate (FP), right circular cylinder (CYL), sphere (SPH), curved edge (CE), and cone-sphere (CS). The RCS is in dBsm (dB relative to a square meter) versus frequency. The RCS is the monostatic RCS viewed on the axis (see Fig. 16-52) of

Table 16-2 Classification of Elementary Scatterers

n	radii of curvature	geometry
-2	two infinite radii	plane
-1	one infinite, one nonzero finite	curved cylinder
0	two nonzero finite	spheroid
0	one infinite, one zero	straight edge
1	one nonzero finite, one infinite	non-planar edge
2	two zero	tip or point
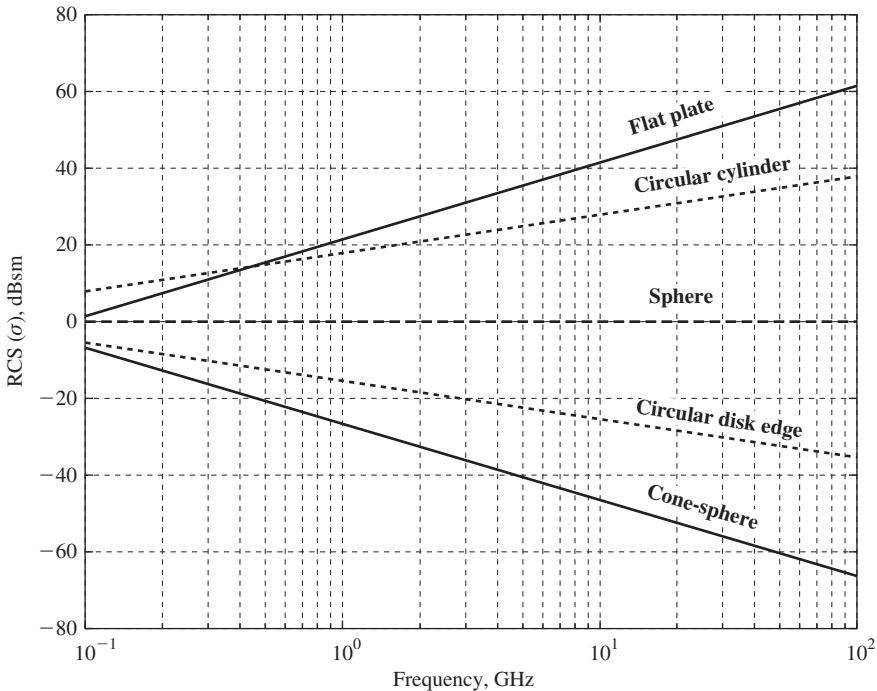
Figure 16-51 Frequency dependence of five simple scattering shapes.

(a) Flat circular plate.

(b) Circular cylinder.

(c) Sphere.

(d) Circular disk.

(e) Cone Sphere.
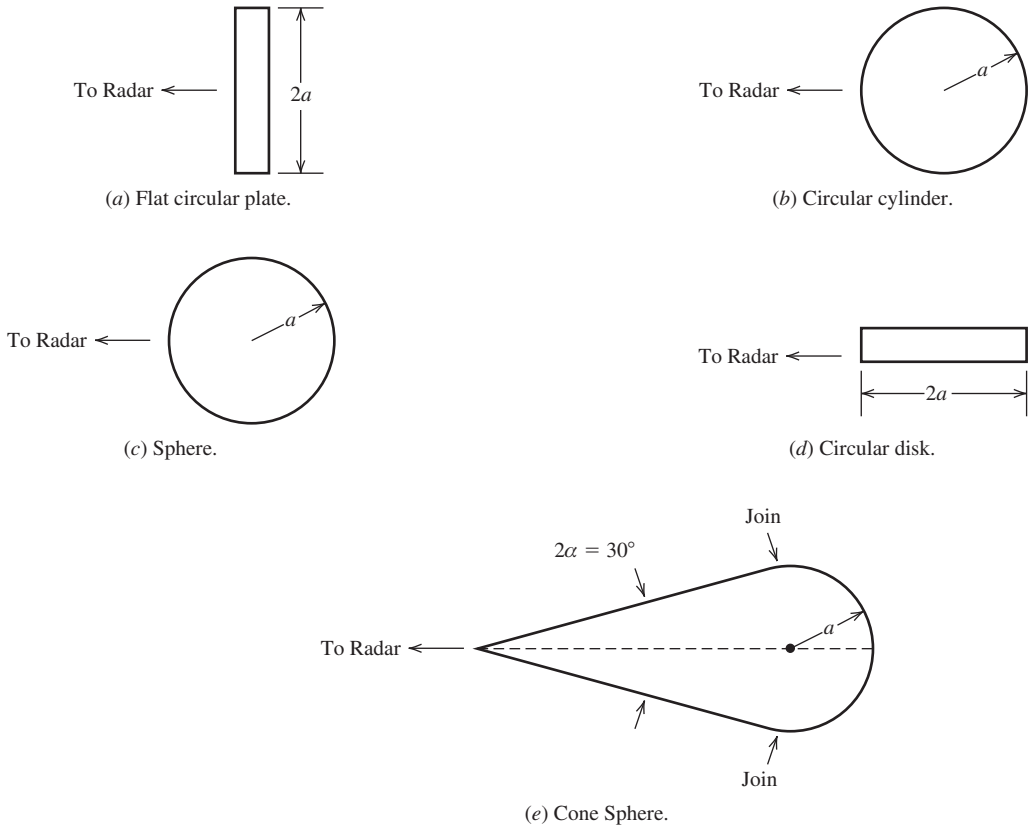
Figure 16-52 Side view of five simple scattering shapes.

each object except for the disk which is viewed normal to the curved edge in the plane containing the edge. The common geometrical feature among the shapes is that all use a radius of 0.5642 m, which is the radius that produces a cross section of 1 square meter, or 0 dBsm, for the sphere. Further, four of the shapes (FP, CYL, SPH, CS) project a physical area of 1 square meter $(1 \, \text{m}^2)$ in the monostatic radar direction. All five first-order radar cross sections can be predicted by physical optics if no higher-order effects such as creeping wave, multiple diffractions, and so on are included. The purpose here is to illustrate the dominant frequency dependence of some common geometrical features exclusive of higher-order effects.

The uppermost curve in Fig. 16-51 is that of the flat circular plate. The wavelength dependence is λ^{-2}, as it is for the rectangular plate in Prob. 16.13-1. The flat plate has two infinite radii of curvature, the $n = -2$ case.

The next curve is that of a finite length, L, circular cylinder viewed normal to the axis of the cylinder as indicated in Fig. 16-48 for the infinite length circular cylinder. The projected area of the finite length circular cylinder viewed normal to the curved surface is $2aL = (2)(0.5642)(0.8862) = 1 \, \text{m}^2$, and the RCS is inversely proportional to wavelength. The level of the curve is influenced by the length to diameter ratio of the cylinder, that is by the radius of the cylinder. The curved surface of the circular cylinder has one infinite radius of curvature and one finite and nonzero, the $n = -1$ case.

The third curve applies to the RCS of a sphere and also to the RCS of a finite length straight edge. The scattering by the sphere is mainly from the specular point, as indicated in Fig. 16-6. The sphere has no infinite radii of curvature, the $n = 0$ case. The zero wavelength dependence of a straight finite edge can be seen by using (5-6) in the RCS definition. The straight finite edge is also the $n = 0$ case.

The fourth curve is the RCS of the flat circular plate viewed edge-on [33] as shown in Fig. 16-52d. This is the case of a non-planar edge, which has one nonzero finite radius of curvature and the RCS is proportional to wavelength, the $n = 1$ case.

The fifth curve is the RCS of a cone-sphere wherein more than half of a sphere is "mated" to a cone such that there is the smoothest possible join between the cone and the sphere, as illustrated in Fig. 16-52e. There are three backscattering contributors. The incident wave first encounters the tip of the cone. Then the incident wave encounters the join between the cone and the spherical portion. Third, the incident wave excites a creeping wave on the spherical surface, some which is capable of creeping around the back side of the cone-sphere and shedding energy in the backscatter direction. This could produce an oscillation in the cone-sphere curve in Fig. 16-51 at the lower frequencies just as it does for the sphere in Fig. 16-7, if it were included here. The tip scattering is weak, leaving the join between the cone and the sphere as the dominant contributor to the backscattered RCS, the $n = 2$ case. Intuitively it might seem that the join between the cone and the sphere, being as smooth as possible, would not be a significant scattering source. However, if one looks at the cone sphere profile in Fig. 16-52e, there is a discontinuity in the derivative of the slope at the cone-sphere join, and this is the cause of the scattering [2]. An approximate physical optics formula for the join scattering is

$$\sigma = \frac{\lambda^2}{16\pi} \sec^4 \alpha \qquad (16\text{-}174)$$

where α is the cone half-angle. An approximate physical optics formula for tip scattering (see Prob. 16.13-3) is

$$\sigma = \frac{\lambda^2}{16\pi} \tan^4 \alpha \qquad (16\text{-}175)$$

Clearly the scattering from the join is much greater than the tip scattering. If the tip is rounded instead of pointed, the tip scattering increases markedly. The cone-sphere curve does not include creeping wave effects.

The scattering contributors for the five elementary shapes are not the only sources of scattering in general, but are representative of the major first-order sources. Not all have been treated analytically in this chapter. Only first-order scattering has been considered in this section. Contributions by higher-order effects such as multiple diffractions and creeping waves have not been included. For additional discussion of scattering sources, refer to the literature [4, 34].

The last section in this text examines the physical theory of diffraction (PTD), a combination of physical optics and currents near edges. To investigate PTD, we need the method of stationary phase, which is the subject of the next section.

16.16   METHOD OF STATIONARY PHASE

As we have seen many times earlier in this book, integrals describing radiation have integrands that consist of an amplitude function times a phase function. In many cases, an asymptotic evaluation is possible if the amplitude function is slowly varying and the exponential function is rapidly varying.

Consider the integral

$$I = \int_a^b f(x) e^{j\beta\gamma(x)} \ dx \qquad (16\text{-}176)$$

in which $f(x)$ and $\gamma(x)$ are real functions. If $f(x)$ is slowly varying and $\beta\gamma(x)$ is a rapidly varying function over the interval of integration due to β being large, the major contribution from the integral comes from the point or points of stationary phase. A point of stationary phase is defined as a point where the first derivative of the phase function γ vanishes:

$$\frac{d\gamma}{dx} = 0, \quad \text{at} \quad x = x_o \tag{16-177}$$

To expand the phase function in a Taylor series about the point of stationary phase, write

$$\gamma(x) \simeq \gamma_o + (x - x_o)\gamma_o' + (x - x_o)^2 \gamma_o''/2 + \cdots \tag{16-178}$$

where γ_o' and γ_o'' represent the derivatives of γ with respect to x, evaluated at x_o. Now γ_o' is zero by definition [see (16-177)], and in the neighborhood of the point of stationary phase the quantity $(x - x_o)$ is small so that the high-order terms (i.e., order 3 and higher) indicated in (16-178) may be neglected. If there is one and only one stationary point x_o in the interval from a to b, and x_o is not near either a or b, (16-176) thus becomes

$$I = \int_{x_o-\delta}^{x_o+\delta} f(x)e^{j\beta\gamma_o} e^{j\beta(x-x_o)^2 \gamma_o''/2} \; dx \quad (\text{if } \gamma_o'' \neq 0) \tag{16-179}$$

where δ represents a small number. Thus, the range of integration has been reduced to a small neighborhood about the point of stationary phase. If $f(x)$ is slowly varying, it may be approximated by $f(x_o)$ over this small interval. Thus, (16-179) becomes

$$I_o = f(x_o)e^{j\beta\gamma_o} \int_{-\infty}^{\infty} e^{j\beta(x-x_o)^2 \gamma_o''/2} \; dx = f(x_o)e^{j\beta\gamma_o} \int_{-\infty}^{\infty} e^{j\beta z^2 \gamma_o''/2} \; dz \tag{16-180}$$

where $(x - x_o) = z$. For convenience, the limits of integration have been changed again, this time to infinity. This introduces a little error if the chief contribution to the integral comes from the neighborhood of the point of stationary phase. In other regions, the rapid phase variations cause the contribution from one half-period to be nearly canceled by that from the next half-period if the amplitude $f(x)$ is constant or varies slowly.

Now consider the integral

$$\int_{-\infty}^{\infty} e^{jaz^2} \; dz = \int_{-\infty}^{\infty} (\cos az^2 + j \sin az^2) \; dz = \sqrt{\frac{\pi}{|a|}} \, e^{j(\pi/4) \, \text{sgn}(a)} \tag{16-181}$$

where

$$\text{sgn}(a) = \begin{cases} 1 & \text{if} \quad a > 0 \\ -1 & \text{if} \quad a < 0 \end{cases}$$

If we use (16-181) to evaluate (16-180), the stationary phase approximation is

$$\int_{x_o-\delta}^{x_o+\delta} f(x)e^{j\beta\gamma(x)} \; dx \approx f(x_o) \; e^{j\beta\gamma(x_o)} \sqrt{\frac{2\pi}{\beta|\gamma_o''|}} e^{j(\pi/4) \, \text{sgn}(\gamma_o'')} \tag{16-182}$$

If two or more points of stationary phase exist in the interval of integration (a to b) and there is no coupling between them, the total value of the integral is obtained by summing the contributions from each such point as given in (16-182).

Eq. (16-182) is not valid if the second derivative of the phase function vanishes at the point of stationary phase. In this event, $\gamma_o'' = 0$ and it is necessary to retain the third-order term in the Taylor series in (16-178).

Eq. (16-182) also becomes invalid if one of the limits of integration, a or b, is close to the point of stationary phase x_o. In this event, however, it is possible to express the integral in the form of a Fresnel integral as discussed later. A problem also arises if there exist two or more stationary points close together in the range of integration.

To obtain the endpoint contribution, it is best to write (16-176) as

$$I = \int_{-\infty}^{\infty} f(x)e^{j\beta\gamma(x)} \ dx - \int_b^{\infty} f(x)e^{j\beta\gamma(x)} \ dx - \int_{-a}^{\infty} f(-x)e^{j\beta\gamma(-x)} \ dx \qquad (16\text{-}183)$$

or

$$I = I_o - I_b - I_{-a} \qquad (16\text{-}184)$$

The evaluation of I_o has been done in (16-182). I_b, for instance, can be evaluated via integration by parts (see Prob. 12.14-1) by allowing the wave number to be complex and to have a small amount of loss (i.e., small α) so that the contribution to the integral by the upper limit at infinity vanishes, and then letting the wave number be approximated by β, as before. Thus,

$$I_b \cong -\frac{1}{j\beta} \frac{f(b)}{\gamma'(b)} \ e^{j\beta\gamma(b)} \qquad (16\text{-}185)$$

A similar expression can be found for I_{-a}. Eq. (16-185) is valid when b is not near (or coupled) to x_o. When the stationary point is coupled to the endpoint, we have [H.10.4: James]

$$I_b \cong U(-\varepsilon_1)I_o + \varepsilon_1 f(b)e^{j\beta\gamma(b)\mp jv^2}\sqrt{\frac{2}{\beta|\gamma''(b)|}}F_{\pm}(v) \qquad (16\text{-}186)$$

where $\gamma''(b) \neq 0$, $\varepsilon_1 = \mathrm{sgn}(b - x_o)$, $v = \sqrt{\frac{\beta}{2|\gamma''(b)|}}|\gamma'(b)|$, $U =$ unit step function, and $F_{\pm}(v)$ is the Fresnel integral.

There also are formulas for the stationary phase evaluation of double integrals [2]. However, (16-182) and (16-185) are sufficient to develop the physical theory of diffraction in the next section.

EXAMPLE 16-6   *Echo Width of a Circular Cylinder*

Consider the radar echo width of an infinite circular cylinder about the z-axis (see Fig. 16-53). We employ (16-182) to do this; (16-185) is used in the next section. Starting with the two-dimensional counterpart to (16-182), write for ψ in the cylindrical system:

$$\psi = \frac{1}{4j}H_0^{(2)}(\beta\rho) \cong \frac{1}{4j}\sqrt{\frac{2j}{\pi\beta|\boldsymbol{\rho}-\boldsymbol{\rho}'|}}e^{-j\beta|\boldsymbol{\rho}-\boldsymbol{\rho}'|} \qquad (16\text{-}187)$$
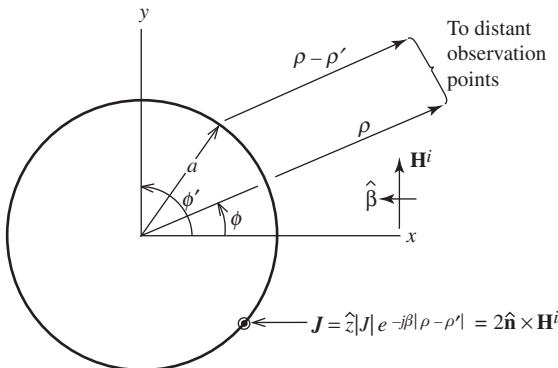
Figure 16-53 Geometry for radiation by a circular cylinder.

where $|\boldsymbol{\rho} - \boldsymbol{\rho}'| = \rho - \rho' \cos(\phi - \phi')$ and the asymptotic or large argument approximation for the Hankel function, $H_0^{(2)}(\beta\rho)$, of the second kind and zeroth order has been used. For amplitude purposes in the far field, $|\boldsymbol{\rho} - \boldsymbol{\rho}'| \approx \rho$. Thus, similar to (16-183), we obtain

$$\boldsymbol{\nabla}\psi = \hat{\boldsymbol{\rho}}\frac{1}{4j}\sqrt{\frac{2j}{\pi\beta}}\left[-\frac{1/2 + j\beta\rho}{\rho^{3/2}}\right]e^{-j\beta|\rho-\rho'|} \tag{16-188}$$

In the far field, the $\frac{1}{2}$ is insignificant compared to $\beta\rho$. Thus,

$$\mathbf{H}^s \approx \sqrt{\frac{j}{8\pi\beta}}\int(\mathbf{J}\times\hat{\boldsymbol{\rho}})\left[j\frac{\beta}{\rho^{1/2}}\right]e^{-j\beta|\rho-\rho'|}\ dc' \tag{16-189}$$

where dc' is an incremental element on a circumferential line of the cylinder, $dc' = a\,d\phi'$. Applying the definition of echo width (see Sec. 16.11) and taking $\mathbf{J} = 2\hat{\mathbf{n}}\times\mathbf{H}^i$ give

$$\sigma_w = \frac{\beta}{4}\frac{1}{|\mathbf{H}^i|^2}\left|\int((2\hat{\mathbf{n}}\times\mathbf{H}^i)\times\hat{\boldsymbol{\rho}})e^{-j\beta|\rho-\rho'|}\ dc'\right|^2 \tag{16-190}$$

The phase of the current on the cylinder is $e^{+j\beta a \cos(\phi-\phi')}$ (i.e., advanced relative to the origin as indicated in Fig. 16-46). Thus,

$$\sigma_w = \beta\left|\int e^{j\beta a \cos(\phi-\phi')}(\hat{\boldsymbol{\rho}}\cdot\hat{\mathbf{n}})e^{-j\beta|\rho-\rho'|}\ dc'\right|^2 \tag{16-191}$$

Noting that $(\hat{\boldsymbol{\rho}}\cdot\hat{\mathbf{n}}) = \cos\phi'$ and $|\boldsymbol{\rho} - \boldsymbol{\rho}'| = \rho - a\,\cos\phi'$ when $\rho' = a$ and $\phi = 0$ yields

$$\sigma_w = \beta\left|e^{-j\beta\rho}\int_{-\pi/2}^{\pi/2}\cos\phi'\,e^{+j2\beta a \cos\phi'}\,a\,d\phi'\right|^2 \tag{16-192}$$

The integral can be evaluated in a straightforward manner by the method of stationary phase. Using (16-182), identify $f'(\phi') = a\,\cos\phi'$ and $\gamma(\phi') = -2a\,\cos\phi'$. To find the stationary point(s) ϕ_o, use $f'(\phi') = 0 = -a\,\sin\phi'$ and determine that $\phi'_o = 0,\ \pi$. Due to the physical optics assumption of no current at $\phi'_o = \pi$, the value of π is discarded. Therefore, $f(\phi'_o) = a\,\cos(0°) = a$. Since $\gamma''(\phi') = 2a\,\cos\phi'$, then $\gamma''(\phi'_o) = 2a$. Therefore,

$$\sigma_w = \beta\left|ae^{-j\beta\rho}e^{j\beta a}\sqrt{\frac{2\pi}{\beta 2a}}e^{j\pi/4\,\mathrm{sgn}\,(\gamma''_o)}\right|^2 = \pi a \tag{16-193}$$

Thus, our stationary phase evaluation has produced the same result for the echo width as we obtained in Sec. 16.11 using geometrical optics.

It is interesting to compare the treatment here for the cylinder with the treatment of the sphere in the previous section. In the case of the sphere, the projection of the currents onto a plane was integrated. This produced an integral that could be evaluated in closed form. Had the sphere problem been formulated in a manner similar to that used here for the cylinder by integrating on the actual surface, it would have been necessary to use stationary phase for double integrals. The single integral treatment is, however, sufficient for our development of PTD in the next section.

16.17 PHYSICAL THEORY OF DIFFRACTION

The physical theory of diffraction (PTD) is an extension of physical optics (PO) that refines the PO surface field approximation just as GTD or UTD refines the geometrical optics surface-field approximation. The original PTD formulation was developed by Ufimtsev [35] for surfaces with perfectly conducting edges. Ufimtsev's original work was done at about the same time as the ray-optical work of Keller and independently of Keller.

However, Ufimtsev was aware of the work of Sommerfeld [7] and Pauli [9] and used the asymptotic form in (16-42) as did Keller.

In his work, Ufimtsev postulated that there was a nonuniform component of the current that would include effects not accounted for in the physical optics current, called the uniform current in his work (see Fig. 16-54). Ufimtsev did not actually find the nonuniform current for the wedge, but instead found the field due to the nonuniform component of the current by indirect means. (More recently, expressions for these nonuniform currents have been found [36, 37].) He found the non-uniform current contribution to the field by subtracting the PO field from the known total field for the wedge. The result was the field due to what was left, that from the nonuniform current. To see this, write

$$\mathbf{E}_{\text{total}}^s = \mathbf{E}_{\text{GO}}^r + \mathbf{E}_K^d \tag{16-194}$$

where

$\mathbf{E}_{\text{total}}^s$ = the total scattered field

$\mathbf{E}_{\text{GO}}^r$ = the reflected field obtained by geometrical optics

$\mathbf{E}_K^d$ = the diffracted field found using the Keller diffraction coefficient in (12-59)

Then write $\mathbf{E}_{\text{PO}}$ for the field due to the physical optics current, as

$$\mathbf{E}_{\text{PO}} \simeq \mathbf{E}_{\text{PO}}^r + \mathbf{E}_{\text{PO}}^d \tag{16-195}$$

where $\mathbf{E}_{\text{PO}}^d$ is called the physical optics diffracted field (which is not the total diffracted field) and is due to the abrupt termination of the physical optics current at, for example, the edge of the half-plane in Fig. 16-54. $\mathbf{E}_{\text{PO}}^r$ is the reflected field obtained by integration of the currents and is theoretically equal to $\mathbf{E}_{\text{GO}}^r$ since both represent the reflected field, although by different means.

Subtracting (16-195) from (16-194) or ($\mathbf{E}_{\text{total}}^s - \mathbf{E}_{\text{PO}}$) yields the field due to the nonuniform current $\mathbf{E}^{\text{nu}}$ as

$$\mathbf{E}^{\text{nu}} = \mathbf{E}_K^d - \mathbf{E}_{\text{PO}}^d \tag{16-196}$$

The field due to the nonuniform current, when added to the field due to the uniform current (e.g., the specular contribution), gives the total scattered field:

$$\mathbf{E}_{\text{total}}^s = \mathbf{E}^{\text{unif}} + \mathbf{E}^{\text{nu}} \tag{16-197}$$

If we take $\mathbf{E}_{\text{GO}}^r = \mathbf{E}_{\text{PO}}^r$, it follows from (16-195) that

$$\mathbf{E}_{\text{GO}}^r = \mathbf{E}_{\text{PO}} - \mathbf{E}_{\text{PO}}^d \tag{16-198}$$
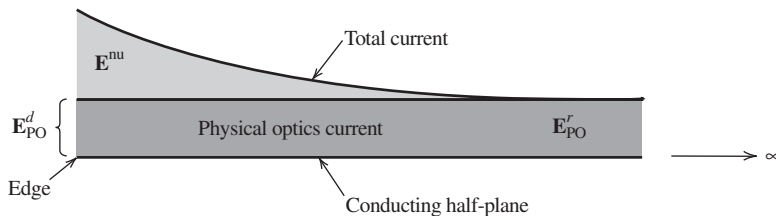
Figure 16-54 Conceptualization of PTD currents. Also indicated are the fields generated by those currents.

Substituting (16-198) into (16-194) gives

$$\mathbf{E}_{\text{total}}^s = \mathbf{E}_{\text{PO}} + (\mathbf{E}_K^d - \mathbf{E}_{\text{PO}}^d) = \mathbf{E}^{\text{unif}} + (\mathbf{E}^{\text{nu}}) \tag{16-199}$$

Next, we present expressions for the fields due to the nonuniform current. As in GTD, each polarization is considered separately. Thus for the nonuniform contribution in the two-dimensional case, the parallel and perpendicular polarizations (where $\parallel$ and $\perp$ refer to the orientation of the *electric* field) are

$$E_{\parallel}^{\text{nu}} = E^i f' \frac{e^{j(\beta\rho + \pi/4)}}{\sqrt{2\pi\beta\rho}} \tag{16-200}$$

$$H_{\perp}^{\text{nu}} = H^i g' \frac{e^{j(\beta\rho + \pi/4)}}{\sqrt{2\pi\beta\rho}} \tag{16-201}$$

Thus, $E_{\parallel}^{\text{nu}}$ and $H_{\perp}^{\text{nu}}$ are known in terms of two simple functions f' and g', where $f' = f - f_o$ and $g' = g - g_o$ and

$$\left\{\begin{matrix} f \\ g \end{matrix}\right\} = \frac{1}{n} \sin\frac{\pi}{n} \left[\frac{1}{\cos\dfrac{\pi}{n} - \cos\dfrac{\phi - \phi'}{n}} \right] \mp \left[\frac{1}{\cos\dfrac{\pi}{n} - \cos\dfrac{\phi + \phi'}{n}} \right] \tag{16-202}$$

$$f_o = \begin{cases} f_a & 0 < \phi' \le \pi - \phi_{\text{int}} \\ f_a + f_b & \pi - \phi_{\text{int}} \le \phi' \le \pi \\ f_b & \pi < \phi' < 2\pi - \phi_{\text{int}} \end{cases} \tag{16-203}$$

$$g_o = \begin{cases} g_a & 0 < \phi' \le \pi - \phi_{\text{int}} \\ g_a + g_b & \pi - \phi_{\text{int}} \le \phi' \le \pi \\ g_b & \pi < \phi' < 2\pi - \phi_{\text{int}} \end{cases} \tag{16-204}$$

$$\left\{\begin{matrix} f_a \\ g_a \end{matrix}\right\} = \left\{\begin{matrix} \sin\phi' \\ -\sin\phi \end{matrix}\right\} \frac{1}{\cos\phi + \cos\phi'} \tag{16-205}$$

$$\left\{\begin{matrix} f_b \\ g_b \end{matrix}\right\} = \left\{\begin{matrix} \sin(2\pi - \phi_{\text{int}} - \phi') \\ -\sin(2\pi - \phi_{\text{int}} - \phi) \end{matrix}\right\}$$
$$\times \frac{1}{\cos(2\pi - \phi_{\text{int}} - \phi) + \cos(2\pi - \phi_{\text{int}} - \phi')} \tag{16-206}$$

with $\phi_{\text{int}} = (2 - n)\pi$, the interior wedge angle. The a subscript denotes that the A face is illuminated and b denotes that the B face is illuminated as in Fig. 16-55.

Clearly, f and g correspond to the Keller GTD diffraction coefficients. Even though we know that f and g tend to infinity at the reflection and shadow boundaries, f' and g' do not because the singularities in f and g are canceled by identical singularities in f_o and g_o. The quantities f_o and g_o are the PO diffraction coefficients.

The expressions for f_o and g_o are obtained from the stationary phase endpoint contribution in (16-185). To demonstrate this [38], consider the wedge in Fig. 16-55 for the perpendicular (TE) case with only face A illuminated. Write for the incident magnetic field [see Eq. (16-37)]

$$\mathbf{H}^i = \hat{\mathbf{z}}\, H_o e^{j\beta\rho \cos(\phi - \phi')} \tag{16-207}$$
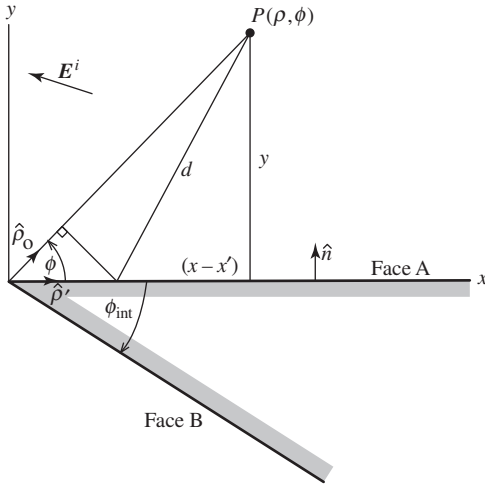
Figure 16-55 TE plane wave incidence on the A face of a wedge.

The physical optics current on the illuminated face of the wedge, face A in Fig. 16-55, is given by

$$\mathbf{J}_{PO} = 2\hat{\mathbf{n}} \times \mathbf{H}^i\big|_{\rho=x',\phi=0}$$

$$= 2(\hat{\mathbf{y}} \times \hat{\mathbf{z}})H^i e^{j\beta x' \cos \phi'} = 2\hat{\mathbf{x}} \ H_o e^{j\beta x' \cos \phi'} \tag{16-208}$$

The vector potential for the two-dimensional wedge is

$$\mathbf{A} = \mu \int_0^\infty \mathbf{J}_{PO} \frac{1}{4j} H_0^{(2)}(\beta|\boldsymbol{\rho}_o - \boldsymbol{\rho}'|) \ dx' \tag{16-209}$$

Using the asymptotic expression of the Hankel function gives

$$\mathbf{A} \simeq \mu \frac{2H^i}{4j}\hat{\mathbf{x}} \int_0^\infty e^{j\beta x' \cos \phi'} \sqrt{\frac{2j}{\pi\beta|\boldsymbol{\rho}_o - \boldsymbol{\rho}'|}} e^{-j\beta|\boldsymbol{\rho}_o - \boldsymbol{\rho}'|} \ dx' \tag{16-210}$$

From Fig. 16-55, note that $|\boldsymbol{\rho}_o - \boldsymbol{\rho}'| \cong \rho_o - \hat{\boldsymbol{\rho}} \cdot \boldsymbol{\rho}' = \rho_o - x' \cos \phi$. Thus, $\mathbf{A}$ becomes, with the usual far-field approximations,

$$\mathbf{A} \cong \mu \frac{H^i}{2j}\hat{\mathbf{x}} \sqrt{\frac{2j}{\pi\beta\rho_o}} \int_0^\infty e^{-j\beta\rho_o} e^{j\beta x'(\cos \phi' + \cos \phi)} \ dx' \tag{16-211}$$

Considering just the integral with $d \approx \rho_o - x' \cos \phi$, we get

$$I = \int_0^\infty e^{j\beta(x' \cos \phi' - d)} dx' \cong I_o + I_b \tag{16-212}$$

where I_o is the stationary contribution discussed in the previous section and I_b is the endpoint contribution in Eq. (16-185). Evaluation of I_o is not of immediate interest to us here and is left as an exercise for the reader. Our attention is turned to I_b, where

$$I_b = -\frac{1}{j\beta} \frac{f(0)}{\gamma'(0)} \ e^{j\beta\gamma(0)} \tag{16-213}$$

Here,

$$f(0) = 1 \tag{16-214a}$$

$$\gamma(x') = x' \cos \phi' - d \tag{16-214b}$$

$$d = \{(x - x')^2 + (y)^2\}^{1/2} \tag{16-214c}$$

$$\frac{d}{dx'}(d) = -\frac{(x - x)'}{d} \tag{16-214d}$$

$$\gamma'(x') = \cos \phi' - \frac{d}{dx'}(d) = \cos \phi' + \frac{x - x'}{d} \tag{16-214e}$$

$$\gamma'(0) = \cos \phi' + \frac{x}{\rho_o} = \cos \phi' + \cos \phi \tag{16-214f}$$

$$\gamma(0) = -d(0) = -\rho_o \tag{16-214g}$$

Thus, the endpoint contribution to the integral is

$$I_b = -\frac{1}{j\beta} \frac{1}{\cos \phi' + \cos \phi} e^{-j\beta \rho_o} \tag{16-215}$$

which means that the endpoint contribution $\mathbf{A}_\perp^{\text{ep}}$ is

$$\mathbf{A}_\perp^{\text{ep}} = -\mu \frac{H^i \hat{\mathbf{x}}}{\sqrt{2\pi j\beta \rho_o}} e^{j\beta \rho_o} \frac{-1}{j\beta} \frac{1}{\cos \phi' + \cos \phi} \tag{16-216}$$

Since $\mathbf{E}_\perp^{\text{ep}} = -j\omega \mathbf{A}_\perp^{\text{ep}}$ and $\eta H^i = E^i$,

$$\mathbf{E}_\perp^{\text{ep}} = E^i \frac{1}{\sqrt{2\pi j\beta \rho_o}} \frac{1}{\cos \phi + \cos \phi'} e^{-j\beta \rho_o} \hat{\mathbf{x}} \tag{16-217}$$

And finally for the magnetic field, we have

$$\mathbf{H}_\perp^{\text{ep}} = (\hat{\boldsymbol{\rho}}_o \times \hat{\mathbf{x}}) H^i \frac{1}{\sqrt{2\pi j\beta \rho_o}} \frac{1}{\cos \phi + \cos \phi'} e^{-j\beta \rho_o}$$

$$= -H^i \frac{1}{\sqrt{2\pi j\beta \rho_o}} \frac{\sin \phi}{\cos \phi + \cos \phi'} e^{-j\beta \rho_o} \hat{\mathbf{z}} = H_{\text{PO}}^d \, \hat{\mathbf{z}} \tag{16-218}$$

which is the postulated result given in (16-200) through (16-206).

Suppose that PTD is to be used to calculate the results in Fig. 16-14a. First, it must be kept in mind that PTD uses the equivalence principle, as does physical optics, where all conducting media are replaced with equivalent currents radiating in free space. This should be contrasted with GTD, where all material media are retained. Starting with region I (Fig. 16-12), one can write

$$\mathbf{E}_{\text{total}} \cong \mathbf{E}^i + \mathbf{E}^{\text{unif}} + \mathbf{E}^{\text{nu}} \tag{16-219}$$

or

$$\mathbf{E}_{\text{total}} = \mathbf{E}^i + (\mathbf{E}_{\text{PO}}) + (\mathbf{E}_K^d - \mathbf{E}_{\text{PO}}^d)$$

$$= \mathbf{E}^i + (\mathbf{E}_{\text{PO}}^r + \mathbf{E}_{\text{PO}}^d) + (\mathbf{E}_K^d - \mathbf{E}_{\text{PO}}^d) \tag{16-220}$$

which in terms of GTD is

$$\mathbf{E}_{\text{total}} = \mathbf{E}^i + \mathbf{E}^r_{\text{GO}} + \mathbf{E}^d_K \quad \text{in region I} \tag{16-221}$$

To apply PTD, (16-219a) is used, whereas the use of (16-219c) is the application of GTD. In the former, fields from currents are used, whereas in the latter, ray-optical fields are used. More specifically, to apply PTD as given in (16-219a) to the half-plane problem of Fig. 16-14, $\mathbf{E}^i$ is represented by (16-37), $\mathbf{E}^r_{\text{PO}}$ by (16-182), $\mathbf{E}^d_{\text{PO}}$ by (16-185), and $(\mathbf{E}^d_k - \mathbf{E}^d_{\text{PO}})$ by (16-201). Fig. 16-56 is helpful in understanding the PTD calculation in region I. As the reflection boundary is approached in Fig. 16-56, (16-186) must be used since the stationary point and endpoint become coupled.

For region II, as in Fig. 16-56, all the quantities in (16-219b) are also present, except $\mathbf{E}^r_{\text{PO}}$ that is absent outside the two transition regions. Here with PTD, $\mathbf{E}^{\text{unif}}$ is continuous across the reflected field shadow boundary. $\mathbf{E}^{\text{nu}}$ is also continuous. Thus, $\mathbf{E}_{\text{total}}$ is continuous across the reflected field shadow boundary. (See Prob. 16.17-2.)

With regard to Fig. 16-56, in region III all PTD quantities in (16-219a) are present just as in region II. Moving across the incident field shadow boundary, $\mathbf{E}^{\text{nu}}$ will, of course, be continuous. Both $\mathbf{E}^i$ and $\mathbf{E}^{\text{unif}}$ are also continuous and, therefore, $\mathbf{E}_{\text{total}}$ is continuous. Moving deeper into the shadow region, $\mathbf{E}^i$ will (theoretically) be canceled by $\mathbf{E}^r_{\text{PO}}$, leaving $\mathbf{E}^d_K$. This is not surprising because it is known from Sec. 16.2 that $\mathbf{E}^d_K$ gives the correct field in the deep shadow region. To calculate the field in PTD in region III, note from Fig. 16-56 that $\mathbf{E}^r_{\text{PO}}$ appears in the shadow region and (16-185) is used in its evaluation away from the shadow boundary, just as it was used in region I.

(a) Observation point in region I.

(b) Observation point in a transition region (e.g., near to the reflection boundary). Stationary point and end point are coupled.

(c) Observation point in region II, no stationary point.

(d) Observation point in region III. The incident field will be approximately canceled by the stationary point contribution.
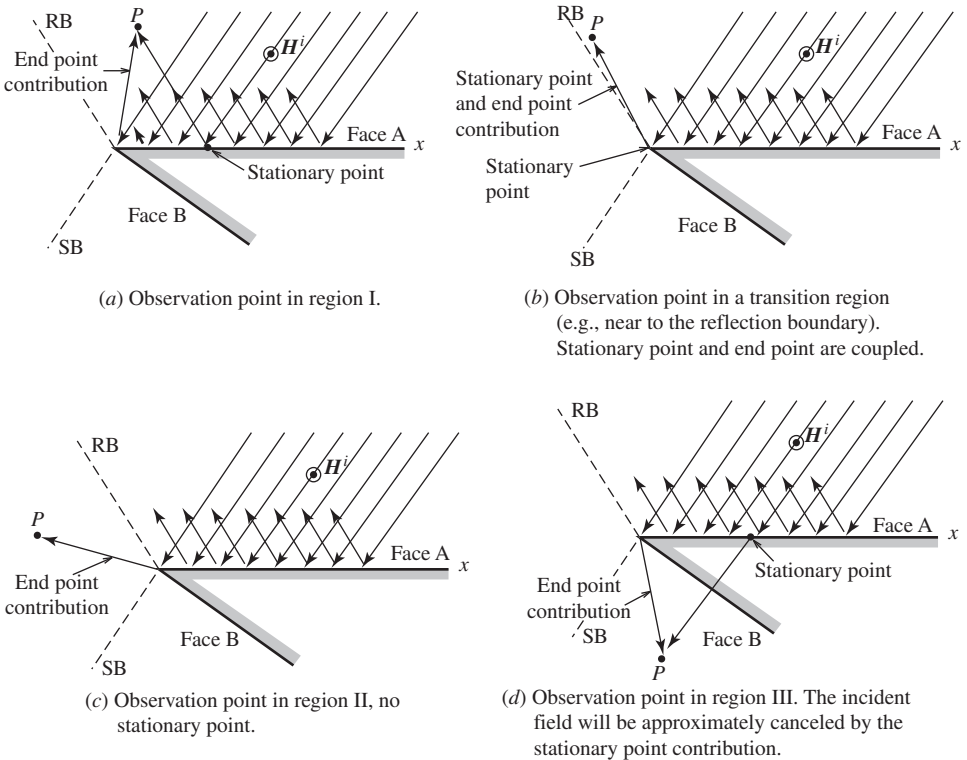
Figure 16-56 The relationship between stationary phase and geometrical optics for a conducting wedge.

PTD, just like GTD, applies only to scattering directions lying on the cone of diffracted rays shown in Fig. 16-15. This restriction for PTD may be overcome by using incremental length diffraction coefficients (ILDC) devised by Mitzner [39]. The ILDCs do for PTD what equivalent currents in Sec. 16.9 do for GTD [40].

16.18   CYLINDRICAL PARABOLIC REFLECTOR ANTENNAS—PTD

As a second example of the application of the PTD, consider again the cylindrical parabolic reflector antenna shown in Fig. 16-20. In the half-space where z is positive (region A in Fig. 16-57), we use aperture integration for $0 < \zeta < \pi/2$. This gives the contribution from $\mathbf{E}^{\text{unif}}$. The contribution from $\mathbf{E}^{\text{unif}}$ will increasingly disagree with the UTD result in Fig. 16-58 as ζ moves from 30 to 90°. This discrepancy may be removed by including $\mathbf{E}^{\text{nu}}$ in the calculation as Fig. 16-58 shows for the PTD case.

To obtain the field or pattern in the deep shadow region (the shaded part of region C in Fig. 16-57), we cannot use the field from the aperture integration. An examination of the equivalence principle shows that this is so. For example, from an examination of the half-plane example in the previous section, we know that in the deep shadow region the field may be found from just $\mathbf{E}_k^d$ if $\mathbf{E}^i$ is canceled by $\mathbf{E}_{\text{PO}}^r$. However, if $\mathbf{E}_{\text{PO}}^r$ is taken to be from the equivalent currents in the aperture, $\mathbf{E}_{\text{PO}}^r$ gives a collimated beam in the negative z-direction that clearly cannot be canceled by the field from the feed. However, if $\mathbf{E}_{\text{PO}}^r$ is taken to be the field from the currents on the parabolic surface, it will cancel $\mathbf{E}^i$ in the shaded region.
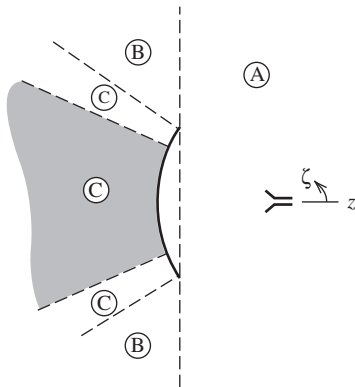
Figure 16-57 Regions of a cylindrical parabolic reflector antenna (see Fig. 16-20).
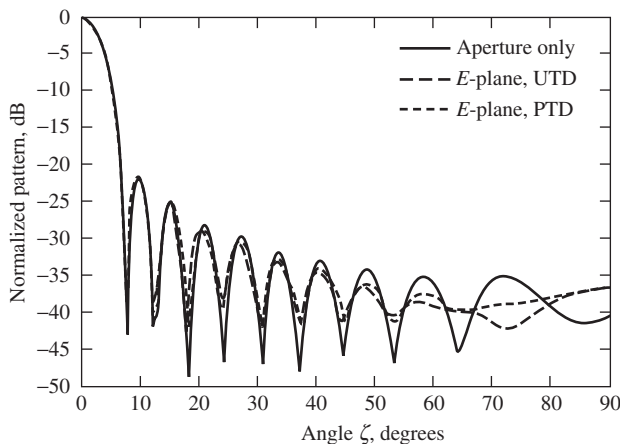
Figure 16-58 *E*-plane comparison of physical optics aperture integration with UTD and PTD in the side-lobe region for the geometry of Fig. 16-20.

To get the field in the unshaded part of regions C and B, we need to use $\mathbf{E}^{unif}$ and $\mathbf{E}^{nu}$, which means the integration over the parabolic surface itself must be done. Generally, it is easier to integrate over the aperture than over the parabolic surface, but that is not a valid option here. In obtaining the field in regions B and C, the need to integrate over the parabolic surface can be avoided by simply using the UTD diffraction coefficients in a GTD model rather than the Keller coefficients in a PTD model, realizing that UTD gives the correct fields without the singularity at the incident field shadow boundary. That is to say, in this problem the simpler model is a GTD model with UTD coefficients for regions C, B, and part of A.

Fig. 16-58 shows the E-plane radiation pattern for the cylindrical parabolic reflector antenna of Fig. 16-20. Fig. 16-58 shows a comparison of aperture integration, PTD, and UTD for the TE (or perpendicular) case. The agreement between classical aperture integration and the two asymptotic theories is excellent in the region of the main beam and the first few side lobes; thereafter, there is increasing disagreement because aperture integration does not fully account for edge diffraction effects that are increasingly important at larger angles.

Fig. 16-59a shows a comparison between UTD single diffraction, UTD double diffraction, and PTD for the full E-plane pattern. Aperture integration is used for the main

(a) Perpendicular case.

(b) Parallel case.
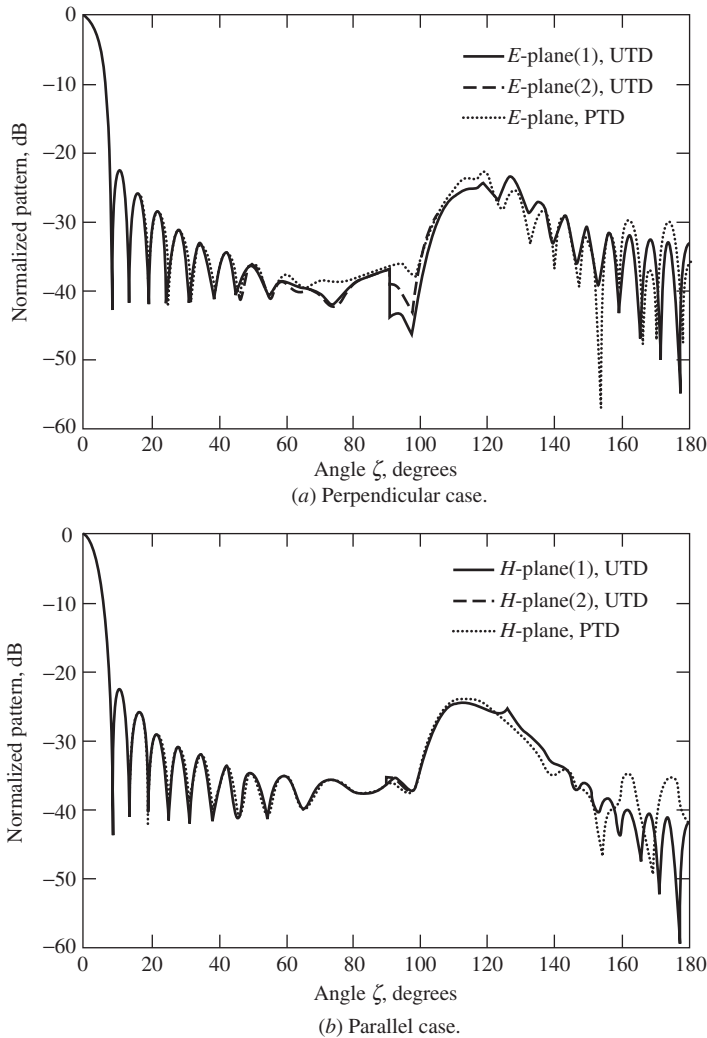
Figure 16-59 A comparison of PTD with UTD single (1) and double (2) diffraction calculations for the geometry of Fig. 16-20.

beam and the first side lobe in the UTD calculations; thereafter, starting in the second null, UTD alone is used. With two exceptions, the agreement is quite good between UTD and PTD. First, in the vicinity of 90°, there is some disagreement since edge diffraction effects are particularly strong for this polarization and one edge is in the transition region of the other (see Fig. 16-18c). Note, however, that the pattern is near the −40-dB level. Second, there is some disagreement in and near the back-lobe region. This disagreement is not due to a deficiency in the diffraction calculations, but the inability of E^{unif} from the integration over the currents on the parabolic surface to exactly cancel E^i from the feed in the PTD calculation. If the feed had less taper (e.g., $\cos\theta_s$ instead of $\cos^2\theta_s$), the diffracted field would be stronger and the incomplete cancellation of E^{unif} by E^i less apparent at the higher back-lobe level. For this geometry, disagreement in the back-lobe region is greater in the H-plane case than the E-plane case because the diffracted field in the former is weaker than in the latter, resulting in a more noticeable incomplete cancellation effect.

16.19   SUMMARY

In this chapter on high-frequency methods, a variety of techniques have been presented for predicting both the near- and far-zone fields from perfectly conducting bodies whose dimensions are large in terms of the wavelength. The GTD/UTD approach is ray-based and relatively simple when the number of rays is not large. The PTD approach is current-based and requires the integration of currents. Since integration is a smoothing process, a geometrical surface description in PTD does not have to be as accurate as in UTD. In both UTD and PTD, the most difficult part to calculate is usually the most basic part: GO in the case of UTD and PO in the case of PTD. An example of this is the calculation of scattering by an infinite cylinder in Sec. 16.11 where, for small radii, the creeping wave contribution was more accurate than that of GO.

The importance of the UTD method in antenna and scattering problems stems from the significant advantages to be gained from its use: (1) it is simple to use and yields accurate results; (2) it provides some physical insight into the radiation and scattering mechanisms involved; (3) it can be used to treat problems for which exact analytical solutions are not available. UTD is also used in acoustic problems such as SONAR and problems involving inhomogeneous or anistropic media [H.6: Felsen].

The importance of the PTD method is mostly in scattering problems. An advantage of the PO is that it provides scattering information in directions that are not necessarily in the specular direction or on the cone of diffracted rays. PTD is thought to have played a key part in the development of the B-2 stealth bomber.

The methods of this chapter tend to complement the intermediate frequency moment method techniques presented in Chap. 14 and the FDTD technique in Chap. 15. And, as was seen in Sec. 16.12, the moment method can be formally combined with UTD into a hybrid technique that extends the class of problems to which moment methods can be applied. This can be done not only because both MoM and UTD are highly practical techniques, but also because they are inherently flexible in their application to analysis and design problems. Outside the scope of this text are other more recently developed hybrid techniques. These include FEM/BEM, MoM/PO, FEM/PO, and others.

REFERENCES

1. M. Born and E. Wolf, *Principles of Optics*, Pergamon, New York, 1959.
2. G. T. Ruck, Ed., *Radar Cross Section Handbook*, Plenum, New York and London, 1970.
3. J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi, Eds., *Electromagnetic and Acoustic Scattering by Simple Shapes*, North-Holland, Amsterdam, 1969.
4. J. W. Crispin and K. M. Siegel, Eds., *Methods of Radar Cross-Section Analysis*, Academic Press, New York and London, 1968.

5. J. B. Keller, "Geometrical Theory of Diffraction," *J. Opt. Soc. Amer.*, Vol. 52, pp. 116–130, 1962.

6. L. M. Graves, Ed., "A Geometrical Theory of Diffraction," in *Calculus of Variations and Its Applications*, McGraw Hill, New York, pp. 27–52, 1958.

7. A. Sommerfeld, *Optics*, Academic Press, New York, 1954.

8. P. M. Russo, R. C. Rudduck, and L. Peters, Jr., "A Method for Computing *E*-plane Patterns of Horn Antennas," *IEEE Trans. on Ant. & Prop.*, Vol. AP-13, pp. 219–224, March 1965.

9. W. Pauli, "On Asymptotic Series for Functions in the Theory of Diffraction of Light," *Phys. Rev.*, Vol. 54, pp. 924–931, 1938.

10. R. G. Kouyoumjian and P. H. Pathak, "A Uniform Geometrical Theory of Diffraction for an Edge in a Perfectly Conducting Surface," *Proc. IEEE*, Vol. 62, pp. 1448–1461, 1974.

11. R. G. Kouyoumjian, "Asymptotic High Frequency Methods," *Proc. IEEE*, Vol. 53, pp. 864–876, 1965.

12. R. G. Kouyoumjian, "The Geometrical Theory of Diffraction and Its Application," in *Numerical and Asymptotic Techniques in Electromagnetics*, Springer-Verlag, New York, 1975.

13. M. Albani, et al., "UTD Vertex Diffraction Coefficient for the Scattering by Perfectly Conducting Faceted Structures," *IEEE Trans. on Ant. & Prop.*, Vol. 57, No. 12, pp. 3911–3925, Dec. 2009.

14. A. R. Lopez, "The Geometrical Theory of Diffraction Applied to Antenna and Impedance Calculations," *IEEE Trans. on Ant. & Prop.*, Vol. 14, pp. 40–45, Jan. 1966.

15. C. E. Ryan, Jr. and L. Peters, Jr., "Evaluation of Edge Diffracted Fields Including Equivalent Currents for the Caustic Regions," *IEEE Trans. Ant. & Prop.*, Vol. AP-17, pp. 292–299, March 1969. (See also correction in Vol. AP-18, p. 275, 1970.)

16. R. F. Harrington, *Time-Harmonic Electromagnetic Fields*, McGraw-Hill, New York, 1961.

17. N. Wang, "Self-consistent GTD Formulation for Conducting Cylinders with Arbitrary Convex Cross Section," *IEEE Trans. on Ant. & Prop.*, Vol. AP-24, pp. 463–468, July 1976.

18. M. Levy and J. B. Keller, "Diffraction by a Smooth Object," *Comm. Pure Appl. Math.*, Vol. 12, pp. 159–209, 1959.

19. P. H. Pathak and R. G. Kouyoumjian, "An Analysis of the Radiation from Apertures in Curved Surfaces by the Geometrical Theory of Diffraction," *Proc. IEEE*, Vol. 62, pp. 1438–1447, 1974.

20. D. R. Voltmer, "Diffraction of Doubly Curved Convex Surfaces," Ph.D. diss., Ohio State University, 1970.

21. L. W. Henderson and G. A. Thiele, "A Hybrid MM-GTD Technique for the Treatment of Wire Antennas Near a Curved Surface," *Radio Sci.*, Vol. 16, pp. 1125–1130, Nov.–Dec. 1981.

22. R. Luebbers, "Finite Conductivity Uniform GTD versus Knife Edge Diffraction in Prediction of Propagation Path Loss," *IEEE Trans. Ant. & Prop.*, Vol. 32, No. 1, pp. 70–76, Jan. 1984.

23. J. B. Andersen, "UTD Multiple-Edge Transition Zone Diffraction," *IEEE Trans. on Ant. & Prop.*, Vol. 45, No. 7, pp. 1093–1097, July 1997.

24. C. Tzaras and S. Saunders, "An Improved Heuristic UTD Solution for Multiple-Edge Transition Zone Diffraction," *IEEE Trans. on Ant. & Prop.*, Vol. 49, No. 12, pp. 1678–1682, Dec. 2001.

25. D. Erricolo and P. L. E. Uslenghi, "Two-Dimensional Simulator for Propagation in Urban Environments, "IEEE Trans. Veh. Tech., Vol. 50, No. 4, pp. 1158–1168, July 2001.

26. D. Erricolo, G. D'Elia, and P. L. E. Uslenghi, "Measurements on Scaled Models of Urban Environments and Comparisons with Ray-Tracing Propagation Simulation," *IEEE Trans. on Ant. & Prop.* Vol. 50, No. 5, pp. 727–735, May 2002.

27. T. Sarkar et al., "A Survey of Various Propagation Models for Mobile Communication," *IEEE Trans. on Ant. & Prop. Mag*, Vol. 45, No. 3, pp. 51–82, June 2003.

28. G. A. Thiele and T. H. Newhouse, "A Hybrid Technique for Combining Moment Methods with the Geometrical Theory of Diffraction," *IEEE Trans. on Ant. & Prop.*, Vol. AP-23, pp. 62–69, Jan. 1975.

29. G. A. Thiele, "Hybrid Methods in Antenna Analysis," *IEEE Proc.*, Vol. 80, No. 1, pp. 66–78, Jan. 1992.

30. J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi, "Electromagnetic and Acoustic Scattering by Simple Shapes," Hemisphere Publishing Corp, New York, 1987, pp. 593, 618.

31. G. A. Thiele, "Fundamentals of Radar Cross Section," short course notes, Southeastern Center for Electrical Engineering Education (SCEEE), St. Cloud, FL, 1981–1992.

32. J. W. Crispin, Jr. and A. L. Maffett, "Radar Cross Section Estimation for Simple Shapes," Proc. IEEE, Vol. 53, No. 8, pp. 833–848, August 1965.

33. T. B. A. Senior, "Disk Scattering at Edge-On Incidence," *IEEE Trans. on Ant. & Prop.*, Vol. AP-17, No. 6, Nov. 1969.

34. A. K. Bhattacharyya and D. L. Sengupta, "Radar Cross Section Analysis & Control," Artech House, Boston, 1991.

35. P. Ia Ufimtsev, "Approximate Computation of the Diffraction of Plane Electromagnetic Waves at Certain Metal Bodies. Part I. Diffraction Patterns at a Wedge and a Ribbon," *Zh. Tekhn. Fiz.* (USSR), Vol. 27, No. 8, pp. 1708–1718, 1957.

36. P. K. Murthy and G. A. Thiele, "Non-Uniform Currents on a Wedge Illuminated by a TE Plane Wave," *IEEE Trans. on Ant. & Prop.,* Vol. AP-34, pp. 1038–1045, Aug. 1986.

37. K. M. Pasala, "Closed-form Expressions for Nonuniform Currents on a Wedge Illuminated by TM Plane Wave," *IEEE Trans. on Ant. & Prop.*, Vol. 36, pp. 1753–1759, Dec. 1988.

38. K. M. Pasala, unpublished notes.

39. K. M. Mitzer, "Incremental Length Diffraction Coefficients," Tech. Rep. No. AFAL-TR-73-296, Northrop Corporation.

40. E. F. Knott, J. F. Shaeffer, and M. T. Tuley, *Radar Cross Section*, Artech House, Boston, 1985.

PROBLEMS

16.1-1 It can be shown [10] that the principal radii of curvature of the geometrical optics reflected wavefront are given by

$$\frac{1}{\rho_1} = \frac{1}{2}\left(\frac{1}{\rho_1^i}+\frac{1}{\rho_2^i}\right)+\frac{1}{f_1} \quad \text{and} \quad \frac{1}{\rho_2} = \frac{1}{2}\left(\frac{1}{\rho_1^i}+\frac{1}{\rho_2^i}\right)+\frac{1}{f_2}$$

where ρ_1^i and ρ_2^i are the principal radii of curvature of the incident wavefront and ρ_1 and ρ_2 are the principal radii of curvature of the reflected wavefront. General expressions for f_1 and f_2 are given in the literature [10]. However, for an incident spherical wave,

$$\frac{1}{f_{1,2}} = \frac{1}{\cos\theta_i}\left(\frac{\sin^2\theta_2}{r_1^c}+\frac{\sin^2\theta_1}{r_2^c}\right)$$
$$\pm \sqrt{\frac{1}{\cos^2\theta_i}\left(\frac{\sin^2\theta_2}{r_1^c}+\frac{\sin^2\theta_1}{r_2^c}\right)^2-\frac{4}{r_1^c r_2^c}}$$

where θ_1 and θ_2 are the angles between the incident ray and principal directions (i.e., tangent unit vectors) associated with the principal radii of curvature of the surface r_1^c and r_2^c, respectively.

(a) Show that for $\theta_1 = \theta_o$ and $\theta_2 = 90°$, the first equation reduces to (16-20) and the second to

$$\frac{1}{\rho_2} = \frac{1}{\ell_o}+\frac{2\,\cos\theta_o}{r_2^c}$$

(b) Without using (16-20) or the expression for ρ_2 immediately above, show that in the case of plane wave illumination

$$\sqrt{\rho_1\rho_2}=\tfrac{1}{2}\sqrt{r_1^c r_2^c}$$

16.1-2 An infinite elliptical paraboloid is described by the equation

$$\frac{x^2}{2r_1^c}+\frac{y^2}{2r_2^c} = -z$$

where r_1 and r_2 are the principal radii of curvature at the specular point. Using geometrical optics, show that the radar cross section for axial incidence is

$$\sigma = \pi r_1^c r_2^c$$

Actually, this result applies to any surface expressible in terms of a second-degree polynomial, where r_1^c and r_2^c are the principal radii of curvature at the reflection point [2, 3]. Is the above result valid for a cylindrical surface or flat plate? Why not?

16.1-3 A plane wave is incident on a smooth three-dimensional conducting convex body. The two principal radii of curvature of the body at the specular point are $r_1^c = 5\lambda$ and $r_2^c = 10\lambda$. Write expressions for the electric and magnetic backscattered fields if the incident plane wave fields are

$$\mathbf{E}^i = \hat{\mathbf{y}} e^{-j\beta x} \qquad \text{and} \qquad \mathbf{H}^i = \hat{\mathbf{z}} \, \frac{e^{-j\beta x}}{\eta}$$

16.1-4 (a) Determine the monostatic RCS of a convex infinite paraboloid of revolution for on axis incidence if the radii of curvature at the specular point are given by $r^c = 2F$, where F is the focal length of the paraboloid [H.3: Milligan, 2nd ed., p. 382].

(b) Under what conditions would the result of (b) hold for the monostatic RCS of a circularly symmetric parabolic reflector for on axis incidence upon its *concave* surface?

16.2-1 A cylindrical wave is incident on a cylindrical parabolic reflector as shown. To obtain the diffracted field from the top edge (only) at any point in space, the edge may be analyzed as if a half-plane were tangent to the uppermost portion of the parabolic surface. Divide the space around the top edge into three separate regions and write general expressions (with numerical values for ϕ') for the total electric field from the top edge in those three regions of space. In which of the three regions is the *total* geometrical optics field zero?
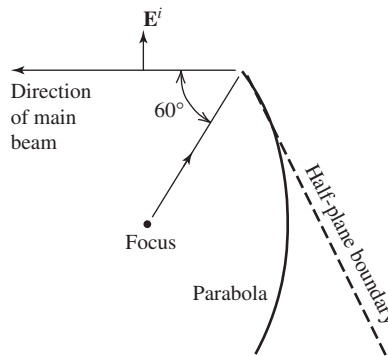
16.2-2 Evaluate the following Fresnel integrals:

(a) $\displaystyle\int_0^\infty e^{-j\tau^2} \, d\tau$

(b) $\displaystyle\int_0^5 e^{-j\tau^2} \, d\tau$

(c) $\displaystyle\int_5^\infty e^{-j\tau^2} \, d\tau$

16.2-3 Find $v_B(\rho, \phi^\pm)$, using both (16-42) and (16-44) for a 90° interior angle wedge when:
  (a) $\phi' = 45°$, $\rho = 10\lambda$, $\phi = 220°$
  (b) $\phi' = 45°$, $\rho = 10\lambda$, $\phi = 230°$
Compare your results in (a) and (b) and explain any differences. What is v_* in parts (a) and (b)?

16.2-4 Find $v_B(\rho, \phi^\pm)$ for a 90° interior wedge angle (both polarizations) when:
  (a) $\phi' = 45°$, $\rho = 10\lambda$, $\phi = 90°$
  (b) $\phi' = 45°$, $\rho = 10\lambda$, $\phi = 138°$
  (c) $\phi' = 45°$, $\rho = 10\lambda$, $\phi = 180°$
Comment on your results and justify the formulas you used to evaluate the diffracted field in each case.

16.2-5 Substitute (16-34) and (16-35) into (16-33) and explain the physical significance of each of the four terms you obtain.

16.2-6 Draw a sketch that illustrates the first postulate of Keller's theory. Include both a direct ray and a diffracted ray in your sketch.

16.3-1 Consider a magnetic line source parallel to the edge of a half-plane as shown. In this situation, the diffracted field appears to originate from a magnetic line source located at the edge. Using the flux tube concept of Fig. 16-3, show that the diffracted field may be written as

$$E_\perp^d(\rho) = -D_\perp E_\perp^i(Q)\frac{e^{-j\beta\rho}}{\sqrt{\rho}}$$

where $E_\perp^i(Q)$ is the value of the incident field at the edge.
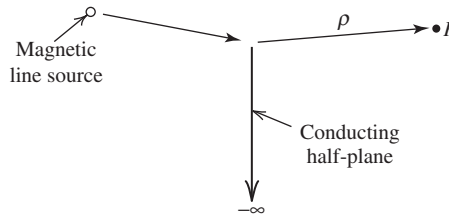
16.3-2 Repeat Prob. 16.3-1 when the magnetic line source is replaced by an electric line source and show that

$$E_\parallel^d(\rho) = -D_\parallel E_\parallel^i(Q)\frac{e^{-j\beta\rho}}{\sqrt{\rho}}$$

16.3-3 Consider the situation where a point source illuminates the edge of a half-plane at normal incidence. Unlike the previous two problems, in this case there will be spreading in both principal planes. Using the flux tube concept of Fig. 16-3, show that the diffracted field may be written as either

$$E_\parallel^d(s) = -D_\parallel E_\parallel^i(Q)\sqrt{\frac{s'}{s'+s}}\frac{e^{-j\beta s}}{\sqrt{s}}$$

or

$$E_\perp^d(s) = -D_\perp E_\perp^i(Q)\sqrt{\frac{s'}{s'+s}}\frac{e^{-j\beta s}}{\sqrt{s}}$$

16.3-4 Show that the diffraction coefficient matrix $[D]$ in (16-46) will generally have seven nonvanishing coefficients if an edge-fixed coordinate system is used rather than a ray-fixed system.

16.4-1 Derive (16-59) from (16-52) and show that (16-59) is the same as (16-42).

16.4-2 Consider the case where a half-plane is illuminated by a plane wave and the observation point is near the edge of the wedge.
  (a) Show that UTD reduces to the Sommerfeld-Pauli result in (16-44) and hence the UTD is exact.
  (b) Is UTD an exact solution if the source is near the wedge edge and the observation point is at a very large distance? Why?
  (c) If both the source and observation points are near the wedge edge, the UTD solution will not be exact. Why? (Although the solution may not be exact, the results may still be useful—see Sec. 16.12.)

16.4-3 Show that an alternative to (16-55) would be to define N^+ as the value of $[(\phi \pm \phi') + \pi]/2\pi n$ rounded to the nearest integer. Define a similar alternative to (16-56).

16.4-4 Consider a wedge illuminated by either an electric or magnetic line source parallel to the edge and at some distance from it $(\rho' \gg \lambda)$.

(a) At the reflection boundary (or incident boundary), show that the diffraction coefficient must have a discontinuity of magnitude $\sqrt{\rho'\rho/(\rho+\rho')}$.

(b) Show that at the reflection boundary (or incident boundary), the UTD diffraction coefficient is discontinuous by an amount $\pm\sqrt{L}$. What determines the sign of the discontinuity? The following approximation is useful:

$$F(X) = \left[\sqrt{\pi X} - 2Xe^{j(\pi/4)} - \frac{2}{3}X^2 e^{-j(\pi/4)} \right] e^{j(\pi/4+X)}$$

which is valid when X is small

(c) From the results of (a) and (b), show that the total field is continuous across the reflection (or incident) shadow boundary.

16.4-5 A plane wave is incident at an angle of $\gamma'_o = 45°$, $\phi' = 30°$ on the edge of a $90°\left(n=\frac{3}{2}\right)$ conducting wedge.

(a) Use (16-52) and Fig. 16-17 to calculate $E^d_\perp$ at a distance $s = 2\lambda$ when $\phi = 120°$, $132°$, $138°$, $180°$, $222°$, $228°$, and $260°$ when $E^i_\perp = 1\,\text{V/m}$.

(b) Repeat (a) for $E^d_{||}$ when $E^i_{||} = 1\,\text{V/m}$.

16.4-6 (a) Use (16-36), (16-37), and (16-42) to compute the total field in Fig. 16-14a. Your result will differ from that in Fig. 16-14a at the reflection and shadow boundaries. Why?

(b) Recompute (a), but use (16-52) instead of (16-42).

(c) Comment on the difference between the results in (a) and (b) above.

16.4-7 (a) Using trigonometric identities, put (16-42) into a form similar to (16-52) but without the transition functions, F.

(b) Comment on the purpose of the transition functions.

16.5-1 Use the E-plane model in Fig. 16-18b and a computer program for wedge diffraction to verify the curves in Fig. 9-16 that were obtained by aperture integration.

16.5-2 Explain why the rays in Fig. 16-18c make a negligible contribution to the radiation pattern except when $\zeta \approx 90°$.

16.5-3 Show that the doubly diffracted field from Q_1 in Fig. 16-18c can be written as

$$E^d_{1,2}(P) = \frac{1}{2} \frac{e^{-j\beta\rho_E}}{\sqrt{\rho_E}} D_\perp(Q_2)D_\perp(Q_1) \frac{e^{-j\beta 2a}}{\sqrt{2a}} \frac{e^{-j\beta r}}{\sqrt{r}} e^{j\beta a \sin\zeta}.$$

16.5-4 Redo Prob. 15.8-5.

16.5-5 Calculate the diffracted field from the 194° interior wedge angle edge on the outside of the horn antenna in Fig. 15-18 that is formed by the join of the horn wall with the waveguide. Use $\rho = \lambda$. Assume that the electric field incident on the join is $1\,\text{V/m}$. Compare your result to what you observe in Fig. 11-21d of [H.3: Stutzman, 2nd ed., Chap. 11].

16.6-1 Use a computer program for wedge diffraction to calculate the total diffracted field for $0 \le \zeta \le 2\pi$ for the antenna of Fig. 16-20. Compare your results with Fig. 16-21. Why is there a difference?

16.6-2 Draw a sketch of the "creeping wave" rays (see Sec. 16.11) on the back side of the parabolic reflector of Fig. 16-20. Now draw rays that originate at Q_1 or Q_2 and reflect several times along the *inside* surface parabolic reflector. These rays are called whispering gallery rays.

16.6-3 Show that the doubly diffracted ray from Q_1 in Fig. 16-20 can be written as

$$E^d_{1,2}(P) = \frac{e^{-j\beta\rho_E}}{\sqrt{\rho_E}} F(\theta_o)D_{||}(Q_2)D_{||}(Q_1) \cdot \frac{e^{-j\beta 2a}}{\sqrt{2a}} \frac{e^{-j\beta r}}{\sqrt{r}} e^{j\beta a \sin\zeta}$$

16.6-4 Derive (16-71).

16.6-5 If the line source in Fig. 16-20 is a magnetic line source, calculate the far-field pattern. Your result will be similar to that in Fig. 16-21, except that the discontinuity at $\zeta = 90°$ will be greater and the back lobes will be about 8 dB higher. Why?

16.6-6 Use the UTD to calculate the H-plane pattern of a 90° corner reflector antenna with a dipole feed. The dipole feed is 0.5λ from the apex of the reflector, the reflector sides are 1.0λ long, and the aperture of the corner reflector is 1.414λ across.

16.7-1 The diffracted field that is neglected in (16-86) may be written generally as [12]

$$E^d_{SD}(P) = \frac{1}{2j\beta} \frac{\partial E^i(Q)}{\partial n} \frac{\partial}{\partial \phi'} D_\| \bigg|_{\phi'=0} \sqrt{\frac{\rho}{s(\rho+s)}} e^{-j\beta s}$$

Compare the value of this slope diffracted field with the direct field in (16-86) when $\theta = 90°$.

16.7-2 (a) Using (16-36) and (16-37) in (16-33), show how a factor of 2 arises in $E(\rho, \phi)$ for the perpendicular polarization in the infinite ground plane case ($v_B = 0$) when the plane wave has grazing incidence to the ground plane ($\phi' = 0$).

   (b) Then verify, in general, that at grazing incidence the diffracted field must be multiplied by $\frac{1}{2}$, as in (16-80) and (16-82), to obtain the correct value of the diffracted field. To do this, use either the asymptotic form in (16-42) or (16-59) to show that $D_\| \rightarrow 0$ and a factor of 2 naturally arises in $D_\perp$.

16.8-1 A short monopole (stub antenna) is mounted at the center of a square ground plane 6λ on a side as shown in Fig. 16-27a.

   (a) Using the two-point approximation, show that the relative diffracted field in the region $200° < \phi_1 < 340°$ can be expressed by

$$E^d = \frac{e^{-j(\beta r + \pi/4)}}{\sqrt{2\pi\beta r}} \left[\frac{1}{\cos\dfrac{\phi_1}{2}} - \frac{e^{-j12\pi \cos\phi_2}}{\cos\dfrac{\phi_2}{2}} \right]$$

   where $\phi_2 = 2\pi - (\phi_1 - \pi)$
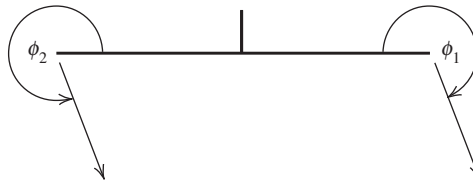
   (b) Why must the diffracted field be zero when $\phi_1 = 270°$ for this problem? Use a sketch and physical reasoning to explain why.

   (c) Calculate and plot a graph of the diffracted field for $200° < \phi_1 < 340°$. Compare your results with Fig. 16-28.

16.9-1 Derive (16-97) and (16-98).

16.9-2 Derive (16-101) and (16-102).

16.9-3 A short monopole (stub antenna) is mounted at the center of a circular ground plane 6λ in diameter as shown in Fig. 16-27b.

   (a) Using the equivalent concept, show that the relative diffracted field in the region $90° \leq \theta \leq 180°$ can be expressed by

$$E^d = -\frac{e^{-j(\beta r + \pi/4)}}{\sqrt{2\pi\beta r}} \frac{1}{\cos(\phi/2)} 2\pi j J_1(6\pi \sin\theta)$$

   where J_1 is the first-order Bessel function Note that

$$\int_0^{2\pi} \cos(\xi - \xi') e^{jx \cos(\xi-\xi')} \, d\xi' = 2\pi j J_1(x)$$

   (b) Calculate the diffracted field and compare with that calculated in Prob. 16.8-1.

16.10-1 A triangular cylinder is illuminated by a line source as shown. Apply the self-consistent method to this problem by setting up (16-110) in a form similar to (16-109). Note that

some of the matrix elements will be zero. Check your solution with that in [17].
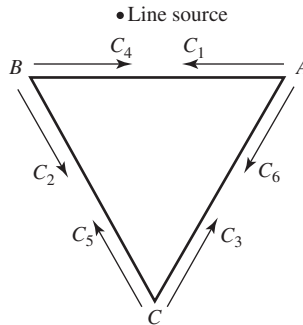
16.11-1 (a) Use the information in Table 16-1 to compute σ_w in (16-127). Plot σ_w vs. βa for $0.1 \leq \beta a \leq 10$ for both polarizations on the same graph.

(b) Repeat part (a) when the geometrical optics contribution (only) is multiplied by $\left[1 - j\frac{5}{16(\beta a)} + \frac{127}{512(\beta a)^2}\right]$ in the parallel case and by $\left[1 + j\frac{11}{16(\beta a)} - \frac{353}{512(\beta a)^2}\right]$ in the perpendicular case [H.10.4: James]. Plot both results on the same graph. Compare with Fig. 16-34.

(c) Plot the parallel polarization results from (a) and (b) on one graph and the perpendicular polarization results on another. Comment on your results.

16.11-2 Which of the following applies to a creeping wave? (a) has no attenuation on a flat surface; (b) propagates in the shadowed region of a target; (c) has a larger attenuation per unit length for small radii of curvature than for large radii of curvature.

16.12-1 (a) Show that diffraction by a straight edge is independent of frequency (wavelength).

(b) Can the results in Figs. 16-37 and 16-38 be scaled to a frequency different than that used in the figures? Why?

16.12-2 (a) A vertically polarized cellular antenna transmits 20 W at 860 MHz. A receiving an- tenna is shadowed by a 0.3-km-high ridge normal to a line drawn between the two antennas as shown. How much power is available at the terminals of the receiving antenna if the gain of the receiving antenna in the direction of the ridge is 4 dB and that of the trans- mitting antenna is 15 dB toward the ridge? As a rough approximation, assume the ridge is perfectly conducting.
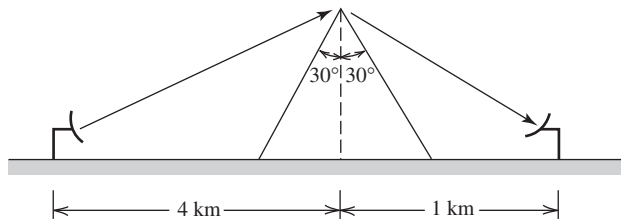
(b) Compare your result to the power that would be available if the ridge was not present.

(c) What is the excess path loss due to the presence of the ridge?

16.13-1 Consider a monopole at the center of a square ground plane whose sides are $\lambda/2$ long. The monopole is to be represented using pulse basis functions and delta weighting functions. The four sides of the ground plane are to be accounted for using wedge diffraction. Diffraction by the four corners is to be ignored. Derive the necessary equations that would enable you to calculate Z_{mn}^g in (16-132).

16.13-2 Derive (16-138).

16.13-3 For the problem in Fig. 16-42, show all possible ray paths that do not involve double (or higher-order) diffractions.

16.13-4 A dipole of length ℓ is located a distance d from the surface of an infinitely long circular cylinder of radius a. The dipole is parallel to the axis of the cylinder. Show how you

would account for the presence of the cylinder if only the dipole is represented by the method of moments.

16.14-1 Using physical optics, show that the radar cross section of a flat rectangular plate at normal incidence is $\sigma = 4\pi(A^2/\lambda^2)$ where A is the area of the plate.

16.14-2 Eq. (16-156) can be converted to a different and often useful form by noting that $(\hat{\mathbf{z}} \cdot \hat{\mathbf{n}})\ ds$ is the projection of the element of surface area ds onto the xy-plane. Thus, $(\hat{\mathbf{z}} \cdot \hat{\mathbf{n}})\ ds = ds_z = (ds_z/d\ell)\ d\ell$ where ds_z is the projection of ds onto the xy-plane. Then (16-156) becomes

$$\sigma = \frac{4\pi}{\lambda^2} \left(\int_0^L e^{-j2\beta\ell} \frac{ds_z}{d\ell} d\ell \right)$$

where ℓ is the distance from the reference plane to the surface. Use the above expression for the radar cross section to derive the physical optics expression for the RCS of the sphere.

16.14-3 Show that the RCS of an infinite cone (as shown) is $\sigma = (\lambda^2\ \tan^4\ \alpha)/16\pi$
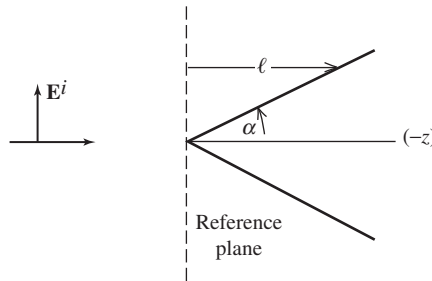
16.14-4 Show that the RCS of a square flat plate with edges parallel to the x- and y-axes, and direction of incidence in the xz-plane, is

$$\sigma = \frac{4\pi a^4}{\lambda^2} \left[\frac{\sin(\beta a\ \sin\theta)}{\beta a\ \sin\theta} \right]^2 \cos^2\theta$$

where a is the length of one side. Compare the angular variation of this result with that of the uniformly illuminated line source in Chap. 5.

16.14-5 Show that the RCS of a circular flat plae, or disk, in the xy-plane is

$$\sigma = \frac{\pi a^2}{\tan^2\theta} \left[J_1\left(\frac{4\pi a \sin\theta}{\lambda}\right) \right]^2$$

where a in the radius of the disk and $J_1(x)$ the Bessel function of order one. Also show that at $\theta = 0°$, the above expression reduces to $\sigma = (4\pi/\lambda^2)A^2$ where A is the area of the disk [4].

16.14-6 Show that in Example 16-5, the dot product $(\hat{\mathbf{z}} \cdot \hat{\mathbf{n}})$ is the same for both the convex and concave cases.

16.14-7 The radar cross section of a target depends on which of the following? (a) the waveform of the radar; (b) the range to target; (c) the value of the incident field at the target

16.14-8 Show that RCS determined by the physical optics method contains no polarization information.

16.14-9 Which of the following is true? The field $2\hat{\mathbf{n}} \times \mathbf{H}^i$ at the surface of a perfectly conducting object is: (a) exact in all cases; (b) without polarization information; (c) the geometrical optics surface field.

16.16-1 After writing

$$I_b = \int_b^{\infty} f(x)e^{j\beta\gamma(x)}dx = \frac{1}{j\beta}\int_b^{\infty} \frac{f(x)}{\gamma'(x)}j\beta\gamma'(x)e^{j\beta\gamma(x)}\ dx$$

integrate by parts to obtain (16-185).

16.16-2 Show that

$$I_{-a} \cong \frac{1}{j\beta} \frac{f(-a)}{\gamma'(-a)} e^{j\beta\gamma(-a)}$$

16.16-3 Show that (16-183) follows from (16-180) and (16-181).

16.16-4 Interpret the discussion between (16-29) and (16-30) in terms of the concept of stationary phase.

16.16-5 Using physical reasoning based on the concept of stationary phase, explain why the monostatic axial incidence scattering by a parabolic surface is the same for both convex and concave cases. (See Prob. 16.1-4.)

16.17-1 Show that $\left\{ \begin{matrix} f \\ g \end{matrix} \right\}$ in (12-202) can be expressed as

$$\left\{ \begin{matrix} f \\ g \end{matrix} \right\} = \frac{-1}{2n} \left[\cot\left(\frac{\pi + \phi^-}{2n}\right) + \cot\left(\frac{\pi - \phi^-}{2n}\right) \right] \pm \frac{1}{2n} \left[\cot\left(\frac{\pi + \phi^+}{2n}\right) + \cot\left(\frac{\pi - \phi^+}{2n}\right) \right]$$

16.17-2 For the half-plane case in Fig. 16-14, analytically show that at the reflection and shadow boundaries, the singularity in f is canceled by the singularity in f_o making f' in (16-200) continuous.

HINT: Identify which cotangent term in the solution of Prob. 16.5-1 that is singular at the reflection boundary and which term is singular at the shadow boundary. Cancel these terms with the singularity in f_o by letting $\phi^+ = \pi \pm \delta$ at the GO reflection boundary and by letting $\phi^- = \pi \pm \delta$ at the GO shadow boundary.

16.17-3 For the half-plane case in Fig. 16-14a show by numerical computation that f' in (12-202) is continuous across the reflection and shadow boundaries. Also compute separately f and f_o at these boundaries. Use the geometry in Fig. 16-14.

16.17-4 Evaluate I_o in (16-212).

16.17-5 Show that (16-196) substituted into (16-197) leads to (16-194).

Appendix A

Frequency Bands

A.1 RADIO FREQUENCY BANDS

Frequency ——————————→

3 kHz	30 kHz	300 kHz	3 MHz	30 MHz	300 MHz	3 GHz	30 GHz	300 GHz
	VLF	LF	MF	HF	VHF	UHF	SHF	EHF	
100 km	10 km	1 km	100 m	10 m	1 m	10 cm	1 cm	1 mm

←—————————— Wavelength
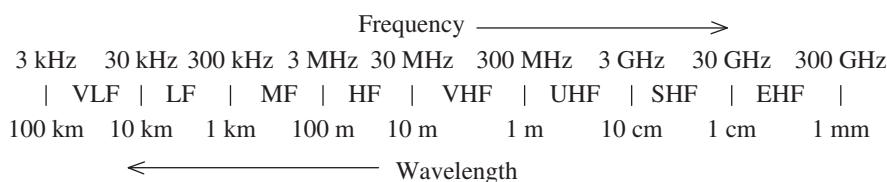
A.2 TELEVISION CHANNEL FREQUENCIES (IN NORTH AMERICA)

VHF

Channel No.	Frequency Range (MHz)	Channel No.	Frequency Range (MHz)
2	54–60	8	180–186
3	60–66	9	186–192
4	66–72	10	192–198
5	76–82	11	198–204
6	82–88	12	204–210
7	174–180	13	210–216

UHF

Channel No.	Frequency Range (MHz)	Channel No.	Frequency Range (MHz)	Channel No.	Frequency Range (MHz)
14	470–476	23	524–530	32	578–584
15	476–482	24	530–536	33	584–590
16	482–488	25	536–542	34	590–596
17	488–494	26	542–548	35	596–602
18	494–500	27	548–554	36	602–608
19	500–506	28	554–560	37	608–614
20	506–512	29	560–566	38	614–620
21	512–518	30	566–572	39	620–626
22	518–524	31	572–578	40	626–632

(*Continued*)

UHF (Continued)

Channel No.	Frequency Range (MHz)	Channel No.	Frequency Range (MHz)
41	632–638	47	668–674
42	638–644	48	674–680
43	644–650	49	680–686
44	650–656	50	686–692
45	656–662	51	692–698
46	662–668		

Note: The carrier frequency for the video portion is the lower frequency plus 1.25 MHz. The audio carrier frequency is the upper frequency minus 0.25 MHz. All channels have a 6-MHz bandwidth. For example, the Channel 2 video carrier is at 55.25 MHz and the audio carrier is at 59.75 MHz.

A.3   CELLULAR TELEPHONE BANDS

Band Name (s)	Transmit/Receive (MHz)	Region: Technology
Cellular (GSM 850)	824–849/869–894	U.S.: 2G, 3G, LTE
GSM 900 (WCDMA)	880–915/925–960	non-U.S.: 2G, 3G
DCS (GSM 1800)	1710–1785/1805–1880	non-U.S.: 2G
DCS 1900	1710–1755/2110–2155	U.S.: 3G, LTE
PCS 1900 (GSM 1900)	1850–1910/1930–1990	Americas: 2G; U.S.: 2G, 3G, LTE
UMTS (WCDMA)	1910–1980/2110–2170	Europe, Asia: 3G, LTE

A.4   RADAR BANDS

IEEE Band Designations

HF	3–30 MHz
VHF	30–300 MHz
UHF	300–1000 MHz
L-band	1–2 GHz
S-band	2–4 GHz
C-band	4–8 GHz
X-band	8–12 GHz
Ku-band	12–18 GHz
K-band	18–27 GHz
Ka-band	27–40 GHz
V-band	40–75 GHz
W-band	75–110 GHz
Millimeter	110–300 GHz

Appendix B

Material Data and Constants

B.1 CONDUCTIVITIES OF GOOD CONDUCTORS

Conductor	Conductivity (S/m)
Silicon steel	2×10^6
Brass	1.1×10^7
Aluminum	3.5×10^7
Gold	4.1×10^7
Copper	5.8×10^7
Silver	6.1×10^7

B.2 WIRE DATA

Wire Size AWG	Diameter in mm (in.)	Single Copper Wire Continuous Duty Current Capacity (A)	Copper Wire dc Resistance per Unit Length (Ω/100 m)
8	3.264 (0.1285)	73	0.1952
9	2.906 (0.1144)	—	0.2462
10	2.588 (0.1019)	55	0.3103
11	2.305 (0.0907)	—	0.3914
12	2.053 (0.0808)	41	0.4935
13	1.828 (0.0720)	—	0.6224
14	1.628 (0.0641)	32	0.7849
16	1.291 (0.0508)	22	1.248
18	1.024 (0.0403)	16	1.984
20	0.812 (0.0320)	11	3.155
22	0.644 (0.0253)	—	5.017
24	0.511 (0.0201)	—	7.98
26	0.405 (0.0159)	—	12.69
28	0.321 (0.0129)	—	20.17
30	0.255 (0.0100)	—	32.06

B.3   DIELECTRIC CONSTANT: PERMITTIVITY OF FREE SPACE

$$\varepsilon_o = 8.854 \times 10^{-12} \text{ F/m} \approx 10^{-9}/36\pi \text{ F/m}$$

B.4   PERMEABILITY OF FREE SPACE

$$\mu_o = 1.2566 \times 10^{-6} \text{ H/m} \approx 4\pi \times 10^{-7} \text{ H/m}$$

B.5   VELOCITY OF LIGHT IN FREE SPACE

$$c = 1/\sqrt{\mu_o \varepsilon_o} = 2.997925 \times 10^8 \text{ m/s}$$

B.6   INTRINSIC IMPEDANCE OF FREE SPACE

$$\eta_o = \sqrt{\frac{\mu_o}{\varepsilon_o}} = 376.73 \ \Omega \approx 120\pi \ \Omega$$

B.7   PROPERTIES OF SOME COMMON DIELECTRICS

Dielectric		Permittivity at 10 GHz	
Type	Name	ε_r'	tan δ
Foam	Styrofoam	1.03	0.00015
Plastic	Teflon (PTFE)	2.08	0.00037
Plastic	Duroid 5870	2.33	0.0012
Plastic	Polystyrene	2.54	0.00043
Plastic	Duroid 6006	6.15	0.0027
Ceramic	Aluminum Oxide	8.79	0.0018

Note: The loss in dielectrics is represented using complex relative permittivity: $\varepsilon_r = \varepsilon_r' - j\varepsilon_r'' = \varepsilon_r'(1 - j \tan \delta)$, where tan δ is called *loss tangent*.

Appendix C

Coordinate Systems and Vectors

C.1 THE COORDINATE SYSTEMS AND UNIT VECTORS

C.1.1 The Coordinate Systems
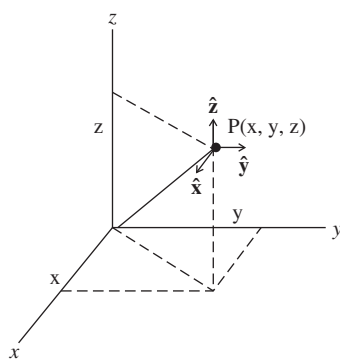
Rectangular Coordinates
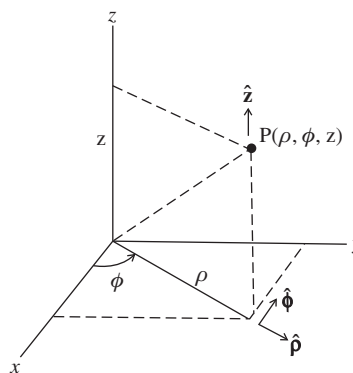
Cylindrical Coordinates
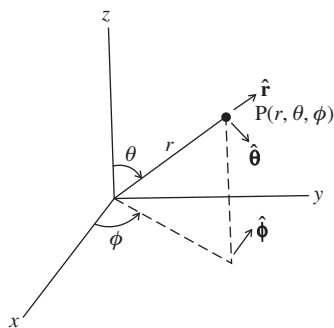
Spherical Coordinates

C.1.2   Unit Vector Representations

$$\hat{\mathbf{x}} = \hat{\mathbf{r}} \sin \theta \cos \phi + \hat{\boldsymbol{\theta}} \cos \theta \cos \phi - \hat{\boldsymbol{\phi}} \sin \phi \tag{C-1}$$

$$\hat{\mathbf{y}} = \hat{\mathbf{r}} \sin \theta \sin \phi + \hat{\boldsymbol{\theta}} \cos \theta \cos \phi + \hat{\boldsymbol{\phi}} \cos \phi \tag{C-2}$$

$$\hat{\mathbf{z}} = \hat{\mathbf{r}} \cos \theta - \hat{\boldsymbol{\theta}} \sin \theta \tag{C-3}$$

$$\hat{\mathbf{r}} = \hat{\mathbf{x}} \sin \theta \cos \phi + \hat{\mathbf{y}} \sin \theta \sin \phi + \hat{\mathbf{z}} \cos \theta \tag{C-4}$$

$$\hat{\boldsymbol{\theta}} = \hat{\mathbf{x}} \cos \theta \cos \phi + \hat{\mathbf{y}} \cos \theta \sin \phi - \hat{\mathbf{z}} \sin \theta \tag{C-5}$$

$$\hat{\boldsymbol{\phi}} = -\hat{\mathbf{x}} \sin \phi + \hat{\mathbf{y}} \cos \phi \tag{C-6}$$

C.2   VECTOR IDENTITIES

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{A} \cdot \mathbf{B})\mathbf{C} \tag{C-7}$$

$$(\mathbf{A} \times \mathbf{B}) \times \mathbf{C} = (\mathbf{C} \cdot \mathbf{A})\mathbf{B} - (\mathbf{C} \cdot \mathbf{B})\mathbf{A} \tag{C-8}$$

$$\nabla \cdot (\nabla \times \mathbf{G}) = 0 \tag{C-9}$$

$$\nabla \times \nabla g = 0 \tag{C-10}$$

$$\nabla \cdot \nabla g = \nabla^2 g \tag{C-11}$$

$$\nabla(f + g) = \nabla f + \nabla g \tag{C-12}$$

$$\nabla \cdot (\mathbf{F} + \mathbf{G}) = \nabla \cdot \mathbf{F} + \nabla \cdot \mathbf{G} \tag{C-13}$$

$$\nabla(fg) = g\nabla f + f\nabla g \tag{C-14}$$

$$\nabla \cdot (f\mathbf{G}) = \mathbf{G} \cdot (\nabla f) + f(\nabla \cdot \mathbf{G}) \tag{C-15}$$

$$\nabla \times (f\mathbf{G}) = (\nabla f) \times \mathbf{G} + f(\nabla \times \mathbf{G}) \tag{C-16}$$

$$\nabla \times (\nabla \times \mathbf{G}) = \nabla(\nabla \cdot \mathbf{G}) - \nabla^2 \mathbf{G} \tag{C-17}$$

$$\nabla^2 \mathbf{G} = \hat{\mathbf{x}}\nabla^2 G_x + \hat{\mathbf{y}}\nabla^2 G_y + \hat{\mathbf{z}}\nabla^2 G_z \tag{C-18}$$

$$\nabla \cdot (\mathbf{F} \times \mathbf{G}) = \mathbf{G} \cdot (\nabla \times \mathbf{F}) - \mathbf{F} \cdot (\nabla \times \mathbf{G}) \tag{C-19}$$

$$\mathbf{F} \cdot (\mathbf{G} \times \mathbf{H}) = \mathbf{G} \cdot (\mathbf{H} \times \mathbf{F}) = \mathbf{H} \cdot (\mathbf{F} \times \mathbf{G}) \tag{C-20}$$

$$\mathbf{\nabla} \times (\mathbf{F} \times \mathbf{G}) = \mathbf{F}(\mathbf{\nabla} \cdot \mathbf{G}) - \mathbf{G}(\mathbf{\nabla} \cdot \mathbf{F}) + (\mathbf{G} \cdot \mathbf{\nabla})\mathbf{F} - (\mathbf{F} \cdot \mathbf{\nabla})\mathbf{G} \tag{C-21}$$

$$\mathbf{\nabla}(\mathbf{F} \cdot \mathbf{G}) = (\mathbf{F} \cdot \mathbf{\nabla})\mathbf{G} + (\mathbf{G} \cdot \mathbf{\nabla})\mathbf{F} + \mathbf{F} \times (\mathbf{\nabla} \times \mathbf{G}) + \mathbf{G} \times (\mathbf{\nabla} \times \mathbf{F}) \tag{C-22}$$

$$\iiint_v \mathbf{\nabla} \cdot \mathbf{G} \, dv = \oiint_s \mathbf{G} \cdot ds \quad \text{divergence theorem} \tag{C-23}$$

$$\iiint_s (\mathbf{\nabla} \times \mathbf{G}) \cdot ds = \oint_l \mathbf{G} \cdot dl \qquad \text{Stokes' theorem} \tag{C-24}$$

C.3   VECTOR DIFFERENTIAL OPERATORS

C.3.1   Rectangular Coordinates

$$\mathbf{\nabla}g = \hat{\mathbf{x}}\frac{\partial g}{\partial x} + \hat{\mathbf{y}}\frac{\partial g}{\partial y} + \hat{\mathbf{z}}\frac{\partial g}{\partial z} \tag{C-25}$$

$$\mathbf{\nabla} \cdot \mathbf{G} = \frac{\partial G_x}{\partial x} + \frac{\partial G_y}{\partial y} + \frac{\partial G_z}{\partial z} \tag{C-26}$$

$$\mathbf{\nabla} \times \mathbf{G} = \hat{\mathbf{x}}\left(\frac{\partial G_z}{\partial y} - \frac{\partial G_y}{\partial z}\right) + \hat{\mathbf{y}}\left(\frac{\partial G_x}{\partial z} - \frac{\partial G_z}{\partial x}\right) + \hat{\mathbf{z}}\left(\frac{\partial G_y}{\partial x} - \frac{\partial G_x}{\partial y}\right) \tag{C-27}$$

$$\mathbf{\nabla}^2 g = \frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} + \frac{\partial^2 g}{\partial z^2} \tag{C-28}$$

C.3.2   Cylindrical Coordinates

$$\mathbf{\nabla}g = \hat{\boldsymbol{\rho}}\frac{\partial g}{\partial \rho} + \hat{\boldsymbol{\phi}}\frac{1}{\rho}\frac{\partial g}{\partial \phi} + \hat{\mathbf{z}}\frac{\partial g}{\partial z} \tag{C-29}$$

$$\mathbf{\nabla} \cdot \mathbf{G} = \frac{1}{\rho}\frac{\partial}{\partial \rho}(\rho G_\rho) + \frac{1}{\rho}\frac{\partial G_\phi}{\partial \phi} + \frac{\partial G_z}{\partial z} \tag{C-30}$$

$$\mathbf{\nabla} \times \mathbf{G} = \hat{\boldsymbol{\rho}}\left(\frac{1}{\rho}\frac{\partial G_z}{\partial \phi} - \frac{\partial G_\phi}{\partial z}\right) + \hat{\boldsymbol{\phi}}\left(\frac{\partial G_\rho}{\partial z} - \frac{\partial G_z}{\partial \rho}\right) + \hat{\mathbf{z}}\frac{1}{\rho}\left[\frac{\partial}{\partial \rho}(\rho G_\phi) - \frac{\partial G_\rho}{\partial \phi}\right] \tag{C-31}$$

$$\mathbf{\nabla}^2 g = \frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho \frac{\partial g}{\partial \rho}\right) + \frac{1}{\rho^2}\frac{\partial^2 g}{\partial \phi^2} + \frac{\partial^2 g}{\partial z^2} \tag{C-32}$$

C.3.3   Spherical Coordinates

$$\nabla g = \hat{\mathbf{r}}\frac{\partial g}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial g}{\partial \theta} + \hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\frac{\partial g}{\partial \phi} \tag{C-33}$$

$$\nabla \cdot \mathbf{G} = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2 G_r) + \frac{1}{r\sin\theta}\frac{\partial}{\partial \theta}(G_\theta \sin\theta) + \frac{1}{r\sin\theta}\frac{\partial G_\phi}{\partial \phi} \tag{C-34}$$

$$\begin{aligned}
\nabla \times \mathbf{G} = \ &\hat{\mathbf{r}}\frac{1}{r\sin\theta}\left[\frac{\partial}{\partial \theta}(G_\phi \sin\theta) - \frac{\partial G_\theta}{\partial \phi}\right] \\
&+ \hat{\boldsymbol{\theta}}\frac{1}{r}\left[\frac{1}{\sin\theta}\frac{\partial G_r}{\partial \phi} - \frac{\partial}{\partial r}(rG_\phi)\right] \\
&+ \hat{\boldsymbol{\phi}}\frac{1}{r}\left[\frac{\partial}{\partial r}(rG_\theta) - \frac{\partial G_r}{\partial \theta}\right]
\end{aligned} \tag{C-35}$$

$$\nabla^2 g = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial g}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial g}{\partial \theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 g}{\partial \phi^2} \tag{C-36}$$

Appendix D

Trigonometric Relations

$$\sin(\alpha \pm \beta) = \sin \alpha \cos \beta \pm \cos \alpha \sin \beta \tag{D-1}$$

$$\cos(\alpha \pm \beta) = \cos \alpha \cos \beta \mp \sin \alpha \sin \beta \tag{D-2}$$

$$\sin\left(\frac{\pi}{2} \pm \alpha\right) = \cos \alpha \tag{D-3}$$

$$\cos\left(\frac{\pi}{2} \pm \alpha\right) = \mp \sin \alpha \tag{D-4}$$

$$\sin \alpha \cos \beta = \tfrac{1}{2}\left[\sin(\alpha + \beta) + \sin(\alpha - \beta)\right] \tag{D-5}$$

$$\cos \alpha \sin \beta = \tfrac{1}{2}\left[\sin(\alpha + \beta) + \sin(\alpha - \beta)\right] \tag{D-6}$$

$$\cos \alpha \cos \beta = \tfrac{1}{2}\left[\cos(\alpha + \beta) + \cos(\alpha - \beta)\right] \tag{D-7}$$

$$\sin \alpha \sin \beta = -\tfrac{1}{2}\left[\cos(\alpha + \beta) - \cos(\alpha - \beta)\right] \tag{D-8}$$

$$\sin \alpha = 2 \sin \frac{\alpha}{2} \cos \frac{\alpha}{2} \tag{D-9}$$

$$\sin 2\alpha = 2 \sin \alpha \cos \alpha \tag{D-10}$$

$$\cos \alpha = 2 \cos^2 \frac{\alpha}{2} - 1 = 1 - 2 \sin^2 \frac{\alpha}{2} \tag{D-11}$$

$$\cos 2\alpha = 2 \cos^2 \alpha - 1 = \cos^2 \alpha - \sin^2 \alpha = 1 - 2 \sin^2 \alpha \tag{D-12}$$

$$\cos 3\alpha = 4 \cos^3 \alpha - 3 \cos \alpha \tag{D-13}$$

$$\cos 4\alpha = 8 \cos^4 \alpha - 8 \cos^2 \alpha + 1 \tag{D-14}$$

$$\cos\, m\alpha = 2^{m-1}\cos{}^m\alpha - \frac{m}{1!}2^{m-3}\cos{}^{m-2}\alpha$$
$$+ \frac{m(m-3)}{2!}2^{m-5}\cos{}^{m-4}\alpha + \cdots \tag{D-15}$$

$$1 = \sin^2\alpha + \cos^2\alpha \tag{D-16}$$

$$\sec^2\alpha = \frac{1}{\cos^2\alpha} = 1 + \tan^2\alpha \tag{D-17}$$

$$\sin\alpha = \alpha - \frac{\alpha^3}{3!} + \frac{\alpha^5}{5!} - \frac{\alpha^7}{7!} + \cdots \tag{D-18}$$

$$\cos\alpha = 1 - \frac{\alpha^2}{2!} + \frac{\alpha^4}{4!} - \frac{\alpha^6}{6!} + \cdots \tag{D-19}$$

$$e^{\pm j\alpha} = \cos\alpha \pm j\sin\alpha \tag{D-20}$$

$$\tan\alpha = \frac{\sin\alpha}{\cos\alpha} \tag{D-21}$$

$$\sin\alpha = \frac{e^{j\alpha} - e^{-j\alpha}}{2j} \tag{D-22}$$

$$\cos\alpha = \frac{e^{j\alpha} + e^{-j\alpha}}{2} \tag{D-23}$$

$$\sin^{-1}\alpha = \alpha + \frac{\alpha^3}{2\cdot 3} + \frac{3\alpha^5}{2\cdot 4\cdot 5} + \cdots \tag{D-24}$$

$$\cos^{-1}\alpha = \frac{\pi}{2} - \left(\alpha + \frac{\alpha^3}{2\cdot 3} + \frac{3\alpha^5}{2\cdot 4\cdot 5} + \cdots\right) \tag{D-25}$$

Appendix **E**

Hyperbolic Relations

$$\sinh \alpha = \frac{e^{\alpha} - e^{-\alpha}}{2} = \alpha + \frac{\alpha^3}{3!} + \frac{\alpha^5}{5!} + \frac{\alpha^7}{7!} + \cdots \tag{E-1}$$

$$\cosh \alpha = \frac{e^{\alpha} + e^{-\alpha}}{2} = 1 + \frac{\alpha^2}{2!} + \frac{\alpha^4}{4!} + \frac{\alpha^6}{6!} + \cdots \tag{E-2}$$

$$\tanh \alpha = \frac{\sinh \alpha}{\cosh \alpha} = \frac{1}{\coth \alpha} \tag{E-3}$$

$$\sinh(\alpha \pm j\beta) = \sinh \alpha \, \cos \beta \pm j \cosh \alpha \sin \beta \tag{E-4}$$

$$\cosh(\alpha \pm j\beta) = \cosh \alpha \, \cos \beta \pm j \sinh \alpha \sin \beta \tag{E-5}$$

$$\sinh(j\alpha) = j \sin \alpha = \frac{e^{j\alpha} - e^{-j\alpha}}{2} \tag{E-6}$$

$$\cosh(j\alpha) = \cos \alpha = \frac{e^{j\alpha} + e^{-j\alpha}}{2} \tag{E-7}$$

Appendix F

Mathematical Relations

F.1 DIRAC DELTA FUNCTION

The Dirac delta function (or impulse function) is zero everywhere except when the argument is zero

$$\delta(x - x_o) = 0 \qquad \text{for} \qquad x \neq x_o \qquad \text{(F-1)}$$

For the zero argument case, the function is singular but in a special way: The area is unity, that is,

$$\int_{-\infty}^{\infty} \delta(x - x_o)dx = 1 \qquad \text{(F-2)}$$

Another useful property of the Dirac delta function follows:

$$\int_{-\infty}^{\infty} g(x)\delta(x - x_o)dx = g(x_o) \qquad \text{(F-3)}$$

F.2 BINOMIAL THEOREM

$$(a + b)^n = a^n + na^{n-1}b + \frac{n(n-1)}{2!}a^{n-2}b^2$$
$$+ \frac{n(n-1)(n-2)}{3!}a^{n-3}b^3 + \cdots \qquad \text{(F-4)}$$

$$(1 \pm x)^n \approx 1 \pm nx \qquad \text{for} \qquad x \ll 1 \qquad \text{(F-5)}$$

F.3 BESSEL FUNCTIONS

$$J_o(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{jx \cos \alpha} \, d\alpha \qquad \text{(F-6)}$$

$$J_n(x) = \frac{j^{-n}}{2\pi} \int_0^{2\pi} e^{jx \cos \alpha} \cos (n\alpha) \, d\alpha$$

$$= \sum_{m=0}^{\infty} \frac{(-1)^m x^{2m+n}}{m!(m+n)!2^{2m+n}}$$

(F-7)

$$J_n(x) = \frac{2(n-1)}{x} J_{n-1}(x) - J_{n-2}(x)$$

(F-8)

$$\int x^{n+1} J_n(x) dx = x^{n+1} J_{n+1}(x)$$

(F-9)

$$\int_0^1 (1 - x^2)^n x J_0(bx) \, dx = \frac{2^n n!}{b^{n+1}} J_{n+1}(b)$$

(F-10)

F.4 SOME USEFUL INTEGRALS

$$\int \sin (a + bx) e^{cx} \, dx = \frac{e^{cx}}{b^2 + c^2} [c \sin (a + bx) - b \cos (a + bx)]$$

(F-11)

$$\int_{-\infty}^{\infty} \frac{\sin^2 x}{x^2} dx = \pi$$

(F-12)

$$\text{Si}(x) = \int_0^x \frac{\sin \tau}{\tau} d\tau \qquad \text{sine integral}$$

(F-13)

$$\text{Ci}(x) = -\int_x^{\infty} \frac{\cos \tau}{\tau} d\tau \qquad \text{cosine integral}$$

(F-14)

$$\text{Cin}(x) = \int_0^x \frac{1 - \cos \tau}{\tau} d\tau$$

(F-15)

$$\text{Cin}(x) = 0.5772 + \ln(x) - \text{Ci}(x)$$

(F-16)

$$C(x) = \int_0^x \cos \left(\frac{\pi}{2} \tau^2 \right) d\tau; \qquad C(-x) = -C(x)$$

(F-17a)

Fresnel integrals

$$S(x) = \int_0^x \sin \left(\frac{\pi}{2} \tau^2 \right) d\tau; \qquad S(-x) = -S(x)$$

(F-17b)

Appendix G

Computing Tools for Antennas

There are several well-developed commercial math applications packages, such as MATLAB, Mathcad, and Mathematica. Students and practicing engineers regularly use at least one of these computing tools. Any of these packages can be used to code formulas presented in this book to obtain numerical and graphical and even symbolic analytical results. This book does not present code statement listings. This is because there is no one standard package in use, and because there is learning value in developing one's own codes.

The Preface addressed the topic of *Simulations as a Learning Aid*. Simulation of antennas is now common, and students should become familiar with at least one package while taking a course from this text. Several chapter-end problems include simulation aspects. The area of simulation packages is very dynamic. New packages and upgrades are introduced frequently, and older packages become unsupported. Thus, no one package is recommended. Section H.10 lists several books focused on topics in computational methods, and some sources include codes. Following is a list of several suggested low-cost simulation computational packages. Students should acquire and use at least one wire antenna package. Most have excellent user interfaces that handle three-dimensional geometries and also have well-organized outputs.

G.1 WIRE ANTENNA SIMULATION PACKAGES

G.1.1 NEC Family

NEC (Numerical Electromagnetics Code) is the first large-problem-capable moment method code. The following website has links to several NEC-based codes: www.nec2.org.

G.1.2 EZNEC Family

EZNEC has packages of various capability levels based on the NEC moment method algorithm, but all are affordable. There is a free demonstration version with limited capability. EZNEC is an excellent choice for beginners. See [H.3: Huang and Boyle, Sec. 6.3] for examples using EZNEC. Go to: www.eznec.com.

G.1.3 GRADMAX

GRADMAX is a totally online interactive wire code that does not require purchase or download to use. It uses the Java applet technology and runs directly from a web browser

in any common operating system. It is moment method based and includes optimization features. See the following reference for details: R. Ravelo and M. Terada, "Analysis and Optimization of Wire Antennas over the Internet," *IEEE Ant. & Prop. Mag.*, Vol. 52, pp. 188–193, Feb. 2010. Go to: www.ene.unb.br/∼terada/antennas.

G.1.4   WIRE

WIRE is a free moment method based code. WIRE permits the user to specify arbitrary arrangements of straight wires of finite size, with or without lumped loads, and with arbitrary connectivity. Both antenna and scattering problems can be solved. Many antenna configurations can be modeled, including arrays. The method of moments solution approach is used, so full mutual coupling is accounted for. The following are available outputs: values for current distribution on the wires, input impedance, radiation patterns, and gain; and plots of current distributions and patterns. Go to: https://filebox.vt.edu /users/wadavis/.

G.2   PARABOLIC REFLECTOR ANTENNA SIMULATION PACKAGES

G.2.1   WebPRAC—Web-based Parabolic Reflector Antenna Code

WebPRAC is a user-friendly program for analysis of reflector antennas that is run interactively online. No purchase or download is required. It uses the Java applet technology and runs directly from a web browser in any common operating system. The main reflector geometry, including offset reflectors, as well as the desired illumination is specified by the user. The program returns the required feed pattern, gain, and radiation patterns, including cross-polarization patterns. See the following reference for details: R. Ravelo, M. Terada, and W. Stutzman, "Analysis of Reflector Antennas through the World Wide Web," *IEEE Ant. & Prop. Mag.*, Vol. 49, pp. 113–116, April 2007. Go to: www.ene.unb.br/∼terada/antennas.

G.2.2   GRASP—General Reflector and Antenna Farm Analysis Program

GRASP is a full-capability reflector antenna code. A free student version is available for download. Go to: www.ticra.com.

G.3   WEB SITES WITH ANTENNA CALCULATION AND MODELING TOOLS

- Dipole and array pattern calculation and plotting program.
  Go to: www.amanogawa.com.
- ANCAN—Analysis and Characterization of Antenna Arrays.
  Go to: www.tsc.uniovi.es/∼flasheras/ancan.html.
- Inexpensive hardware modeling approaches.
  Go to: www.freeantenans.com.

Appendix **H**

Book List

H.1 INTRODUCTION

This appendix lists nearly all antenna books ever written in English. The list is organized by technical topic, and within each topic then ordered alphabetically by the lead author's last name. An article with historical notes and useful information on locating the books is available.[1]

H.2 ANTENNA DEFINITIONS

The IEEE standard listed here is the major reference for antenna terms and their meaning. However, a reader cannot assume that a term found in books and articles follows the IEEE standard and should look in the book or article for the author's definition.

IEEE Standard Definitions of Terms for Antennas:
IEEE Standard 145-1993, IEEE, 28 pp., 1993.

H.3 FUNDAMENTAL BOOKS ON ANTENNAS

Fundamental books are textbooks or books covering several antenna topics. These references, for the most part, have consistent notation and have connectivity between topics.

J. Aharoni, *Antennae*, Oxford, 265 pp., 1946.

C. A. Balanis, *Antenna Theory: Analysis and Design*, 1st ed., Harper & Row, 790 pp., 1982; 2nd ed., Wiley, 941 pp., 1997; 3rd ed., Wiley, 1117 pp., 2005.

S. Best, *Introduction to Antennas*, SciTech Pub., 3 CDs, 2004.

Lamont Blake, *Antennas*, 1st ed., Wiley, 415 pp., 1966; 2nd ed., Artech House, 461 pp., 1984; 3rd ed., L. Blake and M. W. Long, *Antennas: Fundamentals, Design, and Measurements*, Scitech Pub., 650 pp., 2009.

G. A. Briggs and R. S. Roberts, *Aerial Handbook*, Rank Wharfedale, 175 pp., 1964. (revised, 1968)

R. A. Burberry, *VHF and UHF Antennas*, IEE Electromagnetic Waves Series No. 35-Peter Peregrinus, 301 pp., 1992.

---

[1] W. L. Stutzman, "Bibliography for Antennas: A List of Every English-Language Book Ever Written," *IEEE Antennas & Propagation Magazine*, Vol. 50, pp. 128–143, Aug. 2008. An updated version of the article is posted at http://www.wiley.com/college/stutzman/std.htm.

R. Chatterjee, *Antenna Theory and Practice*, 1st ed., Wiley, 314 pp., 1988; 2nd ed., New Age International Ltd., 394 pp., 1998.

C. Christodoulou and P. Wahid, *Fundamentals of Antennas: Concepts and Applications*, SPIE Press, 94 pp., 2001.

R. E. Collin and F. J. Zucker, Eds., *Antenna Theory*, McGraw-Hill, Part 1, 666 pp., 1969; Part 2, 683 pp., 1969.

F. R. Connor, *Antennas*, 1st ed., Edward Arnold, 99 pp., 1972; 2nd ed., 125 pp., 1989.

L. M. Dezettel, *Introduction to Antennas*, Radio Shack, 127 pp., 1972.

S. Drabowitch, A. Papiernik, and J. Encianas, *Modern Antennas*, 1st ed., Chapman & Hall, 631 pp., 1998; 2nd ed. (with H. Griffiths), Springer, 689 pp., 2005.

R. S. Elliott, *Antenna Theory and Design*, 1st ed., Prentice-Hall, 594 pp., 1981. Revised ed., Wiley–IEEE Press, 594 pp., 2003.

A. Z. Fradin, *Microwave Antennas*, Pergamon Press, 668 pp., 1961.

D. W. Fry and F. K. Goward, *Aerials for Centimetre Wave-Lengths*, Cambridge University Press, 172 pp., 1950.

V. Fusco, *Foundations of Antenna Theory and Techniques*, Prentice-Hall, 230 pp., 2005.

W. Gosling, *Radio Antennas and Propagation*, Newnes (Elsevier), 192 pp., 1998.

Heinrich Hertz, *Electric Waves*, MacMillan Company, 279 pp., 1893. *Comment*: This is the first book on antennas.

Y. Huang and K. Boyle, *Antennas: From Theory to Practice*, Wiley, 378 pp., 2008.

P.-S. Kildal, *Foundations of Antennas: A Unified Approach*, Studentlitteratur, 394 pp., 2000.

G. J. King, *The Practical Aerial Handbook*, 1st ed., Odhams, 224 pp., 1967; 2nd ed., Newnes-Butterworth, 232 pp., 1970.

R. W. P. King and C. W. Harrison, *Antennas and Waves: A Modern Approach*, MIT Press, 778 pp., 1969.

J. D. Kraus, *Antennas*, 1st ed., McGraw-Hill, 553 pp., 1950; 2nd ed., McGraw-Hill, 892 pp., 1988; J. Kraus and R. Marhefka, *Antennas for All Applications*, 3rd ed., McGraw-Hill, 938 pp., 2002.

E. A. Laport, *Radio Antenna Engineering*, McGraw-Hill, 563 pp., 1952.

Kai Fong Lee, *Principles of Antenna Theory*, Wiley, 324 pp., 1984.

T .S. M. Maclean, *Principles of Antennas—Wire and Aperture*, Cambridge Press, 360 pp., 1986.

Thomas A. Milligan, *Modern Antenna Design*, 1st ed., McGraw-Hill, 408 pp., 1985; 2nd ed., Wiley–IEEE Press, 614 pp., 2005.

George Monser, *Antenna Design: A Practical Guide*, McGraw-Hill, 163 pp., 1996.

E. B. Moullin, *Radio Aerials*, Oxford U. Press, 514 pp., 1949.

H. Page, *Principles of Aerial Design*, Van Nostrand, 172 pp., 1966.

E. Roubine and J. C. Bolomey, *Antennas, Volume 1: General Principles*, Hemisphere, 218 pp., 1987. S. Drabowitch and C. Ancona, *Antennas, Volume 2: Applications*, Hemisphere, 305 pp., 1988.

S. A. Schelkunoff, *Advanced Antenna Theory*, Wiley, 216 pp., 1952.

S. A. Schelkunoff and H. T. Friis, *Antenna Theory and Practice*, Wiley, 639 pp., 1952.

S. Silver, Editor, *Microwave Antenna Theory and Design,* MIT Radiation Laboratory Series Vol. 12, McGraw-Hill,, 623 pp., 1949.

Carl E. Smith, *Theory and Design of Directional Antenna Systems*, Cleveland Inst. of Radio Elec., 1946.

Martin S. Smith, *Introduction to Antennas*, Springer-Verlag, 128 pp., 1988.

Robert A. Smith, *Aerials for Metre and Centimetre Wave-Lengths*, Cambridge U. Press, 218 pp., 1949.

Woodrow Smith, *Antenna Manual*, Editors and Engineers, 301 pp., 1948.

W. L. Stutzman and G. A Thiele, *Antenna Theory and Design*, 1st ed., Wiley, 598 pp., 1981; 2nd ed., Wiley, 648 pp., 1998; 3rd ed., Wiley, 820 pp., 2012.

L. Thourel, *The Antenna* (Translation of *Les Antennes*, 1956), Wiley, 407 pp., 1960.

W. H. Watson, *The Physical Principles of Wave Guide Transmission and Antenna Systems*, Oxford U. Press, 208 pp., 1947.

W. L. Weeks, *Antenna Engineering*, McGraw-Hill, 370 pp., 1968.

H. Paul Williams, *Antenna Theory and Design*, Vol. 1: *Foundations of Antenna Theory*; Vol. 2: *The Electrical Design of Antennae*, 1st ed., Pitman, 1950; 2nd ed., Pitman, 795 pp., 1966.

E. Wolff, *Antenna Analysis*, Wiley, 514 pp., 1966.

H.4   BOOKS ON ANTENNAS WITH PROPAGATION

S. Cloude, *An Introduction to Electromagnetic Wave Propagation & Antennas*, Springer, 176 pp., 1996.

R. E. Collin, *Antennas and Radiowave Propagation*, McGraw-Hill, 508 pp., 1985.

John Griffiths, *Radio Wave Propagation and Antennas: An Introduction*, Prentice-Hall, 384 pp., 1987.

A. Harish and M. Sachidananda, *Antennas and Wave Propagation,* Oxford, 402 pp., 2007.

E. Jordan, *Electromagnetic Waves and Radiating Systems*, 1st ed., Prentice-Hall, 1950; E. Jordan and K. Balmain, 2nd ed., Prentice-Hall, 753 pp., 1968.

G. Raju, *Antennas and Wave Propagation*, Akhil Books, 402 pp., 2007.

S. R. Saunders, *Antennas and Propagation for Wireless Communication Systems*, 1st ed., Wiley, 409 pp., 1999; 2nd ed. (with A. Aragon-Zavala), Wiley, 524 pp., 2007.

K. Siwiak, *Radiowave Propagation and Antennas for Personal Communication*, 1st ed., Artech House, 320 pp., 1995; 2nd ed., Artech House, 418 pp., 1998; 3rd ed., (with Y. Bahreini), Artech House, 464 pp., 2007.

W. Sinnema, *Electronic Transmission Technology: Lines, Waves, and Antennas*, 1st ed., Prentice-Hall, 264 pp., 1979; 2nd ed., Prentice-Hall, 400 pp., 1988.

Albert Smith, *Radio Frequency Principles and Applications: The Generation, Propagation, and Reception of Signals and Noise*, Wiley, 217 pp., 1998.

R. Vaughan and J. Bach Anderson, *Channels, Propagation, and Antennas for Mobile Communications*, IET, 753 pp., 2003.

J. Wait, *Introduction to Antennas and Propagation*, Peregrinus, 256 pp., 1986.

G. Welch, *Wave Propagation and Antennas*, Van Nostrand, 257 pp., 1958.

H.5   BOOKS ON ANTENNAS WITH OTHER TOPICS

R. G. Brown, R. A. Sharpe, and W. L. Hughes, *Lines, Waves, and Antennas: The Transmission of Electrical Energy*, 1st ed., Ronald Press, 279 pp., 1961; 2nd ed., 471 pp., 1973.

R. W. P. King, H. R. Mimno, and A. H. Wing, *Transmission Lines, Antennas and Waveguides*, McGraw-Hill, 347 pp., 1945.

J. A. Kuecken, *Antennas and Transmission Lines*, Howard Sams, 320 pp., 1969.

H. Mott, *Antennas for Radar and Communications: A Polarimetric Approach*, Wiley, 521 pp., 1992.

S. Roman, *Transmission Lines and Antennas*, Holt, Rinehart, Winston, 146 pp.,1969.

B. Rulf and G. Robertshaw, *Understanding Antennas for Radar, Communications, and Avionics*, Van Nostrand Reinhold, 335 pp.,1987.

K. Warnick and P. Russer, *Problem Solving in Electromagnetics, Microwave Circuits, and Antenna Design in Communication Engineering*, 2nd ed., Artech House, 688 pp., 2006.

H.6   HANDBOOKS AND GENERAL REFERENCE BOOKS ON ANTENNAS

Antenna Engineering Handbook
1st ed.: H. Jasik, Ed., McGraw-Hill, 1021 pp., 1961.
2nd ed.: R. C. Johnson and H. Jasik, Eds., McGraw-Hill, 1408 pp., 1984.

Chapters 24 through 41 of 2nd ed. also published as: R. C. Johnson and H. Jasik, Eds., *Antenna Applications Reference Guide*, McGraw-Hill, 480 pp., 1987.

3rd ed.: R. C. Johnson, Ed., McGraw-Hill, 1392 pp., 1993.

4th ed.: J. Volakis, Ed., McGraw-Hill, 1900 pp., 2007.

C. Balanis, Ed., *Modern Antenna Handbook*, Wiley, 1650 pp., 2008.

R. Bansal, Ed., *Handbook of Engineering Electromagnetics*, Marcel Decker, 706 pp., 2004.

J. Carr, *Practical Antenna Handbook*, 1st ed., TAB Books, 439 pp., 1989; 2nd ed., TAB Books, 560 pp., 1994; 3rd ed., McGraw-Hill, 574 p., 1998; 4th ed., McGraw-Hill, 609 pp., 2001.

Kai Chang, Ed., *Encyclopedia of RF and Microwave and Engineering*, Wiley, 6 volumes, 5832 pp., 2005.

Kai Chang, Ed., *Handbook of RF/Microwave Components and Engineering*, Wiley, 1784 pp., 2003.

Kai Chang, Ed., *Handbook of Microwave and Optical Components*, Vol. 1: Microwave Passive and Antenna Components, Wiley, 907 pp., 1989.

P. J. B. Clarricoats, Ed., *Advanced Antenna Technology*, Microwave Exhibitions and Publications, 454 pp., 1981.

L. B. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves*, Prentice-Hall, 888 pp., 1973.

L. C. Godara, Ed., *Handbook of Antennas in Wireless Communications*, CRC Press, 936 pp., 2002.

R. C. Hansen, Ed., *Microwave Scanning Antennas*, Academic Press, Vol. I–Apertures, 442 pp., 1964; Vol. II–Arrays, 400 pp., 1966; Vol. III–Frequency Scanning Arrays, 422 pp., 1966.

R. C. Johnson, *Designer Notes for Microwave Antennas*, Artech House, 195 pp., 1991.

E.C. Jordan, Ed., *Symposium on Electromagnetic Theory and Antennas*, Part 2, 669 pp., Pergamon Press, 1963.

Y. T. Lo and S. W. Lee, Eds., *Antenna Handbook: Theory, Applications, and Design*, Van Nostrand Reinhold, 2284 pp., 1988.

A. W. Rudge, K. Milne, A. D. Olver, and P. Knight, Eds., *The Handbook of Antenna Design*, Peregrinus, Vol. I, 708 pp., 1981 and Vol. II, 945 pp., 1983.

B. L. Smith and M. H. Carpentier, Eds., *The Microwave Engineering Handbook*, Vol. 2: *Microwave Circuits, Antennas and Propagation*, Van Nostrand Reinhold, 1993.

H.7   BOOKS ON ANTENNA MEASUREMENTS

H. Arai, *Measurement of Mobile Antenna Systems*, Artech House, 214 pp., 2001.

G. E. Evans, *Antenna Measurement Techniques*, Artech House, 229 pp., 1990.

S. Gregson, J. McCormick, and C. Panini, *Principles of Planar Near-Field Antenna Measurements*, IET, 424 pp., 2007.

J. E. Hansen, Ed., *Spherical Near-Field Antenna Measurements,* IET, 387 pp., 1988.

T. B. Hansen and A. D. Yaghijian, *Plane-Wave Theory of Time-Domain Fields: Near-Field Scanning Applications*, IEEE Press, 376 pp., 1999.

L. H. Hemming, *Electromagnetic Anechoic Chambers: A Fundamental Design and Specification Guide*, Wiley, 220 pp., 2002.

J. S. Hollis, T. J. Lyon, and L. Clayton, Eds., *Microwave Antenna Measurements,* Scientific-Atlanta, 604 pp., 1970.

IEEE Standard Test Procedures for Antennas, IEEE Standard 149-1979, IEEE, 143 pp., 1979.

D. Kerns, *Plane-Wave Scattering-Matrix Theory of Antennas and Antenna-Antenna Interactions*, U.S. Government Printing Office, 162 pp., 1981.

A. D. Kuz'min and A. E. Solmenovich, *Radio-Astronomical Methods of Antenna Measurement*, Academic Press, 182 pp., 1966.

E. Skomal and A. A. Smith, *Measuring the Radio Frequency Environment*, Van Nostrand Reinhold, Chapter 6, "Antennas," 45 pp., 1985.

Dan Slater, *Near-Field Antenna Measurements*, Artech House, 310 pp., 1991.

H.8   BOOKS ON SPECIFIC ANTENNA TOPICS

H.8.1   Wire Antennas

A. Banos, *Dipole Radiation in the Presence of a Conducting Half-Space*, Pergamon Press, 245 pp., 1966.

R. M. Bevensee, *Handbook of Conical Antennas and Scatterers*, Gordon and Breach Science, 173 pp., 1973.

M. L. Burrows, *ELF Communications Antennas*, Peregrinus, 245 pp., 1978.

J. Carr, *Antenna Toolkit*, 1st ed., Elsevier, 215 pp., 1997; 2nd ed., Elsevier, 253 pp., 2001.

J. E. Cunningham, *The Complete Broadcast Antenna Handbook: Design, Installation, Operation & Maintenance*, G/L Tab Books, 450 pp., 1977.

A. E. Harper, *Rhombic Antenna Design*, Van Nostrand, 111 pp., 1941.

R. W. P. King, *Tables of Antenna Characteristics*, IFI/Plenum, 393 pp., 1971.

R. W. P. King, *The Theory of Linear Antennas*, Harvard University Press, 944 pp., 1956.

F. M. Landstorfer and R.R. Sacher, *Optimisation of Wire Antennas*, Wiley, 174 pp., 1985.

D. B. Miron, *Small Antenna Design*, Elsevier, 283 pp., 2006.

L. A. Moxon, *HF Antennas for All Locations*, Radio Society of Great Britain, 260 pp., 1982.

B. D. Popovic, M. B. Dragovic, and A. R. Djordjevic, *Analysis and Synthesis of Wire Antennas*, Research Studies Press, 304 pp., 1982.

B. D. Popovic and B. M. Kolundzija, *Analysis of Metallic Antennas and Scatterers*, IET-Stevenage, 193 pp., 1994.

J. Raines, *Folded Unipole Antennas: Theory and Applications*, McGraw-Hill, 400 pp., 2007.

J. R. Wait, *Electromagnetic Radiation from Cylindrical Structures*, Pergamon, 200 pp., 1959.

M. M. Weiner, *Monopole Antennas*, Marcel Dekker, 740 pp., 2003.

M. M. Weiner, S. P. Cruze, C. C. Li, and W. J. Wilson, *Monopole Elements on Circular Ground Planes*, Artech House, 306 pp., 1987.

S. Uda and Y. Mushiake, *Yagi-Uda Antenna*, Saski Printing and Publishing Co., 198 pp., 1954.

American Radio Relay League (ARRL) publications on antennas:

The ARRL Antenna Book, frequent editions.

The ARRL Antenna Compendium, several volumes.

W. N. Caron, *Antenna Impedance Matching*, ARRL, 195 pp., 1989.

J. L. Lawson, *Yagi-Antenna Design*, ARRL, 196 pp., 1986.

D. B. Leeson, *Physical Design of Yagi Antennas*, ARRL, 340 pp., 1992.

M. Walter Maxwell, *Reflections: Transmission Lines and Antennas*, ARRL, 435 pp., 1990.

H.8.2   Arrays

N. Amitay, V. Galindo, and C. P. Wu, *Theory and Analysis of Phased Array Antennas*, Wiley, 443 pp., 1972.

A. K. Bhattacharyya, *Phased Array Antennas: Floquet Analysis, Synthesis, BFNs, and Active Array Systems*, Wiley, 496 pp., 2006.

E. Brookner, Ed., *Practical Phased Array Systems*, Artech House, 258 pp., 1991.

N. Fourikis, *Advanced Array Systems, Applications, and RF Technologies*, Academic Press, 323 pp., 2000.

N. Fourikis, *Phased Array-Based Systems and Applications*, Wiley, 426 pp., 1997.

R. C. Hansen, *Phased Array Antennas*, 1st ed., Wiley, 486 pp., 1998; 2nd ed., Wiley, 547 pp., 2009.

R. C. Hansen, Ed., *Significant Phased Array Papers*, Artech House, 271 pp., 1973.

Randy Haupt, *Antenna Arrays: A Computational Approach*, Wiley, 534 pp., Feb. 2010.

S. Haykin, Ed., *Array Signal Processing*, Prentice-Hall, 433 pp., 1985.

Don H. Johnson and D. E. Dugeon, *Array Signal Processing: Concepts and Techniques*, Prentice-Hall, 512 pp., 1993.

L. Josefsson and P. Persson, *Conformal Array Antenna Theory and Design*, Wiley–IEEE Press, 472 pp., 2006.

R. W. P. King, G. J. Fikioris, and R. B. Mack, *Cylindrical Antennas and Arrays*, Cambridge, 628 pp., 2002.

R. W. P. King, R. B. Mack, and S. S. Sandler, *Arrays of Cylindrical Dipoles*, Cambridge, 494 pp., 1968.

A. Kumar, *Antenna Design with Fiber Optics*, Artech House, 201 pp., 1996.

J. Litva and T. K-Y Lo, *Digital Beamforming and Wireless Communications*, Artech House, 301 pp., 1996.

M. T. Ma, *Theory and Application of Antenna Arrays*, Wiley, 413 pp., 1974.

M. T. Ma and D. C. Hyovalti, *A Table of Radiation Characteristics of Uniformly Spaced Optimum Endfire Arrays with Equal Sidelobes*, National Bureau of Standards, 65 pp., 1965.

B. Mahafza and A. Elsherbeni, *MATLAB Simulations for Radar and Systems Design*, Chapman & Hall (CRC Press), 682 pp., 2004.

R. J. Mailloux, *Phased Array Antenna Handbook*, 1st ed., Artech House, 534 pp., 1994. 2nd ed., Artech House, 496 pp., 2005.

M. Mikavica and A. Nesic, *CAD for Linear and Planar Antenna Arrays of Various Radiating Elements*, Disks and Users Manual, Artech House, 1991.

A. Mortazawi, T. Itoh, and J. Harvey, *Active Antennas and Quasi-Optical Arrays*, Wiley–IEEE, 337 pp., 1998.

B. Munk, *Finite Antenna Arrays and FSS*, Wiley, 357 pp., 2003.

A. A. Oliner and G. H. Knittel, Eds., *Phased Array Antennas*, Artech House, 394 pp., 1972.

J. N. Sahalos, *Orthogonal Methods for Array Synthesis*, Wiley, 216 pp., 2006.

J. P. Scherer, *LAARAN: Linear Antenna Array Analysis Software and User's Manual*, Artech House, 1989.

B. D. Steinberg, *Microwave Imaging with Large Antenna Arrays*, Wiley, 308 pp., 1983.

B. D. Steinberg, *Principles of Aperture and Array System Design, Including Random and Adaptive Arrays*, Wiley, 356 pp., 1976.

H. J. Visser, *Array and Phased Array Antenna Basics*, Wiley, 359 pp., 2005.

D. I. Voskresenskii, A. I. Grinev, and E. N. Voronin, *Electrooptical Arrays*, Springer-Verlag, 318 pp., 1989.

H.8.3   Broadband and Ultrawideband Antennas

B. Allen, M. Dohler, E. Okon, W. Malik, A. Brown, and D. Edwards, Eds., *Ultra Wideband Antennas and Propagation for Communications*, *Radar, and Imaging*, Wiley, 475 pp., 2007.

Z. N. Chen and M. Y. W. Chia, *Broadband Planar Antennas*, Wiley, 243 pp., 2006.

R. G. Corzine and J. A. Mosko, *Four-Arm Spiral Antennas*, Artech House, 265 pp., 1989.

Y. Mushiake, *Self-Complementary Antennas*, Springer-Verlag, 128 pp., 1996.

H. Nakano, *Helical and Spiral Antennas — A Numerical Approach*, Wiley–Research Studies Press, 261 pp., 1987.

V. Rumsey, *Frequency Independent Antennas*, Academic Press, 150 pp., 1966.

Hans Schantz, *The Art and Science of Ultrawideband Antennas*, Artech House, 331 pp., 2005.

Carl E. Smith, *Log Periodic Antenna Design Handbook*, Smith Electronics, Inc., Cleveland, 1966.

D. H. Werner and Raj Mittra, Eds., *Frontiers in Electromagnetics*, IEEE Press, 787 pp., 2000.

H.8.4   Microstrip Antennas and Printed Antennas

I. J. Bahl and P. Bhartia, *Microstrip Antennas*, Artech House, 348 pp., 1980.

R. Bancroft, *Microstrip and Printed Antenna Design*, Noble, 250 pp., 2004; 2nd ed., Scitech Pub., 300 pp., 2009.

P. Bhartia, K. V. S. Rao, and R. S. Tomar, *Millimeter-Wave Microstrip and Printed Circuit Antennas*, Artech House, 322 pp., 1991.

A. Bhattacharyya, *Electromagnetic Fields in Multilayered Structures: Theory and Applications*, Artech House, 179 pp., 1994.

G. Dubost, *Flat Radiating Dipoles and Applications to Arrays*, Wiley–Research Studies Press, 103 pp., 1981.

D. G. Fang, *Antenna Theory and Microstrip Antennas*, CRC Press, 311 pp., 2009.

R. Garg, P. Bhartia, I. Bahl, and A. Ittipiboon, *Microstrip Antenna Design Handbook*, Artech House, 845 pp., 2001.

K. C. Gupta and A. Benella, Eds., *Microstrip Antenna Design*, Artech House, 395 pp., 1988.

N. Herscovici, *CAD of Aperture-Fed Microstrip Transmission Lines and Antennas*, Artech House, 1996.

J. R. James and P. S. Hall, Eds., *Handbook of Microstrip Antennas*, Vols. I and II, IET-Peregrinis, 1350 pp., 1989.

J. R. James, P. S. Hall, and C. Wood, *Microstrip Antenna Theory and Design*, IET-Peregrinus, Electromagnetic Waves Series 12, 290 pp., 1981.

G. Kumar and K. P. Ray, *Broadband Microstrip Antennas*, Artech House, 424 pp., 2003.

K. F. Lee and W. Chen, Eds., *Advances in Microstrip and Printed Antennas*, Wiley, 599 pp., 1997.

D. M. Pozar and D. H. Schaubert, Eds., *Microstrip Antennas: The Analysis and Design of Microstrip Antennas and Arrays*, Wiley-IEEE Press, 431 pp., 1995.

R. A. Sainati, *CAD of Microstrip Antennas for Wireless Applications*, Artech House, 255 pp., 1996.

R. B. Waterhouse, *Microstrip Patch Antennas: A Designer's Guide*, Kluwer, 421 pp., 2003.

K-L Wong, *Compact and Broadband Microstrip Antennas*, Wiley, 327 pp., 2002.

K-L Wong, *Design of Nonplanar Microstrip Antennas and Transmission Lines*, Wiley, 372 pp., 1999.

J.-F. Zurcher and F. Gardiol, *Broadband Patch Antennas*, Artech House, 209 pp., 1995.

H.8.5   Reflector and Lens Antennas

J. Baars, *The Parabaloidal Reflector Antenna in Radio Astronomy and Communication: Theory and Practice*, Springer, 253 pp., 2007.

J. Brown, *Microwave Lenses*, Wiley, 125 pp., 1953.

B. S. Cornbleet, *Microwave Optics: The Optics of Microwave Antenna Design*, Academic Press, 416 pp., 1976.

A. Fradin, *Microwave Antennas*, Pergamon, 668 pp., 1961.

J. Y. Guo and S. K. Barton, *Fresnel Zone Antennas*, Springer, 120 pp., 2002.

H. D. Hristov, *Fresnel Zones in Wireless Links, Zone Plate Lenses and Antennas*, Artech House, 323 pp., 2000.

J. Huang and J. A. Encinar, *Reflectarray Antennas*, Wiley–IEEE Press, 216 pp., 2007.

T. Kitsuregawa, *Advanced Technology in Satellite Communication Antennas*, Artech House, 515 pp., 1990.

Roy Levy, *Structural Engineering of Microwave Antennas for Electrical, Mechanical, and Structural Engineers*, IEEE Press, 354 pp., 1996.

A. W. Love, Ed., *Reflector Antennas*, Wiley–IEEE Press, 427 pp., 1978.

W. V. T. Rusch and P. D. Potter, *Analysis of Reflector Antennas*, Academic Press, 178 pp., 1970.

Craig R. Scott, *Modern Methods of Reflector Antenna Analysis and Design*, Artech House, 130 pp., 1990.

C. J. Sletten, Ed., *Reflector and Lens Antennas: Analysis and Design Using Personal Computers*, Artech House, 432 pp., 1988. *Reflector and Lens Antennas: Analysis and Design Using PCs, Software and Users Manual*, Version 2.0, 1991.

B. S. Westcott, *Shaped Antenna Reflector Design*, Wiley–Research Studies Press, 186 pp., 1983.

P. J. Wood, *Reflector Antenna Analysis and Design*, Peregrinus–IET, 221 pp., 1986.

H.8.6   Horns/Feeds

P. J. B. Clarricoats and A. D. Olver, *Corrugated Horns for Microwave Antennas*, IET, 232 pp., 1983.

E. V. Jull, *Aperture Antennas and Diffraction Theory*, IET-Peregrinus, 173 pp., 1981.

A. W. Love, Ed., *Electromagnetic Horn Antennas*, IEEE Press, 451 pp., 1976.

A. D. Olver, P. J. B. Clarricoats, A. A. Kishk and L. Shafai, *Microwave Horns and Feeds*, IET, 486 pp., 1994.

I. Shumljansky, *Horn Radiators of Complex Configuration*, World Scientific, 160 pp., 1993.

J. Uher, J. Bornemann, and U. Rosenberg, *Waveguide Components for Antenna Feed Systems: Theory and CAD*, Artech House, 457 pp., 1993.

H.8.7   Dielectric Antennas

J. Bach Anderson, *Metallic and Dielectric Antennas*, Polyteknisk Forlag, 217 pp., 1971.

Rajeswari Chatterjee, *Dielectric and Dielectric-Loaded Antennas*, Wiley–Research Studies Press, 1985.

D. G. Kiely, *Dielectric Aerials*, Methuen, 132 pp., 1953.

K. M. Luk and K. W. Leung, Eds., *Dielectric Resonator Antennas*, Research Studies Press, 388 pp., 2003.

A. Petosa, *Dielectric Resonator Antenna Handbook*, Artech House, 336 pp., 2007.

C. Salema, C. Fernandes, and R. K. Jha, *Solid Dielectric Horn Antennas*, Artech House, 244 pp., 1998.

H.8.8   Traveling Wave Antennas

C. H. Walter, *Traveling Wave Antennas*, McGraw-Hill, 429 pp., 1965.

H.8.9   Slot Antennas

James R. Wait, *Electromagnetic Radiation from Cylindrical Structures*, IEE-Peregrinus, 202 pp., 1988.

H.8.10   Antennas in (or near, or filled with) Materials

J. Galejs, *Antennas in Inhomogeneus Media*, Pergamon Press, 294 pp., 1969.

Volkert W. Hansen, *Numerical Solution of Antennas in Layered Media*, Wiley, 294 pp., 1989.

R. W. P. King and G. S. Smith, *Antennas in Matter: Fundamentals, Theory, and Applications*, MIT Press, 868 pp., 1981.

B. Munk, *Metamaterials: Critique and Alternatives*, Wiley, 189 pp., 2009.

D. Liu, U. Pfeiffer, J. Grzyb, and B. Gaucher, Eds., *Advanced Millimeter-Wave Technologies: Antennas, Packaging, and Circuits*, Wiley, 850 pp., 2009.

Fan Yang and Y. Rahmat-Samii, *Electromagnetic Band Gap Structures in Antenna Engineering*, Cambridge U. Press, 280 pp., 2008.

H.8.11   Electrically Small Antennas

K. Fujimoto, A. Henderson, K. Hirasawa, and J. James, *Small Antennas*, Wiley, 300 pp., 1987

R. C. Hansen, *Electrically Small, Superdirective, and Superconducting Antennas*, Wiley, 168 pp., 2006.

J. Volakis, C-C Chen, and K. Fujimoto, *Small Antennas: Miniaturization Techniques & Applications*, McGraw-Hill, 428 pp., 2010.

H.8.12   Miscellaneous Antenna Topics

A. Kumar and H. D. Hristov, *Microwave Cavity Antennas*, Artech House, 471 pp., 1989.

D. R. Rhodes, *Synthesis of Planar Antenna Sources*, Clarendon Press, 210 pp., 1974.

Y. Shifrin, *Statistical Antenna Theory*, Golem Press, 370 pp., 1971.

H.9   BOOKS ON ANTENNAS FOR SPECIFIC APPLICATIONS

H.9.1   Antennas for Mobile and Personal Wireless Communications

Z. N. Chen, *Antennas for Portable Devices*, Wiley, 290 pp., 2007.

Z. N. Chen and K-M Luk, *Antennas for Base Stations in Wireless Communications*, McGraw-Hill, 412 pp., 2009.

K. Fujimoto and J. R. James, Eds., *Mobile Antenna Systems Handbook*, 1st ed., Artech House, 617 pp., 1994; 2nd ed., Artech House, 680 pp., 2001; 3rd ed., Artech House, 710 pp., 2008.

B. L. Grananstein, *Physical Principles of Wireless Communications*, Auerbach Publications, 288 pp., 2007.

P. S. Hall and Y. Hao, *Antennas and Propagation for Body-Centric Wireless Communications*, Artech House, 291 pp., 2006.

K. Hirasawa and M. Haneishi, *Analysis, Design, and Measurement of Small and Low-Profile Antennas*, Artech House, 270 pp., 1992.

H. Hu, Y. Zhang, and J. Luo, *Distributed Antenna Systems: Open Architecture for Future Wireless Communications*, SciTech Publishing, 496 pp., 2007.

K.-C. Huang and D. J. Edwards, *Millimetre Wave Antennas for Gigabit Wireless Communications: A Practical Guide to Design and Analysis in a System Context*, Wiley, 286 pp., 2008.

A. Kumar, *Fixed and Mobile Terminal Antennas*, Artech House, 357 pp., 1991.

Preston E. Law, *Shipboard Antennas*, 1st ed., Artech House, 561 pp., 1983; 2nd ed., Artech House, 566 pp., 1986.

P. Neelakanta and R. Chatterjee, *Antennas for Information Super-Highways–An Exposition on Outdoor and Indoor Wireless Antennas*, Research Studies Press, 525 pp., 2003.

V. Rabinovich, N. Alexandrov, and B. Alkhateeb, *Automotive Antenna Design and Applications*, CRC Press, 330 pp., 2010.

D. A. Sanchez-Hernandez, Ed., *Multiband Integrated Antennas for 4G Terminals*, Artech House, 336 pp., 2008.

L. Setian, *Practical Communication Antennas with Wireless Applications*, Prentice-Hall, 569 pp., 1998.

D. Thiel and S. Smith, *Switched Parasitic Antennas for Cellular Communications*, Artech House, 254 pp., 2002.

R. Waterhouse, Ed., *Printed Antennas for Wireless Communications*, Wiley, 472 pp., 2007.

K-L Wong, *Planar Antennas for Wireless Communications*, Wiley, 301 pp., 2003.

H.9.2   Adaptive, Electronic Warfare, and Signal Processing Antennas, Including Smart Antennas

B. Allen and M. Ghavami, *Adaptive Array Systems: Fundamentals and Applications*, Wiley, 253 pp., 2005.

N. Blaunstein and C. Christodoulou, *Radio Propagation and Adaptive Antennas for Wireless Communication Links: Terrrestrial, Atmospheric, and Ionospheric*, Wiley, 614 pp., 2007.

J. Blogh and L. Hanzo, *Third-Generation Systems and Intellectual Wireless Networking: Smart Antennas and Adaptive Modulation*, IEEE Press–Wiley, 413 pp., 2002.

S. Chandran, Ed., *Adaptive Antenna Arrays*, Springer, 660 pp., 2004.

R. T. Compton, Jr., *Adaptive Antennas: Concepts and Performance*, Prentice-Hall, 448 pp., 1988.

A. Farina, *Antenna-Based Signal Processing Techniques for Radar Systems*, Artech House, 370 pp., 1992.

A. Fenn, *Adaptive Antennas and Phased Arrays for Radar and Communications*, Artech House, 394 pp., 2008.

L. C. Godara, *Smart Antennas*, CRC Press, 448 pp., 2004.

F. Gross, *Smart Antennas for Wireless Communications: With MATLAB*, McGraw-Hill, 270 pp., 2006.

M. A. Halim, *Adaptive Array Measurements in Communications*, Artech House, 308 pp., 2001.

L. Hanzo, J. Blogh, and S. Ni, *3G, HSPA, and FDD Networking: Smart Antennas and Adaptive Modulation*, 1st ed., Wiley, 413 pp., 2002; 2nd ed., Wiley, 596 pp., 2008.

P. Hoole, *Smart Antennas and Signal Processing for Communications, Biomedical, and Radar Systems*, WIT Press, 304 pp., 2001.

J. E. Hudson, *Adaptive Array Principles*, IEE–Peregrinus, 253 pp., 1981.

R. Janaswamy, *Radiowave Propagation, and Smart Antennas for Wireless Communications*, Kluwer Publishing, 312 pp., 2001.

V. Kuhn, *Wireless Communications over MIMO Channels: Applications to CDMA and Multiple Antenna Systems*, Wiley, 363 pp., 2006.

J. C. Liberti and T. S. Rappaport, *Smart Antennas for Wireless Communications*, Prentice-Hall, 376 pp., 1999.

M. Martone, *Multiantenna Digital Radio Transmission*, Artech House, 401 pp., 2002.

R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*, 1st ed., Wiley, 543 pp., 1980; 2nd ed. (with R. Haupt), 530 pp., 2011.

E. Nicolau and D. Zaharia, *Adaptive Arrays*, Elsevier, 297 pp., 1989.

G. T. Okamoto, *Smart Antenna Systems and Wireless LANs*, Kluwer Academic Publishers, 207 pp., 1999.

B. Pattan, *Robust Modulation Methods and Smart Antennas in Wireless Communications*, Prentice-Hall, 304 pp., 2000.

T. S. Rappaport, *Smart Antennas: Adaptive Arrays, Algorithms, and Wireless Position Location*, IEEE Press, 817 pp., 1998.

T. K. Sarkar, M. Wicks, M. Salazar-Palma, and R. Bonneau, *Smart Antennas*, Wiley, 452 pp., 2003.

T. K. Sarkar and M. Salazar-Palma, *Physics of Multiantenna Systems and Broadband Processing*, Wiley, 562 pp., 2008.

G. Tsoulos, Ed., *Adaptive Antennas for Wireless Communications*, Wiley, 764 pp., 2001.

M. Weiner, Ed., *Adaptive Antennas and Receivers*, CRC, 1204 pp., 2006.

A. Zooghby, *Smart Antenna Engineering*, Artech House, 324 pp., 2005.

H.9.3 Satellite Communications Antennas

R. Dybdal, *Communication Satellite Antennas: System Architecture, Technology, and Evaluation*, McGraw-Hill, 340 pp., 2009.

W. A. Imbriale, *Large Antennas of the Deep Space Network*, Wiley, 302 pp., 2003.

W. A. Imbriale and J. H. Yuen, Eds., *Spaceborne Antennas for Planetary Exploration*, Wiley, 570 pp., 2006.

T. Kitsuregawa, *Advanced Technology in Satellite Communication Antennas: Electrical and Mechanical Design*, Artech House, 415 pp., 1989.

R. Mittra, W. Imbriale, and, I. Maanders, Eds., *Satellite Communication Antenna Technology*, Elsevier, 652 pp., 1983.

D. Mudgway, *Big Dish: Building America's Deep Space Connection to the Planets*, U. of Florida Press, 253 pp., 2005.

T. Y. Otoshi, *Noise Temperature Theory and Applications for Deep Space Communications Antenna Systems*, Artech House, 300 pp., 2008.

D. Rogstad, A. Mileant, and T. Pham, *Antenna Arraying Techniques in the Deep Space Network*, Wiley, 166 pp., 2003.

H.10 BOOKS ON COMPUTATIONAL METHODS FOR ANTENNAS

H.10.1 Books Addressing Multiple Methods

Richard C. Booton, Jr., *Computational Methods for Electromagnetics and Microwaves*, Wiley, 182 pp., 1992.

W. C. Chew, J-M Jin, E. Michielssen, and J. Song, *Fast and Efficient Algorithms in Computational Electromagnetics*, Artech House, 931 pp., 2001.

D. B. Davidson, *Computational Electromagnetics for RF and Microwave Engineering*, Cambridge, 411 pp., 2005.

F. Gustrau and D. Manteuffel, *EM Modeling of Antennas and RF Components for Wireless Communication Systems*, Springer, 276 pp., 2006.

H. Kogure, Y. Kogure, and J. Rautio, *Introduction to Antenna Analysis Using EM Simulators*, Artech House, 250 pp., 2011.

R. Mittra, Ed., *Topics in Applied Physics*, Vol. 3, Numerical and Asymptotic Techniques in Electromagnetics, Springer-Verlag, 260 pp., 1974.

D. Poljak, *Electromagnetic Modelling of Wire Antennas Structures*, WIT Press, 162 pp., 2002.

M. N. O. Sadiku, *Numerical Techniques in Electromagnetics*, 1st ed., CRC Press, 690 pp., 1992; 2nd ed., CRC Press, 743 pp., 2000; 2nd ed., CRC Press, 743 pp., 2000; 3rd ed., CRC Press, 648 pp., 2009.

H. Visser, *Approximate Antenna Modeling for CAD*, Wiley, 320 pp., 2009.

H.10.2 Moment Methods and Software for Wire Antennas

NEC

J. L. Smith, *Basic NEC with Broadcast Applications*, Elsevier–Focal Press, 304 pp., 2008.

AWAS (Analysis of Wire Antennas and Scatterers)

A. R. Djordjevic, M. B. Bazdar, V. Petrovic, D. Olcan, T. K. Sarkar, and R. F. Harrington, *AWAS for Windows 2.0: Analysis of Wire Antennas and Scatterers, Software and User's Manual*, Artech House, 280 pp., 2002.

A. R. Djordjevic, M. B. Bazdar, T. K. Sarkar, and R. F. Harrington, *AWAS for Windows: Analysis of Wire Antennas and Scatterers*, Artech House, 1995.

A. R. Djordjevic, M. B. Bazdar, G. M. Vitosevic, T. K. Sarkar, and R. F. Harrington, *Analysis of Wire Antennas and Scatterers: Software and User's Manual*, Artech House, 228 pp., 1990.

MININEC

S. T. Li, J. W. Rockway, J. C. Logan, and D. W. S. Tam, *Microcomputer Tools for Communications Engineering*, Artech House, 269 pp., 1983.

J. Rockway, J. Logan, D. Tam, and S. Li, *The MININEC SYSTEM: Microcomputer Analysis of Wire Antennas*, Artech House, 256 pp., 1988.

WIPL-D

B. K. Kolundzija, J. S. Ognjanovic, and T. K. Sarkar, Book 1; B. K. Kolundzija and A. Djordjevic, *WIPL-D Microwave: Circuit and 3D EM Simulation for RF & Microwave Applications — Software and User's Manual*, Artech House, 2006.

B. K. Kolundzija, J. S. Ognjanovic, T. K. Sarkar, and R. F. Harrington, *WIPL: Electromagnetic Modeling of Composite Wire and Plate Structures — Software and User's Manual*, Artech House, 1995.

W. C. Gibson, *The Method of Moments in Electromagnetics*, CRC Press, 272 pp., 2008.

C. Hafner, *The Generalized Multipole Technique for Computational Electromagnetics*, Artech House, 297 pp., 1990.

Robert C. Hansen, *Moment Methods in Antennas and Scattering*, Artech House, 465 pp., 1990.

R. F. Harrington, *Field Computation by Moment Methods*, Macmillan, 229 pp., 1968.

B. Kolundzija and A. Djordjevic, *Electromagnetic Modeling of Composite Metallic and Dielectric Structures*, Artech House, 408 pp., 2002.

E. K. Miller, L. Medgyesi-Mitschang, and E. H. Newman, Eds., *Computational Electromagnetics: Frequency-Domain Method of Moments*, Wiley–IEEE Press, 508 pp., 1991.

J. Moore and R. Pizer, Eds., *Moment Methods in Electromagnetics*, Wiley–Research Studies Press, 398 pp., 1984.

B. D. Popovic, *CAD of Wire Antennas and Related Radiating Structures*, Wiley–Research Studies Press, 324 pp., 1991.

R. Mittra, Ed., *Computer Techniques for Electromagnetics*, Pergamon Press, 403 pp., 1973.

B. J. Strait, Ed., *Applications of the Method of Moments to Electromagnetic Fields*, SCEEE Press, 574 pp., 1980.

K. Umashankar and A. Taflove, *Computational Electromagnetics*, Artech House, 717 pp., 1993.

Johnson J. H. Wang, *Generalized Moment Methods in Electromagnetics*, Wiley, 553 pp., 1991.

H.10.3   Finite-Difference Time-Domain (FDTD) Methods

A. Elsherbeni and V. Demir, *The Finite Difference Time Domain Method for Electromagnetics: With Matlab Simulations*, SciTech Publishing, 500 pp., 2009.

K. S. Kunz and R. J. Luebbers, *Finite Difference Time Domain Method for Electromagnetics*, CRC Press, 448 pp., 1993.

W. Yu, R. Mittra, T. Su, Y Liu, and X. Yang, *Parallel Finite-Difference Time-Domain Method*, Artech House, 262 pp., 2006.

A. Taflove, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 1st ed., Artech House, 599 pp., 1995; A. Taflove and S. Hagness, 2nd ed., Artech House, 852 pp., 2000; A. Taflove and S. Hagness, 3rd ed., Artech House, 1006 pp., 2005.

H.10.4   High-Frequency Methods

V. A. Borovikov and B. Kinber, *Geometrical Theory of Diffraction*, IET, 390 pp., 1994.

R. H. Clarke and J. Brown, *Diffraction Theory and Antennas*, Wiley, 292 pp., 1980.

R. C. Hansen, Ed., *Geometric Theory of Diffraction*, IEEE Press, 406 pp., 1981.

G. L. James, *Geometrical Theory of Diffraction for Electromagnetic Waves*, P. Peregrinus, 253 pp., 1976; Revised ed., P. Peregrinus, 253 pp.,1980; 3rd ed., P. Peregrinus-IET, 293 pp., 1986.

D. A. McNamara, C. W. Pistorius, and J. A. G. Malherbe, *Introduction to the Uniform Geometrical Theory of Diffraction*, Artech House, 471 pp., 1990.

P. Ufimtsev, *Fundamentals of the Physical Theory of Diffraction*, Wiley, 329 pp., 2007.

P. Ufimtsev, *Theory of Edge Diffraction in Electromagnetics: Origination and Validation of the Physical Theory of Diffraction*, SciTech Publishing, 444 pp., 2009.

H.10.5   Finite Element Numerical Methods

J. Jin and D. Riley, *Finite Element Analysis of Antennas and Arrays*, Wiley, 452 pp., 2008.

J. Volakis, A. Chatterjee, and L. Kempel, *Finite Element Method Electromagnetics: Antennas, Microwave Circuits, and Scattering Applications*, Wiley, 344 pp., 1998. Update volume: J. Volakis, K. Sertel, and B. Usner, *Frequency Domain Hybrid Finite Element Methods for Electromagnetics*, Morgan & Claypool Publishers, 148 pp., 2006.

H.10.6   Other Numerical Methods

C. Christodoulou and M. Georgiopoulos, *Applications of Neural Networks in Electromagnetics*, Artech House, 512 pp., 2001.

R. L. Haupt and D. H. Werner, *Genetic Algorithms in Electromagnetics*, Wiley, 301 pp., 2007.

Y. Rahmat-Samii and E. Michielson, *Electromagnetic Optimization by Genetic Algorithms*, Wiley, 480 pp., 1999.

H.10.7   Software for General Antenna Applications

L. Diaz and T. Milligan, *Antenna Engineering Using Physical Optics: Practical CAD Techniques and Software*, Artech House, 343 pp., 1996.

A. Elsherbeni and M. Inman, *Antenna Design and Visualization Using MATLAB*, SciTech Publishing, 2006.

J. A. Kuecken, *Exploring Antennas and Transmission Lines by Personal Computer*, Van Nostrand, 318 pp., 1986.

S. Makarov, *Antenna and EM Modeling with Matlab*, Wiley, 273 pp., 2002.

D. Pozar, *PCAAD — Personal Computer Aided Antenna Design*, Antenna Design Associates, updated versions released periodically.

D. Pozar, *Antenna Design Using Personal Computers*, Artech House, 141 pp., 1985.

H.11   BOOKS ON TOPICS CLOSELY RELATED TO ANTENNAS

H.11.1   Polarization

P. Beckmann, *The Depolarization of Electromagnetic Waves*, Golem Press, 214 pp., 1968.

H. Mott, *Polarization in Antennas and Radar*, Wiley, 297 pp., 1986.

W. L. Stutzman, *Polarization in Electromagnetic Systems*, Artech House, 239 pp., 1993.

H.11.2   **Radomes**

D. Lynch, *Introduction to RF Stealth*, Scitech, 573 pp., 2004.

H. L. Hirsch and D. C. Grove, *Practical Simulation of Radar Antennas and Radomes*, Artech House, 287 pp., 1987.

D. J. Kozakoff, *Analysis of Radome Enclosed Antennas*, Artech House, 265 pp., 1997.

J. D. Walton, Jr., Ed., *Radome Engineering Handbook*, Marcel Dekker, 592 pp., 1970.

H.11.3   **Artificial Materials Useful in Antennas**

C. Caloz and T. Itoh, *Electromagnetic Metamaterials: Transmission Line Theory and Microwave Applications*, Wiley–IEEE Press, 352 pp., 2005.

F. Capolino, *Artificial Materials Handbook*, Vol. I: Theory and Phenomena of Artificial Materials, 800 pp., 2008; Vol. II: Applications of Artificial Materials, 800 pp., 2008.

G. V. Eleftheriandes and K. G. Balmain, *Negative-Refraction Metamaterials: Fundamental Principles and Applications*, Wiley–IEEE Press, 440 pp., 2005.

N. Engheta, and R. Ziolkowski, Eds., *Metamaterials: Physics and Engineering Explorations*, Wiley, 414 pp., 2006.

B. Munk, *Frequency Selective Surfaces: Theory and Design*, Wiley, 440 pp., 2000.

H.11.4   **Other**

K. C. Gupta and P. S. Hall, Eds., *Analysis and Design of Integrated Circuit-Antenna Modules*, Wiley, 440 pp., 1999.

S. E. Lipsky, *Microwave Passive Direction Finding*, Wiley, 298 pp., 1987.

T. Macnamara, *Handbook of Antennas for EMC*, Artech House, 328 pp., 1995.

T. Macnamara, *Introduction to Antenna Placement and Installation*, Wiley, 394 pp., 2010.

J. A. Navarro and Kai Chang, *Integrated Active Antennas and Spatial Power Combining*, Wiley, 368 pp., 1996.

G. A. Savitskii, *Calculations for Antenna Installations*, Amerind Pub., 181 pp., 1982.

G. Wiskin, R. Manton, and J. Causebrook, *Masts, Antennas, and Service Planning*, Focal Press, 233 pp., 1992.

B. S. Yarman, *Design of Ultra Wideband Antenna Matching Networks*, Springer, 310 pp., 2008.

Index

810

Symbols

SYMBOLS[1]

a = wire radius [m]

$\mathbf{A}$ = magnetic vector potential [V-s/m]

AF = array factor

A_e = effective aperture = $e_r A_{em}$ [m^2]

A_{em} = maximum effective aperture [m^2]

A_p = physical aperture = physical area of an aperture antenna [m^2]

AR = polarization ellipse axial ratio ($1 \leq |AR|$)

b = mean radius of a loop [m]

B = fractional bandwidth = BW/f_C

B_p = percent bandwidth = $B \times 100 = 2\dfrac{f_U - f_L}{f_U + f_L} \times 100$ [%]

B_r = ratio bandwidth = f_U/f_L

BW = antenna bandwidth = $f_U - f_L$ [Hz]

BWFN = beamwidth between first nulls [rad]

c = velocity (speed) of electromagnetic waves in free space $\approx 3 \times 10^8$ m/s

c = directivity factor used in $D = c\ 2\ L/\lambda$

d = interelement spacing in an equally spaced array [m]

D = diameter of an antenna [m]

D = directivity; directivity in dB = 10 log D

D_u = directivity of a uniform aperture (constant amplitude and phase)

e_r = radiation efficiency

$\mathbf{E}$ = electric field vector [V/m]

E_a = aperture electric field [V/m]

E^i = incident electric field vector [V/m]

EIRP = effective (or equivalent) isotropically radiated power [W]

f = frequency [Hz]

$f_L\ (f_U)$ = lower (upper) operating frequency of an antenna [Hz]

$f\ (\ )$ = pattern factor or array factor normalized to unity maximum

$F(\ )$ = radiation pattern normalized to unity maximum

$g\ (\ )$ = element pattern in a continuous current antenna normalized to unity maximum

$g_a\ (\ )$ = pattern of a single element in an array normalized to unity maximum

$g_{ae}\ (\ )$ = average active element pattern normalized to a typical element in the array with all elements active

$g_i\ (\ )$ = isolated-element pattern normalized to unity maximum

G = (maximum) gain of an antenna; gain in dB = 10 log G

G_o = gain with no phase error

G_{ref} = gain of a reference antenna

G_r = gain of a receive antenna

G_t = gain of a transmit antenna

h = height of an antenna above ground [m]

$\mathbf{h}$ = vector effective length of an antenna [m] = $h\ \hat{\mathbf{h}}$

H = magnetic field vector [A/m]

HP = half-power beamwidth [rad]; HP_{deg} = half-power beamwidth [deg]

$HP_{E°}$ ($HP_{H°}$) = half power beamwidth in degrees in *E*-plane (*H*- plane) [deg]

I = (electric) current [A]

I_{in} = current at input terminals of an antenna [A]

I_m = maximum current appearing on an antenna [A]

I^m = magnetic current [V]

J$_S$ = surface current density [A/m]

k = beamwidth factor used in HP = *k* λ/L

K = antenna factor [m^{-1}]

L = inductance [H]

L = mean length of a wire loop [m]

L = length of a line source antenna or linear array [m]

M = magnetic current density vector [V/m^2]

n = index of refraction (except in Chap. 16)

n̂ = unit normal vector

p = polarization efficiency; polarization mismatch factor ($0 \leq p \leq 1$)

P = power radiated [W]

P_D = power delivered from an antenna to a load (after impedance and polarization mismatches) [W]

P_{in} = power into an antenna (includes ohmic loss) [W]

P_o = power lost to ohmic resistance [W]

P_N = power available due to noise [W]

P_r = power available from a receiving antenna if properly matched [W]

P_t = power from a transmitter delivered to an antenna [W]

q = charge [C]

q = impedance mismatch factor ($0 \leq q \leq 1$)

Q = radiation quality factor

r = spherical coordinate radius [m]

r′ = distance from a point in a source distribution to an observation point P [m]

r_p = distance from the coordinate origin to the observation point P [m]

r_{ff} = distance to far field from a coordinate origin centered on an antenna [m]

R = distance between two points in three dimensional space [m]

R = line of sight separation on a radio link [m]

R_A = input resistance at antenna terminals [Ω]

R_{ri} = radiation resistance referred to the input terminals [Ω]

R_S = surface resistance [Ω]

s = surface area (except Chap. 9, where S is used) [m^2]

S = time-average power flux density [W/m^2]

S = Poynting vector [W/m^2]

SLL = side lobe level relative to main beam peak; side lobe level in dB = 20 log (SLL)

t = time [s]

T_A = antenna output noise temperature [K]

T_r = receiver input noise temperature [K]

U = radiation intensity [W/steradian]

v = volume (except Chap. 9, where *V* is used)

v = phase velocity [m/s]

V_A = open circuit voltage across the terminals of an antenna [V]

VSWR = voltage standing wave ratio (VSWR $\geq$ 1)

X_{in} = input reactance at antenna terminals [Ω]

XPOL = cross polarization level (usually relative to pattern main beam peak)

Z = impedance [Ω]

Z_A = impedance at input to antenna [Ω]

Z_m = active impedance of mth element in an array with all elements excited [Ω]

Z_{mn} = mutual impedance between the mth and n^{th} elements of an array antenna [Ω]

GREEK SYMBOLS

α = interelement phase shift in an array; the amount an element leads its left
    neighboring element [rad]

α = pitch angle of a helix antenna [rad]

α_h = turn-to-turn phase shift in a helix antenna in the axial mode [rad]

β = phase constant = $2\pi/\lambda$ [rad/m]

β_g = phase constant in a waveguide [rad/m]

β_o = excitation phase per meter across a line source [rad/m]

Γ = voltage reflection coefficient

Δz = length of an ideal dipole [m]

ε = polarization ellipticity angle [rad] = $\cot^{-1}(-AR)$

ε = permittivity [F/m]

ε_r = relative permittivity = $\varepsilon/\varepsilon_o$ where ε_o = permittivity of free space
    = $10^{-9}/36\pi$ F/m

ε' = effective permittivity = $\varepsilon - j\,(\sigma/\omega)$ [F/m]

ε_{ap} = aperture efficiency

ε_t = aperture taper efficiency

η = intrinsic impedance = $\sqrt{\dfrac{\mu}{\varepsilon}}$ [Ω]

θ_o = angle of the pattern main beam peak relative to the endfire direction in one-dimensional
    antennas and relative to broadside for multidimensional antennas [rad]

λ = free-space wavelength [m]

μ = permeability [H/m]

μ_r = relative permeability = μ/μ_o where μ_o = permeability of free space
    = $4\pi \times 10^{-7}$ H/m

ρ = charge density [C/m^3]

σ = conductivity [S/m]

τ = polarization tilt angle [rad]

Φ = (electric) scalar potential [V]

ψ = array factor argument = $\beta d \cos\theta + \alpha$ [rad]

ω = radian frequency = $2\pi f$ [rad/s]

Ω_A = beam solid angle [steradians]

Ω_M = main beam solid angle [steradians]