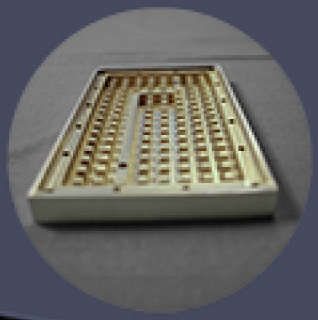
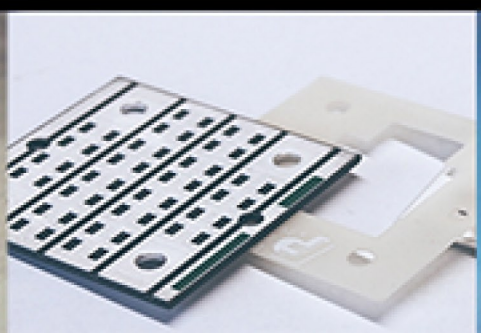
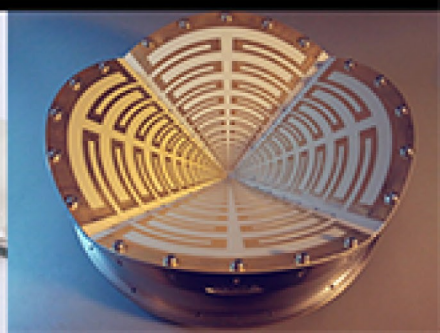


FOUNDATIONS of ANTENNA ENGINEERING

A Unified Approach for Line-Of-Sight And Multipath

Per-Simon Kildal





Foundations of Antenna Engineering: A Unified Approach for Line-of-Sight and Multipath

Per-Simon Kildal
Chalmers University of Technology

PDF created October 14, 2015

KILDAL

Published by Kildal Antenn AB in April 2015.

There are updated free PDF versions of the book available subject to registration via `www.kildal.se`, as well as additional supporting material such as the MATLAB handbook and presentations slides.

Lecturers who want to provide printed copies of the book to their students can get a free license to do so by registering via the same web page. They will then also be able to download a grey-scale PDF version of the book which is more suitable for printing than the color PDF version.

ISBN: 978-91-637-8515-3

To Cécile, Madeleine and Susanne

Contents

Preface

General part

I started to write this book in 1994 when I took over the teaching of the antenna course at Chalmers University of Technology (Chalmers). That is 20 years ago now, and a lot of developments have happened with the antenna technology during these years, as a result of the digital revolution and the growth of the mobile wireless communications. My research group has been a major front line player in this development, and my textbook has gradually evolved together with the technological development. My intention is that it shall continue to evolve in the years to come. There are more opportunities in the wireless revolution that requires rethinking and reordering, and not the least cleaning up the terminology. This book is the start of such a process.

The antenna engineer of today must know the basic antenna theories and principles in order to find good and innovative initial antenna geometries, he or she must know the electromagnetic (EM) simulation tools in order to verify the initial design, and he or she must know how to optimize the antenna on system level in order to provide the best overall antenna solution. The latter requires also some knowledge of propagation, signal processing, and communication and radar systems. Therefore, my textbook puts a lot of attention to the *characterization on system level*. This has been done by introducing a reference environment for multipath, referred to as *Rich Isotropic MultiPath* (*RIMP*) (Chapter 3). RIMP is a logical compliment to the normal "free space" or anechoic reference environment (Chapter 2), which in propagation terminology could be called a pure Line-Of-Sight (LOS environment. There is one chapter associated with each of these reference environments. Chapter 3 naturally handles MIMO arrays, i.e., the multi-port antenna arrays used in the digital mobile communications systems. Such arrays link to the classical array antennas in Chapter 10 via the embedded element far-field function, and the embedded element efficiency. The latter represents a fundamental limitation on gains of dense arrays and explains the classical *element-gain paradox* in classical arrays (Chapter 11).

When introducing new and modern material in a textbook, some traditional parts need to be removed or compacted. I did this very early during the 20 years writing period, by building up a new EM foundation for antenna theory based on *radiation from incremental current sources* (Chapter 4). Note the term incremental. This send the right signals to the reader: All radiation problems can be reduced to an integral or summation over the incremental sources. To make this complete I decided to introduce the Huygens source, which is a specific linear combination of an electric and magnetic current. *The Huygens source* plays a major role in determining the fundamental limitations of the directivity on small antennas (Chapter 11).

The terms that characterize this textbook the best are *rethinking and reordering*. Several new terms have been defined, and they have during the years gradually been accepted. The most important being the fact that wireless devices with multiport antennas can be characterized in chambers with reflecting metal walls (reverberation chambers). The reverberation chambers emulate RIMP. This is a complete contrast to traditional anechoic "free-space like" chambers, so it took 10-14 years before this was generally accepted, after it was introduced in year 2000. This introduction has together with the definition of a digital threshold receiver (Chapter3) opened up a complete new world of opportunities. By the latter definition the rethinking can continue.

The intermediate chapters between Chapters 4 and 11 contain *the theory of the most common antenna types*. These chapters are supported by a MATLAB handbook, which contains the programs that have been used to produce all the design curves. Therefore, the students can use them during their work with the exercises, and later in professional work situations to make initial antenna designs. The theories in these chapters contain in principle classical design formulas, but the derivations have been modernized. This was done quite early by starting from the incremental currents (introduced in Chapter 4), and using the modern *Method of Moments* (MoM) introduced in Chapter 4 as a general concept for deriving traditional impedance formulas. This can be achieved by using one physical basis function. The MoM is generally often referred to as *Galerkin's method*. I have myself a background in MoM via R.F. Harrington's book, and I have also several papers myself related to MoM combined with spectral domain approaches (Chapter 6). I find the MoM theory very good for understanding radiation phenomena. The whole Chapter 4 is built up using MoM as a way of thinking.

I have at the end of this preface, proposed a list of learning goals for students using this textbook.

Previous book versions

The book was published first time in 2000 by a publisher in Sweden, who has no more rights to the material. The present version is updated and the major Chapters 3 and 11 have been added. This update started in 2009, and the manuscript has since then been used as a compendium in my *Antenna Engineering* course at Chalmers.

About the citations

The reader will from the previous versions of it understand that the textbook is very special, because it contains a lot of rethinking. As a result, there are many self-citations, with the purpose of having scientific documentation of all the rethinking. This is of course sensitive. Therefore, I have asked my friend Professor Stefano Maci to add references to alternative well proven and accepted approaches, at the end of almost all chapters under the title *Complementary Comments*. He is a Director of the European School of Antennas and as such he is very well aware of most approaches used in antenna theory and design, so he is the right person to do this. I hope that this way of giving credit to others' approaches will make my book more acceptable.

Acknowledgments

Several persons have been of help in developing and giving feedback related to this material during the years, most of them once being a member of my antenna research group at Chalmers and having a PhD from me. I am thankful to them all. I will here only specifically mention the major later contributions: My younger daughter Susanne transferred the whole manuscript from Framemaker to LaTeX in 2012. This included also an improvement and extention of Chapter 3 and the new Chapter 11. Dr Nima Jamaly generously offered himself to go carefully through the new manuscript with pencil, pointing out typographical errors and places where the explanations could be better, and proposing many specific language improvements. He did a fantastic job, being thorough and accurate. Dr Ulf Carlberg is the author of the latest version of the MATLAB handbook. He rebuilt and improved everything and extended it to be able to produce 16 pages with illustrative 3D color graphs of 3D radiation patterns and contour *uv*-plots. My older daughter Madeleine implemented all these improvements in the LaTeX files, updated the book with the new color graphs, and made corrections in page layout, text and figures, in particular in Chapter 3 (multipath) and 11 (fundamental limitations). The nice cover page of the 2014 compendium was designed by Philip Sjöqvist, and this is also used in the final version. I am also very thankful to Master student Wenjie Yu who in Autumn 2013 gave me a list of 36 small errors in the equations. It was a remarkable achievement to go through everything so carefully. I should also specifically mention 2 Master students from 2003, Roman Iustin and Lorens Al-Hazam, who really went through the book in detail and gave good feedback. As a result they got also very high scores on the exam. No-one has ever done so good exams in the course as they then did, both of them.

The following persons have helped finalizing the book manuscript for printing during Autumn 2014: Madeleine helped with some MATLAB programs and graphs and implemented all changes carefully in the book using LaTeX Dr Xiaoming Chen helped to further extend Chapter 3, Abbas Vosoogh helped with producing new results on realized gain limitations of classical arrays in Chapter 11, and I had several long discussions with Professor Stefano Maci in particular on the material in Chapters 4 and 11.

I got very good help with the final proofreading of different chapters in February 2015 from Professor Jian Yang, Dr Rob Maaskant, Dr Ashraf Uz Zaman and Dr Andrés Alayón Glazunov, all members of my research group.

The MATLAB handbook was originally written in Mathcad by myself. Ahmed Hussain transformed the Mathcad programs to MATLAB in 2009, with an emphasis on the pedagogical aspects. I am thankful to Professor Zvonimir Sipus for his thorough discussions and help with in particular Chapter 6 during the first years. Jan Olof Yxell at Chalmers made several of the original drawings in the book. I am thankful for numerous discussions with Professor Jan Carlsson about many topics in the book, and in particular for writing Appendix F with matrix descriptions on N-port circuits.

Some more characteristics of the textbook

In antennas as well as in other sciences, *knowledge and know-how* is important. However, there is another art which is even more important, and without which we cannot create new knowledge: *imagination*. Imagination is defined in Webster's dictionary as "the act or

power of creating mental images of what has never been actually experienced, or of creating new images or ideas by combining previous experiences". My emphasis when writing the book has been to make compact vector formulations of the theory and to illustrate the interpretation of the theory. Hopefully, this will enhance the reader's imagination. I make frequent use of equivalent sources, making it easier to generate mental images of complicated equations.

Modern antenna designs are more and more often based on computer-aided design (CAD). This involves running computer programs that have been bought from a commercial software vendor or acquired with or without cost from specialists in the field. Therefore, the antenna designer of today needs an overall knowledge and understanding of the characteristics and limitations of the different analysis methods. In addition, some analytical skill is needed in order to make *fast initial designs* and to develop simple solutions for special cases for verification of the computations. Also, there may often be a need to *combine different analysis methods* and software. In such cases, the designer needs enough analytical and theoretical background to do the combination correctly. In order to satisfy these needs I have tried to make the mathematical formulations as compact and simple as possible. This has been done by consequently using the vector notation, as vectors are very easily programmed nowadays. In addition, the mathematical notation is considerably simplified by making use of equivalent currents. The *equivalence theorem* makes it easier to structure a complicated antenna problem into more manageable smaller subproblems. The radiation field can always be found by integration over a given distribution of impressed and induced currents located in free space. These currents may be physical or equivalent. The radiation integrals can be solved analytically only for special cases. I have chosen to include such analytical solutions only if they are so simple that they can improve the physical interpretation, or if they drastically decrease computation time. Otherwise, I think numerical solutions are preferable.

The main reason for the *compact vector notation* is that it is independent of the choice of coordinate system, and thereby it is easier to interpret the expressions. And, physical interpretations are important for improving the imagination and for building up intuition. Both imagination and intuition play an important role in research and development because they make it simpler to foresee working solutions. They improve creativity.

Maxwell's field equations based on the electric and magnetic vector potentials contain the parameters permittivity, permeability and angular frequency (ε, μ and ω), which we rarely use in practice. The relative values ε_r and μ_r have much more convenient values than ε and μ. In addition, ε_0, μ_0 and ω can always be replaced by the wave number and wave impedance (k and η). Therefore, my book consequently uses k and η instead of ε_0, μ_0 and ω in all equations. Actually, *I am of the opinion that the commonly used vector potentials in most cases are inconvenient and unnecessary, and I have therefore chosen to avoid them by using the direct vector integral formulas for the fields resulting from current sources.* This has contributed significantly to the compactness of the formulations.

It is always difficult to remember where the *wave impedance* η should appear in the radiation field formulas. In order to improve this, I have chosen wherever it is natural to begin all formulas for the H-field with $1/\eta$, and to always use η in front of the electric current distribution $\mathbf{J}$. Thereby, η comes in no other places. With this notation it is very easy to normalize the H-field and the current in such a way that the wave impedance in free space becomes unity, if desired.

The electric and magnetic fields can always be expressed as integrals over impressed and induced electric and magnetic currents. If these currents are unknown, the integrals can be used to formulate integral equations by applying the boundary condition for the fields. Such integral equations are normally solved by the MoM. This book is not about numerical MoM solutions. Still, I have chosen to *present the MoM as a basic principle in field theory*, and give an introduction to it in Chapter 4. The reason is that the classical integral formulas for the impedances of dipoles, slots and patches can be derived more simply and understandable by using a MoM approach (with one expansion function) than in classical ways (such as by "eletromagnetic force"). The MoM approach makes it also easier to understand that the results can be improved by increasing the number of expansion functions.

I have chosen to begin the book with a chapter on characterization of antennas (Chapter 2), instead of introducing Maxwell's equations and the far-field of incremental sources. This characterization includes definitions of the *far-field function*, the *phase reference point* and the *co- and cross-polar polarization vectors*. This makes it possible later in Chapter 4, when Maxwell's equations and the incremental sources are introduced, to characterize also the latter in terms of their radiation characteristics. Thereby, the incremental sources will become easier to use when designing antennas.

The terms far-field function and phase reference point cannot be found in other text books. They are evident and self-explanatory to experienced people working in the antenna field. Still, I have often been surprised about how many do not know that a phase pattern has a reference point, and that this has nothing to do with the phase center. Therefore, I have chosen to define these important terms clearly in Chapter 2.

Antennas are normally studied in the transmit mode, for which case the *equivalent circuit* can be found in many textbooks. Recently, and in particular in relation to signal processing antennas, it has become common to perform a detailed receive mode analysis. Such analyses become much more accurate if a complete equivalent circuit of the receiving antenna is used. Therefore, I have chosen to also present the equivalent circuits for reception, which as far as I know cannot be found in other textbooks. The induced voltage source is expressed in terms of the far-field function in the transmit mode as well as the direction and complex amplitude of the incident wave.

Several antennas are rotationally symmetric. They are more commonly referred to as *bodies of revolution (BOR)*. These antennas can be analyzed more simply than others by taking advantage of the symmetry. I have chosen to divide them into two types, BOR_0 and BOR_1 antennas, where the index denotes the order of variation which the field has in φ-direction. I have added a Section 2.4 with the mathematic descriptions of BOR_0 and BOR_1 antennas.

Aperture antennas are most conveniently analyzed in terms of *Huygens sources* located at the wavefronts of the aperture field. In order to handle this approach in an efficient way, I have chosen to define a Huygens equivalent and a Huygens approximation, which relate to Huygens study made in 1690.

When analyzing plane apertures, I introduce the Fourier transform of the aperture field. I have chosen the sign of the exponent of the integrand in such a way that the *spectral domain fields* can be conveniently interpreted as waves propagating in the direction of $\mathbf{k} = k_x\hat{\mathbf{x}} + k_y\hat{\mathbf{y}} + k_z\hat{\mathbf{z}}$, where k_x and k_y are the spectral variables. This definition of the Fourier transform is not in agreement with the standard definition, but it is easier to physically interpret herein.

I have in Chapter 5 chosen to interpret the Fourier transforms of planar current distributions as *infinite current sheets*. This has the advantage of reducing the complexity of the resulting field formulas. It is also easier to generalize the formulas to fields in dielectric multilayer structures (not included in this book).

Microstrip antennas are most accurately analyzed by the MoM and by accounting for the substrate by the spectral domain approach. I have developed the reaction formula for the patch impedance. Still, I have chosen to use the simpler two-slot transmission line model to calculate Mathcad results for *patch antennas* (Chapter 6)), to make the MATLAB program simpler and faster. I use in fact an *improved transmission line model* valid for rectangular patches on thin substrates. This gives an accuracy which is comparable with that of the MoM for this special case. The derived reaction formula for the patch impedance contains a probe correction for both line-fed and probe-fed patches.

The book contains a section about *Gaussian beams*. The Gaussian beam is a very illustrative way of describing how fields transform from radiating near-fields to far-fields. It is very useful for fast initial calculations of beamwidths from any aperture antenna at any distance from the aperture, and the results are in many cases quite accurate. The Gaussian beam is in particular useful in designing corrugated horn antennas. The Gaussian beam formulas are rarely found in basic antenna textbooks, as they are too complex for doing calculation by hand without introducing errors. To improve this, I have separated the formulas into independent parts, all being physically interpretable, such as for example "diffraction cone radius" and "geometrical optics cone radius", and the formulas are programmed in the MATLAB handbook, by which results easily can be achieved. This separation in two parts makes then also the Gaussian beam a source for understanding the phenomenon of aperture diffraction.

In Chapter 1, I have included a section about terminology where I describe the difference between *rays and phase paths*, and where I introduce aperture diffraction as well as edge diffraction. I have chosen to do this already in the introductory Chapter 1, because most people have a physical feeling for rays even without having read any theoretical work on ray techniques, and because the concept of diffraction is so essential in antennas and so strongly linked to ray interpretation. The concept of aperture diffraction is in particular important. Without it people may get a completely wrong impression of how antennas work, from knowing only about ray interpretations.

Chapter 10 on linear and planar arrays presents two ways of computing the far-field of an array: By the element-by element sum (resulting in the classical array factor), and as a grating-lobe sum expression. The latter converges very fast for regular arrays, and is related to the Floquet mode sum expression. Thereby it is possible to treat array antennas as apertures and use the aperture theories in Chapter 7. The classical array links to the modern MIMO array via the far-field function of the embedded element, as explained in Chapter 3.

Antenna analysis is always subject to approximations. I hope that my approaches will help in choosing the approximations that give the most accurate valid results.

Learning outcome of Antenna Engineering course at Chalmers

The major parts of the book are used as literature in the course *Antenna Engineering* at Chalmers. This is offered within two international Master programs at Chalmers, on *i)*

Wireless, Photonics and Space Engineering and *ii)* Communication Engineering. The Master program administration required that we defined some expected learning outcomes for the students. I have chosen to list these learning outcomes below, because they may be useful for others. This course has 14 double lecture hours spread over 7 weeks, and a corresponding number of tutorial hours. There are also three antenna laboratory exercises related to *i)* design of microstrip patch antenna, *ii)* measurements in anechoic chamber, and *iii)* measurements in reverberation chamber. There are not enough lectures to cover all the material in the book, so the most theoretical parts of Chapters 4, 6 and 11 are excluded from the course, and some additional material is not mandatory. I have chosen to teach the excluded parts for my PhD students in an advanced course called "EM theory for antenna specialists". This includes also supplementary material from several journal articles.

The expected learning outcome after having followed the Antenna Engineering course at Chalmers are: The overall aim of the textbook is to provide the reader with an understanding of antennas for use in both traditional line-of-sight (LOS) systems and in modern wireless communication systems with multipath and Rayleigh fading, ranging from initial design with simple classical design formulas to numerical design and characterization with measurements. With this understanding the reader should be able to:

a) Describe how antennas for line-of-sight (LOS) systems work and are characterized. Examples of LOS systems are radio telescopes, radar, radio links (point-to-point and point-to-multipoint communications) and satellite communication systems.

b) Describe how antennas in multipath environment with fading behave and are characterized, such as antennas for mobile terminals and devices such as mobile phones, including also the characterization of the whole mobile terminal and the user interaction. This is unique for the book!

c) Describe the most common materials used in numerical antenna analysis as well as in practical antenna design.

d) Explain the different factors contributing to the efficiency and gain of different types of antennas.

e) Explain the physical limitations of antennas; such as miniaturization and bandwidth limits of small antennas, maximum gain limits of large antennas including supergain, and correlation and efficiency limits of multiport/multibeam array antennas. This is unique for the book!

f) Explain how different antennas can be analyzed in terms of classical incremental elementary sources, by using a modern and compact non-differential vector notation and numerical integration. The incremental elementary sources are the electric current, the equivalent magnetic current and the directive Huygens source. This is unique for the book!

g) Apply his knowledge about antenna analysis to design antennas using classical formulas and design curves for the most traditional antenna types; such as dipoles, slots, horns, reflectors and phased arrays. Good initial designs with classical formulas are important for a successful numerical design with a professional antenna CAD tool.

h) Apply his knowledge about characterization of antennas for LOS and fading environment to measure antennas, both in classical anechoic chambers and in modern reverberation chambers, respectively. The reverberation chamber is a multipath emulator, in which also active

mobile terminals such as mobile phones can be measured.

April 2015
Per-Simon Kildal

Chapter 1

Introduction

An antenna is a device for radiating or receiving electromagnetic waves. It was originally called an aerial before the name antenna became more popular, and sometimes simply called a radiator. It is a transducer for coupling electromagnetic energy between free space and a waveguide, transmission line, or receiver/transmitter. An antenna is often directional in the sense that it concentrates waves in certain directions in space relative to waves in other directions. In this sense it represents a directional filter in space.

Antennas are needed and used in wireless ground-based telecommunication systems such as direct broadcasting, point-to-point radio links, mobile telephony, local area networks, traffic toll systems and navigation. They are also used in satellite communication systems for transmission of data, video and telephone signals between earth stations connected to the ground-based network, and for broadcasting satellite-TV signals. Antennas are used in mobile earth terminals. And, we even have mobile hand held phones with small antennas for receiving (and transmitting) signals from (and to) satellites, which relay the signals down to ground again all around the earth. Furthermore, antennas find applications in military and civilian surveillance radars, such as ground-based weather radars and air traffic radars, and air-borne military search and scouting radars. There exist air- and satellite-borne synthetic aperture radars (SARs) for very high resolution surveillance of land resources and pollution, and of military and civilian activity on sea and land (e.g., for surveillance of national fishing borders). SARs make it possible to synthesize a larger "antenna" than the actual antenna's physical size by moving the antenna and using signal processing. Antennas are also used in scientific radars (e.g., for ionospheric and planetary research). Finally, antennas are used in receive systems for radio astronomy, meteorology, and for detection of signals from air-borne radars (early warning). Some radio telescopes for radio astronomical observations are large spectacular examples of high technology antenna designs.

The art of antenna design is special and very interesting because so many different disciplines are involved. The antenna is often the largest and most expensive structural part of a microwave system, and it must, for physical reasons, have a certain size in terms of wavelengths in order to satisfy the system's directional filter requirements. Therefore, good and compact antenna solutions require designs which are thoroughly worked through, both from a mechanical and an electromagnetic points of view. Knowledge of material technology and manufacturing processes is needed. It may often be desirable (but not always be possible) to

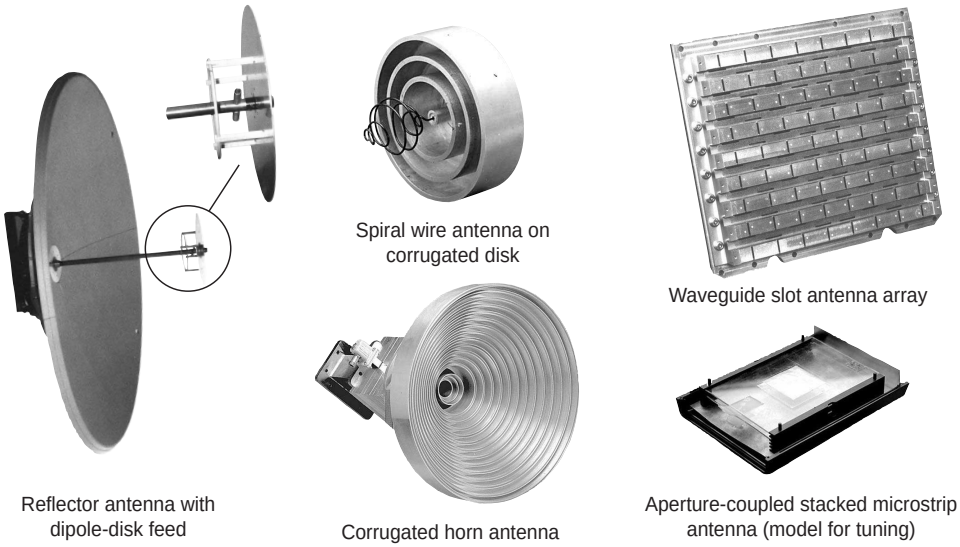


Figure 1.1: Examples of different antenna types.

modify the system requirements in order to find the most cost-effective solutions. Therefore, the design of advanced and optimized antennas involves systems engineering, mechanical engineering, and electromagnetic engineering.

Today, industrial antenna design involves a combination of computer simulations (by different commercial or semi-commercial computer programs) and experimental developments. In both cases, it is advantageous for engineers to have a reasonable background in antenna theory in order to increase their creativity and productivity. All numerical implementations of different analysis methods suffer from approximations, so the results must always be investigated with respect to convergency and assumption validity. In addition, the methods themselves are only valid under certain conditions. For example, ray tracing methods are only valid when the structure is smooth and large in terms of wavelengths.

Accurate measurements of radiation patterns and in particular directivities are complicated and require theoretical knowledge and experience. Such know-how may even be more important than good measurement equipment during the development of an antenna. Still, a modern measurement equipment is necessary in order to be able to verify the performance with sufficient accuracy to satisfy a critical customer. No quality-conscious customer will accept an antenna based on theory or simulations alone. Accurate measurements are always needed to validate the model predictions. There are certain electrical (or rather electromagnetic) system requirements for an antenna, such as gain and sidelobe levels. Common for these constrains is that they are easier to satisfy when the antenna is larger in terms of wavelengths. However, the antenna is often the largest mechanical part in a microwave system, so the specific application puts requirements on its size, weight and cost, thus calling for small antennas. Very often the antenna engineer's task is to do a trade-off between size and/or cost and performance. His job is to find the smallest or cheapest antenna which can satisfy the electric requirements.

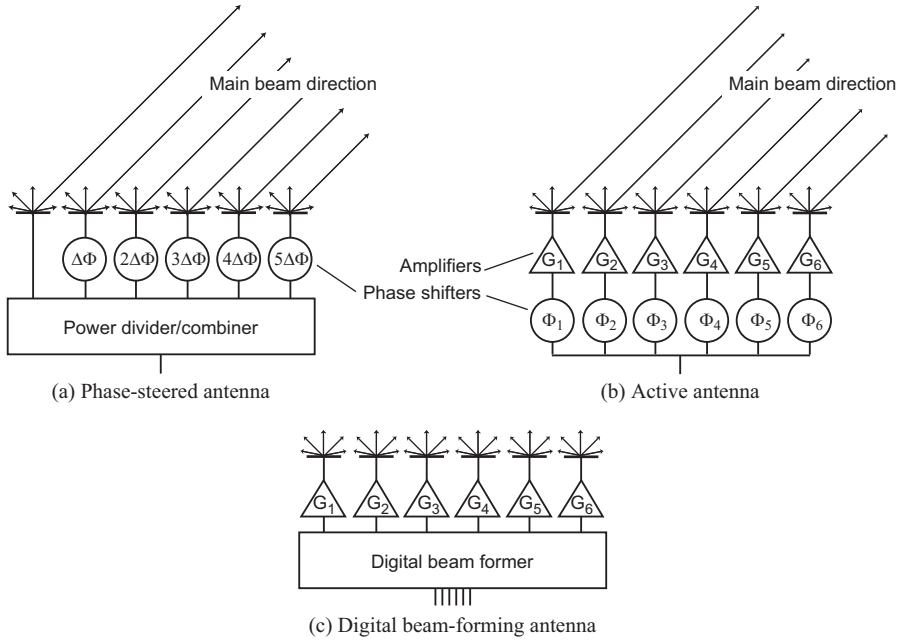


Figure 1.2: Examples of different phased-array antenna types in transmitting mode. The latter is used in MIMO systems.

1.1 Antenna types and classes

There exist several different basic antenna types (Fig. 1.1). The ones which will be treated in this book are *wire antennas*, *slot antennas*, *microstrip antennas*, *horn antennas*, *reflector antennas*, and linear and planar *arrays* of them. There are also other types which are not so common (e.g., dielectric resonator antennas and leaky wave antennas), and a large class of ultra-wideband and frequency independent antennas that will not be treated either, except for being mentioned in terms of their fundamental limitations in Chapter 11. In principle any structure of arbitrary material can be designed to radiate in free space if its size is larger than typically a few tenths of the free-space wavelength. Some antennas may radiate well even when they are apparently shorter than this. However, in such cases they are located on a box or a ground plane (or held by a hand), and it is the box or ground plane (or hand) which gives the sufficient size for the antenna to radiate. A human finger may actually work well as an antenna. Old radios without built-in antennas often work (or work better) if a finger touches the antenna input connector.

A linear or planar array antenna consists of several antenna elements, which are fed via a feed network (Fig. 1.2). If an adjustable phase shifter is connected to each element, the direction of the main beam of the antenna can be phase steered by changing the phase of each of the phase shifters. There may even be active microwave modules at each element amplifying the transmit (or receive) signals. Then, the antenna is referred to as an *active antenna*. The transmitted or received amplitude of each element can be changed by adjusting the gain of each module. The active antenna can be phase steered by including a phase shifter in each module. We can have control of both the shape and direction of the beam by changing both

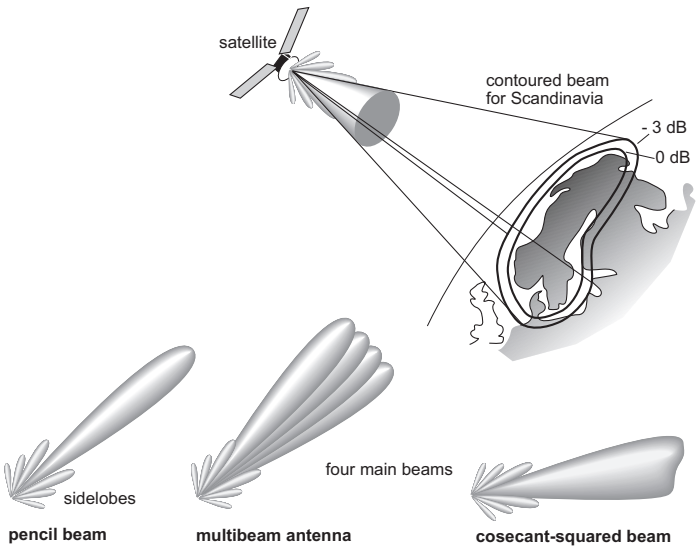


Figure 1.3: Examples of different classes of antenna beams.

the amplitude and phase of each element. The element modules may be connected directly to a digital processor (computer) via an analog-to-digital (AD) converter (on reception) or a digital-to-analog (DA) converter (on transmission). In these cases the antenna beam can be controlled digitally. We refer to this technique as *digital beam-forming* and to the antennas as *adaptive* antennas or *signal processing* antenna systems. Such antennas are also called *multi-port* antennas, because the port of each element is accessible. They are in particular used at both the transmitting and receiving sides in so-called *MIMO* (Multiple-Input Multiple-Output) communication systems. Then, the communication speed can be improved, in particular through environments which cause large time-varying signal variations, so-called *fading*. The active modules in such arrays are often realized as microwave monolithic integrated circuits (MMICs) in order to reduce the size and the cost of manufacturing modules in large numbers. In the upper microwave and millimeter wave regions, active antennas may have thousands of elements. At millimeter wave frequencies the antennas may also be integrated with the active module on the same chip. We refer to such antennas as integrated antenna systems.

Antennas are also classified in terms of the shapes of their radiation patterns (Fig. 1.3). Most antennas are *pencil-beam antennas* with a narrow main beam and low sidelobes. Pencil-beam antennas on satellites illuminate a relatively small spot on the earth's surface, and they are called spot-beams antennas. On satellites there may also be *contoured-beam antennas* with beams which are shaped to cover certain geographical areas on the earth. Radars may have *cosecant-squared antennas* with a radiation pattern in the elevation plane which varies as $\csc^2(\theta) = 1/\sin^2(\theta)$. The reason for this is that the sensitivity in this case becomes independent of the height of the target above the ground. Most antennas have only one main beam, but *multi-beam antennas* are also attractive in many applications. There exist also *omnidirectional antennas*, sector-beam antennas, antennas with hemispherical coverage, and many more.

1.2 Brief history of antennas and analysis methods

The first antenna was designed and manufactured by Heinrich Hertz when he experimentally discovered electromagnetic waves in 1888, twenty four years after Maxwell's theoretical predictions were revealed to the Royal Society in London. That was a parabolic metal cylinder with a dipole antenna along its focal line, see [1] and [2]. Marconi achieved a successful transatlantic transmission in 1901 by using an array of 50 copper wires for the transmitting antenna in England. The theories of radiating apertures, reflectors and array antennas were developed during World War II in connection with the design of the military radar. These were later published in the book by S. Silver [3] in the M.I.T. Radiation Laboratory Series. Among other early important antenna books are the book on helical antennas by J.O. Kraus [4], the book by S.K. Schelkunoff which contains both theory and experiments [5], and the book on linear antennas by R.W.P. King [6]. Several later antenna books are included in the reference list [7]-[29].

The development of the antenna area has been strongly influenced by the development of theories, analysis methods and numerical techniques, and by the growth of computer technology. The oldest analysis method still in use is geometrical optics, which is a ray technique. The classical reflection law was described already by Euclid 300 years BC. The lens formula was introduced by Alhazan around year 1000 AD. Snell found the refraction law in 1621, and further developments were done during the seventeenth (Fermat) and nineteenth century (Hamilton) to what we today refer to as classical geometrical optics. The classical description is based on straight rays and power densities. The geometrical optics used today is commonly referred to as modern *Geometrical Optics* (GO) and includes phase, polarization and field strength descriptions (Luneberg 1944 and Kline 1951). The GO is a high frequency approximation that can be used when the wavelength is small compared to the structure variations and dimensions. The GO fields propagate along straight lines (rays) in free-space regions and reflect from structure boundaries. Several scientists including Sommerfeld (1884) studied the effect of edges and found that there exist diffraction effects not accounted for by GO. In order to overcome this problem, straight diffracted rays were added to the GO description (Keller [30]). This theory was later extended by Kouyoumjian and Pathak (1974-81) to a Uniform Geometrical Theory of Diffraction [31]. This is commonly known under the abbreviation *UTD* and is widely used in antenna analysis. The GO and UTD are more generally referred to as asymptotic techniques, in the sense that they are asymptotic solutions of field integrals for high frequency. Similar asymptotic solutions can be derived for cases other than reflectors and edges by using the principle of stationary phase. The alternative and related saddle-point method is also applicable. The ray descriptions of GO, UTD and other asymptotic techniques are not valid in *caustic* regions[1]. This is a severe limitation of such high frequency methods.

The wave theories do not suffer from the lack of validity in special regions, such as is the case for GO and UTD. They were developed by Huygens (1690), Hook, Grimaldi, Young and Fresnel (1830). Finally, in 1864, Maxwell connected these classical wave theories together with the theories of optics in what we now refer to as Maxwell's equations. These are the basis of all electromagnetic modeling today. In modern antenna theory and numerical methods, equivalent sources play an important role in the formulations. The *equivalence theorem* was introduced by S.A. Schelkunoff in 1936 and is related to the principle of Huygens from

[1] This means regions where neighboring rays intersect.

1690.

The GO and UTD are well suited for numerical modeling and several computer codes exist. However, they can only be used when all structure parts are large in terms of the wavelength. Maxwell's equations can be rearranged to provide electric and magnetic fields as integrals over a product of a current distribution and a *Green's function*. The latter is the field due to an incremental current source, which we may alternatively call a point source response. The current distribution is unknown and induced by other impressed sources, and it may or may not represent physical current. In the latter case we call it an equivalent current. This may be specific for a field formulation only valid in some part of solution space. When we apply a boundary condition to the integral expressions for the fields, we get an integral equation. Integral equations are nowadays solved by the *Method of Moments* (MoM), as introduced to the antenna community by Harrington [32]-[33]. Several computer programs for antenna design based on the MoM are available. The MoM cannot be used when the structure is very large in terms of the wavelength due to the large computer memory and long computation time needed. Therefore, an antenna type of a given size may often be analyzed by MoM at low frequencies and by UTD at high frequencies.

Modern numerical methods also involve solving the differential forms of Maxwell's equations directly. When implemented for harmonic time variation these are referred to as Finite Element Methods (*FEM*), and when implemented in the time domain they are referred to as Finite Difference Time Domain (*FDTD*) methods. Several general commercial and semi-commercial computer codes for antenna analysis using FEM or FDTD are available. The computer times of FEM and FDTD codes are normally much larger than for MoM codes. Three-dimensional FEM and FDTD are only applicable to structures which hardly exceed a few wavelengths in extent.

An analytical method which has been very popular for the last decade is the *spectral domain method*. This method is conveniently used to analyze microstrip antennas on planar multi-layer structures in combination with MoM. The approach is based on Fourier transformation of the fields in the two uniform directions of the structure to obtain a spectrum of plane waves or rather a spectrum of one dimensional (1D) solutions. The fields have given harmonic variations in the two dimensions representing the spectral domain. The method is extendable to other similar three-dimensional (3D) field problems involving radiating elements on structures which are uniform in two directions of the coordinate system, such as circular cylindrical and spherical multilayer structures. All these 3D problems can be solved by using a spectrum of 1D solutions (i.e., in terms of a spectrum of plane, cylindrical or spherical waves, respectively). The 1D solutions for the cylindrical and spherical structures represent a discrete spectrum, and are therefore also referred to as modal solutions. Spectral domain methods can be used to analyze antennas on structures that are uniform in only one direction of the coordinate system, such as cylindrical structures with arbitrary cross-sectional shape and *Bodies of Revolution* (BOR). In these cases the solutions are formulated as a spectrum of two-dimensional (2D) solutions. The FEM is also often used in combination with such spectral domain methods. There is some theory about spectral domain approaches in Chapter 6.

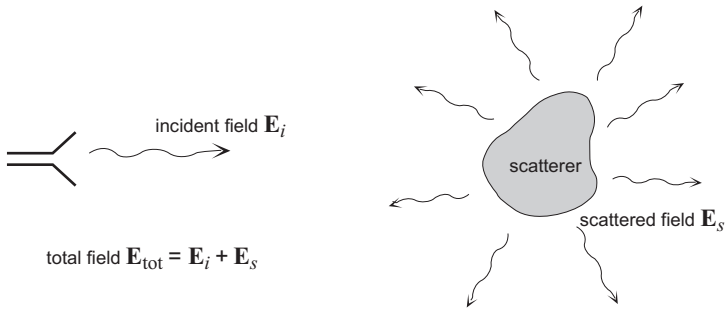


Figure 1.4: Definition of a scattering problem.

1.3 Terminology, quantities, units and symbols

The Institute for Electrical and Electronics Engineers (IEEE) have defined a standard terminology for antennas, [34]. Most of the terms for characterizing antennas, which will be introduced in Chapter 2, follow this standard. In addition, in the present section we define some special terms in antenna theory, which are commonly used to describe different physical phenomena. This is because these terms are easily confused with each other. Also there is a need to define the symbols used for the pertinent quantities and explain the units of the different field quantities to appear in later chapters. Therefore, this section also introduces some common terminology for quantities, symbols and units, used in this book.

1.3.1 Radiation or scattering

The area of electromagnetic radiation is often divided into antennas and *scattering*. Simply speaking, the area of scattering is electromagnetic radiation (or reradiation) excluding antennas. This means prediction and reduction of Radar Cross Sections (RCS) of objects, and determining the type and characteristics of a passive object from measurement of electromagnetic fields. The latter is normally referred to as inverse scattering or electromagnetic signature. The word scattering comes from the definition of a field problem where an object located in free space is illuminated by an incident field, where the incident field $\mathbf{E}_i$ is the observed field when the object is removed (Fig. 1.4). The total field $\mathbf{E}_{\text{tot}}$ is the observed field when the object (i.e., the scatterer) is present, from which we define the scattered field due to the object as $\mathbf{E}_s = \mathbf{E}_{\text{tot}} - \mathbf{E}_i$. The scattered field can be expressed as an integral over equivalent sources at the surface of the scatterer. The scattered field is radiating in all directions away from the scatterer. The scattered field is, in particular, large in shadow regions. In fact, in shadow directions it has nearly the same size as the incident field and nearly opposite phase, so as to give almost zero total field (i.e., shadow). The scattering in the direction of the shadow is referred to as *forward scattering* (i.e., scattering in the direction of propagation of the incident wave). *Backscatter* is the scattering in the direction of the distant source of the incident field. The distinction between the antenna and scattering areas lies in that antennas radiate energy whereas in scattering a known incident wave is redirected (scattered). However, the distinction between the antenna and scattering problem is not so evident. Antenna analysis involves also scattering analysis.

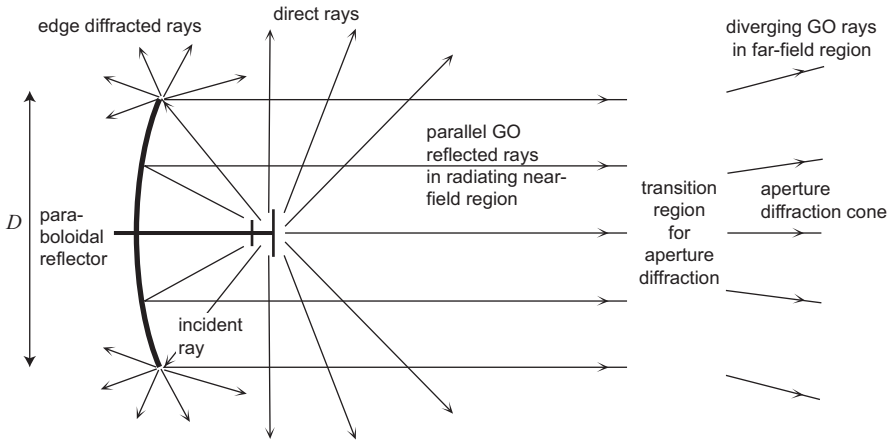


Figure 1.5: Illustration of reflection, edge diffraction and aperture diffraction.

1.3.2 Reflection, refraction and diffraction

The terms *reflection*, *refraction* and *diffraction* are associated with ray descriptions such as GO and UTD. Rays are always straight in free space. In a GO ray field problem we have an incident field which is propagating along rays originating from the source. When such a ray hits the smooth surface of an object, it is reflected. The field propagating along the GO reflected ray is called the reflected ray field. There may also be a so-called refracted ray continuing into the object. This GO refracted ray will change direction by refraction at the interface, but it will thereafter be straight until it meets a new material interface, provided the material of the object is homogeneous.

Keller described in 1985 diffraction as any process whereby electromagnetic wave propagation differs from *GO*. In particular, we have the term edge diffraction associated with the rays diffracted from the edges of objects as described by *UTD*. The fields along these rays are called the diffracted ray fields. GO and UTD predict infinite fields in points where neighboring rays intersect (and hence GO and UTD are not valid there). Such points are referred to as caustics, and the actual field behavior is described as caustic diffraction. Parallel rays (i.e., plane phase-fronts) often originate from apertures. Parallel rays intersect at infinity. Therefore, GO cannot be used to describe the wave propagation from a plane aperture with a constant phase distribution to a large distance from the aperture. The actual radiation field behavior at large distance is in this case referred to as *aperture diffraction*. Aperture diffraction causes a bundle of parallel rays to transform to a bundle of diverging rays with an associated angular field variation, when the distance from the aperture is larger than $2D^2/\lambda$, with D the aperture diameter and λ the wavelength. Most of the diverging rays are contained within a *diffraction cone* with a cone angle of approximately $2\lambda/D$ radians. Ray fields cannot easily be separated into incident and scattered fields as defined in the previous subsection. The incident field of GO is reflected, shadowed, refracted and diffracted by an object, whereas the incident field of a scattering problem is the field when the object is removed. The latter is therefore not shadowed by the object.

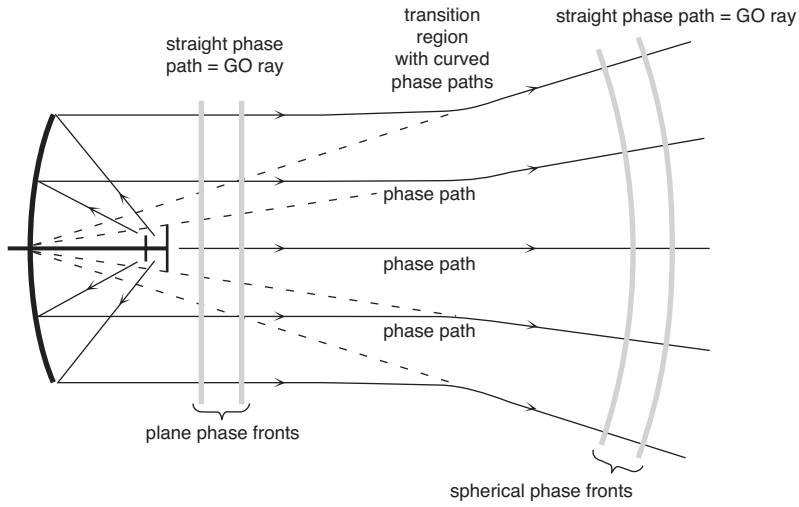


Figure 1.6: Illustration of GO rays, phase-fronts and phase paths.

1.3.3 Rays, waves, phase-fronts and phase paths

The ray technique is inherently a high-frequency approximation which breaks down in certain regions. The ray fields have a local description following straight lines in space. In spite of this, single rays cannot exist. Single rays are not physical. They always exist in ray bundles which are either converging, diverging or parallel. Therefore, single rays cannot propagate through a small hole in a metal screen and continue straight forward. They rather diffract through the hole in such a way that a spectrum of straight rays appears on the shadow side of the screen, propagating in all directions.

The wave description is different. The wave description is global in the sense that the description itself links together fields at one point in space with the fields in the rest of space. Still the wave and the ray descriptions are related. A wave has *phase-fronts* (i.e., surfaces in space over which the phase is constant). The normals to a phase-front define the propagation direction of the wave. The phase velocity depends on the shape of the phase-front. If we follow the propagation direction along normals to successive phase-fronts, we get a curve which is called the *phase path*. In spite of this, they are strongly related to rays. In field regions where the phase paths are straight, they represent GO rays. However, the phase paths bend through *caustics* and focal regions; whereas, the GO rays go straight through such regions. The GO description is not valid in the caustic region but outside it. A single phase path may coincide with one GO ray on one side of a caustic or focal region, and with another different GO ray on the other side. Inside a caustic region it is possible to use a separate GO ray description that is valid only inside the caustic. The diverging rays in the far-field region of Fig. 1.5 and 1.6 are valid GO descriptions in the far-field region, which is a caustic region for the parallel rays closer to the antenna.

1.3.4 SI units for fields and sources, and dB

The International System of units (SI) are for length meter (m), for time second (s), for voltage volt (V), and for current ampere (A). The units of both the *electric field* strength **E** and the equivalent *magnetic surface current density* **M** are volt/meter (V/m). The *magnetic field* strength **H** and the *electric surface current density* **J** have both the unit ampere/meter (A/m). However, these units are rarely used in antenna work. Instead, the fields and currents are presented as relative values normalized to their maxima or to other values. The unit of radiation impedances is ohm (Ω), and radiation admittances are in Siemens (S), and similarly for mutual impedances and admittances. These units are commonly used, but in many cases the radiation impedance and admittance are more conveniently replaced by a dimensionless amplitude reflection coefficient. The unit of power is Watt (W) or milliWatt (mW). We have $1\,\mathrm{W} = 1\,\mathrm{VA}$.

In antenna work relative numbers are most often presented in *dB*. The dB value is calculated as

$$|A/A_{\mathrm{ref}}|_{\mathrm{dB}} = 20 \cdot \log|A/A_{\mathrm{ref}}| \text{ dB} ,$$

where A is the amplitude of the voltage, current, reflection coefficient, field strength or surface current density, and A_{ref} is the reference value. Alternatively, it is calculated as

$$|P/P_{\mathrm{ref}}|_{\mathrm{dB}} = 10 \cdot \log|P/P_{\mathrm{ref}}| \text{ dB} ,$$

where P is the power or power density and P_{ref} is the reference power or power density. The resulting dB values are the same in both cases because

$$|A/A_{\mathrm{ref}}|^2 = |P/P_{\mathrm{ref}}| .$$

Sometimes a letter is added to the dB abbreviation in order to explain the reference value. For example, dBm means a power ratio in dB where the reference level is 1 mW, and dBW means a reference value of 1 W. In antennas we measure the antenna gain in dBi, which means dB relative to isotropic radiation, i.e, radiation that is spread equally much in all directions in space. We measure the figure of merit of a receiving antenna in dB/K (i.e., "dB over K") which means in dB relative to 1/Kelvin. The latter two units will be described in Chapter 2.

Table 1.1: Relative efficiencies and corresponding values in dB.

Relative efficiency:	1.0	0.99	0.95	0.90	0.80	0.64	0.50
Efficiency in dB:	0.0	-0.04	-0.22	-0.50	-1.0	-2.0	-3.0

All antennas can be characterized by different efficiencies and subefficiencies. These are often presented as the relative value itself or in percent (%). In the present book we consequently choose to give all such values in dB. The reason is that percent values often give wrong impressions. For example, increasing an efficiency by 10 % sounds the same when increasing it from 40 % to 50 % as when increasing it from 80 % to 90 %. In reality, the former increase is worth double of the latter. This is clearly seen from the dB values which are 1.0 dB in the former case and 0.5 dB in the latter. Therefore, results are more meaningful and more easily

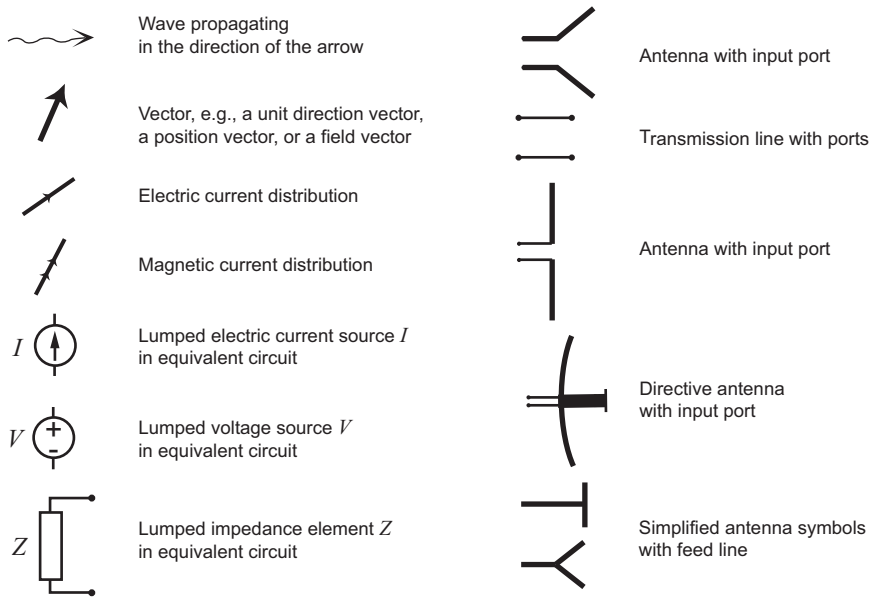


Figure 1.7: Symbols used in the drawings in this book.

interpreted when presented in dB. Tables 1.1 and 1.2 show some amplitude and power ratios and their corresponding dB values.

Table 1.2: Amplitude and power ratios and corresponding dB values.

Amplitude ratio:	1.0	0.32	0.1	0.032	0.01
Power ratio:	1.0	0.1	0.01	0.001	0.0001
Ratio:	0 dB	-10 dB	-20 dB	-30 dB	-40 dB

1.3.5 Symbols

The symbols used in the drawings in this book are explained in Fig. 1.7.

1.4 Vector notation and coordinate transformations

This book uses the vector notation in all formulations. This makes the formulas compact, and easy to interpret and remember. Another advantage is that the vector formulations are independent of the coordinate system chosen. Some may comment that vector formulas are not easy to use in practice or to program, but this is not true. The vector notation makes it easier to get a good structure of a computer program with reusable subroutines. Thereby, the time needed for developing a program is reduced.

1.4.1 Some vector formulas

In this book, a vector $\vec{E}$ is written in bold type as $\mathbf{E}$ and a unit vector $\vec{n}$ as $\hat{\mathbf{n}}$. A vector field is a function of the observation point coordinates. This is written as $\mathbf{E}(\mathbf{r})$ where $\mathbf{r} = [x, y, z]$ is the coordinates of the field point in the coordinate system defined by the unit vectors $\hat{\mathbf{x}} = [1, 0, 0]$, $\hat{\mathbf{y}} = [0, 1, 0]$ and $\hat{\mathbf{z}} = [0, 0, 1]$. $\mathbf{E} = [E_x, E_y, E_z]$ contains the three components of the vector field in the same coordinate system. $\mathbf{E}$, $\mathbf{r}$, $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ are readily programmed as complex or real one-dimensional arrays with 3 elements having the values of their x-, y- and z-components in the coordinate system. In this book expressions for the H-field are starting with $1/\eta$ where η is the wave impedance. In computer programming it is possible to use $\eta_0\mathbf{H}$ as an array variable instead of $\mathbf{H}$, where η_0 is the free space wave impedance. Then, the values of the normalized H-field, $\eta_0\mathbf{H}$, get the same order of magnitude as the values of the E-field, which can make it easier to interpret the results of a computation.

In order to simplify mathematical formulations, we choose local coordinate systems in which the local geometry is easily described. When a vector field $\mathbf{E} = [E_{x_l}, E_{y_l}, E_{z_l}]$ is known in the local coordinate system, its components in the global coordinate system are found by the formula

$$\mathbf{E} = E_{x_l}\hat{\mathbf{x}}_l + E_{y_l}\hat{\mathbf{y}}_l + E_{z_l}\hat{\mathbf{z}}_l \ , \tag{1.1}$$

where $\hat{\mathbf{x}}_l = [a_x, a_y, a_z]$, $\hat{\mathbf{y}}_l = [b_x, b_y, b_z]$ and $\hat{\mathbf{z}}_l = [c_x, c_y, c_z]$ are the unit vectors defining the local coordinate system with a_x, a_y, a_z, the components of $\hat{\mathbf{x}}_l$ in the global coordinate system and correspondingly for $\hat{\mathbf{y}}_l$ and $\hat{\mathbf{z}}_l$. Eq. (1.1) is to be used as follows to calculate the component E_x of $\mathbf{E}$ in the global coordinate system,

$$E_x = E_{x_l}a_{x_l} + E_{y_l}b_{x_l} + E_{z_l}c_{x_l} \ , \tag{1.2}$$

and corresponding for E_y and E_z. The observation point $\mathbf{r} = [r_{x_l}, r_{y_l}, r_{z_l}]$ in the local coordinate system is transformed to $\mathbf{r} = [r_x, r_y, r_z]$ in the global system by

$$\mathbf{r} = \mathbf{r}_0 + r_{x_l}\hat{\mathbf{x}}_l + r_{y_l}\hat{\mathbf{y}}_l + r_{z_l}\hat{\mathbf{z}}_l \ , \tag{1.3}$$

where $\mathbf{r}_0 = [r_{x_0}, r_{y_0}, r_{z_0}]$ is the coordinates of the origin of the local coordinate system in the global coordinate system. Eq. (1.3) is to be used as follows:

$$r_x = r_{x_0} + r_{x_l}a_{x_l} + r_{y_l}b_{x_l} + r_{z_l}c_{x_l} \tag{1.4}$$

and similarly for r_y and r_z.

Considering a vector field $\mathbf{E}$ in the global coordinate system, we can find its component (i.e., projection) E_r in any direction (defined by the unit vector $\hat{\mathbf{a}} = [a_x, a_y, a_z]$) by the dot- (or *scalar*-) product

$$E_r = \mathbf{E} \cdot \hat{\mathbf{a}} = E_x a_x + E_y a_y + E_z a_z \ . \tag{1.5}$$

If $\hat{\mathbf{a}}$ is a complex unit vector, we define the component of a complex vector field $\mathbf{E}$ along $\hat{\mathbf{a}}$ by the *dot-product* between $\mathbf{E}$ and the complex conjugate of $\hat{\mathbf{a}}$, i.e.,

$$E_a = \mathbf{E} \cdot \hat{\mathbf{a}}^* = E_x a_x^* + E_y a_y^* + E_z a_z^* \ . \tag{1.6}$$

This dot-product is normally referred to as a *symmetric product* in mathematical text books.

We also make frequent use of the vector (or cross) product. This is defined by

$$\begin{aligned}\mathbf{M} = \mathbf{E} \times \hat{\mathbf{r}} =&(E_y a_{r_z} - E_z a_{r_y})\hat{\mathbf{x}} - (E_x a_{r_z} - E_z a_{r_x})\hat{\mathbf{y}} + \\ &+(E_x a_{r_y} - E_y a_{r_x})\hat{\mathbf{z}} \ .\end{aligned} \tag{1.7}$$

This can be evaluated more rapidly as follows. The result of a *cross product* between two vectors $\mathbf{E}$ and $\hat{\mathbf{r}}$ is a new vector which is orthogonal to both $\mathbf{E}$ and $\hat{\mathbf{r}}$ with magnitude

$$|\mathbf{M}| = |\mathbf{E}||\hat{\mathbf{r}}| \sin\alpha \ , \tag{1.8}$$

where α is the angle between $\mathbf{E}$ and $\hat{\mathbf{r}}$. Thus, if $\hat{\mathbf{r}}$ is a unit vector and $\mathbf{E}$ is orthogonal to $\hat{\mathbf{r}}$, (i.e., $\mathbf{E} \perp \hat{\mathbf{r}}$), $\mathbf{M}$ gets the same magnitude as $\mathbf{E}$. The direction of $\mathbf{M}$ can be rapidly found by using the *right-hand rule* as follows. Straighten out the four joint fingers of the right hand and orient them in the direction of the first vector $\mathbf{E}$. Let your thumb point normal to the other fingers in the plane of the flat hand. Orient your flat hand in such a way that if you bend the joint fingers 90° at their roots they will point along the second vector $\hat{\mathbf{r}}$. Then, the direction of the resulting vector $\mathbf{M}$ is the normal to $\mathbf{E}$ and $\hat{\mathbf{r}}$ in the direction of the thumb. It is also possible to find the direction of $\mathbf{M}$ by considering a screw with right-handed threads.

The following vector formula will also be frequently used,

$$\mathbf{E} = K[\mathbf{J} - (\mathbf{J} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}] \ , \tag{1.9}$$

where K is a constant. This operation gives a vector $\mathbf{E}$ which is in the plane of the vectors $\mathbf{J}$ and $\hat{\mathbf{r}}$, but with no component along $\hat{\mathbf{r}}$. We can express this as $\mathbf{E}$ being proportional to $\mathbf{J}$ after $\mathbf{J}$'s projection, $\mathbf{J} \cdot \hat{\mathbf{r}}$, along $\hat{\mathbf{r}}$ has been subtracted.

1.4.2 Coordinate transformations

In order to transform a vector field from one local coordinate system to the global coordinate system, we need to know the unit vectors of the local coordinate system in terms of their components in the global system. Such transformations of unit vectors of coordinate systems are well known. They are given in Appendix C for the sake of convenience.

1.4.3 Dyads

Dyads are used to provide compact notations for given vector operations. They are often used in connection with formulations of Green's functions (i.e., field solutions due to point sources). For example, the vector formula

$$\mathbf{E} = K[\mathbf{J} - (\mathbf{J} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}] \tag{1.10}$$

may be expressed in terms of a dyad $\bar{\mathbf{G}}$ as

$$\mathbf{E} = \mathbf{J} \cdot \overline{\mathbf{G}}(\hat{\mathbf{r}}) \ , \tag{1.11}$$

where $\overline{\mathbf{G}}(\hat{\mathbf{r}})$ is defined by the following vector operation on an arbitrary vector $\mathbf{J}$,

$$\mathbf{J} \cdot \overline{\mathbf{G}}(\hat{\mathbf{r}}) = K[\mathbf{J} - (\mathbf{J} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}] \ . \tag{1.12}$$

Thus, the purpose of introducing dyads is to simplify vector expressions. When a dyadic operator is applied to a vector $\mathbf{J}$, the resulting quantity is a vector with an amplitude and a direction which are generally different from those of $\mathbf{J}$.

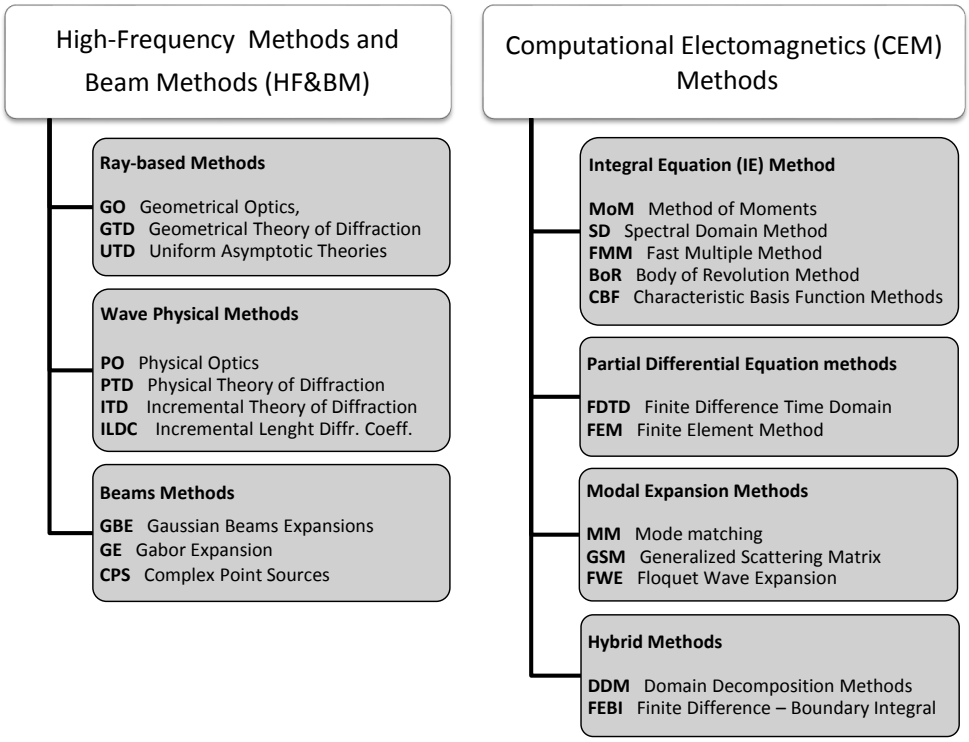


Figure 1.8: Classification of EM analysis methods.

1.5 Overview on EM analysis methods by S. Maci

The methods of analysis of antennas can be divided into high-frequency and beam (HF&BM) methods and computational electromagnetics (CEM) methods. Both can be implemented in software analysis tools; HF&BM are based on analytical formulas, whereas CEM methods are based on numerical algorithms. A synthetic overview is shown in Fig. 1.8, where the most used acronyms are given.

HF&BM are used when the size of the problem is very large in terms of wavelengths. Among the great variety of HF Methods, one can distinguish ray-methods (GTD, UTD, GO), which are based on asymptotic solutions of the wave equation. These methods represent the asymptotic field in terms of rays. Ray tracing algorithms must be implemented to treat ray-fields in complex environments, and this constitutes the main drawback in numerical implementation. These methods are suited for treating problems of thousands of wavelengths in size, like those encountered when analyzing antennas or the scattering from complex structures, such as aircrafts or ships. Also, they play an important role when simulating scattering of complex multipath scenarios, like in urban environments.

Wave Physical Methods (WPM, e.g., PO, PTD, ITD) represent the scattered or radiated fields in terms of radiation integrals and provide an asymptotic HF approximation of the integrand. The most used among them is Physical Optics (PO) [35] (treated in Chapter 7

of this book), which describes the field scattered by metallic smooth scatterers in terms of a free-space radiation integral of equivalent electric currents. At each point of the smooth surface, these currents are identified as those induced on a flat infinite metal plate at the tangential point. PO is widely used in the analysis of regular and shaped reflector antennas. Among WPM methods, the incremental (PTD, ITD, ILDC) methods provide diffracted or fringe field representations in terms of line integrals along edge-type discontinuities. Beam methods are suitable for representing fields from near to far fields and are used for problems involving lenses, multiple reflectors and quasi-optical systems. The latter are often referred to as beam waveguides. Gaussian Beams (GB) and Complex Point Sources (CPS) will be described later in this book.

Further historical details on high frequency techniques are given here. The Geometrical Theory of Diffraction (GTD) by Keller [30] postulates the existence of diffracted rays compensating the discontinuity of GO field at the so-called incidence or reflection Shadow Boundaries (SB). These latter are those boundaries where incident or reflected rays abruptly emerge or disappear due to an edge discontinuity. Keller's original GTD formulation of diffracted rays was invalid exactly at the SBs, where the contribution of diffracted rays is more important. The UTD by Kouyoumjian and Pathak [31] improves this description for wedge-type discontinuities by introducing a transition function that can be multiplied with the GTD-formulated diffracted rays. Thereby, this gives a uniform compensation of the discontinuity at the GO ray-field SB. The UTD transition function tends to unity outside a narrow transition region across the SB and then recovers the GTD description. In [36]-[38] Pathak extended GTD to the problem of curved surfaces, making use of a canonical solution for cylinders.

Ray field theories like GO, GTD and UTD, fail in the neighborhood of *caustics* of reflected or diffracted rays. In fact, the power density associated with ray-fields become infinite there, and a wave physical description like PO should be invoked to overcome this impairment. PO is valid also at caustics of reflected and diffracted rays and this is why it is useful in calculating the field radiated by parabolic reflector antennas. Far from caustics, the asymptotic evaluation of the PO radiation integral reduces to the GO description, plus approximate diffracted rays from the edges. These diffracted rays have a different amplitude coefficient than those predicted by UTD and they predict a less accurate field. To correct the inaccuracy of the asymptotic PO integral, Ufimtsev introduced the Physical Theory of Diffraction (PTD) [39]. PTD additionally improves the PO currents close to edges by "non-uniform" or "fringe" currents related to those induced by diffractive effects on the surface of the scatterer. The description of the PTD is given in terms of PO plus line integral of "elementary edge waves" along edge discontinuities. The PTD elementary edge waves were also referred to as Incremental Length Diffraction Coefficients (ILDC) by other authors [40]-[41]. An historical overview on the origin and development of PTD can be found in [42]. A different approach is suggested in the Incremental Theory of Diffraction (ITD) [43]-[44], which in contrast to PTD satisfies reciprocity.

While HF&BM, which have roots in optics, have been developed since the sixties, CEM methods had their significant evolution since the nineties, with a great improvement in the last fifteen year [45]-[46]. CEM methods are versatile and general since they can be applied to a variety of structures of any shape and material structure. On the other hand, the electrical size of the solution problem increases enormously, and therefore their disadvantage is the large demand of dynamic memory and CPU time. CEM methods can be divided in Integral Equation (IE) Methods, Partial Differential Equation (PDE) Methods, and Modal Expan-

sion (MEX) Methods. These three methods are today often combined in many commercial software tools because each method is appropriate for different purposes. In particular, IE are mostly used in open domains, while PDE and MEX are mostly used in closed domains. However, IE methods can be used in closed domain as well, but they are not competitive with PDE and MEX methods. Reversely, PDE can be used in open domains, provided that the domain is closed with absorbing boundary conditions or other alternative techniques. Analogously, MEX can be used in open domains (for instance using Spherical Mode Expansions), but limited to specific domain shapes.

IE methods express boundary conditions in terms of radiation from equivalent currents on a closed surface. These unknown currents are thereafter expanded as a sum of basis functions with unknown coefficients. Next, these coefficients are determined by averaging the boundary condition of the integral equation using so-called testing functions. This solution scheme is known as Method of Moments (MoM) [32], treated in Chapter 4 of this book. The MoM can be formulated in both space and spectral (wavenumber) domains. The Spectral Domain (SD) methods are mainly used for multilayer structures and patch antennas, and are described in Chapter 6.

MoM reduces the integral equation to a set of linear algebraic equations whose solution requires the inversion of a square matrix. Depending on the matrix size (namely on the number of unknown), different strategies are used, spanning from direct matrix inversion to iterative processes (some more details are given in Chapter 4, Section 4.8). Due to memory storage requirements, when the number of unknowns exceeds some hundreds of thousands (i.e., size of some tenth of wavelengths), it is not possible to formulate the problem in a conventional way. Then, groups of unknowns should be used to render a sparse MoM matrix, with the advantage of reduced memory occupation and number of numerical operations. This concept is pursued in the Fast Multipole Method (FMM) [47]-[48]. To treat large problems, a valid alternative to matrix sparsification is matrix compression. This can be obtained by first combining the unknowns of subportions of the problem into special basis functions, called Characteristic Basis Functions (CBF) [49]-[50] or Synthetic Basis Functions (SFX) [51]. For specific problems with circular symmetry, MoM may be formulated in a special way, which reduces the complexity to the one of a 2D problem. This formulation is applied successfully to *Body of Revolution* (*BoR*) antennas, like conical corrugated horns and circularly symmetric reflectors. In most of the practical cases, IE methods are formulated through surface current unknown, and rarely by volumetric unknowns (polarization currents), being in the latter case not competitive with respect to finite element approaches.

A second class of CEM methods is based on the solution of Partial Differential Equations (PDE) derived from Maxwell's Equations directly. In contrast to Integral Equations, which use currents on surface as unknowns, PDE uses the EM fields in the volume of interest as unknowns. There are two different classes of PDE methods arising from two different solution schemes: the Finite Difference and the Finite Element schemes. In the first scheme, finite difference equations are used to approximate the partial derivatives. This is done in both space and time, leading to the Finite Difference Time Domain (FDTD) method [52]-[53]. Since FDTD is implemented in time, it can cover wide frequency ranges in a single simulation run when a broadband time-pulse is used, and it can also accurately treat non-linear material properties. The finite difference equations are solved in leapfrog manner, where the electric field vector components are solved at a given instant of time and the magnetic vector components in the same cell are solved at the next instant of time. The

process is repeated continuously until the transient or steady-state condition is fully evolved. Yee applied this concept for the first time to Maxwell's equation in the seminal paper [54]. The widely known acronym FDTD is indeed due to Taflove [55]. FDTD is of simple implementation and gives a good physical understanding of the field evolution when using a valid graphic interface. It can also treat a wide variety of complex structures and combination of linear and not linear materials. Furthermore, it fits well for parallel processing in CPUs with such architectures. On the other hand, FDTD presents the following drawbacks: *i)* the domain under study is gridded, and therefore a large volume requires very long, and sometimes unacceptable computation times. *ii)* It does not treat resonances in an efficient way so that convergence is very slow. Finally, *iii)* it can suffer from instability. In order to truncate the space-grid, it is possible to use a Perfectly Matched Layer (PML) [56]. PML, introduced by Berenger in '94, greatly improved the capability of FDTD, rendering it applicable (and sometimes competitive with MoM) when analyzing antenna problems.

There exists an alternative formulation of the finite-difference case making use of a Transmission Line Matrix (TLM) approach [57]. This was very early used in a vector two-dimensional (V2D) implementation [58]. The BOR problem is also a vector 2D problem. Therefore, this V2D approach can be used to analyze BOR antenna structures with a computational effort of smaller order than a full 3D solution [59].

The second solution scheme for PDE is constituted by the Finite Element Method (FEM) [60]-[61]. FEM replaces the volumetric continuous domain into subdomains (finite elements). The simplest finite element is a linear tetrahedron with four nodes, located at the four corners. The unknown functions are then represented by simple interpolating functions with unknown coefficients. The interpolating functions over the tetrahedron are usually a polynomial of first (linear) or second (quadratic) order, written in terms of the node positions. Thus, the original boundary value problem with an infinite number of degrees of freedom is converted into a problem with a finite number of degrees of freedom, thus leading to a finite number of unknowns. The solution scheme is obtained by testing the equations with functions equal to interpolating functions (Galerkin testing) in a similar way as it is done in MoM. The appropriate numbering of the nodes in a global system leads to a linear system which possesses a limited banded matrix, with relevant advantages in terms of inversion time and memory storage. This renders FEM more convenient than a volumetric MoM, which does not satisfy this property. The difficulty in using FEM in antenna problems is, as in FDTD, the termination of the discretized domain. Absorbing boundary conditions can be used for this purpose as well as a formulation referred to as Finite Element Boundary Integral (FEBI) method [62]-[63]. The latter is based on a separation into internal and external regions through the equivalence theorem, followed by the imposition of the continuity of the tangential components of the fields at the interface. The internal region is treated by FEM, and the external region is treated by IE and solved by MoM. Therefore, FEBI can be also seen as a hybridization of the FEM and MoM scheme, and it is implemented in various commercial solvers. However, due to the MoM coupling, FEBI matrices are not banded.

We finally mention those methods which, although not being general as the others, are widely used because they are extremely fast when applicable. These are the Modal Expansion (MEX) Methods [64]. They consist in dividing a closed domain in several subdomains where the modes are known in analytical forms. The modal representations in the different regions are matched together by applying continuities of tangential fields at the domain boundaries, which also determines the unknown coefficients of the expansion. The original application of

this method to waveguide discontinuities and corrugated horns is known under the name of Mode Matching (MM). Several generalizations of this method are formulated in literature, which are referred to as Generalized Scattering Matrix (GSM) or Generalized Admittance Matrix (GAM) [65]-[66]. In both these methods, the domains are more general than simple waveguide regions, although closed form analytical modes must exist for the fields inside them. The advantage of these classical methods is that they are extraordinary fast and accurate, since the problem can be formalized as a microwave network with a limited number of ports. They are so fast that they can be directly applied during the optimization process of designing antenna or microwave devices (e.g., corrugated horns and microwave waveguide filters). The disadvantages consist in their limited range of applicability. However, the applicability can be extended by hybridization with FEM in domains where modes are not known in analytical form. The MEX technique applied to periodic structure can be formulated by the use of Floquet Wave (FW) expansions, thus leading to a method successfully used since the eighties for the study of Frequency Selective Surfaces (FSS) [67].

A class of hybrid methods, very popular nowadays, is known under the name of Domain Decomposition Methods (DDM) [68]-[70]. These methods solve large boundary value problems in terms of several smaller domain boundary value domain problems, with overlapped or disjoined domains. DDM allows for the use of different techniques in the various subdomains, and thereby massive numerical parallel processing. DDM can be thought as the generalization of a MEX method to open domains [71], for those cases where the domains are disjoined and spherical wave or beam expansions are used in the network description.

Today, all the main software companies for EM solvers have implemented MoM, FEM, FDTD and Mode Matching in their products, and most of them also have HF ray-based solvers. The formulations inside the solvers are rapidly evolving in successive releases. However, each of the most popular software became famous for one particular method. For instance the ANSYS-HFSSTM [72] has an original release based on FEM, FEKOTM [73] on MoM CSTTM [74], on a special version of FTDT, GRASPTM [75] on PO and UTD.

1.6 References

[1] J.F. Ramsay, "Microwave antenna and waveguide techniques before 1900", *Proceedings of the IRE*, Vol. 46, No. 2, pp. 405-415, 1958.

[2] A.W. Love, "Some highlights in reflector antenna development", *Radio Science*, Vol. 11, pp. 671-684, August-September 1976. (Reprinted in A. W. Love, *Reflector Antennas*. New York, IEEE, 1978.)

[3] S. Silver (Ed.), *Microwave Antenna Theory and Design*, M.I.T. Radiation Laboratory Series, Vol. 12, McGraw-Hill, New York, 1949.

[4] J.D. Kraus, *Antennas*, 2nd edition, McGraw-Hill, New York, 1988.

[5] S.K. Schelkunoff and H.T. Friis, *Antennas: Theory and Practice*, Wiley, New York, 1952.

[6] R.W.P. King, *Theory of Linear Antennas*, Harvard University Press, Cambridge, Massachusetts, 1956.

[7] R.C. Johnson and H. Jasik, *Antenna Engineering Handbook*, 2nd edition, McGraw Hill, New York, 1984.

[8] H. Jasik, *Antenna Engineering Handbook*, McGraw-Hill, New York, 1961.

[9] E.A. Wolff, *Antenna Analysis*, Wiley, New York, 1966.

[10] R.E. Collin and F.J. Zucker, *Antenna Theory, Parts 1 and 2*, McGraw- Hill, New York, 1969.

[11] W.N. Christiansen and J.A. Høgbom, *Radiotelescopes*, Cambridge Univiversity Press, Cambridge, Massachusetts, 1969.

[12] S. Drabowitch and C. Ancona, *Antennas*, Vol. 2, Ecole Superieure d'Electricite, Masson, Paris, 1978.

[13] E. Zentner, *Radio communication (Antennas, Propagation and Radar)* (in Croatian), Radiokomunikacije (Antene, Radiomreze in radarski sustavi), Skolska Knjiga, Zagreb, 1980.

[14] R.S. Elliott, *Antenna Theory and Design*, Prentice-Hall, 1981.

[15] W.L. Stutzman and G.A. Thiele, *Antenna Theory and Design*, 2nd edition, Wiley, New York, 1997.
[16] C.A. Balanis, *Antenna Theory and Design*, 2nd edition, Wiley, New York, 1996.
[17] A.W. Rudge, K. Milne, A.D. Olver and P. Knight (Ed.), *The Handbook of Antenna Design*, Vol. 1-2, Peregrinus, 1982.
[18] J.R. Wait, *Introduction to Antennas and Propagation*, Peregrinus, 1986.
[19] S.O. Andreassen, *Radio Antenner* (in Norwegian), Tele-mobil AS, Oslo, Norway, 1994.
[20] S. Drabowitch, A. Papiernik, H. Griffiths, J. Encinas and Bradford L. Smith, *Modern Antennas*, Chapman & Hall, 1998.
[21] T. Kitsuregawa, *Advanced Technology in Satellite Communication Antennas: Electrical and Mechanical Design*, Artech House, 1990.
[22] G.E. Evans, *Antenna Measurement Techniques*, Artech House, 1990.
[23] K. Fujimoto and J.R. James (Ed.), *Mobile Antenna Systems Handbook*, Artech House, 1995.
[24] T. Macnamara, *Handbook of Antennas for EMC*, Artech House, 1995.
[25] D.J. Kozakoff, *Analysis of Radome-Enclosed Antennas*, Artech House, 1997.
[26] Y.T. Lo and S. W. Lee (Ed.), *Antenna Handbook*, Van Nostrand Reinhold, 1988.
[27] D.M. Sazonov, *Microwave Circuits and Antennas*, Mir Publisher Moscow, 1990.
[28] B.L. Smith and M.H. Carpentier (Ed.), *Microwave Engineering Handbook, Vol. 1: Microwave Components, Vol. 2: Microwave Circuits, Antennas and Propagation, Vol. 3: Microwave Systems and Applications*, IEEE Press/Chapman and Hall, 1993.
[29] T.A. Milligan, *Modern Antenna Design*, 2nd edition, IEEE Press, Wiley-Interscience, 2005.
[30] J.B. Keller, "Geometrical Theory of Diffraction", *Journal of Optical Society of America*, 52, pp. 116-130, 1962.
[31] R.G. Kouyoumjian and P.H. Pathak, "A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface", *Proceedings of the IEEE*, Vol. 62, No. 11, pp. 1448-1461, November 1974.
[32] R.F. Harrington, *Field Computation by Moment Methods*, Macmillan, New York, 1968.
[33] R.F Harrington, "Matrix methods for field problems", *Proceedings of the IEEE*, Vol. 55, No. 2, pp. 136-149, February 1967.
[34] IEEE, "IEEE standard definitions of terms for antennas." *IEEE std 145-1993*, Institute of Electrical and Electronics Engineers, New York, June 21, 1993.
[35] A. Jenkins, H.E. White, "Fundamentals of Physical Optics", McGraw-Hill, New York, 1937.
[36] P.H. Pathak, "An Asymptotic Analysis of the Scattering of Plane Waves by a Smooth Convex Cylinder," *Radio Science*, 14, pp. 419-435, May-June 1979.
[37] P.H. Pathak, "High-Frequency Techniques for Antenna Analysis", *Proceedings of the IEEE*, 80, 1, pp. 44-65, January 1992.
[38] P.H. Pathak, G. Carluccio, M. Albani, "The Uniform Geometrical Theory of Diffraction and Some of Its Applications", *IEEE Antennas and Propagation Magazine*, Vol. 55, No. 4, August 2013.
[39] P.Y. Ufimtsev, *Fundamentals of the Physical Theory of Diffraction*, Whiley, New Jersey, Second edition, 2014.
[40] Michaeli, Arie, "Equivalent edge currents for arbitrary aspects of observation", *IEEE Transactions on Antennas and Propagation*, 32.3 (1984): 252-258.
[41] Shore, Robert A., and Arthur D. Yaghjian, "Incremental diffraction coefficients for planar surfaces", *IEEE Transactions on Antennas and Propagation*, 36.1 (1988): 55-70.
[42] P.Y. Ufimtsev, "50 year anniversary of PTD: coments on origination and development", *IEEE Antennas and Propagation Mag.*, 55 (3), pp. 18-28.
[43] R. Tiberio, A. Toccafondi, A. Polemi, S. Maci, "Incremental Theory of Diffraction. A new improved formulation", *IEEE Transactions on Antennas and Propagation*, Vol. 52, pp. 2234–2243, 2004.
[44] R. Tiberio, S. Maci, "An incremental theory of Diffraction. Scalar formulation" (pubblicazione su rivista, 1994), *IEEE Transactions on Antennas and Propagation*, Vol. AP-42/5, pp. 600–612.
[45] Chew, W-C, et al., *Fast and efficient algorithms in computational electromagnetics*, Artech House, Inc., 2001.
[46] Peterson, A. F., Ray, S. L., Mittra, *Computational methods for electromagnetics*, New York: IEEE press, 1998.
[47] R. Coifman, V. Rokhlin, and S. Wandzura, "The fast multipole method for the wave equation: A pedestrian prescription", *IEEE Antennas and Propagation Magazine*, pp. 7-12, 1993.
[48] J. Song, C.-C. Lu, and W.C. Chew. "Multilevel fast multipole algorithm for electromagnetic scattering by large complex objects", *IEEE Transactions on Antennas and Propagation*, 45.10 (1997): 1488-1493.
[49] VVS Prakash, and R. Mittra, "Characteristic basis function method: A new technique for efficient solution of method of moments matrix equations", *Microwave and Optical Technology Letters*, 36.2 (2003): 95-100.

[50] R. Maaskant, R. Mittra, A. Tijhuis, "Fast analysis of large antenna arrays using the characteristic basis function method and the adaptive cross approximation algorithm", *IEEE Transactions on Antennas and Propagation*, Vol. 56, No. 11, pp. 3440-3451, November 2008.

[51] L. Matekovits, A. Valeriu Laza, and G. Vecchi, "Analysis of large complex structures with the synthetic-functions approach", *IEEE Transactions on Antennas and Propagation*, 55.9 (2007): 2509-2521.

[52] Taflove, Allen, and Susan C. Hagness, "Computationai Electrodynamics", Boston, Artech House (2000).

[53] Kunz, K. S., & Luebbers, R. J. (1993). *The finite difference time domain method for electromagnetics*. CRC press.

[54] K.S. Yee "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media", *IEEE Transactions on Antennas and Propagation*, 14 (3): 302–307.

[55] A. Taflove, M.E. Brodwin, "Numerical solution of steady-state electromagnetic scattering problems using the time-dependent Maxwell's equations", *IEEE Transactions on Microwave Theory and Techniques*, 23 (8), 623-630, 1975.

[56] J.-P. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves", *Journal of computational physics*, Vol. 114, No. 2, pp. 185-200, 1994.

[57] M. Celuch-Marcysiak, W. G. Gwarek, "Generalized TLM algorithms with controlled stability margin and their equivalence with finite-difference formulations for modified grids", *IEEE Transactions on Microwave Theory and Techniques*, Vol. 43, No. 9, pp. 2081-2089, September 1995.

[58] W.K. Gwarek, T. Morawski, and C. Mroczkowski, "Application of the FDTD method to the analysis of circuits described by the two-dimensional vector wave equation", *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-41, No. 2, pp. 311–317, February 1993.

[59] M. Celuch-Marcysiak, W. G. Gwarek, "Industrial Design of Axisymmetrical Devices Using a Customized FDTD Solver from RF to Optical Frequency Bands", *IEEE Microwave Magazine*, pp. 150-159, December 2008.

[60] Silvester and Ferrari, *Finite Element for Electrical Engineering*, Cambridge, Cambridge University Press, 1983 (2nd ed. 1990, 3rd ed 1996).

[61] J.M. Jin, *The finite element method in Electromagnetics*, New York, Wiley, 1993 (2nd Ed 2002, 3rd ed 2014).

[62] J.L Volakis, A. Chatterjee, and L. Kempel "Finite Element Methods for Electromagnetics: Antennas Microwave circuits and Scattering Applications", *IEEE Press*, Piscataway, NY, 1998.

[63] Jin, Jian-Ming, and John Leonidas Volakis, "A finite-element-boundary integral formulation for scattering by three-dimensional cavity-backed apertures", IEEE Transactions on Antennas and Propagation, 39.1 (1991): 97-104.

[64] Conciauro, G., Guglielmi, M., & R. Sorrentino, *Advanced modal analysis*, J. Wiley, 2000.

[65] Alessandri, F., G Bartolucci, and R Sorrentino, "Admittance matrix formulation of waveguide discontinuity problems: Computer-aided design of branch guide directional couplers", Microwave Theory and Techniques, IEEE Transactions on 36.2 (1988): 394-403.

[66] Hall, R.C., Mittra, R., & Mitzner, K. M. (1988). "Analysis of multilayered periodic structures using generalized scattering matrix theory", Antennas and Propagation, IEEE Transactions on, 36(4), 511-517.

[67] Mittra, R., Chan, C. H., & Cwik, T. (1988). "Techniques for analyzing frequency selective surfaces-a review". Proceedings of the IEEE, 76(12), 1593-1615.

[68] Stupfel, B., "A fast-domain decomposition method for the solution of electromagnetic scattering by large objects", Antennas and Propagation, IEEE Transactions on , Vol. 44, No. 10, pp. 1375, 1385, October 1996.

[69] A. Alonso-Rodriguez and L. Gerardo-Giorda "New nonoverlapping domain decomposition methods for the harmonic Maxwell system", SIAM J. Sci. Comput., Vol. 28, No. 1, pp. 102–122, 2006.

[70] S.-C. Lee, M.N. Vouvakis, Jin-Fa Lee, "A non-overlapping domain decomposition method with non-matching grids for modeling large finite antenna arrays", *Journal of Computational Physics*, Vol. 203, No. 1, pp. 1–21, 10 February 2005.

[71] G. Carli, S. Maci, E. Martini, "A domain decomposition method based on a generalized scattering matrix formalism and a complex source expansion", *Progress in Electromagnetic Research Waves*, Vol. 19, pp. 445-473, 2010.

[72] ANSYS, `http://www.ansys.com/`

[73] FEKO - EM Simulation Software, `https://www.feko.info`

[74] CST - Computer Simulation Technology, `https://www.cst.com/`

[75] GRASP, `www.ticra.com/products/software/grasp`

Chapter 2

Characterization of directive antennas

The purpose of the present chapter is to define and describe in general terms the different quantities that are normally used to characterize microwave antennas. All these quantities are functions of electromagnetic fields or currents, and these fields and currents are everywhere in this book assumed to be time-harmonic. Time-harmonic fields are most conveniently described by complex vectors. Therefore, we start by defining these as well as the related time average power density (Section 2.1).

Radiation fields are waves propagating in the radial direction away from the antenna. Far away from the antenna they can locally be thought of as plane waves. So we describe the plane wave and its possible polarizations in the rectangular coordinate system in Section 2.2. This description comes as a simple introduction to the more complicated treatment of the polarization of the radiation field in the actual spherical coordinate system in Section 2.3. This section also deals with the far-field function and the different radiation patterns extracted from it, as well as the directive gain and the directivity.

Several antennas consist of rotationally symmetric structures that are excited in some way, e.g., by a dipole located on the symmetry axis. Rotationally symmetric structures are also called *Bodies of Revolution* (BOR). For BOR antennas which are excited with azimuthal field of no or simple variation, there are some important and useful relations between the radiation fields in different planes. These are described in Section 2.4.

The antenna is always part of a system which also includes at least a transmitter or receiver. The system performance is best characterized in terms of the antenna gain G (also called power gain), effective aperture, ohmic losses, antenna noise temperature T and the figure of merit. In a communication system, the latter is also called G/T (say "G over T"). These quantities are defined in Section 2.5. We will also present the equivalent circuits of transmitting and receiving antennas.

2.1 Time-harmonic electromagnetic fields

We will in this course calculate the electric and magnetic fields generated by given electric or magnetic source currents[1]. When the source current varies fast with time, we may get radiation of propagating electromagnetic fields. Such a wave propagation is most simply studied for the case that the sources have a steady sinusoidal time variation of the form

$$\cos(\omega t + \phi) = \Re\{e^{j(\omega t+\phi)}\} \ , \tag{2.1}$$

where ϕ is the phase in radians and $\omega = 2\pi f = 2\pi/T$ is the angular frequency in radians/seconds with f the *frequency* in Hz and T the period in seconds. Then, all the resulting fields must be time-harmonic as well. The fields are generally vector functions of all space coordinates and the time, so we may write

$$\vec{E}(x,y,z,t) = \Re\{\mathbf{E}(x,y,z)e^{j\omega t}\} \ , \tag{2.2}$$

$$\vec{H}(x,y,z,t) = \Re\{\mathbf{H}(x,y,z)e^{j\omega t}\} \ , \tag{2.3}$$

where $\vec{E}(x,y,z,t)$ and $\vec{H}(x,y,z,t)$ are instantaneous values of the E- and H-fields, respectively, at a time t and location $\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}$. $\mathbf{E}(x,y,z)$ and $\mathbf{H}(x,y,z)$ are time-harmonic vector fields of the form

$$\mathbf{E}(x,y,z) = E_x(x,y,z)\hat{\mathbf{x}} + E_y(x,y,z)\hat{\mathbf{y}} + E_z(x,y,z)\hat{\mathbf{z}} \ , \tag{2.4}$$

where $E_x(x,y,z)$, $E_y(x,y,z)$ and $E_z(x,y,z)$ are complex functions, and similarly for $\mathbf{H}(x,y,z)$. All the formulas and the results in this book are expressed in terms of such time-harmonic vector fields (and currents). We do normally not need to generate the instantaneous field and current functions from the time-harmonic versions, although it can be done by using (2.2) and (2.3). The instantaneous *power* $P(t)$ crossing a closed surface S with a normal $\hat{\mathbf{n}}$ can be calculated from (see [1, Section 8-5.1])

$$P(t) = \iint_S \vec{E} \times \vec{H} \cdot \hat{\mathbf{n}} dS \ , \tag{2.5}$$

where $\vec{E} \times \vec{H}$ is called the instantaneous Poynting vector. The fields are periodic with period T, so the power averaged over a period T must be constant independent of t. It becomes

$$P_{\text{ave}} = \frac{1}{T}\int_0^T P(t)dt = \iint_S (\mathbf{W}_{\text{ave}} \cdot \hat{\mathbf{n}})dS \ , \tag{2.6}$$

where

$$\mathbf{W}_{\text{ave}} = \frac{1}{2}\Re\{\mathbf{E} \times \mathbf{H}^*\} \ , \tag{2.7}$$

is the time average *Poynting vector* or power density vector. The factor 1/2 in (2.7) comes from time averaging over the product of two cosine functions, containing the harmonic time dependences of the E- and H-fields, respectively. We could also have removed this 1/2 factor if we had defined $\mathbf{E}(x,y,z)$ and $\mathbf{H}(x,y,z)$ as the Root-Mean-Square (RMS). Also called effective complex fields instead of peak values. The consequence of assuming time-harmonic electromagnetic fields is that the analysis needs to be done for one frequency at the time. In practice the analysis must be done for some discrete frequencies over the operational bandwidth of the antenna. It is often sufficient to choose the three frequencies representing the lower band-edge, the center, and the upper band-edge.

[1] For more information about radiation from current sources, see Chapter 4.

2.2 Plane waves and their polarization

At an observation point far away from an antenna we can always regard the radiation field locally as a propagating plane wave. Therefore, the fields incident on a receiving antenna can normally be considered as a plane wave. The time-harmonic E- and H-fields of a plane wave propagating in positive z-direction in free space are described by[2]

$$\mathbf{E} = \mathbf{E}_t e^{-jkz} = [E_x \hat{\mathbf{x}} + E_y \hat{\mathbf{y}}] e^{-jkz} \ , \tag{2.8}$$

$$\mathbf{H} = \frac{1}{\eta}\hat{\mathbf{z}} \times \mathbf{E} = \frac{1}{\eta}\hat{\mathbf{z}} \times \mathbf{E}_t e^{-jkz} = \frac{1}{\eta}[-E_y \hat{\mathbf{x}} + E_x \hat{\mathbf{y}}] e^{-jkz} \ , \tag{2.9}$$

where $\eta = 377\,\Omega \approx 120\pi\,\Omega$ is the *wave impedance* in free space,

$$k = 2\pi/\lambda$$

is the *wavenumber* and λ is the *wavelength*. E_t is the E-field vector with the propagation factor e^{-jkz} omitted. The wavelength in free space can be calculated from the frequency f by using

$$\lambda = \frac{c}{f} \ ,$$

i.e., $\lambda[\text{mm}] \approx 300/f[\text{GHz}]$ where $c = 2.99790 \times 10^8$ m/s is the *phase velocity in free space* (also called the speed of light). The power density vector becomes

$$\begin{aligned} \mathbf{W}_{\text{ave}}(x,y,z) &= \frac{1}{2}\Re\{\mathbf{E} \times \mathbf{H}^*\} = \frac{1}{2\eta}[\mathbf{E} \times \hat{\mathbf{z}} \times \mathbf{E}^*] \\ &= \frac{1}{2\eta}|\mathbf{E}_t|^2 \hat{\mathbf{z}} = \frac{1}{2\eta}[|E_x|^2 + |E_y|^2]\hat{\mathbf{z}} \ , \end{aligned} \tag{2.10}$$

which means that there is a propagation of power in the positive z-direction.

The electric (and magnetic) fields have both x- and y-components, i.e., E_x and E_y, and H_x and H_y, respectively. The electromagnetic wave is said to be polarized. This property makes it possible to simultaneously modulate different signals on E_x and E_y, at the same frequency. Thus, we are able to re-use the frequencies in a communication system and essentially double the capacity. Generally, we may re-use frequency by making use of any pair of orthogonal polarizations, not only the x- and y-polarizations. In order to enable use of both polarizations, both the transmitting and receiving antennas must be designed for good isolation between the two polarizations. This means that if we have modulated different signals on the two field components, we must be able to detect them without mutual interference. Such requirements are formulated by treating one of the *polarizations* as the desired (*co-polar*) polarization, and, when this is excited, the second orthogonal polarization is undesired (*cross-polar*). The isolation is then the ratio between the co-polar and cross-polar components. When the second polarization is excited, this is co-polar and the first is cross-polar. The isolation for this case will be the same if we consider the same dual-polarized antenna (due to reciprocity).

The polarization is always determined from the characteristics of the E-field and can be described in terms of a desired co-polar component E_{co} which is parallel with a co-polar unit vector $\hat{\mathbf{co}}$, and an undesired cross-polar component E_{xp} which is parallel with a cross-polar

[2] See any textbook on basic theory of time-harmonic electromagnetic fields, e.g., [1].

unit vector $\hat{\mathbf{xp}}$, which is orthogonal to $\hat{\mathbf{co}}$. Both $\hat{\mathbf{co}}$ and $\hat{\mathbf{xp}}$ are orthogonal to the propagation direction $\hat{\mathbf{z}}$. Thus, we can recast the total E-field as

$$\mathbf{E} = (E_{\text{co}}\hat{\mathbf{co}} + E_{\text{xp}}\hat{\mathbf{xp}})e^{-jkz} \ . \tag{2.11}$$

The co- and cross-polar unit vectors are generally orthonormal complex vectors, satisfying

$$|\hat{\mathbf{co}}|^2 = \hat{\mathbf{co}} \cdot \hat{\mathbf{co}}^* = 1 \ , \qquad \hat{\mathbf{co}} \cdot \hat{\mathbf{xp}}^* = 0 \ , \tag{2.12}$$

$$|\hat{\mathbf{xp}}|^2 = \hat{\mathbf{xp}} \cdot \hat{\mathbf{xp}}^* = 1 \ , \qquad \hat{\mathbf{xp}} \cdot \hat{\mathbf{co}}^* = 0 \ . \tag{2.13}$$

Therefore, the co- and cross-polar components of the E-field ($\mathbf{E}_t$) can be found by scalar multiplication of the E-field with $\hat{\mathbf{co}}^*$ and $\hat{\mathbf{xp}}^*$, respectively, i.e.,

$$E_{\text{co}} = \mathbf{E}_t \cdot \hat{\mathbf{co}}^* \ , \qquad E_{\text{xp}} = \mathbf{E}_t \cdot \hat{\mathbf{xp}}^* \ . \tag{2.14}$$

2.2.1 Linear polarization

When the E-field is y-directed, we refer to it as linearly y-polarized. If this is the desired co-polar polarization, we have[3]

$$\hat{\mathbf{co}} = \hat{\mathbf{y}} \ , \qquad \hat{\mathbf{xp}} = \hat{\mathbf{x}} \ , \tag{2.15}$$

$$E_{\text{co}} = \mathbf{E}_t \cdot \hat{\mathbf{y}}^* = \mathbf{E}_t \cdot \hat{\mathbf{y}} = E_y \ , \tag{2.16}$$

$$E_{\text{xp}} = \mathbf{E}_t \cdot \hat{\mathbf{x}}^* = \mathbf{E}_t \cdot \hat{\mathbf{x}} = E_x \ . \tag{2.17}$$

For a co-polar *linear polarization* in x-direction we choose correspondingly[4]

$$\hat{\mathbf{co}} = \hat{\mathbf{x}} \ , \qquad \hat{\mathbf{xp}} = -\hat{\mathbf{y}} \ . \tag{2.18}$$

A linear polarization is often referred to as being *horizontal* (HOR) or *vertical* (VER). A HOR polarization is horizontal relative to the ground. It is defined mathematically by a real $\hat{\mathbf{co}} \perp \hat{\mathbf{n}} \perp \hat{\mathbf{z}}$ where $\hat{\mathbf{n}}$ is the normal to the ground and $\hat{\mathbf{z}}$ is the direction of propagation. The E-field of a VER polarization is not necessarily normal to the ground (i.e., we normally do not require that $\hat{\mathbf{co}} = \hat{\mathbf{n}}$, but rather $\hat{\mathbf{co}}$ lies in the plane defined by $\hat{\mathbf{z}}$ and $\hat{\mathbf{n}}$). This makes it possible to use the terms HOR and VER polarizations also of waves not propagating parallel with the ground, such as a wave leaving ground in a certain elevation direction (Fig. 2.1). The horizontal and vertical polarizations are often also referred to as *perpendicular* and *parallel* polarizations, respectively, relative to the elevation plane.

Generally, an arbitrary linear polarization can be defined by co- and cross-polar unit vectors of the form

$$\hat{\mathbf{co}} = \cos\xi\hat{\mathbf{x}} + \sin\xi\hat{\mathbf{y}} \ , \tag{2.19}$$

$$\hat{\mathbf{xp}} = \sin\xi\hat{\mathbf{x}} - \cos\xi\hat{\mathbf{y}} \ . \tag{2.20}$$

We see that $\xi = \pi/2$ for a desired y-polarization and $\xi = 0$ for a desired x-polarization. If we do not require a specific direction of the linear polarization, we can choose ξ in such a way that the cross-polar component of the field is minimized.

[3] We could also have chosen $\hat{\mathbf{xp}} = -\hat{\mathbf{x}}$ or more generally $\hat{\mathbf{xp}} = e^{j\phi}\hat{\mathbf{x}}$ with ϕ on arbitrary phase.

[4] We could also here have chosen $\hat{\mathbf{xp}} = \hat{\mathbf{y}}$ or $\hat{\mathbf{xp}} = e^{j\phi}\hat{\mathbf{y}}$ in the same way as for the co-polar y-polarization.

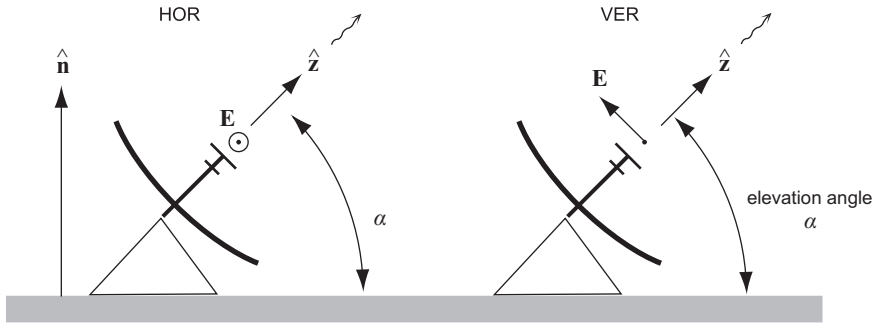


Figure 2.1: Illustration of horizontal (left) and vertical (right) polarization of an antenna with a certain elevation angle α measured from the horizon.

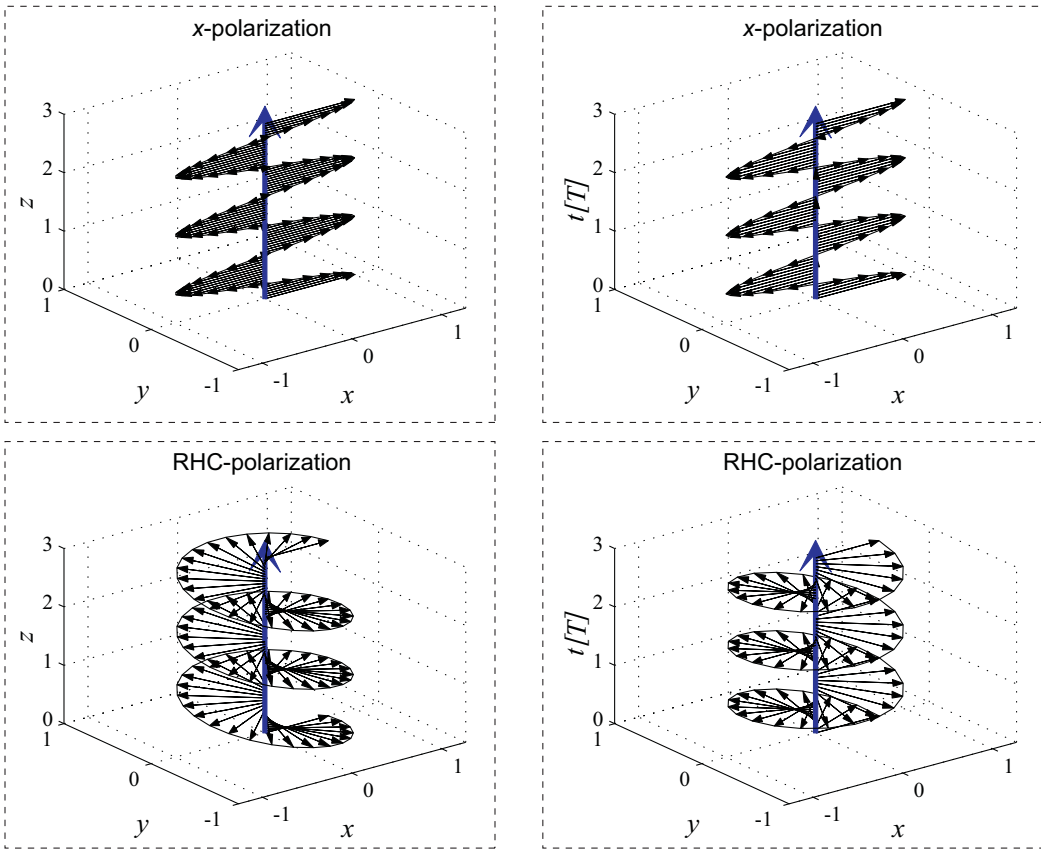


Figure 2.2: ⋆A linear x-polarized wave (upper) and a right-hand circular (RHC) polarized wave (lower), plotted as a function of z for fixed t (left) and as a function of t for fixed z (right). There exist MATLAB code which generates videos of the field vector as a function of z when the time t increases.

2.2.2 Circular polarization

A plane wave can also have an E-field that rotates an angle 2π in the xy-plane when the time passes through a period T. This rotation of the E-field may either be observed *i)* at a given position in space as a function of increasing time; or, *ii)* at a given time as a function of z along the direction towards which the wave propagates. In both cases the fields may have either a right-hand or left-hand rotation, corresponding to the directions of the fingers of respectively the right or left hand when the thumb is pointing in the direction of propagation of the wave (i.e., the positive z-axis). However, the two observation methods *i)* and *ii)* give different answers to whether the rotation direction is right-hand or left-hand. Observation method *i)* is always chosen for the definition by convention (Fig. 2.2[5]).

We thus simply define *right-hand circular* (RHC) polarization for waves propagating in the z-direction by the unit vector

$$\hat{\mathbf{co}} = (\hat{\mathbf{x}} - j\hat{\mathbf{y}})/\sqrt{2} \ , \tag{2.21}$$

which means that the y-component has a phase factor $-j = e^{-j\pi/2}$ compared to the x-component. In other words, the y-component of the field is delayed by $t_0 = T/4$ (i.e., a quarter of a period) compared to the x-component. Let us study in more detail a RHC-polarized plane wave of the form

$$\mathbf{E} = \hat{\mathbf{co}}\, e^{-jkz} = (\hat{\mathbf{x}} - j\hat{\mathbf{y}})e^{-jkz}/\sqrt{2} \ . \tag{2.22}$$

The time-varying field is

$$\vec{E}(z,t) = \Re\{\mathbf{E}e^{j\omega t}\} = E_x(z,t)\hat{\mathbf{x}} + E_y(z,t)\hat{\mathbf{y}} \ , \tag{2.23}$$

with

$$\begin{aligned} E_x(z,t) &= \Re\{\mathbf{E}\cdot\hat{\mathbf{x}}e^{j\omega t}\} = \cos(\omega t - kz)/\sqrt{2} \ , \\ E_y(z,t) &= \Re\{\mathbf{E}\cdot\hat{\mathbf{y}}e^{j\omega t}\} = \cos\left(\omega t - \frac{\pi}{2} - kz\right)/\sqrt{2} \\ &= \cos\left(\omega\left(t - \frac{T}{4}\right) - kz\right)/\sqrt{2} \\ &= \cos\left(\omega t - k\left(z + \frac{\lambda}{4}\right)\right)/\sqrt{2} \ . \end{aligned} \tag{2.24}$$

The above $\vec{E}(z,t)$ for RHC circular polarization is plotted in Fig. 2.2 as a function of z for $t = 0$ and as a function of t for $z = 0$. $\vec{E}(z,t)$ for linear x-polarization is also plotted. We see that the plot shows a right-hand circular spiral when $\vec{E}(z,t)$ is plotted as a function of t. The cross-polar unit vector which corresponds to $\hat{\mathbf{co}}$ in (2.21) is[6]

$$\hat{\mathbf{xp}} = (\hat{\mathbf{x}} + j\hat{\mathbf{y}})/\sqrt{2} \tag{2.25}$$

When the propagating E-field has the form in (2.11), we find the complex amplitudes of its co- and cross-polar fields to be

$$E_{\mathrm{co}} = \mathbf{E}\cdot\hat{\mathbf{co}}^* = (E_x + jE_y)/\sqrt{2} \ , \tag{2.26}$$

[5] There exist Matlab code for all figures of which the caption start with ⋆.

[6] We could also have chosen $\hat{\mathbf{xp}} = (\hat{\mathbf{y}} - j\hat{\mathbf{x}})/\sqrt{2}$ or (2.25) multiplied with any complex constant of unit amplitude.

$$E_{\rm xp} = \mathbf{E} \cdot \hat{\mathbf{xp}}^* = (E_x - jE_y)/\sqrt{2} \ . \tag{2.27}$$

The complex values can be expressed in terms of their amplitudes and phases, according to

$$E_x = |E_x|e^{j\phi_x} \ , \qquad E_y = |E_y|e^{j\phi_y} \ . \tag{2.28}$$

We see that $E_{\rm xp} = 0$ if $|E_x| = |E_y|$ and $\phi_y - \phi_x = -\pi/2$. This corresponds to the condition described in the next paragraph.

The RHC circularly polarized wave can be excited by two orthogonal, but otherwise equal, linearly polarized antennas. Assume that the two antennas are x-and y-polarized, respectively. Then, the 90° delay of the signal on the y-polarized antenna relative to the x-polarized one can be obtained by connecting the x- and y-polarized antennas to the in-phase (0°) and quadrature (−90°) outputs of a 3 dB hybrid power divider, respectively. Alternatively, we can feed the y-polarized antenna through a cable which is $\Delta z = \lambda_c/4$ longer than the cable feeding the x-polarized antenna where, λ_c is the wavelength in the cable. We also can generate circular polarization by using helical antennas and many other ways.

We define *left-hand circular* (LHC) polarization correspondingly by co and cross-polar unit vectors of the forms

$$\hat{\mathbf{co}} = (\hat{\mathbf{x}} + j\hat{\mathbf{y}})/\sqrt{2} \ , \tag{2.29}$$

$$\hat{\mathbf{xp}} = (\hat{\mathbf{x}} - j\hat{\mathbf{y}})/\sqrt{2} \ . \tag{2.30}$$

Generally, we may define a desired *elliptical polarization* by the unit vectors

$$\hat{\mathbf{co}} = [\hat{\mathbf{x}} + Ae^{j\Delta\phi}\hat{\mathbf{y}}]/\sqrt{1+A^2} \ , \tag{2.31}$$

$$\hat{\mathbf{xp}} = [-Ae^{-j\Delta\phi}\hat{\mathbf{x}} + \hat{\mathbf{y}}]/\sqrt{1+A^2} \ , \tag{2.32}$$

where A and $\Delta\phi$ are real constants. However, the elliptical polarization has no practical interest as a *desired* polarization, although in practice linearly and circularly polarized fields are often elliptically polarized as the cross-polarization is never identically zero. An elliptical polarization is obtained by adding two orthogonal circular polarizations with different amplitudes, or by adding two orthogonal linear polarizations with different phases. Adding two orthogonal circular polarizations with equal amplitude gives linear polarization. Adding two orthogonal linear polarizations with 90° phase difference and equal amplitudes gives circular polarization.

2.2.3 Axial ratio and cross-polarization

The polarization ellipse is the curve which the peak of the E-field describes when the time varies in a plane normal to the propagation direction of the wave. The ellipse can be characterized by its maximum and minimum field values, $|E_{\rm max}|$ and $|E_{\rm min}|$, respectively. The *Axial Ratio* (AR) of the ellipse is defined by

$$({\rm AR})_{\rm dB} = 10\log\left|\frac{E_{\rm max}}{E_{\rm min}}\right|^2 \ {\rm dB} \ . \tag{2.33}$$

For a desired circular polarization the AR in dB and the amplitudes of the co- and cross-polar fields are related by

$$({\rm AR})_{\rm dB} = 10\log\left[\frac{|E_{\rm co}| + |E_{\rm xp}|}{|E_{\rm co}| - |E_{\rm xp}|}\right]^2 \ {\rm dB} \ . \tag{2.34}$$

For an ideally circularly polarized wave the AR is unity (0 dB). The axial ratio is infinity for an ideally linearly polarized wave.

The *cross-polar* level can be quantified in four different ways: by the *cross-polar decoupling* (also called the *cross-polar isolation*)

$$(\mathsf{XPD})_{\mathrm{dB}} = 10\log\left|\frac{E_{\mathrm{co}}}{E_{\mathrm{xp}}}\right|^2 \ \mathrm{dB}\ , \tag{2.35}$$

by the *relative cross-polar level*

$$(\mathsf{XP})_{\mathrm{dB}} = 10\log\left|\frac{E_{\mathrm{xp}}}{E_{\mathrm{co}}}\right|^2 \ \mathrm{dB}\ , \tag{2.36}$$

by the *polarization efficiency* which is the ratio in dB between the co-polar power density and the total power density of the wave, i.e.,

$$(e_{\mathrm{pol}})_{\mathrm{dB}} = 10\log\left(\frac{|E_{\mathrm{co}}|^2}{|E_{\mathrm{co}}|^2 + |E_{\mathrm{xp}}|^2}\right) \ \mathrm{dB}\ . \tag{2.37}$$

2.2.4 Example: Amplitude and phase errors in circular polarization excitations

Circular polarization is often generated by using two linearly polarized antennas. In this case, it is important to know how accurate the two linear excitations need to be relative to each other. We will study this.

Consider an x-polarized plane wave of unit amplitude and a y-polarized plane wave of amplitude $jAe^{j\Delta\phi}$, both propagating in the positive z-direction. The combined wave has left-hand circular (LHC) polarization when $A = 1$ and $\Delta\phi = 0$. Find the requirements to the maximum acceptable deviations of A (in dB) or $\Delta\phi$ (in degrees) from these values which give a cross-polar decoupling better than 25 dB. What is the polarization efficiency and axial ratio in this case?

SOLUTION:

The vector amplitude of the combined z-propagating wave is (see (2.8))

$$\mathbf{E}_t = E_x\hat{\mathbf{x}} + E_y\hat{\mathbf{y}} = \hat{\mathbf{x}} + jAe^{j\Delta\phi}\hat{\mathbf{y}}\ ,$$

with $A = 10^{(A)_{\mathrm{dB}}/20}$. The co-polar LHC component is found by using (2.14) with (2.29) for the co-polar unit vector:

$$E_{\mathrm{co}} = \mathbf{E}_t \cdot \hat{\mathbf{co}}^* = (\hat{\mathbf{x}} + jAe^{j\Delta\phi}\hat{\mathbf{y}}) \cdot (\hat{\mathbf{x}} - j\hat{\mathbf{y}})/\sqrt{2} = (1 + Ae^{j\Delta\phi})/\sqrt{2}\ .$$

The cross-polar RHC component is correspondingly found from (2.14) using (2.30) for the cross-polar unit vector:

$$E_{\mathrm{xp}} = \mathbf{E}_t \cdot \hat{\mathbf{xp}}^* = (\hat{\mathbf{x}} + jAe^{j\Delta\phi}\hat{\mathbf{y}}) \cdot (\hat{\mathbf{x}} + j\hat{\mathbf{y}})/\sqrt{2} = (1 - Ae^{j\Delta\phi})/\sqrt{2}\ .$$

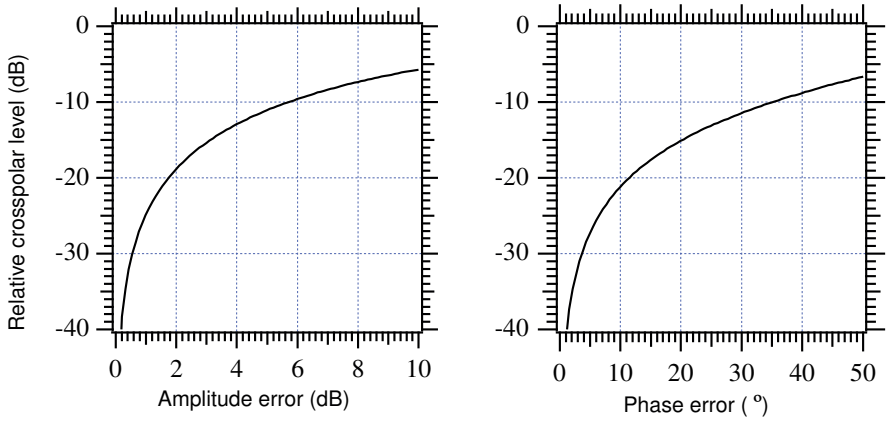


Figure 2.3: ⋆Amplitude error (left scale) and phase error (right scale) of circular polarization excitations as a function of resulting relative cross-polar level.

The resulting relative cross-polar level is

$$\left|\frac{E_{\rm xp}}{E_{\rm co}}\right| = \left|\frac{1 - Ae^{j\Delta\phi}}{1 + Ae^{j\Delta\phi}}\right| \ .$$

We can now evaluate this for different $(A)_{\rm dB}$ and $\Delta\phi$ in order to find the requirements on $(A)_{\rm dB}$ and $\Delta\phi$. We have done this and plotted the results in Fig. 2.3⋆.The polarization efficiency as obtained by (2.37) is plotted in Fig. 2.4⋆.

In addition, let us use a more elegant approach by deriving approximate expressions valid for small $(A)_{\rm dB}$ and $\Delta\phi$. These will be useful for rapid calculations by hand. We first consider the case that there are only amplitude errors (i.e., $\Delta\phi = 0$). Then, by using one of the series expansions in Appendix B, we get

$$A = 10^{(A)_{\rm dB}/20} \approx (1 + 0.115(A)_{\rm dB}) \qquad \text{for } (A)_{\rm dB} << 4.3\,{\rm dB} \ .$$

This gives from (2.36)

$$\mathsf{XP} = \left|\frac{E_{\rm xp}}{E_{\rm co}}\right| \approx \left|\frac{-0.115(A)_{\rm dB}}{2 + 0.115(A)_{\rm dB}}\right| \approx 0.0575|(A)_{\rm dB}| \ ,$$

$$(\mathsf{XP})_{\rm dB} \approx -25 + 20\log|(A)_{\rm dB}|\ {\rm dB} \ .$$

Thus, the cross-polarization is $-25\,{\rm dB}$ when the amplitude error is $(A)_{\rm dB} = 1\,{\rm dB}$. The axial ratio is found by using (2.34) with $E_{\rm max} = 1$ and $E_{\rm min} = A$ or opposite. Thus,

$$(\mathsf{AR})_{\rm dB} = |(A)_{\rm dB}| \ .$$

Let us now consider the case that $(A)_{\rm dB} = 0\,{\rm dB}$ and $\Delta\phi \ll 180°$, i.e., only phase errors. Then, we get by using an expansion formula in Appendix B

$$Ae^{j\Delta\phi} = e^{j\Delta\phi} = 1 + j(\Delta\phi)_{rad} = 1 + j\frac{\pi}{180°}(\Delta\phi)° \ .$$

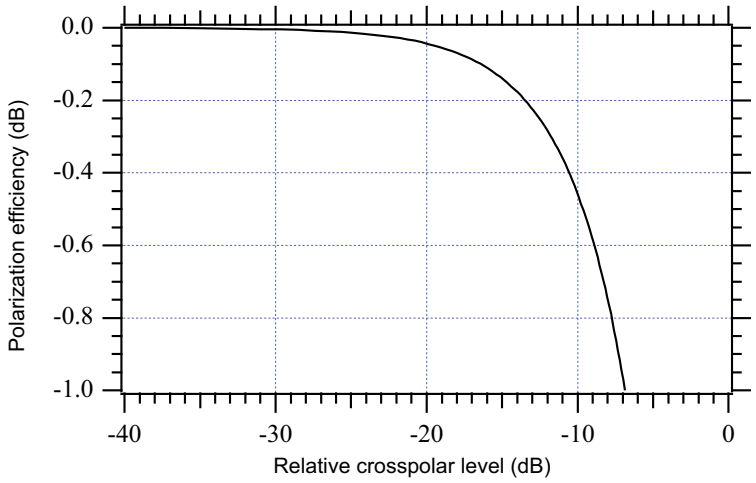


Figure 2.4: *Polarization efficiency as a function of relative cross-polar level.

Further, using (2.36)

$$\mathsf{XP} = \left|\frac{E_{\rm xp}}{E_{\rm co}}\right| = \left|\frac{-j(\Delta\phi)_{\rm rad}}{2 + j(\Delta\phi)_{\rm rad}}\right| = \frac{\pi}{360}(\Delta\phi)^\circ \ ,$$

$$(\mathsf{XP})_{\rm dB} = 20\log \mathsf{XP} = -41 + 20\log|(\Delta\phi)^\circ| \ {\rm dB} \ .$$

Thus, when the phase error is 1°, the relative cross-polarization is $-41\,$dB. The axial ratio is found by using (2.34). We get

$$(\mathsf{AR})_{\rm dB} = 20\log\left|\frac{1 + |E_{\rm xp}/E_{\rm co}|}{1 - |E_{\rm xp}/E_{\rm co}|}\right| \ {\rm dB} \approx 20\log\left[1 + \frac{\pi}{180^\circ}|(\Delta\phi)|^\circ\right] \ {\rm dB}$$

$$\approx 20(\log {\rm e})\frac{\pi}{180^\circ}(\Delta\phi)^\circ \ {\rm dB} = |(\Delta\phi)^\circ|/6.3 \ {\rm dB} \ ,$$

where we have made use of another series expansion from Appendix B. Thus, the axial ratio in dB due to phase errors can readily be calculated by dividing the phase error (in degrees) by 6.3.

For low cross-polarizations, the polarization efficiency in (2.37) becomes,

$$(e_{\rm pol})_{\rm dB} = 10\log\left(1 - \left|\frac{E_{\rm xp}}{E_{\rm co}}\right|^2\right) \ {\rm dB} = -4.3|E_{\rm xp}/E_{\rm co}|^2 \ {\rm dB} \ ,$$

by using also an expansion formula from Appendix B. We see that common relative cross-polar power levels below 0.01 (i.e., $-20\,$dB) cause very low losses. Studying the above formulas, we see that our numerical example gives the following answer. In order to get the cross-polar decoupling better than $25\,$dB (i.e., the cross-polar level lower than $-25\,$dB), we need

$$|(A)_{\rm dB}| < 1\,{\rm dB} \qquad \text{when } \Delta\phi = 0 \ ,$$

$$(\Delta\phi)^\circ \ < 6.3^\circ \qquad \text{when } (A)_{\rm dB} = 0 \ .$$

The axial ratio is $1\,$dB in both cases, and the polarization loss is $0.014\,$dB. The results can also be read out of Fig. 2.3. The polarization efficiency is plotted in Fig. 2.4 as a function of the relative cross-polar level.

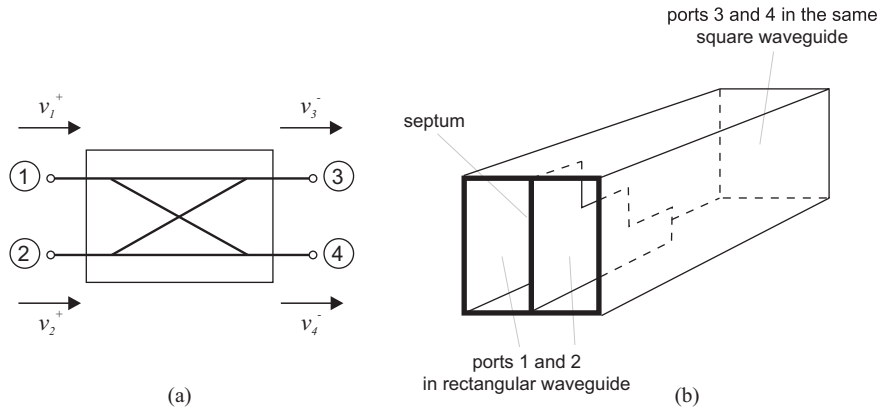


Figure 2.5: Illustration of four-port polarizer (a) and drawing of septum polarizer (b) with rectangular and square waveguide ports.

2.2.5 Polarizer for generating circular polarization

In order to generate dual circular polarization from two linear ones we need a polarizer. This can be realized in many different ways. In most cases we can represent it as a passive circuit component with four ports (Fig. 2.5), which for the ideal lossless case is described by the scattering matrix

$$[S_{ij}] = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 1 & -j \\ 0 & 0 & -j & 1 \\ 1 & -j & 0 & 0 \\ -j & 1 & 0 & 0 \end{bmatrix} ,$$

where $S_{ij} = V_i^-/V_j^+$, with V_i^- is the amplitude of the wave leaving port i, and V_j^+ is the amplitude of the wave exciting port j, both of them for the case when $V_k^+ = 0$ for $k \neq j$.

The polarizer works as follows: We connect ports 3 and 4 to two similar and lossless but orthogonally x- and y-polarized antennas which are impedance matched to the characteristic impedance of these ports. Then, from this excitation, there comes a propagating wave of the form

$$\mathbf{E}e^{-jkz} = (V_3^-\hat{\mathbf{x}} + V_4^-\hat{\mathbf{y}})e^{-jkz} .$$

If we excite port 1 with unit amplitude, i.e., $V_1^+ = 1$, the signal out of port 3 has amplitude $V_3^- = S_{31}V_1^+ = 1/\sqrt{2}$ and that out of port 4 has $V_4^- = S_{41}V_1^+ = -j/\sqrt{2}$. Thus, the combined field has the form

$$\mathbf{E}_1 = S_{31}\hat{\mathbf{x}} + S_{41}\hat{\mathbf{y}} = (\hat{\mathbf{x}} - j\hat{\mathbf{y}})/\sqrt{2} ,$$

which represents RHC polarization according to (2.21). When port 2 is excited, we get correspondingly

$$\mathbf{E}_2 = S_{32}\hat{\mathbf{x}} + S_{42}\hat{\mathbf{y}} = (-j\hat{\mathbf{x}} + \hat{\mathbf{y}})/\sqrt{2} = (-j)(\hat{\mathbf{x}} + j\hat{\mathbf{y}})/\sqrt{2} ,$$

which represents LHC polarization according to (2.29).

A polarizer can for instance be realized as a microwave circuit with four coaxial ports. It can also be realized with circular waveguide input and output. Then, each of the two orthogonal

linear polarizations in the single mechanical input of the circular waveguide represent ports 1 and 2, respectively. The single output circular waveguide contains both the two orthogonal linearly polarized output ports 3 and 4. Thus, in this case both input circuit ports 1 and 2 are present in the same physical circular input waveguide port, and similarly both output circuit ports 3 and 4 are present in the same physical output circular waveguide port. The septum polarizer is a very compact polarizer. This is realized with two physically separated rectangular waveguide input ports 1 and 2 (see Fig. 2.5b), and a physical quadratic or circular output waveguide containing both the circuit ports 3 and 4. The quadratic or circular waveguide output is very convenient, in particular for feeding circular waveguide horn antennas with dual circular polarization. The physically separate input ports are convenient as no extra orthomode transducer is needed.

2.2.6 Example: Mismatch in polarizer

a) Assume that the x- and y-polarized antennas on ports 3 and 4 in Fig. 2.5 have $75\,\Omega$ input impedances, whereas the polarizer ports have $50\,\Omega$ impedances. Determine the relative cross-polar level and the relative radiated power when we excite port 1 with a unit wave and port 2 is terminated with $50\,\Omega$.

b) Repeat the calculations for the case when port 2 is short-circuited.

SOLUTION:

a) When ports 3 and 4 are terminated by $75\,\Omega$, the reflection coefficients are

$$r = \frac{V_3^+}{V_3^-} = \frac{V_4^+}{V_4^-} = \frac{75-50}{75+50} = 0.2 \quad \text{(i.e., } -14\,\text{dB)} \ .$$

If port 1 is excited, the waves leaving ports 3 and 4 are $S_{31} = 1/\sqrt{2}$ and $S_{41} = -j/\sqrt{2}$ respectively. These are reflected and create output signals on ports 1 and 2 according to

$$V_1^- = rS_{13}(1/\sqrt{2}) + rS_{14}(-j/\sqrt{2}) = (r/2) - (r/2) = 0 \ ,$$
$$V_2^- = rS_{23}(1/\sqrt{2}) + rS_{24}(-j/\sqrt{2}) = -j(r/2) - j(r/2) = -jr \ .$$

Thus, equal reflections on ports 3 and 4 cause a wave out of port 2 (i.e., coupling to the orthogonal polarization). If port 2 is ideally impedance matched, $V_2^+ = 0$ and there is no reflection back into the polarizer. Therefore, the wave amplitudes on ports 3 and 4 do not change, so the polarization is ideally RHC as in *a)* with zero cross-polar level. The total radiated power, which is the mismatch factor of the polarizer circuit, becomes one minus the relative power lost in the dummy load at port 2:

$$e_{\text{rad}} = 1 - |r|^2 = 0.96 \quad \text{(i.e., } -0.18\,\text{dB)} \ .$$

b) In this case we still have $75\,\Omega$ on ports 3 and 4, but we now consider the case when port 2 is short-circuited. Then, $V_2^+ = -V_2^- = jr$, which gives a new output on ports 3 and 4 providing an E-field of the form

$$\mathbf{E}_2 = V_2^+ S_{32}\hat{\mathbf{x}} + V_2^+ S_{42}\hat{\mathbf{y}} = (r/\sqrt{2})(\hat{\mathbf{x}} + j\hat{\mathbf{y}}) \ ,$$

which is LHC and therefore cross-polar. This is again reflected by the 75 Ω load, giving rise to multiple reflections in the polarizer, but these do not destroy the cross-polar level determined already by $\mathbf{E}_2$. The relative cross-polar level becomes

$$(\mathsf{XP})_{\mathrm{dB}} = 20\log\left(\left|\frac{\mathbf{E}_2}{\mathbf{E}_1}\right|\right) = 20\log|r| = -14\,\mathrm{dB}\ .$$

And from (2.34) the axial ratio becomes

$$(\mathsf{AR})_{\mathrm{dB}} = 20\log\left|\frac{1+|r|}{1-|r|}\right| = 20\log\left|\frac{1.2}{0.8}\right| = 3.5\,\mathrm{dB}\ .$$

Therefore, mismatches after a polarizer circuit will either cause attenuation or create cross-polarization or both together.

2.3 Radiation fields

This section generally describes the radiation fields of antennas and how we characterize them. When we observe the radiated fields at a point r at a very large distance $\mathbf{r}$ from the antenna, the field variations with distance and direction $\hat{\mathbf{r}} = \mathbf{r}/r$ from the antenna become separable according to $\mathbf{E}(\mathbf{r}) = \mathbf{G}(\hat{\mathbf{r}})e^{-jkr}/r$. Therefore we need only to measure the direction dependence $\mathbf{G}(\hat{\mathbf{r}})$ of the amplitude and phase of the field in order to reconstruct it at any point far away from the antenna. This property is of course very useful, and it is of importance to know the distance outside which we can use it safely.

Moreover, there exist other theoretical cases for which the r dependence is separable from the direction dependence, and for which the r dependence is different from e^{-jkr}/r. For example, in two-dimensional (2D) field problems the field varies as $e^{-jkr}/\sqrt{r}$ at large distances. However, ideal 2D antennas are infinitely long and hence not realizable, so we will in real life always observe the e^{-jkr}/r dependence for large r. Still, some long antennas and scatterers may be treated approximately as 2D structures in order to facilitate the analysis within certain distances from them, but we will not consider such cases in this book.

Note that the above e^{-jkr}/r dependence is valid only when the antenna is located in free space (i.e., in vacuum). In reality there will be surrounding structures and the atmosphere that affects the radiation field. Still, the theoretical free-space environment is very useful in characterizing antennas. The effect of neighboring structures is often negligible if they are not located in the direction of the main beam. Thus, the vacuum is a good model for the atmosphere unless there is heavy rain, fog, smoke or atmospheric disturbances.

2.3.1 Field regions

The radiation field is most often interpreted to be the field in the *far-field* region at which the r and $\hat{\mathbf{r}}$ dependences of the fields are separable. This appears gradually when r increases, but it has appeared to a very high accuracy when [7]

$$r \geq 2D^2/\lambda\ . \tag{2.38}$$

[7] This relation will be derived in Section 4.2.

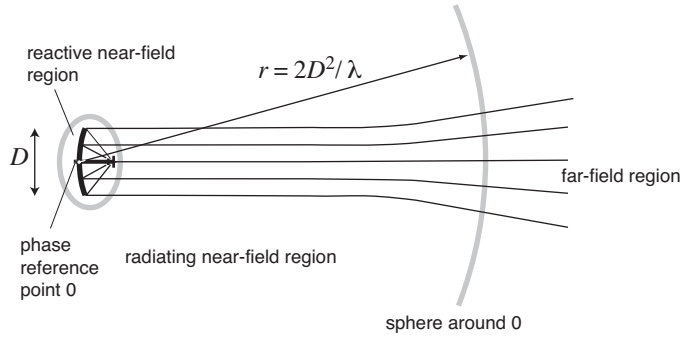


Figure 2.6: Illustration of different field regions around the antenna.

Here D is the largest diameter of the antenna, or more precisely the diameter of the smallest sphere that contains all structural parts of the antenna. Note that this sometimes may include a ground plane on which the antenna is located. The above condition is referred to as the *far-field* (or *Fraunhofer*) condition or criterion. The condition in (2.38) is not unique. Sometimes $r \geq (D^2/\lambda)$ is sufficient. It depends strongly on the antenna type and the environment. Also, sometimes D in (2.38) is replaced by an effective diameter. Therefore, in practical work it is important to state clearly the far-field condition being used. In theoretical work, we may simply assume that $r \to \infty$ in the far-field.

The region for which r is smaller than D^2/λ or $2D^2/\lambda$ is called the near-field. The near-field is also a radiation field. Indeed, the radiating power is dominating everywhere in the near-field except very close to the antenna, where reactive oscillating and nonradiating power may dominate. The extent of this inner *reactive near-field region* cannot be defined in general because it depends strongly on the antenna geometry. However, often the reactive effects vanish at 2-3 wavelengths away from the antenna. The outer and largest part of the near-field region is correspondingly called the *radiating near-field*. These different field regions are illustrated in Fig. 2.6.

We will only discuss far-fields in this book, except when studying the Gaussian beam in Section 5.7.

2.3.2 "Radiation fields" of receiving antennas

Most of this book is devoted to electromagnetic field analysis of antennas in order to determine their radiation patterns and equivalent circuits. All this analysis is done by considering radiating antennas (i.e., antennas in transmitting mode). In practice, most antennas are used both for transmission and reception, so the receiving mode analysis is equally important as the transmitting mode analysis. Despite this, the analysis methods and their terminology are closely linked to those used for the transmitting antenna. This can be justified by the fact that wave propagation satisfies reciprocity. There is an exception related to waves in some magnetic materials, but such materials are traditionally only used in nonradiating waveguide components and not in the radiating parts of an antenna.

Reciprocity[8], is a general relation between sources and resultant fields, or between the related currents and voltages. Reciprocity states that if a current source which is present at the port of an antenna a causes a certain voltage measured at the port of another antenna b (which may be located far away from antenna a), then, if we instead excite antenna b with the same current as antenna a had, we will measure the same voltage at the port of antenna a which we first measured at antenna b.

Physically, this means that the antenna structures and the environments between the two antennas affect the amplitude and phase of the transferred fields between the two ports in the same way independent of which direction the wave propagates. This is true independent of which antenna is transmitting and which is receiving. Therefore, it is sufficient only to analyze the antennas in transmitting mode. The performance on reception is reciprocal. Some papers (mainly old ones) calculate the properties of antennas when they operate in the receiving mode, but today the radiation analysis dominates because it is much simpler. Also it is far more easily extended from simple approximate solutions to more accurate and complicated numerical solutions. It is not at all recommended to try to perform field analysis of antennas in the receiving mode. Once the equivalent circuits of the antenna have been determined, an antenna can easily be analyzed as a part of the system, even in the receiving mode, by using network and circuit theory[9].

2.3.3 Far-field function and radiation intensity

From the discussion at the beginning of Section 2.3 we may write the far-field at a point $\mathbf{r}$ generally as

$$\mathbf{E}(\mathbf{r}) = \frac{1}{r} e^{-jkr} \mathbf{G}(\hat{\mathbf{r}}) \ , \tag{2.39}$$

where $1/r$ is the divergence factor, e^{-jkr} is the phase factor and $\mathbf{G}(\hat{\mathbf{r}})$ is the complex *far-field function* (we can also call it the radiation field function), and where

$$\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}} \ , \qquad r = \sqrt{x^2 + y^2 + z^2} \qquad \text{and} \qquad \hat{\mathbf{r}} = \mathbf{r}/r \ . \tag{2.40}$$

The radiation field is most conveniently represented in a spherical coordinate system with origin somewhere inside the antenna structure (Fig. 2.7). This gives

$$\hat{\mathbf{r}} = \sin\theta\cos\varphi\hat{\mathbf{x}} + \sin\theta\sin\varphi\hat{\mathbf{y}} + \cos\theta\hat{\mathbf{z}} \ , \tag{2.41}$$

where θ is the polar angle and φ is the azimuth angle. Therefore, we may also express the far-field as

$$\mathbf{E}(r,\theta,\varphi) = \frac{1}{r} e^{-jkr} \mathbf{G}(\theta,\varphi) \ . \tag{2.42}$$

Eq. (2.39) and (2.42) express the E-field both *a)* as a spherical wave originating from $r = 0$, and, b) locally for large r as a plane wave. In this book we use both the notations in (2.39) and (2.42). We will express the fields and far-field functions henceforth in terms of the unit vector $\hat{\mathbf{r}}$ or in terms of (θ,φ). That is, $\mathbf{G}(\hat{\mathbf{r}}) = \mathbf{G}(\theta,\varphi)$.

The corresponding H-field is

$$\mathbf{H}(r,\theta,\varphi) = \frac{1}{\eta}\hat{\mathbf{r}} \times \mathbf{E}(r,\theta,\varphi) \ , \tag{2.43}$$

[8] Reciprocity will be studied further in Section 4.5.

[9] For more information see Section 2.6.1

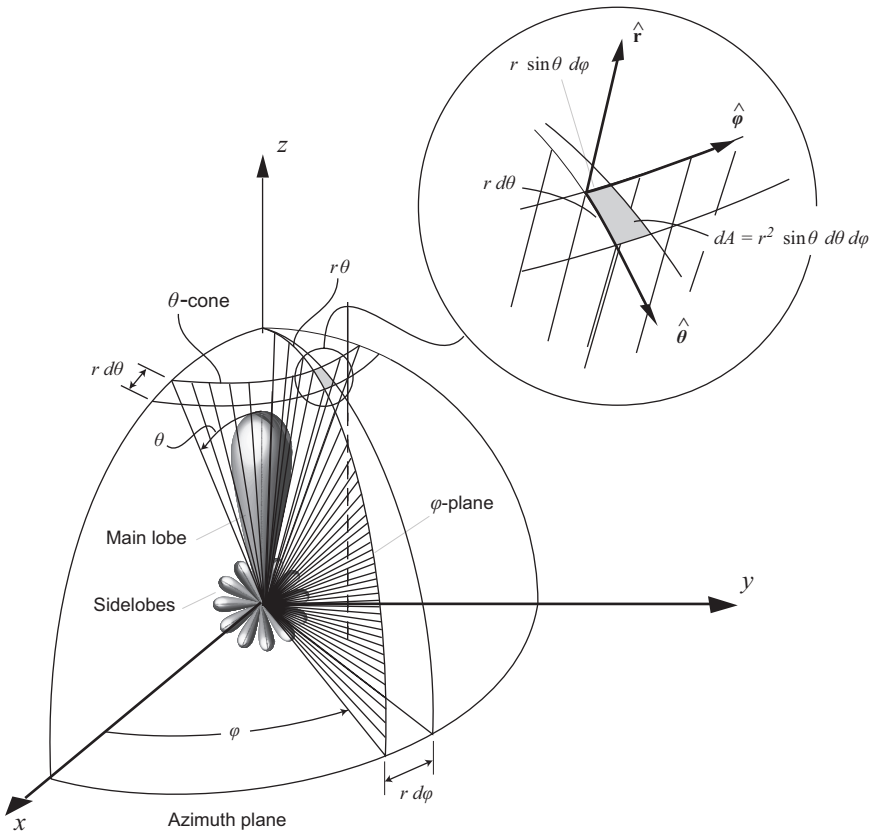


Figure 2.7: The spherical coordinate system for representation of radiation fields.

which is the same relation that was used for the plane wave in (2.9). The power density vector $\mathbf{W}_{\text{ave}}$ is found by using (2.7):

$$\mathbf{W}_{\text{ave}} = \frac{1}{2}\Re\{\mathbf{E}\times\mathbf{H}^*\} = \frac{1}{2}\frac{1}{\eta r^2}|\mathbf{G}(\theta,\varphi)|^2\hat{\mathbf{r}}\ . \tag{2.44}$$

For antennas, it is common to introduce the *radiation intensity* $U(\theta,\varphi)$, which is the radiated power per unit solid angle. This becomes

$$\begin{aligned} U(\theta,\varphi) &= r^2(\mathbf{W}_{av}\cdot\hat{\mathbf{r}}) = \frac{1}{2\eta}|\mathbf{G}(\theta,\varphi)|^2 \\ &= \frac{1}{2\eta}\left[|G_{\text{co}}(\theta,\varphi)|^2 + |G_{\text{xp}}(\theta,\varphi)|^2\right]\ , \end{aligned} \tag{2.45}$$

where the co- and cross-polar far-field functions $G_{\text{co}}(\theta,\varphi)$ and $G_{\text{xp}}(\theta,\varphi)$ are found by using co-and cross-polar unit vectors $\hat{\mathbf{co}}(\theta,\varphi)$ and $\hat{\mathbf{xp}}(\theta,\varphi)$ that are orthogonal to $\hat{\mathbf{r}}$, according to

$$G_{\text{co}}(\theta,\varphi) = \mathbf{G}(\theta,\varphi)\cdot\hat{\mathbf{co}}^*(\theta,\varphi)\ , \tag{2.46}$$

$$G_{\text{xp}}(\theta,\varphi) = \mathbf{G}(\theta,\varphi)\cdot\hat{\mathbf{xp}}^*(\theta,\varphi)\ . \tag{2.47}$$

The unit vectors $\hat{\mathbf{co}}(\theta,\varphi)$ and $\hat{\mathbf{xp}}(\theta,\varphi)$ can be defined for wave propagation in a direction $\hat{\mathbf{r}} = \hat{\mathbf{r}}(\theta,\varphi)$ in the same way as the corresponding unit vectors in Section 2.2 for a plane wave propagating in a direction $\hat{\mathbf{z}}$. This will be treated in Section 2.3.5.

2.3.4 Phase reference point and Fraunhofer approximation

The phase of the radiation field is determined by the imaginary part of the exponent of the phase factor $(-kr)$, plus the phase of the far-field function $\mathbf{G}(\theta,\varphi)$. The phase of the field at a given point r must be independent of where we locate the origin of the coordinate system, from where r is measured, as long as the antenna is located at the same physical position. Therefore, as the phase factor e^{-jkr} varies rapidly with even small relative changes in r (when r is large), the phase of the far-field function will change rapidly as well. From this it is clear that the far-field function has a phase *reference point*, which actually is the point from which r is measured. This is normally the center of the coordinate system.

Let us now see how the far-field function changes when the phase reference point is moved from the origin of the coordinate system to a point $\mathbf{r}_0 = x_0\hat{\mathbf{x}} + y_0\hat{\mathbf{y}} + z_0\hat{\mathbf{z}}$. The E-field $\mathbf{E}(r,\theta,\varphi)$ at a point $\mathbf{r} = \mathbf{R} + \mathbf{r}_0$ must be independent of the phase reference point for a given antenna. Therefore,

$$\mathbf{E}(r,\theta,\varphi) = \mathbf{E}(R,\theta,\varphi) \ , \tag{2.48}$$

where $\mathbf{E}(R,\theta,\varphi)$ is the E-field expressed in the translated coordinate system with origin at $\mathbf{r}_0$. By using (2.42) on both sides of (2.48) we get

$$\frac{1}{r}e^{-jkr}\mathbf{G}(\theta,\varphi) = \frac{1}{R}e^{-jkR}\mathbf{G}'(\theta,\varphi) \ , \tag{2.49}$$

where $R = |\mathbf{r} - \mathbf{r}_0|$. Let us now use the following two approximations, which are commonly known as the *Fraunhofer approximations*[10], and substitute

$$\frac{1}{r} \approx \frac{1}{R} \quad \text{and} \quad r \approx R + \mathbf{r}_0 \cdot \hat{\mathbf{r}} \ , \tag{2.50}$$

in the amplitude and phase expressions, respectively. These approximations are easily understood by studying Fig. 2.8. They are according to (2.38) valid when $R > 2r_0^2/\lambda$. Using them, the far-field function $\mathbf{G}'(\theta,\varphi)$ with the new phase reference point $\mathbf{r}_0$ becomes

$$\mathbf{G}'(\theta,\varphi) = \mathbf{G}(\theta,\varphi)e^{-jk(r-R)} = \mathbf{G}(\theta,\varphi)e^{-jk\mathbf{r}_0\cdot\hat{\mathbf{r}}} \ , \tag{2.51}$$

which is a simple phase transformation of the original far-field function $\mathbf{G}(\theta,\varphi)$.

If instead of moving the phase reference point, we move the antenna itself from the origin to a point $\mathbf{r}_A$ in the coordinate system, then the far-field function of the antenna in its new position becomes

$$\mathbf{G}_A(\theta,\varphi) = \mathbf{G}(\theta,\varphi)e^{jk\mathbf{r}_A\cdot\hat{\mathbf{r}}} \ , \tag{2.52}$$

when both $\mathbf{G}_A(\theta,\varphi)$ and $\mathbf{G}(\theta,\varphi)$ have their phase reference points in the origin of the coordinate system. This follows from similar arguments as those which were used to derive (2.51), using $\mathbf{r}_A = -\mathbf{r}_0$. When the coordinate system is moved, the field is the same in a given fixed point relative to the antenna. In contrast, given a coordinate system, when the antenna is moved over a small distance relative to a fixed observation point, the field at the observation point changes only by a phase factor. Recall that the latter holds only if $r > 2r_A^2/\lambda$ and $r > 2D^2/\lambda$.

[10] See Section 4.2.2.

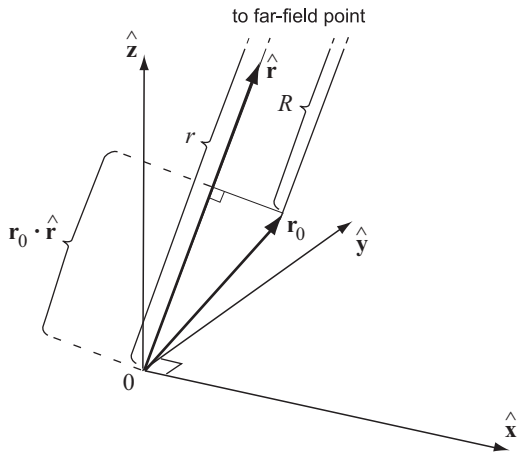


Figure 2.8: Moving the phase reference point from the origin to a point $\mathbf{r}_0$.

2.3.5 Polarization of radiation fields

The far-field is locally a plane wave, so we may use the formulas in Section 2.2 to find the co- and cross-polar components of it. To enable this we replace the unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ by those of a primed coordinate system with $\hat{\mathbf{z}}'(\theta,\varphi) = \hat{\mathbf{r}}$, and

$$\hat{\mathbf{x}}'(\theta,\varphi) = \cos\varphi\hat{\boldsymbol{\theta}} - \sin\varphi\hat{\boldsymbol{\varphi}} \ , \tag{2.53}$$

$$\hat{\mathbf{y}}'(\theta,\varphi) = \sin\varphi\hat{\boldsymbol{\theta}} + \cos\varphi\hat{\boldsymbol{\varphi}} \ , \tag{2.54}$$

where $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\varphi}}$ are the unit vectors in the directions of increasing θ and φ, respectively. Then, when $\theta = 0$, we see that $\hat{\mathbf{z}}' = \hat{\mathbf{r}} = \hat{\mathbf{z}}$, $\hat{\mathbf{x}}'(0,\varphi) = \hat{\mathbf{x}}$, and $\hat{\mathbf{y}}'(0,\varphi) = \hat{\mathbf{y}}$. Based on Eq. (2.19) and (2.20), in this new coordinate system we can define an arbitrary linear polarization by the co- and cross-polar unit vectors

$$\hat{\mathbf{co}}(\theta,\varphi) = \cos\xi\hat{\mathbf{x}}'(\theta,\varphi) + \sin\xi\hat{\mathbf{y}}'(\theta,\varphi) \ ,$$

$$\hat{\mathbf{xp}}(\theta,\varphi) = \sin\xi\hat{\mathbf{x}}'(\theta,\varphi) - \cos\xi\hat{\mathbf{y}}'(\theta,\varphi) \ ,$$

where $\xi = 0$ ($\xi = \pi/2$) corresponds to x-polarization (y-polarization) on the axis. We can substitute (2.53) and (2.54) and express $\hat{\mathbf{co}}(\theta,\varphi)$ and $\hat{\mathbf{xp}}(\theta,\varphi)$ in the alternative forms

$$\hat{\mathbf{co}}(\theta,\varphi) = \quad \cos(\varphi-\xi)\hat{\boldsymbol{\theta}} - \sin(\varphi-\xi)\hat{\boldsymbol{\varphi}} \ ,$$

$$\hat{\mathbf{xp}}(\theta,\varphi) = -\sin(\varphi-\xi)\hat{\boldsymbol{\theta}} - \cos(\varphi-\xi)\hat{\boldsymbol{\varphi}} \ .$$

These unit vectors define a pure linear polarization for all real choices of the angle ξ. The polarization performance of the antenna naturally will depend on how we choose ξ (i.e., on how we align $\hat{\mathbf{co}}$ and $\hat{\mathbf{xp}}$ relative to the radiated field of the antenna). One way to choose ξ is to optimize it for each direction (θ,φ) in such a way that the cross-polar level is as low as possible. Fortunately, the cross-polar characteristics of an antenna are mainly of interest inside the main beam. Inside narrow beams, the optimum value of ξ does not vary significantly. Therefore, for all pencil-beam antennas we may adjust ξ (e.g., align the

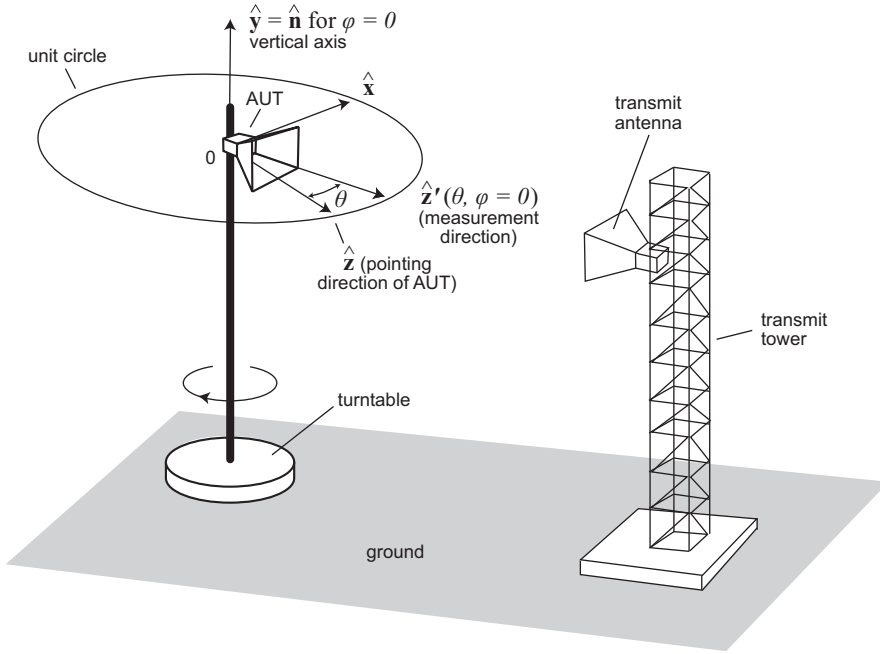


Figure 2.9: Coordinate systems for measuring co- and cross-polar fields according to Ludwig's third definition using a far-field measurement range. Note: The antenna is shown for the case when $\hat{\mathbf{y}} = \hat{\mathbf{n}}$, which means that we measure the radiation pattern in the $\varphi = 0^\circ$ plane.

antenna) for minimum cross-polarization on the axis and use this value of ξ inside the whole main beam.

Using this approach means that a y-polarized antenna is characterized by co- and cross-polar *unit vectors* of the respective forms

$$\hat{\mathbf{co}}(\theta, \varphi) = \hat{\mathbf{y}}'(\theta, \varphi) = \sin\varphi\hat{\boldsymbol{\theta}} + \cos\varphi\hat{\boldsymbol{\varphi}} \ , \tag{2.55}$$

$$\hat{\mathbf{xp}}(\theta, \varphi) = \hat{\mathbf{x}}'(\theta, \varphi) = \cos\varphi\hat{\boldsymbol{\theta}} - \sin\varphi\hat{\boldsymbol{\varphi}} \ . \tag{2.56}$$

Similar definitions can be used for linear x-polarization. The definition given in (2.55) - (2.56) is often referred to as *Ludwig's third definition* [2], which applies to pencil-beam antennas. It also applies to feeds used in reflector antennas, even when they have broad beams.[11]

For RHC polarization we get

$$\hat{\mathbf{co}}(\theta, \varphi) = [\hat{\mathbf{x}}'(\theta, \varphi) - j\hat{\mathbf{y}}'(\theta, \varphi)]/\sqrt{2} = e^{-j\varphi}[\hat{\boldsymbol{\theta}} - j\hat{\boldsymbol{\varphi}}]/\sqrt{2} \ , \tag{2.57}$$

$$\hat{\mathbf{xp}}(\theta, \varphi) = [\hat{\mathbf{x}}'(\theta, \varphi) + j\hat{\mathbf{y}}'(\theta, \varphi)]/\sqrt{2} = e^{j\varphi}[\hat{\boldsymbol{\theta}} + j\hat{\boldsymbol{\varphi}}]/\sqrt{2} \ , \tag{2.58}$$

and correspondingly for LHC polarization

$$\hat{\mathbf{co}}(\theta, \varphi) = [\hat{\mathbf{x}}'(\theta, \varphi) + j\hat{\mathbf{y}}'(\theta, \varphi)]/\sqrt{2} = e^{j\varphi}[\hat{\boldsymbol{\theta}} + j\hat{\boldsymbol{\varphi}}]/\sqrt{2} \ , \tag{2.59}$$

[11] See Section 9.2.2 and Section 9.3.1.

$$\mathbf{x\hat{p}}(\theta,\varphi) = [\hat{\mathbf{x}}'(\theta,\varphi) - j\hat{\mathbf{y}}'(\theta,\varphi)]/\sqrt{2} = e^{-j\varphi}[\hat{\boldsymbol{\theta}} - j\hat{\boldsymbol{\varphi}}]/\sqrt{2}\ . \tag{2.60}$$

The co- and cross-polar field components for linear polarization in (2.55) and (2.56) are easily measured by pointing the antenna horizontally and rotating it around the vertical axis. This can be done in most far-field measurement ranges, such as the one shown in Fig. 2.9. This consists of a tower with a fixed antenna at one end, and at the other end an Antenna Under Test (*AUT*) which is located on a table being rotated around a vertical axis $\hat{\mathbf{n}}$. Assume that the coordinate system of the antenna under test is located with the z-axis pointing in the same horizontal direction as the antenna radiates. The y-axis is aligned with the linear polarization of the antenna. Then, when we rotate the antenna around the vertical axis $\hat{\mathbf{n}}$, we observe the radiation field of the AUT at the port of the fixed antenna in the tower. This is recorded as a function of θ for a fixed φ. In fact, the φ-plane is determined from $\tan\varphi = (\hat{\mathbf{n}}\cdot\hat{\mathbf{x}})/(\hat{\mathbf{n}}\cdot\hat{\mathbf{y}})$. We measure perpendicular (*HOR*) polarization when the E-field direction $\hat{\mathbf{y}}$ is horizontal (i.e., $\hat{\mathbf{n}}\cdot\hat{\mathbf{y}} = 0$), corresponding to the plane $\varphi = 90°$. We measure parallel (*VER*) polarization when the E-field is vertical (i.e., $\hat{\mathbf{n}}\cdot\hat{\mathbf{y}} = 1$), corresponding to the plane $\varphi = 0°$ as shown in Fig. 2.9. The polarization of the antenna in the tower must be purely linear. Furthermore, when $\theta = 0$, we must align it with $\hat{\mathbf{y}}$ for measuring the co-polar component, and with $\hat{\mathbf{x}}$ for measuring the cross-polar component. These two orientations of the polarization of the antenna in the tower will give the directions $\hat{\mathbf{y}}'(\theta,\varphi)$ and $\hat{\mathbf{x}}'(\theta,\varphi)$ in the coordinate system of the AUT when this is rotated an angle θ around the vertical $\hat{\mathbf{n}}$ axis. Therefore, we measure the co- and cross-polar field components of the antenna under test as they are defined by the unit vectors in (2.55) and (2.56).

2.3.6 Co- and cross-polar radiation patterns

The co-polar radiation pattern is a graphical representation of the co-polar far-field function $|G_{\rm co}(\theta,\varphi)|$, the cross-polar radiation pattern is a graphical representation of the cross-polar far-field function $|G_{\rm xp}(\theta,\varphi)|$, and the co-polar phase pattern is a graphical representation of the copolar phase function $\Phi_{\rm co}(\theta,\varphi)$ which is defined by the relation

$$G_{\rm co}(\theta,\varphi) = |G_{\rm co}(\theta,\varphi)|e^{j\Phi_{\rm co}(\theta,\varphi)}\ .$$

The cross-polar phase pattern is generally of no interest.

The radiation patterns are normally presented as *polar or rectangular plots* showing the value of the far-field function in dB as a function of the polar angle θ for a given constant azimuth angle (i.e., in a given φ-plane as illustrated in Fig. 2.10). Polar plots usually are only used for antennas with very broad beams. The radiation patterns also can be presented as contour plots in terms of rectangular *uv-coordinates*, with $u = \sin\theta\cos\varphi$ and $v = \sin\theta\sin\varphi$ along the axes (Fig. 2.11). Such plots are also called uv-plots. *Contour plots* are mainly used for narrow beams, for which it is also common to use $\theta_x = \theta\cos\varphi$ and $\theta_y = \theta\sin\varphi$ in degrees along the axes. For several times in this book, we present *universal radiation patterns* in a rectangular plot as a function of a normalized projected direction $ka\sin\theta$, where a is the aperture diameter or width of the antenna (in the plane of the presented pattern) and $k = 2\pi/\lambda$ is the free-space wavenumber. Such patterns are valid for a large range of values of a.

Radiation patterns are always normalized in some way, e.g., to the level on axis at $\theta = 0$,

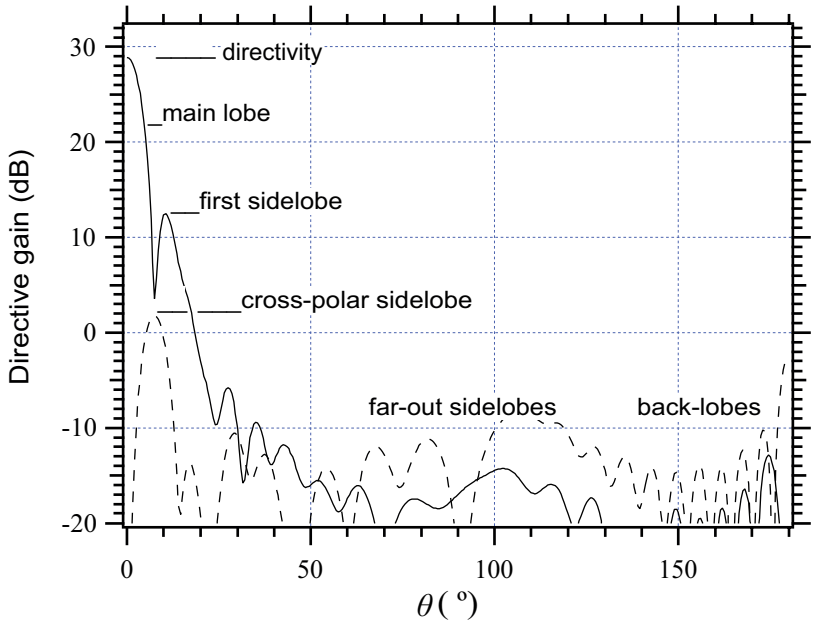


Figure 2.10: Example of rectangular plot of radiation pattern in any φ-plane.

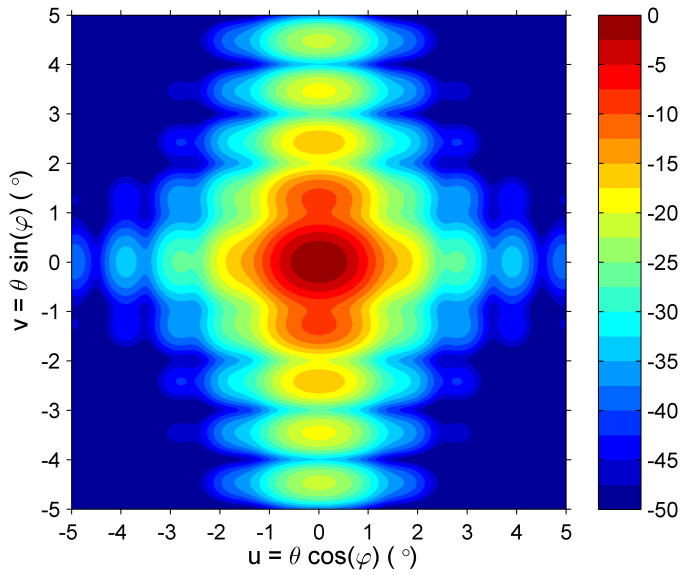


Figure 2.11: *Example of contour plot of the co-polar far-field function $G_{\mathrm{co}}(\theta, \varphi)$, in dB relative to $G_{\mathrm{co}}(0, \varphi)$.

according to

$$(G_{\rm co}(\theta,\varphi))_{\rm dB} = 10\log\left|\frac{G_{\rm co}(\theta,\varphi)}{G_{\rm co}(0,\varphi)}\right|^2 \ {\rm dB}\ , \tag{2.61}$$

$$(G_{\rm xp}(\theta,\varphi))_{\rm dB} = 10\log\left|\frac{G_{\rm xp}(\theta,\varphi)}{G_{\rm co}(0,\varphi)}\right|^2 \ {\rm dB}\ . \tag{2.62}$$

The best is to normalize the radiation patterns to the isotropic radiation level by using the total radiated power, as shown in Section 2.3.9. This normalized radiation intensity is referred to as the *directive gain*, and the unit is dBi="dB relative to isotropic level". The directive gain patterns are easy to use for design of microwave systems, because the power density can be calculated directly from the dBi levels and the total radiated power in Watts at any direction and distance.[12] Thus, antenna designers should therefore always present their radiation patterns in dBi. For this purpose the total radiated power needs to be evaluated from the measured or calculated radiation fields. This is laborious if the far-field function has many sidelobes, but there exist techniques to reduce the effort. These will be discussed in later chapters. It is also possible to present measured radiation patterns in dBi by comparing them against the measured gain in dBi of a reference antenna. Gain measurements are normally done by comparing the on-axis performance by that of a well-calibrated antenna (e.g., a so-called *standard gain horn*).

A typical radiation pattern has a main beam, nulls and sidelobes. The pointing direction of an antenna is the direction of the center of its main beam, often called the *boresight* direction. An observation *on axis* means in the center of the main beam.

2.3.7 Phase center

The *phase reference point*, which we already have defined, is often mistakenly called the *phase center* (and vice versa). However, it is very important to be aware of the difference between them. The phase center is the location of the center of curvature of the wavefront of the radiation fields. In other words, the phase center is the phase reference point which makes the phase of the far-field function constant. Such ideal phase centers seldom can be determined, so the following definition is more practical: The phase centre is the particular phase reference point which minimizes the phase variation of the co-polar far-field function $G_{\rm co}(\theta,\varphi)$ over a given solid angle of interest. The implementation of this definition will be different for different applications, and we will later in Chapter 8 introduce a specific one valid for feeds for reflector antennas. A good approximate formula for the location of the phase center of the co-polar radiation field within an angular region $0 < \theta < \theta_{\rm max}$ in a given φ-plane ($\varphi = \varphi_0$) is:

$$\frac{z_{\rm pc}}{\lambda} = \frac{\Phi_{\rm co}(0,\varphi_0) - \Phi_{\rm co}(\theta_{\rm max},\varphi_0)}{360^\circ(1-\cos\theta_{\rm max})}\ , \tag{2.63}$$

when $\Phi_{\rm co}(\theta,\varphi_0)$ is in degrees. This formula is often used in practice and is valid for symmetrical patterns; i.e., when $\Phi_{\rm co}(\theta,\varphi_0) = \Phi_{\rm co}(-\theta,\varphi_0)$ or more stringently when $\Phi_{\rm co}(\theta,\varphi_0) = \Phi_{\rm co}(\theta,\pi+\varphi_0)$. The formula can be derived from (2.51) by requiring that the co-polar phases of $\mathbf{G}'(\theta,\varphi)$ shall be equal at $\theta = 0$ and $\theta = \theta_{\rm max}$.

[12] See Section 2.5.3.

2.3.8 Total radiated power

The straightforward way to find the *total radiated power*, P_{rad}, is to evaluate the integral of the radiation intensity $U(\theta,\varphi)$ over the whole far-field sphere:

$$\begin{aligned} P_{\rm rad} &= \iint\limits_{\text{far-field sphere}} (\mathbf{W}_{\rm ave}\cdot\hat{\mathbf{r}})dA = \iint\limits_{4\pi} (\mathbf{W}_{\rm ave}\cdot\hat{\mathbf{r}})r^2\sin\theta d\theta d\varphi \\ &= \iint\limits_{4\pi} U(\theta,\varphi)\sin\theta d\theta d\varphi \\ &= \frac{1}{2\eta}\iint\limits_{4\pi}[|G_{\rm co}(\theta,\varphi)|^2+|G_{\rm xp}(\theta,\varphi)|^2]\sin\theta d\theta d\varphi\ . \end{aligned} \tag{2.64}$$

In this book we will also use another version of this integral in which we suppress the wave impedance η and the factor $1/2$. This new total *power integral* P is simply the integral of the square of the amplitude of the far-field function over the far-field sphere:

$$P = 2\eta P_{\rm rad} = \iint\limits_{4\pi}[|G_{\rm co}(\theta,\varphi)|^2+|G_{\rm xp}(\theta,\varphi)|^2]\sin\theta d\theta d\varphi\ . \tag{2.65}$$

The total power integral is sometimes very difficult to evaluate, in particular for large antennas with a lot of sidelobes. However, for rotationally symmetric antennas we can expand the φ-variation in a Fourier series and solve the φ-integral analytically. This simplifies the integration considerably.[13] For large-aperture antennas (e.g., horn antennas), we can replace (2.65) by an integral over the limited aperture plane[14] over which the power integral even may be given by a simple analytic expression. Furthermore, in reflector antennas, we often may perform the power integral over the simple far-field of the feed instead of over the complex far-field of the total antenna.[15]

The isotropic radiation level is used as a reference for the directive gain. This is defined by zero cross-polarization and

$$G_{\rm co}(\theta,\varphi) = G_{\rm ISO} = constant\ . \tag{2.66}$$

The power integral in this case is easily evaluated to be

$$P = |G_{\rm ISO}|^2\iint\limits_{4\pi}\sin\theta d\theta d\varphi = |G_{\rm ISO}|^2 4\pi\ . \tag{2.67}$$

2.3.9 Directive gain and directivity

The *directive gain* and its unit (dBi) was introduced in Section 2.3.6. It is defined by the normalization of the far-field function to that of an isotropic co-polar radiator which has the same total radiated power. Thus, the *directive gain* of the co-polar far-field function is defined by (using (2.67))

$$\begin{aligned} (D_{\rm co}(\theta,\varphi))_{\rm dBi} &= 10\log(|G_{\rm co}(\theta,\varphi)|^2/|G_{\rm ISO}|^2)\ \text{dBi} \\ &= 10\log(4\pi|G_{\rm co}(\theta,\varphi)|^2/P)\ \text{dBi}\ . \end{aligned} \tag{2.68}$$

[13] See Section 2.4.2.
[14] See Section 7.3.3.
[15] See Section 9.2.4.

The value of the directive gain in the center of the main beam is called the *directivity*. It is given by

$$\begin{aligned} &(D_0)_{\text{dBi}} = 10\log(4\pi|G_{\text{co}_0}|^2/P) \text{ dBi} , \\ \text{where} \qquad &|G_{\text{co}_0}| = \begin{cases} |G_{\text{co}}(0,0)| & \text{for centered beams} \\ |G_{\text{co}}(\theta,\varphi)|_{\max} & \text{otherwise} \end{cases} . \end{aligned} \tag{2.69}$$

where a centered beam means that the desired (but not necessarily the actual) main beam maximum is *on axis* at $\theta = 0°$. This may not necessarily coincide with the actual maximum of $G_{\text{co}}(\theta,\varphi)$. The direction of the desired (or actual) maximum of the main beam is referred to as the *pointing direction* of the antenna.

2.3.10 Beamwidth

There are often requirements for the width of the main beam of radiation patterns. The beamwidth is the angle in degrees between two points where the directive gain has the same value. This value may either be specified in dBi, or as a level (also called *taper*) *in dB relative to the directivity*. Examples of the latter are the 3 dB *beamwidth* $\theta_{3\,\text{dB}}$ (also called the half-power beamwidth) and the 10 dB beamwidth $\theta_{10\,\text{dB}}$, for which the levels are -3 dB and -10 dB, respectively, below the main beam maximum. In order to be extra clear, we may specify whether we mean the half or full beamwidth (i.e., from the center of the main beam at $\theta = 0$ and out or between the two opposite taper points). In practical work, the full beamwidth is most used. In this book we will always mean the half beamwidth when we write $\theta_{3\,\text{dB}}$ and similar, as we always have $\theta = 0$ on axis.

For narrow beam antennas, the directivity can be estimated from the beamwidth.[16]

2.3.11 Cross-polarization

The *directive gain* of the cross-polar far-field is defined in the same way as the co-polar directive gain, by

$$(D_{\text{xp}}(\theta,\varphi))_{\text{dBi}} = 10\log(4\pi|G_{\text{xp}}(\theta,\varphi)|^2/P) \text{ dBi} . \tag{2.70}$$

Some specific values of the cross-polarization are of interest, such as the *relative cross-polar level* on axis:

$$(\mathsf{XP}(0°))_{\text{dB}} = 10\log\left|\frac{G_{\text{xp}}(0°,0°)}{G_{\text{co}}(0°,0°)}\right|^2 \text{ dB} . \tag{2.71}$$

The corresponding power loss is represented by the *polarization efficiency* defined in (2.37):

$$e_{\text{pol}} = \frac{1}{1+|\mathsf{XP}(0°)|^2} = \frac{|G_{\text{co}}(0°,0°)|^2}{|G_{\text{co}}(0°,0°)|^2+|G_{\text{xp}}(0°,0°)|^2} . \tag{2.72}$$

The relative level of a *cross-polar sidelobe* showing up at $\theta = \theta_s$ in a plane $\varphi = \varphi_s$, becomes

$$(\mathsf{XP}(\theta_s))_{\text{dB}} = 10\log\left|\frac{G_{\text{xp}}(\theta_s,\varphi_s)}{G_{\text{co}}(0°,0°)}\right|^2 \text{ dB} . \tag{2.73}$$

[16] See Section 2.4.3.

The latter level is most conveniently given in dB. Such cross-polar requirements originate from a desire to avoid interference between two orthogonal communication channels on the same frequency for polarization diversity transmission (or reception), while using the same antenna.

2.3.12 Beam efficiency

Sometimes we need to know the relative power within a certain cone angle θ_0 around the pointing direction of the antenna (e.g., within the 20 dB beamwidth of the main beam) of a pencil-beam antenna. This relative power is referred to as the *beam efficiency* within θ_0, and is given by:

$$(e_{\text{BE}})_{\text{dB}} = 10 \log \left[\frac{P_{\text{co}}(\theta_0)}{P_{\text{co}}(\pi) + P_{\text{xp}}(\pi)} \right] \text{ dB} , \tag{2.74}$$

$$\text{where} \qquad P_{\text{co}}(\theta) = \int_0^{\theta} \int_0^{2\pi} |G_{\text{co}}(\theta, \varphi)|^2 \sin\theta d\varphi d\theta \tag{2.75}$$

$$\text{and} \qquad P_{\text{xp}}(\theta) = \int_0^{\theta} \int_0^{2\pi} |G_{\text{xp}}(\theta, \varphi)|^2 \sin\theta d\varphi d\theta \ . \tag{2.76}$$

The total *power integral* is given in this notation as

$$P = P_{\text{co}}(\pi) + P_{\text{xp}}(\pi) \ . \tag{2.77}$$

2.3.13 E- and H-plane patterns

Many antennas have two planes of symmetry. If such antennas are excited for linear polarization in such a way that the E-field on axis lies in one of the symmetry planes (e.g., if the E-field is oriented in $\hat{\mathbf{y}}$ direction so that $\hat{\mathbf{co}} = \hat{\mathbf{y}}$ for $\theta = 0°$), then the E-field will lie in the yz-plane for all θ-directions in the φ=90° plane. Therefore, this far-field function in the yz-plane is called the *E-plane pattern* and it is given by $|G_{\text{co}}(\theta, 90°)|$. The far-field function in the xz-plane is therefore called the *H-plane pattern*, and it is given by $|G_{\text{co}}(\theta, 0°)|$.

In practice we determine the E- and H-planes as follows. The mechanical structure of the antenna must have two planes of symmetry through the radiation axis (z-axis). This means that the antenna can also be rotationally symmetric. We study the exciting waveguide or electric current probe or dipole. The probe also could be penetrating into the waveguide of a coax-to- waveguide transition. Then, the E-field on the radiation axis has the same direction as the E-field in the center of waveguide, or the same as the electric current probe or dipole. The φ-plane, which coincides with this E-field direction on axis, is called the E-plane. The orthogonal plane coincides with the H-field direction on axis and is called the H-plane.

2.3.14 Fourier expansion of the radiation field

The φ-variation of the far-field function can always be expanded in a Fourier series, because φ is periodic with period 2π:

$$\begin{aligned}\mathbf{G}(\theta,\varphi) &= G_{\mathrm{co}}(\theta,\varphi)\hat{\mathbf{co}} + G_{\mathrm{xp}}(\theta,\varphi)\hat{\mathbf{xp}} \\ &= \sum_{n=0}^{\infty}[\mathsf{CO}_{\mathrm{sn}}(\theta)\sin(n\varphi) + \mathsf{CO}_{\mathrm{cn}}(\theta)\cos(n\varphi)]\hat{\mathbf{co}} + \\ &+ \sum_{n=0}^{\infty}[\mathsf{XP}_{\mathrm{sn}}(\theta)\sin(n\varphi) + \mathsf{XP}_{\mathrm{cn}}(\theta)\cos(n\varphi)]\hat{\mathbf{xp}}\ .\end{aligned} \tag{2.78}$$

Alternatively, we may expand it in this way;

$$\begin{aligned}\mathbf{G}(\theta,\varphi) &= G_{\theta}(\theta,\varphi)\hat{\boldsymbol{\theta}} + G_{\varphi}(\theta,\varphi)\hat{\boldsymbol{\varphi}} \\ &= \sum_{n=0}^{\infty}[A_n(\theta)\sin(n\varphi) + B_n(\theta)\cos(n\varphi)]\hat{\boldsymbol{\theta}} + \\ &+ \sum_{n=0}^{\infty}[C_n(\theta)\sin(n\varphi) - D_n(\theta)\cos(n\varphi)]\hat{\boldsymbol{\varphi}}\ ,\end{aligned} \tag{2.79}$$

where the minus sign in front of $D_n(\theta)$ is chosen of symmetry reasons. If the antenna is y-polarized with two symmetry planes, the latter Fourier expansion reduces to

$$\mathbf{G}(\theta,\varphi) = \sum_{n=0}^{\infty} A_n(\theta)\sin(n\varphi)\hat{\boldsymbol{\theta}} + \sum_{n=0}^{\infty} C_n(\theta)\sin(n\varphi)\hat{\boldsymbol{\varphi}}\ . \tag{2.80}$$

These expansions are convenient, and they can be used to analytically extend the far-field function to all φ-angles if it is known in only a few φ- planes.

2.3.15 Example: Phase reference point for asymmetric phase pattern

Consider an antenna with a given measured radiation pattern in the xz-plane. The antenna and its excitation have two planes of symmetry, so we expect a radiation pattern in the xz-plane which is symmetric around $\theta = 0^\circ$. However, the measured phase $\Phi_{\mathrm{co}}(\theta)$ shows an asymmetric pattern (see Fig. 2.12) with a slope at $\theta = 0^\circ$ which is

$$\left(\frac{\partial\Phi_{\mathrm{co}}(\theta)}{\partial\theta}\right)_{\theta=0} = a\ .$$

The reason for this is that the location of the rotation axis in the measurement setup does not cross the z-axis of the antenna. Find the x-coordinate of the *phase reference point*.

SOLUTION:

Let us assume that the location of the phase reference point (i.e., where the rotation axis cross the xz-plane) is at $\mathbf{r}_0 = x_0\hat{\mathbf{x}} + z_0\hat{\mathbf{z}}$. The phase of the co-polar radiation pattern when

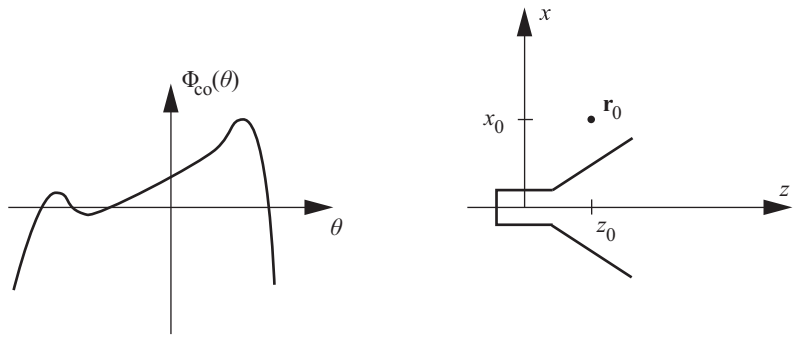


Figure 2.12: Illustration of asymmetric phase pattern (left) and antenna with coordinate system and location of rotation axis.

referred to $\mathbf{r}_0$ is, from (2.51):

$$\Phi'_{co}(\theta) = \Phi_{co}(\theta) - k\mathbf{r}_0 \cdot \hat{\mathbf{r}} \ ,$$

where $\hat{\mathbf{r}}$ is in a direction in the xz-plane (see Appendix C)

$$\hat{\mathbf{r}} = \sin\theta\hat{\mathbf{x}} + \cos\theta\hat{\mathbf{z}}$$

and $\Phi_{co}(\theta)$ is the phase pattern of the antenna when referred to a point on the symmetry axis. We get

$$\Phi'_{co}(\theta) = \Phi_{co}(\theta) - kx_0 \sin\theta - kz_0 \cos\theta \ .$$

Both $\Phi_{co}(\theta)$ and $\cos\theta$ are symmetric with θ-derivatives equal to zero when $\theta = 0$. Therefore,

$$\left(\frac{\partial\Phi'_{co}(\theta)}{\partial\theta}\right)_{\theta=0} = -kx_0(\cos\theta)_{\theta=0} = -kx_0 \ .$$

Thus, if the slope of the phase pattern at $\theta = 0$ is a, the x-coordinate of the phase reference point is

$$x_0 = -a/k = -\frac{a}{2\pi}\lambda.$$

2.3.16 Example: Calculation of phase center of a symmetric beam

Consider a quadratic phase pattern of the form

$$\Phi_{co}(\theta) = \Phi_{ctr} + \Phi_0(\theta/\theta_0)^2 \ ,$$

where $\theta = \pm\theta_0$ are the directions where the phase has changed by Φ_0 compared to the phase at $\theta = 0$. Find the location of the *phase center* within the angular region $0° < \theta < 30°$ when $\Phi_0 = 180°$ and $\theta_0 = 60°$. Also, determine the phase center in the limit when θ approaches 0.

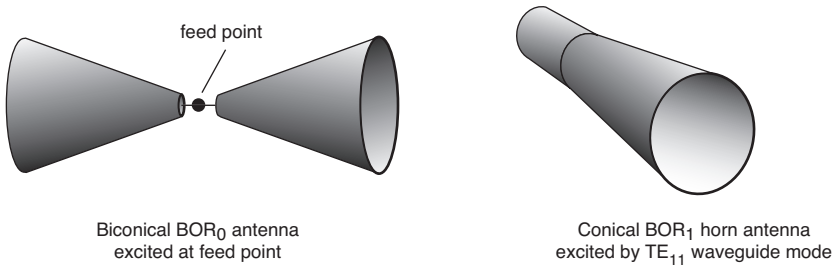


Figure 2.13: Examples of BOR_0 and BOR_1 antennas.

SOLUTION:

We have

$$\Delta\Phi = \Phi_{\mathrm{co}}(0) - \Phi_{\mathrm{co}}(\theta_{\max}) = -180^\circ(30^\circ/60^\circ)^2 = -45^\circ \ .$$

The location of the phase center is found by using (2.63):

$$z_{\mathrm{pc}} = \frac{-\lambda \cdot 45^\circ}{360^\circ(1-\cos\theta_{\max})} = -0.93\lambda \ ,$$

which means that the phase center is 0.93λ behind the phase reference point. The phase center at $\theta = 0$ can be found from the same formula by letting $\theta_{\max} \to 0$. This gives

$$\Delta\Phi(\theta_{\max}) = \Phi_{\mathrm{co}}(0) - \Phi_{\mathrm{co}}(\theta_{\max}) = -\Phi_{_0}(\theta_{\max}/\theta_{_0})^2$$

and

$$\begin{aligned} z_{\mathrm{pc}} &= \lim_{\theta_{\max}\to 0}\left(\frac{-\Phi_{_0}(\theta_{\max}/\theta_{_0})^2}{360^\circ(1-\cos\theta_{\max})}\right)\lambda \\ &= -\frac{\Phi_{_0}\lambda}{180^\circ\theta_{_0}^2}\left(\frac{180^\circ}{\pi}\right)^2 = -\left(\frac{180^\circ}{\pi\cdot 60^\circ}\right)^2\lambda = -0.91\lambda \ , \end{aligned}$$

which is obtained by using the expansion $\cos\theta = 1 - (\theta^2/2)$, which is valid for small θ in radians. Thus, the phase center for the direction $\theta = 0$ is located 0.91λ behind the phase reference point.

2.4 Rotationally symmetric antennas (BOR)

The mechanical structures of several antennas are rotationally symmetric, or, in other words, they are *bodies of revolution* (BOR), such as circular horn antennas. Still, BOR antennas do not necessarily have rotationally symmetric far-field functions. BOR antennas can be divided in two basic types, depending on how they are excited (Fig. 2.13).

2.4.1 BOR_0 antennas with rotationally symmetric radiation fields

In order to obtain a rotationally symmetric radiation field, the BOR antenna structure must be excited by a rotationally symmetric source, such as an electric or magnetic dipole aligned

along the symmetry axis. In the former case, the far-field function will be of the form, see (2.79),

$$\mathbf{G}_e(\theta,\varphi) = B_{_0}(\theta)\hat{\boldsymbol{\theta}} \tag{2.81}$$

and in the latter

$$\mathbf{G}_m(\theta,\varphi) = C_{_0}(\theta)\hat{\boldsymbol{\varphi}} \tag{2.82}$$

corresponding to entirely θ-directed and φ-directed E-fields, respectively. We will refer to such antennas as BOR_0 antennas. This is because the antenna structure is a body of revolution (BOR) and that the radiation field is of zero order variation in φ-direction according to the expansion in (2.79).

2.4.2 BOR_1 antennas

If a BOR antenna is excited by a short transverse current on the symmetry axis, the radiation field will only contain the $n = 1$ terms of the expansion in (2.79). The current source can be an incremental electric current[17] located on the z-axis and directed along $\hat{\mathbf{y}}$, where the far-field function is

$$\mathbf{G}_y(\theta,\varphi) = G_{_E}(\theta)\sin\varphi\hat{\boldsymbol{\theta}} + G_{_H}(\theta)\cos\varphi\hat{\boldsymbol{\varphi}} \ , \tag{2.83}$$

where $G_{_E}(\theta) = A_{_1}(\theta)$ and $G_{_H}(\theta) = C_{_1}(\theta)$ are seen to be the complex far-field functions in the E- and H-planes, respectively. It is important to note that always $G_{_E}(0) = G_{_H}(0)$, because the *E- and H-plane patterns* coincide for $\theta = 0$. We will refer to BOR antennas of this kind as y-polarized *BOR_1 antennas*, because the far-field function has only the first order variation in φ according to (2.80). In practice the equation in (2.83) is valid also when the exciting dipole has finite length (up to a half wavelength) provided it is centered on the symmetry z-axis of the structure. All rotationally symmetric antennas excited by TE_{11} (or similar) circular waveguide modes are also BOR_1 antennas (see Fig. 2.13, right side).

The BOR_1 form in (2.83) is important because it allows us to construct the whole far-field function from the E- and H-plane patterns only. By using *Ludwig's third definition* in (2.55) and (2.56), the co- and cross-polar far-field functions resulting from the BOR_1 form become

$$G_{\text{co}}(\theta,\varphi) = \mathbf{G}_y(\theta,\varphi)\cdot\hat{\mathbf{co}}^* = G_{\text{co}_{45^\circ}}(\theta) - G_{\text{xp}_{45^\circ}}(\theta)\cos 2\varphi \ , \tag{2.84}$$

$$G_{\text{xp}}(\theta,\varphi) = \mathbf{G}_y(\theta,\varphi)\cdot\hat{\mathbf{xp}}^* = G_{\text{xp}_{45^\circ}}(\theta)\sin 2\varphi \ , \tag{2.85}$$

where

$$G_{\text{co}_{45^\circ}}(\theta) = \frac{1}{2}[G_{_E}(\theta) + G_{_H}(\theta)] \tag{2.86}$$

$$G_{\text{xp}_{45^\circ}}(\theta) = \frac{1}{2}[G_{_E}(\theta) - G_{_H}(\theta)] \ , \tag{2.87}$$

are the complex co- and cross-polar far-field functions, respectively, in the $\varphi = 45^\circ$-plane. We refer to (2.84) through (2.87) as the *BOR_1 relations*. Fig. 2.14 shows E- and H-plane and 45°-plane radiation patterns of an example BOR_1 antenna. If we excite the BOR_1 antenna for linear x-polarization, we get

$$\mathbf{G}_x(\theta,\varphi) = \mathbf{G}_y(\theta,\varphi+\pi/2) = G_{_E}(\theta)\cos\varphi\hat{\boldsymbol{\theta}} - G_{_H}(\theta)\sin\varphi\hat{\boldsymbol{\varphi}} \ . \tag{2.88}$$

[17] See Section 4.4.1.

If we excite it for RHC polarization we get

$$\mathbf{G}_c(\theta,\varphi) = (\mathbf{G}_x - j\mathbf{G}_y)/\sqrt{2} = (G_{_E}(\theta)e^{-j\varphi}\hat{\boldsymbol{\theta}} - jG_{_H}(\theta)e^{-j\varphi}\hat{\boldsymbol{\varphi}})/\sqrt{2}\ . \tag{2.89}$$

The co- and cross-polar far-field functions in this RHC case become (using (2.57) and (2.58))

$$G_{\rm co}(\theta,\varphi) = \mathbf{G}_c \cdot \hat{\mathbf{co}}^* = G_{{\rm co}_{45^\circ}}(\theta)\ , \tag{2.90}$$

$$G_{\rm xp}(\theta,\varphi) = \mathbf{G}_c \cdot \hat{\mathbf{xp}}^* = G_{{\rm xp}_{45^\circ}}(\theta)e^{-j2\varphi}\ , \tag{2.91}$$

where $G_{{\rm co}_{45^\circ}}(\theta)$ and $G_{{\rm xp}_{45^\circ}}(\theta)$ are the same as in (2.86) and (2.87). Therefore, the cross-polarization when the $\rm BOR_1$ antenna is ideally excited for circular polarization is the same as the cross-polarization in the 45°-plane when it is ideally excited for linear polarization. Note that the cross-polarization equals to half of the difference between the complex E- and H-plane far-field functions. This means that the E- and H-plane far-field functions must be equal in both amplitude and phase in order to get zero cross-polarization in a $\rm BOR_1$ antenna.

The *power integral* of the $\rm BOR_1$ antenna has the following two equivalent forms:

$$P = \pi \int_0^\pi \left\{|G_{_E}(\theta)|^2 + |G_{_H}(\theta)|^2\right\} \sin\theta d\theta\ , \tag{2.92}$$

$$P = 2\pi \int_0^\pi \left\{|G_{{\rm co}_{45^\circ}}(\theta)|^2 + |G_{{\rm xp}_{45^\circ}}(\theta)|^2\right\} \sin\theta d\theta\ . \tag{2.93}$$

We can conveniently express the latter as the sum of the powers in the co-polar and cross-polar fields;

$$P = P_{\rm co} + P_{\rm xp}\ , \tag{2.94}$$

where for circular polarization

$$P_{\rm co} = P_{\rm coc} = 2\pi \int_0^\pi |G_{{\rm co}_{45^\circ}}(\theta)|^2 \sin\theta d\theta\ , \tag{2.95}$$

$$P_{\rm xp} = P_{\rm xpc} = 2\pi \int_0^\pi |G_{{\rm xp}_{45^\circ}}(\theta)|^2 \sin\theta d\theta\ . \tag{2.96}$$

For linear polarization, the co- and cross-polar parts $P_{\rm col}$ and $P_{\rm xpl}$ of the power integral for linear polarization are related to $P_{\rm coc}$ and $P_{\rm xpc}$ for circular polarization defined above by $P_{\rm co} = P_{\rm col} = P_{\rm coc} + P_{\rm xpc}/2$ and $P_{\rm xp} = P_{\rm xpl} = P_{\rm xpc}/2$.

2.4.3 Example: Directivity of $\rm BOR_1$ antenna with low sidelobes

It is often convenient to be able to represent an experimental radiation pattern by a simple analytical expression. There are three common expressions which well approximate main lobes of pencil-beams. These are $e^{-(\theta/\theta_0)^2}$, $\cos^n(\theta)$ and $\cos^n(\theta/2)$, where θ_0 and n are chosen in such a way that the analytical expression resembles the main beam as much as possible (e.g., in such a way that the 3 dB widths of the experimental and analytical patterns are equal). The three expressions represent narrow beams equally well, but the $\cos^n(\theta/2)$ *pattern* is the best for representing broad beams; therefore, we choose it here. Express the directivity as a function of the 3 dB beamwidth when the co-polar far-field function is $\cos^n(\theta/2)$ and there is no cross-polarization.

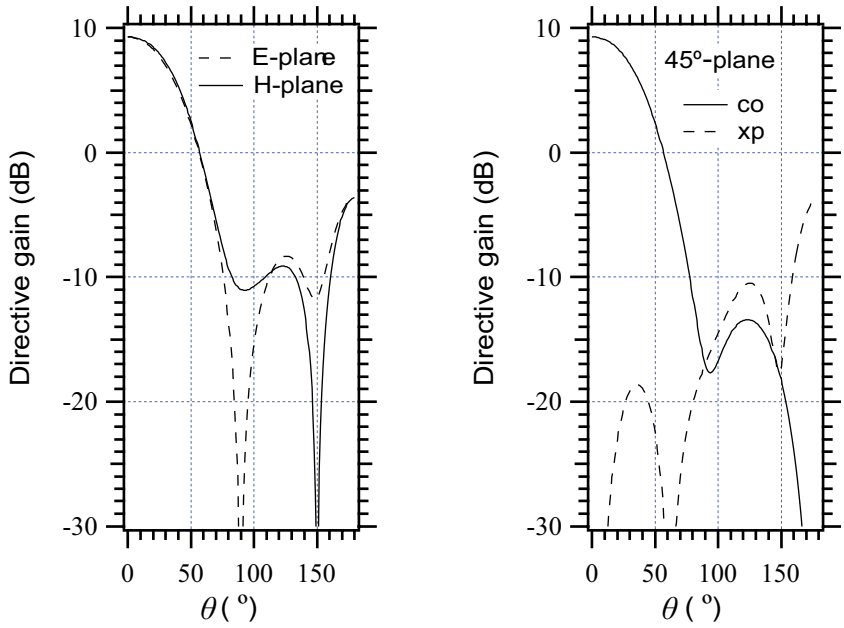


Figure 2.14: Examples of E- and H-plane (left) and 45°-plane co- and cross-polar (right) radiation patterns of BOR_1 antenna.

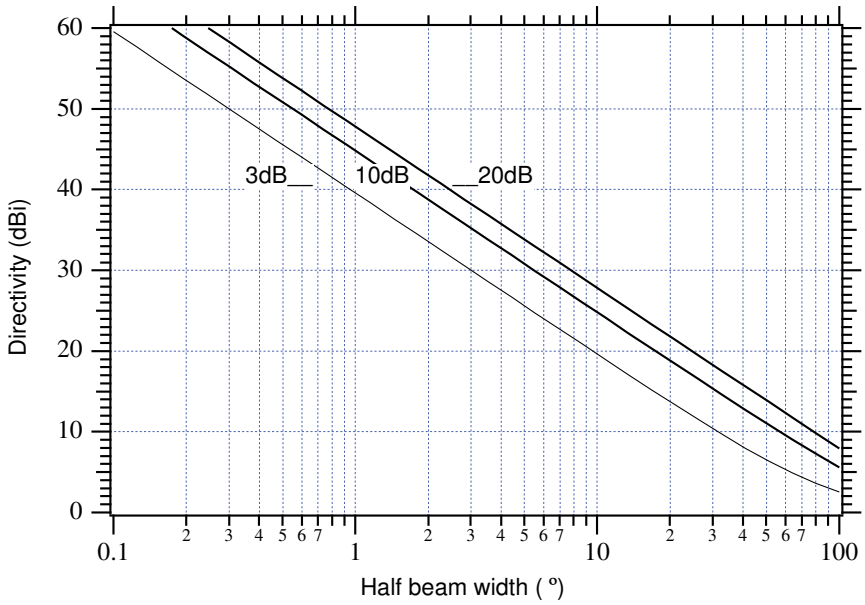


Figure 2.15: *Directivity of rotationally symmetric $\cos^n(\theta/2)$ radiation pattern as a function of half beamwidths in degrees for different tapers.

SOLUTION:

When the antenna is y-polarized, the approximate far-field function is

$$\mathbf{G}(\theta,\varphi) = \cos^n(\theta/2)(\sin\varphi\hat{\boldsymbol{\theta}} + \cos\varphi\hat{\boldsymbol{\varphi}}) \ .$$

The co-polar far-field function is $G_{\rm co}(\theta) = \cos^n(\theta/2)$. The parameter n can be expressed in terms of the half 3 dB beamwidth $\theta_{3\,\rm dB}$ of the experimental pattern by using $n20\log(\cos(\theta_{3\,\rm dB}/2))$ $= -3$, i.e.,

$$n = -3/(20\log(\cos(\theta_{3\,\rm dB}/2)) \ .$$

The power integral in (2.93) becomes

$$\begin{aligned} P &= 2\pi\int_0^\pi \cos^{2n}(\theta/2)\sin\theta d\theta \\ &= 8\pi\int_0^\pi \cos^{2n+1}(\theta/2)\sin(\theta/2)d(\theta/2) = 4\pi/(n+1) \ . \end{aligned}$$

Thus, the directivity is

$$D_{_0} = \frac{4\pi}{P} = (n+1) \ ,$$

which is plotted in dB in Fig. 2.15⋆. A more handy formula can be obtained by series expanding the result for small $\theta_{3\,\rm dB}$ (see Appendix B);

$$\log(\cos(\theta_{3\,\rm dB}/2)) \approx \log(1-(\theta_{3\,\rm dB})^2/8) \approx (\log{\rm e})(\theta_{3\,\rm dB})^2/8 \ .$$

This gives:

$$D_{_0} \approx 9070/(\theta^\circ_{3\,\rm dB})^2 \qquad \text{and} \qquad (D_{_0})_{\rm dB} \approx 39.6 - 10\log(\theta^\circ_{3\,\rm dB})^2{\rm dB} \ ,$$

where $\theta_{3\,\rm dB}$ in the latter formulas is in degrees. If the 3 dB half beamwidth in E- and H-planes are different, we may still use the same formulas but with $\theta_{3\,\rm dB}$ being the square root of the average of the square of the 3 dB half beamwidths in the two planes.[18]

2.4.4 Example: Directivity of BOR1 antenna with high far-out sidelobes

The theoretical $\cos^n(\theta/2)$ pattern has no sidelobe which practical antennas have. Assume that the antenna has a 3 dB half beamwidth of 5°. Show how much the directivity is reduced if the antenna has an increased sidelobe level envelope of $\mathsf{SE} = -20$ dB between $\theta_{_1} = 60^\circ$ and $\theta_{_2} = 80^\circ$.

SOLUTION:

We express the far-field function of the y-polarized antenna as

$$\begin{aligned} &\mathbf{G}(\theta,\varphi) = G(\theta)(\sin\varphi\hat{\boldsymbol{\theta}} + \cos\varphi\hat{\boldsymbol{\varphi}}) \ , \\ \text{with} \qquad &G(\theta) = \begin{cases} A + \cos^n(\theta/2) & \theta_{_1} < \theta < \theta_{_2} \\ \cos^n(\theta/2) & elsewhere \end{cases} \ . \end{aligned}$$

[18] See also Subsection 2.4.5 and ⋆.

By developing the square of $G(\theta)$ the power integral gets three terms;

$$P_1 = 2\pi \int_0^{\pi} \cos^{2n}(\theta/2) \sin\theta d\theta = 4\pi/(n+1) \ ,$$

$$P_2 = 2\pi|A|^2 \int_{\theta_1}^{\theta_2} \sin\theta d\theta = 2\pi|A|^2[\cos\theta_1 - \cos\theta_2] \ ,$$

$$P_3 = 4\pi|A| \int_{\theta_1}^{\theta_2} \cos^{n}(\theta/2) \sin\theta d\theta \ .$$

In order to get a useful approximate formula, we will here assume that $A \gg \cos^n(\theta/2)$ in the region $\theta_1 < \theta < \theta_2$ so that $P_3 \ll P_2$ and can be neglected. Furthermore, we will approximate P_2 by

$$P_2 \approx 2\pi|A|^2 \sin\theta_m \Delta\theta \ ,$$

where $\theta_m = (\theta_2 + \theta_1)/2$ is the average value of θ in the high sidelobe region and $\Delta\theta = \theta_2 - \theta_1$. This approximate P_2 is valid when $\theta_2 - \theta_1$ is small. The resulting directivity is

$$D = \frac{4\pi}{P_1 + P_2} = D_0 e_{\rm sl} \ ,$$

where $D_0 = n+1$ is the directivity when $A = 0$, and $e_{\rm sl}$ is the directivity reduction due to the sidelobes, given by

$$e_{\rm sl} = \frac{P_1}{P_1 + P_2} = \left(1 + \frac{P_2}{P_1}\right)^{-1} ;$$

$$\frac{P_2}{P_1} = \frac{1}{2}|A|^2 \sin\theta_m \left(\frac{\pi\Delta\theta}{180^\circ}\right) D_0 \ ,$$

where $\Delta\theta = \theta_2 - \theta_1$ is in degrees.

In the numerical example we have $\mathsf{SE} = -20\,\text{dB}$ between $\theta_1 = 60^\circ$ and $\theta_2 = 80^\circ$. The best result of the above formula is obtained if we calculate $|A|^2$ according to

$$|A|^2 = 10^{(\mathsf{SE}-3)/10} \ .$$

This can be justified by the fact that A is a sidelobe envelope, so that the sidelobes vary periodically under a maximum level of $|A|^2$. Taking the square of this amplitude and integrating corresponds to using a constant average sidelobe level which is 3 dB below the sidelobes' peaks (i.e., their envelope).

Thus, we reduce SE by 3 dB to the average power density level when calculating $|A|^2$. Finally,

$$n = -3/(20\log(\cos((5^\circ)/2))) = 363 \ , \qquad \text{gives} \qquad D_0 = 364 \quad (\text{i.e., } 25.6\,\text{dB}) \ ,$$

$$|A|^2 = 10^{-(20+3)/10} = 0.005 \ ,$$

$$\text{gives} \qquad P_2/P_1 = \frac{1}{2} 0.005 \sin 70^\circ \left(\frac{\pi \cdot 20^\circ}{180^\circ}\right) 364 = 0.30 \ ,$$

$$\text{and} \qquad e_{\rm sl} = (1 + 0.3)^{-1} = 0.77 \quad (\text{i.e., } -1.1\,\text{dB}) \ .$$

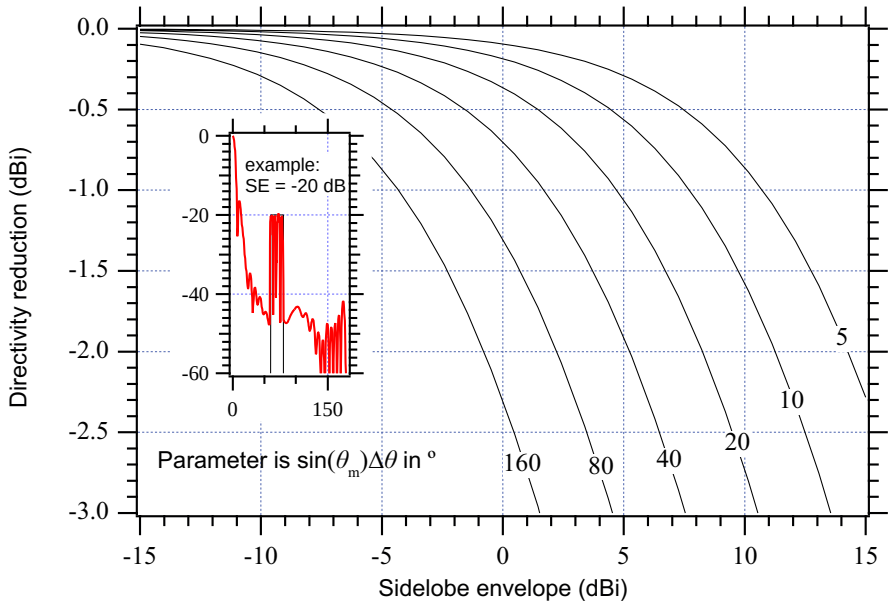


Figure 2.16: *Reduction in directivity as a function of uniform sidelobe level in dBi. The parameters of the curves are $\sin(\theta_m)\Delta\theta$ in degrees.

Thus, increased sidelobes may have a severe effect on the directivity. In our case the directivity reduces by 1.1 dB. The results of the above are presented in Fig. 2.16* as a function of a general sidelobe level in dBi, i.e., as a function of $10\log(|A|^2 D_0)$. Using the figure for our example we have a sidelobe envelope level $(|A|^2 D_0)_{\text{dBi}} = (25.6 - 20)\,\text{dBi} = 5.6\,\text{dBi}$ and the figure parameter is $\sin(70°) \cdot 20° \approx 17°$, from which we can read a directivity reduction of about -1.1 dB.

2.4.5 Example: BOR_1 antenna with different E- and H-plane patterns

If an antenna has different E- and H-plane patterns we can model it as

$$\mathbf{G}(\theta, \varphi) = \cos^{n_1}(\theta/2)\sin\varphi\hat{\boldsymbol{\theta}} + \cos^{n_2}(\theta/2)\cos\varphi\hat{\boldsymbol{\varphi}} \ .$$

Determine the directivity when the E- and H-plane patterns have 3 dB half-beamwidths of 20° and 25°, respectively. Find the relative cross-polar level at $\theta = 20°$ in the 45°-plane.

SOLUTION:

The power integral is from (2.92)

$$P = P_E + P_H;$$
$$P_E = \pi \int_0^{\pi} (\cos^{n_1}(\theta/2))^2 \sin\theta d\theta = 2\pi/(n_1 + 1) \ , \qquad P_H = 2\pi/(n_2 + 1) \ .$$

The directivity is

$$D_0 = 4\pi/(P_E + P_H) = \frac{2(n_1+1)(n_2+1)}{(n_1+n_2+2)} \ .$$

The numerical values are

$$n_1 = -3/20\log(\cos(20/2)) = 22.6 \ , \qquad n_2 = 14.4 \ , \qquad D_0 = 18.6 \quad (\text{i.e., } 12.7\,\text{dB}) \ .$$

The cross-polar pattern in the 45°-plane is from (2.87)

$$G_{\text{xp}_{45^\circ}}(\theta) = \frac{1}{2}(\cos^{n_1}(\theta/2) - \cos^{n_2}(\theta/2)).$$

We evaluate it for $\theta = 20^\circ$;

$$G_{\text{xp}_{45^\circ}}(20^\circ) = \frac{1}{2}(0.7075 - 0.8022) = 0.047 \ .$$

The relative cross-polar level is from (2.73)

$$|G_{\text{xp}_{45^\circ}}(20^\circ)/G_{\text{co}_{45^\circ}}(0)|^2 = (0.047)^2 \quad (\text{i.e., } -26.5\,\text{dB}) \ .$$

See also ⋆ for Fig. 2.15.

2.4.6 Example: BOR_1 antenna with different E- and H-plane phase patterns

If there are phase differences in E- and H-planes we may extend the $\cos^n(\theta/2)$ feed model according to

$$\mathbf{G}(\theta,\varphi) = \cos^n(\theta/2)[e^{jk\Delta(1-\cos\theta)}\sin\varphi\hat{\boldsymbol{\theta}} + e^{-jk\Delta(1-\cos\theta)}\cos\varphi\hat{\boldsymbol{\varphi}}] \ .$$

Determine the locations of the *phase centers* in E- and H-planes. Determine the directivity and the cross-polar level at $\theta = 20^\circ$ in the 45°-plane when the 3 dB half-beamwidth is 20° and $\Delta = 2\lambda$.

SOLUTION:

The phase center location Z_{pc_E} in E-plane ($\varphi = 90^\circ$) and Z_{pc_H} H-plane ($\varphi = 0^\circ$) can be found from (2.63). We have

$$\Phi_E(\theta) = k\Delta(1-\cos\theta) \qquad \text{and} \qquad \Phi_H(\theta) = -k\Delta(1-\cos\theta)$$

which yield

$$Z_{\text{pc}_E} = \frac{\Phi_E(0) - \Phi_E(\theta)}{k(1-\cos\theta)} = -\Delta$$

$$\text{and} \qquad Z_{\text{pc}_H} = \frac{\Phi_H(0) - \Phi_H(\theta)}{k(1-\cos\theta)} = \Delta \ .$$

Thus, the E-plane phase center is located a distance Δ behind the phase reference point, and the H-plane phase center a distance Δ in front of it.

The power integral is most easily evaluated by using (2.92), i.e.,

$$P = P_{_E} + P_{_H} = 2P_{_E} = 2\pi \int_0^\pi \cos^{2n}(\theta/2) \sin\theta d\theta = 4\pi/(n+1) \ .$$

Thus, the directivity is

$$D_{_0} = n + 1 \ .$$

The cross-polar far-field function in the 45°-plane is from (2.87)

$$\begin{aligned} G_{\mathrm{xp}_{45^\circ}}(\theta) &= \frac{1}{2} \cos^n(\theta/2)[e^{jk\Delta(1-\cos\theta)} - e^{-jk\Delta(1-\cos\theta)}] \\ &= \cos^n(\theta/2) j \sin(k\Delta(1-\cos\theta)) \ . \end{aligned}$$

The numerical values are

$$n = -3/20 \log(\cos(20/2)) = 22.6 \ ,$$

$$D_{_0} = n + 1 = 23.6 \quad \text{(i.e., 13.7 dBi)} \ ,$$

$$\begin{aligned} |G_{\mathrm{xp}_{45^\circ}}(20^\circ)/G_{\mathrm{co}_{45^\circ}}(0)|^2 &= \cos^{2n}(10^\circ) \sin^2(k\Delta(1-\cos(20^\circ))) \\ &= 0.5 \sin^2(360^\circ \cdot 2(1-0.95)) \\ &= 0.24 \quad \text{(i.e., } -6.3 \text{ dB)} \ . \end{aligned}$$

Thus, phase differences between E- and H-plane patterns cause cross-polarization in the 45°-plane.⋆[19]

2.5 System characteristics of the antenna

From the system's point of view an antenna can be regarded as a black box with one (or more) input ports and one or more outputs to free space. The system designer wants to know the characteristics of the antenna referred to the input port (terminal).

2.5.1 Antenna gain

The directivity was defined from the radiation intensity in the main beam direction and the total radiated power. The *antenna gain* (also called *realized gain* or *power gain* or simply the *gain*) $G_{_0}$ has the same definition, except that the total radiated power is replaced by the total power delivered to the antenna port. This may be expressed by using the directivity $D_{_0}$ as

$$G_{_0} = e_{\mathrm{rad}} e_{\mathrm{pol}} D_{_0} \ , \tag{2.97}$$

where e_{rad} is the *total radiation efficiency* and e_{pol} is the *polarization efficiency*. The total radiation efficiency is the ratio between the radiated power and the power incident on the antenna port. It can be further factorized in two subefficiencies, according to

$$e_{\mathrm{rad}} = e_r e_{\mathrm{abs}} \ , \tag{2.98}$$

[19] Phase differences between the E- and H-planes are also included in the MATLAB code for Fig. 2.15.

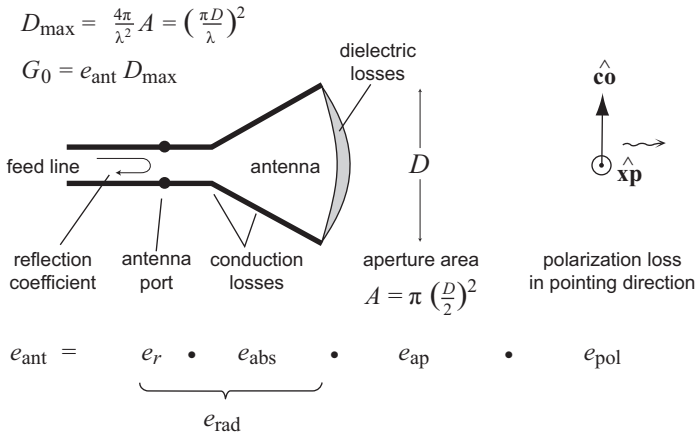


Figure 2.17: Example of lens-corrected horn antenna with input reflection coefficient r, conduction losses and dielectric losses.

where $e_r = 1 - |r|^2$ is the *mismatch factor* caused by a reflection coefficient r at the antenna port, and e_{abs} is the *radiation efficiency* due to ohmic losses (i.e., *absorption*) in the conductive and dielectric parts of the antenna structure (Fig. 2.17). The *polarization efficiency* e_{pol} may or may not be included already in D_0, depending on the on-axis characteristics of the far field function used to determine D_0. If this had no cross-polarization on axis, we may use e_{pol} to account for an actual system cross-polarization on axis, due to, e.g., a nonideal polarizer or tolerances. It is given by

$$e_{\text{pol}} = \left(\frac{\mathbf{E}(0°, 0°) \cdot \hat{\mathbf{co}}^*}{|\mathbf{E}(0°, 0°)|} \right)^2 . \tag{2.99}$$

2.5.2 Aperture efficiency and effective area

Several antennas have more or less well defined apertures, through which all or most of the radiated power propagates. The apertures may be real apertures as in horn antennas or virtual apertures as in front of reflector antennas or over planar array antennas[20]. Let us consider a plane aperture of area A. In Section 7.3.3 we will show that the *maximum available directivity* of such an aperture when radiating as an antenna is

$$D_{\max} = \frac{4\pi}{\lambda^2} A . \tag{2.100}$$

receiving antennas are often characterized in terms of their effective aperture instead of their antenna gain. The *effective aperture* A_e is defined as the ratio between the total power P_r received at the antenna port (or absorbed in a load with the required load impedance) and the power density W_t of the plane wave coming in from the *pointing direction* of the antenna:

$$A_e = P_r / W_t . \tag{2.101}$$

[20] For further information see Chapter 5.

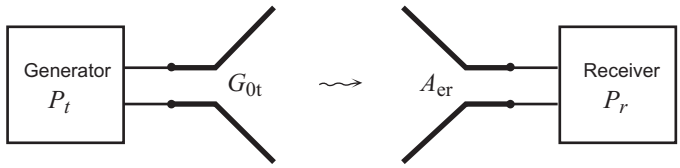


Figure 2.18: A communication system with transmitting (left) and receiving (right) antenna systems.

Eq. (2.100) can be used to express A_e in terms of the antenna gain, according to

$$A_e = \frac{\lambda^2}{4\pi} G_0 \ . \tag{2.102}$$

Eq. (2.101) and (2.102) are also valid for small antennas of which it is not possible to define any aperture area, e.g., for short dipoles. We will in Section 5.1.5 derive it from the input impedance and directivity of a short dipole. The effective aperture can be used to calculate the received power from an incident plane wave with known power density, as shown in the next subsection.

Let us now define the *aperture efficiency* e_{ap}. The most common definition is:

$$e_{\text{ap}} = \frac{D_0}{D_{\max}} \tag{2.103}$$

and therefore *not* $e_{\text{ap}} = G_0/G_{\max} = A_e/A$ which could be expected. Using (2.103) and (2.97) we can express the antenna gain as

$$G_0 = e_{\text{ant}} D_{\max} \ ; \qquad e_{\text{ant}} = e_r e_{\text{abs}} e_{\text{pol}} e_{\text{ap}} \ , \tag{2.104}$$

where e_{ant} is the total or overall *antenna efficiency*. This antenna efficiency includes four different subefficiencies. It is also possible to further factorize the aperture efficiency in further subefficiencies. We will postpone it to Section 9.4.1.

We see that the definition of the aperture efficiency gives the following relation between the aperture efficiency and the effective aperture;

$$A_e = e_{\text{ant}} A = e_{\text{rad}} e_{\text{pol}} e_{\text{ap}} A \ . \tag{2.105}$$

All the subefficiencies in (2.104) and (2.105) are conveniently used in practice to characterize both receiving and transmitting antennas, due to reciprocity. Almost all of them are, however, most easily determined and interpretable for transmitting antennas. When we present the values of all the different efficiencies we should always do it in dB according to $(e_{\text{ant}})_{\text{dB}} = 10\log(e_{\text{ant}})$.

2.5.3 Friis transmission equation and the radar equation

We will now show how the antenna gain and the effective aperture can be used to calculate the transmission loss between two antennas in a communication system. Let us consider a transmitting antenna with antenna gain G_{0t}, and a receiving antenna with effective aperture

A_{er}. The two antennas are pointing towards each other with coinciding and opposite directed pointing directions (Fig. 2.18). We assume that a power generator delivers the power P_t to the port of the transmitting antenna with no reflection. Then, if the antenna is a co-polar lossless isotropic radiator, the power density in the far-field at the location of the receiving antenna would be

$$W_{\mathrm{ISO}} = \frac{P_t}{4\pi r^2} \ , \tag{2.106}$$

where r is the spacing between the two antennas. The antenna has an antenna gain G_{0t} relative to the isotropic radiator, so we get a corresponding increase in the power density:

$$W_{_0} = \frac{P_t}{4\pi r^2} G_{0t} \ . \tag{2.107}$$

The power available from the port of the receiving antenna is now easily determined from the definition of the effective aperture:

$$P_r = W_{_0} A_{\mathrm{er}} = W_{_0} \frac{\lambda^2}{4\pi} G_{0r} = P_t \frac{1}{4\pi r^2} \frac{\lambda^2}{4\pi} G_{0t} G_{0r} \ , \tag{2.108}$$

where G_{0r} is the antenna gain of the receiving antenna. Finally, we get

$$\frac{P_r}{P_t} = \left(\frac{\lambda}{4\pi r} \right)^2 G_{0t} G_{0r} \ , \tag{2.109}$$

which is commonly referred to as the *Friis transmission formula*. The $(\lambda/4\pi r)^2$ factor is often called the *free space attenuation*. The effective aperture and the antenna gain can also be used to derive the ratio between P_t and P_r of a radar antenna, resulting in the so-called *radar equation*:

$$\frac{P_r}{P_t} = \left(\frac{1}{4\pi r^2} \right)^2 G_{0t} A_{\mathrm{er}} \cdot (\mathsf{RCS}) \ ,$$

where RCS is the radar cross section of the object (target) in square meters. The target is located a distance r away from the antenna, and

$$A_{\mathrm{er}} = \frac{\lambda^2}{4\pi} G_{0t} \ ,$$

since the same antenna is used for both transmit and receive (monostatic case).

2.5.4 Antenna noise temperature and G/T

The actual figure of merit of a communication or radar system is the *Signal-to-Noise Ratio* (SNR). SNR is the ratio of the signal power to the noise power. In antennas, the noise power P_n is most conveniently measured in terms of its equivalent *noise temperature* T, which is related to P_n through

$$P_n = kT\Delta f \ , \tag{2.110}$$

where k is Bolzmann's constant and Δf is the bandwidth of the system. T is measured in Kelvin (K). We do not need to know the value of Bolzmann's constant. It is sufficient in antenna analysis to know that the noise power is proportional to the noise temperature. The relation between the Kelvin and Centigrade (°C) temperature scales is linear, with $0\,\mathrm{K} = -273\,^{\circ}\mathrm{C}$ and $293\,\mathrm{K} = 20\,^{\circ}\mathrm{C}$.

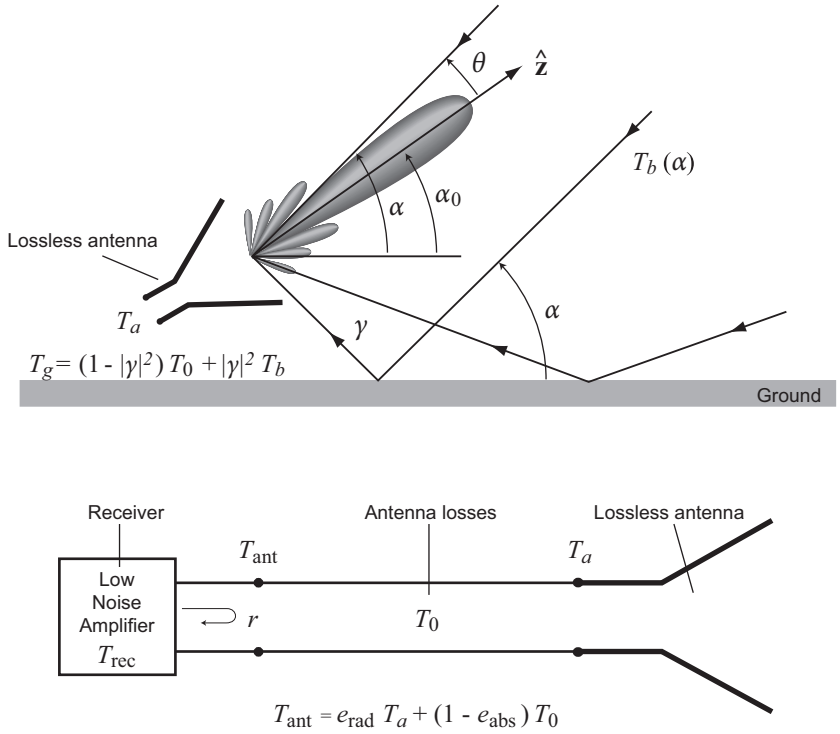


Figure 2.19: Geometry (upper) and system block diagram (lower) for calculation of antenna noise temperature. T_0 is the ambient temperature.

The received signal power is proportional to the *antenna gain* G_0, so a commonly used figure of merit for receiving antennas is

$$G_0/T = (e_{\text{ant}} D_{\max})/T_{\text{sys}} \ . \tag{2.111}$$

This 'G over T' is always expressed in dB/K (pronounced 'dB over K') which is dB relative to 1/K ('one over K'). The *system noise temperature* T_{sys} can be divided in the *receiver noise* T_{rec} and the *antenna noise* T_{ant} according to

$$T_{\text{sys}} = T_{\text{rec}} + T_{\text{ant}} \ . \tag{2.112}$$

Note that T_{rec} is a function of the reflection coefficient r, the physical temperature of the receiver and its noise parameters. The latter T_{ant} contains many contributions. For example, losses in the atmosphere and the surrounding ground, as well as in conducting or dielectric parts of the antenna. Also, the sun and other discrete noise sources in the sky may contribute. The basic noise model of a lossy medium and its physical explanation are as follows. We consider first a wave propagating through the medium. If this has a power density of unity when it enters the medium, the power density at the end of the medium is less, i.e., $e_{\text{abs}} < 1$. The lost power is absorbed, corresponding to excitation of resonances in the molecules or atoms in the medium, and it is transformed to heat. When no impressed wave is present, the molecular resonances will keep the same balance between heat (temperature) and microwave power density. Therefore, the medium always will radiate incoherent microwaves related to

its absolute temperature

$$T_{\text{abs}} = (1 - e_{\text{abs}})T_{\text{med}} \ , \tag{2.113}$$

where T_{med} is the physical temperature of the medium in Kelvin. T_{med} will normally be the ambient temperature T_0, which is used in the illustration in Fig. 2.19. The noise temperature may vary with frequency, but only if e_{abs} does. Thus, a medium which absorbs all the incident waves so that $e_{\text{abs}} = 0$ (often referred to as a black body) has an equivalent noise temperature equal to its physical temperature.

The sky, including the Earth's atmosphere, contributes to the antenna noise temperature through its *brightness temperature* T_b which is due to the physical molecular temperature, scattering and ohmic losses in the atmosphere. The brightness temperature is the noise seen by a narrow antenna beam pointing to the sky in a certain direction. T_b is a function of the elevation pointing direction α_0 of the antenna and the azimuth angle. The azimuth variation is normally negligible, and the elevation variation can be found tabulated in certain frequency bands. Such tables sometimes include negative elevation angles, which corresponds to the case that the main lobe points towards the ground. However, these values will in reality depend strongly on the conditions of the ground, for instance, its reflection coefficient. The *ground noise* contribution from a ground direction can be expressed as

$$T_g = [1 - |\gamma|^2]T_0 + |\gamma|^2 T_b \ , \tag{2.114}$$

where T_0 is the *ambient temperature*, standardized to $T_0 = 293\,\text{K}$, which corresponds to $20\,^\circ\text{C}$, γ is the reflection coefficient of the ground for the given incidence angle, and T_b is the brightness temperature of the sky in the direction of the ground reflection. From T_g and T_b it is possible to define a noise temperature profile $T_{\text{bg}}(\alpha)$, where α is the elevation angle, which is valid also for negative α, by

$$T_{\text{bg}}(\alpha) = \begin{cases} T_b(\alpha) & \text{for} \quad 0^\circ < \alpha < 90^\circ \\ T_g(\alpha) & \text{for} \quad -90^\circ < \alpha < \ 0^\circ \end{cases} \ . \tag{2.115}$$

An actual antenna has a far-field function $\mathbf{G}(\theta, \varphi)$ which illuminates the whole sky and ground. Therefore, the noise temperature contribution T_a from the far-field function of the antenna (we may call it the *ideal-antenna noise temperature*) has to be calculated as a weighted average over all directions to the sky and ground (see Fig. 2.19):

$$T_a = \frac{\iint_{4\pi} T_{\text{bg}}(\alpha(\theta,\varphi))[|G_{\text{co}}(\theta,\varphi)|^2 + |G_{\text{xp}}(\theta,\varphi)|^2] \sin\theta d\theta d\varphi}{\iint_{4\pi} [|G_{\text{co}}(\theta,\varphi)|^2 + |G_{\text{xp}}(\theta,\varphi)|^2] \sin\theta d\theta d\varphi} \ , \tag{2.116}$$

where $\alpha(\theta, \varphi)$ is the elevation angle calculated in the direction (θ, φ) in the coordinate system of the far-field function, $G_{\text{co}}(\theta, \varphi)$ is the co-polar far-field function, and $G_{\text{xp}}(\theta, \varphi)$ is the cross-polar. It is common in simple calculations to approximate (2.116) by

$$T_a = (1 - p_{\text{grd}})T_{\text{mb}} + p_{\text{grd}} T_0 \ , \tag{2.117}$$

where p_{grd} is the relative power hitting the ground and T_{mb} is the brightness temperature in the direction of the main beam. In this case we have assumed reflection coefficient $\gamma = 0$ at the ground, so that $T_g = T_0$ from (2.114).

An actual non-ideal antenna will also have losses and mismatch, so that the *actual-antenna noise temperature* referred to the receiving port becomes

$$T_{\text{ant}} = e_{\text{rad}} T_a + T_{\text{abs}} \ ; \qquad T_{\text{abs}} = (1 - e_{\text{abs}})T_0 \ , \tag{2.118}$$

where $e_{\rm abs}$ is the *radiation efficiency* due to absorption and $e_{\rm rad} = e_{\rm abs}(1-|r|^2)$ is the *total radiation efficiency*[21]. A lossy transmission line can be included in $e_{\rm abs}$. The following simple relation between $(e_{\rm abs})_{\rm dB}$ and $T_{\rm abs}$ can be derived for small losses and is convenient for quick calculations,

$$T_{\rm abs} = -70(e_{\rm abs})_{\rm dB}{\rm K} \ . \tag{2.119}$$

Thus, an ohmic loss of 0.1 dB gives 7 K contribution to the system noise. The receiver is often a *Low Noise Amplifier* (LNA) and its noise performance may be given in terms of the *noise figure*

$$\mathsf{NF} = 10\log\left(1+\frac{T_{\rm rec}}{T_0}\right) \ ; \qquad T_{\rm rec} = T_0[10^{(\mathsf{NF})/10}-1] \ . \tag{2.120}$$

The noise figure of a good transistor receiver can be lower than 1 dB, corresponding to 76 K. In radio telescopes the receivers are often cooled with liquid Helium to give noise temperatures down to 5 K.

2.5.5 Bandwidth

The characteristics of antennas are normally specified over a certain desired *bandwidth* Δf around the center frequency f_0. The operational frequency band is then defined by

$$f_0 - \Delta f/2 < f < f_0 + \Delta f/2 \ . \tag{2.121}$$

The performance of most antennas can be scaled in frequency by scaling all the dimensions in such a way that the dimensions become the same in terms of wavelengths. Therefore, it is very convenient to describe different antennas in terms of their relative bandwidth. This can be defined in percent by

$$(\Delta f)_{\%} = (100\Delta f/f_0)\% \ . \tag{2.122}$$

For large bandwidths the relative bandwidth is more clearly defined by the ratio

$$(f_{\rm max}/f_{\rm min}) : 1 \ , \tag{2.123}$$

where $f_{\rm max}$ and $f_{\rm min}$ are the highest and lowest frequencies of the frequency band. This means that

$$(f_{\rm max}/f_{\rm min}) = \frac{f_0+\Delta f/2}{f_0-\Delta f/2} \approx 1+\left(\frac{\Delta f}{f_0}\right) \ , \tag{2.124}$$

where the latter approximation is valid only for small $\Delta f/f_0 \ll 1$.

2.5.6 Tolerances

The electrical characteristics of antennas are always linked to their mechanical *tolerances*. In some cases (e.g., in large reflector antennas) the system characteristics may be described in terms of the tolerances of a major critical part of the antenna (e.g., the main reflector). The tolerances of a reflector is often given in terms of its *Root-Mean-Square* (*RMS*) value[22][11].

[21] Which was seen before in Section 2.5.1.

[22] RMS = the square root of the mean square error from an ideal surface.

This is obtained by calculating the deviations of the actual surface from its best fit theoretical ideal choice, averaging the square of this deviation over the whole surface, and finally taking the square root of the averaged value. A reflector must typically have an RMS accuracy better than $\lambda/50$ for loss of gain less than 0.3 dB.

The tolerance requirements for other mechanical parts depend strongly on the antenna type and how strong the requirements for sidelobes and return loss are. The requirements are also very different for critical and noncritical parts of the antennas. The critical parts are normally those having the strongest current densities or fields. In a critical part of a high performance antenna, the tolerance requirement may be as strong as $\lambda/500$. Noncritical dimensions may deviate several wavelengths from the optimum.

2.5.7 Environmental effects

A customer would like to specify the performance when the antenna is located in its actual operational *environment*. So, the antennas need to be characterized under certain temperature, wind, rain, snow and ice conditions. They also may need to withstand shaking and acceleration. In particular, there are strong environmental requirements for space antennas, referred to as space qualification requirements. Therefore, antenna design involves a lot of mechanical considerations and knowledge of material characteristics. Sometimes antennas are protected towards environmental effects by an enclosing dielectric shell, referred to as a *radome*. Radomes can be of different types, such as space-frame radomes (which are supported by a metallic space-frame structure), thin dielectric radomes (with wall-thickness much smaller than a half wavelength), thick tuned dielectric radomes (with wall thickness equal to half a wavelength of the radome material), and sandwich foam radomes (consisting of thin dielectric sheets around a foam core of a quarter or three quarters wavelength thickness).

Rain, snow and hail have two different effects on an antenna system. First, the wave propagation in the atmosphere is attenuated due to both scattering from the individual particles (i.e., drops, flakes and hailstones) and losses in the water molecules. Second, there may exist direct effects on the antenna, such as:

a) Rain generating a water film on the antennas or their radomes, causing attenuation due to wave propagation through the film or reflection from it.

b) Rain, or melted snow and hail, penetrating into cables, waveguides or antenna elements, causing increased loss, strong impedance mismatches and even short-circuiting of antenna components.

c) Accumulation of snow; dry snow has very little effect on the wave propagation in the low frequency region, but wet snow may be almost as bad as water.

d) Accumulation of ice is very severe because it is heavy and may destroy the whole mechanical antenna or radome construction, and the wave propagation through ice is very different from air.

The atmospheric effects of rain, snow and hail are not considered as a part of the antenna characteristics, rather as a part of the loss budget of the communication system. All the effects in steps **a)** to **d)** above are parts of the antenna characteristics. This means that

engineers often need to analyze and design antennas as if they are radiating in free space, but with included layers of water, snow or ice on places where these may accumulate.

2.5.8 Example: Noise temperature and G/T

Consider an antenna with an antenna gain of 28.5 dBi. The system noise temperature is 100 K. The brightness temperature of the sky is 5 K. Such low brightness temperatures are present at L-band when the antenna is not pointed towards a source of cosmic noise.

a) How much will the G/T in dB improve if we remove a cable with 0.1 dB loss between the receiver and the antenna?

b) How much will the G/T in dB improve if we reduce the relative power intersecting the ground from 20 % to 10 %?

SOLUTION:

The G/T of the system is

$$(G/T)_{\mathrm{dB}} = (G)_{\mathrm{dB}} - (T_{\mathrm{sys}})_{\mathrm{dB/K}} = 28.5\,\mathrm{dB} - 20\,\mathrm{dB/K} = 8.5\,\mathrm{dB/K}\ .$$

a) A cable with 0.1 dB loss represents a noise temperature of (use (2.119))

$$T_{\mathrm{cbl}} = 70 \cdot (0.1\,\mathrm{dB})\mathrm{K} = 7\,\mathrm{K}\ .$$

If we remove the cable, the gain increases by 0.1 dB and the system noise temperature decreases by 7 K. The new G/T becomes

$$\begin{aligned}(G/T)_{\mathrm{dB}} &= (G)_{\mathrm{dB}} + 0.1\,\mathrm{dB} - 10\log((T_{\mathrm{sys}} - 7\,\mathrm{K})/(1\,\mathrm{K})) \\ &= 28.6\,\mathrm{dB} - 19.7\,\mathrm{dB/K} = 8.9\,\mathrm{dB/K}\ .\end{aligned} \tag{2.125}$$

Thus, the G/T increases by 0.4 dB when we remove a cable with 0.1 dB loss. If the system noise temperature was less, the increase would have been even more.

b) Using (2.117) with 20 % of the power intersecting the ground, for the antenna noise we have

$$T_{a_1} = (1 - 0.2) \cdot 5\,\mathrm{K} + 0.2 \cdot 293\,\mathrm{K}\ .$$

Correspondingly, with 10 % power intersecting the ground

$$T_{a_2} = (1 - 0.1) \cdot 5\,\mathrm{K} + 0.1 \cdot 293\,\mathrm{K}\ .$$

Thus, the antenna noise temperature decreases by

$$\Delta T_a = T_{a_1} - T_{a_2} = 0.1 \cdot 5\,\mathrm{K} + 0.1 \cdot 293\,\mathrm{K} = 28.8\,\mathrm{K}\ ,$$

and the new G/T becomes

$$\begin{aligned}(G/T)_{\mathrm{dB}} &= (G_{\mathrm{dB}}) - 10\log((T_{\mathrm{sys}} - 28.8\,\mathrm{K})/(1\,\mathrm{K})) \\ &= 28.6\,\mathrm{dB} - 18.1\,\mathrm{dB/K} = 10.5\,\mathrm{dB/K}\ .\end{aligned}$$

Thus, the G/T increases by 2 dB when we reduce the power hitting the ground from 20 % to 10 %. If the system noise temperature were less, the increase would be even more.

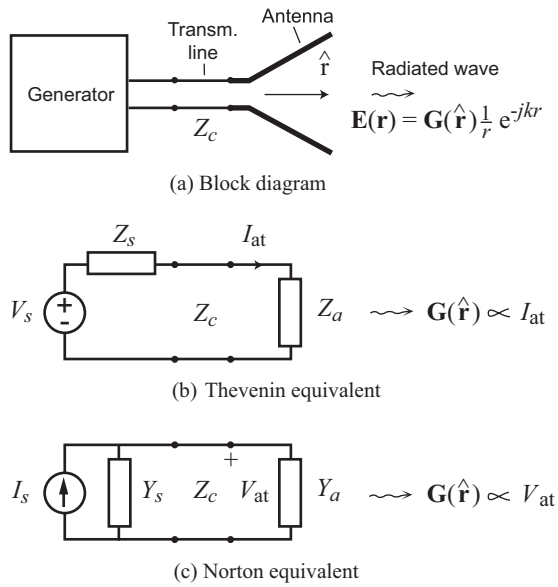


Figure 2.20: Block diagram, Thevenin and Norton equivalent circuits of a transmitting antenna.

2.6 Equivalent circuits of single-port antennas

In general, an antenna will be connected to a transmitter or receiver system, often via a transmission line such as a coaxial cable or rectangular waveguide. In order to analyze the antenna in this network, we need a circuit representation of it, both in its receiving and transmitting modes [3].

2.6.1 Transmitting antennas

The *equivalent circuit* of a *transmitting* antenna with signal generator and transmission line is shown in Fig. 2.20. The transmitter is modeled by its *Thevenin* (or *Norton equivalent*), and the antenna is modeled by a complex radiation impedance $Z_a = R_a + jX_a$ (or radiation admittance $Y_a = G_a + jB_a$). The *radiation impedance* (*admittance*) is also called the *antenna impedance* (admittance) and *input impedance* (admittance) of the antenna. The radiation impedance is normally used when the expression for the far-field function $\mathbf{G}(\hat{\mathbf{r}})$ is proportional to the excitation current $I_{\rm at}$ at the antenna port, and the radiation admittance is used when $\mathbf{G}(\hat{\mathbf{r}})$ is proportional to the excitation voltage $V_{\rm at}$ at the antenna port. The power dissipated in the resistive (or conductive) part R_a (or G_a) of Z_a (or Y_a), is equal to the sum of the total radiated power and the power dissipated as actual heat in the antenna structure due to losses. X_a (or B_a) is related to the reactive power, i.e., the power oscillating back and forth inside the antenna and its close neighborhood. The *reflection coefficient* at the input port is

$$r = (Z_a - Z_c)/(Z_a + Z_c)\ , \tag{2.126}$$

where Z_c is the characteristic impedance of the transmission line. This can also be referred to as the *port impedance* of the antenna. The reflection coefficient is normally presented

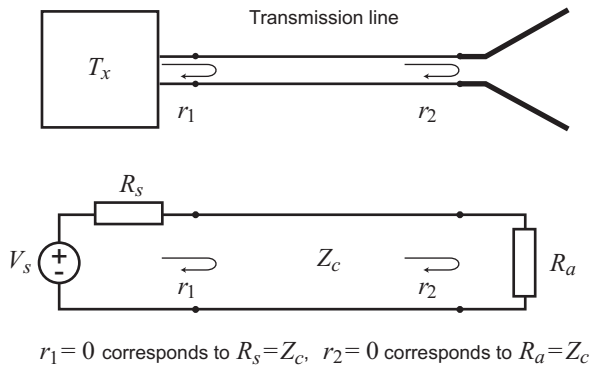


Figure 2.21: Matching of an antenna and a generator (or receiver) to the transmission line in between them.

as $20\log(|r|)$ in dB. The inverse of the reflection coefficient is called the *return loss* being $-20\log(|r|)$ in dB.

2.6.2 Impedance matching to transmission line

Most often we wish to design an antenna with small reflection coefficient, in order to avoid standing waves on the transmission line (Fig. 2.21). Such standing waves increase the ohmic losses on the transmission line because the wave propagates the same distance several times. Also, reflections on the line reduce the bandwidth. The signal generator may also get overheated as a result of mismatch because the power dissipated in the source resistance R_s will increase. Ideally we want $r = 0$, which we refer to as the antenna being impedance matched to the characteristic impedance of the transmission line. This appears when

$$R_a = Z_c \qquad \text{and} \qquad X_a = 0 \ , \tag{2.127}$$

or equivalently, when

$$G_a = 1/Z_c \qquad \text{and} \qquad B_a = 0 \ . \tag{2.128}$$

2.6.3 Receiving antenna

The *equivalent circuit* of an antenna in the *receiving* mode is shown in Fig. 2.22. The antenna is the same as the one in Fig. 2.20 and is modeled either by its *Thevenin* (or *Norton*) *equivalent circuit*. The source impedance Z_a (admittance Y_a) is equal to the radiation impedance (admittance) for the transmitting mode, and the equivalent source voltage $V_{\rm ar}$ (current $I_{\rm ar}$) is proportional to the amplitude of the incoming wave.

We use the Thevenin equivalent when the expression for the far-field function $\mathbf{G}(\hat{\mathbf{r}})$ is proportional to the excitation current $I_{\rm at}$ in the transmitting mode. Then, the induced voltage in the equivalent circuit for the receive case is

$$V_{\rm ar} = -\frac{2j\lambda}{\eta I_{\rm at}} \mathbf{G}(\hat{\mathbf{r}}) \cdot \mathbf{E}_0 \ , \tag{2.129}$$

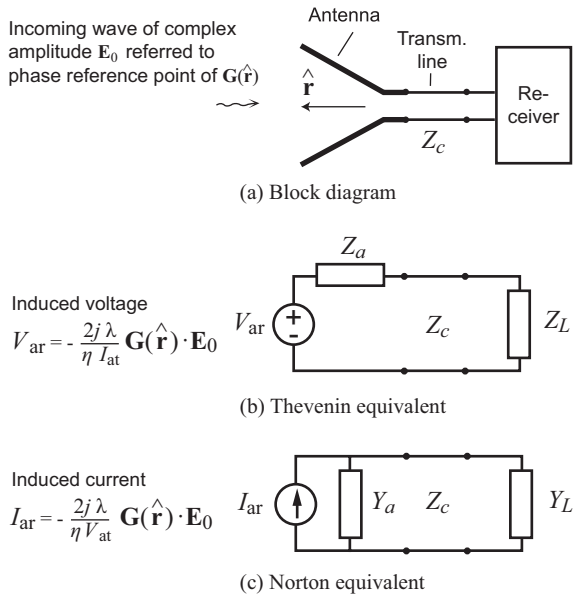


Figure 2.22: Block diagram, Thevenin and Norton equivalents of the antenna in Fig. 2.20 when it is used as receiving antenna.

where $\eta = 377\,\Omega \approx 120\pi\,\Omega$ is the free space wave impedance. $\mathbf{E}_0$ is the electric field vector of an incident plane wave propagating in the direction $-\hat{\mathbf{r}}$. In fact, $\mathbf{E}_0$ must be the complex vector of this wave at the location of the phase reference point of $\mathbf{G}(\hat{\mathbf{r}})$.

We use the Norton equivalent when the expression for the far-field function $\mathbf{G}(\hat{\mathbf{r}})$ is proportional to the excitation voltage $V_{\rm at}$ in the transmitting mode. The induced current in the equivalent circuit for the receive case is

$$I_{\rm ar} = -\frac{2j\lambda}{\eta V_{\rm at}} \mathbf{G}(\hat{\mathbf{r}}) \cdot \mathbf{E}_0 \ , \tag{2.130}$$

where η and $\mathbf{E}_0$ is the same as in (2.129).

We have here not derived the expressions in (2.129) and (2.130). This can be done by assuming that $\mathbf{E}_0$ is produced by an incremental dipole in infinity and using the reciprocity relation between the reactions between this dipole and the excitations of the antenna[23].

2.6.4 Conjugate impedance matching

Absorbed power in the load $Z_L = R_L + jX_L$ or $Y_L = G_L + jB_L$ is referred to as received power. There will also be power dissipated in the radiation resistance of the receiving antenna. This power corresponds to the power that is radiated by the currents that the incident wave induces on the antenna. The antenna is also in this case matched to the transmission line if (2.127) or (2.128) is satisfied. We normally also require that the receiver load satisfies

[23] See Section 4.5.

$R_{_L} = Z_c$ and $X_{_L} = 0$ to match the load to the transmission line. If so, the powers dissipated in R_a and $R_{_L}$ are equal.

If the transmission line is short, or if the receiver is connected directly to the antenna port, we may use conjugate matching of the load to the antenna. Then, we choose the load impedance in such a way that

$$Z_{_{Lt}} = Z_a^* \ , \quad \text{i.e.,} \quad R_{_{Lt}} = R_a \ , \quad \text{and} \quad X_{_{Lt}} = -X_a \ , \tag{2.131}$$

where $Z_{_{Lt}} = R_{_{Lt}} + jX_{_{Lt}}$ is the load impedance transformed through the transmission line to the antenna port. This load maximizes the power transferred to the load for a given Z_a and a given power density of the incident wave. The principle of *conjugate matching* is known from circuit theory.

2.6.5 Impedance and reflection coefficient transformations

When analyzing an equivalent circuit such as that in Fig. 2.21, we need to know how to transform reflection coefficients and impedances along transmission lines. These formulas can be found in the textbooks on transmission line theory, but we will include them here for the sake of completeness. The reflection coefficient $r_{_2}$ in Fig. 2.21 is expressed in terms of the antenna impedance Z_a and the characteristic impedance Z_c of the transmission line as

$$r_{_2} = \frac{Z_a - Z_c}{Z_a + Z_c} \ . \tag{2.132}$$

Alternatively, if we know $r_{_2}$, Z_a can be found from

$$Z_a = Z_c \left(\frac{1 + r_{_2}}{1 - r_{_2}} \right) \ . \tag{2.133}$$

The *reflection coefficient* $r_{_2}$ when measured at the input port 1 of the *transmission line* is

$$r_{_2}(l) = r_{_2} e^{-j2\beta l} \ , \tag{2.134}$$

where β is the propagation constant on the transmission line and l is its length. The *impedance* seen at this point is

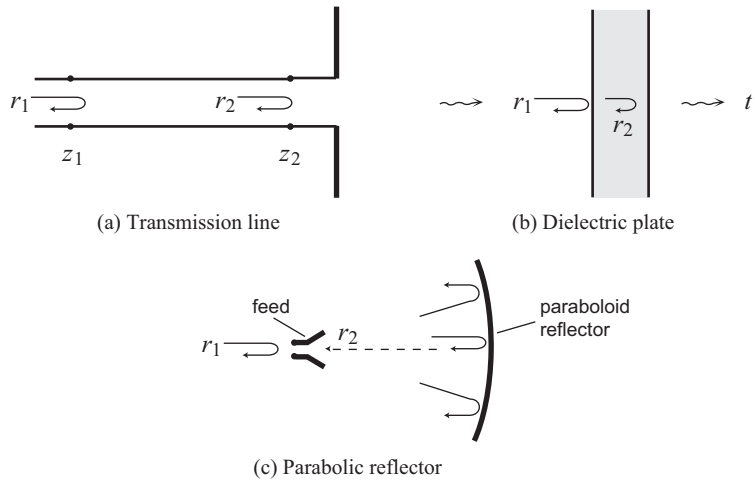
$$Z_a(l) = Z_c \left(\frac{1 + r_{_2}(l)}{1 - r_{_2}(l)} \right) \ , \tag{2.135}$$

This can also be expressed as

$$Z_a(l) = \left(\frac{Z_a + jZ_c \tan \beta l}{Z_c + jZ_a \tan \beta l} \right) Z_c \ . \tag{2.136}$$

The corresponding formulas for the admittances are

$$Y_a(l) = Y_c \left(\frac{1 + r_a(l)}{1 - r_a(l)} \right) = \left(\frac{Y_a + jY_c \tan \beta l}{Y_c + jY_a \tan \beta l} \right) Y_c \ . \tag{2.137}$$



(a) Transmission line (b) Dielectric plate

(c) Parabolic reflector

Figure 2.23: Examples of small reflections r_1 and r_2 on a transmission line, for a plane wave incident on a dielectric plate, and at the terminal of a feed for a parabolic reflector.

2.7 Periodic reflection coefficients $r(f)$

An important part of the process of designing an antenna is to match its impedance to the transmission line connected to its port. This ensures that the power of the signal entering from the transmission line is accepted by the antenna and not reflected back. The impedance matching may be done numerically, but very often an additional experimental tuning in a measurement setup is needed. In both cases it will be helpful to know how different reflections interfere. This can help in sorting out the location of a reflection. The interference between reflections can be treated approximately in a simple manner which is well suited for estimations by using the theory of small reflections, as described below.

Consider the transmission of a wave inside a transmission line or in free space. At two positions z_1 and z_2 along the transmission line there are some small reflections r_1 and r_2, respectively (Fig. 2.23). This may be for instance a wave inside a coaxial line which has a non-ideal connector at one end, z_1. At the other end, z_2, it is connected to the input port of a monopole antenna. It can also be a wave between the port of a small feed horn at z_1 and a parabolic reflector at z_2, in which part of the reflected wave from the paraboloid is received by the feed horn giving rise to the reflection r_2. It may also be a plane wave incident normally on a dielectric sheet, in which case the reflection r_1 appears from the front side of the sheet and $r_2 = -r_1$ from the rear side.

For an accurate analysis of the interference between two reflections we need to use the impedance transformation formulas in Section 2.6.5. When the reflections are small, typically $|r_1| < 0.2$ and $|r_2| < 0.2$, we can find the total *reflection* r by using the approximate formula

$$|r| = \left| r_1 + r_2 e^{-j2\beta(z_2 - z_1)} \right| , \tag{2.138}$$

where β is the propagation constant on the transmission line between z_2 and z_1. This formula is very convenient for quick calculations by hand, and it is easily interpreted. The absolute

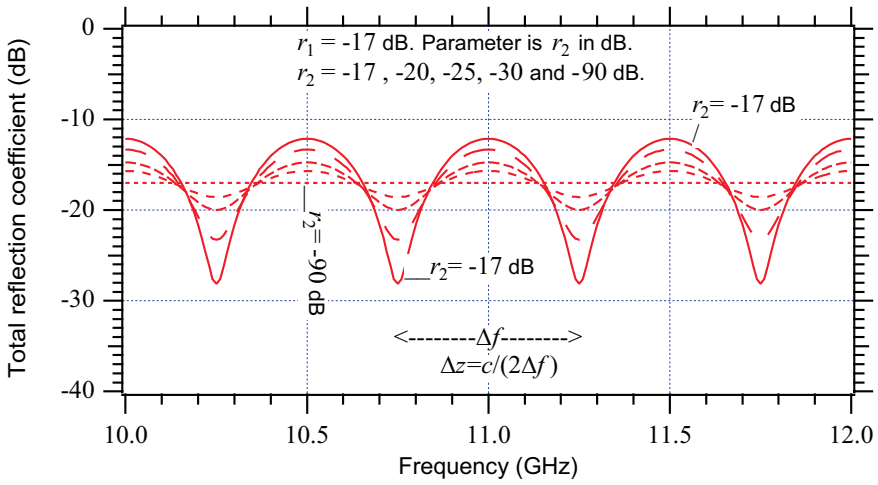


Figure 2.24: *Example of interference between two reflections separated by $\Delta z = z_2 - z_1 = 30\,\text{cm}$ when $|r_1|$ is $-17\,\text{dB}$ for different levels of $|r_2|$.

value of r is a periodic function which varies with frequency, because

$$\beta = 2\pi/\lambda = 2\pi f/c \ ,$$

with c the phase velocity of the wave on the transmission line between z_1 and z_2. Note that, for simplicity we consider a nondispersive transmission line with β proportional to f.

The frequency difference $\Delta f = f_2 - f_1$ between two interference maxima (or minima) of $|r|$ in (2.138) is found by letting the phase term change by 2π, i.e.,

$$2\beta_2(z_2 - z_1) - 2\beta_1(z_2 - z_1) = 2\pi \ , \qquad \beta_2 - \beta_1 = \frac{\pi}{(z_2 - z_1)} \ ,$$

which gives

$$\Delta f = f_2 - f_1 = \frac{c}{2(z_2 - z_1)} \ .$$

The formula for Δf is the same as that of calculating the frequency when the wavelength is $2(z_2 - z_1)$. The maximum value of $|r|$ is $|r_1| + |r_2|$, and the minimum is $|r_1| - |r_2|$. The VSWR is

$$\text{VSWR} = \frac{1 + |r|}{1 - |r|} \approx 1 + 2|r| \ ,$$

where the approximation is valid for $|r| < 0.1$.

Fig. 2.24 shows $|r|$ in dB when it is calculated by the exact transmission line formulas for $r_1 = 0.2$ (i.e., $-17\,\text{dB}$) and different r_2 with $z_2 - z_1 = 10\,\text{cm}$. The agreement with the above approximate formulas for small reflections is very good.*[24]

[24] A MATLAB code exists, for calculation by both the exact and the approximate formulas.

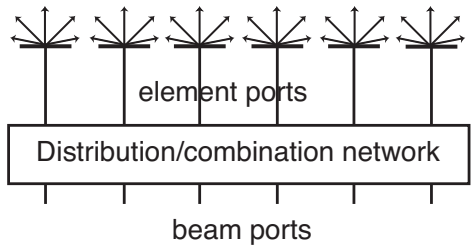


Figure 2.25: Illustration of multi-port array antenna with element ports and beam ports.

2.8 Equivalent circuits of multi-port array antennas

Array antennas have more than one radiating element. Then, the *element ports* may be combined to one or more *beam ports* via a power divider network for the transmitting case or a power combination network for the receiving case[25]. The power combiner/divider networks can be used to generate multiple beams excited via the beam ports. If the element ports are combined to only one port, the antenna may be characterized on this port as a single-port antenna.

In a multi-port antenna there will generally be mutual coupling between the ports, meaning that if we excite one port with a voltage we will be able to measure an induced voltage at the other ports. This mutual coupling can be characterized by mutual impedances Z, mutual admittances, Y, or so-called scattering parameters, S. The scattering parameters are defined as the ratio between the incoming and reflected waves in a transmission line connected to the ports, and they are commonly referred to as S-parameters.

The equivalent circuits and definitions of Z- and S-parameters of a two port antenna are given in Appendix F.

2.9 Further reading

Several system aspects of antennas are described in various chapters in the classic radar book by Skolnik [4]. See also the reference list to Chapter 1.

2.10 Complementary comments by S. Maci

The far-field definition is usually adopted for antennas with $D > 2\lambda$. For smaller antennas the distinction between reactive near-field distance and far-field distance looses meaning. It is indeed more appropriate defining the far-field distance by the factorization of the field in (2.39), namely when the field can be seen as the product between e^{-jkr}/r and a function depending only on θ and φ. For small antennas the far field distance can be defined by looking at where the quadrupole moment contribution becomes negligible [5], or equivalently as discussed in Section 4.2.1 of this book, when the higher order terms of the $1/kr$ expansion

[25] This was already explained on page 3 and shown in Fig. 2.25

of the field from the incremental current sources become negligible. In [6] it is suggested to use the practical rules $R_{\text{far}} = \lambda$ for $D < \lambda/4$, i.e., small antennas, which agrees with the approximations in Section 4.2.1.

The far-field function $\mathbf{G}(\hat{\mathbf{r}})$ defined in (2.39) in this book may have other names in other textbooks, and many authors prefers to normalizing it like $\mathbf{f}(\hat{\mathbf{r}}) = \mathbf{G}(\hat{\mathbf{r}})/|\mathbf{G}(\hat{\mathbf{r}})|_{\max}$. Stutzman and Thiele [7] call it a *normalized far-field pattern*, and Franceschetti calls it a *radiation vector* [8]. The power integral P defined in (2.65), normalized with respect to the maximum vector function value $|\mathbf{G}(\hat{\mathbf{r}})|^2_{\max}$ is sometimes called the *beam solid angle*; i.e., $\Omega = P/|\mathbf{G}(\hat{\mathbf{r}})|^2_{\max}$ (e.g., [5, page 35] or [9, page 45]). This quantity is interpreted as the solid angle of an equivalent antenna having uniform radiation intensity $|\mathbf{G}(\hat{\mathbf{r}})|^2_{\max}$ in any direction inside Ω and zero elsewhere, and radiating the total power of the true antenna. The beam solid angle is related to the maximum directivity by $D_0 = 4\pi/\Omega$. For antennas with directivity larger than 20 dB, the beam solid angle is approximately equal to the product of the 3 dB beam angle $\theta^{(E)}_{3\,\text{dB}}$ and $\theta^{(H)}_{3\,\text{dB}}$ in the two principal planes, when these are expressed in radians; i.e.,

$$\Omega \approx \theta^{(E)}_{3\,\text{dB}} \theta^{(H)}_{3\,\text{dB}} \ .$$

This gives the relation

$$\theta^{(E)}_{3\,\text{dB}} \theta^{(H)}_{3\,\text{dB}} \approx 4\pi/D_0 \ ,$$

between the 3 dB beam angles and directivity D_0 [9, page 46]. This expression is less accurate than the ones derived in Sections 2.4.4-2.4.6, but its simplicity allows for a fast estimate of the 3 dB angles in directive antennas.

Whenever it is possible to define a current I_{at} at the antenna terminals, some authors, (i.e., [10]) define an *effective vector height* $\mathbf{h}_t(\hat{\mathbf{r}})$ for the *transmitting mode* by

$$\mathbf{E}(\hat{\mathbf{r}}) = jk\eta \frac{e^{-jkr}}{4\pi r} I_{\text{at}} \mathbf{h}_t(\hat{\mathbf{r}}) \ .$$

This definition originates from the far-field radiated by an elementary dipole. In fact, for an elementary dipole $\mathbf{h}_t(\hat{\mathbf{r}}) = \Delta z \hat{\boldsymbol{\theta}} \sin\theta$. According to the notation followed in the present book, we have $\mathbf{h}_t(\hat{\mathbf{r}}) = \frac{4\pi}{I_0 jk\eta} \mathbf{G}_t(\hat{\mathbf{r}})$. In the receiving mode, it is possible to define an *effective vector height for the receiving mode*, by the relationship $\mathbf{h}_r(\hat{\mathbf{r}}) \cdot \mathbf{E}_0 = V_{\text{ar}}$, where V_{ar} is the voltage of the Thevenin equivalent circuit in Fig. 2.22. Following this definition, the antenna reciprocity can be expressed as $\mathbf{h}_t(\hat{\mathbf{r}}) = \mathbf{h}_r(\hat{\mathbf{r}})$, namely by requiring that the effective vector height for the transmitting mode is identical to the one for receiving mode.

2.11 Exercises

1. **Polarization and dB:**

   a) The axial ratio of a circularly polarized wave is 1 dB. Find the relative cross-polar level and the polarization efficiency.

   b) A desired linearly polarized wave has an axial ratio of 20 dB. Find the relative cross-polar level and the polarization efficiency.

   c) Consider a desired linear x-polarization. The wave is elliptically polarized with 20 dB axial ratio. However, the major axis of the polarization ellipse makes an angle of 20° with the x-axis. What is now the relative cross-polar level and the polarization efficiency? How much do the cross-polar level and the polarization efficiency improve if we tilt the desired polarization in a way that it becomes aligned with the major axis of the polarization ellipse?

d) Consider an elliptically polarized wave with 3 dB axial ratio. Find the relative cross-polar level and the polarization efficiency when we consider the desired polarization to be the best linear. Compare the values with those we obtain when the desired polarization is the best circular.

2. **Polarization:** For a circular polarization excitation, make a MATLAB program which calculates the relative cross-polar level caused by simultaneous amplitude and phase errors. Present the results in a contour plot with amplitude and phase as variables. Start with the existing MATLAB document for Fig. 2.3.
3. **Phase center:** Derive Eq. (2.63) for the phase center location of a symmetrical radiation pattern.
4. **Phase center:** The table below shows a measured phase pattern.

$\boldsymbol{\theta}$	$0°$	$\pm 10°$	$\pm 20°$	$\pm 30°$	$\pm 40°$
Phase	$95°$	$105°$	$140°$	$-175°$	$-90°$

a) Calculate the phase center location within the $\pm 20°$ sector. Calculate the phases for all angles when the phase reference point is moved to this phase center.

b) Calculate the phase center location within the $\pm 30°$ sector. Calculate the phases for all angles when the phase reference point is moved to this phase center.

5. **Polarizer and noise temperature:** Consider the example in Subsection 2.2.6 on the mismatch in the polarizer. Assume that port 2 is terminated by $50\,\Omega$ and that the impedances on ports 3 and 4 are $75\,\Omega$.

a) How large is the contribution to the system noise temperature due to the mismatch on ports 3 and 4 when the antenna noise temperature is 40 K?

b) Assume that we locate the termination on port 2 inside a cold box with temperature 20 K. How large is now the contribution to the system noise temperature due to the mismatch on ports 3 and 4? What is the axial ratio of the polarization ellipse due to the same mismatch?

c) Assume that port 2 is terminated by a $75\,\Omega$ load in room temperature. What is now the contribution to the system noise temperature and what is the axial ratio of the polarization ellipse?

6. **Directivity and phase reference point:** Consider a hypothetical omnidirectional antenna with constant phase, i.e.,

$$\mathbf{G}(\theta, \varphi) = \text{const.}(\sin\varphi\hat{\boldsymbol{\theta}} + \cos\varphi\hat{\boldsymbol{\varphi}}) \ .$$

a) What is the polarization of this antenna along the z-axis?

b) Find the phase patterns in the E- and H-planes if the phase reference point is moved to $z = 0.5\lambda$. Sketch it.

c) Find the phase patterns in the E- and H-planes if the phase reference point is moved to $x = 0.5\lambda$. Sketch it.

d) Find the directivity of the omnidirectional source. Does it depend on the location of the phase reference points.

e) Find the cross-polarization in the above three cases.

f) Find the directivity when the antenna only radiates into the upper hemisphere, i.e.,

$$\mathbf{G}(\theta, \varphi) = 0 \qquad \text{for } \pi/2 < \theta < \pi \ .$$

7. **Fraunhofer approximation:** Derive the condition for which the Fraunhofer approximation is valid in Section 2.3.4 by assuming that $\mathbf{r}_0 = x_0\hat{\mathbf{x}}$ and $\mathbf{r} = r\hat{\mathbf{z}}$, and expanding $r' = |\mathbf{r} - \mathbf{r}_0|$ in a power series for small x_0. What is the maximum phase error due to the approximation when $r' = x_0^2/\lambda$ and $r' = 2x_0^2/\lambda$? The largest phase error appears always in the direction where $\hat{\mathbf{r}} \perp \mathbf{r}_0$.

8. **BOR$_1$ antenna:** Consider a BOR$_1$ antenna with a 3 dB half-beamwidth of 15° in both E- and H-planes.

   a) Find the directivity from the approximate $\cos^n(\theta/2)$ formulas.

   The mechanical design of the antenna is modified and cause the following changes of the radiation pattern in E-plane: the 3 dB half-beamwidth increases to 17° and sidelobes which are -25 dB less than the main beam maximum, appear between $\theta = 60°$ and $\theta = 110°$. The pattern does not change in H-plane. There is no change in the phase patterns.

   b) Find the new directivity by using the formulas in the examples in Subsection 2.4.4 and 2.4.5.

   c) Find the largest value of the relative cross-polar level at $\theta = 15°$. In which φ-plane does it appear?

9. **BOR$_1$ antenna:** We excite the unmodified BOR$_1$ antenna in exercise 8 for circular polarization with an axial ratio of 1 dB on axis.

   a) Find the polarization efficiency in dB.

   b) What is the largest value of the relative cross-polar level within the main beam?

10. **BOR$_1$ antenna:** If there are phase differences in the E- and H-planes, they can be included in the approximate theoretical far-field function in the following way,

$$\mathbf{G}(\theta,\varphi) = \cos^n(\theta/2)(e^{-jk\Delta(\cos\theta-1)}\sin\varphi\hat{\boldsymbol{\theta}} + e^{jk\Delta(\cos\theta-1)}\cos\varphi\hat{\boldsymbol{\varphi}}) \ ,$$

    where Δ is adjusted to account for the actual measured phase difference at the θ-value for which the beam pattern is 10 dB down.

    a) Derive a formula to calculate Δ from this phase difference. Where are the phase center locations of the E- and H-plane patterns?

    b) Derive the co- and cross-polar radiation patterns in the 45°-plane. Simplify the formula for the cross-polar field by expanding the resulting equation for small Δ. Sketch the patterns. What is the level of the cross-polar sidelobe when $\Delta = \lambda/4$ and the 10 dB widths of the E- and H-plane patterns are 60°?

    c) Assume that we excite the antenna for circular polarization. Where is the phase center of the co-polar pattern? What are now the cross-polar patterns in the 0°- and 90°-planes? Sketch them.

11. **Conjugate matching:** Derive the conjugate matching condition in (2.131) by using the equivalent circuit in Fig. 2.22b and by maximizing the power delivered to the load for a given radiation impedance Z_a.

2.12 References

[1] D.K. Cheng, *Field and Wave Electromagnetics*, Addison-Wesley, 1989.

[2] A.C. Ludwig, "The definition of cross-polarization", *IEEE Transactions on Antennas and Propagation*, Vol. AP-21, pp. 116-119, January 1973.

[3] P.-S. Kildal, "Equivalent circuits of receive antennas in signal processing arrays", *Microwave and Optical Technology Letters*, Vol. 21, No 4, pp. 244-246, May 20 1999.

[4] M.I. Skolnik, *Introduction to Radar System*, McGraw-Hill, 1962.

[5] J.A Stratton, *Electromagnetic Theory*, IEEE Press, Piscataway, NJ 08854.

[6] S. Maci, "Notes for the antenna course" , Universty of Siena internal notes for students of the Antenna course, in Italian.

[7] W.L. Stutzman and G.A. Thiele, *Antenna Theory and Design*, 2nd edition, Wiley, New York, 1997.

[8] G. Franceschetti, *Electromagnetics: Theory, Techniques, and Engineering Paradigms*, Plenum Press, New York, 1997.

[9] C.A. Balanis, *Antenna Theory and Design*, 2nd edition, Wiley, New York, 1996.

[10] G. Franceschetti, *Campi Elettromagnetici*, Bollati Boringhieri (in Italian), 1983.

[11] J. Ruze, "Antenna tolerance theory - A review", *Proceedings of the IEEE*, Vol. 54, pp. 633-640, April 1966.

Chapter 3

Characterization in multipath

Mobile wireless terminals such as handset or phones are subject to strong fading due to multipath propagation. This is true in particular when they are used in urban and indoor environments. Then, the performance can be significantly improved by making use of antenna diversity. This means that the signals on two antennas (with different positions, polarizations or radiation patterns) are combined in a way that there are shallower fading minima in the combined signal than either of the contributing signals. Thereby, the *Signal-to-Noise Ratio* (SNR) in the fading dips will be better, so that the fading margins in the system link budget can be reduced. The increased SNR will also allow the capacity of the communication channel to be increased by using a higher order modulation, if the system allows this. In addition, if we have several antennas on both the transmitting and the receiving sides, we can establish several communication channels (referred to as *bitstreams*) through the multipath environment. This allows for even larger capacity increase. Such systems are often referred to as *Multiple Input Multiple Output* (*MIMO*) systems. The present chapter describes how to characterize both single-port and multi-port antennas for use in multipath and in particular for achieving diversity and MIMO performance.

The so-called *reverberation chamber* [1] has since the nineteen-eighties been used for EMC testing of radiated emissions and immunity. It is a metal cavity that is sufficiently large to support many resonant modes at the frequency of operation. It contains some proper tools to stir the modes so that statistical field variations appear. It has been shown that the reverberation chamber represents a multipath environment of a similar type that we find in urban and indoor environments. And, since year 2000, its application has been extended to characterize antennas and terminals designed for use in multipath environments. Therefore, this chapter also describe how reverberation chambers work and how they can be used to measure performance of antennas and active terminals, in particular when used with diversity and MIMO capability.

3.1 Multipath without Line-of-Sight (LOS)

Traditionally, antennas were designed for use in environments where there is a *Line-of-Sight* (*LOS*) between the two sides of the communication link. Therefore, antennas were charac-

terized equivalently, by measurements in anechoic chambers. Still, there may be additional wave paths between the transmitting and receiving sides. For instance, large smooth objects located outside the LOS cause reflection, edges of large objects cause diffraction, and small or irregular objects cause scattering of the waves originating from the transmitting antenna. The wave contributions via these paths will add at the receiving side. They have independent complex amplitudes (i.e., amplitudes and phases), so that they may add up constructively or destructively, or something between these two extremes. In addition, the wave paths and their complex amplitudes change with time, due to the moving of the terminal or objects of the environment. Even small changes of half wavelength between two paths may cause the received level to change from constructive to destructive interferences. Such level variations are therefore called small-scale fading [2, Sections 2 & 4], or more commonly *fast fading*. There may also be large-scale fading, more commonly called *slow fading*, due to level variations of a single dominant wave contribution. This can be caused by variation of the distance to the transmitting antenna, variation of the far-field function of the receiving antenna due to rotation of the receiving antenna, or shadowing due to blocking objects such as e.g. the user's body.

3.1.1 Rayleigh fading and CDF

At the receiving side the multipath environment can be characterized by several independent incoming plane waves. The independency means that their amplitudes, phases and polarizations as well as the *Angles of Arrival* (*AoA*) are statistically arbitrary relative to each other. If the LOS is absent, and if the number of incoming waves is large enough (typically a few hundred), or if we move the antenna around in a less rich environment, the in-phase and quadrature components of the received complex signal become normally distributed, i.e., a *complex Gaussian* distribution. This corresponds to a *Rayleigh* distribution of the signal magnitudes, and an exponential distribution of the power. The phase is uniformly distributed over 2π. The complex Gaussian distribution is a direct result of the central limit theorem.

The left graph in Fig. 3.1[1] shows an example of a Rayleigh fading signal. We see that the signal varies by more than 25 dB. The Rayleigh fading is most conveniently illustrated by the Cumulative probability Distribution Function (CDF), showing the cumulative probability of signal amplitudes in dB, which is shown in the right graph in Fig. 3.1. In this figure the reference level for the dB values is obtained as the sum of the received powers of all the samples divided by the number of samples, i.e., the average received power. We see that there is a probability of 0.1 % of having deeper dips than −30 dB, and of 1 % of having deeper dips than −20 dB, and of 10 % of having deeper dips than −10 dB. These values are descriptive for the Rayleigh distribution and very easy to remember.

3.1.2 Angle of Arrival (AoA), XPD and polarization imbalance

The arriving waves may have a certain AoA distribution in the elevation and azimuth planes. It is natural to assume that the mobile terminal can be oriented arbitrarily relative to directions in the horizontal plane, which means that the azimuth angle is uniformly distributed.

[1] There exist MATLAB code for all figures of which the caption start with ⋆.

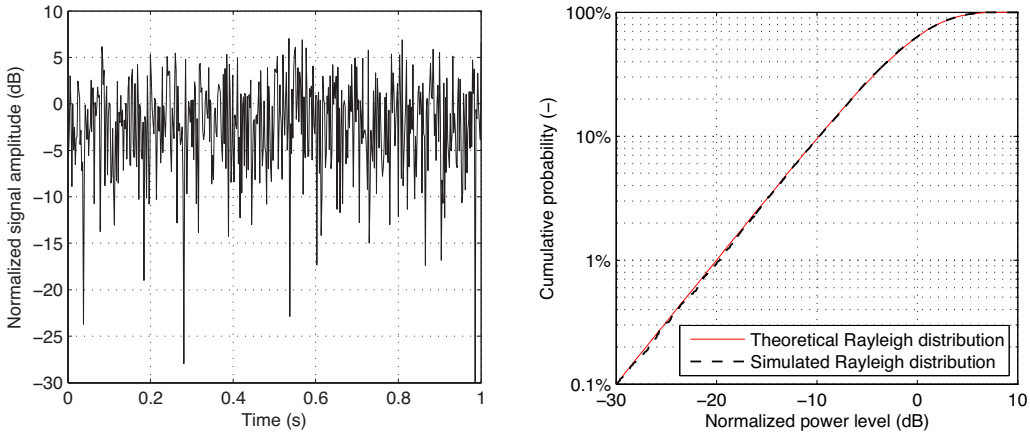


Figure 3.1: *Example of a fading signal (left) and its CDF (right). The signal levels are presented in dB after being normalized to the time averaged power.

The terminals may under normal use have a certain preferred (or most probable) orientation relative to the vertical axis, and common environments (in particular outdoor) have larger probability of waves coming in from close-to-horizontal directions than close-to-vertical. Therefore, we may need an elevation distribution function to describe real multipath environments. Real environments normally also have a larger content of vertical polarization than horizontal, because most base stations are vertically polarized. This is in propagation literature characterized by a Cross-polar Power Discrimination (XPD) [3]. We here instead refer to this as a polarization imbalance in the environment, as it represents a deterioration of the isotropic reference environment defined later. Both the AoA distribution and XPD are different in different real-life environments. This is problematic because the performance of the antennas and wireless terminals then depend on where they are used or measured. So the results of measurements in one environment cannot directly be transferred to another. Therefore, it is convenient to define a reference environment, which can easily be reproduced in practice, giving repeatable results for the performance. The rich isotropic multipath environment has this characteristic, being a result of polarization balance and a uniform AoA distribution over the whole sphere.

3.1.3 Rich Isotropic Multipath (RIMP)

As explained above it is desirable to have a rich isotropic reference environment with polarization balance and a uniform distribution of AoA in both azimuth and elevation. The latter means an environment in which all AoAs over the whole unit sphere are equally probable. This simplifies the characterization of antennas and terminals in the sense that the performance becomes independent of the orientation of the antenna in the environment. The reverberation chamber emulates such a rich isotropic multipath (RIMP) environment, provided it is large enough. This reference environment has no true counterpart in reality, but it is still quite representative as any environment appears isotropic if the terminal is used with arbitrary orientations in the environment. It is here important to point out that normally an antenna is not located symmetrically on the phone. Therefore, the phone has different orien-

tations in the environment when used in talk positions on the left and right sides of the head. This is because opposite sidewalls of the phone point upwards for these two talk positions. A *rich* environment means that typically at least 20 waves are present instantaneously.

The RIMP environment will produce *independent and identically distributed* (*i.i.d.*) channels if the multi-port antennas being considered have uncoupled ports with the same total radiation efficiency. The i.i.d. case is a well-established reference case in theoretical works on communication systems and propagation.

3.2 Characterization of single-port antennas in RIMP

3.2.1 Antenna impedance, port impedance and reflection coefficient

All antennas are characterized by their far-field function[2] $\mathbf{G}(\theta, \varphi)$ and antenna impedance[3] Z_a. The far-field function can be used to plot the co-and cross-polar radiation patterns and to find traditional pure LOS quantities such as directivity and sidelobe levels. For antennas operating in a RIMP environment the shape of the far-field function does not play any significant role, in particular not for single antenna systems. In multi-port antennas it has some significance as it determines the correlation between the received signals on the ports.

In order to connect to other components like receivers and transmitters the antenna must have a well-defined port, often in the form of a transmission line or a connector for a transmission line. The characteristic impedance Z_c of this transmission line defines the port impedance, normally $50\,\Omega$. When the port impedance is known, the antenna impedance can be transformed to a complex reflection coefficient r on the transmission line, where[4]

$$r = (Z_a - Z_c)/(Z_a + Z_c)\ . \tag{3.1}$$

In an S-parameter representation of the antenna, $r = S_{11}$.

3.2.2 Mean Effective Gain (MEG) and Directivity (MED)

Antennas in fading environments are sometimes characterized by the so-called Mean Effective Gain (MEG), the formula of which is given in [3] and [4]. This can be calculated from the far-field function of the antenna, and it is a function of the orientation of the antenna, its polarization, and the azimuth and elevation distributions of the AoA in the environment, as well as the XPD of the environment. For the RIMP environment the MEG becomes equal to half the classical total radiation efficiency e_{rad}, i.e.,

$$\mathsf{MEG} = e_{\mathrm{rad}}/2\ . \tag{3.2}$$

Generally, the MEG of an antenna in an arbitrary environment can be decomposed in two factors: the classical total radiation efficiency and the mean effective directivity (MED), i.e.,

[2] Definition of the far-field function is on page 35.
[3] Antenna impedance is defined on page 65.
[4] For impedance and reflection coefficient transformations, see page 68.

$$\mathsf{MEG} = \mathsf{MED} \cdot e_{\mathrm{rad}} \ . \tag{3.3}$$

In this way the MED solely contains the effects of the environment and the shape of the radiation pattern, whereas the radiation efficiency contains the effects of losses and impedance mismatch. The formula for MED is the same as for MEG, but the "realized gain" far-field functions in the formula for MEG must be replaced by the "directive gain" far-field functions in the formula for MED[5]. Therefore, the ratio between MEG and MED becomes the same as between the classical realized and directive gains for LOS systems, i.e., the total radiation efficiency. The total radiation efficiency can be measured in a reverberation chamber. We will not give the formulas for the MEG and MED here, because we have limited our considerations to the RIMP environments as explained before. The MEG is discussed in detail in [40].

3.2.3 Total radiation efficiency and transmission formula

From the previous subsections we can state that for antennas in RIMP environment the performance is determined only by the total radiation efficiency, and not by the shape of the radiation pattern and the realized gain. The total radiation efficiency includes both the radiation efficiency due to absorption (ohmic losses) in the antenna, and the mismatch factor[6].

This means that the average transferred power between two antennas, when both of them are located in RIMP, is proportional to the product of the radiation efficiencies of both of them, whereas in free space it is proportional to the product of their realized gains[7]. The attenuation in multipath environment is also much larger than in free space.

Often, the base station antenna is located above the scattering environment, and the terminal is located inside the multipath environment. Then, the average power transfer function will be proportional to the realized gain of the base station antenna, and to the total radiation efficiency of the terminal.

3.3 Characterization of multi-port antennas in RIMP

In LOS systems we can increase the signal to noise ratio by increasing the directivity of the antennas, because the received signal increases with the gains of both the transmitting and receiving antennas. However, in the previous section we have explained that antennas in RIMP environments do not improve performance by increasing directivity. Still, we can use several antenna elements to form a diversity or MIMO system. However, to achieve a better performance we must process the received signals, in a dynamic manner.

To realize antenna diversity and MIMO systems we need several antenna elements located together in an array, as illustrated in Fig. 3.2. This is not a classical antenna array in the sense that the elements are combined with specific amplitude and phase relations between them. Therefore, we prefer to call the diversity or MIMO antenna a multi-port antenna rather than an array. To be more precise, each port can be characterized as a single antenna,

[5] The directive and realized gains are defined in Section 2.3.9 and 2.5.1, respectively.
[6] The radiation efficiency and mismatch factor were already defined on page 56.
[7] For instance, see Friis transmission equation on page 59.

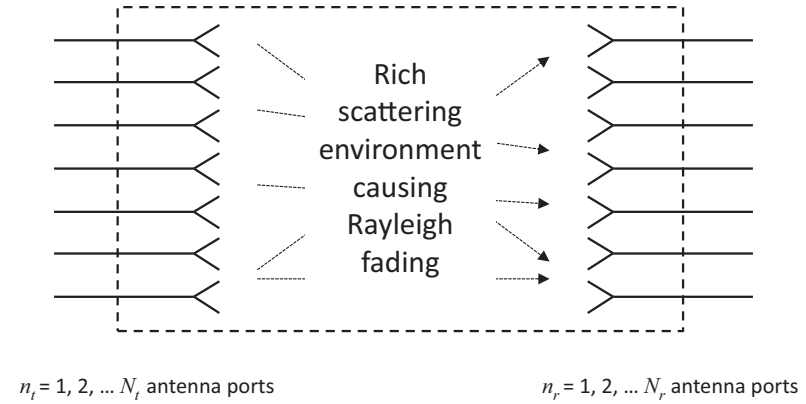


Figure 3.2: Definition of MIMO system with multi-port antennas on both the transmitting and receiving sides. There are in total $N_r \times N_t$ channels through the environment, each one defined between one input port of the transmitting antenna and one output port of the receiving antenna.

and the signal inserted (transmitting case) or being present (receiving case) at each port has no prescribed amplitude and phase relation to the signals on the other ports. Instead, the signals are continuously processed in an adaptive way to optimize system performance. The processing is based on first measuring the channel between the transmitting and receiving antenna ports. This corresponds to a full calibration of the communication link from one port of the transmitting antenna to a port on the receiving antenna, including the environment. It is commonly referred to as channel estimation, and is done continuously to account for all variations of the environment.

3.3.1 Definition of channel

In communication and propagation literature the term channel is used to denote the complex signal amplitude received at a specific port of a receiving antenna relative to the signal inserted at a specific port of the transmitting antenna. Thus, there are $N_r \times N_t$ channels if the transmitting antenna has N_t ports and the receiving antenna has N_r ports. This means that the channels represent the S-parameters between the ports of the transmitting and receiving antennas. Thus, the propagation channel is defined with the antennas included.

Classical single-beam *array* antennas for LOS systems have one port. They are characterized by their far-field functions and antenna impedances in the same way as other single-port antennas. The latter is the impedance seen at the array port when all elements are excited with the amplitude and phase that give the desired shape of the far-field function. The far-field function of a classical array antenna is the product of two factors: the far-field function of a single element and the array factor[8]. There is more about classical arrays in Chapter 10.

In theoretical work on diversity and MIMO systems the element antennas are sometimes

[8] This assumes that the far-field function, of a single embedded element is the same independent of its location in the array, see Section 10.1.1 and 10.3.1.

treated as being isolated. This means that neighboring elements are ignored (removed) when dealing with each element. However, such analysis must be done with caution because this isolated element approach is not strictly valid when analyzing multi-port antennas. It only gives correct results if the ports are completely uncoupled, which they rarely will be in reality, except for two-port antennas with orthogonal polarizations on the two ports. The correct analysis is done by using the embedded element approach as described below[9].

3.3.2 Embedded elements

The embedded element approach has also its background in classical array analysis. However, compared to the *isolated element approach* it is regarded as an advanced topic. Thus, it is rarely covered in introductory antenna courses. The *embedded element approach* considers the far-field of each single array element when all the other elements are present, but they are not excited, rather terminated with their port impedances. The far-fields of such embedded elements are sometimes called active element patterns, a terminology introduced already in [5], but they are now more commonly (and descriptively) referred to as embedded element far-field functions, a term used already in [6].

The excited antenna element induces radiating currents on the terminated non-excited elements. Therefore, the far-field function of the embedded element may be very different from that of the isolated element, in particular if the elements are closely spaced with large mutual coupling. The far-field functions of the embedded elements are used to describe blindness in classical arrays, whereas in MIMO and diversity antennas they play a more significant role.

In a multipath environment the received signals on every port are detected and digitalized independently of each other and thereafter they are combined dependent on the channel estimates. Therefore, each port receives signals through their embedded far-field functions. There may also be multi-port processing on the transmitting side. Then, we need to use embedded element patterns in the transmitting mode analysis as well. In this way, the radiation efficiency at each port as well as the correlation between their signals, are determined by their corresponding embedded far-field functions. Both these quantities (radiation efficiency and correlation) are needed in order to quantify performance of diversity and MIMO antenna systems. Such analysis in terms of far-field functions of embedded elements was introduced in [7] and [8] based on the analysis of antenna diversity in [9].

The far-field functions of embedded elements can be computed by most commercial computer programs for antenna analysis, by exciting one port and terminating all non-excited ports with their port impedances (normally $50\,\Omega$). The embedded element patterns can also be measured in anechoic chambers. However, it is also possible to characterize diversity and MIMO antennas in RIMP without knowing the embedded element patterns explicitly. For example, the reverberation chamber provides a way of measuring radiation efficiency and correlation without going via the embedded element patterns[10].

[9] There is more about embedded and isolated element approaches in Chapter 11.
[10] There is more about embedded and isolated element approaches in Chapter 11.

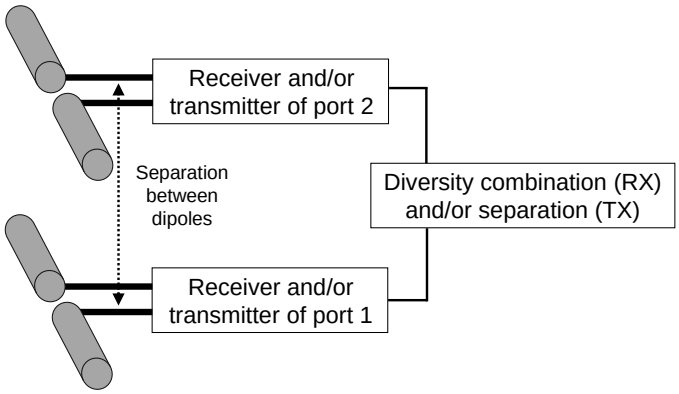


Figure 3.3: llustration of diversity antenna consisting of two parallel dipoles. There are receivers and/or transmitters at each of the two dipole ports.

3.3.3 Embedded radiation efficiency and decoupling efficiency

We will now explain how to calculate the total radiation efficiency of an element in a multi-port antenna. Fig. 3.3 illustrates a diversity antenna consisting of two dipoles connected to their receivers/transmitters. This means that we have assumed that antenna diversity is implemented. This configuration was analyzed in [6] by using a classical semi-analytical approach represented by the equivalent circuit in the upper part of Fig. 3.4. The approach in [6] is generally applicable to any two-port antenna consisting of single-mode radiating elements, such as dipoles, slots and to some extent patches. The most accurate will always be to perform a complete numerical computation using an electromagnetic solver of any kind. Then, there is no restriction to the elements being single-mode. The general form of the equivalent circuit is shown in the lower part of Fig. 3.4. This latter equivalent circuit is too simple to show the induced currents, so it cannot be used to calculate or understand why the far-field function is different from that of an isolated antenna element. This general simplified equivalent circuit is simply an illustration of the result of a numerical calculation of Z_a and $\mathbf{G}(\theta,\varphi)$[11], which is a one-port equivalent circuit. The equivalent circuit will look the same at port 2, but it will generally have different $\mathbf{G}(\theta,\varphi)$ and Z_a.

If the diversity antenna itself consists of materials with very small or no losses, the only contributions to the total radiation efficiency of the embedded element will be the absorption in the termination on port 2, and the reflections on the excited port 1. Using the upper equivalent circuit in Fig. 3.4, the input impedance of the embedded element is

$$Z_a = Z_{11} + \frac{Z_{12}I_2}{I_1} = Z_{11} - \frac{Z_{12}^2}{Z_{22}+Z_c} \ , \tag{3.4}$$

where the latter is obtained directly from the right loop of the equivalent circuit. The reflection coefficient r at port 1 and the corresponding mismatch factor e_{ref_1} become

$$r = \frac{Z_a - Z_c}{Z_a + Z_c} \qquad \text{and} \qquad e_{\mathrm{ref}_1} = 1 - |r|^2 \ . \tag{3.5}$$

[11] For more information see also page 65.

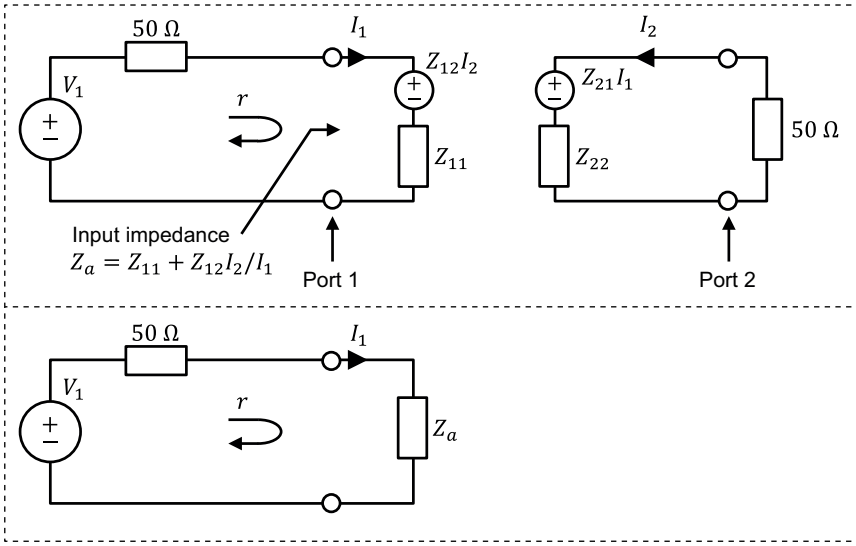


Figure 3.4: Complete equivalent circuit (upper figure) for calculation of the far field function and radiation efficiency of the embedded element 1 of a diversity antenna consisting of two single-mode elements, each with its own port. Port 1 is excited, and port 2 is terminated. Z_{11} and Z_{22} are the impedances of the two isolated element antennas. $Z_{12} = Z_{21}$ is their mutual impedance. $r = S_{11}$ is the complex reflection coefficient on port 1. The lower figure shows the equivalent circuit of the same two-port antenna as in the upper figure, but on the general Thevenin form for one-port antennas. The port impedances are $Z_c = 50\,\Omega$.

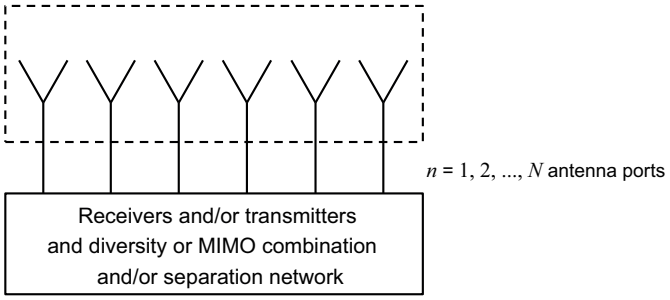


Figure 3.5: Illustration of N-port antenna that can be characterized by its S-parameters between all ports.

The total power that is accepted by port 1 is found as $P_{\rm acc} = \Re\{Z_{\rm in}\}|I_1|^2/2$. The difference between this and the power $P_2 = Z_c|I_2|^2/2$, absorbed in the load on port 2, is the radiated power $P_{\rm rad}$, i.e., $P_{\rm rad} = P_{\rm acc} - P_2$. Finally, the total radiation efficiency at port 1 becomes

$$e_{\rm rad} = e_{\rm ref_1} e_{\rm abs} \qquad \text{with} \qquad e_{\rm abs} = \frac{P_{\rm rad}}{P_{\rm acc}} = 1 - \frac{Z_c|I_2|^2}{\Re\{Z_{\rm in}\}|I_1|^2} \ , \tag{3.6}$$

where $e_{\rm abs}$ is the efficiency due to the power dissipated in the load on port 2. This e_{abs} is therefore the radiation efficiency of port 1. This can be expressed much simpler by using the S-parameters between ports 1 and 2, according to

$$\begin{aligned} e_{\rm rad} &= 1 - |S_{11}|^2 - |S_{21}|^2 \ , \qquad e_{\rm ref_1} = 1 - |S_{11}|^2 \ , \\ e_{\rm abs} &= \frac{1 - |S_{11}|^2 - |S_{21}|^2}{1 - |S_{11}|^2} \ . \end{aligned} \tag{3.7}$$

These radiation efficiencies can also be called *embedded element efficiencies*, reflecting the fact that the port is connected to one element of an array with the other elements present and terminated. The S-parameter expressions can readily be extended to any multi-port antenna, such as that illustrated in Fig. 3.5. The *total embedded radiation efficiency* of element number i when there are a total of N ports becomes

$$e_{\rm rad} = 1 - \sum_{j=1}^{N} |S_{ji}|^2 \ . \tag{3.8}$$

This total embedded radiation efficiency can also be called a decoupling efficiency. There will be an additional efficiency factor if there are losses in the materials that the antenna is made of. Therefore, (3.8) is only valid as a total radiation efficiency if the antenna is lossless.

Eq. (3.8) also defines a fundamental limitation of elements in classical dense arrays [10]. If the elements are very close (typically closer than 0.5 wavelengths), the mutual couplings cause a severe efficiency reduction of each embedded element. This fundamental limitation causes the directivity of a dense classical array to be smaller than the directivity of each isolated element multiplied with the number of the elements[12].

[12] See Section 11.3.2.

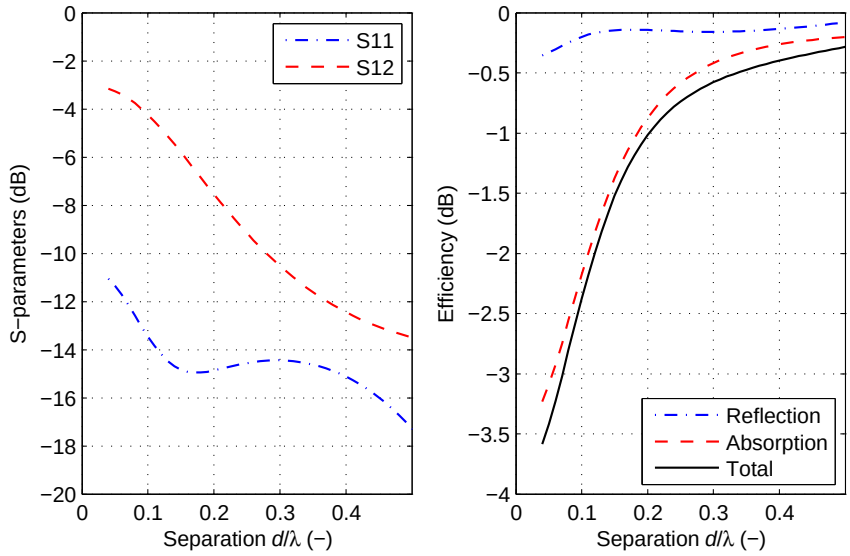


Figure 3.6: *S-parameters of two parallel 0.47 λ dipoles and their embedded total radiation efficiency.

The mismatch factor and embedded total radiation efficiency of the two parallel dipoles are plotted together with the S-parameters in Fig. 3.6 as a function of dipole spacing. The S-parameters have been obtained using the formulas for self impedance and mutual impedance between two dipoles given in Chapter 5 and 9, presented in Fig. 5.9 and 10.10 respectively. We see that the efficiency degrades severely when the dipoles approach each other.

3.3.4 Correlation between ports

The fading environment has statistical properties, so we need to characterize the performance in terms of the CDFs of the received signals on the different ports. The most random case is that the received voltages on the ports are completely independent, i.e., uncorrelated. In practice there will be some correlation between them. This is characterized by a complex *correlation coefficient* ρ defined by[13]

$$\rho = \frac{\sum V_1 V_2^*}{\sqrt{\sum V_1 V_1^* \sum V_2 V_2^*}} \tag{3.9}$$

where V_1 and V_2 are sequences of the samples of simultaneously received voltages of two ports numbered 1 and 2, and the sums are taken over all the samples. Sometimes we have not access to the phases of V_1 and V_2. Then, we can evaluate the envelope correlation

$$\rho_e = \frac{\sum |V_1 V_2^*|}{\sqrt{\sum |V_1|^2 \sum |V_2|^2}} \tag{3.10}$$

[13] This correlation formula is only valid for quantities with a zero mean value, like the received voltages in RIMP.

and it is easy to replace this in the formulas for the diversity gain in Section 3.4.4 by using that for Gaussian variables

$$|\rho| = \sqrt{|\rho_e|^2} \tag{3.11}$$

Formula 3.9 can readily be transformed to deterministic integrals in terms of the normalized scalar-products between the far-field functions of the two ports, i.e., according to [3],

$$\rho = \frac{\iint_{4\pi} \mathbf{G}_1(\theta,\varphi) \cdot \mathbf{G}_2^*(\theta,\varphi) d\Omega}{\sqrt{\iint_{4\pi} \mathbf{G}_1(\theta,\varphi) \cdot \mathbf{G}_1^*(\theta,\varphi) d\Omega \iint_{4\pi} \mathbf{G}_2(\theta,\varphi) \cdot \mathbf{G}_2^*(\theta,\varphi) d\Omega}} \ , \tag{3.12}$$

where $\mathbf{G}_1(\theta,\varphi)$ and $\mathbf{G}_2(\theta,\varphi)$ are the embedded far-field functions of ports 1 and 2, respectively. The effect of the correlation on diversity gain is analyzed below and quantified by (3.21). Eq 3.12 is valid between two ports of any antenna in RIMP.

It is also possible to express the same complex correlation coefficient in terms of the S-parameters measured at the antenna ports, *but only if there are no ohmic losses in the antenna*. The appropriate formula for the envelope correlation is given in [11, Eq. (5)] for the multi-port case. A careful derivation corresponding to that in [11] gives for the complex correlation

$$\rho = \frac{-(S_{11}^* S_{12} + S_{21}^* S_{22})}{\sqrt{[1-(|S_{11}|^2+|S_{21}|^2)][1-(|S_{12}|^2+|S_{22}|^2)]}} \ . \tag{3.13}$$

The denominator is readily seen to be equal to the square root of the product of the embedded element efficiencies on each of the two ports, see (3.7). The absolute value of this complex correlation is plotted in Fig. 3.12 for the two dipoles introduced in Section 3.3.3.

3.4 Characterization of diversity performance

With diversity, two antennas are used, located sufficiently far from each other (space diversity) or otherwise with orthogonal polarizations (polarization diversity), or orthogonal embedded far-field functions (pattern diversity), so that there is low coupling between them. The received statistical signals on the two ports will then be uncorrelated in RIMP, and it is very unlikely that there will be fading dips simultaneously on both ports. Therefore, by an appropriate combination of the two signals, the probability of a fading dip in the combined signal will be considerably reduced. There are several different possible combination schemes, such as switch diversity, *Selection Combining* (*SC*) and *Maximal Ratio Combining* (*MRC*), and the improvement of the CDF will be as large as 12 dB at the 1 % level of the CDF. We will here not detail the different diversity schemes and instead refer to [3]. We assume MRC in the discussions to follow.

3.4.1 Channel estimation and digital MRC processing

In order to characterize diversity and MIMO antenna systems we need to know the algorithms employed in the signal processing. We will here explain *Maximal Ratio Combining* (*MRC*).

The basis of a digital diversity and MIMO system is channel estimation. This means that the transfer functions are measured between the N_t transmitting and N_r receiving ports in

Fig. 3.2, i.e., all the $N_r \times N_t$ channels. These correspond to the S-parameters S_{ij} between the transmitting and receiving sides. We could in traditional antenna terms call this a calibration of the whole antenna system including the environment.

The channels can only be measured on the receiving sides (based on measuring known pilot signals transmitted from the transmitting side), unless the channels are the same in both directions. Such reciprocity is only present if the uplink (i.e. transmitting from terminal and receiving at base station) and downlink (i.e. the opposite) transmissions appear at the same time and frequency. This will in practice mean the time and bandwidth within which there is no fading, i.e., the coherence time and coherence bandwidth of the channels, respectively. However, this is generally not the case. Therefore, we have to deal with two cases referred to as: *a)* Channel State Information (CSI) being known and *b)* CSI being unknown, on the transmitting side. The former allows for adaptivity by signal processing also on the transmitting side. This will improve performance compared to the processing on the receiving side only. The second case is more common and will be illustrated here.[14]

The data stream is transmitted after the channels have been estimated. This is generally done by a complex amplitude modulation of the channels over time. The modulation must generally be fast compared to the coherence time of the environment (after which the channel must be re-estimated), and slow enough to be within its coherence bandwidth. Otherwise, the estimated channels will not be correct. Thus, best data transmission is present as long as the received channels are the same as the estimated channels (except for the modulation of the former). The reason is that the estimated channels are used to optimize the detection of the signal (i.e., data) on the modulated channels. However, the user and the environment change with time, so the channel matrix must be regularly re-estimated.

Let us use h_i' to denote the estimated channels on the receiving ports, and $C(t)h_i$ to denote the modulated received channels, where the time variation is used to denote the modulation. We have by the latter assumed that the modulation is fast compared to the coherence time and narrowband compared to the coherence bandwidth of the environment. Let us assume that we have a total of $N_r \times N_t$ channels with the same modulation $C(t)$. Then, the optimum way to receive the modulated signal $C(t)$ is to combine the received channels in this way

$$C_{\text{opt}}(t) = \sum_{\text{all } i} C(t)h_i \cdot w_i \tag{3.14}$$

where $w_i = h_i'^*$ are the weights with $h_i'^*$ denoting complex conjugation of the channel estimates h_i'. Thus, we combine the modulated channels by using the complex conjugate of the estimated channel values as weights, see Fig. 3.7. This sum is referred to as *MRC*. This corresponds to the *conjugate matching* in Section 2.6.4. The conjugate matching of a load ensures maximum power transfer to the load, and MRC ensures maximum power transfer of the combined power available in all the N_r received channels. It is important to be aware that this MRC is done digitally after detection and digitalization of the channels, and that the combined signal is a complex voltage quantity and not a power quantity. The MRC will also generally represent a maximization of the signal-to-noise ratio. The conjugate matching and MRC correspond to what in general EM field theory is referred to as *conjugate field matching* (*CFM*). Above, we explained MRC for when combining channels available on different

[14] CSI is known only on the receiving side in Frequency Division Duplex (FDD) systems, because the frequencies are different on uplink and downlink. It is in principle known on both sides in Time-Division Duplex (TDD) systems.

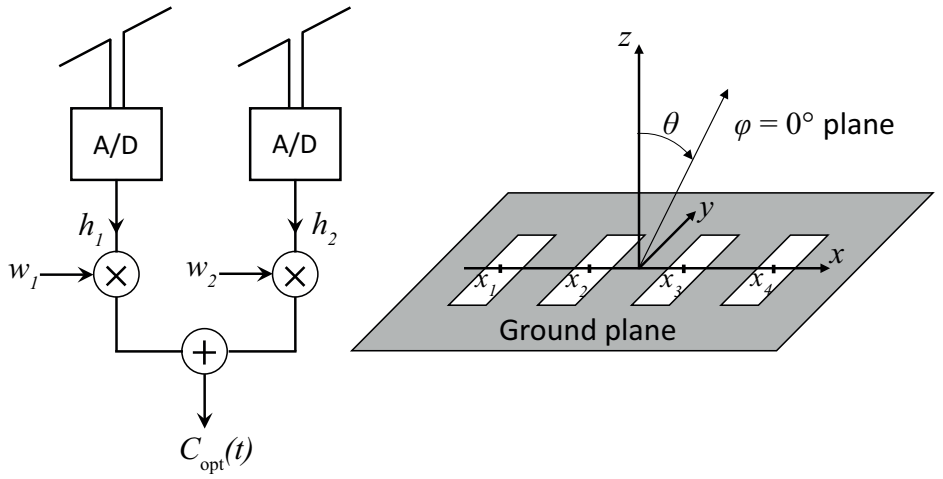


Figure 3.7: Block diagram of MRC algorithm applied to four-port antenna for which $w_1 = h_1'^*$ and $w_2 = h_2'^*$ with h_1' and h_2' being the channel estimates. The A/D box is a receiver with an analogue to digital converter. The example four-port antenna is shown to the right.

receiving ports, but the approach is general and we can combine channel values originating from different transmitting ports in the same way.

3.4.2 Example: MRC applied to 2D slot antenna case

Let us illustrate MRC for the case of a transmitting antenna with one port and a receiving antenna consisting of N_r isolated[15] short slots with 0.5λ spacing in an infinite ground plane, as shown in Fig. 3.7. The sources are then represented by short magnetic dipoles that have omnidirectional far-field functions in one plane, and we orient them so that this is the xz-plane. The ground plane limits the radiation to $\pm 90°$. We will present the formulas of the analysis for the more general case of N_r elements equally spaced along the x-axis with spacing d. The co-polar far-field function of each element i in the xz-plane is then

$$G_i(\theta) = e^{jkx_i \sin\theta} \tag{3.15}$$

by using (2.52) and (2.41) for $\varphi = 0°$. For simplicity we have here and in the rest of this section omitted constants as they do not affect the shapes of the radiation patterns. Let us now for this study choose the very simple case that we have Q incident waves from the directions θ_q having phase Φ_q and equal amplitudes. We want the total available incident power to be the same independent on the number of waves, so we use accordingly complex wave amplitudes

$$E_q = \sqrt{1/Q} e^{j\Phi_q} \ . \tag{3.16}$$

[15] Isolated elements means that we do not consider mutual coupling in this analysis.

Then, we can use (2.129) (see also (3.30)) to get the following total induced voltage at the port of element i (this is the *channel estimation*):

$$V_i = \sum_{q=1}^{Q} G_i(\theta_q) e^{j\Phi_q} = \sum_{q=1}^{Q} e^{j(kx_i \sin\theta_q + \Phi_q)} \tag{3.17}$$

These first V_i are therefore the estimated channels h_i'. After the estimation has been completed we can combine the ports in the optimum way for receiving the signal modulated on the Q waves by using the MRC algorithm in (3.14). The received combined signal becomes then, by using the weights $w_i = h_i'^*$,

$$\mathsf{MRC} = \sum_{i=1}^{N_r} V_i h_i'^* = \sum_{i=1}^{N_r} \left(V_i \sum_{q=1}^{Q} e^{-j(kx_i \sin\theta_q + \phi_q)} \right) \ . \tag{3.18}$$

The far-field function for the whole antenna is obtained by summing up all far-field functions in (3.15) over all elements by using the estimated channel weights in (3.17) as port weights. This, gives

$$G_{\text{all}}(\theta) = \sum_{i=1}^{N_r} V_i^* e^{jkx_i \sin\theta_k} \ . \tag{3.19}$$

Let us now determine the radiation patterns of this far-field function and discuss them, when $N_r = 4$ and 10. We assume first that there is one incoming wave like in a LOS environment, see the two upper-most graphs in Fig. 3.8. Note that there are two curves in each graph, corresponding to two different AoAs of the wave, 0° (solid curve) and 60° (dashed curve), respectively. We see that MRC corresponds to forming a main beam from the elements that point in the AoA of the incoming wave, like in a classical array. The main beam is very wide for the 4-element case, but it is much narrower for the 10-element case. For the case of a plane wave coming in from 60° we also see a large lobe at −90°, i.e., along the ground plane. This is the beginning of what in classical array theory a grating lobe. If the spacing is larger than 0.5λ, there may be more such grating lobes having the same level as the main lobe[16].

Let us now assume that there are two incoming waves with equal phase, see the two middle graphs in Fig. 3.8. Then, the MRC will result in one main lobe (with a shallow dip between them) covering both AoAs for the 4-element array case, and two clear main lobes for the 10-element case, one in each AoA direction. There are three sidelobes and four deep dips between these main lobes, which means that the phase in the two main lobes are the same (the phase changes by 180° when the pattern goes through a null). If the phase of one of these plane waves now is changed to be 180° relative to the other (dashed curve), two main lobes with a deep null between them are formed for the 4-element case. This allows then for constructive reception of the two incident waves, by making their corresponding received voltages have the same phase.

If there are more than two plane waves we will not be able to distinguish lobes in the directions of the waves for the 4-element array, but we see four clear main lobes for the 10-element array, as shown in the two lower graphs in Fig. 3.8.

For larger arrays the conclusion will be the same. We will not be able to form directive beams towards the AoAs unless there are more antenna ports than waves in the environment. The

[16] See Chapter 10.

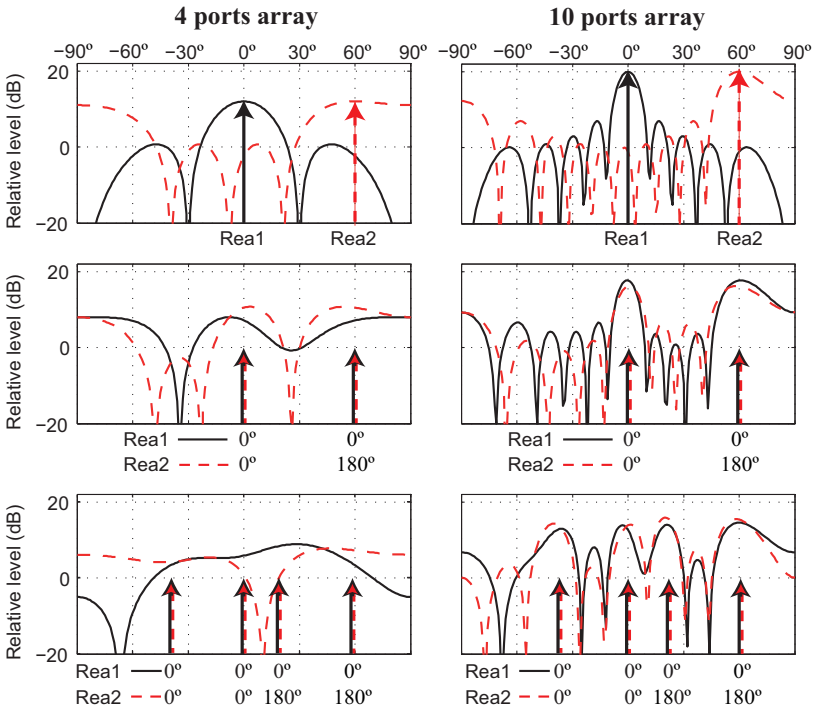


Figure 3.8: *Illustration of the far-field patterns (beams) formed on the receiving side of two different MIMO arrays when the MRC algorithm is used. The two arrays consists of 4 (left graphs) and 10 (right graphs) small slots, respectively. The slots have, half wavelength spacing on an infinite plane (see Fig. 3.7). The number of incident waves, their angles of arrival (AoAs) in the xz-plane, and their relative phases in the center of the array are explained by the arrows (AoAs) and curve legends below each graph. The AoA direction of 0° corresponds to broadside. The two curves in each graph corresponds to two different phase relations between the incident waves, except for the upper two graphs where the two curves represent different AoAs. The phases of the incident waves are explained in the legends below the AoA arrows.

MRC algorithm is still generally referred to as beam-forming, and it is the optimum way of combining the channels to achieve the highest signal-to-noise ratio.

The *Selection Combining* (*SC*) algorithm is much simpler than MRC. In SC combining, the receiving port with the strongest signal-to-noise ratio is selected at each time interval, and the others are not used. SC does not work as a classical array for the case of a single incident wave. In RIMP it is still quite efficient.

Note that the radiation patterns in Fig. 3.8 only are examples. The wave amplitudes and AoAs are statistical in nature, so we need to characterize the performance of MIMO arrays by CDFs. The presented patterns are only intended to illustrate that we only get interpretable lobes in the directions of the waves when there are fewer incoming waves than antenna ports.

3.4.3 Diversity gains (apparent, effective and actual)

The quality of a diversity antenna system is determined by the improved shape of its CDF in the environment. Improved shape means that the slope of the CDF is steeper, and any shift towards higher signal-to-noise values. Therefore, we need to specify the environment, and we need to know the CDFs at each port. We limit ourselves here to the well-defined RIMP environment. The CDFs are determined from the discrete channel samples measured on each port of the receiving antenna, and from the channel samples at the port of a reference antenna in the same environment. The samples are measured[17] or simulated for different channel realizations. The channel realizations can be obtained, e.g., by moving the antenna around in the RIMP environment, by locating it at many different fixed locations, or by changing statistically the amplitudes and AoAs of the waves in the environment. The reference antenna is typically a wideband single-port antenna with known total radiation efficiency.

A CDF of a channel is obtained from the channel samples in the following way:

1. Produce a set of M channel samples, where M is a number preferably larger than or equal 1000.
2. Compute the average received power from all samples, i.e., take the square of the absolute value of all samples, sum them up, and divide by M.
3. Normalize all samples to the square root of the average received power of the *ideal reference antenna*. This is the average received power of the reference antenna divided by its known total radiation efficiency.
4. Arrange all normalized samples from the lowest to the highest level and number them successively from 1 to M.
5. The CDF is then the curve obtained by plotting the sample number divided by the total number of samples (i.e., the cumulative probability) versus the level of the corresponding samples in dB.

Such a CDF will in an ideal RIMP environment have the theoretical Rayleigh shape except at the lowest levels where the probabilities are close to $1/M$. Therefore, we can produce more

[17] The channel samples can, e.g., be measured in a reverberation chamber, see Section 3.7, or they can be simulated as explained in Section 3.6.5.

accurate CDFs by using more samples. The shape of the CDFs in RIMP will also deviate from Rayleigh shape if the environment contains a small LOS contribution, and if it is not rich enough.

We now use the two parallel dipoles in Fig. 3.3 as an example, with a given separation between them of 0.045λ so that there is a significant mutual coupling and a low embedded radiation efficiency. Then, Fig. 3.9 shows the CDF of the fading channel amplitudes in RIMP environment. The theoretical Rayleigh distribution is included as a reference. The CDFs observed on any antenna port have the same shape as this theoretical Rayleigh, but they are shifted horizontally with respect to the Rayleigh by the value of their total embedded radiation efficiencies in dB. If we apply MRC to the signal samples on the two ports, we get the improved MRC CDF shown in the figure, having a steeper slope.

The diversity gain is the difference in dB between the MRC CDF and a reference Rayleigh CDF at a certain CDF-level, normally chosen to be 1 %*. We can distinguish between apparent, effective and actual diversity gains [7] and [12], depending on what kind of reference we use. The reference is a Rayleigh-shaped CDF corresponding to an ideal antenna with* 100 % *efficiency, and therefore the dB value is referred to as a dBR value.*

Apparent diversity gain: The reference CDF is the CDF at the port with the strongest average power levels.

Effective diversity gain: The reference CDF is the CDF at the port of the ideal reference antenna, i.e., the theoretical Rayleigh CDF.

Actual diversity gain: The reference CDF is the CDF at the port of an existing practical single-port antenna. The latter is then normally the single-port antenna that is to be replaced by the diversity antenna under test, so that we can determine the actual improvement over an existing solution. Both antennas must therefore be measured at the same location, e.g., relative to a head phantom.

In the case of actual diversity gain, the practical antenna reference shall be located in the position relative to an object (e.g., a head phantom) that corresponds to the desired position of operation of the terminal at which the antennas are located. Under these definitions, the actual diversity gain in dB is the effective diversity gain in dB minus the radiation efficiency in dB of the single-port existing antenna that the diversity antenna shall replace. The latter value is negative, so the apparent diversity gain appears much larger than it actually is if the reference was an efficient single-port antenna. The effective diversity gain represents the gain over an ideal single-port reference antenna, where the latter is characterized with no additional antenna in its proximity.

We see that in our example in Fig. 3.9 the apparent diversity gain at 1 % CDF level is 8 dBR, whereas the effective diversity gain compared to the ideal single-port antenna reference is only about 3 dBR. This means that if the receiver system allows for a fading margin of 20 dB to be able to receive with sufficient quality 99 % of the time or for 99 % of the users (i.e., 1 % CDF level), we can reduce the fading margin by 3 dB if we use this specific diversity antenna (two parallel dipoles with 0.04λ spacing) instead of a very good single antenna. This 3 dB diversity gain could easily be made larger by using larger dipole spacing, or by choosing two orthogonal dipole antennas. The theoretical maximum is 12 dB by MRC (and 10 dB by selection combining). The discrepancy between 3 dB and 12 dB is in the example mainly due to the low radiation efficiency when the dipoles are so close. This is caused by mutual

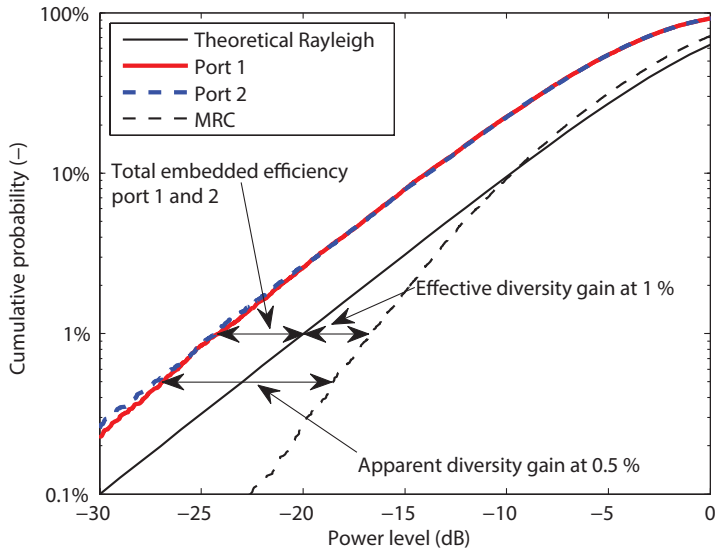


Figure 3.9: ⋆CDF of the ideal reference antenna (corrected for its finite radiation efficiency) and of a diversity antenna consisting of two parallel dipoles separated by 0.05λ.

coupling, giving large absorption in the $50\,\Omega$ load of the not-excited dipole. In contrast, the reduced diversity gain due to correlation has minor effect, as explained below.

The published results of diversity gains as a function of dipole spacing can be seen in Fig. 3.10[18] with MRC. Similar results are shown for selection combining in [7] including verification by measurements in reverberation chamber. There it is also shown significant actual diversity gain for antennas used near the human body (using a simple cylindrical head phantom). The actual diversity gain is shown to be more than 6 dB, even when the dipole separation is only 0.06λ. Antennas with uncoupled orthogonally polarized ports will naturally have larger diversity gains.

3.4.4 Theoretical determination of diversity gain

Theoretically, we can determine the effective diversity gain G_{eff} by the following formula

$$G_{\text{eff}} = e_{\text{rad}} G_{\text{app}} \ , \tag{3.20}$$

where G_{app} is the apparent diversity gain. This depends on the correlation coefficient ρ obtained from (3.12) by using the far-field functions of the two embedded ports, and for a lossless antenna the simpler equation (3.13) can be used. The relation between apparent diversity gain and correlation is given by the following approximate formula in [13, pp. 474],

$$G_{\text{app}} = 10.5 \cdot e_\rho \qquad \text{with} \qquad e_\rho = \sqrt{1 - |\rho|^2} \ , \tag{3.21}$$

[18] There exist MATLAB code for all figures of which the caption start with ⋆.

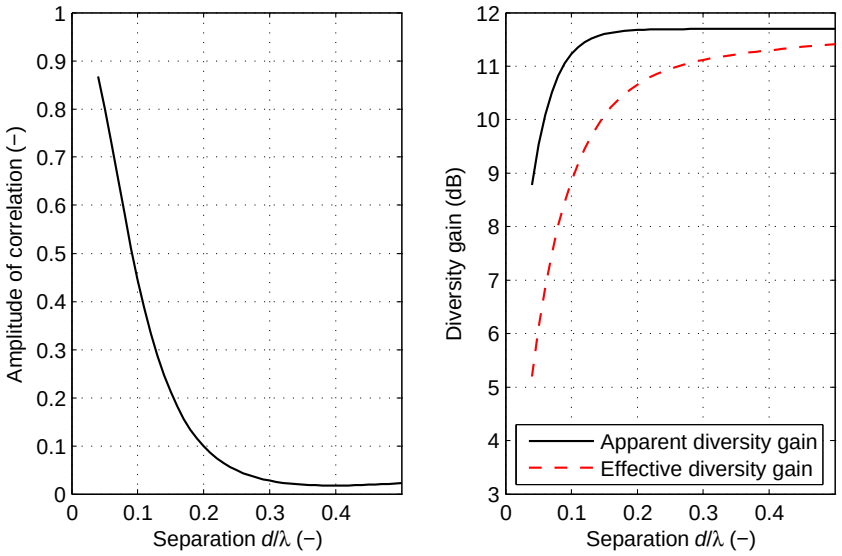


Figure 3.10: *Correlation between the ports of two parallel halfwave dipoles in RIMP, and the apparent and effective diversity gains by MRC versus dipole spacing.

where 10.5 is the maximum apparent diversity gain at 1 % CDF level with selection combining, and e_ρ is an approximate expression for the *decorrelation efficiency*, i.e., the reduction in diversity gain due to correlation between the signals on the two ports. This formula is not very accurate for correlations close to unity when compared with the more accurate formulas in [13, pp. 470]. However, if we scale ρ with a factor 0.99, the formula becomes

$$e_\rho = \sqrt{1 - |0.99 \cdot \rho|^2} \ , \tag{3.22}$$

which differs from the more accurate expression for the apparent diversity gain at 1 % CDF by less than 0.1 dB as shown in Fig. 3.11. It is worthwhile to note that the correlation must be very strong in order to cause a significant reduction of the diversity gain.

Eq. (3.21) can only be used when the embedded total radiation efficiencies are the same on the two ports. In [14] there is presented an empirical formula that can be used also when the efficiencies on the two ports are different. This is

$$G_{\text{app}} = \sqrt{\left(1 + \frac{e_{\min}}{e_{\max}}\right)^2 + 105\frac{e_{\min}}{e_{\max}}(1 - |\rho|^2)} \tag{3.23}$$

where $e_{\min}$ and $e_{\max}$ are the embedded total radiation efficiencies on the two ports.

Otherwise, and for more ports, the diversity gain can be determined from measured and computed CDFs [15]. The formulas in (3.22) and (3.23) are simple, but we still need to perform channel measurements or simulations in order to determine the complex correlation ρ from (3.9), (3.12) or (3.13).

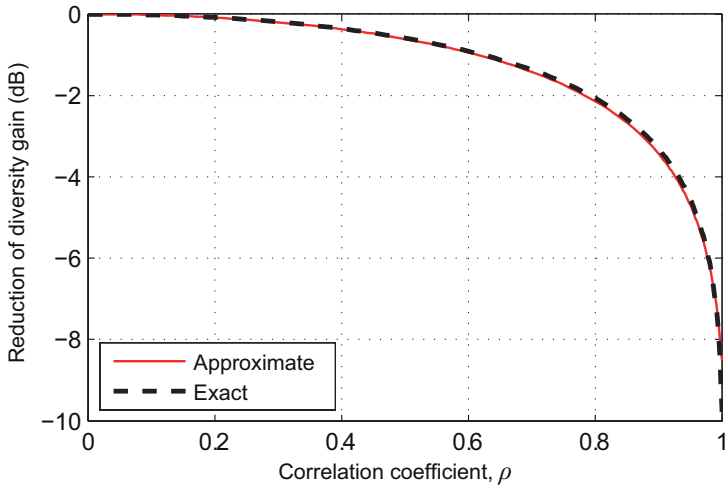


Figure 3.11: *Correlation efficiency factor at 1 % CDF, i.e., reduction of diversity gain at 1 % cumulative probability level, as a function of the pattern correlation in (3.12), obtained with the approximate equation (3.22) and the exact expression in [13].

3.5 Maximum available capacity from Shannon

In modern MIMO mobile communication systems there will be multi-port antennas on both the base station and terminal sides to form several communication channels between them, referred to as spatial multiplexing. These different communication channels are commonly referred to as *bitstreams* to separate them from the larger number of channels available to achieve diversity. For instance, four antennas on both sides form $4 \times 4 = 16$ possible channels in the way they are defined in Section 3.4.3, but there will only be four bitstreams available. This correspond to communication via the eigenvectors of the whole channel matrix, where the eigenvalues are the signal strength of each bitstream. The most common way of diagonalizing a matrix is *Singular Value Decomposition*(*SVD*). However, the SVD requires preprocessing on the transmitting side and therefore *CSI* to be known. Therefore, other bitstream decomposition methods are more common, requiring only processing on the receiving side, such as *Zero Forcing* (*ZF*).

The data on the transmitting side is distributed among the different bitstreams, which therefore contain completely different modulated information in contrast to the diversity MRC case in Section 3.5.1 that maximizes the received power for one bitstream. If CSI is known on the transmitting side it is possible to transfer much more information.

The maximum possible average channel capacity in a MIMO system can be calculated by using Shannon's fundamental formula for maximum available capacity. We illustrate this basic Shannon capacity with two simple examples in LOS. The wireless channel is treated in much more details in [16].

3.5.1 Single-port system

The maximum available capacity of a communication channel through any environment is given by Shannon's classical formula

$$C = \log_2 |1 + \mathsf{SNR}| \quad (\text{bits/s/Hz}) \ , \tag{3.24}$$

where SNR is the signal-to-noise ratio at the receiving side, and where $\log_2$ means the logarithm with base 2. This means that the capacity of a system not only depend on its bandwidth, but also depends on the SNR. Thus, if we increase the transmit power, we can transfer more bits/s/Hz over the channel[19]. This is done in practice by increasing the order of the modulation. However, we have to know the channel and its SNRs in order to benefit from this. Therefore, channel estimation and its feedback to the transmitter is important in communication engineering. The Shannon formula is also valid under fading, but then the SNR of the channel varies with time, in which case channel estimation must be performed continuously, and hence becomes a very important issue.

3.5.2 Parallel channels in LOS

If we use two parallel communication systems instead of one, we can naturally double the capacity if they are mounted in a way that there is no interference. This can be done, e.g., by locating the antennas of the extra system in the nulls of the radiation patterns of the other, or to use orthogonal polarization of the two systems. Thus, the maximum available capacity of two such systems corresponding to two parallel channels (i.e., two bitstreams) becomes

$$\begin{aligned} C = C_1 + C_2 &= \log_2 \left| 1 + \frac{1}{2}\mathsf{SNR}_1 \right| + \log_2 \left| 1 + \frac{1}{2}\mathsf{SNR}_2 \right| \\ &= \log_2 \left(\left| 1 + \frac{1}{2}\mathsf{SNR}_1 \right| \left| 1 + \frac{1}{2}\mathsf{SNR}_2 \right| \right) \ , \end{aligned} \tag{3.25}$$

where the indices on C and SNR denote systems 1 and 2. The factor $1/2$ on all the SNRs means that we have divided the total transmitted power equally between the two channels. We will now discuss whether or not it is possible to increase the capacity by combining the two antennas on each side to a classical directive array so that a single communication system with larger SNR is formed.

The signal to noise ratio of a LOS system is given by

$$\mathsf{SNR} = \frac{P_r}{N} = \frac{P_t G_t G_r \left(\frac{\lambda}{4\pi r}\right)^2}{N} \ , \tag{3.26}$$

where P_r is the received power, P_t is the transmitted power, G_r is the realized gain of the receiving antenna, G_t is the realized gain of the transmit antenna, N is the noise power, and $\lambda/4\pi r$ is called the free space attenuation. We now double the gains on both the transmitting and receiving sides by combining on each side two equal uncoupled directive antennas to a

[19] This is valid if the modulation of the signal is allowed to change adaptively.

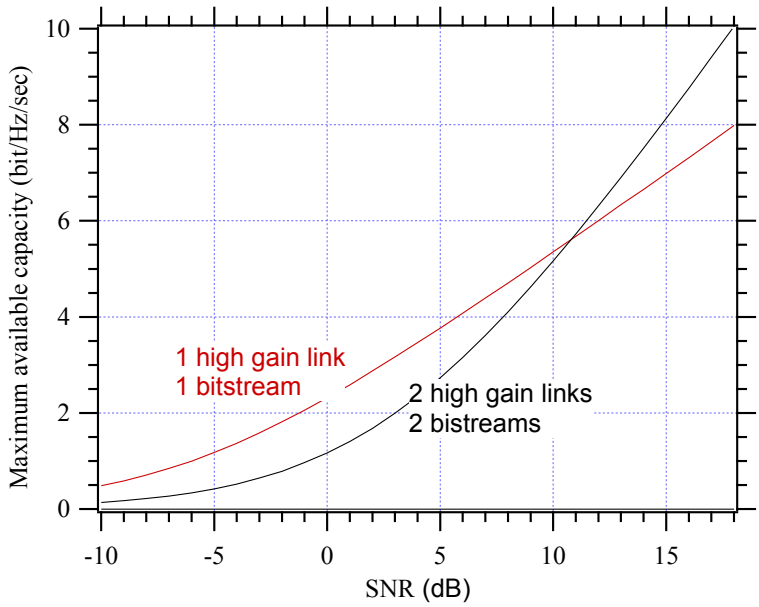


Figure 3.12: Maximum available capacity according to Shannon's formula for two parallel 1×1 systems (one antenna on each side), and one 2×2 antenna system where the two antennas on each side are combined to a directive array of double directivity on each side.

classical directive array with the double gain. Then, the SNR of the new single-bitstream system with classical two-element arrays on both sides becomes

$$\text{Classical arrays:} \quad \mathsf{SNR}_{2\times 2} = 4\mathsf{SNR}\,, \tag{3.27}$$

i.e., four times larger than the SNR of each of the two systems. The condition for this equation to be valid is that the noise power is independent of the antenna gain. This assumption is normally true in practical ground-based communication systems.

The capacities for the two cases; two parallel low gain links with the power equally distributed between the two transmitting antennas, and one high gain link, become

$$\begin{array}{ll} \text{Two parallel low gain links:} & C_{2\text{lg}} = 2\log_2|1+\tfrac{1}{2}\mathsf{SNR}| \\ \text{One high gain link:} & C_{1\text{hg}} = \log_2|1+4\mathsf{SNR}| \end{array} \tag{3.28}$$

in (bits/s/Hz). These two equations are plotted and compared in Fig. 3.12 to show that as long as $\mathsf{SNR} > 10.8\,\text{dB}$ in each channel it is advantageous (for increasing capacity) to use parallel systems with low gain antennas rather than one system with high gain antennas. The maximum available capacity increases by up to a factor 2 when we double the number of channels (for large SNR). These are theoretical results. In practice there will be other limitations that may change this conclusion, e.g., hardware and cost constraints.

3.5.3 Parallel channels in multipath

In a fading environment the instantaneous capacity is given by the same formula as for the LOS case, but the formula for the SNR is different. The space attenuation is much larger

due to the scattering from all blocking objects, and the SNR will fade with the fading signal level. Therefore, it is common to use an average maximum available capacity (also called ergodic capacity) in characterization of systems used in multipath, obtained by averaging the Shannon capacity over the distributed variations of the power samples of the signal. The power samples are distributed exponentially in Rayleigh fading.

In a MIMO system with multiple-port antennas on both the transmitting and receiving sides there are many channels through the environment. Consider the two-port antennas from the example, in Section 3.5.2, but now in a multipath environment. Then, the signal (from each of the two transmitting ports) will be received on both the receiving antenna ports. In principle we will have the same situation as in LOS. If SNR is small it is better to combine the ports to one bitstream using diversity whereas if SNR is large we can use spatial diversity to get two bitstreams (links) and thereby higher capacity.

The general formula for the maximum available capacity of a MIMO system with an N_t-port transmit antenna and an N_r-port receiving antenna is

$$\widetilde{C} = \log_2 \left(\det \left(I_{\mathrm{N_r \times N_t}} + \frac{\mathsf{SNR}}{N_t} \widetilde{\mathbf{H}}_{\mathrm{N_r \times N_t}} \widetilde{\mathbf{H}}^*_{\mathrm{N_r \times N_t}} \right) \right) , \tag{3.29}$$

where $I_{\mathrm{N_r \times N_t}}$ is a unit matrix, $\widetilde{\mathbf{H}}_{\mathrm{N_r \times N_t}}$ is a normalized complex channel matrix, and $\widetilde{\mathbf{H}}^*_{\mathrm{N_r \times N_t}}$ is the complex conjugate transpose of $\widetilde{\mathbf{H}}_{\mathrm{N_r \times N_t}}$. The normalization must be done with respect to average received power on a reference antenna with known radiation efficiency. The $\widetilde{C}$ and $\widetilde{\mathbf{H}}$ notation with the tilde symbol means that C and $\mathbf{H}$ fade with time, so that we need to average the $\widetilde{C}$s over the distribution function of the signal power in order to compare the channels. The channel matrix contains fading samples of the channels which means that we need to average $\widetilde{C}$ over the distribution of $\widetilde{\mathbf{H}}_{\mathrm{N_r \times N_t}}$.

The Shannon formula represents the maximum available capacity, and this is very far from what is achievable in practice. In practical systems typically 6 dB more transmitted power is needed than given by Shannon's formula, for a given desired capacity. Therefore, the Shannon capacity is not very useful for characterization of antenna systems. We will instead choose the ideal digital threshold receiver for such characterization, to be introduced in Section 3.8.

3.5.4 Normalization

The correct normalization of the channel matrix is very important. It must be done with respect to the square root of the time averaged power received on a single-port reference antenna with 100 % radiation efficiency[20]. And this reference antenna must be located in the same environment, and we must require that the total power transmitted from the antenna on the opposite side of the environment is the same as when we produced the channel matrix. We will in the next subsection show how typical channel matrices can be computed and measured.

[20] See §3 in the procedure in Section 3.4.3.

3.5.5 Numerical simulation of channels in multipath

We will now explain how to simulate a fading environment numerically. First, we generate a set $\{\mathbf{E}_k\}$ of $k = 1, 2, ..., K$ incident plane waves with AoAs (θ_k, φ_k) randomly and uniformly distributed over a sphere surrounding the MIMO array to be investigated. The sources $\mathbf{E}_k$ in the set have independent and complex Gaussian-distributed amplitudes of the θ- and φ-polarized components. We compute the combined open-circuit received voltage V_i on each embedded antenna port due to all these waves by using the equivalent circuit of the antenna in receiving mode (see Fig. 2.22), i.e.,

$$V_i = -\frac{2j\lambda}{\eta I} \sum_{k=1}^{K} \mathbf{G}_i(\theta_k, \varphi_k) \cdot \mathbf{E}_k \ , \tag{3.30}$$

where all parameters have been explained previously[21] and $\mathbf{G}_i(\theta, \varphi)$ is the embedded far-field function on port i. The power absorbed in the load $R_{_L}$ becomes

$$P_{_i} = \frac{1}{2} R_{_L} \left| \frac{V_i}{Z_a + Z_{_L}} \right|^2 \ , \tag{3.31}$$

which has its maximum value when the antenna is matched to the load, i.e., when $Z_a^* = Z_L$, as shown in Section 2.6.4.

In order to determine the performance of a specific multi-port antenna with N_r ports in an $N_r \times N_t$ MIMO system, we assume that the N_t ports at the opposite transmitting side are completely uncorrelated with equal amplitude. We can model this by generating N_t sets of incident plane waves, with $k = 1, 2, ..., 20$ plane waves in each set, and calculating N_t values of received voltages V_i at each receive antenna port i. The received voltages are then put row-wise into the channel matrix $\mathbf{H}$.

We compute the received voltages of the reference antenna in the same way by using (3.30) and one set of incident plane waves, but now using the far-field function of the reference antenna. This should as explained Section 3.5.4 and Section 3.4.3 be a matched antenna with 100 % radiation efficiency.

Then, we repeat the calculation of the received voltages V_i on the N_r receive antenna port, and the received voltage on the reference antenna port $i = \text{ref}$, a large number of times M_{CDF}, for example $M_{\text{CDF}} = 10000$ times. The larger M the better convergence of the CDF, and the better estimate of the average received power. The average received power on the reference antenna is evaluated using

$$P_{\text{av}} = \frac{1}{M_{\text{CDF}}} \sum_{m=1}^{M_{\text{CDF}}} (P_{\text{ref}})_m \ . \tag{3.32}$$

Now we have a reference level and can determine the CDF and the total embedded radiation efficiency at each port i of the multi-port that we are investigating, and thereby evaluate the diversity gains as explained in Section 3.4.3.

[21] See Section 2.6.3.

The embedded radiation efficiency on port i of the multi-port antennas becomes e.g.,

$$e_{\mathrm{rad}_i} = \frac{1}{P_{\mathrm{av}}} \frac{1}{M_{\mathrm{CDF}}} \sum_{m=1}^{M_{\mathrm{CDF}}} (P_i)_m \ , \tag{3.33}$$

where P_i is given by (3.31).

3.6 Emulation of RIMP using reverberation chamber

The reverberation chamber is a large metal cavity provided with mode stirrers and one or more antennas, as shown in Fig. 3.13. It has been used for more than 30 years for *Electromagnetic Compatibility* (*EMC*) measurements of electromagnetic susceptibility and radiated emissions of electronic devices, as described in the overview article in [1]. The basic theories of it have also been well understood [13], [17]-[19]. The main EMC application has been to generate high field strength for susceptibility testing. In recent years the reverberation chamber has also been developed to a more accurate instrument for measuring the characteristics of desired radiation of small antennas and active mobile terminals. In particular when these are intended for use in Rayleigh fading, such as for wireless/mobile communications in urban or indoor environments. We here describe how the reverberation chamber works, and how to measure the performance of antennas as defined in the previous chapters.

3.6.1 Mode stirring (mechanical, platform, polarization)

The basic measurement setup in a reverberation chamber is illustrated in Fig. 3.13(a). This setup is used both for calibration of the chamber and measuring passive antenna performance. We measure the transmission coefficient S_{21} between two antennas in the chamber. One of these is the reference antenna or the Antenna Under Test (AUT), and the other is part of the chamber and could be wall-fixed. The reference antenna is used during calibration. The AUT is used during tests. The chamber is so large that several cavity modes are excited at the frequency of test, so the level of S_{21} may be larger than in free space, but it may also be much lower, depending on how the modes are excited and how they interfere (combine) resulting in fading. Reverberation chambers are provided with different ways of stirring the modes and thereby the S_{21} level varies, i.e., emulation of fading. The most common mechanical mode stirring methods are to use fans or other large rotating or translating mechanical structures.

The chamber in Fig. 3.13 has the following stirring capabilities:

1. Mechanical stirring: This is achieved by means of two plate-shaped stirrers, which can be moved along a complete wall and along the ceiling using electrical stop motors. The larger volume the stirrers cover, the better.

2. Platform stirring [20]: The AUT test is located on a rotatable platform which moves the antenna in the chamber. This stirring method is very effective in small chambers.

3. Polarization stirring [21]: There are three orthogonal wall-fixed chamber antennas, and S_{21} is successively measured between each of them and the AUT. Thus, the chamber gets a good polarization balance.

There is also a fourth very effective stirring method, referred to as frequency stirring [22]. This corresponds to averaging S_{21} (complex frequency stirring) or $|S_{21}|^2$ (power frequency stirring) over a frequency band during the processing of the results. This will be treated in the next section.

3.6.2 The S-parameters of the chamber and of the antennas

Fig. 3.14 illustrates how S_{11} and S_{12} are measured between two antennas inside a reverberation chamber. We can physically argue that S_{11} of the AUT must consist of two contributions: one contribution being the S_{11}^a from the antenna itself as if it was located in free space, and another S_{11}^c from the chamber, i.e.,

$$S_{11} = S_{11}^a + S_{11}^c \ . \tag{3.34}$$

The former of these contributions is deterministic, whereas the latter is random as a result of the stirring. If the number of independent samples of S_{11}^c is large enough, S_{11}^c gets a complex Gaussian distribution with zero mean. Therefore, it is possible to determine S_{11}^a by complex averaging of the measured S_{11} over all stirrer positions i, and if needed by additional complex frequency stirring, as shown in [23], i.e.,

$$S_{11}^a = \frac{1}{M} \sum_{\text{all } i} S_{11} = \overline{S}_{11} \ , \tag{3.35}$$

where $\overline{S}_{11}$ is a compact way of writing the average of S_{11}. Similarly we can find S_{22}^a of the wall-mounted chamber antenna.

In a similar way, we can argue that S_{21} must consist of two contributions, one deterministic contribution S_{21}^d being the same as in free space, and another statistic contribution S_{21}^c coming from the chamber, i.e.,

$$S_{21} = S_{21}^d + S_{21}^c \ . \tag{3.36}$$

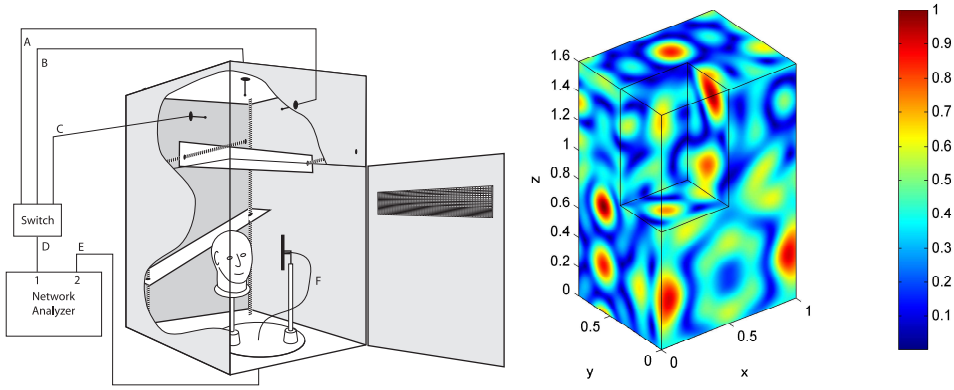
3.6.3 Rayleigh fading, Rician fading and AoA distribution

We would like the reverberation chamber to provide the rich isotropic reference environment RIMP with Rayleigh fading. This requires the direct coupling in (3.36) to be as low as possible. The direct coupling follows the free space transmission equation, also referred to as Friis transmission equation[22], yielding

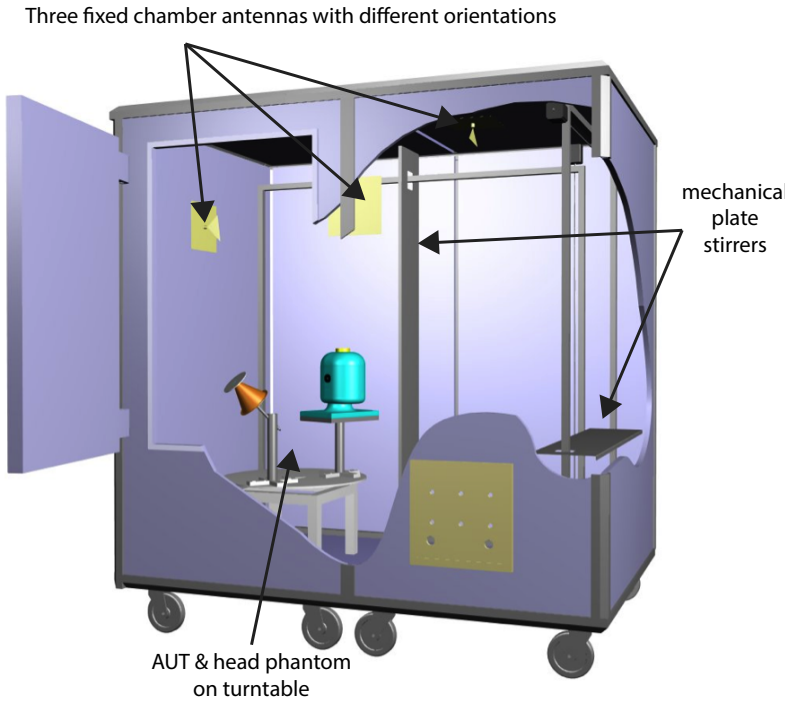
$$|S_{21}^d|^2 = \left(\frac{\lambda}{4\pi r}\right)^2 G_t G_r \ , \tag{3.37}$$

where r is the distance between the antennas, λ is the wavelength, and G_t and G_r are the realized gains of the wall-mounted antenna and AUT, respectively, in the direction of the opposite

[22] See page 58.



(a) (b)



(c)

Figure 3.13: Typical reverberation chamber (RC) for Over-The-Air (OTA) measurements. (a) Sketch of RC and measurement setup for calibration and measurements of performance of multi-port antennas (sketch shows two parallel dipoles). (b) *Color plot of relative E-field distribution computed at the walls of simplified RC and on three orthogonal surfaces meeting in the center of the chamber, for illustration of the statistical nature of the fields. When the stirrers move, the field distribution will change. (c) Drawing of RC with wideband disk-cone antenna for calibration of the transfer function. The calibration antenna can be replaced by a passive Antenna Under Test (AUT) or an active wireless Device Under Test (DUT).

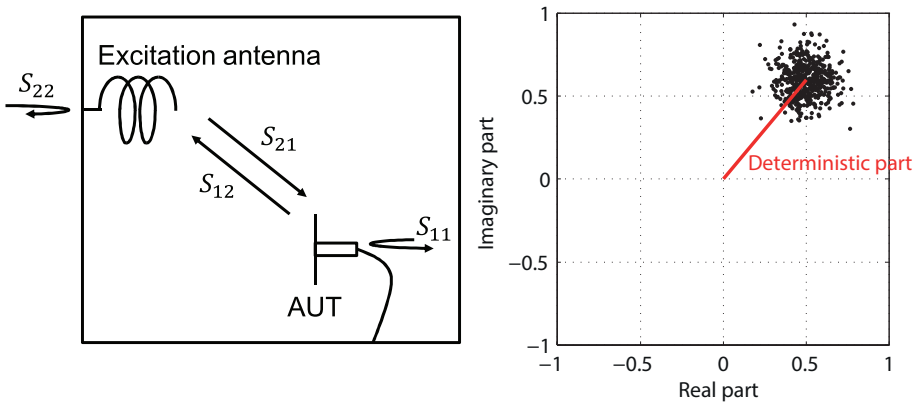


Figure 3.14: *Example of measured S-parameters of two antennas in a reverberation chamber as a result of mode stirring, and illustration of the deterministic and random contributions to them. Note that the deterministic part in reality is much smaller.

antenna. Therefore, the main lobes of the antennas should not point towards each other. The polarization stirring is very effective to reduce direct coupling, as the direction of the AUT will be different in different stirrer positions. The direct coupling together with loading the chamber can alternatively be used constructively to get a controlled Rician distribution for test purposes, as proposed in [24].

From the discussion in the previous subsection, it is clear that the reverberation chamber provides Rayleigh fading if the direct coupling is reduced to be sufficiently smaller than the chamber contribution to S_{21}. This can be done by platform stirring, using a metal shield to block the direct coupling, or by using a wall fixed antenna that has a null in the direction of the AUT. In [25], it was shown that each mode in a rectangular cavity can be expressed as eight plane waves, and that the arrival directions of these plane waves are uniformly distributed over the unit sphere, if there are enough excited modes. Therefore, the reverberation chamber represents a RIMP environment if it is large enough, i.e., the desired reference environment defined in Section 3.1.3.

3.6.4 Average transmission level (Hill's formula) and calibration

The transmission between two antennas in free space follows Friis transmission equation (3.37). The corresponding formula for transmission between two antennas located in a reverberation chamber is Hill's transmission formula, see [18]. This is valid when there is no direct coupling, and it presumes a large chamber with many excited modes. We choose to refer to Hill's formula as the chamber power transfer function, and to present it in the following form

$$G_{\text{chm}} = |S_{21}^c|^2 = \frac{P_r}{P_t} = \frac{c^3 e_{\text{rad}_1} e_{\text{rad}_2}}{16\pi^2 V f^2 \Delta f} \ , \tag{3.38}$$

where f is the frequency, c is the velocity of light, V is the chamber volume, e_{rad_1} and e_{rad_2} are the *total radiation efficiencies* of the two antennas, and Δf is the *average mode*

bandwidth. The latter consists of four additive contributions due to: the wall losses, power leakage from the chamber, the antennas present in the chamber, and any absorbing objects in the chamber, i.e.,

$$\Delta f = \sum_{\substack{\text{all}\\\text{walls}}} \Delta f_{\text{wal}} + \sum_{\substack{\text{all}\\\text{slots}}} \Delta f_{\text{slt}} + \sum_{\substack{\text{all}\\\text{ant}}} \Delta f_{\text{ant}} + \sum_{\substack{\text{all}\\\text{obj}}} \Delta f_{\text{obj}} \ , \tag{3.39}$$

with

$$\begin{aligned} \Delta f_{\text{wal}} &= \frac{2A}{3V}\sqrt{\frac{c\rho f}{\pi\eta}} \ , & \Delta f_{\text{lek}} &= \frac{c\sigma_{_1}}{4\pi V} \ , \\ \Delta f_{\text{ant}} &= \frac{c^3 e_{\text{rad}}}{16\pi^2 V f^2} \ , & \Delta f_{\text{obj}} &= \frac{c}{2\pi V}\sigma_a \ , \end{aligned} \tag{3.40}$$

where η is the free space wave impedance, A the area of a conducting surface (such as a chamber wall) with surface resistance ρ, $\sigma_{_1}$ is the leakage cross section of a narrow slot in the chamber wall, and σ_a is the absorption cross section of an absorbing object. $\sigma_{_1}$ and σ_a are defined in [18][23] and have slow frequency variation compared to the explicit frequency variation in the formulas. The total Δf can be very different in different practical chambers, corresponding to chamber *quality factors* (Q) of between 30 and several thousands ($Q = f/\Delta f$).

In practice a measurement of radiation efficiency goes as follows. First, the chamber transfer function is determined by calibration, using a reference antenna with known radiation efficiency e_{rad_2}, and with the AUT present in the chamber with its port match-terminated. The actual measurement will then be done by match-terminating the reference antenna and mounting the cable to the port of the AUT. The ratio between the two average power transfer functions of the chamber in the two cases equals the ratio between the radiation efficiencies of the reference antenna and the AUT. The radiation efficiency e_{rad_1} of the chamber-fixed antenna does not need to be known because it will be the same both when measuring the reference antenna and the AUT.

It is also possible to calibrate the chamber without having the AUT inside. However, then, the reference antenna must also be removed when the AUT is measured. This is a simplified procedure, because we can use the same calibration when measuring several AUTs, and in particular it is simpler when measuring many active terminals. However, this simplified procedure only works if the chamber is loaded so much that the reference antenna and AUT does not represent any significant contribution to the Δf of the chamber. If they do so, the results of the measurements may be wrong, as the chamber transfer function does no longer have a linear dependence on e_{rad} of the reference antenna, and the AUT. This may happen in particular if the reference antenna or AUT are made of materials which absorb radiation and thereby increase Δf even when the antenna ports are open or short-circuited.

[23] It should be noted that here we have expressed the average power transfer function in terms of the average mode bandwidth Δf rather than in terms of the quality factor (Q) of the modes, which Hill used in [18]. The reason is that the formula for Δf is much more compact, because the different Δf contributions are additive, which Q contributions are not. Also, for specific chambers (at least loaded ones), the average mode bandwidth will not vary much with frequency, and therefore the value of Δf characterizes the chamber better than Q, over a large frequency band.

3.6.5 Frequency stirring on net transfer function

We have found that if we remove the free space mismatch factor from S_{21} in (3.38), we get an average chamber transfer function that varies slower with frequency. This can then be frequency stirred for better accuracy without losing resolution due to variations in the mismatch factor. The free space input reflection coefficients of the two antennas can be obtained by complex averaging of the S_{11} and S_{22} measured in the reverberation chamber, as explained before. By removing the two mismatch factors we get the following formula for the maximum available (or net) chamber power transfer function

$$G_{\text{chm}} = \frac{1}{N} \sum_{N} \frac{|S_{21}^{c}|^2}{(1 - |\overline{S}_{11}|^2)(1 - |\overline{S}_{22}|^2)} \ , \tag{3.41}$$

where N is the number of stirrer positions. This function can be frequency stirred to improve accuracy.

3.6.6 Number of independent samples & accuracy

In order to perform accurate measurements in the reverberation chamber we need the chamber transfer function in (3.38) to be proportional to the radiation efficiency independent of which antenna or terminal we use, like in Hill's theoretical transmission formula. This is possible only if the mode stirring creates enough independent samples. The S_{21} samples are complex Gaussian distributed if the chamber is well stirred. Then, the relative accuracy by which we can estimate G_{chm} has a standard deviation of [6]

$$\sigma = 1/\sqrt{N_{\text{ind}}} \ , \tag{3.42}$$

where N_{ind} is the number of independent samples. This means that we need $N_{\text{ind}} = 100$ for an accuracy of $\pm 10\,\%$, i.e., $\pm 0.5\,\text{dB}$. Thus, it is crucial that we can obtain at least 100 independent samples by the mode stirring.

The number of independent samples is determined primarily by the mode density in the chamber, i.e., by the number of modes per MHz. This is approximately given by the classical formula

$$\frac{\partial N_{\text{mod}}}{\partial f} = V f^2 8\pi/c^3 \ . \tag{3.43}$$

The number of independent samples is proportional to the mode density, but the proportionality constant is not known. It depends on chamber loading, mode stirring methods, mechanical stirrer shapes, and chamber shape (at least for small chambers). However, we can preliminary state the following approximate relation between the mode density and number of independent samples:

$$N_{\text{ind}} \leq 8 \left[\frac{\partial N_{\text{mod}}}{\partial f} \right] (\Delta f + B_{\text{fs}} + B_{\text{mch}}) \ , \tag{3.44}$$

where B_{fs} is the bandwidth of the frequency stirring, and B_{mch} is a mechanical stirring bandwidth used to characterize the stirrers. The factor 8 is due to platform stirring and is rooted both empirically and by physical reasoning. The latter argument goes as follows: Each cavity mode can be written as a sum of 8 plane waves [23], and therefore we can get

8 times more independent samples than modes by moving the antenna or terminal under test around in the environment, which we do by platform stirring. The $\leq$ sign in (3.44) means that this is an upper bound when we have enough samples and the mechanical mode stirring is sufficiently strong. A thorough study of uncertainties in reverberation chambers using (3.44) and different loads to control Δf can be found in [26].

A good reverberation chamber can provide measurements of efficiency-related quantities with standard deviations better than 0.5 dB, and even approaching 0.1 dB [26] without doing any frequency stirring. This requires that the chamber should not be loaded too heavily, i.e., Δf should be small compared to the frequency of operation. The problem with frequency stirring B_{fs} is that the resolution becomes worse (resolution bandwidth increases), and then, we cannot resolve variations in the radiation efficiency which are faster than B_{fs}. In practice the resolution is somewhat better by using the mismatch correction to the transfer function before power averaging, like in (3.41). The reason is that normally mismatch efficiencies varies faster with frequency than efficiencies due to ohmic losses.

3.7 Measurements in reverberation chamber

3.7.1 Calibration and characterizing multi-port antennas

The procedure for measuring a single- or multi-port antenna in a reverberation chamber is briefly described as follows. The AUT is located inside the reverberation chamber in such a way that it is more than 0.5 wavelengths from the walls and mechanical stirrers[24]. We also locate a single reference antenna with known radiation efficiency far enough from the AUT to avoid significant direct coupling[25]. We connect one of the AUT ports to a source, i.e., a network analyzer, and terminate all the other ports and the reference antenna in 50 Ω. We gather S-parameters between the port and the three chamber antennas (used for polarization stirring) for all positions of the platform and the mechanical stirrers and for all frequency points. The measurement procedure is then repeated for every antenna port, also with the unconnected ports terminated in 50 Ω, for exactly the same stirrer positions and position of the array inside the chamber. Thus, the field environment is exactly the same when measuring every port. The complex transmission coefficients S_{21} between the connected port and each of the three fixed chamber antennas, as well as the reflection coefficients S_{11} of each of the chamber antennas and S_{22} of the array port, are stored for every stirrer position and frequency point. Finally, we connect the reference antenna to the network analyzer and perform the same measurements as for the array. During the reference measurements, the AUT with all its ports terminated in 50 Ω must be present in the chamber. This is necessary because the loading of the chamber (and thus the Q-factor) needs to be the same during the measurements of both the reference antenna and the AUT. A further reason is that the AUT itself loads the reverberation chamber considerably even when there is a lossy object such as a head phantom inside the chamber[26].

In a small chamber it is advantageous to use frequency stirring (averaging) to improve accuracy. In such cases we correct the complex samples of S_{21} with mismatch factors due to

[24] For directive antennas longer distances will be needed in the main lobe direction.

[25] For non-directive antennas a spacing of half to one wavelengths is sufficient.

[26] A simplified procedure with only one antenna at the time in the chamber is described at the end of Section 3.6.4. However, this is not recommended when measuring multi-port antennas.

both S_{11} and S_{22} before the frequency stirring, see (3.41). We also normalize the corrected S_{21} samples to the reference level corresponding to 100 % radiation efficiency. This is obtained from the corrected S_{21} samples measured for the reference antenna, and its known radiation efficiency. We refer briefly to these corrected and normalized samples of S_{21} as the normalized S_{21} values. The normalized S_{21} values represent estimates of the channel matrix $\mathbf{H}$ between the wall antennas and the AUT inside the chamber. Therefore, from the measured S-parameters the diversity gain and capacity can be obtained as explained in preceding sections.

In the above explanations the MIMO channel is considered between the chamber antennas and the multi-port AUT. It is of course also possible to locate two multi-port antennas inside the chamber and measure the MIMO channels between the ports of these two antennas, and evaluate them as a complete MIMO system.

3.7.2 Radiated power, receiver sensitivity and data throughput

Complete mobile devices and cell phones can also be measured in reverberation chambers. The total radiated power is readily measured as follows: first find the chamber transfer function in (3.41) by using a reference antenna with known efficiency. Next measure the radiated power from the phone by connecting a spectrum analyzer or a wireless communication test instrument to the chamber-fixed antennas.

Even more important for wireless internet and multimedia terminals is the receiver sensitivity, because this will directly affect the time for downloading data. The receiver sensitivity is characterized by the signal level needed to give a certain specified Bit Error Rate (BER), or Frame Error Rate (FER) for CDMA systems. When averaged over many directions in the radiation pattern, the limiting static signal level is referred to as the *Total Isotropic Sensitivity* (*TIS*). An alternative in reverberation chambers is to measure BER/FER during continuous fading to determine what is referred to as *Average Fading Sensitivity* (*AFS*) [27]. The AFS is a more realistic performance parameter than TIS, and, it is much faster to measure. The TIS and AFS are related to each other for the case of flat fading. Flat fading appears when the signal bandwidth is much smaller than the coherence bandwidth of the multipath channel [26]. The research on measurements of active devices in reverberation chamber is reviewed in [28]. The modern LTE (i.e., 4G) systems are very flexible and allows for a dynamically changing data rate. The reverberation chamber has been successively applied to measure data rate throughput.

The work in [29] describes a simple ideal threshold receiver, which explains advanced digital receivers very well. Using this, it is also possible to model throughput in RIMP that is in agreement with what is measured in a reverberation chamber. We will in the next section explain this useful threshold receiver and how it is used for modeling the throughput.

3.8 System modeling using digital threshold receiver

The LOS system has a simple path loss formula, known as Friis transmission equation[27]. There exists also a simple average transmission formula for the RIMP environment, for the

[27] See Section 2.5.3.

special case when this is emulated in the reverberation chamber. This is referred to as Hill's equation (3.38). Real environments may also be RIMP, but the transmission formula is of course very different in real-life multipath, because of all the scattering objects. Generally, the received power level decays faster through a scattering environment than in free space. Still the RIMP environment is relevant, and the reverberation chamber emulates its physical properties. We can also remove the effect of the large additional attenuation by calibration with a reference antenna with known total radiation efficiency.

In a modern communication system like the LTE 4G system, the quality of its wireless devices is best characterized by their throughput. This is the data rate versus received power during continuous fading for a fixed modulation and bandwidth. Therefore, it is important to be able to perform repeatable quantitative measurements of throughput. We will here show how this can be done by using a reverberation chamber, and we will also show how we can model it by a simple threshold receiver model [29].

3.8.1 The digital threshold receiver

The modern digital receivers contains an RF[28] receive amplifier and an analogue to digital transformer. Let us connect a cable (thus, the conducted case) between a base station emulator (i.e., a wireless communication test instrument) and a digital LTE receiver as shown to the right in Fig. 3.15. This corresponds to the stationary case of no fading, referred to by the term *Additive White Gaussian Noise*(*AWGN*) in propagation and wireless communication system literature. Then, we observe throughput curves that are very steep, almost like a threshold, see Fig. 3.16. The threshold varies linearly with the system bandwidth. Therefore, system modeling during fading is simplified a lot by introducing an ideal theoretical threshold as shown by the dashed curves in the same graph. Thus, if we know the threshold P_t, we have for the conducted case

$$\mathsf{TPUT}_{\mathrm{con}}(P) = \mathsf{TPUT}_{\max} \begin{cases} 0 & \text{when } P < P_t \\ 1 & \text{when } P > P_t \end{cases} , \tag{3.45}$$

where P is the maximum available power at the port of the receiver, and $\mathsf{TPUT}_{\max}$ is the maximum throughput.

When we locate the digital threshold receiver in a dynamic fading environment, we observe throughput variations due to this threshold. Sometimes the received level is above the threshold and otherwise not, even if the transmitted power is constant. This variation depends on the CDF of the channel. Actually, the relative throughput becomes equal to counting the number of observations that the instantaneous power (i.e., channel) is above the threshold, compared to the total number of observations. Therefore, the relative throughput becomes equal to the *Probability of Detection* (*PoD*) , i.e., the complement of the outage probability. We can express these relations as follows

$$\mathsf{TPUT}(P_{\mathrm{av}}) = \mathsf{TPUT}_{\max}\mathsf{PoD}(P_{\mathrm{av}}/P_t) = \mathsf{TPUT}_{\max}\left\{1 - \mathsf{CDF}(P_t/P_{\mathrm{av}})\right\} , \tag{3.46}$$

where P_{av} is the maximum available average received power in the environment, and the CDF is the CDF of the channel. The maximum available average received power is found by *i)* calibrating the chamber with a single-port reference antenna to find the average transfer function,

[28] RF is an abbreviation for Radio Frequency.

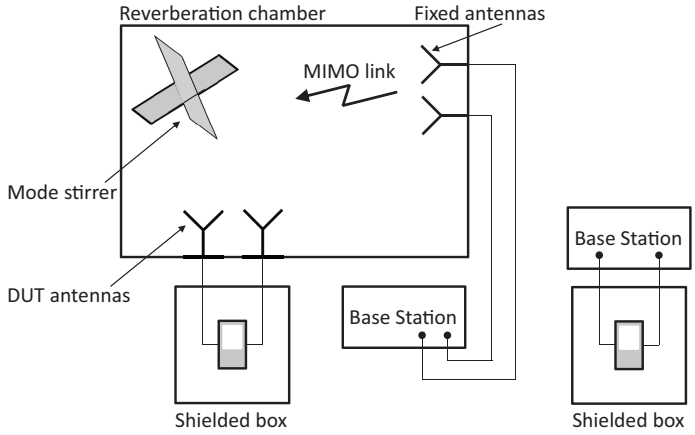


Figure 3.15: Illustration of measurement setup for measuring an LTE device in reverberation chamber (left) and conducted (right). The device is here shown to be located inside a shielded box outside the reverberation chamber. This makes it possible to use external antennas with known radiation efficiency and correlation, so that we can model the effect of them. However, this is not necessary when characterizing actual devices.

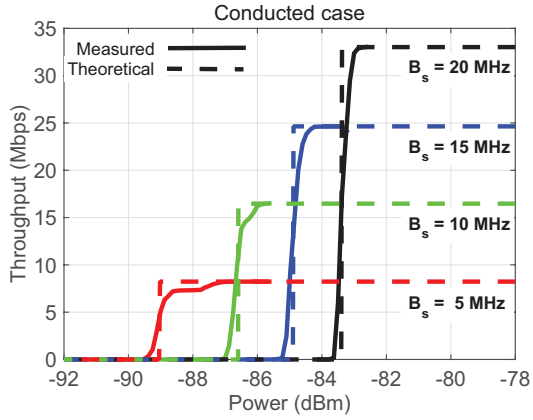


Figure 3.16: Measured (solid) and theoretical (dashed) throughput curves for an LTE device for the conducted AWGN case, i.e., when a cable is connected to the receiving port of the wireless device. The theoretical curves represent the threshold model.

ii) normalizing this average by the total radiation efficiency of the reference antenna, and *iii)* multiplying this normalized average with the fixed transmitted power. The normalization corresponds to using a reference antenna with 100 % total radiation efficiency, and thereby get the maximum available average received power. The CDF changes depending upon the signal processing, and therefore we can model the corresponding TPUT improvement if we know the processing algorithms. We have already introduced the MRC algorithm for antenna diversity, and we will now show how this also can be used to model the *Orthogonal Frequency Division Multiplexing* (*OFDM*) in the LTE system.

3.8.2 Modeling OFDM in LTE 4G system

Previous wireless communication systems suffered from irreducible bit errors if the environment had strong fading with long *time delay spreads*, i.e., a small coherence bandwidth[29]. The 4G LTE system is benefitting a lot from the OFDM, which represents diversity in the frequency domain. If we have a fading null at one frequency, we can instead benefit from a strong level at another frequency, providing the system bandwidth is larger than the *coherence bandwidth*. The coherence bandwidth is in a reverberation chamber equal to the average mode bandwidth, and it has a specific inverse relation to the time delay spread [30]. It is possible to achieve realistic coherence bandwidths in a reverberation chamber.

The OFDM is realized as a spectrum of narrow frequency bands. Each band is narrower than normal real-life coherence bandwidths to avoid irreducible bit errors. Together the spectra of narrow bands make up the system bandwidth. The different OFDM bands are in principle combined by using the MRC algorithm, i.e., by estimating the channels in each band, weight the signal in each band by the estimates, and add them. Thus, this corresponds to diversity in the frequency domain, or simply frequency diversity. We get an SNR improvement if some part of the system bandwidth contains uncorrelated frequency channels. We will model the relative number of uncorrelated frequency bands by

$$N_{\text{fd}} = \mathsf{NINT}\left(\frac{B_s}{B_c}\right) \ , \tag{3.47}$$

where NINT is a function taking the nearest integer of its argument, B_s is the system bandwidth and B_c is the coherence bandwidth. Thus, we get a higher diversity order the larger the B_s/B_c is.

3.8.3 Theoretical and measured results for i.i.d. diversity case

The threshold model has been used to produce theoretical throughput curves by using *a)* the measured thresholds in Fig. 3.16, and *b)* the formula for the frequency diversity in (3.47). The results are plotted for three different system bandwidths and coherence bandwidth in Fig. 3.17a, and the corresponding measured curves in 3.17b. The measurements have been done in a reverberation chamber. We see that the simple threshold receiver can model very accurately the throughput measured in a RIMP environment. The throughput curves corresponds to the *i.i.d.* case[30], which we have achieved in the measurements by using *i)* external

[29] The descriptors of the environment is overview in [28].
[30] See Section 3.1.3.

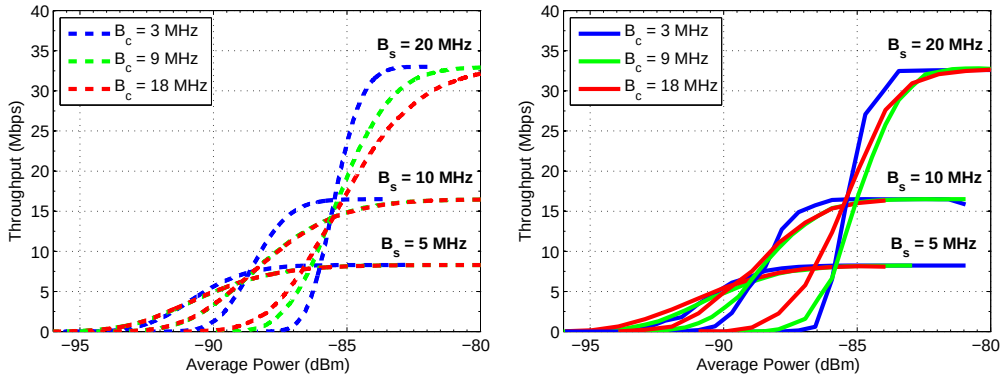


Figure 3.17: Theoretical (a) and measured (b) throughput curves for different system bandwidths and coherence bandwidths of the RIMP (i.i.d.) environment.

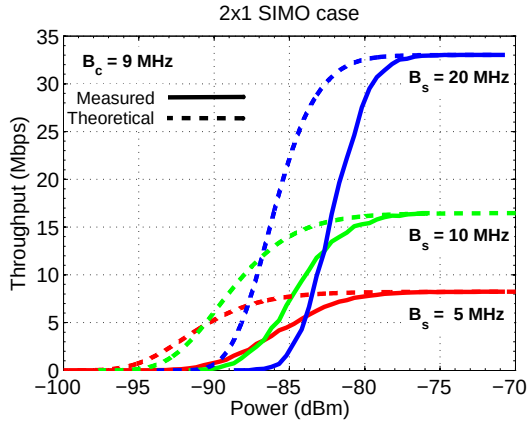


Figure 3.18: Measured (solid) and theoretical (dashed) OTA LTE throughput for 2×1 MIMO system (also called SIMO) with externally connected antenna for different LTE system bandwidths and coherence bandwidths. The measured results are for lossy antennas with efficiency of $-4\,\mathrm{dB}$ while theoretical results are for $0\,\mathrm{dB}$ efficient antennas.

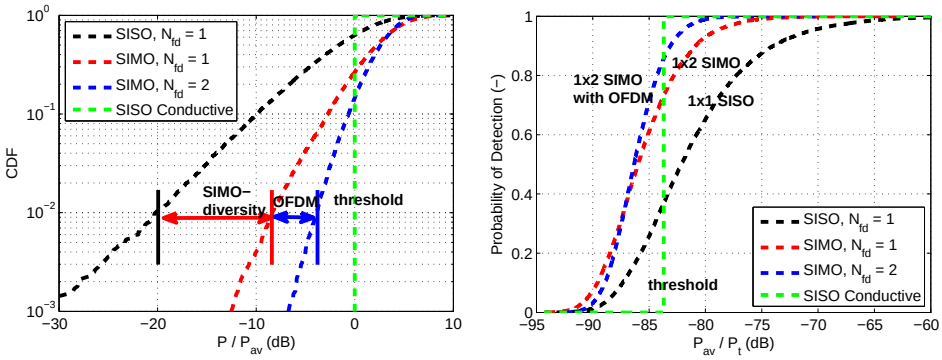


Figure 3.19: Theoretical CDF (left) and PoD (right) for the same channel matrix and i.i.d. case. The effect of the antenna diversity (difference between 1×1 and 2×1 curves) and the OFDM can clearly be seen in both curves. The diversity gains are illustrated at 1 % level in the CDF graphs. Similar diversity gains can be found at 99 % level in the PoD graphs, but it is better then to read the diversity gains at 10 % level, which has better accuracy.

uncoupled antennas on the device, and *ii)* the emulated RIMP environment of the reverberation chamber. Similar curves can be found in [31]. The threshold receiver model shows similar agreement with measured throughput for 2-port diversity antennas, see Fig. 3.18 but note the different efficiencies of the theoretical and measured curves to make them easier to distinguish. Such throughput data rate curves for the i.i.d. case are well known in communication theory, and they can also now be used to characterize commercial LTE devices by measurements in RIMP. The curves represent i.i.d. curves, which means that the performance will be worse for practical implementations of the antenna hardware (due to embedded radiation efficiencies and correlation) and the MIMO and OFDM algorithms.

The threshold receiver has till now only been used for RIMP environment, but it is very useful for producing throughput curves also in other environments, if the statistics of the environments are known. The relative throughput is also equal to a PoD, see (3.46), which makes it very easy to understand and extend, such as to modeling higher level system characteristics on the overall network level. The relation between the CDF and the PoD can be seen in Figure 3.19 for RIMP including both antenna diversity and OFDM.

The theories presented here are shown only for single bitstreams, i.e., diversity. Extensions to more bitstreams are described in the next section (see also [32]-[33]).

3.9 MIMO multiplexing to obtain multiple bitstreams

This section deals with the multi-bitstreams case (i.e., spatial multiplexing). In order to use the throughput model in (3.46) we need to diagonalize the channel matrix for each channel realization, i.e., at each instance of it, to determine the SNR of each bitstream at each realization of the channel. Thereafter, we obtain the CDF of the signal of each bitstream by observing the SNR over several realizations, i.e., over several instances such as over an

observation time interval, or over a distribution of users. The receiver will usually estimate the channel via a training sequence, and this *Channel State Information* (*CSI*) may or may not be available to the transmitter. The CSI cannot be transferred to the transmitting side if the coherence time of the fading is too short. Then, the required short intervals of the feedback will cause too large overhead on the capacity. Nevertheless, for Time Division Duplex (TDD) system the channel will satisfy reciprocity so that the CSI will be known on the transmit side without any feedback. Depending on the availability of the CSI on the transmitting side, the MIMO diagonalization can take different forms, which we will discuss in the following subsection.

3.9.1 Diagonalizing the channel matrix

For the case when only the receiver knows the CSI, we will assume that the transmitter transmits each bitstream on different ports of the transmitting antenna. Then, we also assume that the diagonalization is done by processing using the *Zero-Forcing* (*ZF*) algorithm [34], as illustrated in Fig. 3.20. For each *OFDM* subcarrier the MIMO channel is assumed to be flat in the same way as in Section 3.8.2, i.e.,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \ , \tag{3.48}$$

where $\mathbf{H}$ is the MIMO channel matrix at the subcarrier frequency, $\mathbf{x}$ and $\mathbf{y}$ are the transmitted and received signal vectors (i.e., linear matrices), respectively, and $\mathbf{n}$ is the noise vector with independent identically distributed (i.i.d.) Gaussian elements with a variance of unity[31]. Note also that we do not use any index for the subcarrier in the flat channel model in (3.48), just of notational convenience. Let now $\mathbf{h}_i$ be the i^{th} column of $\mathbf{H}$ and x_i be the i^{th} element of $\mathbf{x}$, where $i = 1, ..., N_t$ (with the number of transmit antennas denoted as N_t). Then, (3.48) can be rewritten as

$$\mathbf{y} = \mathbf{h}_i x_i + \sum_{j \neq i} \mathbf{h}_j x_j + \mathbf{n} \ . \tag{3.49}$$

The first term in the right side of (3.49) corresponds to the i^{th} (intended) bitstream, and the second term represents the interference with respect to the i^{th} bitstream caused by all the other bitstreams. A ZF receiver projects the i^{th} stream into the subspace orthogonal to the one spanned by $\mathbf{h}_1, ..., \mathbf{h}_{i-1}, \mathbf{h}_{i+1}, ..., \mathbf{h}_{N_t}$. Note that this projection can be done for any numbers of bitstreams up to the rank of the channel matrix. The rank depends on how independent the different rows of the channel matrix are, or rather how uncorrelated they are when observed over several channel realizations. When the CSI is known at both transmitting and receiving sides, the MIMO channel diagonalization can be done via normal *Singular Value Decomposition* (*SVD*). Then, we need to apply beamforming (i.e., precoding) on the transmitting side in addition to a decoding on the receiving side, and the results are that each bitstream will correspond to a distinct eigenmode of the MIMO channel matrix [34]. Let the SVD of $\mathbf{H}$ be

$$\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H \ , \tag{3.50}$$

where $\mathbf{U}$ and $\mathbf{V}$ are the unitary matrices, and $\mathbf{\Lambda}$ is a diagonal matrix consisting of the singular values of $\mathbf{H}$. Then, the precoding and power allocation (distribution of signal power between

[31] The notation in (3.48) is in agreement with the common notation in MIMO literature, so we adopt it here even though it does not fit well to the common antenna theory notation. Thus, the vectors in (3.48) are linear matrices not representing geometrical positions in three-dimensional space.

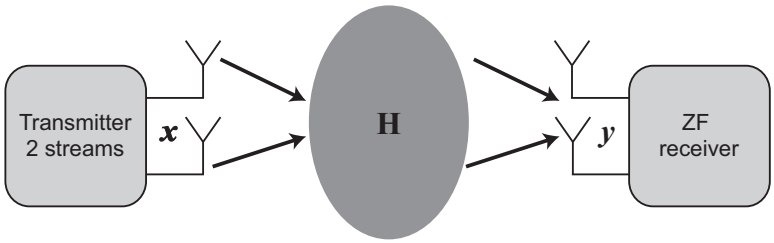


Figure 3.20: *Illustration of a 2×2 MIMO system with ZF receiver and two transmitted bitstreams.

the two bitstreams) is done by multiplying the signal vector $\mathbf{s}$ by $\mathbf{VP}$, i.e., $\mathbf{x} = \mathbf{VPs}$, where $\mathbf{P}$ is a diagonal matrix whose elements correspond to the allocated power for each bitstream. The decoding is done by multiplying $\mathbf{y}$ by $\mathbf{U}^H$, $\mathbf{r} = \mathbf{U}^H\mathbf{y}$. The resulting parallel MIMO channels are then the rows of this matrix equation

$$\mathbf{r} = \mathbf{\Lambda P s} + \mathbf{z} \ , \tag{3.51}$$

where $\mathbf{z} = \mathbf{U}^H\mathbf{n}$. Note that $\mathbf{s}$ and $\mathbf{z}$ have the same statistics as $\mathbf{x}$ and $\mathbf{n}$, respectively, because $\mathbf{V}$ and $\mathbf{U}$ are unitary matrices. Depending on the power allocation matrix $\mathbf{P}$, different performance can be achieved (please refer to [34] for details).

The data rate throughput can be modeled for the i.i.d. case in the same way as for single bitstream with MRC algorithm. The CDF is determined by assuming a Gaussian channel, and we apply the threshold receiver to each instance of the two bitstream channels. The OFDM is included with MRC algorithm making the CDF steeper[32]. The throughput for a given number of bitstreams is then obtained as the throughput of all bitstreams when all of them have enough signal level to come above the threshold. Therefore, the throughput of two bitstreams will improve for this case of fixed number of bitstream, by allocating more power to the weakest bitstream channels, so that they all reach the threshold at the same time for each realization. This is referred to as *inverse power allocation*, see next subsection. This inverse scheme is a result of a practical constraint with a fixed receiver threshold (this change in steps when system bandwidths is changed). In Shannon's formula there is no such practical constraints, and then the optimum is to allocate more power to the strongest bitstream channel. This latter is in communication systems literature referred to as *water-filling power allocation*.

The resulting beams on the transmitting side (for SVD only) and on the receiving side (for both ZF and SVD) can be plotted as radiation patterns, in the same way as for the MRC single-beam case in Fig. 3.8. However, they become very difficult to interpret in scattering environments with many waves present. They also depend very strongly on the phase relations between the incoming waves, like for the MRC case. We therefore omit presenting such patterns here, and instead present the modeled and measured throughput in RIMP for the 2-bitstream 2×2 MIMO case in the next section.

[32] Section 3.8.2 deals with modeling OFDM (frequency diversity).

3.9.2 Measurements of two bitstreams in reverberation chamber

We will here show measurements of the throughput for a 2-bitstream case by using both ZF and SVD algorithms. For the MRC case we showed measurements on a commercial LTE device and the agreement with the modeling in RIMP was good, even though the algorithm actually used in the commercial devices was not known. We will now show 2-bitstream results achieved when we have full control of the algorithm used in the measurements. This is possible with Software-Defined Radio (SDR), and we use a commercial SDR system by which we can use LabVIEW for the programming of the algorithm.

The sampling rate was set to 400 kHz and each symbol was selected as 8 samples, resulting in a symbol rate of $50k$ symbols/s. This allows for a maximum bit rate of 100 kbps with gray coded Quadrature Phase Shift Keying (QPSK). However, due to a large packet overhead, the achievable bit rate was only 33 kbps. The systems settings were operating frequency of 915 MHz, a root raised cosine as the pulse shaping filter with a roll off factor of 0.5 and a length of 6, and a system bandwidth of 75 kHz. The latter, is much smaller than the 3.5 MHz coherence bandwidth for the measurement setup in the Reverberation Chamber (RC) [7]-[8]. The used RC has a size of $1.75 \times 1.80 \times 1.25\,\mathrm{m}^3$, and is equipped with two translating plate stirrers with sizes of $0.97 \times 0.40\,\mathrm{m}^2$ and $0.88 \times 0.30\,\mathrm{m}^2$, respectively, and a turntable platform with a diameter of about 0.6 cm. It is shown in Fig. 3.13 and was used also for the measurements in Section 3.8. Both transmitting and receiving antennas are wideband triangular-shaped monopole antennas that are orthogonally polarized with sufficient separation to ensure no correlation. The stirrers in the RC run stepwise to 300 positions, and at each stirrer position 200 packets are sent. The measurements are performed over the power range of -60 dBm to -90 dBm. Fig. 3.21 shows the simulated and measured throughputs of 2×2 MIMO systems using the SDR device and reverberation chamber [33]. It shows the relative throughputs of the SVD-based 2×2 MIMO systems both with equal and inverse power allocation between the two bitstreams. As a reference, we also plot the relative throughput of the 2×2 ZF-based MIMO system in the same figure. The solid curves in the figure represent measured throughput and the dashed curves correspond to the simulated throughput using the throughput model. There is good agreement between measurements and simulations. Furthermore, we see that, without the *inverse power allocation*, the SVD-based MIMO throughput for a fixed MCS is even slightly worse than that of the ZF-based MIMO. The reason is that with SVD, the best eigen-channel is better than for the ZF-case, but the worst eigen-channel is worse (and the probability of detecting two streams equals that of detecting the worst one, when the data rate is fixed).

The normal *water-filling power allocation* technique puts more power on the best eigen-channel and less in the worst one. Therefore, the relative throughput of an SVD-based MIMO system with water-filling technique will be even worse than the equal power allocation case. Therefore, the water-filling technique is not used in practical systems such as LTE. We see from Fig. 3.21 that the SVD-based MIMO with the inverse power allocation is about 2.5 dB better (in terms of power cost) than that without power allocation and about 1.5 dB better than the ZF-based MIMO system. The results also show that the measurements in reverberation chambers are equal to the i.i.d. case if the antennas are uncoupled. Uncoupled antennas makes the received single amplitudes uncorrelated in RIMP according to (3.13).

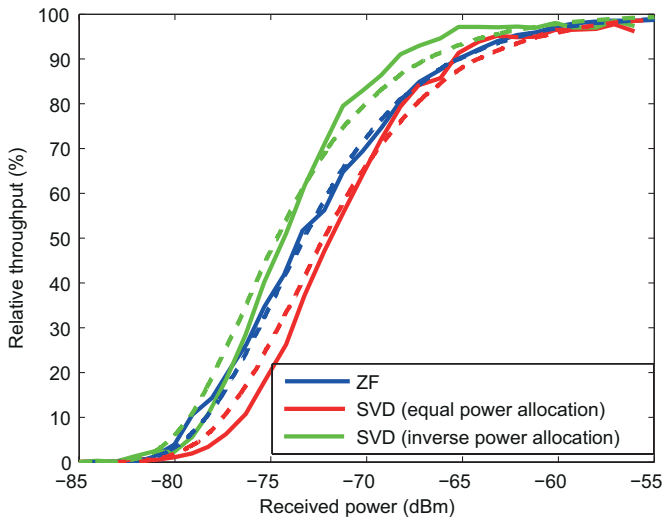


Figure 3.21: *Measured (solid curves) and theoretical (dashed curves) throughputs of ZF- and SVD-based 2×2 MIMO system with and without inverse power allocation.

3.9.3 Quality of throughput in terms of MIMO efficiency

The quality of reception in RIMP of a single bitstream was in Section 3.4 characterized in terms of a diversity gain. This is the improvement in dB of the CDF relative to the Rayleigh-distributed CDF received on an *ideal reference antenna*. The improvement will in practice be reduced, due to correlation and the embedded element efficiency as explained in Section 3.4.4. Such degradation can be quantified by a diversity-efficiency in *dBiid*, i.e., the level of the achieved CDF at a certain CDF level relative to the CDF of the i.i.d. case. The multi-bitstream case can similarly be quantified by an MIMO efficiency in dBiid. We will not give details here, but refer to [35].

3.10 Example: Polarization diversity and multiplexing in LOS

We will here show a simple LOS example that illustrates the difference between diversity gain and multiplexing efficiency. We choose a very clean case when there are two antennas locating in free space and pointing towards each other. Each of the antennas has two orthogonal linearly polarized ports, but the polarizations on the transmitting and receiving sides are not aligned with each other. Then, we will show how the different diversity and MIMO algorithms handles this case.

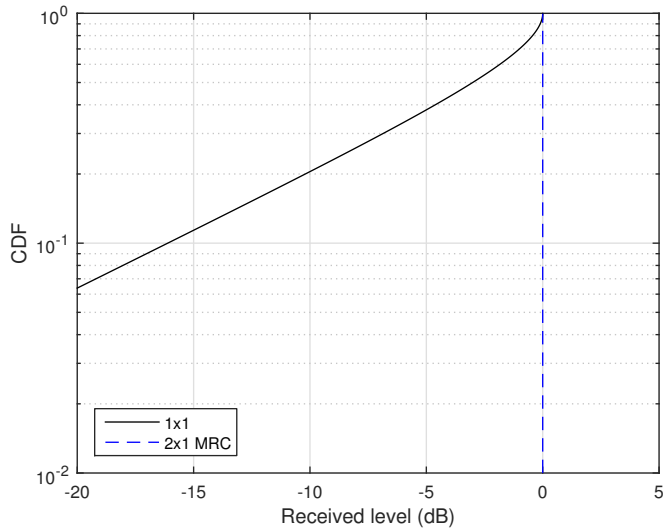


Figure 3.22: CDF of the received voltage at the port of a linearly polarized antenna receiving an incident wave of arbitrary linear polarization. If two orthogonally polarized receive antenna ports are used, the MRC algorithm gives a constant received level, i.e., the CDF is a vertical line at the 0 dB level.

3.10.1 Single bitstream

Let us first use only one port on the transmitting and receiving sides. Assume that the polarizations corresponding to these ports are $\hat{p}_{t_1}$ on the transmitting side and $\hat{p}_{r_1}$ on the receiving side. Then, the channel value that corresponds to the received voltage on the receiving port, will be[33]

$$h_{11} = \hat{p}_{r_1} \cdot \hat{p}_{t_1} = \cos(\phi_0) \ , \tag{3.52}$$

where ϕ_0 is the angle between the polarization vectors of the transmitting antenna (incident wave) and the receiving antenna, and where we have normalized the channel value to its maximum value for aligned polarizations. For an arbitrary polarization, ϕ_0 will be distributed uniformly between 0° and 360°, but we can reduce this to the interval 0° to 90° by using the symmetric properties of $|\cos(\phi_0)|$. Then, we can use probability theory to determine the CDF, which becomes

$$\mathsf{CDF} = 1 - \arccos(|h_{11}|)(2/\pi) \ .$$

This is plotted versus h_{11} in dB in Fig. 3.22. Alternatively, we could evaluate (3.52) in dB for N uniformly distributed values of ϕ_0 on the interval $0° - 90°$ and produce the CDF directly from these values. We see from Fig. 3.22 that there is a 20 % probability of levels being more than 10 dB below the maximum level, and a 10 % probability of levels being lower than 16 dB below the maximum.

Consider now the case of two receiving antenna ports with orthogonal polarizations $\hat{p}_{r_1}$ and $\hat{p}_{r_2}$. Then, the received voltage on the two receiving ports will be

$$h_{11} = \hat{p}_{r_1} \cdot \hat{p}_{t_1} = \cos(\phi_0) \qquad \text{and} \qquad h_{21} = \hat{p}_{r_2} \cdot \hat{p}_{t_1} = \cos(90° - \phi_0) = \sin(\phi_0) \ . \tag{3.53}$$

[33] According to the equivalent circuit in Fig. 2.22 where we have normalized to the case when the two polarizations are aligned.

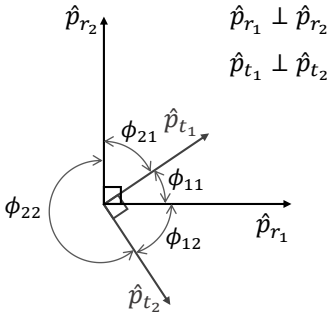


Figure 3.23: Illustration of polarizations of transmitting and receiving antennas.

The optimum weights become by using the MRC algorithm the same as the channels themselves, because the channels are real. Thus, by adding the two weighted channel values the received combined voltage becomes (according to (3.14))

$$h_{\text{MRC}} = \cos^2(\phi_0) + \sin^2(\phi_0) = 1 \ . \tag{3.54}$$

Therefore, by using a 2-port dual-polarized receive antenna, we get a constant received signal in LOS independent of the polarization of the transmitting antenna. Thus, the MRC algorithm automatically combines the two polarizations on the receiving side to that of the incoming wave. Thus, we have got a polarization diversity gain in LOS of 16 dB at the 10 % CDF level, and 10 dB at the 20 % CDF level, by using MRC in combination with a dual-polarized receiving antenna, see Fig. 3.22. This is a much larger diversity gain than what we have in RIMP environment with normal fading.

3.10.2 Two bitstreams

Let us now use the second port of the transmitting antenna to generate one more bitstream. Then, the four channels making up the channel matrix become (see Fig. 3.23)

$$\begin{aligned} h_{11} &= \hat{p}_{r_1} \cdot \hat{p}_{t_1} = \cos(\phi_{11}) \ , & h_{12} &= \hat{p}_{r_1} \cdot \hat{p}_{t_2} = \cos(\phi_{12}) \ , \\ h_{21} &= \hat{p}_{r_2} \cdot \hat{p}_{t_1} = \cos(\phi_{21}) \ , & h_{22} &= \hat{p}_{r_2} \cdot \hat{p}_{t_2} = \cos(\phi_{22}) \ . \end{aligned} \tag{3.55}$$

The optimum excitation (i.e., weight) for receiving bitstream 1 with MRC will be (see Fig. 3.7 and (3.14))

$$w_{11} = h_{11}^* = \hat{p}_{r_1} \cdot \hat{p}_{t_1} = \cos(\phi_{11}) \ , \qquad w_{21} = h_{21}^* = \hat{p}_{r_2} \cdot \hat{p}_{t_1} = \cos(\phi_{21})$$

and correspondingly for bitstream 2

$$w_{12} = h_{12}^* = \hat{p}_{r_1} \cdot \hat{p}_{t_2} = \cos(\phi_{12}) \ , \qquad w_{22} = h_{22}^* = \hat{p}_{r_2} \cdot \hat{p}_{t_2} = \cos(\phi_{22}) \ .$$

This gives then the two combined received voltages

$$\begin{aligned} h_{b_1} &= h_{11} \cdot w_{11} + h_{21} \cdot w_{21} = \cos^2(\phi_{11}) + \cos^2(\phi_{21}) = 1 \ , \\ h_{b_2} &= h_{12} \cdot w_{12} + h_{22} \cdot w_{22} = \cos^2(\phi_{12}) + \cos^2(\phi_{22}) = 1 \ . \end{aligned}$$

The reason both becomes unity is the same as for (3.54), i.e., the two polarizations on the transmitting side are orthogonal so that ϕ_{11} and ϕ_{21} are separated by 90°, as well as ϕ_{12} and ϕ_{22}. The interference between the two bitstreams is determined by

$$h_{b_1} = h_{11} \cdot w_{12} + h_{21} \cdot w_{22} = \cos(\phi_{11})\cos(\phi_{12}) + \cos(\phi_{21})\cos(\phi_{22}) = 0 \ .$$

The reason of the zero value is that ϕ_{11} and ϕ_{21} are separated by 90°, as well as ϕ_{11} and ϕ_{22} equal. Thus, for this special case it worked to use MRC twice to determine the bitstream channels.

Thus, the MRC algorithm is able to ideally decouple any two orthogonal linearly polarized incoming waves, provided we use orthogonal linearly polarized receive antennas with the same directive gain for both polarizations. We get two ideal CDFs of the same level, but we have increased the power by a factor two when introducing the second polarization.

This ideal transmission of two bitstreams would not have worked if the two receiving antennas had not been orthogonal. Then, we have to use the ZF algorithm in order to decouple them, and the CDF performance would degrade. If the transmitting antennas had not been orthogonal, we could have used SVD to decouple the channels with best performance, but that would have required even a processing of the excitations on the transmitting side.

3.11 Antennas for use on handsets

The antennas that are used on handsets can be of any type. The most important is that they can be miniaturized. The size fundamental size limitations on small antennas are treated in Chapter 11. The smaller they are the more narrow bandwidth. We will not explain any examples of small antennas in this textbook, because there are so many of them. They are known under names such as inverted-F antennas, Planar Inverted-F Antennas (PIFAs), dielectric resonator antennas. The inverted-F antenna can be made of metal wires or strips. The PIFA is a half microstrip patch antenna grounded at the end[34]. Many small antennas are reviewed in [36].

The mutual coupling analysis of multi-port small antennas are important, because the correlation in RIMP decrease if the mutual coupling decrease as already seen from equation (3.13). Several techniques exist to reduce it in order to reduce the mutual coupling between small antennas, such as those in [37]-[39].

3.12 Exercises

1. **Diversity gain from S-parameters:** The diversity gain can be found from the S-parameters of lossless 2-port antennas. Eq. 3.13 requires that we also know the phase of the S-parameters, but often we do not know that. Use the S-parameters in Fig. 3.6 for 0.1λ and 0.4λ spacing between the two dipoles, and the results in Fig. 3.9 to determine what the phase of $S_{11}^* S_{21}$ must be. The two dipoles are identical.

   We now assume that the two dipoles are fed by coaxial cables with 0.5 dBloss. If the S-parameters were measured with these cables, what is then the effective diversity gain?

[34] See also the reference list of Chapter 6 about microstrip antennas.

2. **Radiation patterns of MRC arrays:** Look at the far-fields of short slots in Chapter 5, and modify the equations in Section 3.4.2 for the case that the two slots are oriented along the x-axes. Make a small MATLAB program and see how the new element pattern affects the far-field of the 2-port arrays when MRC is used for the same incident wave examples as in Section 3.4.2. Make simulations also for a larger array of 4 ports. What can we say about the main lobes of the radiation patterns when we have 1, 2, 4 and 6 incident waves? Change the phases of the waves and observe how the lobes change.
3. **Diversity gain from incomplete CDFs:** We need to have almost 1000 samples of the channels in order to have a reasonable accuracy of the CDF at 1 % level. This takes a long time to measure, and if we do not have the far-field patterns we cannot simulate it either. Also, most small antennas for mobile phones are quite lossy, so we cannot use the simpler S-parameter equations. An alternative is then to measure the embedded total radiation efficiency and the correlation, and to estimate the diversity gain from the equations in Section 3.4.4. The reason is that the embedded efficiency and the correlation converges much faster than the CDFs at 1 % level. Make a simulation of two CDFs and determine the difference in accuracy as a function of the number of samples for both approaches. Compare the uncertainty of the diversity gain with the uncertainty of the estimate of the average power level of the CDF, see (3.42) in Section 3.6.6.
4. **Average power transfer function in reverberation chamber:** Hill's formula (3.38) describes the average power transfer function in a reverberation chamber. Use MATLAB and plot Hill's formula as a function of frequency for the reverberation chamber in [26] for different mode bandwidths. You will in [26, Table I] find mode bandwidths between 1 and 10 MHz obtained by different loadings of the chamber. Plot also the free space attenuation from one of the wall antennas to the AUT when you assume that the distance is half the largest chamber width. How much larger is the transfer function through the chamber than the free space attenuation at 1 GHz for 1 MHz, 5 MHz and 10 MHz average mode bandwidths?
5. **Polarization diversity and multiplexing in LOS:** The diversity by MRC combination and MIMO multiplexing concepts are very easy to illustrate and understand for directive antennas in Line-Of-Sight (LOS). Consider the same as in Section 3.10, i.e., two antennas pointing against each other and both being located in the far-field of each other. Note that some of the points below were treated already in Section 3.10, but we have extended them with circular and non-orthogonal polarizations.

   a) Let the two antennas be linearly polarized, and let the polarizations of both antennas be aligned. Write the expression for the transferred power between them. Take then and rotate the polarization of one of the antennas. Plot the received power level in dB relative to the maximum, i.e., the reference level.

   b) Assume that the transmit antenna is circularly polarized. What is now the received power in dB relative to the maximum in *a)* when the transmit antenna is rotated around the pointing axis towards the receiving antenna?

   c) The level variation in *a)* is representative for the case when the polarization of the incident wave on the receiving antenna is random. Introduce now a 2-port antenna on the receiving side, and combine the two ports with the MRC algorithm. How does now the received power vary with rotation angle of the transmitting antenna, for the cases in *a)* and *b)*? Use the same reference level as before.

   d) Introduce now a 2-port antenna also on the transmitting side. The two ports excite orthogonal linear polarizations. Divide the transmitting power between the two ports. Use the MRC algorithm to combine all four channels. What is not the received power variation for the cases in *a)* and *c)*?

   e) Repeat *d)* when the polarizations of the two receiving antennas are not orthogonal, with an angular error of $\Delta\Phi_{\mathrm{pol}}$. How does now the received power vary?

   f) Repeat *d)* and *e)*, when the ZF algorithm is used, to generate two bitstreams.

3.13 References

[1] M. Bäckström, O. Lundén and P.-S. Kildal, "Reverberation chambers for EMC susceptibility and emission analyses", *Review of Radio Science 1999-2002*, pp. 429-452.

[2] William H. Tranter (Ed.), Brian D. Woerner (Ed.), Theodore S. Rappaport (Ed.) and Jeffrey H. Reed (Ed.), *Wireless Personal Communications: Channel Modeling and Systems Engineering*, Kluwer Academic Publisher, 2000.

[3] R. Vaughan and J.B. Andersen, *Channels, Propagation and antennas for Mobile Communications*, IEE Electromagnetic Waves Series 50, The Institution of Electrical Engineers, London 2003.

[4] T. Taga, "Analysis for mean effective gain of mobile antennas in land mobile radio environments", *IEEE Transactions on Vehicular Technology*, No. 39, pp. 117-131, May 1990.

[5] R.C. Hansen, *Microwave Scanning Antennas, Vol II: Array Theory and Practice*, Academic Press 1964.

[6] A.C. Ludwig, "Mutual coupling, gain and directivity of an array of two identical antennas", *IEEE Transactions on Antennas and Propagation*, pp. 837- 841, November 1976

[7] P.-S. Kildal and K. Rosengren, "Correlation and capacity of MIMO systems and mutual coupling, radiation efficiency and diversity gain of their antennas: Simulations and measurements in reverberation chamber", *IEEE Communications Magazine*, Vol. 42, No. 12, pp. 102-112, December 2004.

[8] K. Rosengren and P.-S. Kildal, "Radiation efficiency, correlation, diversity gain, and capacity of a six monopole antenna array for a MIMO system: Theory, simulation and measurement in reverberation chamber", Proceedings IEE, Microwave Antennas Propagation, Vol. 152, No. 1, pp. 7-16, February 2005, see also Erratum published in August 2006.

[9] P.-S. Kildal and K. Rosengren, "Electromagnetic analysis of effective and apparent diversity gain of two parallel dipoles", IEEE Antennas and Wireless Propagation Letters, Vol. 2, No. 1, pp. 9-13, 2003.

[10] P.W. Hannan, "The element-gain paradox for a phased-array antenna", IEEE Transactions on Antennas and Propagation, Vol. AP-12, No. 7, pp. 423-433, July 1964.

[11] S. Blanch, J. Romeu, I. Corbella, "Exact representation of antenna system diversity performance from input parameter description", *Electronics Letters*, Vol. 39, No. 9, pp. 705-707, May 2003.

[12] P.-S. Kildal, K. Rosengren, J. Byun and J. Lee, "Definition of effective diversity gain and how to measure it in a reverberation chamber", Microwave and Optical Technology Letters, Vol. 34, No. 1, pp. 56-59, July 5, 2002.

[13] M. Schwartz, W.R. Bennett and S. Stein, *Communication System and Techniques*, New York, McGraw-Hill, 1965.

[14] N. Jamaly, P.-S. Kildal and J. Carlsson, "Compact formulas for diversity gain of two-port antennas", IEEE Antenna and Wireless Propagation Letters, Vol. 9, pp. 970-973, November 2010.

[15] X. Chen, P.-S. Kildal, J. Carlsson and J. Yang, "MRC diversity and MIMO Capacity Evaluations of Multi-Port Antennas Using Reverberation Chamber and Anechoic Chamber," IEEE Transactions on Antennas and Propagation, Vol. 61, No. 2, pp. 917-926, February 2013.

[16] D. Tese and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press 2005.

[17] J.G. Kostas and B. Boverie, "Statistical model for a mode-stirred chamber", IEEE Transactions on Electromagnetic Compatibility, Vol. 33, No. 4, pp. 366-370, November 1991.

[18] D.A. Hill, M.T. Ma, A.R. Ondrejka, B.F. Riddle, M.L. Crawford and R.T. Johnk, "Aperture excitation of electrically large, lossy cavities", IEEE Transactions on Electromagnetic Compatibility, Vol. 36, No. 3, pp. 169-178, August 1994.

[19] D.A. Hill, "Linear dipole response in a reverberation chamber", IEEE Transactions on Electromagnetic Compatibility, Vol. 41, No. 4, pp. 365-368, November 1999.

[20] K. Rosengren, P.-S. Kildal, C. Carlsson and J. Carlsson, "Characterization of antennas for mobile and wireless terminals in reverberation chambers: Improved accuracy by platform stirring", Microwave and Optical Technology Letters, Vol. 30, No. 20, pp. 391-397, September 2001.

[21] P.-S. Kildal and C. Carlsson, "Detection of a polarization imbalance in reverberation chambers and how to remove it by polarization stirring when measuring antenna efficiencies", Microwave and Optical Technology Letters, Vol. 34, No. 2, pp. 145-149, July 20, 2002.

[22] D.A. Hill, "Electronic mode stirring for reverberation chambers",IEEE Transactions on Electromagnetic Compatibility, Vol. 36, No. 4, pp. 294-299, November 1994.

[23] P.-S. Kildal, C. Carlsson and J. Yang, "Measurement of free space impedances of small antennas in reverberation chambers", Microwave and Optical Technology Letters, Vol. 32, No. 2, pp. 112-115, January 2002.

[24] C.L. Holloway, D.A. Hill, J.M. Ladbury, P. Wilson, G. Koepke and J. Coder, "On the use of reverberation chambers to simulate a controllable Rician radio environment for the testing of wireless devices", to appear in a Special issue of IEEE Transactions on Antennas and Propagation, Autumn 2006.

[25] K. Rosengren and P.-S. Kildal, "Study of distributions of modes and plane waves in reverberation chambers for characterization of antennas in multipath environment", Microwave and Optical Technology Letters, Vol. 30, No. 20, pp. 386-391, September 2001.

[26] P.-S. Kildal, X. Chen, C. Orlenius, M. Franzén and C. Lötbäck Patané, "Characterization of reverberation chambers for OTA measurements of wireless devices: Physical formulations of channel matrix and new uncertainty formula", IEEE Transactions on Antennas and Propagation, Vol. 60, No. 8, pp. 3875-3891, August 2012.

[27] C. Orlenius, P.-S. Kildal, and G. Poilasne, "Measurements of total isotropic sensitivity and average fading sensitivity of CDMA phones in reverberation chamber", IEEE AP-S International Symposium, Washington D.C., 3-8 July 2005 (Orlenius is with Bluetest AB, Gothenburg, Sweden, Poilasne is with Kyocera Wireless Corp., San Diego, CA, USA).

[28] P.-S. Kildal, C. Orlenius and J. Carlsson, "OTA testing in multipath of antennas and wireless devices with MIMO and OFDM", Proceedings of the IEEE, Vol. 100, No. 7, pp. 2145-2157, July 2012.

[29] P.-S. Kildal, A. Hussain, X. Chen, C. Orlenius, A. Skårbratt, J. Åsberg, T. Svensson and T. Eriksson, "Threshold receiver model for throughput of wireless devices with MIMO and frequency diversity measured in reverberation chamber", IEEE Antennas and Propagation Wireless Letters, Vol. 10, pp. 1201-1204, 2011.

[30] X. Chen, P.-S. Kildal, C. Orlenius and J. Carlsson, "Channel sounding of loaded reverberation chamber for Over-the-Air testing of wireless devices - coherence bandwidth versus average mode bandwidth and delay spread", IEEE Antennas and Wireless Propagation Letters, Vol. 8, pp. 678-681, 2009. See also correction in Vol. 12, 2013.

[31] A. Hussain, P.-S. Kildal, "Study of OTA Throughput of LTE Terminals for Different System Bandwidths and Coherence Bandwidths", Proceedings of 7th European Conference on Antennas and Propagation (EuCAP 2013), Gothenburg, Sweden, April 2013.

[32] X. Chen, P.-S. Kildal, and M. Gustafsson, "Characterization of implemented algorithm for MIMO spatial multiplexing in reverberation chamber," IEEE Transactions on Antennas and Propagation, Vol. 61, No. 8, pp. 4400-4404, August 2013.

[33] X. Chen, B.T. Einarsson and P.-S. Kildal, "Improved MIMO Throughput with Inverse Power Allocation Study using USRP Measurement in Reverberation Chamber", *IEEE Antennas and Wireless Propagation Letters* (AWPL), Vol. 13, pp. 1494-1496.

[34] A. Paulraj, R. Nabar and D. Gore, *Introduction to space-time wireless communication*, Cambridge University Press, 2003.

[35] P.-S. Kildal, X. Chen, M. Gustafsson, Z. Shen, "MIMO Characterization on System Level of 5G Micro Base Stations Subject to Randomness in LOS", *IEEE Access*, Vol. 22, pp. 1062-1075, 18$^{\text{th}}$ September 2014.

[36] K. Fujimoto, H. Morishita, *Modern Small Antennas*, Cambridge University Press, 2014.

[37] A. Diallo, C. Luxey, P. Le Thuc, R. Staraj, G. Kossiavas, "Study and Reduction of the Mutual Coupling Between Two Mobile Phone PIFAs Operating in the DCS1800 and UMTS Bands", *IEEE Transactions on Antennas and Propagation*, Vol. 54, No. 11, pp. 3063-3074, November 2006.

[38] A. Chebihi, C. Luxey, A. Diallo, P. Le Thuc, R. Staraj, "A Novel Isolation Technique for Closely Spaced PIFAs for UMTS Mobile Phones", *IEEE Antennas and Wireless Propagation Letters*, Vol. 7, pp. 665-668, 2008.

[39] H. Li, B. K. Lau, Z. Ying, S. He, "Decoupling of Multiple Antennas in Terminals With Chassis Excitation Using Polarization Diversity, Angle Diversity and Current Control", *IEEE Transactions on Antennas and Propagation*, Vol. 60, No. 12, pp. 5947-5957, December 2012.

[40] A. A. Glazunov, A. F. Molisch, and F. Tufvesson, "Mean effective gain of antennas in a wireless channel," *IET Microwaves, Antennas and Propagation*, Vol. 3, No. 2, pp. 214–227, 2009.

Chapter 4

The theory of radiation from current sources

The purpose of this chapter is to introduce Maxwell's equations, which govern radiation, scattering and propagation of electromagnetic fields. We will also present some general concepts which facilitate the use of these equations such as the equivalence and imaging principles. We will emphasize the integral form of Maxwell's equations from which radiation fields from known current distributions can be directly calculated. The integrands are expressed in terms of vector dyadic *Green's functions* for sources in free space, which make the notation more compact and easier to interpret. The equivalence principle allows us to introduce equivalent sources, making the vector dyadic Green's functions even more usable. The term Green's function is commonly used to denote a point source response.

In most of the calculations in this book, we treat antennas under the presumption that we know approximately the physical or equivalent current distributions. Therefore, we can use the vector dyadic Green's functions to find the radiation fields by direct analytical or numerical integration over these currents. For the antennas in this book, this approach is quite accurate if the dimensions are within the limits where the approximate current distributions are valid.

Advanced antenna design involves calculation of the actual current or field distributions. This is normally done numerically (e.g., by the *Method of Moments*). We will in the last section of this chapter introduce the basics of the Method of Moments. This will be useful in later chapters to derive formulas for the self-impedances of dipoles, slots and microstrip antennas using only one or two expansion functions. These formulas lead to the classical impedance formulas for dipoles and slots.

4.1 Maxwell's equations

We introduce *Maxwell's equations* for time-harmonic fields and the standard boundary conditions that apply to these fields at material interfaces. We also describe the related "soft" and "hard" boundary conditions that are commonly used in acoustics and in diffraction theory.

They facilitate the interpretation of electromagnetic field behavior in the vicinity of some material boundaries. Furthermore, we introduce the electric and magnetic vector potentials and the related integral forms of Maxwell's equations. These vector potentials are commonly used in the antenna theory, and therefore must be studied. However, we will not use them in this book. Instead we introduce the alternative direct vector integral forms of the E- and H-fields in Section 4.2, and continue using these in the rest of the book. These forms are often referred to as dyadic Green's functions and are derived in Appendix A.

4.1.1 Differential form

The original Maxwell's equations contain time-differentiation of the electric and magnetic fields. However, for linear materials and harmonic time variation, these differentiations can be replaced by the factor $j\omega$. This gives the following time-harmonic form of the four basic Maxwell's equations for single-valued, smooth, bounded fields:

$$\nabla \times \mathbf{E} = -j\omega\mathbf{B} - \mathbf{M} \ , \tag{4.1}$$

$$\nabla \times \mathbf{H} = j\omega\mathbf{D} + \mathbf{J} \ , \tag{4.2}$$

$$\nabla \cdot \mathbf{D} = q_e \ , \tag{4.3}$$

$$\nabla \cdot \mathbf{B} = q_m \ , \tag{4.4}$$

where $\mathbf{J}$ and $\mathbf{M}$ are associated electric and magnetic source current densities representing the sources of the E-, D-, H- and B-fields, and q_e and q_m are the electric and magnetic charge densities. These basic equations contain four different fields; the electric field E, most simply referred to as the E-field, the magnetic field H referred to as the H-field, and two proportional D- and B-fields, respectively, that only play a role in specific materials. Therefore, we will not treat D- and B-fields explicitly in this text, but only implicitly via their relation to the E- and H-fields in (4.6) and (4.7) below.

The so-called charge *continuity equation* can be derived from (4.2) and (4.3) to be

$$\nabla \cdot \mathbf{J} = -j\omega q_e \ , \tag{4.5}$$

and correspondingly for magnetic currents and charge densities. The magnetic source currents and charges do not exist in reality, and they were not included in the original Maxwell's equations. But modern field calculations are considerably simplified by using equivalent magnetic currents.

The D- and B-fields are functions of the E- and H-fields. For fields in linear and isotropic media, they are simply proportional to them:

$$\mathbf{D} = \varepsilon\mathbf{E} \ , \tag{4.6}$$

$$\mathbf{B} = \mu\mathbf{H} \ , \tag{4.7}$$

where ε is the *permittivity* and μ is the *permeability* of the medium at the point where fields are being considered. It is most common to work with the relative material parameters which are defined by the relations

$$\varepsilon = \varepsilon_r\varepsilon_0 \ , \qquad \varepsilon_r = \varepsilon_r' - j\varepsilon_r'' \ , \tag{4.8}$$

$$\mu = \mu_r \mu_0 \ , \qquad \mu_r = \mu_r' - j\mu_r'' \ , \tag{4.9}$$

where ε_r is the *relative permittivity* and ε_0 is the permittivity in free space, and where μ_r is the *relative permeability* and μ_0 is the permeability in free space[1]. For lossy media the ε_r and μ_r are complex-valued, with the imaginary parts ε_r'' and μ_r'' representing losses.

In conducting materials, the following relation is valid inside the material,

$$\mathbf{J} = \sigma \mathbf{E} \ , \tag{4.10}$$

where σ is the *conductivity* of the medium in Siemens/meter (S/m).[2] In the time-harmonic form of Maxwell's equations (see (4.2)), it is possible to represent σ as an imaginary part ε_r'' of ε_r, and visa versa, by using

$$\sigma = \omega \varepsilon_0 \varepsilon_r'' \ , \qquad \text{or alternatively} \qquad \sigma = \varepsilon_r' \cdot \tan\delta/(60\lambda) \ ,$$

where λ is the free space wavelength defined on page 131, and $\tan\delta = \varepsilon_r''/\varepsilon_r'$ is called the loss tangent or loss factor. The latter expression of σ is obtained from the first one by using (4.29)[3]. In this book we will only treat *Perfect Electric Conductors* (*PEC*s) for which $\sigma \to \infty$. Most conductors can be approximated as PECs in antenna problems provided they are thicker than the penetration depth of the fields for the conducting material. The *penetration depth* (also called *skin depth*) can be derived for plane wave incidence on planar interfaces of good conductors [1] to be

$$\tau = \sqrt{\frac{\lambda}{\pi \eta \sigma \mu_r}} \ , \tag{4.11}$$

where $\eta = \sqrt{\mu_0/\varepsilon_0} = 377\,\Omega$ is the free space wave impedance[4].

The material parameters usually have different values in different frequency bands (i.e., they exhibit discontinuities). They can be found tabulated in handbooks (e.g., in [2]). Some sample values are given in the tables in Appendix D. These tables also include corresponding propagation losses or penetration depths at different frequencies. Note that the D- and B-fields are of no importance in antenna analysis due to the simple material relations. We do not use them in this book.

4.1.2 Standard boundary conditions

The *boundary conditions* that apply to the electromagnetic fields on material interfaces can be derived from Maxwell's equations. Here we simply state them. We assume that

$$\hat{\mathbf{n}} \times (\mathbf{E}_2 - \mathbf{E}_1) = -\mathbf{M}_S \ , \qquad \hat{\mathbf{n}} \times (\mathbf{H}_2 - \mathbf{H}_1) = \mathbf{J}_S \ , \tag{4.12}$$

where $\mathbf{E}_1$ and $\mathbf{H}_1$ are the fields in region 1, $\mathbf{E}_2$ and $\mathbf{H}_2$ are the fields in region 2, $\hat{\mathbf{n}}$ is the normal vector to the interface pointing into region 2, and $\mathbf{J}_S$ and $\mathbf{M}_S$ are the electric

[1] Note that we never need the values of ε_0 and μ_0 explicitly, because they will always in this text enter the equations via η and c, see (4.29).

[2] Siemens $= 1/\Omega$.

[3] Note that $\eta = 377\Omega \approx 120\pi\Omega$, which is a convenient form to use.

[4] Note that $\eta = 377\Omega \approx 120\pi\Omega$, which is a convenient form to use.

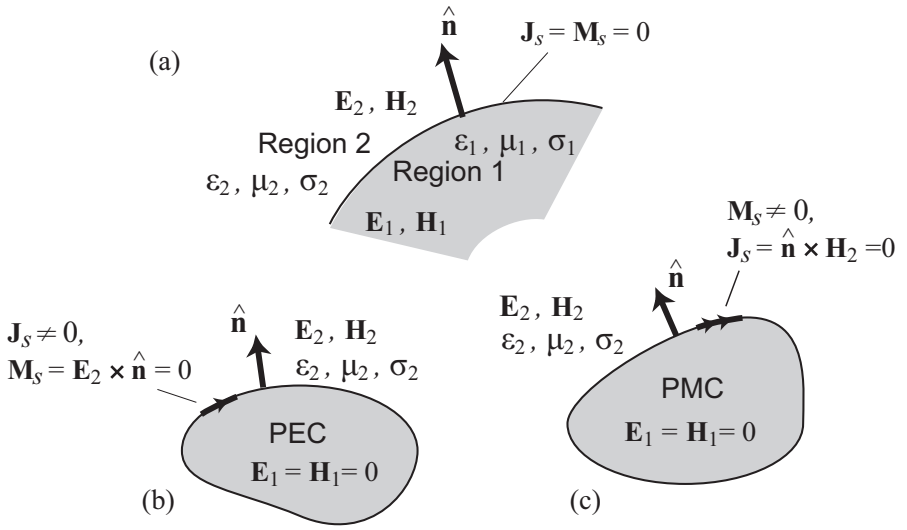


Figure 4.1: The boundary conditions between two regions filled with different materials, (a) Two dielectric regions, (b) PEC in region 1, (c) PMC in region 1.

and magnetic surface current densities at the boundary between these regions (see Fig. 4.1). These boundary conditions are written in vector form. What they actually state is that if the tangential components of $\mathbf{E}$ and $\mathbf{H}$ are discontinuous at a surface, the discontinuity must be equal to the surface current densities $\mathbf{M}_S$ and $\mathbf{J}_S$, respectively.

At the surfaces of *dielectric and magnetic materials* there can be no physical currents, so we get

$$\begin{array}{l} \hat{\mathbf{n}} \times (\mathbf{E}_2 - \mathbf{E}_1) = 0 \\ \hat{\mathbf{n}} \times (\mathbf{H}_2 - \mathbf{H}_1) = 0 \end{array} \quad \text{i.e.,} \quad \begin{array}{l} (\mathbf{E}_2 - \mathbf{E}_1)_{\tan} = 0 \\ (\mathbf{H}_2 - \mathbf{H}_1)_{\tan} = 0 \ . \end{array} \tag{4.13}$$

These conditions mean that the tangential E- and H-field components are continuous. Moreover, inside a *PEC* all fields are zero, and at its surface there can only exist electric currents. Therefore, if region 1 is a PEC, we get from (4.12)

$$\hat{\mathbf{n}} \times \mathbf{E}_2 = 0 \ , \qquad \hat{\mathbf{n}} \times \mathbf{H}_2 = \mathbf{J}_S \ , \tag{4.14}$$

which means that the tangential E-field is zero at the PEC, and that the tangential H-field equals the surface current density. Magnetic conductors do not exist naturally, but sometimes they are convenient tools in electromagnetic analysis. Therefore, it is important to know that the boundary conditions on a *hypothetical Perfect Magnetic Conductor* (*PMC*) are

$$\hat{\mathbf{n}} \times \mathbf{E}_2 = -\mathbf{M}_S \ , \qquad \hat{\mathbf{n}} \times \mathbf{H}_2 = 0 \ , \tag{4.15}$$

which means that the tangential H-field is zero. Inside PMCs all fields are zero, in the same way as for PECs.

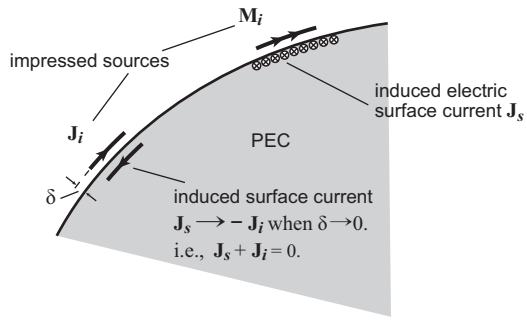


Figure 4.2: Short-circuited nonradiating electric current source on PEC, and radiating magnetic current source on the same PEC.

4.1.3 Impressed current sources on PECs

Consider an impressed incremental electric current source $\mathbf{J}_i = I_0\delta(\mathbf{r}-\mathbf{r}_S)\hat{\mathbf{t}}$ located tangentially at a point $\mathbf{r}_S$ at the surface of a PEC of arbitrary shape[5], see Fig. 4.2. This will induce electric currents on the PEC in such a way that the boundary conditions of the PEC are satisfied, i.e., $\hat{\mathbf{n}} \times \mathbf{E} = 0$ at the surface and $\mathbf{E} = \mathbf{H} = 0$ inside the PEC.

Let us assume that the induced electric current is given by

$$\mathbf{J}_S = -\mathbf{J}_i = -I_0\delta(\mathbf{r}-\mathbf{r}_S)\hat{\mathbf{t}} \ . \tag{4.16}$$

This makes the total current

$$\mathbf{J}_{\text{tot}} = \mathbf{J}_i + \mathbf{J}_S = 0 \ , \tag{4.17}$$

which means that the total field is zero as well, because there is no source for the fields. This clearly satisfies the boundary conditions. Therefore, according to the uniqueness theorem[6], the solution to the field problem is the null-field. Note that the above argumentation is valid for PECs of any shape, and it has nothing to do with imaging[7].

We often express (4.17) as the impressed electric source current being short-circuited when located tangentially at a PEC. This expression probably came from practical half-wave electric dipole antennas. When they are located on a ground plane in direct metal contact with it, the voltage over the feed gap is short-circuited and no radiating currents are induced.

Consider now a magnetic source current $\mathbf{M}_i = M_0\delta(\mathbf{r}-\mathbf{r}_S)\hat{\mathbf{t}}$ located tangentially on the surface of the same PEC. This will also induce electric currents on the PEC in a way that the boundary conditions are satisfied. This situation presents a difficult field problem that normally must be solved numerically, except for some special cases. Such a special case exists when the PEC is a plane, in which case an analytical solution exist. For the plane PEC case, the induced electric currents that make the total field satisfy the boundary condition are $\mathbf{J}_S = \hat{\mathbf{n}} \times \mathbf{H}_t$ where $\mathbf{H}_t$ is the H-field caused by a magnetic current source, $\mathbf{M}_t = 2M_0\delta(\mathbf{r}-\mathbf{r}_S)\hat{\mathbf{t}}$, radiating into homogeneous space (i.e., the H-field caused by $\mathbf{M}_i$ and its image $\mathbf{M}_{\text{img}} = \mathbf{M}_i$)[8]. Note that the E- and H-fields never vanish when a tangential magnetic current is located on a

[5] $\delta(\mathbf{r}) = \delta(x,y,z)$ is the three-dimensional delta function.
[6] The uniqueness theorem can be found in Section 4.3 on page 136.
[7] Imaging can be found in Section 4.6 on page 150.
[8] For more details about imaging see again Section 4.6 on page 150.

PEC, because only electric currents can be induced at the PEC surface. There is no way to add electric and magnetic currents, which flow in the same plane, in such a way that the resulting total field is zero.

4.1.4 Soft and hard boundary conditions

Soft and hard boundary conditions are commonly used to separate two different polarization cases in edge diffraction theory. The terminology itself comes from acoustics (see e.g., [3]), where the acoustic pressure p is zero at the *soft surface*, i.e.,

$$p = 0 \ .$$

Furthermore, the derivative of p normal to a *hard surface* is zero, i.e.,

$$\frac{\partial p}{\partial n} = 0 \ .$$

The latter condition states that the acoustic pressure has a maximum or minimum at the hard surface. The nomenclature comes from the fact that one feels respectively soft or hard when touching these acoustic materials. Both cause total reflection of an acoustic wave.

The soft and hard boundary conditions appear also in electromagnetics. To illustrate this, we study a two-dimensional (2D) PEC structure oriented along the z-axis, and we illuminate it by a plane wave incident normal to the structure (see Fig. 4.3). If the plane wave is polarized in z-direction, i.e., the TM_z (transverse magnetic to z) or H-plane case, the boundary condition in (4.14) gives

$$E_z = 0 \tag{4.18}$$

at the surface of the PEC, which is a soft boundary condition. Consider also a plane wave that is polarized with the H-field in z-direction, i.e., the TE_z (transverse electric to z) or E-plane case. Then, it is possible to show that $\hat{\mathbf{n}} \times \mathbf{E} = 0$ corresponds to

$$\frac{\partial E_n}{\partial n} = 0 \ ; \qquad E_n = \hat{\mathbf{n}} \cdot \mathbf{E} = \sqrt{E_x^2 + E_y^2} \tag{4.19}$$

at the surface of the PEC (see Fig. 4.3), where E_n is the component of the E-field which is normal to the surface of the PEC, and $\partial E_n / \partial n$ means the derivative of E_n normal to the surface. This is a hard boundary condition. Thus, the boundary conditions of the PEC cylinder are soft for the TM_z case and hard for the TE_z case. These two boundary conditions appear also approximately in respectively the H- and E-planes of metallic antenna structures with two planes of symmetry.

The soft boundary condition is readily understood to stop the wave propagation along the surface as it makes the power density vector $(\mathbf{E} \times \mathbf{H}^*)/2$ zero at and along the surface, whereas the hard boundary condition enhances wave propagation along the surface. Both these characteristics are clearly observed in the contour plots of the fields around thin metal cylinders with circular and rhombic cross section in Fig. 4.3. When the cylinder is removed there would only be a plane wave, i.e., a homogenous level of 0 dB. Therefore, the ripples on the contour plots are due to the scattered field from the cylinder. The shadow behind the cylinder is much smaller for the TE_z than for the TM_z case. Also, for the rhombic

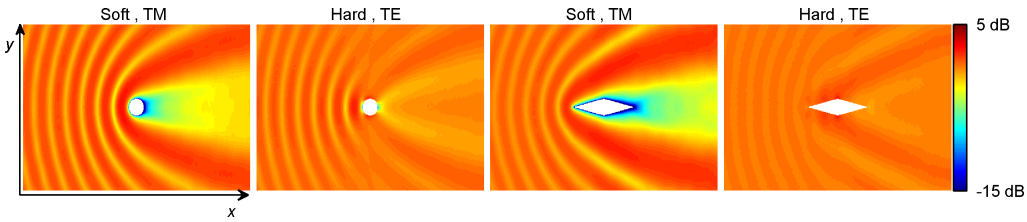


Figure 4.3: Contour plots of total E-field in the cross section around a PEC cylinder when there is an incident plane wave from the left. The two left-most graphs show a cylinder with circular cross section for TM_z (H-plane) and TE_z (E-plane) cases. The two right-most graphs show the same for a cylinder with rhombic cross section. The soft boundary condition appears for TM_z polarization, and the hard boundary condition for TE_z polarization. Both cylinders have a width of half wavelength in y-direction. For larger widths there gradually becomes a shadow also for the TE_z case.

cylinder the shadow is seen to vanish for the hard TE_z case. This will happen for all narrow cross-sections.

There exist ways to create polarization independent *soft surfaces* for electromagnetic waves, e.g., by providing a metal conductor with corrugations. This can be used to make the boundary conditions equal in the E- and H-planes of antennas with two planes of symmetry, and, in particular of rotationally symmetric BOR_1 *antennas*. Corrugated circular horn antennas are considered in Chapter 7 of this book. The polarization-independent *hard surfaces* can be realized by dielectric-filled corrugations, or by metal strips on a dielectric coating. This will reduce the shadow of cylinders with oblong cross sections for any polarization of a wave with known angle of arrival [4].

Once more it must be emphasized that the standard boundary conditions[9] must always be used in all accurate field analyses. In contrast, the soft and hard boundary conditions in (4.18) - (4.19) are most convenient for interpreting field behavior. Such interpretations are in particular important during the initial design of antennas and during problem solving. The boundary conditions in (4.18) - (4.19) are only ideally equivalent to the standard boundary conditions for some special cases, such as for 2D problems. The terms soft and hard boundary conditions are commonly used in diffraction theory [5] where there exists different soft and hard diffraction coefficients for the two cases. The terms can also be used to define artificial soft and hard surfaces which have soft and hard boundary conditions for any polarization [6]. The *soft surface* can be realized by corrugations[10].

The small scattering from a wire for the TE case is used in wire grids to create polarization dependent reflectors. If thin z-directed wires are located close together without touching each other, they will reflect the TM_z polarization, and the TE_z polarization will be transmitted with very low transmission loss.

4.1.5 Auxiliary vector potentials

It is common to introduce a *magnetic vector potential* $\mathbf{A}$ and an *electric vector potential* $\mathbf{F}$ in order to facilitate the solution of Maxwell's equation for sources located in homogeneous

[9] Standard boundary conditions can be found in Section 4.1.2 on page 125.
[10] See Section 8.5 on page 279.

space, see e.g., [1]. Neither $\mathbf{A}$ nor $\mathbf{F}$ have physical interpretations, and they cannot be measured. What makes them attractive is that they are solutions to the inhomogenous Helmholz equations:

$$\nabla^2\mathbf{A}(\mathbf{r}) + k^2\mathbf{A}(\mathbf{r}) = -\mu\mathbf{J}(\mathbf{r}') \ , \tag{4.20}$$

$$\nabla^2\mathbf{F}(\mathbf{r}) + k^2\mathbf{F}(\mathbf{r}) = -\varepsilon\mathbf{M}(\mathbf{r}') \ . \tag{4.21}$$

When $\mathbf{J}(\mathbf{r}')$ and $\mathbf{M}(\mathbf{r}')$ are surface current distributions, the solutions to these two equations in homogeneous space are

$$\mathbf{A}(\mathbf{r}) = \mu \iint_{A_c} \mathbf{J}(\mathbf{r}')\Psi(|\mathbf{r}-\mathbf{r}'|)dS' \ , \tag{4.22}$$

$$\mathbf{F}(\mathbf{r}) = \varepsilon \iint_{A_m} \mathbf{M}(\mathbf{r}')\Psi(|\mathbf{r}-\mathbf{r}'|)dS' \ , \tag{4.23}$$

where the scalar Green's function in these equations is

$$\Psi(R) = \frac{1}{4\pi R}e^{-jkR} \ ; \qquad R = |\mathbf{r}-\mathbf{r}'| \tag{4.24}$$

where k is the wavenumber. The integrals are taken over the surfaces A_e, containing the source currents $\mathbf{J}(\mathbf{r}')$, and A_m containing $\mathbf{M}(\mathbf{r}')$. The E- and H-fields resulting from $\mathbf{A}$ and $\mathbf{F}$ can be derived from Maxwell's equations and known vector operations. They are given by

$$\mathbf{E}(\mathbf{r}) = -j\omega\mathbf{A}(\mathbf{r}) - j\frac{1}{\omega\mu\varepsilon}\nabla(\nabla\cdot\mathbf{A}(\mathbf{r})) - \frac{1}{\varepsilon}\nabla\times\mathbf{F}(\mathbf{r}) \ , \tag{4.25}$$

$$\mathbf{H}(\mathbf{r}) = -j\omega\mathbf{F}(\mathbf{r}) - j\frac{1}{\omega\mu\varepsilon}\nabla(\nabla\cdot\mathbf{F}(\mathbf{r})) + \frac{1}{\mu}\nabla\times\mathbf{A}(\mathbf{r}) \ . \tag{4.26}$$

When $|\mathbf{r}-\mathbf{r}'|$ is large, i.e., in the far-field region, we have

$$\mathbf{E}(\mathbf{r}) = -j\omega[\mathbf{A}(\mathbf{r}) - (\mathbf{A}(\mathbf{r})\cdot\hat{\mathbf{r}})\hat{\mathbf{r}}] + j\omega\eta\hat{\mathbf{r}}\times\mathbf{F}(\mathbf{r}) \ , \tag{4.27}$$

$$\mathbf{H}(\mathbf{r}) = -j\omega[\mathbf{F}(\mathbf{r}) - (\mathbf{F}(\mathbf{r})\cdot\hat{\mathbf{r}})\hat{\mathbf{r}}] - j\frac{\omega}{\eta}\hat{\mathbf{r}}\times\mathbf{A}(\mathbf{r}) \ . \tag{4.28}$$

It is very common to use the vector potentials $\mathbf{A}$ and $\mathbf{F}$ and (4.25) - (4.28) in antenna analyses. The first two of these equations are very laborious to use as they contain several differential operations. The far-field versions are easier to use. The first version also contains the inconvenient parameters ω, μ and ε as separate constants. If Eq. (4.22) - (4.24) are inserted for $\mathbf{A}$ and $\mathbf{F}$ in (4.27) and (4.28), the ω, μ and ε can be removed from the equations and replaced by more practical parameters such as the wavenumber k and the wave impedance η. This removal of ω, μ and ε can be done by using the following *simplification relations*

$$k = \omega\sqrt{\mu\varepsilon} \ , \qquad \eta = \sqrt{\frac{\mu}{\varepsilon}} \ , \qquad \omega\mu = k\eta \ , \qquad \omega\varepsilon = k/\eta \ . \tag{4.29}$$

The most complicated second term in (4.25) can also be expressed in terms of a scalar electric potential ϕ, which is defined by[11]

$$\begin{aligned} \phi_e(\mathbf{r}) &= \frac{1}{\varepsilon}\iint_{A_c} q_e(\mathbf{r}')\Psi(|\mathbf{r}-\mathbf{r}'|)dS' \ ; \\ q_e(\mathbf{r}') &= -\frac{1}{j\omega}\nabla\cdot\mathbf{J}(\mathbf{r}') \ , \end{aligned} \tag{4.30}$$

[11] See e.g., [7, Sec. 4-1].

where q_e is the electric charge density[12]. Then, (4.25) becomes

$$\mathbf{E}(\mathbf{r}) = -j\omega\mathbf{A} - \nabla\phi_e(\mathbf{r}) - \frac{1}{\varepsilon}\nabla \times \mathbf{F}(\mathbf{r}) \ . \tag{4.31}$$

Correspondingly, we can define magnetic charge by

$$q_m(\mathbf{r}') = -\frac{1}{j\omega}\nabla \cdot \mathbf{M}(\mathbf{r}')$$

and (4.26) becomes

$$\mathbf{E}(\mathbf{r}) = -j\omega\mathbf{F}(\mathbf{r}) - \nabla\phi_m(\mathbf{r}) + \frac{1}{\eta}\nabla \times \mathbf{A}(\mathbf{r}) \ , \tag{4.32}$$

$$\phi_m(\mathbf{r}) = \frac{1}{\eta}\iint_{A_m} q_m(\mathbf{r}')\Psi(|\mathbf{r}-\mathbf{r}'|)dS' \ . \tag{4.33}$$

In this book we will choose a more convenient way to calculate the E-and H-fields resulting from known electric and magnetic currents. Through these currents, the E- and H-fields can be obtained directly by integration both in the near and far-field regions. No differentiation is needed. These relations are most commonly known as vector dyadic Green's functions and will be given in their vector integral forms in the next section.

The vector potential formulations have their advantages when dealing with reactive near-fields, such as when computing radiation reactances and susceptances. They are also convenient when deriving integral equations needed for calculating the current distributions accurately by numerical methods. We will not derive such integral equations in this book. However, we will in Section 4.7.1 introduce a numerical algorithm based on (4.32), which we later use for impedance calculations.

4.2 Vector integral forms of the E- and H-fields

The advantage of using the auxiliary vector potentials $\mathbf{A}$ and $\mathbf{F}$ is that they have a scalar Green's function. However, as described at the end of the previous section, the equations by which the E- and H-fields can be calculated from $\mathbf{A}$ and $\mathbf{F}$ are complicated and not convenient. The expressions to follow do not suffer from these disadvantages. They are derived in Appendix A by performing the differential operations in (4.25) - (4.26) on the scalar Green's function instead of on $\mathbf{A}$ and $\mathbf{F}$, and by exchanging the order of integration and differentiation.

We choose in the next subsection to express the vector integral forms for the E- and H-fields in terms of the wavenumber k and the wave impedance η, and not making use of ω, μ and ε. The wavenumber is convenient in equations, whereas its numerical value makes little sense. We instead use the wavelength $\lambda = 2\pi/k$ when presenting practical numerical values. The *wavelength* and the wavenumber depend on the medium which in this book mainly is free space, for which the wavelength λ becomes

$$\lambda = \frac{c}{f} \ , \qquad \text{i.e.,} \qquad \lambda[mm] = 300/f[GHz] \ ,$$

[12] See the continuity equation in (4.5).

where $c = 1/\sqrt{\mu_0\varepsilon_0} = 2.99790 \times 10^8\,\mathrm{m/s} \approx 300\,\mathrm{mm/ns}$ is the *phase velocity in free space* and f is the frequency at which the calculations are done. This frequency is related to the angular frequency ω by $f = \omega/2\pi$. The *wave impedance* in free space is $\eta = \sqrt{\mu_0/\varepsilon_0} = 377\,\Omega$.

4.2.1 General expressions

Let us consider a field problem with known electric and magnetic surface current distributions located at the surface S' described by the primed position vector $\mathbf{r}'$, and everywhere else there is free space. Further, we introduce an observation point $\mathbf{r}$ and

$$\mathbf{R} = \mathbf{r} - \mathbf{r}' \ , \qquad R = |\mathbf{r} - \mathbf{r}'| \ , \qquad \hat{\mathbf{R}} = \mathbf{R}/R \ . \tag{4.34}$$

Then, from Appendix A, we see that the E- and H-fields in the observation point can be calculated in this way

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_{J}(\mathbf{r}) + \mathbf{E}_{M}(\mathbf{r}) \ , \tag{4.35}$$

$$\mathbf{H}(\mathbf{r}) = \mathbf{H}_{J}(\mathbf{r}) + \mathbf{H}_{M}(\mathbf{r}) \ , \tag{4.36}$$

where
$$\mathbf{E}_{J} = C_k \iint_{S'} [\eta\mathbf{J} C_{N_1} - (\eta\mathbf{J}\cdot\hat{\mathbf{R}})\hat{\mathbf{R}} C_{N_2}] \frac{1}{R} e^{-jkR} dS' \ , \tag{4.37}$$

$$\mathbf{H}_{J} = -\frac{1}{\eta} C_k \iint_{S'} (\eta\mathbf{J}\times\hat{\mathbf{R}}) C_N \frac{1}{R} e^{-jkR} dS' \ , \tag{4.38}$$

$$\mathbf{E}_{M} = C_k \iint_{S'} (\mathbf{M}\times\hat{\mathbf{R}}) C_N \frac{1}{R} e^{-jkR} dS' \ , \tag{4.39}$$

$$\mathbf{H}_{M} = \frac{1}{\eta} C_k \iint_{S'} [\mathbf{M} C_{N_1} - (\mathbf{M}\cdot\hat{\mathbf{R}})\hat{\mathbf{R}} C_{N_2}] \frac{1}{R} e^{-jkR} dS' \ , \tag{4.40}$$

$$\text{with} \qquad C_k = -jk/(4\pi) \qquad \text{and,} \tag{4.41}$$

$$C_N = 1 + \frac{1}{jkR} \ , \quad C_{N_1} = 1 + \frac{1}{jkR} - \frac{1}{(kR)^2} \ , \quad C_{N_2} = 1 + \frac{3}{jkR} - \frac{3}{(kR)^2} \ . \tag{4.42}$$

In this book C_k is referred to as the *incremental source constant*, and C_N, C_{N_1} and C_{N_2} as near-field functions, see also (4.43) below. For simplicity we have suppressed the arguments of the different quantities in (4.37)-(4.42)[13]. The expression

$$e^{-jkR}/R$$

is the same scalar Green's function (except for a constant) that is used in the expressions for the vector potentials $\mathbf{A}$ and $\mathbf{F}$. Here, it appears as part of the vector dyadic Green's functions which are used to calculate $\mathbf{H}$ from $\mathbf{J}$, $\mathbf{E}$ from $\mathbf{J}$, $\mathbf{H}$ from $\mathbf{M}$, and $\mathbf{E}$ from $\mathbf{M}$ in free space. In this book we have not written out these Green's functions explicitly, as this is not necessary, but instead we work with the complete vector integrals in the forms given above. Again, note that the above equations are derived from the original Maxwell's equations for sources in free space without making any approximations.

[13] We have written $\mathbf{H}_M$ instead of $\mathbf{H}_M(\mathbf{r})$, $\mathbf{J}$ instead of $\mathbf{J}(\mathbf{r}')$, $\mathbf{R}$ instead of $\mathbf{R}(\mathbf{r},\mathbf{r}')$, C_N instead of $C_N(\mathbf{r},\mathbf{r}')$ and so on.

The above functions C_{N}, C_{N_1} and C_{N_2} are conveniently referred to as *near-field functions*, because

$$C_{N} \approx 1 \ , \quad C_{N_1} \approx 1 \quad \text{and} \quad C_{N_2} \approx 1 \ , \quad \text{when } kR \gg 1 \ . \tag{4.43}$$

This condition appears already when $R \gg 1/k = \lambda/2\pi$, so unless the observation point is very close to or at the current sources we may always use $C_{N} = C_{N_1} = C_{N_2} = 1$. Under these approximations the integrands representing the Green's functions have no radial components in either E- or H-fields. Actually, $\mathbf{E}$ and $\mathbf{H}$ from incremental current sources are orthogonal both to each other and to $\mathbf{R}$ when the approximation in (4.43) is valid.

4.2.2 Radiating far-field expressions

In the far-field we can use the *Fraunhofer approximation* which was introduced and discussed in Section 2.3.4, i.e.,

$$\hat{\mathbf{R}} = \hat{\mathbf{r}} \qquad \text{and} \qquad R = |\mathbf{r} - \mathbf{r}'| = |\mathbf{r}| = r \tag{4.44}$$

in amplitude expressions, and

$$kR = k|\mathbf{r} - \mathbf{r}'| = k(r - (\mathbf{r}' \cdot \hat{\mathbf{r}})) \tag{4.45}$$

in phase expressions. These approximations can be understood by studying Fig. 2.8. In order to see under which conditions they are valid, we will derive the latter. This is most easily done by splitting $\mathbf{r}'$ in terms of its $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\varphi}}$ components and expanding R in a power series valid for small $\mathbf{r}'$, as follows (see Appendix B):

$$\mathbf{r}' = (\mathbf{r}' \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} + (\mathbf{r}' \cdot \hat{\boldsymbol{\theta}})\hat{\boldsymbol{\theta}} + (\mathbf{r}' \cdot \hat{\boldsymbol{\varphi}})\hat{\boldsymbol{\varphi}} \ ,$$

$$\begin{aligned} R^2 = |\mathbf{r} - \mathbf{r}'|^2 &= (r - (\mathbf{r}' \cdot \hat{\mathbf{r}}))^2 + (\mathbf{r}' \cdot \hat{\boldsymbol{\theta}})^2 + (\mathbf{r}' \cdot \hat{\boldsymbol{\varphi}})^2 \\ &= r^2 - 2(\mathbf{r}' \cdot \hat{\mathbf{r}}) + (r')^2 \ , \end{aligned}$$

$$\begin{aligned} R &= r\sqrt{1 - \frac{2}{r}(\mathbf{r}' \cdot \hat{\mathbf{r}}) + \left(\frac{r'}{r}\right)^2} \\ &\approx r\left(1 - \frac{1}{r}(\mathbf{r}' \cdot \hat{\mathbf{r}}) + \frac{1}{2}\left(\frac{r'}{r}\right)^2 - \frac{1}{8}\left(\frac{2}{r}(\mathbf{r}' \cdot \hat{\mathbf{r}})\right)^2\right) \ . \end{aligned}$$

We see that the Fraunhofer approximation gives a phase error

$$\Delta\phi = kR - k(r - (\mathbf{r}' \cdot \hat{\mathbf{r}})) \approx \frac{k}{2r}(r')^2 \ . \tag{4.46}$$

Thus, it is valid with a maximum phase error $\Delta\phi < \pi/8$ when

$$r > \frac{4k}{\pi}(r'_{\max})^2 = 8(r'_{\max})^2/\lambda = 2D^2/\lambda \ , \tag{4.47}$$

where $r'_{\max}$ is the maximum length of $\mathbf{r}'$ and $D = 2r'_{\max}$ is the diameter of the smallest sphere which encloses the antenna. Eq. (4.47) defines the *far-field region* as already discussed in Section 2.3.1.

The Fraunhofer approximation gives the following E-field:

$$\mathbf{E}(\mathbf{r}) = \frac{1}{r}e^{-jkr}\{\mathbf{G}_{J}(\hat{\mathbf{r}}) + \mathbf{G}_{M}(\hat{\mathbf{r}})\} \tag{4.48}$$

$$\text{with} \qquad \mathbf{G}_{J}(\hat{\mathbf{r}}) = \mathbf{I}_{J} - (\mathbf{I}_{J} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} \ , \tag{4.49}$$

$$\mathbf{G}_{M}(\hat{\mathbf{r}}) = \mathbf{I}_{M} \times \hat{\mathbf{r}} \ . \tag{4.50}$$

The above $\mathbf{I}_{J}$ and $\mathbf{I}_{M}$ are the electric and magnetic *radiation integrals*

$$\mathbf{I}_{J} = C_k \iint\limits_{S'} \eta\mathbf{J}(\mathbf{r}')e^{jk(\mathbf{r}'\cdot\hat{\mathbf{r}})}dS' \ , \tag{4.51}$$

$$\mathbf{I}_{M} = C_k \iint\limits_{S'} \mathbf{M}(\mathbf{r}')e^{jk(\mathbf{r}'\cdot\hat{\mathbf{r}})}dS' \ , \tag{4.52}$$

where $C_k = -jk/(4\pi)$ is the convenient *incremental source constant* introduced in (4.41). The H-field becomes

$$\mathbf{H}(\mathbf{r}) = \frac{1}{\eta}\hat{\mathbf{r}} \times \mathbf{E} = \frac{1}{\eta r}e^{-jkr}\{\hat{\mathbf{r}} \times \mathbf{I}_{J} + [\mathbf{I}_{M} - (\mathbf{I}_{M} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}]\} \ . \tag{4.53}$$

The radiation integrals $\mathbf{I}_{J}$ and $\mathbf{I}_{M}$ are taken over the distribution of electric $\mathbf{J}$ and magnetic $\mathbf{M}$ surface currents. For line currents $\mathbf{J}_l$ and $\mathbf{M}_l$ along a curve L, the integrals in (4.51) - (4.52) reduce to the line integrals

$$\mathbf{I}_{J} = C_k \int_{L} \eta\mathbf{J}_l(\mathbf{r}')e^{jk(\mathbf{r}'\cdot\hat{\mathbf{r}})}dl \ , \tag{4.54}$$

$$\mathbf{I}_{M} = C_k \int_{L} \mathbf{M}_l(\mathbf{r}')e^{jk(\mathbf{r}'\cdot\hat{\mathbf{r}})}dl \ . \tag{4.55}$$

If $\mathbf{J}_l$ is distributed along a straight line, $\mathbf{I}_{J}$ becomes C_k multiplied with the Fourier transform $\tilde{\mathbf{J}}_l$ of $\mathbf{J}_l$ and correspondingly for $\mathbf{M}_l$ and $\mathbf{I}_{M}$. This will be used in Chapter 5 when analyzing wire and slot antennas. Similarly, if $\mathbf{J}_l$ is distributed over a plane surface, $\mathbf{I}_{J}$ becomes C_k multiplied with the two-dimensional Fourier transform $\tilde{\mathbf{J}}_l$ of $\mathbf{J}_l$, and correspondingly for $\mathbf{M}_l$ and $\mathbf{I}_{M}$. This property will be used in Chapter 7 when analyzing radiation from apertures.

4.2.3 Duality

The expressions for the field solutions $\mathbf{E}_{M}$ and $\mathbf{H}_{M}$ resulting from magnetic current sources have similarities with the expressions for the fields $\mathbf{E}_{J}$ and $\mathbf{H}_{J}$ from electric current sources. This is due to the symmetry of Maxwell's equations. This symmetry can be formulated in a general *duality theorem* which allows us to find the fields resulting from magnetic currents by using the solution for the fields resulting from corresponding electric currents, and vice versa, using the replacements in Table 4.1. A consequence of duality is that the impedance Z_e at the port of an electric current distribution becomes equal to η^2 times the admittance Y_m at the port of the dual magnetic current distribution of the dual problem[14]. Thus,

$$Z_e = \eta^2 Y_m \ . \tag{4.56}$$

[14] For more information see Section 4.5.4 and Section 5.4.2.

Table 4.1: Duality replacements

Original problem 1	$\eta\mathbf{J}$	$\mathbf{E}_J$	$\eta\mathbf{H}_J$	k	Dual of problem 2
Dual problem 1	$\mathbf{M}$	$\eta\mathbf{H}_M$	$-\mathbf{E}_M$	k	Original of problem 2

4.2.4 Superposition

The integral form of Maxwell's equations is a result of their linearity. The integral vector form also clearly shows that *superposition* applies. We can find the total E- and H-fields of several sources by a complex vectorial addition of the E- and H-fields originating from each of them.

4.2.5 Replacement between electric and magnetic currents

An electric field problem can either have electric or magnetic current sources, or both. In addition, any known electric current distribution can be replaced by an equivalent magnetic current distribution which provides exactly the same E- and H-fields everywhere, and vice versa. The magnetic replacement current distribution $\mathbf{M}_J$ which gives exactly the same field as a known electric current distribution $\mathbf{J}$ is type[15]

$$\mathbf{M}_J = -\frac{1}{j\omega\varepsilon}\nabla_t \times \mathbf{J} = j\frac{\eta}{k}\nabla_t \times \mathbf{J} \tag{4.57}$$

where ∇_t is the curl operator applied only to the coordinates in a plane orthogonal to $\mathbf{J}$. Correspondingly, the electric replacement current distribution $\mathbf{J}_M$, which gives exactly the same field as a known magnetic current distribution $\mathbf{M}$, is

$$\mathbf{J}_M = \frac{1}{j\omega\mu}\nabla_t \times \mathbf{M} = -j\frac{1}{k\eta}\nabla_t \times \mathbf{M}\ . \tag{4.58}$$

These *replacement relations between electric and magnetic currents* can be derived from the differential form of Maxwell's equations. If we take the curl on both sides of (4.1) and use (4.7) and (4.2), we see that a source $j\omega\mu\mathbf{J}$ plays the same role in the equations as $\nabla_t \times \mathbf{M}$. Furthermore, if we take the curl on both sides of (4.2) and use (4.6) and (4.1), we see that $-j\omega\varepsilon\mathbf{M}$ plays the same role as $\nabla_t \times \mathbf{J}$.

This equivalence between $\mathbf{J}$ and $\mathbf{M}$ has no relation to duality and must not be confused with the equivalent sources in Section 4.3 which are used to replace fields over surfaces in space.

4.2.6 Frequency scaling

The principle of *frequency scaling* is an important consequence of Maxwell's equations. This is expressed in the best way by defining a scale model of a known antenna, as an antenna in which the dimensions are a factor S larger than (or smaller than if $S < 1$) the dimensions of

[15] The term replacement currents are specific for this book, but the concept is known and the equations given as (5a) and (5b) in [28]

the known antenna. Then, the radiation characteristics (i.e., the radiation pattern, directive gain, radiation impedance and input reflection coefficient) of the scale model at the frequency $f_{\text{scl}} = f/S$ is exactly the same as the characteristics of the known antenna at the frequency f, provided all materials are the same with exactly the same material parameters ε_r and μ_r. In practice, ε_r and μ_r of a given material may vary slightly with frequency, but unless the scaling factor is too large the changes are normally negligible. If only the imaginary parts of ε_r and μ_r are different, this will effect the antenna noise temperature and ohmic losses, but normally not the radiation pattern and the return loss at the input port.

4.3 Construction of solutions: uniqueness and equivalence

It is possible to show that Maxwell's time-harmonic equations have a unique solution in lossy materials within a finite region. The requirements for this are that all the sources in the region are known in addition to one of the following sets of field components on the boundary of the region:

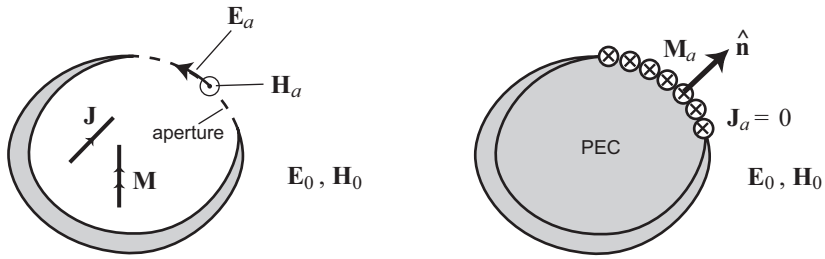
a) the tangential E-field on the whole boundary,
b) the tangential H-field on the whole boundary, or
c) the tangential E-field on part of the boundary and the tangential H-field on the rest of it.

This fact is referred to as the *uniqueness theorem*, and it is also valid for infinite regions by imposing the so-called radiation condition, which states that the fields at infinity are radiation fields decaying as $1/R$. In ideally loss-free finite regions the uniqueness theorem may not be necessarily valid, as internal resonances may be present at certain frequencies. Such internal resonances represent a big problem in numerical field solutions of integral equations by using, e.g., the *Method of Moments*. However, ideally loss-free problems are never present in the real world, so in practice field solutions are always unique.

The uniqueness theorem can be used to construct a field solution within a certain region from the tangential field components at the boundary of the region. This may be used with advantage in order to separate a complicated field problem into subproblems, where each subproblem is associated with a certain field region. We can do this by assuming known tangential field components between the regions. In order to facilitate this approach, the tangential fields at the boundaries between the regions are interpreted as equivalent sources. In the next subsection we will describe the three equivalents that are most commonly used in antenna analysis. We will treat the first two in more detail in Chapter 5 when treating radiation from apertures.

4.3.1 PEC equivalent and magnetic currents

Let us consider a conducting PEC surface which divides space in an outer and an inner field region, and an aperture at the location A in this otherwise closed surface (Fig. 4.4). We assume that there are sources in the inner region that produce fields $\mathbf{E}_a$ and $\mathbf{H}_a$ over the aperture. These aperture fields will in turn create fields $\mathbf{E}_0$ and $\mathbf{H}_0$ in the outer region. It is evident that $\mathbf{E}_a$ and $\mathbf{H}_a$ can be considered as sources for the fields $\mathbf{E}_0$ and $\mathbf{H}_0$ outside the conductor. Let us now construct an equivalent which makes use of this.



(a) Original problem (b) PEC equivalent

Figure 4.4: The PEC equivalent for a radiating aperture A in a conducting surface.

We fill the whole inner region with a PEC material and locate equivalent magnetic source currents

$$\mathbf{M}_a = \mathbf{E}_a \times \hat{\mathbf{n}} \tag{4.59}$$

over the surface of the PEC at the location A where the aperture was. $\hat{\mathbf{n}}$ is the normal to the surface A pointing outwards. In this new field problem, the E-field tangential to the surface of the PEC is zero resulting from the boundary condition in (4.14). However, at the part A of the surface there are impressed magnetic currents $\mathbf{M}_a$. Therefore, we must use the more general boundary condition (4.12), from which we see that $\mathbf{E}_2 \times \hat{\mathbf{n}} = \mathbf{M}_a$ when $\mathbf{E}_1 = 0$. Consequently, the tangential E-field $\mathbf{E}_2 \times \hat{\mathbf{n}}$ outside $\mathbf{M}_a$ is equal to $\mathbf{E}_a \times \hat{\mathbf{n}}$ in (4.59) which is the original tangential E-field in the aperture A before it is closed. Now, it is clear that the new outer field problem with the inner region filled with PEC has the same tangential E-fields at its boundary as the original field problem. Therefore it is equivalent to the original field problem for the outer region according to item (a) of the uniqueness theorem.

This equivalent is referred to as the *PEC equivalent*. It will give the correct field solution in the outer region if $\mathbf{E}_{\text{at}} = (\mathbf{E}_a)_{\tan}$ is exactly known. In practice, we may approximate $\mathbf{E}_{\text{at}}$, where we find an approximate solution in the outer region. This approximate solution is desirable if the approximate solution to $\mathbf{E}_{\text{at}}$ is acceptable. It is easy to find an approximate solution for $\mathbf{E}_{at}$ if the aperture is small in terms of wavelengths.

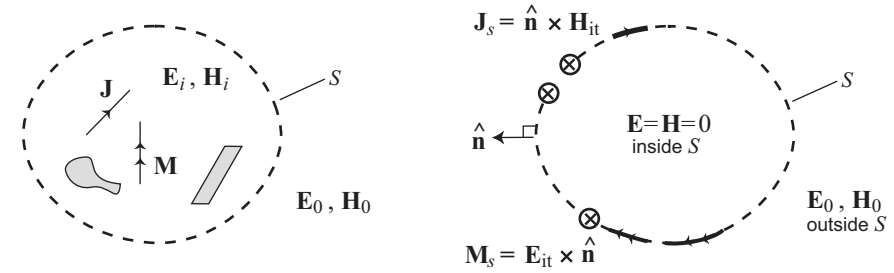
4.3.2 Free space and Huygens equivalents

Let us now consider a field problem with some known sources and scattering objects inside a region in space. We define the boundary S_i of this region in a way that it encloses all sources and scatterers, but does not coincide with any of them (see Fig. 4.5). We denote the fields inside S_i by $\mathbf{E}_i$ and $\mathbf{H}_i$, and outside of it by $\mathbf{E}_0$ and $\mathbf{H}_0$. We further assume that we know the tangential components $\mathbf{E}_{\text{it}} = \mathbf{E}_{0\text{t}}$ and $\mathbf{H}_{\text{it}} = \mathbf{H}_{0\text{t}}$ of both the E- and H-fields at the boundary S_i and want to use them to construct an equivalent problem as follows.

We replace $\mathbf{E}_{\text{it}}$ and $\mathbf{H}_{\text{it}}$ by

$$\mathbf{J}_{S} = \hat{\mathbf{n}} \times \mathbf{H}_{\text{it}} \qquad \text{and} \qquad \mathbf{M}_{S} = \mathbf{E}_{\text{it}} \times \hat{\mathbf{n}} \ , \tag{4.60}$$

where $\hat{\mathbf{n}}$ is the outgoing normal to S_i, and we remove all sources and scatterers in the inner region. By applying the general boundary condition in (4.12) to S_i, we see that if $\mathbf{E}_i = \mathbf{H}_i = 0$



(a) Original problem

(b) Free space equivalent

Figure 4.5: The free space equivalent of a field problem.

we get $\hat{\mathbf{n}} \times \mathbf{H}_{0t} = \mathbf{J}_S$ and $\mathbf{E}_{0t} \times \hat{\mathbf{n}} = \mathbf{M}_S$. Therefore, the tangential field components at S_i of this new problem are the same as those at S_i of the original problem. Thus, the new problem is equivalent to the original one in the outer region according to all items of the uniqueness theorem.

This equivalent is referred to as the *free space equivalent*, since the equivalent sources radiate in free space. If we choose the surface S_i to coincide with the wavefronts of the waves leaving the inner region, it corresponds to the case that Huygens used to illustrate his principle of reconstructing new wavefronts from previous wavefronts. Therefore, we will refer to this special form of the free space equivalent as the *Huygens equivalent*.

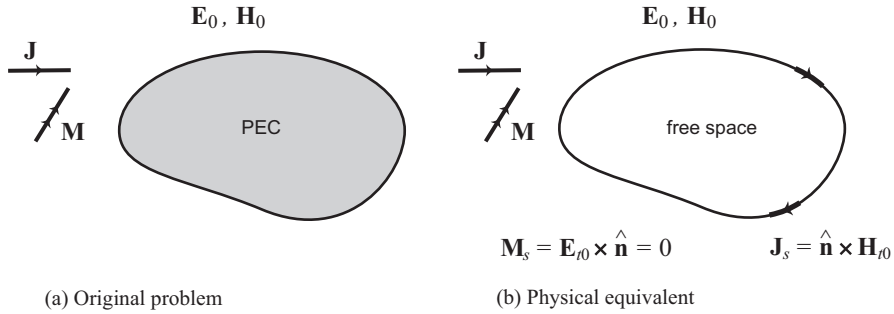
The free space and Huygens equivalents require knowledge of both the tangential E- and H-fields at the boundary. Therefore, they represent over-determined field problems in which we know more than we need in order to get a unique solution. This means that if we use it to find approximate field solutions by estimating $\mathbf{E}_t$ and $\mathbf{H}_t$ on S_i, we may get approximate field solutions that do not give $\mathbf{E}_i = \mathbf{H}_i = 0$ inside S_i. It is important to be aware of this fact. Still, the Huygens equivalent is very convenient to use for creating field solutions, because the sources radiate in free space so that free space Green's functions can be used. In addition, the results are often reasonably accurate over a large region of space. It is commonly used in the analysis of horn and reflector antennas.

4.3.3 Physical equivalent

The third equivalent is commonly used for analysis of antenna problems involving sources and metal scatterers (Fig. 4.6). This equivalent is valid in the region exterior to a PEC scatterer with a boundary defined by its surface with normal $\hat{\mathbf{n}}$, but it also gives correct results of the fields inside the PEC. The equivalent problem consists of the original sources plus some induced surface currents, both radiating in free space with the PEC removed. The induced currents are

$$\mathbf{J}_S = \hat{\mathbf{n}} \times \mathbf{H}_{0t} \tag{4.61}$$

located where the surface of the PEC was before its removal. Thus, the PEC has been replaced by the surface current distribution $\mathbf{J}_S$ located in free space. The *physical equivalent* is a special kind of free space equivalent.



(a) Original problem (b) Physical equivalent

Figure 4.6: The physical equivalent.

The advantage of the physical equivalent is that the PEC is removed, so that we can use free space formulas for the radiation field. The problem is to find the tangential component of $\mathbf{H}_S$ at the surface, so that $\mathbf{J}_S$ can be assumed known. A plausible approximation for large scatterers is the *Physical Optics* (*PO*) approximation which states that $\mathbf{J}_S = 2\hat{\mathbf{n}} \times \mathbf{H}_i$ where $\mathbf{H}_i$ is the incident field on the PEC only due to the original sources. The PO approximation will be treated in more details in Chapter 7. If we want a more precise solution for $\mathbf{J}_S$, we can use the PEC equivalent to formulate an integral equation with $\mathbf{J}_S$ as unknown, and then solve this integral equation by using, e.g., the Method of Moments.

4.4 Incremental current sources

In this section we will describe the fields of incremental current sources. Such sources play an important role in antenna analysis because more complicated sources can be built up as a superposition of incremental sources. The reason is that Maxwell's equations are linear so that the superposition principle applies. The vector fields from the incremental sources represent an evaluation of the vector dyadic *Green's functions* in free space. We will study both electric and magnetic sources. In addition, we introduce a combination of these sources which gives a rotationally symmetric radiation field of the BOR_1 type[16]. This hybrid source is commonly referred to as a Huygens source.

4.4.1 Incremental electric current (or Hertz dipole)

Let us introduce an incremental electric surface current distribution in the xy-plane of the form

$$\mathbf{J}_S(\mathbf{r}') = J_0 dS' \delta(x', y')\hat{\mathbf{y}} \ , \tag{4.62}$$

where $dS' = dx'dy'$ is an incremental surface area and $\delta(x, y)$ is the two-dimensional delta function. $\mathbf{J}_S(\mathbf{r}')$ is equivalent to an incremental electric line current of the form

$$\mathbf{J}(y') = I_0 dy' \delta(y')\hat{\mathbf{y}} \ , \tag{4.63}$$

[16] For more information about BOR_1 antennas, see Section 2.4.2.

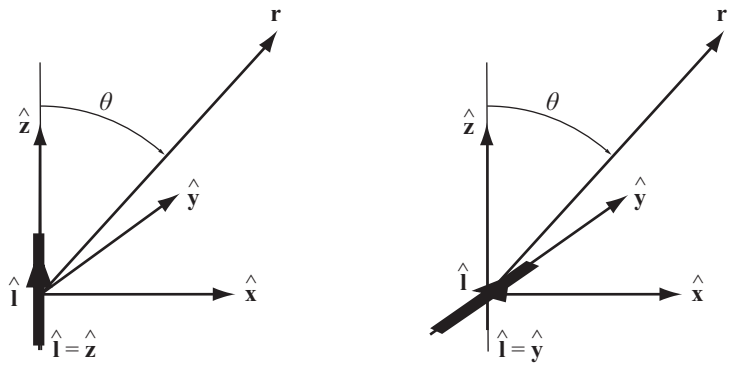


Figure 4.7: Geometry of vertical (left) and horizontal (right) incremental electric currents.

where $\delta(y')$ is the one-dimensional delta function and $I_0 = J_0 dx'$. Let us further change the orientation of the incremental source in a way that it points in an arbitrary direction $\hat{\mathbf{l}}$, and assume that the incremental length is finite and given by l. Then,

$$\mathbf{J}(l') = I_0 \hat{\mathbf{l}} \ , \qquad \text{for } -\frac{l}{2} < l' < \frac{l}{2} \ . \tag{4.64}$$

For historical reasons, such rectangular line current distributions are often referred to as Hertz dipoles. The name refers back to Heinrich Hertz and the spark gap dipoles he used in his experimental discovery of electromagnetic waves in 1888. When $l \ll \lambda$ and $kr \gg 1$, we get [from Eq. (4.34) - (4.43)] the following radiation fields of this incremental electric current

$$\mathbf{E}_{\rm id}(\mathbf{r}) = C_k \eta I_0 l [\hat{\mathbf{l}} - (\hat{\mathbf{l}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}] \frac{1}{r} e^{-jkr} \ , \tag{4.65}$$

$$\eta \mathbf{H}_{\rm id}(\mathbf{r}) = C_k \eta I_0 l (\hat{\mathbf{r}} \times \hat{\mathbf{l}}) \frac{1}{r} e^{-jkr} = \hat{\mathbf{r}} \times \mathbf{E}_{\rm id}(\mathbf{r}) \ , \tag{4.66}$$

where $C_k = -jk/4\pi$ as in (4.41).We can also introduce a far-field function $\mathbf{G}_{\rm id}(\mathbf{r})$ by

$$\mathbf{E}_{\rm id}(\mathbf{r}) = \frac{1}{r} e^{-jkr} \mathbf{G}_{\rm id}(\hat{\mathbf{r}}) \ ; \qquad \mathbf{G}_{\rm id}(\hat{\mathbf{r}}) = C_k \eta I_0 l [\hat{\mathbf{l}} - (\hat{\mathbf{l}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}] \ . \tag{4.67}$$

Now we can use (4.65), (4.66) or (4.67) directly with $\hat{\mathbf{r}} = \sin\theta(\cos\varphi\hat{\mathbf{x}} + \sin\varphi\hat{\mathbf{y}}) + \cos\theta\hat{\mathbf{z}}$ in order to find the far-fields of arbitrary oriented electric currents. The vector operations can easily be understood and evaluated from the explanations to equations (1.7) to (1.9). Yet, here we still study two specific electric current orientations analytically.

When the electric current is oriented vertically along the z-axis, we have $\hat{\mathbf{l}} = \hat{\mathbf{z}}$, which can be found from a simple drawing (Fig. 4.7) that $\hat{\mathbf{z}} - (\hat{\mathbf{z}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} = -\sin\theta\hat{\boldsymbol{\theta}}$. We can also use the equations in Appendix C to obtain the same. Therefore, the far-field function becomes

$$\mathbf{G}_{\rm id} = -C_k \eta I_0 l \sin\theta \hat{\boldsymbol{\theta}} \ . \tag{4.68}$$

Thus, the vertical incremental electric current radiates as a BOR_0 *antenna* with a rotationally symmetric radiation pattern[17].

[17] $\rm BOR_0$ antennas can be found in Section 2.4.1 on page 48.

When the electric current is located in the center of the coordinate system and oriented along the y-axis so that $\hat{\mathbf{l}} = \hat{\mathbf{y}}$, we get the far-field function

$$\mathbf{G}_{\rm id}(\hat{\mathbf{r}}) = C_k \eta I_0 l(\hat{\mathbf{y}} - (\hat{\mathbf{y}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}) \ . \tag{4.69}$$

We can now use $\hat{\mathbf{y}} = \sin\theta \sin\varphi \hat{\mathbf{r}} + \cos\theta \sin\varphi \hat{\boldsymbol{\theta}} + \cos\varphi \hat{\boldsymbol{\varphi}}$ from Appendix C to obtain the following dependency on θ and φ

$$\mathbf{G}_{\rm id}(\theta, \varphi) = C_k \eta I_0 l(\cos\theta \sin\varphi \hat{\boldsymbol{\theta}} + \cos\varphi \hat{\boldsymbol{\varphi}}) \ . \tag{4.70}$$

We see that the horizontal electric current radiates as a BOR_1 *antenna* with respect to the z-axis[18]. The H-plane pattern is omnidirectional, whereas the E-plane pattern is given by $|\cos\theta|$ which has maxima in the $\pm z$-directions and zeros along the positive and negative y-axes, corresponding to $\theta = 90^\circ$ and $\varphi = \pm 90^\circ$ (see Fig. 4.8 and Fig. 4.9). An x-directed electric current gives the corresponding field patterns in its E- and H-planes[19].

From the above, we see that the direction of the E-field at any point can be found by taking the direction $\hat{\mathbf{l}}$ of the electric current and removing its $\hat{\mathbf{r}}$ component in a way that the field becomes normal to $\hat{\mathbf{r}}$. The radiation field is always zero in the directions $\hat{\mathbf{r}} = \pm\hat{\mathbf{l}}$.

4.4.2 Incremental magnetic current

Magnetic current sources are important as they can be used as equivalent sources when analyzing radiation from aperture antennas (see Chapter 7). The radiation field of the incremental magnetic current is readily found in the same way as for the incremental electric current, or it can be constructed from the electric current solutions by using duality. Consider an incremental magnetic current oriented along $\hat{\mathbf{l}}_m$, i.e.,

$$\mathbf{M}(l') = M_0 \hat{\mathbf{l}}_m \ , \qquad \text{for } -\frac{l}{2} < l' < \frac{l}{2} \ . \tag{4.71}$$

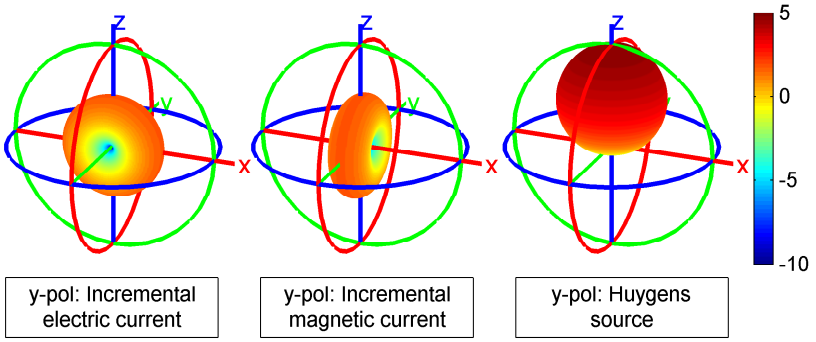
Then, the duality relation[20] states that the E- and H-fields $\mathbf{E}_{\rm im}(\mathbf{r})$ and $\mathbf{H}_{\rm im}(\mathbf{r})$ caused by $\mathbf{M}$ are found from $\mathbf{E}_{\rm id}(\mathbf{r})$ and $\mathbf{H}_{\rm id}(\mathbf{r})$ of the incremental electric currents through

$$\mathbf{E}_{\rm im}(\mathbf{r}) = C_k M_0 l(\hat{\mathbf{l}}_m \times \hat{\mathbf{r}}) \frac{1}{r} e^{-jkr} \ , \tag{4.72}$$

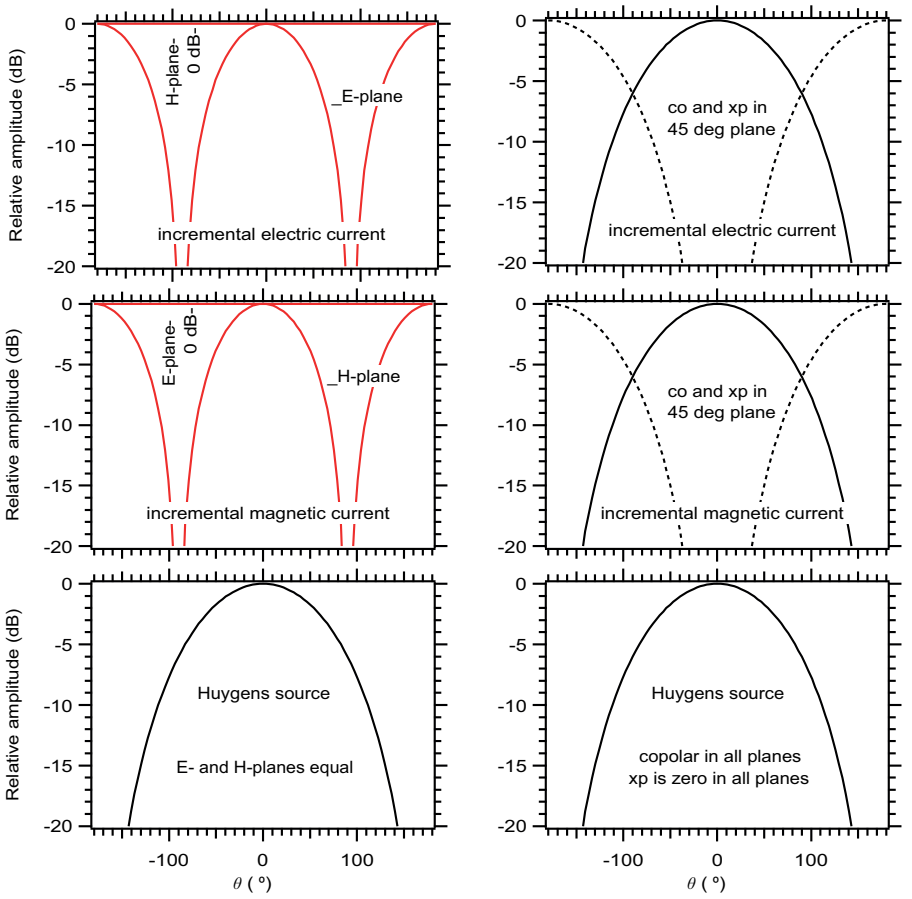
$$\mathbf{H}_{\rm im}(\mathbf{r}) = \frac{1}{\eta} C_k M_0 l[\hat{\mathbf{l}}_m - (\hat{\mathbf{l}}_m \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}] \frac{1}{r} e^{-jkr} \ , \tag{4.73}$$

where $\eta I_0 l$ for the electric current case has been replaced by $M_0 l$, and $C_k = -jk/4\pi$. The fields of both x- and y-directed magnetic currents are written in spherical coordinates later on page 144. It is clear from the discussion of the short electric current that the short magnetic current is omnidirectional in E-plane and has a $\cos\theta$-pattern in H-plane (see Fig. 4.8). The radiation field is zero in the directions $\hat{\mathbf{r}} = \pm\hat{\mathbf{l}}_m$. The direction of the E-field is orthogonal to both $\hat{\mathbf{l}}_m$ and $\hat{\mathbf{r}}$.

[18] $\rm BOR_1$ antennas can be found in Section 2.4.2 on page 49.
[19] See Section 4.4.4 on page 144.
[20] See Section 4.2.3 on page 134.

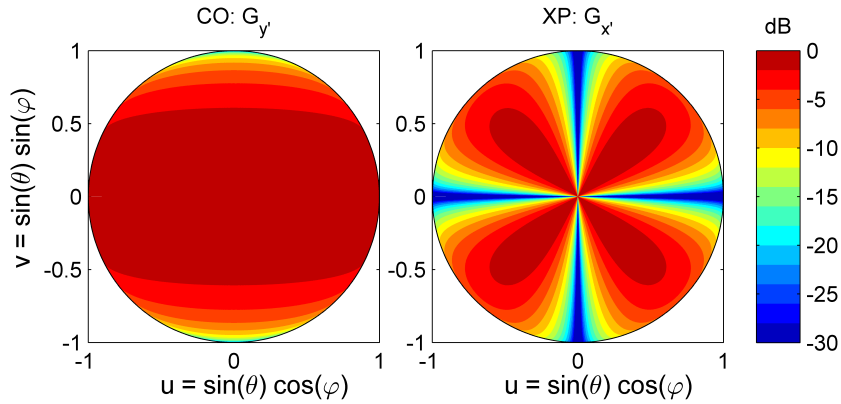


(a)

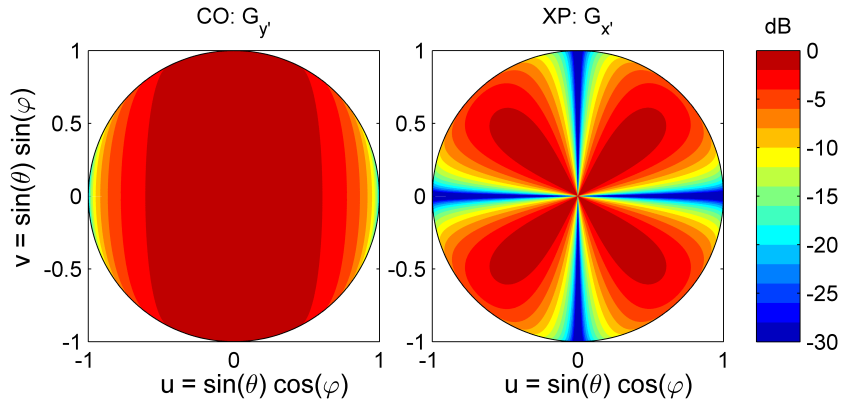


(b)

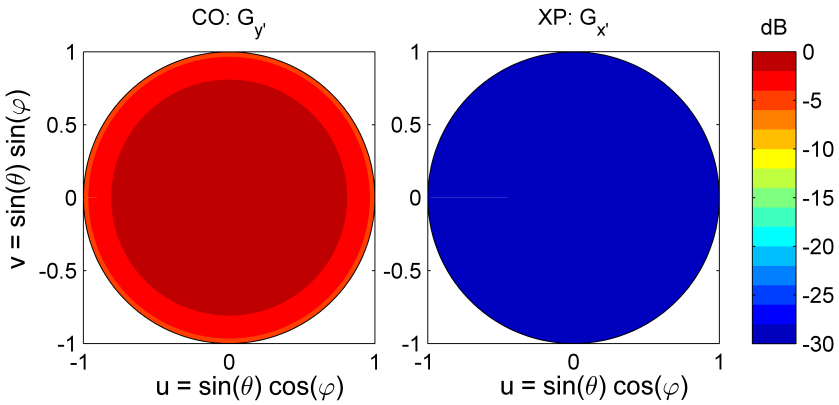
Figure 4.8: Radiation patterns of incremental electric current, magnetic current, and Huygens source: (a) 3D plots when they are polarized in y-direction on the z-axis. (b) E- and H-plane, and co- and cross-polar patterns in the 45°-plane.



(a) Y-polarized incremental electric current.



(b) Y-polarized incremental magnetic current.



(c) Y-polarized Huygens source.

Figure 4.9: Co-polar and cross-polar radiation patterns for y-polarized (a) incremental electric current, (b) magnetic current and (c) Huygens source.

4.4.3 Huygens source

There exists an incremental source which has equal and uni-directional E- and H-plane patterns. This source plays an important role in the analysis of aperture antennas. We refer to it as the Huygens source [8, pp. 44], and describe it as two orthogonal incremental currents, one electric I_0 and one magnetic $M_0 = \eta I_0$. The Huygens source is always present as an incremental area source. Let us assume that it is located in the xy-plane. Then, the y-polarized Huygens source radiating in z-direction is given by the sum of the two sources

$$\mathbf{J}_{H} = J_0 dS' \delta(x', y')\hat{\mathbf{y}} \ , \tag{4.74}$$

$$\mathbf{M}_{H} = \eta J_0 dS' \delta(x', y')(-\hat{\mathbf{x}}) \ , \tag{4.75}$$

where dS' is the incremental surface area.The resultant far-field function becomes

$$\begin{aligned} \mathbf{G}_{H}(\hat{\mathbf{r}}) &= \mathbf{G}_{\mathrm{id}_H}(\hat{\mathbf{r}}) + \mathbf{G}_{\mathrm{im}_H}(\hat{\mathbf{r}}) \\ &= C_k \eta J_0 dS' [\hat{\mathbf{y}} - (\hat{\mathbf{y}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} - \hat{\mathbf{x}} \times \hat{\mathbf{r}}] \ , \end{aligned} \tag{4.76}$$

with $C_k = -jk/4\pi$ the same as before. By using the results of the previous two subsections and $\hat{\mathbf{x}} = \sin\theta\cos\varphi\hat{\mathbf{r}} + \cos\theta\cos\varphi\hat{\boldsymbol{\theta}} - \sin\varphi\hat{\boldsymbol{\varphi}}$ from Appendix C, we get

$$\begin{aligned} \mathbf{G}_{H}(\theta, \varphi) &= C_k \eta J_0 dS' [(\cos\theta \sin\varphi\hat{\boldsymbol{\theta}} + \cos\varphi\hat{\boldsymbol{\varphi}}) + (\sin\varphi\hat{\boldsymbol{\theta}} + \cos\theta\cos\varphi\hat{\boldsymbol{\varphi}})] \\ &= C_k \eta J_0 dS' (1 + \cos\theta)[\sin\varphi\hat{\boldsymbol{\theta}} + \cos\varphi\hat{\boldsymbol{\varphi}}] \\ &= 2C_k \eta J_0 dS' \cos^2(\theta/2)[\sin\varphi\hat{\boldsymbol{\theta}} + \cos\varphi\hat{\boldsymbol{\varphi}}] \ . \end{aligned} \tag{4.77}$$

We see that the Huygens source radiates as a BOR_1 *antenna* with equal E- and H-plane patterns. This means that the Huygens source has no cross-polarization according to *Ludwig's third definition*[21]. We also see that we have a null for $\theta = 180°$, which is the backward direction. These two properties make the Huygens source a desirable *Green's function* for many radiation patterns. It actually also appears as a Green's function in some practical antennas, such as in large aperture antennas (see Chapter 7).

For an arbitrary polarization $\hat{\mathbf{l}}$ of the incremental Huygens source we get the far-field function of the following general vector form

$$\mathbf{G}_{H}(\hat{\mathbf{r}}) = C_k \eta J_0 dS' [\hat{\mathbf{l}} - (\hat{\mathbf{l}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} - (\hat{\mathbf{l}} \times \hat{\mathbf{n}}) \times \hat{\mathbf{r}}] \ , \tag{4.78}$$

where $\hat{\mathbf{n}}$ is the normal to the incremental surface area dS' in the direction of radiation.

4.4.4 Summary

We will often need to know the far-field functions of the incremental electric and magnetic current sources as well as the Huygens source in the spherical coordinate system. Therefore, they are summarized here, for the case when they are lying in the xy-plane in the origin of the coordinate system. In these equations $C_k = -jk/4\pi$ as before.

A. Incremental electric current source of unit magnitude $\eta J_0 l = 1$.

$$\begin{aligned} &y\text{-polarized}, \quad \hat{\mathbf{l}} = \hat{\mathbf{y}}\text{:} \ \mathbf{G}_{\mathrm{id}}(\theta, \varphi) = C_k[\hat{\mathbf{y}} - (\hat{\mathbf{y}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}] = C_k[\cos\theta \sin\varphi\hat{\boldsymbol{\theta}} + \cos\varphi\hat{\boldsymbol{\varphi}}] \\ &x\text{-polarized}, \quad \hat{\mathbf{l}} = \hat{\mathbf{x}}\text{:} \ \mathbf{G}_{\mathrm{id}}(\theta, \varphi) = C_k[\hat{\mathbf{x}} - (\hat{\mathbf{x}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}] = C_k[\cos\theta \cos\varphi\hat{\boldsymbol{\theta}} - \sin\varphi\hat{\boldsymbol{\varphi}}] \end{aligned} \tag{4.79}$$

[21] See Section 2.4.2 on page 49 for more information about BOR_1 antennas.

B. Incremental magnetic current source of unit magnitude $M_0 l = 1$.

$$\begin{aligned} &y\text{-polarized,} \quad \hat{\mathbf{l}}_m = -\hat{\mathbf{x}}\text{:}\ \mathbf{G}_{\mathrm{im}}(\theta,\varphi) = C_k[-(\hat{\mathbf{x}}) \times \hat{\mathbf{r}}] = C_k[\sin\varphi\hat{\boldsymbol{\theta}} + \cos\theta\cos\varphi\hat{\boldsymbol{\varphi}}] \\ &x\text{-polarized,} \quad \hat{\mathbf{l}}_m = \ \ \hat{\mathbf{y}}\text{:}\ \mathbf{G}_{\mathrm{im}}(\theta,\varphi) = C_k[\hat{\mathbf{y}} \times \hat{\mathbf{r}}] = C_k[\cos\varphi\hat{\boldsymbol{\theta}} - \cos\theta\sin\varphi\hat{\boldsymbol{\varphi}}] \end{aligned} \tag{4.80}$$

C. Huygens source of unit magnitude $\eta I_0 dS' = 1$.

$$\begin{aligned} &y\text{-polarized,} \quad \hat{\mathbf{l}} = \hat{\mathbf{y}}\text{:}\ \mathbf{G}_{H}(\theta,\varphi) = 2C_k\cos^2(\theta/2)[\sin\varphi\hat{\boldsymbol{\theta}} + \cos\varphi\hat{\boldsymbol{\varphi}}] \\ &x\text{-polarized,} \quad \hat{\mathbf{l}} = \hat{\mathbf{x}}\text{:}\ \mathbf{G}_{H}(\theta,\varphi) = 2C_k\cos^2(\theta/2)[\cos\varphi\hat{\boldsymbol{\theta}} - \sin\varphi\hat{\boldsymbol{\varphi}}] \end{aligned} \tag{4.81}$$

4.4.5 Example: Directivities of incremental sources

Derive and compare the directivities of the incremental electric current source, the incremental magnetic current source, and the Huygens source.

SOLUTION:

When the three sources are parallel with the xy-plane, their far-field function has the same form as that of BOR_1 antennas, so we may use the power integral in (2.92). For both the electric and magnetic current sources, this gives

$$P_{\mathrm{id}} = \pi|C_k|^2 \int_0^\pi (\cos^2\theta + 1)\sin\theta d\theta = \pi|C_k|^2 \frac{8}{3}$$

as their radiation patterns are equal, but with their E- and H-planes interchanged. The directivities are then also equal, obtained by using (2.69);

$$D_0 = \frac{4\pi|C_k|^2}{\pi|C_k|^2 8/3} = \frac{3}{2} \quad (\text{i.e., } 1.76\,\text{dBi}) \ .$$

The far-field function of the Huygens source is of the form $\cos^n(\theta/2)$ with $n = 2$ and no cross-polarization. In Section 2.4.3, we have already evaluated the directivity of the $\cos^n(\theta/2)$ pattern, and the result is for $n = 2$

$$D_0 = n + 1 = 3 \quad (\text{i.e., } 4.77\,\text{dBi}) \ .$$

Thus, the Huygens source is 3 dB more directive than the incremental electric and magnetic dipoles.

4.5 Reaction, reciprocity and mutual coupling

The equivalent circuit of receiving antennas[22] contains an induced current or voltage source. This is proportional to the amplitude of the incoming plane wave, and thereby also to the current or voltage at the port of the transmitting antenna. This induced source represents

[22] More about this can be found in Section 2.6.1.

therefore the mutual coupling between the two antennas. In this section, we will show how to calculate this coupling by using reciprocity and reaction. The principle of reciprocity was already introduced in Section 2.3.2. Despite that, it will be formulated more completely in terms of reaction integrals in the present section. The *reciprocity theorem* is described mathematically by three different relations valid for different types of sources. These relations can be derived from Maxwell's time-harmonic equations by using known vector operations and by assuming a reciprocal medium, see, e.g., [1, Sec. 3-8]. The three relations will be given below in (4.86) to (4.88) without derivations, and we will show how to use them to calculate mutual coupling, mutual impedances and mutual admittances, as well as self-impedances and self-admittances.

4.5.1 Reaction integrals

The reciprocity relations are formulated most compactly by first introducing the so-called *reaction* between the field of a source a and a source b, which has the form of an integral. The extent of source b is the integration area A_b. The *reaction integrals* has two forms, one which is valid for the case when source b is an electric current distribution $\mathbf{J}_b$, and another when it is a magnetic current distribution $\mathbf{M}_b$. These are

$$\langle \mathbf{E}_a, \mathbf{J}_b \rangle = \iint_{A_b} \mathbf{E}_a \cdot \mathbf{J}_b dA \ , \tag{4.82}$$

$$\langle \mathbf{H}_a, \mathbf{M}_b \rangle = -\iint_{A_b} \mathbf{H}_a \cdot \mathbf{M}_b dA \ . \tag{4.83}$$

In the first definition $\mathbf{E}_a$ is the E-field distribution over A_b due to a source $\mathbf{J}_a$ or $\mathbf{M}_a$, calculated when $\mathbf{J}_b = 0$. In the second definition $\mathbf{H}_a$ is the corresponding the H-field distribution over A_b due to a source $\mathbf{J}_a$ or $\mathbf{M}_a$, calculated when $\mathbf{M}_b = 0$. The reaction integrals are shown as surface integrals, but they become line integrals if the sources $\mathbf{J}_{\rm lb}$ and $\mathbf{M}_{\rm lb}$ are line currents, i.e.,

$$\langle \mathbf{E}_a, \mathbf{J}_{\rm lb} \rangle = \int_{L_b} \mathbf{E}_a \cdot \mathbf{J}_{\rm lb} dl \ , \tag{4.84}$$

$$\langle \mathbf{H}_a, \mathbf{M}_{\rm lb} \rangle = -\int_{L_b} \mathbf{H}_a \cdot \mathbf{M}_{\rm lb} dl \ , \tag{4.85}$$

where L_b defines the extents of the sources.

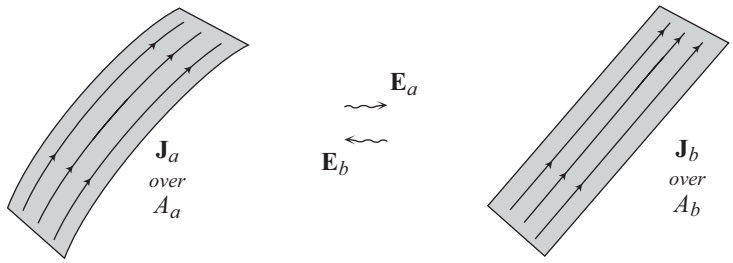
4.5.2 Three reciprocity relations

We are now ready to present the three different *reciprocity relations*, with reference to the drawings in Fig. 4.10:

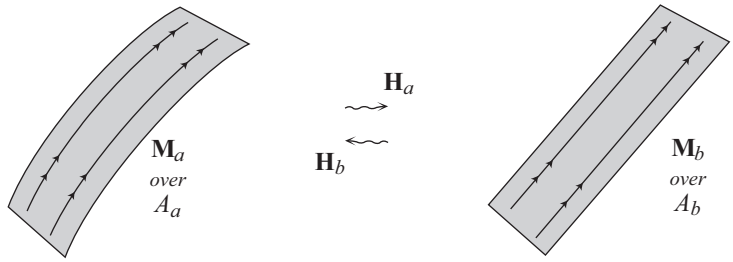
1. If source a is a known electric surface current $\mathbf{J}_a$ distributed over the area A_a, and source b is a known electric surface current $\mathbf{J}_b$ distributed over the area A_b, reciprocity states that

$$\langle \mathbf{E}_a, \mathbf{J}_b \rangle = \langle \mathbf{E}_b, \mathbf{J}_a \rangle \ , \tag{4.86}$$

where $\mathbf{E}_a$ is the field distribution over A_b due to $\mathbf{J}_a$, calculated when $\mathbf{J}_b = 0$, and $\mathbf{E}_b$ is the field distribution over A_a due to $\mathbf{J}_b$, calculated when $\mathbf{J}_a = 0$. In other words,



(a) Electric surface current distributions



(b) Magnetic surface current distribution

Figure 4.10: Illustration of reciprocity between two electric current distributions (upper), and between two magnetic current distributions (lower).

equation (4.86) expresses that the reaction between $\mathbf{E}_a$ and $\mathbf{J}_b$ is equal to the reaction between $\mathbf{E}_b$ and $\mathbf{J}_a$.

2. For magnetic current sources $\mathbf{M}_a$ and $\mathbf{M}_b$ over A_a and A_b, respectively, we have correspondingly

$$\langle \mathbf{H}_a, \mathbf{M}_b \rangle = \langle \mathbf{H}_b, \mathbf{M}_a \rangle \ , \tag{4.87}$$

where $\mathbf{H}_a$ is the H-field distribution over A_b due to $\mathbf{M}_a$, and $\mathbf{H}_b$ is the one due to $\mathbf{M}_b$ evaluated over A_a. In other words, equation (4.87) expresses that the reaction between $\mathbf{H}_a$ and $\mathbf{M}_b$ is equal to the reaction between $\mathbf{H}_b$ and $\mathbf{M}_a$.

3. When source a is an electric current source distributed over A_a, and source b is a magnetic source $\mathbf{M}_b$ distributed over A_b, we have

$$\langle \mathbf{H}_a, \mathbf{M}_b \rangle = \langle \mathbf{E}_b, \mathbf{J}_a \rangle \ , \tag{4.88}$$

where $\mathbf{H}_a$ is the H-field distribution over A_b due to $\mathbf{J}_a$, and $\mathbf{E}_b$ is the field distribution over A_a due to $\mathbf{M}_b$. In other words, the expression (4.88) expresses that the reaction between $\mathbf{H}_a$ and $\mathbf{M}_b$ is equal to the reaction between $\mathbf{E}_b$ and $\mathbf{J}_a$.

4.5.3 Reciprocity between input/output ports of antennas

Let us first use the reciprocity relations to study the coupling between the ports a and b of two antennas (Fig. 4.11). We consider two short electric line currents defined by

$$\mathbf{J}_{\rm la} = I_a \hat{\mathbf{l}}_a \ , \qquad \text{for } -\frac{1}{2}l_a < l < \frac{1}{2}l_a \ , \tag{4.89}$$

$$\mathbf{J}_{\rm lb} = I_b \hat{\mathbf{l}}_b \ , \qquad \text{for } -\frac{1}{2}l_a < l < \frac{1}{2}l_a \ , \tag{4.90}$$

which are located between the poles of the ports a and b of the two antennas, respectively. Each port may be the input/output connector of the antenna. Now, reciprocity states that the following two reactions are equal:

$$\int_{-l_a/2}^{l_a/2} (\mathbf{E}_b \cdot \mathbf{J}_{\rm la}) dl = \int_{-l_b/2}^{l_b/2} (\mathbf{E}_a \cdot \mathbf{J}_{\rm lb}) dl \ . \tag{4.91}$$

We assume that I_a and I_b are constant over the small lengths l_a and l_b, respectively. In addition, we presume that l_a and l_b are so small in terms of the wavelength that E_b and E_a, respectively, can be considered constant over them. This is in particular true between the two wires of a two-wire transmission line. Then we easily evaluate the two coupling integrals in (4.91) and get

$$-V_{\rm ba} I_a = -V_{\rm ab} I_b \ , \tag{4.92}$$

$$\text{where} \qquad V_{\rm ba} = -\frac{1}{I_a}\langle \mathbf{E}_b, \mathbf{J}_{la} \rangle = -\int_{l_a/2}^{l_a/2} (\mathbf{E}_b \cdot \hat{\mathbf{l}}_a) dl \approx E_b l_a \tag{4.93}$$

is the voltage measured at port a due to a source at terminal b, and

$$V_{\rm ab} = -\frac{1}{I_b}\langle \mathbf{E}_a, \mathbf{J}_{lb} \rangle = -\int_{-l_b/2}^{l_b/2} (\mathbf{E}_a \cdot \hat{\mathbf{l}}_b) dl \approx E_a l_b \tag{4.94}$$

is the *voltage* at terminal b due to a source at terminal a. The above two integrals represent voltages, as the voltage between a point $l = l_b/2$ and another point $l = -l_b/2$ is defined by

$$V_b = -\int_{-l_b/2}^{l_b/2} \mathbf{E} \cdot \hat{\mathbf{l}}_b dl \ .$$

The integrals can be taken along any curve between the points $l_b/2$ and $-l_b/2$, in our case a straight line. If the two sources are equal, i.e., $I_a = I_b = I$, the expression (4.92) yields $V_{\rm ba} = V_{\rm ab}$. This means that the response is the same either way. And, if we excite the port of antenna a with a current I, the response at the port of antenna b is a voltage V. If we excite the port of antenna b with the same current I, the response at the port of antenna a is the same, i.e., V. This is only correct when the antennas (and their environments) are exactly the same and have exactly the same locations and orientations, in the two cases.

The ratios $Z_{\rm ba} = V_{\rm ba}/I_b$ and $Z_{\rm ab} = V_{\rm ab}/I_a$ are the mutual impedances between ports a and b, and $Z_{\rm ba} = Z_{\rm ab}$ follows from the reciprocity. According to the above formula for the *mutual impedance* written in terms of the reaction integral is according to the above

$$Z_{\rm ab} = -\frac{1}{I_a I_b}\langle \mathbf{E}_b, \mathbf{J}_{\rm la} \rangle = -\frac{1}{I_a I_b}\langle \mathbf{E}_a, \mathbf{J}_{\rm lb} \rangle \ . \tag{4.95}$$

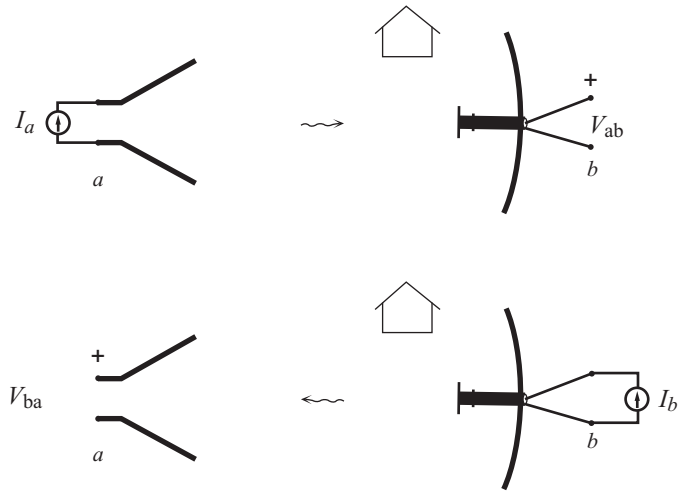


Figure 4.11: Illustration of reciprocity between point current sources I_a and I_b connected to the ports of two antennas.

Now let us consider the self-reaction of the E-field due to $\mathbf{J}_{\rm la}$ at port a. The negative of the self-reaction divided by I_a represents then the induced voltage between the poles of port a due to the current I_a on the same port. Therefore, we may define the *self-impedance* of antenna a as this voltage divided by I_a, i.e.,

$$Z_{\rm aa} = -\frac{1}{I_a I_a}\langle \mathbf{E}_a, \mathbf{J}_{\rm la}\rangle \ . \tag{4.96}$$

The self-impedance of antenna b can be defined correspondingly. The above two formulas are very general and apply even to the extended current distributions[23].

4.5.4 Mutual impedance, mutual admittance and coupling coefficient

The reaction relations can be used to calculate the mutual coupling between all types of sources with known forms, e.g., between the basis functions of a series expansion of the current distribution in a Method of Moments solution. In order to see this and interpret the reaction relations we may consider two discrete circuit ports a and b (associated with current distributions $\mathbf{J}_a$ and $\mathbf{J}_b$) and the equivalent circuits representing these ports, as follows (Fig. 4.12). We define the port b of an electric current distribution $\mathbf{J}_b$ on A_b by an infinitesimal gap located somewhere across and normal to the current distribution, and we define the port current I_b as the total current passing this feed gap, i.e., $I_b = \int \mathbf{J}_b \cdot \hat{\mathbf{l}} dt$ where $\hat{\mathbf{l}}$ is the direction normal to the gap and t is along the gap. Similarly, we can define a port a with current I_a. It is clear that $\mathbf{J}_a$ and $\mathbf{E}_a$ are proportional to I_a, and that $\mathbf{J}_b$ and $\mathbf{E}_b$ are proportional to I_b. We normalize the reaction integrals in (4.82) [which have the dimension Volt-Ampere (VA)]

[23] It will come more about this in the rest of this section.

with the product $I_a I_b$. Then, the reaction integrals get the dimension ohm (Ω) and represent the *mutual impedance* between the two equivalent circuit ports a and b. The formula for the mutual impedance is the same as between point sources in (4.95) and gets into the equivalent circuit as shown in Fig. 4.12[24].

In the same way we can define the port b of a magnetic current distribution over A_b by two opposite points along the rim of $\mathbf{M}_b$. Then we define the port voltage as the total voltage V_b between these two points, i.e., the integrated E-field along a line between the two points. Thus, $V_b = \int \mathbf{M}_b \cdot \hat{\mathbf{l}} dt$ where $\hat{\mathbf{l}}$ is the direction normal to the line and t is along the line (Fig. 4.13). Then, the reaction integrals in (4.83) (which have the dimension Volt-Ampere (VA) as well) get the dimension Siemens (S) and represent a *mutual admittance* $Y_{\rm ab}$ between the two ports a and b if we normalize them with the product $V_a V_b$ of the voltages over these ports, i.e.,

$$Y_{\rm ab} = \frac{1}{V_a V_b} \langle \mathbf{H}_a, \mathbf{M}_b \rangle = \frac{1}{V_a V_b} \langle \mathbf{H}_b, \mathbf{M}_a \rangle \ . \tag{4.97}$$

We can also use the first two reaction integral expressions to calculate the self-reaction, i.e., the reaction between source a and the field of source a. If these are normalized by I_a^2 and V_a^2, respectively, they represent respectively the self impedance and admittance at the above defined ports[25].

By similar arguments, we can obtain the *mutual coupling coefficients* between an electric $\mathbf{J}_a$ and a magnetic $\mathbf{M}_b$ current distribution as follows, see Fig. 4.14:

$$C_{\rm ab} = -\frac{1}{V_b I_a} \langle \mathbf{E}_b, \mathbf{J}_a \rangle = \frac{1}{V_b I_a} \langle \mathbf{H}_a, \mathbf{M}_b \rangle \ . \tag{4.98}$$

This coupling coefficient $C_{\rm ab} = C_{\rm ba}$ represents a voltage coupling coefficient in the equivalent circuit for $\mathbf{J}_a$ and a current coupling coefficient in the equivalent circuit for $\mathbf{M}_b$, as shown in Fig. 2.21[26].

4.6 Imaging

Imaging is a technique which we can use to construct field solutions satisfying the boundary conditions at a specific surface. In practice, we only use imaging in connection with infinite plane PECs or PMCs. For other geometries the imaging equations become so complicated that they hardly represent any simplification over a numerical solution. We will here only show the imaging principle applied to infinite plane PECs and PMCs.

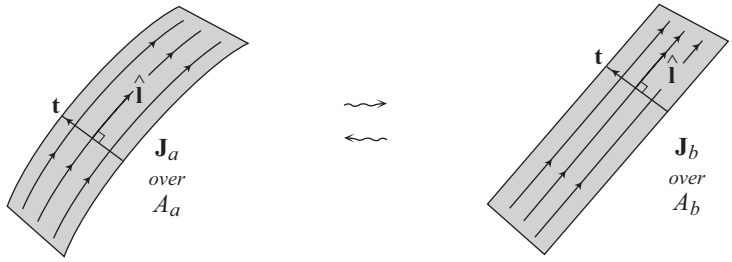
The imaging technique works as follows. The plane *PEC* or *PMC* is replaced by an image plane. The image plane divides space in two regions, the source region and the null-field region. We construct the field in the source region by locating an image source in the null-field region, in a way that the original boundary conditions on the image plane are satisfied[27]. Then by using the uniqueness theorem, the total field solution of the original problem in the source region is equal to the sum of the fields due to the impressed source and its image, i.e.,

[24] See also Section 2.6.1 on page 2.6.1.

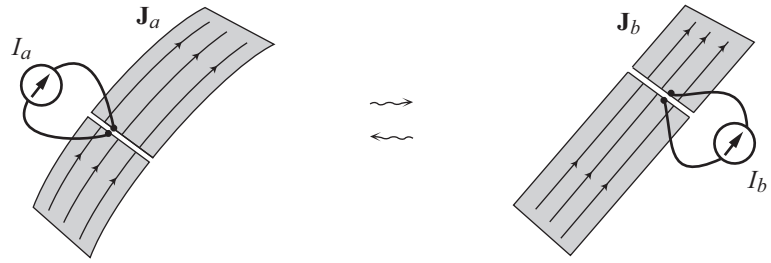
[25] For more information see Section 5.1.7 on page 178.

[26] More information can be found in Section 2.6.1 on page 65.

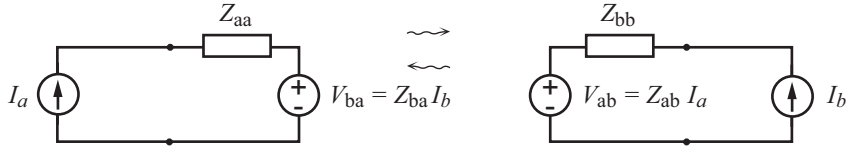
[27] This means $\hat{\mathbf{n}} \times \mathbf{E} = 0$ for a PEC and $\hat{\mathbf{n}} \times \mathbf{H} = 0$ for a PMC where $\hat{\mathbf{n}}$ is the normal to the image plane pointing into the source region.



(a) Electric surface current distributions

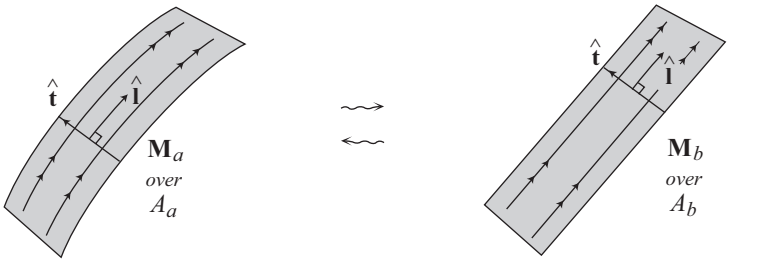


(b) Infinitesimal feed gaps and definitions of ports a and b

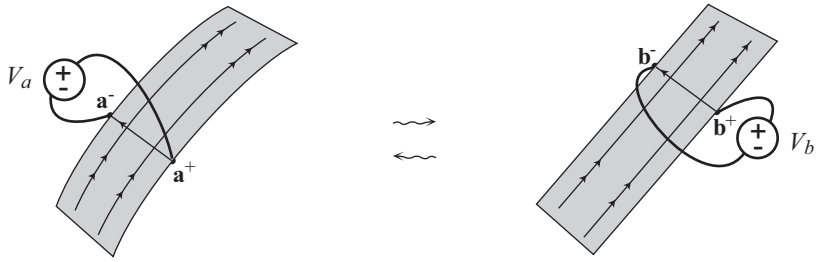


(c) Equivalent circuits

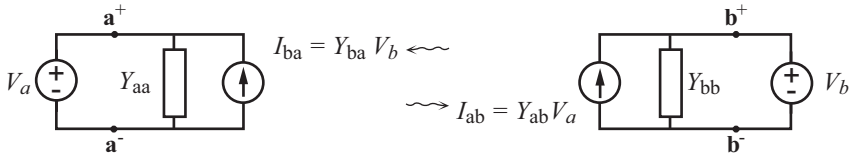
Figure 4.12: Illustration of mutual impedance between electric current distributions $\mathbf{J}_b$ and $\mathbf{J}_a$, and their equivalent circuits.



(a) Magnetic surface current distribution

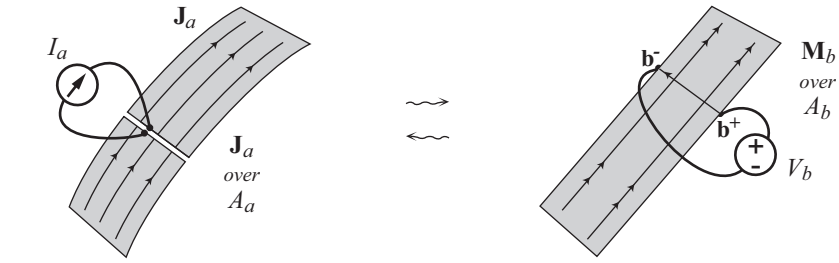


(b) Ports a and b and their poles $\mathbf{a}^+$ and $\mathbf{a}^-$, and $\mathbf{b}^+$ and $\mathbf{b}^-$, respectively

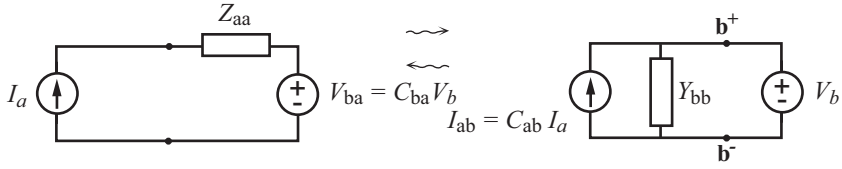


(c) Equivalent circuits

Figure 4.13: Illustration of mutual impedance between magnetic current distributions $\mathbf{M}_a$ and $\mathbf{M}_b$, and their equivalent circuits.



(a) Electric $\mathbf{J}_a$ and magnetic $\mathbf{M}_b$ current distributions



(b) Equivalent circuits

Figure 4.14: Illustration of mutual coupling between electric current distribution $\mathbf{J}_a$ and magnetic current distribution $\mathbf{M}_a$, and their equivalent circuits.

$\mathbf{H}_{\text{tot}} = \mathbf{H}_i + \mathbf{H}_{\text{img}}$, and in the null-field region it is zero by definition. The induced electric currents on the PEC are given by

$$\mathbf{J}_{_S} = \hat{\mathbf{n}} \times \mathbf{H}_{\text{tot}} = \hat{\mathbf{n}} \times (\mathbf{H}_i + \mathbf{H}_{\text{img}}) = 2\hat{\mathbf{n}} \times \mathbf{H}_i \ , \tag{4.99}$$

where the latter equality follows from the symmetry of the problem and also from the equations to follow. Correspondingly, the induced magnetic currents on the PMC are

$$\mathbf{M}_{_S} = \mathbf{E}_{\text{tot}} \times \hat{\mathbf{n}} = (\mathbf{E}_i + \mathbf{E}_{\text{img}}) \times \hat{\mathbf{n}} = 2\mathbf{E}_i \times \hat{\mathbf{n}} \ . \tag{4.100}$$

We will now show the expressions for the image sources when the actual sources are located over a plane PEC and a plane PMC. Consider first an electric current source $\mathbf{J}$ located at a point $\mathbf{r}'$, i.e., $\mathbf{J}(\mathbf{r}')$, and a ground plane defined by $\mathbf{r}_p(u, v)$ when the parameters u and v vary, having a constant surface normal $\hat{\mathbf{n}}$. Then, the image source is located at

$$\mathbf{r}_{\text{img}} = \mathbf{r}' - 2[(\mathbf{r}' - \mathbf{r}_p) \cdot \hat{\mathbf{n}}]\hat{\mathbf{n}} \ , \tag{4.101}$$

where $\mathbf{r}_p$ is any point on the image plane. If the ground plane is a PEC, the image current is given by

$$\mathbf{J}_{\text{img}} = -[\mathbf{J} - 2(\mathbf{J} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}] \ . \tag{4.102}$$

And, if the ground plane is a PMC, the image current is given by

$$\mathbf{J}_{\text{img}} = \mathbf{J} - 2(\mathbf{J} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}} \ . \tag{4.103}$$

If we instead consider a magnetic current source $\mathbf{M}$ located at $\mathbf{r}'$, the image will have the same location as in the $\mathbf{J}$ case, but the image current is

$$\mathbf{M}_{\text{img}} = \mathbf{M} - 2(\mathbf{M} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}} \tag{4.104}$$

for the PEC case. And for the PMC case, we have

$$\mathbf{M}_{\text{img}} = -[\mathbf{M} - 2(\mathbf{M}\cdot\hat{\mathbf{n}})\hat{\mathbf{n}}] \ . \tag{4.105}$$

If the source is incrementally small the far-field function of it (in free space) becomes by using (4.49) and (4.51)

$$\mathbf{G}(\hat{\mathbf{r}}) = C_k\eta[\mathbf{J}(\mathbf{r}') - (\mathbf{J}(\mathbf{r}')\cdot\hat{\mathbf{r}})\hat{\mathbf{r}}]e^{jk(\mathbf{r}'\cdot\hat{\mathbf{r}})} \ . \tag{4.106}$$

From the same two equations the far-field function of the image becomes

$$\mathbf{G}_{\text{img}}(\hat{\mathbf{r}}) = C_k\eta[\mathbf{J}_{\text{img}}(\mathbf{r}_{\text{img}}) - (\mathbf{J}_{\text{img}}(\mathbf{r}_{\text{img}})\cdot\hat{\mathbf{r}})\hat{\mathbf{r}}]e^{jk(\mathbf{r}_{\text{img}}\cdot\hat{\mathbf{r}})} \ . \tag{4.107}$$

Note that we have put no restriction on the location of the center of the coordinate system, i.e., the phase reference point, in these two equations. The total far-field function is then the sum of these, as they already have the same phase reference point, i.e.,

$$\mathbf{G}_{\text{tot}}(\hat{\mathbf{r}}) = \mathbf{G}(\hat{\mathbf{r}}) + \mathbf{G}_{\text{img}}(\hat{\mathbf{r}}) \ . \tag{4.108}$$

We will now show that the boundary conditions actually are satisfied for these cases, by choosing the electric current above the PEC as an example. If the current element $\mathbf{J}$ is infinitely small, the E-field from it at any point $\mathbf{r}_p$ on the surface is given by[28]

$$\mathbf{E}_i = C_k[\eta\mathbf{J}C_{N_1} - \eta(\mathbf{J}\cdot\hat{\mathbf{R}})\hat{\mathbf{R}}C_{N_2}]\frac{1}{R}e^{jkR} \ , \tag{4.109}$$

where we used $\mathbf{R} = \mathbf{r}_p - \mathbf{r}'$. At the same point the E-field of the image source becomes

$$\mathbf{E}_{\text{img}} = C_k[\eta\mathbf{J}_{\text{img}}C_{N1i} - (\eta\mathbf{J}_{\text{img}}\cdot\hat{\mathbf{R}}_i)\hat{\mathbf{R}}_iC_{N_{2i}}]\frac{1}{R_i}e^{-jkR_i} \ , \tag{4.110}$$

where $\mathbf{R}_i = \mathbf{r}_p - \mathbf{r}_{\text{img}}$. By using (4.101) and by studying Fig. 4.15 we get $\mathbf{r}_{\text{img}} - \mathbf{r}_p = \mathbf{r}' - \mathbf{r}_p - 2[(\mathbf{r}' - \mathbf{r}_p)\cdot\hat{\mathbf{n}}]\hat{\mathbf{n}}$ or

$$\mathbf{R}_i = \mathbf{R} - 2(\hat{\mathbf{R}}\cdot\hat{\mathbf{n}})\hat{\mathbf{n}} \ . \tag{4.111}$$

This means that $\mathbf{R} = \mathbf{R}_i$, and consequently that $C_{N_1} = C_{N_{1i}}$, $C_{N_2} = C_{N_{2i}}$.

Let us first consider the case where $\mathbf{J}$ is parallel with $\hat{\mathbf{n}}$. Then, $\mathbf{J}_{\text{img}} = \mathbf{J} \propto \hat{\mathbf{n}}$. The tangential components of the total E-field at the surface can be evaluated by taking $\hat{\mathbf{n}}\times(\mathbf{E}_i + \mathbf{E}_{\text{img}})$. This operation involves the following vector operations [when using $\mathbf{E}_i$ from (4.109) and $\mathbf{E}_{\text{img}}$ from (4.110)];

$$\hat{\mathbf{n}}\times\hat{\mathbf{n}} = 0 \ , \qquad \hat{\mathbf{n}}\times(\hat{\mathbf{n}}\cdot\hat{\mathbf{R}})\hat{\mathbf{R}} \ , \qquad \hat{\mathbf{n}}\times(\hat{\mathbf{n}}\cdot\hat{\mathbf{R}}_i)\hat{\mathbf{R}}_i \ .$$

We see from (4.111) that $\hat{\mathbf{n}}\cdot\hat{\mathbf{R}}_i = -(\hat{\mathbf{n}}\cdot\hat{\mathbf{R}})$ and $\hat{\mathbf{n}}\times\hat{\mathbf{R}}_i = \hat{\mathbf{n}}\times\hat{\mathbf{R}}$, so that

$$\hat{\mathbf{n}}\times(\hat{\mathbf{n}}\cdot\hat{\mathbf{R}})\hat{\mathbf{R}} = \hat{\mathbf{n}}\times(\hat{\mathbf{n}}\cdot\hat{\mathbf{R}}_i)\hat{\mathbf{R}}_i \ .$$

Using these we finally get $\hat{\mathbf{n}}\times(\mathbf{E}_i+\mathbf{E}_{\text{img}}) = 0$ at the image surface, and the boundary condition on the PEC is satisfied.

[28] Compare this with Eq. (4.37).

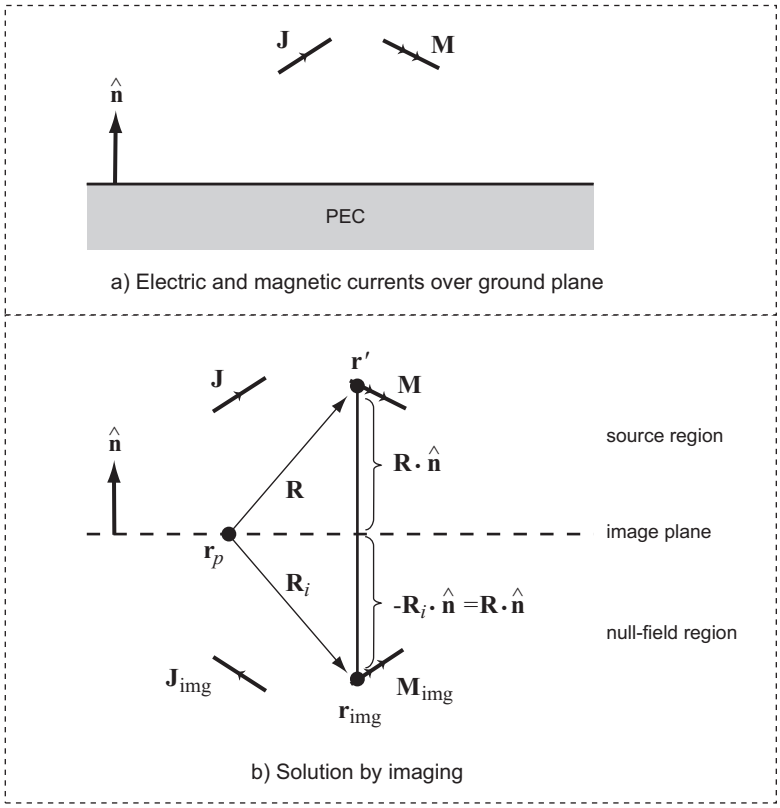


Figure 4.15: Imaging by electric and magnetic current sources in PEC ground plane.

Let us finally consider that $\mathbf{J}$ is orthogonal to $\hat{\mathbf{n}}$. Then, $\mathbf{J}_{\text{img}} = -\mathbf{J}$ and the vector operations needed to determine the tangential E-field at the ground plane become

$$\begin{aligned}\hat{\mathbf{n}} \times \mathbf{J} &= -\hat{\mathbf{n}} \times \mathbf{J}_{\text{img}} \ , \\ \hat{\mathbf{n}} \times (\mathbf{J} \cdot \hat{\mathbf{R}})\hat{\mathbf{R}} &= \hat{\mathbf{n}} \times (\mathbf{J}_{\text{img}} \cdot \hat{\mathbf{R}}_i)\hat{\mathbf{R}}_i \ .\end{aligned}$$

Using these, we also find that the boundary condition $\hat{\mathbf{n}} \times (\mathbf{E}_i + \mathbf{E}_{\text{img}}) = 0$ is satisfied. This proof could also have been done by visual inspection of the drawings in Fig. 4.15 and using the symmetry of the problems.

Thus, our imaging solutions satisfy the boundary conditions which means that they are valid and are unique solutions.

4.7 Integral equations and Method of Moments

We have in Section 4.2 expressed the fields as integrals over electric and magnetic currents located in free space. Actual field problems will also include a structure. However, even in the presence of a structure, it is always possible to define equivalent subproblems. These subproblems consist of equivalent or physical electric and magnetic sources radiating in an unbounded homogeneous material. Therefore, the field expressions for homogeneous regions play an important role even in formulating complicated field problems.

When using equivalent sources to formulate field problems we end up with integral equations which need to be solved in order to determine the equivalent sources. The integral equations are a result of enforcing the original boundary conditions of the fields at the boundaries, i.e., at the locations of the equivalent sources. Such integral equations are conveniently solved numerically by the *Method of Moments* [7]-[10]. Recall that this is not a book on numerical solutions by the Method of Moments. However, using a Method of Moments procedure, some of the classical integral formulas for antenna impedances and admittances can be derived more easily and with better physical insight. Therefore, we introduce this method here in order to use it in the analytic formulations in later sections[29].

We consider as an example an *integral equation* of the form

$$[\mathbf{E}_i(\mathbf{r}) + \mathbf{E}_{\text{S}}(\mathbf{r})]_{\tan} = 0 \qquad \text{for } \mathbf{r} = \mathbf{r}_{\text{S}}(u,v) \ , \tag{4.112}$$

which means that the tangential components of $\mathbf{E}_i(\mathbf{r}) + \mathbf{E}_{\text{S}}(\mathbf{r})$ are zero at the surface S described by $\mathbf{r}_{\text{S}}(u,v)$. $\mathbf{E}_i(\mathbf{r})$ is a known incident field on S, and

$$\mathbf{E}_{\text{S}}(\mathbf{r}) = \iint_{S'} \mathbf{J}(\mathbf{r}') \cdot \overline{\mathbf{G}}(\mathbf{r},\mathbf{r}')dS' \tag{4.113}$$

is the field scattered by the unknown induced current distribution $\mathbf{J}(\mathbf{r}')$, with S' the surface $\mathbf{r}' = \mathbf{r}_{\text{S}}(u,v)$, which is the same as the surface S over which the unknown current $\mathbf{J}(\mathbf{r}')$ is distributed, and $\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}')$ is the dyadic *Green's function* of the field problem. $\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}')$ may be the Green's function in free space (or generally homogeneous space), where we have

$$\mathbf{J}(\mathbf{r}') \cdot \overline{\mathbf{G}}(\mathbf{r},\mathbf{r}') = C_k[\eta\mathbf{J}(\mathbf{r}')C_{N_1} - (\eta\mathbf{J}(\mathbf{r}') \cdot \hat{\mathbf{R}})\hat{\mathbf{R}}C_{N_2}]\frac{1}{R}e^{-jkR} \ , \tag{4.114}$$

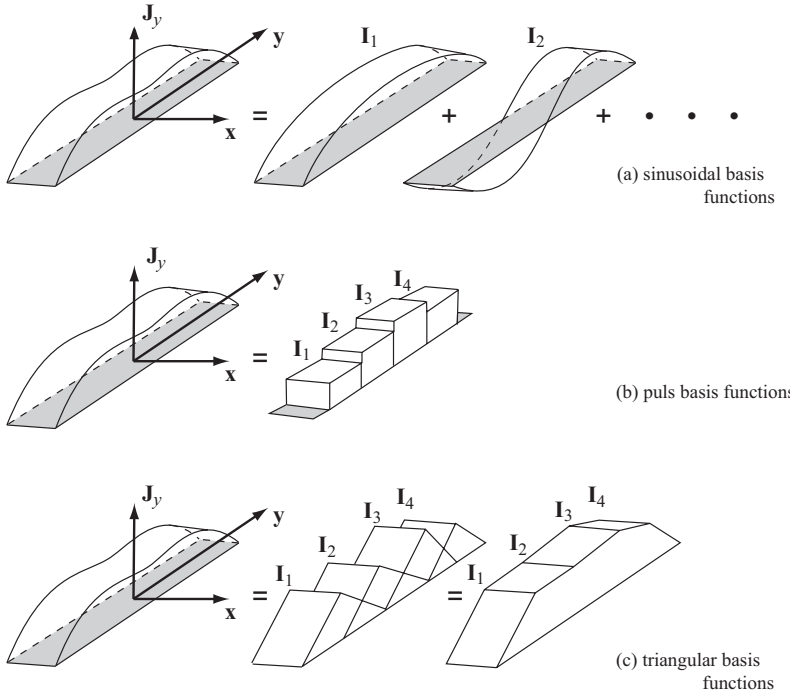


Figure 4.16: Examples of expansions of a current distribution $J_y\hat{\mathbf{y}}$ along a strip in different basis functions. (a) Sinusoidal entire domain basis functions for a Fourier series representation of J_y. (b) Pulse basis functions for a staircase approximation of J_y. (c) Triangular rooftop basis functions for a piecewise linear approximation of J_y.

as given in Section 4.2.1, where $R = |\mathbf{r} - \mathbf{r}'|$ and $\hat{\mathbf{R}} = (\mathbf{r} - \mathbf{r}')/|\mathbf{r} - \mathbf{r}'|$.

We want to determine $\mathbf{J}(\mathbf{r}')$ by solving the integral equation. The standard way to do this is to expand $\mathbf{J}(\mathbf{r}')$ in M *basis functions* $\mathbf{B}_m(\mathbf{r}')$ according to

$$\mathbf{J}(\mathbf{r}') = \sum_{m=1}^{M} I_m \mathbf{B}_m(\mathbf{r}') \ , \tag{4.115}$$

where I_ms are unknown coefficients. The basis function can be of different types. The most common ones are entire domain and subsectional basis functions (Fig. 4.16). Each entire domain basis function covers the whole surface S', and together they may correspond to, e.g., a one or two-dimensional Fourier series expansion of $\mathbf{J}(\mathbf{r}')$ over the surface S'. The subsectional basis functions covers different subsections of the source surface S', and they may or may not overlap with the neighboring basis functions. All together the subsectional basis functions cover the complete surface S' over which the unknown current is distributed. The subsectional basis functions may, e.g., be non-overlapping rectangular pulse functions located side by side, corresponding to a staircase approximation of $\mathbf{J}(\mathbf{r}')$. They may also be triangular rooftop functions which correspond to a staircase approximation of $\mathbf{J}(\mathbf{r}')$ in one direction and a piecewise linear approximation in the other. The piecewise linear approximation of $\mathbf{J}(\mathbf{r}')$ is obtained by letting neighboring triangular sections overlap to 50 %. The basis functions must cover all different components of $\mathbf{J}(\mathbf{r}')$. The Method of Moments solution converges more rapidly if the basis functions are chosen in a way that they represent the major physical characteristics of the current distribution, such as being zero at the ends of wires etc.

Now we insert the basis function expansion of $\mathbf{J}(\mathbf{r}')$ into the integral equation. The result is

$$[\mathbf{E}_i(\mathbf{r})]_{\tan} + \sum_{m=1}^{M} I_m [\mathbf{E}_{B_m}(\mathbf{r})]_{\tan} = 0 \qquad \text{on } S, \tag{4.116}$$

$$\text{with} \qquad \mathbf{E}_{B_m}(\mathbf{r}) = \iint_{S'} \mathbf{B}_m(\mathbf{r}') \cdot \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') dS' \ . \tag{4.117}$$

The above Eq. (4.116) states that the boundary condition must be satisfied continuously over the surface S. This is of course impossible when the current is approximated by a finite series. We have to introduce some kind of weighted average of the boundary condition. To this end, we introduce N weighting functions $\mathbf{W}_n(\mathbf{r}_S)$ over S[30] which are tangential to the surface. We multiply (4.116) with each of the and integrate over S. The result is

$$\sum_{m=1}^{M} I_m Z_{mn} = V_n \qquad \text{for } n = 1, \, N \ , \tag{4.118}$$

$$\text{with} \qquad Z_{mn} = \langle \mathbf{E}_{B_n}(\mathbf{r}_S), \mathbf{W}_n(\mathbf{r}_S) \rangle = \iint_S \mathbf{E}_{B_m}(\mathbf{r}_S) \cdot \mathbf{W}_n(\mathbf{r}_S) dS \ , \tag{4.119}$$

$$V_n = -\langle \mathbf{E}_i(\mathbf{r}_S), \mathbf{W}_n(\mathbf{r}_S) \rangle = -\iint_S \mathbf{E}_i(\mathbf{r}_S) \cdot \mathbf{W}_n(\mathbf{r}_S) dS \ . \tag{4.120}$$

The equations in (4.118) form a set of N linear equations to determine the M unknown expansion coefficients I_m. We need $N = M$ weighting functions in order to get a unique

[29] See Section 5.1.7, 5.4.2, and 6.2.

[30] This is also called test functions.

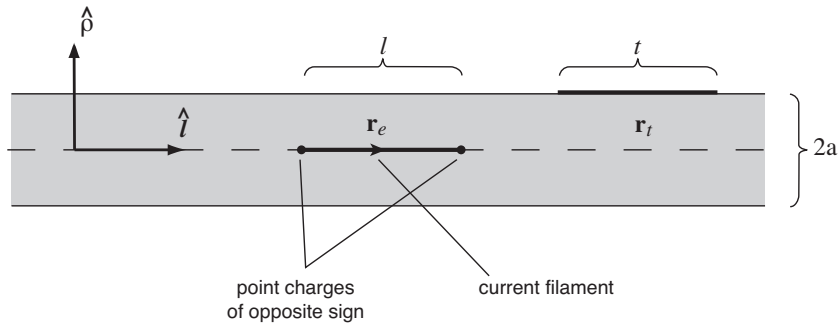


Figure 4.17: Illustration of cross section of wire of radius a and the locations of an incremental current segment, $\mathbf{r}_e$, and a test segment, $\mathbf{r}_t$. The current segment consists of a pulse current filament of length l, plus two point charges at either end.

solution for all I_m. One of the most accurate approaches is to choose the weighting functions equal to the basis functions, i.e., $\mathbf{W}_n(\mathbf{r}_S) = \mathbf{B}_n(\mathbf{r}_S)$, which is referred to as *Galerkin's method*. In this case we recognize Z_{mn} as the *reaction* between the two current sources $\mathbf{B}_m(\mathbf{r}')$ and $\mathbf{B}_n(\mathbf{r}_S)$. Z_{mn} gets the dimension ohm (Ω) and represents a *mutual impedance* if we choose all $\mathbf{B}_n(\mathbf{r}_S)$ to have dimension inverse meters (m^{-1}), so that I_m gets the dimension Ampere (A). Z_{mn} and V_n can be evaluated numerically. There are severe difficulties in evaluating Z_{mn} when $\mathbf{B}_m(\mathbf{r}')$ and $\mathbf{W}_n(\mathbf{r}_S)$ overlap. This is because the Green's function in (4.116) is singular when the field point $\mathbf{r}_S$ and the source point $\mathbf{r}'$ coincide. These singularities are referred to as *source point singularities*. Z_{mm} represents the *self-impedance* of each basis function $\mathbf{B}_m(\mathbf{r}_S)$.

Method of Moments solutions can be implemented numerically and several computer codes are available for analysis of, e.g., wire antennas, two-dimensional scatterers and bodies of revolution, and even three-dimensional scatterers. Infinite multilayer substrates and ground planes can be included via the Green's function[31]. The problem in implementing a Method of Moments solution is often related to the treatment of the source point singularities. In some cases it may be also difficult to get uniform convergence of the results when M increases due to numerical instabilities. In many Method of Moments implementations, there are also problems with structure resonances which give erroneous results at specific frequencies. However, in most cases the Method of Moments gives acceptable and accurate results. For this reason, it is often used as an "ideal" reference for other approximate calculation methods.

In the present book we will only use the Method of Moments with one or two basis functions, in order to derive analytic integral solutions to the input impedance of a dipole, a slot and a microstrip patch antenna. It is possible to get accurate results by using only one or two basis functions if they are chosen correctly. This requires experience and physical insight. For the three mentioned examples, the current distributions are known to be approximated well by half-wave sine or cosine functions. We will in the next two subsections present a numerical algorithm for evaluation of near-field couplings between subsectional basis function for wire and surface currents, and we will later use these to evaluate self-impedances of dipoles, slots and patches. Some basic literature about the Method of Moments in electromagnetics is given in the reference list, see [7]-[10].

[31] For more information see Section 6.3 on page 218.

4.7.1 Simple algorithm for near-field from line current

The vector integral forms of the E- and H-fields given in Section 4.2.1 have strong source point singularities. This fact makes them not suitable for evaluation of self-impedances and Method of Moments matrix elements of wire antennas. Therefore, we here present a near-field formula which is far more convenient for numerical evaluation. We will only show the formula for the E-field, as this is the only one needed when computing impedances of wire antennas.

Consider the same line current $I_0\hat{\mathbf{l}}$ of incremental length $l \ll \lambda$, which was introduced in (4.64), and locate it with its center at the point $\mathbf{r}_l$, see Fig. 4.17. Then, we want to evaluate the field at a location $\mathbf{r}_t + a\hat{\boldsymbol{\rho}}$, at the surface of the wire, where a is the radius of the wire, $\mathbf{r}_t$ is at its center line, and $\hat{\boldsymbol{\rho}}$ is orthogonal to the direction $\hat{\mathbf{t}}$ of the wire, i.e, $\hat{\boldsymbol{\rho}} \perp \hat{\mathbf{t}}$. We need the $\hat{\mathbf{t}}$ component of the E-field at $\mathbf{r}_t + a\hat{\boldsymbol{\rho}}$. This becomes, by assuming a *thin wire approximation*,

$$\mathbf{E}(\mathbf{r}_t + a\hat{\boldsymbol{\rho}}) \cdot \hat{\mathbf{t}}t = C_k \eta I_0 t \left\{ \hat{\mathbf{l}} \cdot \hat{\mathbf{t}} \Psi_{_J} - \frac{1}{k^2 t} \Psi_q(t) \right\} , \tag{4.121}$$

where $\mathbf{l} = l\hat{\mathbf{l}}$,

$$\Psi_{_J} = \frac{1}{l} \int_{-l/2}^{l/2} \frac{1}{R(l')} e^{-jkR(l')} dl' \qquad \text{with} \qquad R(l') = |(\mathbf{r}_t + a\hat{\boldsymbol{\rho}}) - (\mathbf{r}_l + l'\hat{\mathbf{l}})| , \tag{4.122}$$

$$\Psi_q(t) = \frac{1}{R_{\rm pp}} e^{-jkR_{\rm pp}} - \frac{1}{R_{\rm pm}} e^{-jkR_{\rm pm}} - \frac{1}{R_{\rm mp}} e^{-jkR_{\rm mp}} + \frac{1}{R_{\rm mm}} e^{-jkR_{\rm mm}} , \tag{4.123}$$

with 5.16[32]

$$\begin{aligned} R_{\rm pp} &= |(\mathbf{r}_t + (t/2)\hat{\mathbf{t}} + a\hat{\boldsymbol{\rho}}) - (\mathbf{r}_l + (l/2)\hat{\mathbf{l}})| , \\ R_{\rm pm} &= |(\mathbf{r}_t + (t/2)\hat{\mathbf{t}} + a\hat{\boldsymbol{\rho}}) - (\mathbf{r}_l - (l/2)\hat{\mathbf{l}})| , \\ R_{\rm mp} &= |(\mathbf{r}_t - (t/2)\hat{\mathbf{t}} + a\hat{\boldsymbol{\rho}}) - (\mathbf{r}_l + (l/2)\hat{\mathbf{l}})| , \\ R_{\rm mm} &= |(\mathbf{r}_t - (t/2)\hat{\mathbf{t}} + a\hat{\boldsymbol{\rho}}) - (\mathbf{r}_l - (l/2)\hat{\mathbf{l}})| . \end{aligned} \tag{4.124}$$

We have, also in (4.121), weighted the field with the length $t \ll \lambda$ of a subsectional test function. The expression for $\Psi_{_J}$ can be integrated numerically for all locations of $\mathbf{r}_t$ since we have avoided the source point singularity by introducing the wire radius. It can also be conveniently approximated by

$$\Psi_{_J} = \frac{1}{r} e^{-jkr} \qquad \text{with} \qquad r = \sqrt{|\mathbf{r}_t - \mathbf{r}_l|^2 + a^2} . \tag{4.125}$$

The above equations are obtained from [7, Sec. 4-2 and 4-3], except that we have introduced a simpler approximation of $\Psi_{_J}$. This is valid even for $\mathbf{r}_t = \mathbf{r}_l$ provided $l \ll a$, which means that it converges when l decreases for a given wire radius. $\Psi_{_J}$ originates from the line current, the first two terms in (4.123) represent the negative gradient of the potential originating from the point charge at the end $\mathbf{r}_l + (l/2)\hat{\mathbf{l}}$ of the line current, and the last two terms represent the negative gradient of the potential originating from the negative charge at its opposite end $\mathbf{r}_l - (l/2)\hat{\mathbf{l}}$.

[32] We recognize (4.122) as originating from (4.22) and (4.24), and (4.123) as a finite difference approximation of the gradient of the scalar potential in (4.30) (4.31).

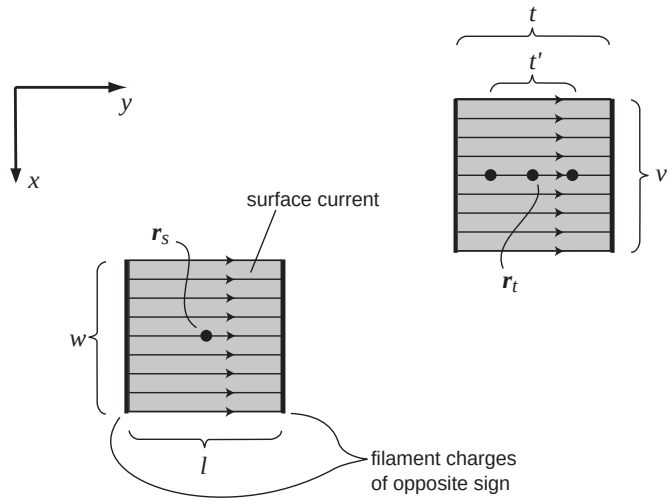


Figure 4.18: Illustration of surface current segment of width w and length l, and test segment of width v and length t.

For $l \leq 0.005$, the algorithm gives plausible accuracy even for the reactive field components when computing impedances and admittances. It has been used to produce the results shown in Fig. 5.9 and Fig. 5.16[33].

4.7.2 Simple algorithm for near-field from surface current

Now we will present a similar formula for convenient evaluation of near-fields of surface current segments. The formula is similar to that presented in [11]. But instead of pulse basis functions for charges, we have chosen line charge distributions at the edges of the pulse current basis functions.

Consider an incremental y-directed surface current segment located at $\mathbf{r}_{_S}$ and of width w and length l in the xy-plane, as shown in Fig. 4.18, i.e.,

$$\mathbf{J}(\mathbf{r}_{_S} + x\hat{\mathbf{x}} + y\hat{\mathbf{y}}) = J_{_0}\hat{\mathbf{y}} \qquad \text{for } -\frac{w}{2} < x < \frac{w}{2}\ ,\ -\frac{l}{2} < y < \frac{l}{2}\ . \tag{4.126}$$

We want to evaluate the field at a location $\mathbf{r}_t$, which may or may not coincide with $\mathbf{r}_{_S}$. We here limit the formula to the y-component of the E-field, and weight this with the area $A = tv$ of a subsectional test function. When $\mathbf{r}_t$ and $\mathbf{r}_{_S}$ coincide, we enforce $t = l$ and $v = w$.

Then, correspondingly to the formula in the previous subsection, we have,

$$\mathbf{E} \cdot \hat{\mathbf{y}} tv = C_k \eta J_{_0} tv \left\{ lw\Psi_{_J} - \frac{w}{k^2 t}\Psi_q(t) \right\}\ , \tag{4.127}$$

where

$$\Psi_{_J} = \frac{1}{lw}\int_{-l/2}^{l/2}\int_{-w/2}^{w/2} \frac{1}{R(x',y')} e^{-jkR(x',y')} dx'dy'\ , \tag{4.128}$$

[33] There exist MATLAB code for all figures of which the caption start with *.

$$\text{with} \qquad R(x',y') = |\mathbf{r}_t - (\mathbf{r}_S + x'\mathbf{x} + y'\mathbf{y})| \tag{4.129}$$

$$\text{and} \qquad \Psi_q(t) = \Psi_{\text{wpp}} - \Psi_{\text{wpm}} - \Psi_{\text{wmp}} + \Psi_{\text{wmm}} \ , \tag{4.130}$$

$$\text{with} \qquad \Psi_{\text{wpp}} = \frac{1}{w}\int_{-w/2}^{w/2} \frac{1}{R_{\text{pp}}(x')} e^{-jkR_{\text{pp}}(x')} dx' \tag{4.131}$$

and similarly for Ψ_{wpm}, Ψ_{wmp} and Ψ_{wpp} with

$$\begin{aligned} R_{\text{pp}}(x') &= |(\mathbf{r}_t + (t'/2)\hat{\mathbf{t}}) - (\mathbf{r}_S + (l/2)\hat{\mathbf{y}} + x'\hat{\mathbf{x}})| \ , \\ R_{\text{pm}}(x') &= |(\mathbf{r}_t + (t'/2)\hat{\mathbf{t}}) - (\mathbf{r}_S - (l/2)\hat{\mathbf{y}} + x'\hat{\mathbf{x}})| \ , \\ R_{\text{mp}}(x') &= |(\mathbf{r}_t - (t'/2)\hat{\mathbf{t}}) - (\mathbf{r}_S + (l/2)\hat{\mathbf{y}} + x'\hat{\mathbf{x}})| \ , \\ R_{\text{mm}}(x') &= |(\mathbf{r}_t - (t'/2)\hat{\mathbf{t}}) - (\mathbf{r}_S - (l/2)\hat{\mathbf{y}} + x'\hat{\mathbf{x}})| \ . \end{aligned} \tag{4.132}$$

The above integrals can readily be evaluated numerically, except for the *three cases* when $\mathbf{r}_t = \mathbf{r}_S$ and $\mathbf{r}_t = \mathbf{r}_S \pm ((l/2) + (t/2))\hat{\mathbf{l}}$. In the first case, the integrands of $\Psi_{_J}$, Ψ_{wpp} and Ψ_{wmp} have singularities. We can avoid the singularity problem in $\Psi_{_J}$ by transforming the $\Psi_{_J}$ integral to the polar coordinates, and evaluate it numerically by discretizing ϕ according to

$$\Psi_{_J} = \frac{4}{lw}\sum_{n=1}^{N} \Delta\phi \frac{j}{k}[e^{-jkR(\phi_n)} - 1] \ ; \qquad \Delta\phi = \frac{\pi}{2N} \ , \tag{4.133}$$

with

$$\phi_n = \frac{1}{2}\Delta\phi + \Delta\phi(n-1) \qquad \text{for } n = 1, 2, ..., N \tag{4.134}$$

$$R(\phi_n) = \begin{cases} \frac{w}{2\cos\phi_n} & \text{for } 0 \ < \phi_n < \phi_{_0} \\ \frac{l}{2\sin\phi_n} & \text{for } \phi_{_0} < \phi_n < \pi/2 \end{cases} \ , \tag{4.135}$$

$$\phi_{_0} = \arctan(l/w) \ .$$

The singularity problem in Ψ_{wpp} and Ψ_{wmm} can be avoided by introducing $t' < t$, see Fig. 4.18, for instance $t' = 0.5t$, and instead evaluating the field by using

$$\mathbf{E}\cdot\hat{\mathbf{y}}tv = C_k\eta J_{_0} tv \left\{ lw\Psi_{_J} - \frac{w}{k^2 t'}\Psi_q(t') \right\} \ . \tag{4.136}$$

Note that the first t in the equation, which represent the length of the test function (i.e., the weight), is kept to its original value, whereas the second and third t are changed to t'. The first t is used to evaluate the gradient of the potentials from the line charges at the ends of the surface current element, and this can of course be chosen independently of the length of the test function.

In the second and third cases, we may correspondingly use (4.136) to avoid the singularities in Ψ_{wpm} or Ψ_{wmp}.

For $l \leq 0.0025$, $N = 1000$ and $t' = t/2$, the algorithm gives acceptable accuracy even for the reactive field components when computing impedances and admittances. It has been used to produce the results in Fig. 5.9 and Fig. 5.24*.

4.8 Complementary comments by S. Maci

The Equivalence theorem is the modern electromagnetic extension of the classical Huygens principle of the XVII century. This principle asserts that *"each element of wavefront can be regarded as a center of secondary disturbance which gives rise to spherical wavelets ... the position of the wavefront at any later time is the envelope of all such wavelets"* [12]. Fresnel extended the Huygens theorem in 1870, leading to the formulation of the so-called Huygens-Fresnel principle, which is the foundation of the theory of diffraction of light. Fresnel supplemented Huygens construction with the postulate that the secondary wavelets can mutually interfere when the wavefront is obstructed by a surface located between the source and the observer. Kirchhoff gave these ideas a mathematical basis in 1883 [13], by showing that the Huygens-Fresnel principle may be regarded as an approximate asymptotic form of a radiation integral that expresses the solution of the homogeneous wave equation. The integral theorem of Kirchhoff can be regarded as the scalar version of the radiation integral of the equivalence theorem. The first modern version of this theorem for electromagnetic waves (Section 4.3.2) was published by A.E.H. Love [14] in 1901 and rigorously proven by McDonald in 1911 [15]. Another physically appealing proof of this theorem is due to Schelkunoff [16]. As a result of the above, the equivalence theorem is today most often referred as Love's equivalent theorem or the Love-Schelkunoff equivalence theorem.

A general treatment of all types of equivalence theorems can be found in [17]. Here, Harrington poses the emphasis on the fact that it is somewhat restrictive to impose the null field inside the equivalence surface S as Love did in his formulation. Any other field would serve as well, leading to the definition of an infinite set of equivalent currents as far as the external region is concerned. Any type of internal field can be constructed, provide the equivalent currents are defined by the jump of the tangential electric and tangential magnetic field at the surface. As a particular choice, this allows for defining a particular set of electric currents only (as well as magnetic current only) radiating in free space, which are equivalent for the external field [18].

Section 4.2.5 presents the expression to replace an electric (magnetic) current distribution with a magnetic (electric) current distribution that radiates exactly the same field. This can also be understood as the existence of non-radiating sources, obtained as the sum of an electric (magnetic) current distribution and its magnetic (electric) replacement current distribution with the opposite sign. The existence of non-radiating current distributions is on the other hand embedded in Love's formulation of the equivalence theorem, which makes the field radiated by the equivalent currents zero inside the equivalence surface.

Section 4.7 summarizes the Method of Moments (MoM) method based on Electric Field Integral Equation (EFIE) for antenna problems. It has been mentioned that the major difficulty in implementing this method is the hyper-singular reaction integral representing the self-impedance. The hyper-singularity is in the dyadic Green's function of the electric field occurring in the kernel. This problem may be overcome by using the Mixed Potential Integral Equation (MPIE) approach [19], where the electric field at the surface boundary is represented by vector and scalar potentials. In this way, the MoM impedances are reduced to the summation of two contributions. The first contribution is the reactions between weight-function currents and vector potentials produced by basis-currents; the second contribution is the reaction integral between weight-function charges and scalar potentials produced by basis-function charges. Both the resulting integrals of the MPIE self-impedances contain

singularity of order one, easy to be calculated. MPIE formulation for multilayer Green's functions, useful for patch antenna problems, is presented in [20]. The MPIE decomposition requires that the vector basis/weight functions are div-conforming, namely their divergence possess a zero at the line-boundary of the subdomain. This property should be satisfied by any MoM Galerkin basis. In particular, the most used div-conforming basis-functions in MoM codes, are the so called Rao-Wilton-Glisson (RWG) basis functions [21], which, thanks to their triangular domain, are adaptable to any type for curved surfaces. Another impairment of MoM is its breakdown for low frequencies. This can be overcome by decomposing the surface currents in a solenoidal and a non-solenoidal remainder [21]-[22].

The direct inversion of a MoM matrix of dimension N requires N^3 numerical operations. Since the density of the sub-domain basis functions should be of the order of ten in a wavelength, the direct inversion becomes rapidly not affordable when increasing the dimension of the antenna in terms of wavelengths. Iterative solutions should be invoked for a more efficient inversion of the problem. This inversion is greatly alleviated if one finds a good pre-conditioner for the specific problem [22]-[23]. For very large antenna problems, MoM can be formulated by the Fast Multiple Method (FMM) [24] or by its multilevel version (MLFMM) [25]. This produces a drastic sparsification of the MoM matrix, which reduces the asymptotic complexity of the problem to $N \log N$ operations where N is the dimension of the MoM matrix. Alternatively, one can reduce the size by compressing the MoM matrix. This compression is obtained by numerical generation from RWG functions of quasi-orthogonal synthetic functions based on a singular value decomposition of the solution currents from a set of independent wave excitations. Depending on the type of wave excitation, different method can be defined [26]-[27].

4.9 Exercises to Chapter 4

1. **Incremental sources:** Evaluate and compare the relative cross-polar levels in dB for the linearly y-polarized incremental electric and magnetic current sources and the Huygens source, at $\theta = 20^\circ$ in the $\varphi = 0^\circ$, 45° and 90° planes. Compare also the level of the back radiation at $\theta = 180^\circ$.

2. **Incremental sources:** Derive the radiation field expressions for RHC polarized incremental electric and magnetic current sources, as well as for the Huygens source. Find and compare the relative cross-polar levels in dB at $\theta = 20^\circ$ in the $\varphi = 0^\circ$, 45° and 90° planes.

3. **Incremental dipoles for circular polarization:** Consider two orthogonal incremental dipoles, located at the same point in space; one pointing in $\hat{\mathbf{x}}$ direction with amplitude I_0 and 0°-phase and the other pointing in $\hat{\mathbf{y}}$ direction with amplitude I_0 and 90°-phase. Derive the expressions for the radiation field. Determine the expressions for the co- and cross-polar patterns for a desired circular polarization. Sketch the two patterns in a figure. What is the relative level of the highest cross-polar sidelobe?

4. **Hard boundary condition:** Extra exercise for those interested in mathematics: Derive the hard boundary condition in Section 4.1.4 from the general boundary conditions in Section 4.1.2 for the TE_z case of a 2D field problem.

5. **Magnetic currents:** Consider a small rectangular radiating slot in an infinite ground plane. Assume that the E-field in the slot is $\mathbf{E} = E_0 \cos(\pi x/(2a))\hat{\mathbf{y}}$ for $-a < x < a$ and $-b < y < b$ where $a \ll \lambda$ and $b \ll \lambda$.

a) Explain how you can find the radiation field of the slot by using the PEC equivalent, imaging and the radiation field of a magnetic current distribution in free space. Write the expression for the radiation field.

b) Find the directivity.

c) Is the boundary condition at the PEC soft or hard in E-plane? What about in H-plane?

6. **Circularly polarized Huygens source:** Derive the expressions for the radiation field of the incremental Huygens source from those of the incremental electric and magnetic dipoles. Construct a Huygens source for circular polarization. Find the co- and cross-polar radiation patterns in the latter case.

7. **Imaging of vertical electric monopole:** Consider a vertical short electric current source on an infinite ground plane. This can, for instance, be the center conductor of a coaxial line coming out of a hole in the ground plane.

   a) Use imaging to find the far-field function, and find thereafter the directivity when we assume that the vertical source is infinitesimal.

   b) Is the boundary condition of this field soft or hard at the PEC?

8. **Imaging of horizontal dipole:** Consider an horizontal incremental electric dipole located d above an infinite ground plane.

   a) Derive the expression for the radiation field by using imaging when $d = \lambda/4$.

   b) Is the boundary condition at the PEC soft or hard in H-plane?

   c) Why is the far-field zero at the PEC in E-plane? Study the near-field. Is the boundary condition actually soft or hard at the PEC in this plane?

4.10 References

[1] R.F. Harrington, *Time-harmonic Electromagnetic Fields*, McGraw- Hill, 1961.

[2] *Reference Data for Radio Engineers*, Howard W. Sams & Co., Inc., a subsidiary of ITT, New York, 1977.

[3] J.J. Bowman, T.B.A. Senior and P.L.E. Uslenghi, *Electromagnetic and Acoustic Scattering by Simple Shape*, Wiley, 1969.

[4] P.-S. Kildal, A.A. Kishk, A. Tengs, "Reduction of forward scattering from cylindrical objects using hard surfaces", *IEEE Transactions on Antennas Propagation*, Vol. AP-44, pp. 1509–1520, November 1996.

[5] R.G. Kouyoumjian and P.H. Pathak, "A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface", *Proceedings of the IEEE*, Vol. 62, No. 11, pp. 1448-1461, November 1974.

[6] P.-S. Kildal, "Artificially soft and hard surfaces in electromagnetics", *IEEE Transactions of Antennas and Propagation*, Vol. 38, pp. 1537-1544, October 1990.

[7] R.F. Harrington, *Field Computation by Moment Methods*, Macmillan, New York, 1968.

[8] R.E. Collin and F.J. Zucker, *Antenna Theory Part 2*, McGraw-Hill, 1969.

[9] R.F. Harrington, "Origin and development of the method of moments for field computations", *Computational Electromagnetics*, Edited by: E.K. Miller, L. Medgyesi-Mitschang and E.H. Newman, IEEE Press, A selected reprint volume, New York 1992.

[10] J.J. Wang, "Generalized Moment Methods in electromagnetics", IEE Proceedings, Vol. 137, Pt. H, No. 2, pp. 127, April 1998.

[11] A.W. Glisson and D.R. Wilton, "Simple and efficient numerical methods for problems of electromagnetic radiation and scattering from surfaces", *IEEE Transactions on Antennas and Propagation*, Vol. AP-28, pp. 593- 603, September 1980.

[12] C. Huygens Traitè de Lumiere, Leyden, 1690. English translation by S.P. Thomson (London, Macmillan & Co., 1912).

[13] G. Kirchhoff, Berl. Ber, p. 641, 1882; Ann. D. Physik , 2, Bd. 18, p. 663, 1883; Ges. Abh. Nachtr., p. 22.

[14] A.E.H. Love, "The integration of equations of propagation of electric waves.", Phil. Trans., Roy. Soc. London, A, Vol. 197, pp. 1-45, 1901.

[15] H.M. MacDonald, "The Integration of the Eequations of Propagation of Electric Waves," Proc London Mathematical SocietySeries E, Vol. 10, pp. 91-95, 1911.

[16] S.A. Schelkunoff, "Some Equivalence Theorems of Electromagnetics and Their Application to Radiation Problems," Bell System Technical Journal, Vol. 15, pp. 102-112, January 1936.

[17] R.F. Harrington, Time Harmonic Electromagnetic Fields, McGraw-Hill, New York, 1961.

[18] E. Martini, G. Carli, S. Maci "An equivalence theorem based on the use of electric currents radiating in free space" IEEE Antennas and Wireless Propagation Letters, Vol. 7, pp. 421-424, November 2008.

[19] J.R. Mosig, "Arbitrarily shaped microstrip structures and their analysis with a mixed potential integral equation" Microwave Theory and Techniques, IEEE Transactions on 36.2 (1988): 314-323.

[20] K.A. Michalski, and J.R. Mosig. "Multilayered media Green's functions in integral equation formulations" Antennas and Propagation, IEEE Transactions on 45.3 (1997): 508-519.

[21] S. Rao, D.R. Wilton, and A.W. Glisson. "Electromagnetic scattering by surfaces of arbitrary shape." Antennas and Propagation, IEEE Transactions on 30.3 (1982): 409-418.

[22] R. Hiptmair, "Operator preconditioning." Computers and mathematics with Applications 52.5 (2006): 699-706.

[23] F. Andriulli, K. Cools, H. Bagci, F. Olyslager, A. Buffa, S. Christiansen, E. Michielssen, (2008). "A multiplicative Calderon preconditioner for the electric field integral equation." Antennas and Propagation, IEEE Transactions on, 56(8), 2398-2412.

[24] R. Coifman, V. Rokhlin, and S. Wandzura. "The fast multipole method for the wave equation: A pedestrian prescription." Antennas and Propagation Magazine, IEEE 35.3 (1993): 7-12.

[25] J. Song, C.-C. Lu, and W.C. Chew. "Multilevel fast multipole algorithm for electromagnetic scattering by large complex objects." Antennas and Propagation, IEEE Transactions on 45.10 (1997): 1488-1493.

[26] V.V.S. Prakash, and R. Mittra. "Characteristic basis function method: A new technique for efficient solution of method of moments matrix equations." Microwave and Optical Technology Letters 36.2 (2003): 95-100.

[27] L. Matekovits, A. Valeriu Laza, and G. Vecchi. "Analysis of large complex structures with the synthetic-functions approach." Antennas and Propagation, IEEE Transactions on 55.9 (2007): 2509-2521.

[28] L.B. Felsen and N. Marcuvitz, "Radiation and scattering of waves", Prentice-Hall, New Jersey, (1973).

Chapter 5

Small wire and slot antennas

In the previous chapter we considered the general theories about how to calculate radiation fields by integration over known current distributions (Chapter 4) and about how to characterize these radiation fields (Chapter 2). We will now start to apply the theories to practical antennas. Later on there will be also some general theory in Chapter 6 about apertures. The present chapter is devoted to small wire radiators such as the monopole, dipole, loop and helical antennas. We will also study radiation from a small slot in a ground plane when this is excited in different ways.

5.1 Electric monopole and dipole

We will in this section describe the *short wire antenna*. In its simplest form it is realized as a center conductor coming out at the end of a coaxial cable, and it is then called a monopole antenna. We will study the monopole when it is mounted on an infinite ground plane, obtained by penetrating the center conductor of the cable through a hole in a large metal plate and by connecting the shield of the cable (outer conductor) to the metal plate (Fig. 5.1). We will also study dipole antennas, which are wire antennas excited near the center of the wire. The excitation of the dipole is normally very complicated since a field transformation is needed from the rotationally symmetric coaxial feed line to the unsymmetrical dipole. Such transformations are done by so-called *baluns*, where the name is an abbreviation resulting from a transformation between a *bal*anced (two wire) line and a *un*balanced (coaxial) line representing the dipole. Two examples of dipoles above ground with different baluns are shown in Fig. 5.1. Several other examples of baluns can be found in the literature [1]-[2].

The present section is based on the theories of incremental electric currents[1]. We start by presenting the approximate current distribution on the monopole and the dipole. Thereafter we find the radiation resistance of the short dipoles. This resistance is very small, but increases with the length of the dipole, so, we need a certain length in order to get an

[1] For more information, see Section 4.4.1.

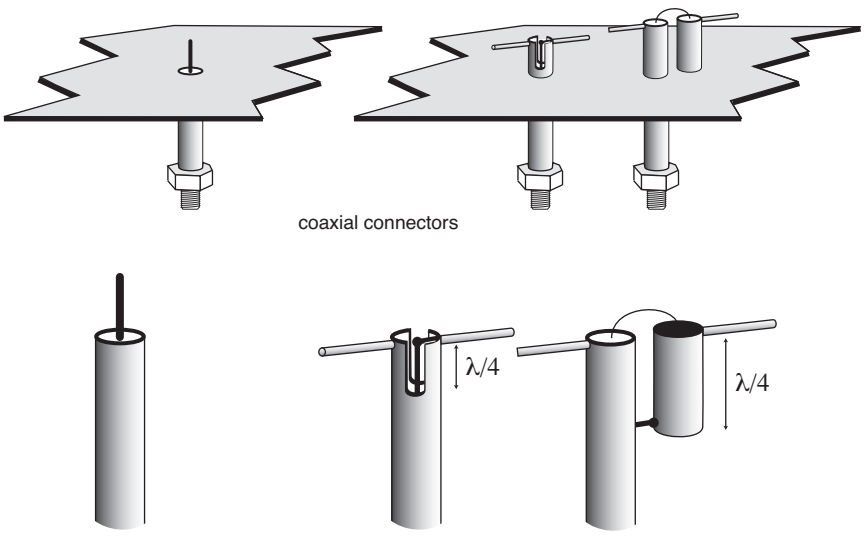


Figure 5.1: Monopole antenna on ground plane (left) and two dipoles on ground plane with different baluns (right).

impedance match to the transmission line. We will show that the optimum length in order to obtain impedance match to a 75 Ω cable is about half a wavelength.

The actual measured radiation impedance of a practical *half-wave dipole* will always differ from the theoretical value. The reason is that the balun and the interaction between the feed gap and the balun cannot be modelled accurately. It is even difficult to model the thickness of the dipole arms correctly. We will limit the analysis to the simplest possible, which is sufficiently accurate for predicting the radiation pattern. It is not accurate for the impedance, but normally advanced numerical methods do not give better results either. Therefore, the development of dipole antennas with good impedance match is always done experimentally. However, calculations are often used with success to estimate the relative change of the impedance with frequency and dimensions in order to facilitate the experimental tuning.

5.1.1 Approximate current distribution of monopole

We consider first a vertical and very thin *monopole* coming up from a little hole in a ground plane, as shown in Fig. 5.2. We assume that the wave on the coaxial line behind the ground plane and the fields in the coaxial hole induce a sinusoidal current distribution on the monopole, of the form

$$\mathbf{J}_l(z') = I_0 \sin\left[k\left(\frac{l}{2} - z'\right)\right] / \sin(kl/2)\hat{\mathbf{z}} \qquad \text{for } 0 < z' < l/2\ , \tag{5.1}$$

where $l/2$ is the length of the monopole. We have chosen the length $l/2$ instead of l in order to make the formula similar to that of the dipole in the next subsection. The normalization

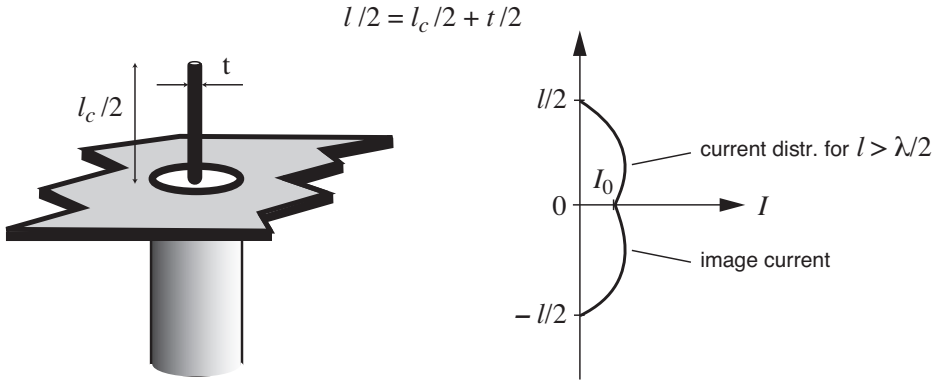


Figure 5.2: Geometry (left) and approximate current distribution (right) of monopole and its image.

in (5.1) is chosen in a way that the current is equal to I_0 in the coaxial opening ($z' = 0$) and zero at the end $z' = l/2$. This current distribution is empirical, but also expected from intuitive arguments as follows. An open coaxial line without any termination will cause total reflection with $I = 0$ at its end and a current varying sinusoidally from the end backwards on the cable. If we flare out the outer conductor, we can look upon the monopole on the ground plane as a conical coaxial continuation of the coaxial transmission line, and we may expect to maintain the sinusoidal current distribution. This expectation is not precise, because the conical coaxial line radiates, but the thin open cylindrical coaxial line almost does not. Accurate analysis by means of numerical methods also shows that the current distribution in reality is very similar to (5.1) for single monopoles on infinite ground planes provided that the monopole is thin and short. In practice the length $l/2$ is around $\lambda/4$ for which (5.1) is still an acceptable approximation. The current distribution gradually deteriorates from (5.1) when $l/2$ becomes longer than $\lambda/4$. See also the discussion at the end of the next subsection. When the reactive part of the input impedance is zero, the monopole is said to be resonant. This appears when $l/2 \approx \lambda/4$.

When the monopole has a finite thickness up to $t = 0.1\lambda$, the current distribution in (5.1) is still a useful approximation if we introduce an equivalent length $l/2 = l_c/2 + t/2$ where $l_c/2$ is the length of the metal cylinder representing the monopole, see Fig. 5.2.

By using (5.1) we also assume that the current is floating along the center of the wire instead of its surface which it does in reality. This assumption is commonly referred to as the *thin wire approximation*.

The far-field function including the effect of the infinite ground plane can be calculated by making use of imaging[2]. The result is a total current as

$$\mathbf{J}_l(z') = I_0 \sin\left[k\left(\frac{l}{2} - |z'|\right)\right] / \sin(kl/2)\hat{\mathbf{z}} \qquad \text{for } 0 < |z'| < l/2 \ . \tag{5.2}$$

[2] See Section 4.6.

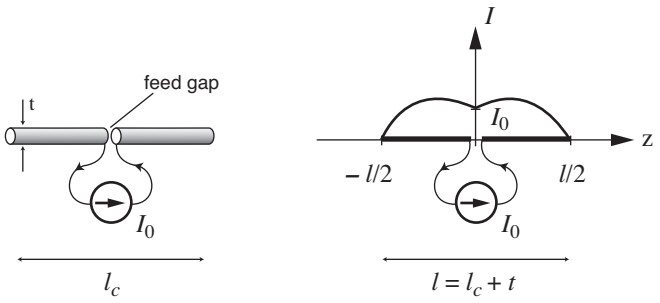


Figure 5.3: Geometry (left) and approximate current distribution (right) of dipole in free space.

5.1.2 Approximate current distribution of dipole

In practice a dipole consists of two aligned quarter wavelength long cylinders, which are called the dipole arms and most often have circular cross-section. The arms are separated by a feed gap where they are in metal contact with the feed cable or with the balun.

In analytical modeling it is common to replace the cylindrical dipole arms by line currents located along the center line of the arms (the *thin wire approximation*), and to assume that the effective length of the line current is $l = l_c + t$ with a short feed gap over which there is an electric current generator providing a constant current I_0 (Fig. 5.3). The variable t is the wire thickness. It is also common to neglect the effect of the connecting transmission line and the balun. The current distribution of this idealized dipole is approximately given by

$$\begin{aligned} & \mathbf{J}_l(z') = I_0 j(z')\hat{\mathbf{z}} \ , \\ \text{with} \qquad & j(z') = \sin\left[k\left(\frac{l}{2} - |z'|\right)\right] / \sin(kl/2) \qquad \text{for } |z'| < l/2 \ . \end{aligned} \tag{5.3}$$

Note that $\mathbf{J}_l = I_0\hat{\mathbf{z}}$ for $z' = 0$ and that $\mathbf{J}_l$ is not smooth at $z' = 0$ unless $l = \lambda/2$. In reality the current distribution will depend on the wire thickness t and length l, the size of the feed gap. Furthermore, it also depends on the surrounding structure, such as the feed line, the balun, a possible shield around the feed line and the balun, a possible ground plane, and possible neighboring dipoles in an array.

The current distribution in (5.3) is a good approximation for center-fed straight thin wire antennas in free space. Unfortunately, real antennas are not located in the free space. Still, the approximate current distribution is usable for the monopole (see Subsection 5.1.1) even when it is significantly longer than $\lambda/4$, provided it is located vertically on a large ground plane. The reason is that the environment of the monopole is very clean, resembling free space after imaging, for which the approximate current distribution is quite accurate. However, when studying half-wave dipoles the surroundings are not so clean. Even when neglecting the feed gap, feed line and balun, there will in most cases be a ground plane or neighboring dipoles present. In this case imaging in the ground plane does not correspond to a pure lengthening of the dipole like for the monopole. The ground plane causes an image dipole at another location which will effect the current distribution by mutual coupling. Still, (5.3) is usable as an approximation when $l < 0.7\lambda$ typically. It represents a good approximation up

to $l \approx 0.5\lambda$ even when there are other disturbing elements present. The most common length of practical dipole antennas is 0.5λ. A dipole is said to be resonant when the reactive part of its input impedance is zero. This appears when $l \approx 0.5\lambda$.

5.1.3 Far-field function of dipole

When the dipole is oriented in an arbitrary direction $\hat{\mathbf{l}}$, we can write

$$\begin{aligned} &\mathbf{J}_l(l') = I_0 j(l')\hat{\mathbf{l}} \ , \\ \text{with} \qquad &j(l') = \sin\left[k\left(\frac{l}{2} - |l'|\right)\right] / \sin(kl/2) \qquad \text{for } |l'| < l/2 \ . \end{aligned} \tag{5.4}$$

When the center of the dipole is located at the origin of the coordinate system, we obtain the following radiation field by using the equations in Section 4.2.2;

$$\mathbf{E}_d(\mathbf{r}) = \frac{1}{r} e^{-jkr} \mathbf{G}_d(\hat{\mathbf{r}}) \ ; \qquad \mathbf{G}_d(\hat{\mathbf{r}}) = \eta I_0 \mathbf{G}_{\mathrm{id}}(\hat{\mathbf{r}}) \tilde{j}(k\hat{\mathbf{l}} \cdot \hat{\mathbf{r}}) \ , \tag{5.5}$$

$$\text{where} \qquad \mathbf{G}_{\mathrm{id}}(\hat{\mathbf{r}}) = C_k[\hat{\mathbf{l}} - (\hat{\mathbf{l}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}] \tag{5.6}$$

$$\text{and} \qquad \tilde{j}(k\hat{\mathbf{l}} \cdot \hat{\mathbf{r}}) = \int_{-l/2}^{l/2} j(l') e^{jkl'\hat{\mathbf{l}}\cdot\hat{\mathbf{r}}} dl' \ , \tag{5.7}$$

with the incremental source constant $C_k = -jk/4\pi$. We see that the *far-field function* of the electric dipole is written as a product of three factors. They are:

1. The current excitation I_0 at the feed gap,
2. the far-field function $\mathbf{G}_{\mathrm{id}}(\hat{\mathbf{r}})$ of the incremental electric current with unit amplitude, which we will refer to as the incremental current factor, and
3. the *Fourier transform* $\tilde{j}(k\hat{\mathbf{l}} \cdot \hat{\mathbf{r}})$ of the current distribution along the dipole, which we will refer to as the current distribution factor $\tilde{j}$.

The latter is easily recognized as a Fourier transform by substituting $x = l'$ and $k_x = k\hat{\mathbf{l}} \cdot \hat{\mathbf{r}}$ in the exponent of the integrand. This *factorization* of the *far-field function* applies to all current distributions that are located along straight lines and radiate in free space.

The Fourier transform in (5.7) can be evaluated analytically by using the integral

$$\int e^{\alpha x} \sin(\beta x + \gamma) dx = e^{\alpha x} \frac{1}{\alpha^2 + \beta^2} [\alpha \sin(\beta x + \gamma) - \beta \cos(\beta x + \gamma)] \ .$$

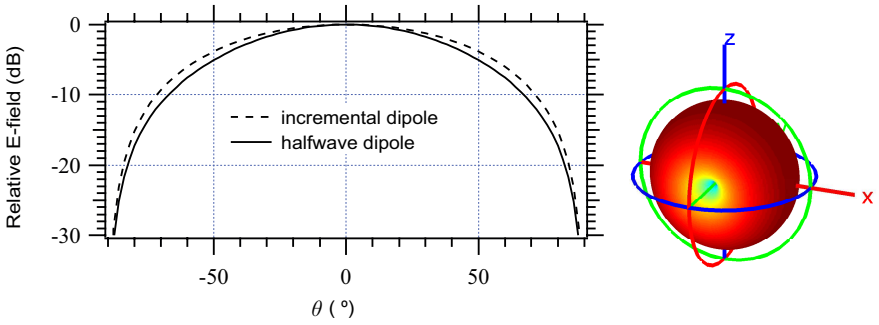
The result is

$$\tilde{j}(k\hat{\mathbf{l}} \cdot \hat{\mathbf{r}}) = \frac{2}{k} \frac{[\cos(kl\hat{\mathbf{l}} \cdot \hat{\mathbf{r}}/2) - \cos(kl/2)]}{[1 - (\hat{\mathbf{l}} \cdot \hat{\mathbf{r}})^2] \sin(kl/2)} \ . \tag{5.8}$$

This expression can conveniently be numerically evaluated for arbitrary orientations $\hat{\mathbf{l}}$. We will still restrict the expression to $l = \lambda/2$ and two specific orientations of $\hat{\mathbf{l}}$ in order to discuss the results.

Let us first consider a vertical dipole oriented in z-direction with $\hat{\mathbf{l}} = \hat{\mathbf{z}}$. This may correspond to a *monopole* over a ground plane. When $l = \lambda/2$, by using $\hat{\mathbf{l}} \cdot \hat{\mathbf{r}} = \cos\theta$ we get

$$\tilde{j}(k\hat{\mathbf{l}} \cdot \hat{\mathbf{r}}) = \tilde{j}(k\cos\theta) = \frac{2}{k} \cos\left(\frac{\pi}{2}\cos\theta\right) / \sin^2\theta \ . \tag{5.9}$$



(a) (b)

Figure 5.4: ⋆Radiation pattern of half-wave dipole: (a) Normalized E-plane pattern, (b) 3D plot when it is y-polarized on the z-axis.

Moreover by using $\hat{\mathbf{z}} - (\hat{\mathbf{z}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} = -\sin\theta\hat{\boldsymbol{\theta}}$, this gives the far-field function

$$\mathbf{G}_{\mathrm{dz}}(\theta) = -C_k \eta I_0 \frac{2}{k} \cos\left(\frac{\pi}{2}\cos\theta\right) / \sin\theta\hat{\boldsymbol{\theta}} \ . \tag{5.10}$$

We see that radiation field of the vertical dipole has complete rotational symmetry. It radiates as a BOR_0 type antenna, and it is applicable as an impressed source for exciting more complex BOR_0 *antennas* made of rotationally symmetric structures.

When the dipole is transverse y-directed with $\hat{\mathbf{l}} = \hat{\mathbf{y}}$ and $l = \lambda/2$ by using $\hat{\mathbf{l}} \cdot \hat{\mathbf{r}} = \sin\theta\sin\varphi$ we obtain

$$\mathbf{G}_{\mathrm{dy}}(\theta,\varphi) = C_k \eta I_0 (\cos\theta\sin\varphi\hat{\boldsymbol{\theta}} + \cos\varphi\hat{\boldsymbol{\varphi}})\tilde{j}(\theta,\varphi) \ , \tag{5.11}$$

$$\text{with} \qquad \tilde{j}(k\hat{\mathbf{y}}\cdot\hat{\mathbf{r}}) = \tilde{j}(\theta,\varphi) = \frac{2}{k}\cos\left(\frac{\pi}{2}\sin\theta\sin\varphi\right) / [1-(\sin\theta\sin\varphi)^2] \ . \tag{5.12}$$

We see that the H-plane pattern is constant, and that the E-plane pattern is $\tilde{j}(\theta, 90^\circ)\cos\theta = \cos((\pi/2)\sin\theta)/\cos\theta$. Fig. 5.4 shows the E-plane pattern and compare it with the corresponding pattern of the incremental dipole[3]. The half-wave dipole is seen to be only slightly more directive than the incremental dipole. The patterns of the short dipole is the same as those of the incremental dipole in (4.70), see also (5.19).

The radiation field of a y-directed dipole is very similar to a BOR_1 *antenna* with different E- and H-plane patterns. It is not precisely a BOR_1 antenna as defined in Section 2.4.2 because the current distribution factor given by (5.12) has a φ-variation in addition to the φ-variation of the incremental dipole factor. Still, the relative cross-polarization in a given φ-plane is only due to the incremental dipole factor. If we evaluate the co- and cross-polar patterns of (5.11), by using the polarization vectors in (2.55) - (2.56) and the BOR_1 relations in (2.83) - (2.87) we achieve

$$G_{\mathrm{co}} = \mathbf{G}_d(\theta,\varphi)\cdot\hat{\mathbf{co}}^* = G_{\mathrm{co}_{45^\circ}}(\theta,\varphi) - G_{\mathrm{xp}_{45^\circ}}(\theta,\varphi)\cos 2\varphi \ , \tag{5.13}$$

$$G_{\mathrm{xp}} = \mathbf{G}_d(\theta,\varphi)\cdot\hat{\mathbf{xp}}^* = G_{\mathrm{xp}_{45^\circ}}(\theta,\varphi)\sin 2\varphi \ , \tag{5.14}$$

[3] There exist MATLAB code for all figures of which the caption start with ⋆.

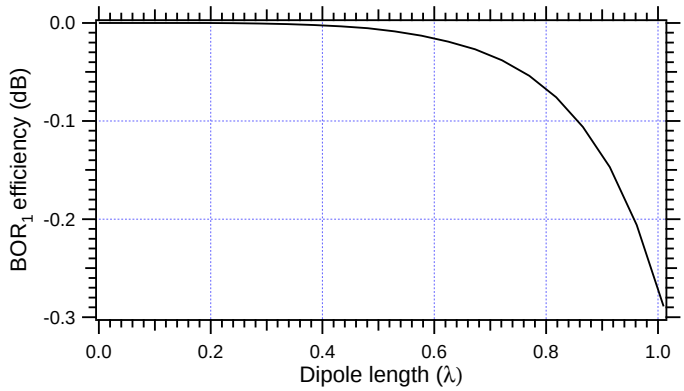


Figure 5.5: Relative power in the BOR_1 mode relative to the total radiated power of a transverse dipole.

$$\text{where} \qquad G_{\mathrm{co}_{45^\circ}}(\theta, 45^\circ) = C_k \eta I_0 \tilde{j}(\theta, 45^\circ)\frac{1}{2}(1+\cos\theta) \tag{5.15}$$

$$\text{and} \qquad G_{\mathrm{xp}_{45^\circ}}(\theta, 45^\circ) = C_k \eta I_0 \tilde{j}(\theta, 45^\circ)\frac{1}{2}(\cos\theta - 1)\ . \tag{5.16}$$

The equations (5.15) and (5.16) describe the co- and cross-polar patterns in the $\varphi = 45^\circ$-plane respectively. If the half-wave dipole had been an ideal BOR_1 antenna, all φ-variations had been in the $\cos 2\varphi$ and $\sin 2\varphi$ factors, i.e., we would have had $G_{\mathrm{co}_{45^\circ}}(\theta, \varphi) = G_{\mathrm{co}_{45^\circ}}(\theta)$ and $G_{\mathrm{xp}_{45^\circ}}(\theta, \varphi) = G_{\mathrm{xp}_{45^\circ}}(\theta)$.

However, if we study the φ variation of the current distribution factor $\tilde{j}(\theta, \varphi)$ we will find that it is negligible for $l < \lambda/2$, in particular near $\theta = 0^\circ$ and $\theta = 180^\circ$. Thus, we may often omit it. This is clear from Fig. 5.5 which shows the relative power in the first order BOR_1 mode compared to the total radiated power as a function of the length of the dipole*. Also, if the half-wave dipole is used to excite complex rotationally symmetric structures, the effect of the higher order φ-modes in the resultant pattern will almost always be negligible, so that the resulting antenna will be a BOR_1 type.

5.1.4 Directivity and radiation resistance of short dipole

When the dipole is much shorter than a half wavelength ($l \ll \lambda/2$), the current distribution in (5.4) becomes triangular. That is,

$$j_l(l') = (1 - 2|l'|/l) \qquad \text{for } |l'| < (l/2)\ . \tag{5.17}$$

Thus, the radiation integral in (5.7) becomes

$$\tilde{j}(k\hat{\mathbf{l}}\cdot\hat{\mathbf{r}}) \approx \int_{-l/2}^{l/2}(1-2|l'|/l)dl' = l - (2/l)(l/2)^2 = l/2\ . \tag{5.18}$$

This result can also be obtained by expanding (5.8) for small l. The far-field function becomes

$$\mathbf{G}_{\mathrm{id}}(\hat{\mathbf{r}}) = \eta I_0 (l/2) C_k[\hat{\mathbf{l}} - (\hat{\mathbf{l}}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}}]\ . \tag{5.19}$$

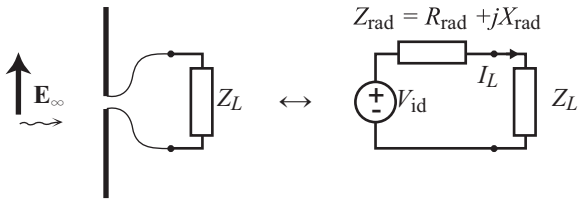


Figure 5.6: Short dipole (left) and its equivalent circuit on reception (right).

This is the same as that of an incremental electric current of length $l/2$. The reason for the discrepancy in the lengths is that the current distribution is constant along the incremental electric current and triangular along the present short dipole. To find the directivity, let us now solve the power integral P as defined in Section 2.3.8. The power integral must be the same for all orientations of the dipole, so we choose $\hat{\mathbf{l}} = \hat{\mathbf{z}}$ as this gives a rotationally symmetric beam which is simpler to integrate. Then, we obtain

$$P = \int_0^{2\pi} \int_0^{\pi} |\mathbf{G}_{\text{id}}|^2 \sin\theta d\theta d\varphi = |C_k \eta I_0 l/2|^2 \frac{8\pi}{3} \ , \tag{5.20}$$

by using that

$$\int_0^{\pi} \sin^3\theta d\theta = \int_0^{\pi} (\sin\theta - \cos^2\theta \sin\theta) d\theta = \left[-\cos\theta + \frac{1}{3}\cos^3\theta \right]_0^{\pi} = \frac{4}{3} \ .$$

From this, the directivity becomes

$$D_0 = \frac{4\pi |\mathbf{G}_{\text{id}}|^2_{\max}}{P} = \frac{4\pi |C_k \eta I_0 l/2|^2}{|C_k \eta I_0 l/2|^2 \frac{8\pi}{3}} = \frac{3}{2} \quad \text{(i.e., } 1.76\,\text{dBi)} \ . \tag{5.21}$$

The total radiated power can also be expressed as $P_{\text{rad}} = P/(2\eta) = (1/2)|I_0|^2 R_{\text{rad}}$, where R_{rad} is the radiation resistance[4]. This relation can be used to calculate R_{rad}, as we know P as a function of I_0 from (5.20). We have (by using also that $\eta \approx 377\,\Omega \approx 120\pi\,\Omega$)

$$R_{\text{rad}} = \frac{2P_{\text{rad}}}{|I_0|^2} = \eta \frac{2\pi}{3} \left(\frac{l}{2\lambda} \right)^2 \approx 20\pi^2 \left(\frac{l}{\lambda} \right)^2 \Omega \ . \tag{5.22}$$

We see that the radiation resistance is very small. When $l = 0.01\lambda$ we get $R_{\text{rad}} = 0.02\,\Omega$, and when $l = 0.1\lambda$ we have $R_{\text{rad}} = 2\,\Omega$. Coaxial cables have characteristic impedances of $50\,\Omega$ or $75\,\Omega$, so we understand that a short dipole will have very low radiation efficiency if it is fed by a coaxial cable. The short dipole has also a large reactance, which will be studied in Sectioin 5.1.7. The reactance is negative, which means that the short dipole is capacitive.

5.1.5 Equivalent circuit and maximum effective aperture of short dipole

The relation between the maximum effective aperture of a receiving antenna and the directivity when the antenna is transmitting was introduced in Section 2.5.2. This relation is

[4] See Section 2.6.1.

general and can be derived by using the reciprocity relations in Section 4.5.2. We study it for a short dipole by using the formulas in the previous Subsection, as follows. The *maximum effective area* $A_{\rm em}$ is defined by

$$A_{\rm em} = \frac{P_{_L}}{W_i} = \frac{\text{Power delivered to conjugate matched load}}{\text{Power density of incident wave}} \ . \tag{5.23}$$

The maximum power is delivered to the load when the load impedance $Z_{_L}$ is *conjugate matched* to the dipole impedance $Z_{\rm rad}$, i.e., when

$$Z_{_L} = Z_{\rm rad}^* = R_{\rm rad} - jX_{\rm rad} \tag{5.24}$$

all according to Section 2.6.1. The power delivered to the load is

$$P_{_L} = \frac{1}{2} R_{_L} |I_{_L}|^2 \ . \tag{5.25}$$

The current in the load can be calculated from the equivalent circuit in Fig. 5.6. Due to the conjugate match we find it to be

$$I_{_L} = \frac{V_{\rm id}}{2R_{\rm rad}} \ . \tag{5.26}$$

In Fig. 5.6 $V_{\rm id}$ is seen to be the induced voltage over the port of the short dipole when the load is removed. We now use (2.129) to express $V_{\rm id}$ in terms of the amplitude E_∞ of the incoming plane wave. The far-field function in (5.19) has the maximum $\eta I_{_0}(l/2)C_k$ in the direction normal to $\hat{\mathbf{l}}$, with $C_k = -jk/4\pi$. The latter together with (2.129) yield

$$V_{\rm id} = -2j\lambda(l/2)C_k E_\infty = -E_\infty(l/2) \ ,$$

where we have assumed that $\mathbf{E}_\infty = E_\infty \hat{\mathbf{l}}$.

Alternatively, by using the reaction integral in (4.82), we can also express $V_{\rm id}$ in terms of E_∞ as follows. Study the four cases (a) to (d) in Fig. 5.7. Assume that there is an antenna with a source I_∞ somewhere in infinity which produces a plane wave with E-field amplitude E_∞ at the location of the dipole (case (a)). This plane wave causes a voltage $V_{\rm id}$ over the input terminals of the dipole. If we instead excite the dipole with the current source $I_{_0}$, and we measure the voltage V_∞ over the terminals of the antenna at infinity (case (b)), from reciprocity we have[5]

$$-V_{\rm id} I_{_0} = -V_\infty I_\infty \ . \tag{5.27}$$

In cases (c) and (d) we have replaced the metal dipole arms by their line current $\mathbf{J}_l(l)$. $\mathbf{J}_l(l)$ has the value $I_{_0}$ at the center, so cases (b) and (c) produce the same field at infinity and hence the same voltage V_∞ at the terminal there. Therefore, the reaction between $\mathbf{J}_l(l)$ and $\mathbf{E}_\infty$ must be equal to the reaction between I_∞ and V_∞, i.e.,

$$\int_{-l/2}^{l/2} (\mathbf{E}_\infty \cdot \mathbf{J}_l(l'))dl' = -V_\infty I_\infty \ . \tag{5.28}$$

By combining (5.27) and (5.28) we obtain the following expression for $V_{\rm id}$

$$V_{\rm id} = -\frac{1}{I_{_0}} \int_{-l/2}^{l/2} (\mathbf{E}_\infty \cdot \mathbf{J}_l(l'))dl' = -\int_{-l/2}^{l/2} E_\infty j_l(l')dl' = -E_\infty l/2 \ . \tag{5.29}$$

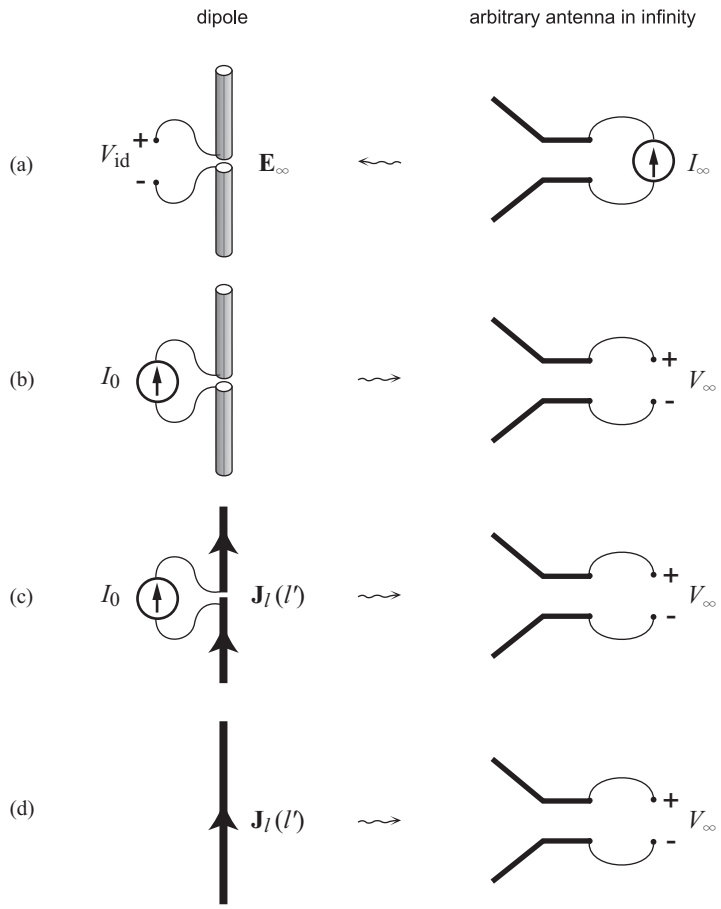


Figure 5.7: Illustration of the *reciprocity relations* in (5.27) and (5.28).

The latter two expressions in (5.29) are obtained by using (5.17) and aligning $\hat{\mathbf{l}}$ with $\mathbf{E}_\infty$. The result in (5.29) is similar to what we obtained by using (2.129) directly.

The power density of the incoming wave is

$$W_i = \frac{1}{2\eta}|E_\infty|^2 \ . \tag{5.30}$$

Finally, by inserting (5.25), (5.26) and (5.30) into (5.23) we arrive at:

$$A_{\text{em}} = \frac{\frac{1}{2}R_L|I_L|^2}{W_i} = \frac{|E_\infty l/2|^2 2\eta}{8R_{\text{rad}}|E_\infty|^2} = \frac{l^2\eta}{16}\left(\frac{12\lambda^2}{2\pi\eta l^2}\right) = \frac{3}{8\pi}\lambda^2 \ . \tag{5.31}$$

Therefore, the short dipole has an effective aperture of

$$A_{\text{em}} = \frac{3}{8\pi}\lambda^2 = 0.119\lambda^2 \ . \tag{5.32}$$

[5] See Section 4.5.3.

Recalling the directivity $D_0 = 3/2$, we see the following general relation holds:

$$A_{\rm em} = \frac{\lambda^2}{4\pi} D_0 \ , \tag{5.33}$$

which was introduced in Section 2.5.2.

5.1.6 Directivity and radiation resistance of half-wave dipole

Let us now consider the *half-wave dipole*. The power integral is most easily evaluated when the dipole is oriented vertically, in the same way as we chose for the short dipole. Then,

$$P_{\rm dp} = 2\pi \int_0^\pi |\mathbf{G}_d(\theta)|^2 \sin\theta d\theta \ . \tag{5.34}$$

This integral has an analytic solution, but this is so complicated that it does not have any advantage over a numerical solution, which is easily evaluated. The result gives the following directivity when $l = \lambda/2$:

$$D_0 = (4\pi|\mathbf{G}_d|^2_{\rm max})/P_{\rm dp} = 1.643 \quad (\text{i.e., } 2.16\,\text{dBi}) \ , \tag{5.35}$$

where $|\mathbf{G}_d|^2_{\rm max} = |\mathbf{G}_d(\pi/2)|^2$. Thus, the directivity of the half-wave dipole is 0.4 dB larger than that of the short dipole.

We can now find the *radiation resistance* $R_{\rm dp}$ as follows. The total radiated power $P_{\rm rad} = P_{\rm dp}/(2\eta)$ according to (2.65). Using (5.35), we can write

$$P_{\rm dp} = 4\pi|\mathbf{G}_d|^2_{\rm max}/D_0 \ ; \qquad |\mathbf{G}_d|^2_{\rm max} = |C_k \eta I_0 2/k|^2 \ ,$$

which was achieved, by setting $\theta = \pi/2$ in Eq. (5.10). The power dissipated in the radiation resistance is

$$P_{\rm rad} = (1/2) R_{\rm dp} |I_0|^2 \ .$$

Therefore,

$$R_{\rm dp} = \frac{2P_{\rm rad}}{|I_0|^2} = \frac{P_{\rm dp}}{\eta |I_0|^2} = \frac{4\pi}{D_0}\eta \left| C_k \frac{2}{k} \right|^2 = \frac{\eta}{\pi D_0} \approx 73\,\Omega \ . \tag{5.36}$$

This is a good number as it is close to the characteristic impedances of commercial coaxial cables which are 50 or 75 Ω. Therefore, the half-wave dipole can easily be matched to the feed cable with high radiation efficiency. The complete radiation impedance including the reactance will be found in the next subsection.

The half-wave dipole has approximately a sinusoidal current distribution independent of how it is excited and where in a structure it is located[6]. This is not true for dipoles of longer lengths, which have almost sinusoidal distribution only when they are excited by a narrow feed gap at the center and radiate in free space. However, even if the current distribution of a half-wave dipole is unchanged by a surrounding structure, its impedance will change due to the coupling.

[6] This was discussed in Section 5.1.2 on page 170.

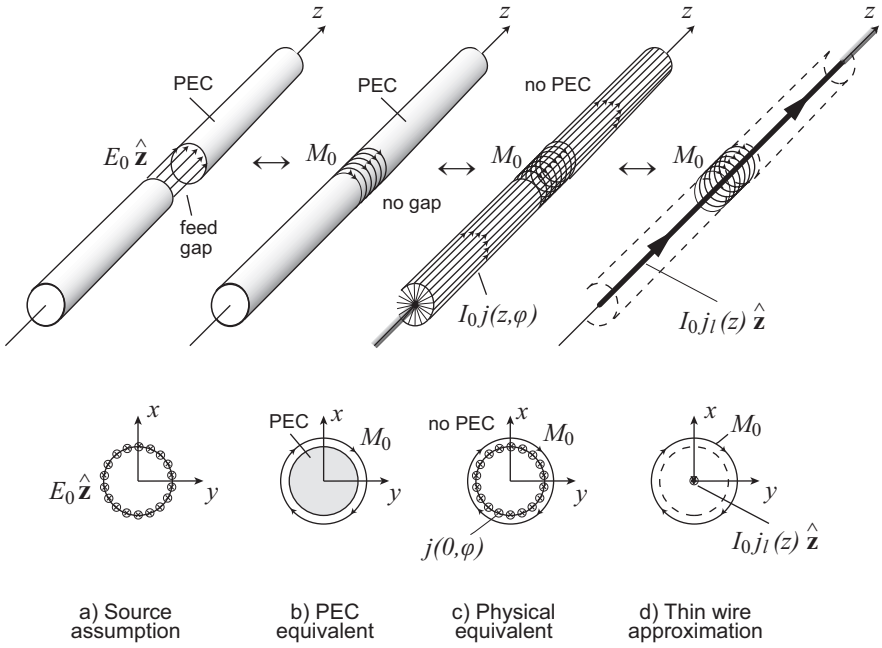


Figure 5.8: Development of the impedance calculation model for an electric dipole (upper) and its cross-section in the feed gap (lower).

5.1.7 Self-impedance of electric dipole

We will now use the Method of Moments and the reciprocity theorem to develop the classical formula for the full complex *radiation impedance* also called self-impedance of a dipole. The explanation will follow cases (a)-(d) in Fig. 5.8. We consider first the physical dipole in case (a), consisting of two metal (PEC) cylinders with a feed gap δ between them, and with an ideal current source I_0 in the feed gap connected to the centers of the cylinders. The current source produces an electric field E_0 over the feed gap, and it is clear that the self-impedance will be given as

$$Z_d = V_0/I_0 \ , \tag{5.37}$$

where $V_0 = -E_0\delta$ is the voltage induced over the feed gap.

In order to find the relation between E_0 and I_0 we need to introduce an electromagnetic analysis model. We do this by using the PEC equivalent to remove the feed gap and replace the field E_0 by a magnetic ring-shaped so-called frill current (case (b))

$$\mathbf{M}_0 = -E_0\hat{\mathbf{z}} \times \boldsymbol{\rho} = -E_0\hat{\boldsymbol{\varphi}} \qquad \text{for } -\delta/2 < z < \delta/2 \ . \tag{5.38}$$

Further, we use the physical equivalent to replace the metal cylinder by an induced electric surface current distribution $\mathbf{J}(z,\varphi)$ (case (c)). Next, we use the *thin wire approximation*. This means that we assume that $\mathbf{J}(z,\varphi)$ is rotationally symmetric and can be replaced by an axial line current $\mathbf{J}_l(z) = 2\pi a\mathbf{J}(z,\varphi) = I_0 j_l(z')\hat{\mathbf{z}}$ (case (d)). The current distribution $\mathbf{J}_l(z)$ is unknown, but can be determined by applying the boundary condition that the total field

must be zero at the surface of the cylinder, i.e.,

$$V_0\mathbf{E}_m(\mathbf{r}_a) + I_0\mathbf{E}_j(\mathbf{r}_a) = 0 \ , \tag{5.39}$$

for $-l/2 < z < l/2$ where $\mathbf{E}_m(\mathbf{r}_a) = \mathbf{E}_M(\mathbf{r}_a)/V_0$ is the field in free space at $\mathbf{r}_a = a\hat{\boldsymbol{\rho}} + z\hat{\mathbf{z}}$ due to a normalized magnetic current $\mathbf{m} = \mathbf{M}_0/V_0 = -\boldsymbol{\varphi}/\delta$. And, $\mathbf{E}_j(\mathbf{r}_a)$ is the field in free space due to the normalized current distribution $j_l(z)\hat{\mathbf{z}} = \mathbf{J}_l/I_0$. Eq. (5.39) defines also an integral equation for determining $j_l(z)$. For thin dipoles shorter than $l \approx \lambda/2$ we already know that the solution is quite accurately given by (5.3). Let us therefore continue by assuming (5.3) for $j_l(z)$, corresponding to solving the integral equation by using the *Method of Moments* with one basis function. The solution will give us the desired relation between I_0 and E_0. We do this by using Galerkin's method, i.e., by satisfying the boundary condition in weighted average over the interval $(-l/2, l/2)$ by using the basis function $j_l(z)$ also as weighting function. Doing this, we get

$$V_0\int_{-l/2}^{l/2}\mathbf{E}_m(\mathbf{r}_a)\cdot j_l(z)\hat{\mathbf{z}}dz + I_0\int_{-l/2}^{l/2}\mathbf{E}_j(\mathbf{r}_a)\cdot j_l(z)\hat{\mathbf{z}}dz = 0 \ . \tag{5.40}$$

We recognize the former of these two integrals as the reaction between the fields $\mathbf{E}_m$ (due to $\mathbf{m} = -\hat{\boldsymbol{\varphi}}/\delta$) and the source $j_l(z)\hat{\mathbf{z}}$. This reaction must according to reciprocity[7] be equal to the reaction between the H-field $\mathbf{H}_j$ (due to $j_l(z)\hat{\mathbf{z}}$) and the source $\mathbf{m} = -\hat{\boldsymbol{\varphi}}/\delta$, i.e.,

$$\langle\mathbf{E}_m(\mathbf{r}_a), j_l(z)\hat{\mathbf{z}}\rangle = \langle\mathbf{H}_j, \mathbf{m}\rangle \ . \tag{5.41}$$

We can find $\mathbf{H}_j$ as follows. Consider the actual current distribution $j(z,\varphi) = j_l(z)/(2\pi a)$ at the circumference of the dipole corresponding to $j_l(z)\hat{\mathbf{z}}$ along its center. The H-field at the surface and the current density are related by $j(z,\varphi)\hat{\mathbf{z}} = \hat{\boldsymbol{\rho}}\times\mathbf{H}_j$ (from Eq. (4.14)) where $\hat{\boldsymbol{\rho}}$ is the radial unit vector normal to the cylindrical surface of the dipole. This gives $\mathbf{H}_j = \hat{\boldsymbol{\varphi}}/(2\pi a)$ at $z = 0$. Furthermore, $\mathbf{m} = -(1/\delta)\hat{\boldsymbol{\varphi}}$, so the right side of (5.41) becomes

$$\langle\mathbf{H}_j, \mathbf{m}\rangle = -\int_{-\delta/2}^{\delta/2}\int_0^{2\pi}\mathbf{H}_j\cdot\mathbf{m}ad\varphi dz = 1 \ . \tag{5.42}$$

We finally obtain from (5.40) the expression for the self-impedance in free space

$$Z_d = \frac{V_0}{I_0} = -\int_{-l/2}^{l/2}\mathbf{E}_j(\mathbf{r}_a)\cdot j_l(z)\hat{\mathbf{z}}dz \ . \tag{5.43}$$

This expression corresponds to calculating Z_d by using the voltage

$$V_0 = \int_{-l/2}^{l/2}\mathbf{E}_j(\mathbf{r}_a)\cdot j_l(z)\hat{\mathbf{z}}dz \ ,$$

which in classical text books is referred to as the voltage calculated by "*electromagnetic force*". What we have done is to derive the electromagnetic force expression by using a modern Method of Moments argumentation. We can also express (5.43) by using the definition of the reaction integral in the following way:

$$Z_d = -\frac{1}{I_0^2}\langle I_0\mathbf{E}_j(\mathbf{r}_a), I_0 j_l(z)\hat{\mathbf{z}}\rangle \ .$$

[7] Use the third reciprocity relation in Section 4.5.2 on page 146.

This expression is seen to be equal to the self-reaction of a current cylinder normalized to the square of the total current $I_0 = \mathbf{J}_l(0) \cdot \hat{\mathbf{z}}$ passing through the feed gap location. The above derivation is general and valid also for other feed gap locations. Therefore, the self-impedance of a dipole fed at center at $z = z_{\text{gap}}$ is

$$Z_d = -\frac{I_0^2}{I_{z_{\text{gap}}}^2} \langle \mathbf{E}_J(\mathbf{r}_a), j_l(z)\hat{\mathbf{z}} \rangle \ , \tag{5.44}$$

where $I_{z_{\text{gap}}} = I_0 j_l(z_{\text{gap}})$ is the value of the total current passing through the feed gap. $\mathbf{E}_j(\mathbf{r}_a)$ are in all these equations the field due to the axial normalized current $j_l(z')$ evaluated at the surface $\mathbf{r}_a = a\hat{\boldsymbol{\rho}} + z\hat{\mathbf{z}}$ of the dipole arms, i.e.,

$$\mathbf{E}_j(\mathbf{r}_a) = C_k \int_{-l/2}^{l/2} [\eta j_l(z')\hat{\mathbf{z}} C_{N_1} - (\eta j_l(z')\hat{\mathbf{z}} \cdot \hat{\mathbf{R}})\hat{\mathbf{R}} C_{N_2}] \frac{1}{R} e^{-jkR} dz' \ , \tag{5.45}$$

with $R = |\mathbf{r}_a - \mathbf{r}'| = \sqrt{a^2 + (z - z')^2}$, and $\hat{\mathbf{R}} = (a\hat{\boldsymbol{\rho}} + (z - z')\hat{\mathbf{z}})/R$. Recall that C_k and the near field functions C_{N_1} and C_{N_2} are given in Section 4.2.1.

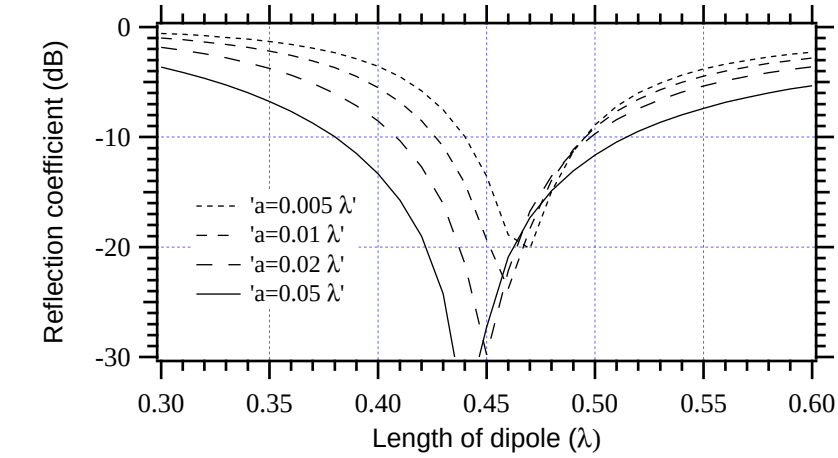
These self-impedance expressions are in principle also valid for infinitely thin dipoles. However, they cannot be evaluated for the foregoing case due to the severe singularities of C_{N_1} and C_{N_2} at $a = 0$. These singularities are called *source point singularities*. We could also have written down the expressions in (5.43) and (5.44) more or less directly by using the general expressions for the self-impedance of current distributions in Section 4.5.4. See also the next subsection.

5.1.8 Impedance of cylindrical and flat electric dipoles

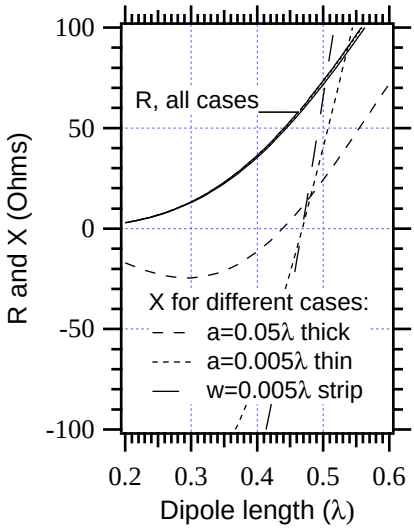
The easiest way to evaluate the impedance formula in (5.44) is not to use (5.45), but rather the near-field algorithm described in Section 4.7.1. The results of such evaluations are presented in Fig. 5.9*. We have used a segment length of $l_s = 0.005\lambda$. The derivations in the previous subsection can also be used to find the self-impedance of an electric strip dipole with a feed gap. In the same way this becomes the self-reaction of the assumed current distribution normalized to the square of the total current passing through the location of the feed gap. In this case we can evaluate the impedance by using the near-field algorithm in Section 4.7.2. Such results are presented in Fig. 5.9 as well*. In this case we have used a segment length of $l = 0.0025\lambda$.

The results are discussed below:

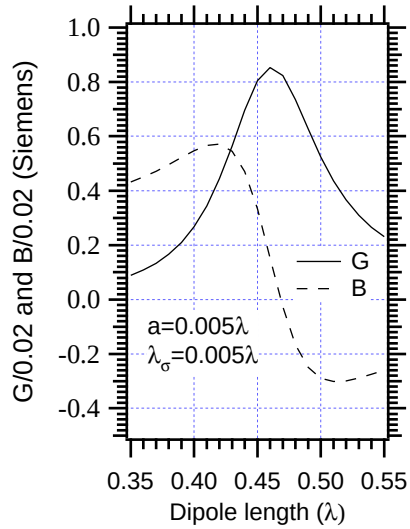
1. The real and imaginary parts (R and X, respectively) of the impedance Z of electric dipoles are presented in Fig. 5.9b. The resistance R increases with length (over the range shown). The values are not at all sensitive to thickness and whether the arms are cylindrical or flat, at least as long as the thickness (or width) is small compared to the length. This is understandable by comparing with the results in Subsection 5.1.6 where we evaluated the radiation resistance from the radiation field, which does not depend on the cross-sectional shape of the dipole arms, as long as the diameter is small. On the other hand, the reactance X is seen to depend strongly on the thickness and cross-sectional shape. The same dependencies are also present in the numerical models used. The resistance converges easily and fast, whereas the reactance is very sensitive



(a)



(b)



(c)

Figure 5.9: *Impedance characteristics of electric dipoles as a function of their length L in wavelengths. Parameters are thickness $2a$ of the cylindrical arms, and width w of flat strip arms. (a) Reflection coefficient, referred to $50\,\Omega$, (b) Resistance R and reactance X, (c) Conductance G and susceptance B.

to the approximations used in the evaluation of the near-field. The resonance frequency is the frequency for which $X = 0$. This reduces when the thickness or width increases, in agreement with the discussions in Subsection 5.1.2.

2. The reflection coefficient in terms of a line impedance of $50\,\Omega$ is shown in Fig. 5.9a. We see that dipoles can be reasonably well matched to $50\,\Omega$, even if the dipole resistance normally is considered to be $73\,\Omega$. The reason is that resonance normally appears when the dipole is shorter, and then the resistance is smaller than when it is half wavelength long. The reflection coefficient is normally presented relative to $50\,\Omega$ because this is a standard cable impedance. We also see that the bandwidth of a half-wave dipole seems to increase when the thickness increases. Note that the curves present variations with $l/2$ for fixed a/λ. This is not really the same as relative variations with frequency for a given dipole with fixed thickness. However, the impedance varies slowly with a/λ, so therefore the curves indicate actual bandwidths.

3. The real (conductivity G) and imaginary (susceptance B) parts of the admittance $Y = 1/Z$ are shown in Fig. 5.9c for a selected arm thickness. The values are normalized to $50\,\Omega = 0.02\,\text{Siemens}$. We see that the admittance has a typical resonant behavior with a peaked real part G at the point where the imaginary part B is zero.

We can draw the following general conclusion from the above discussions: *A resonant antenna which is excited at a current maximum has an input resistance which is not sensitive to variations in the geometry and the analysis model, whereas the reactance is very sensitive to both such variations. The input admittance shows a typical resonant behavior characterized by a conductance peak at the resonance frequency where the susceptance is zero.*

Resonant slot antennas are excited at a voltage maximum and therefore behave in an opposite way, see page 197.

5.1.9 Dipole at arbitrary location

The far-field function in Section 5.1.3 was derived for the case that the dipole had its center in the origin of the coordinate system. We will now find the far-field function when the dipole is moved to a new location $\mathbf{r}_0$ (see Fig. 5.10). Then, the coordinates of the source are described by

$$\mathbf{r}'(l') = \mathbf{r}_0 + l'\hat{\mathbf{l}} \qquad \text{for } -\frac{l}{2} < l' < \frac{l}{2} \ . \tag{5.46}$$

We again use (4.54) for this source, where now

$$\mathbf{r}' \cdot \hat{\mathbf{r}} = \mathbf{r}_0 \cdot \hat{\mathbf{r}} + l'\hat{\mathbf{l}} \cdot \hat{\mathbf{r}} \ , \tag{5.47}$$

with $\mathbf{r}_0 \cdot \hat{\mathbf{r}}$ being constant, independent of l'. This means that a factor $e^{jk\mathbf{r}_0 \cdot \hat{\mathbf{r}}}$ can be taken outside the integral sign in (4.54), and the remaining integrand is the same as for the dipole located in the origin. Therefore, we may write the far-field function by using (5.5) as

$$\mathbf{G}_d(k, I_0, l, \hat{\mathbf{l}}, \mathbf{r}_0, \hat{\mathbf{r}}) = C_k \eta I_0 [\hat{\mathbf{l}} - (\hat{\mathbf{l}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}] \tilde{j}(k\hat{\mathbf{l}} \cdot \hat{\mathbf{r}}) e^{jk\hat{\mathbf{r}}_0 \cdot \hat{\mathbf{r}}} \ , \tag{5.48}$$

where all the parameters have been defined before and $\tilde{j}(k\hat{\mathbf{l}} \cdot \hat{\mathbf{r}})$ is the same radiation integral as in (5.8). Compare this change of location of the dipole with the change of the phase reference point in Section 2.3.4. Note that moving the source and moving the phase reference point

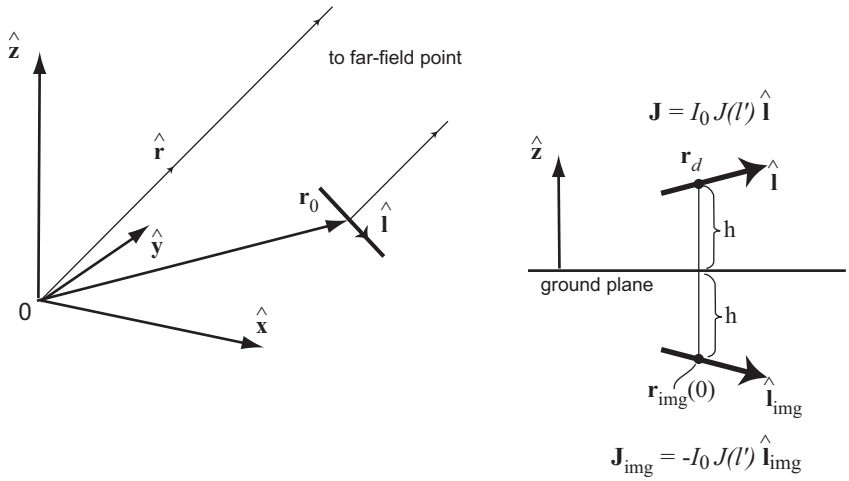


Figure 5.10: Electric dipole at arbitrary location $\mathbf{r}_0$ (left), and dipole above ground with its image (right).

are opposite actions. The phase reference point of the far-field function in (5.48) is still the origin of the coordinate system. Expression (5.48) is very general as the dipole can have any location $\mathbf{r}_0$ and orientation l, so it is well suited for programming. All the vectors can be programmed in rectangular coordinates.

Several dipoles with different orientations and locations in the same coordinate system can be superimposed as follows. If the far-field function of each of them is calculated by (5.48), they will have the same phase reference point (the origin). Therefore, the total far-field function is obtained by simply adding the contributions of each one.

5.1.10 Arbitrary dipole above ground

Dipole antennas are often located above an electric conducting plate which is commonly referred to as the ground plane. In consequence, the radiation field is directed into half-space becoming a unidirectional radiation pattern. We will here use the image theory[8] to calculate the radiation field. This means that we will neglect the finite size of the ground plane. The ground plane is assumed to coincide with the xy-plane with normal $\hat{\mathbf{n}} = \hat{\mathbf{z}}$.

We assume that the electric dipole with far-field function given by (5.48) is located at a point $\mathbf{r}_d = h\hat{\mathbf{z}}$, i.e., a height h above the ground (see Fig. 5.10). Then, the coordinates of the dipole is described by $\mathbf{r}'(l') = h\hat{\mathbf{z}} + l'\hat{\mathbf{l}}$, and its far-field function is obtained from (5.48) to be

$$\mathbf{G}_d(k, I_0, l, \hat{\mathbf{l}}, h\hat{\mathbf{z}}, \hat{\mathbf{r}}) = C_k \eta I_0 [\hat{\mathbf{l}} - (\hat{\mathbf{l}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}] \tilde{j}(k\hat{\mathbf{l}} \cdot \hat{\mathbf{r}}) e^{jkh\hat{\mathbf{z}}\cdot\hat{\mathbf{r}}} \ . \tag{5.49}$$

The location of the image dipole is described by (4.101) and (5.46), giving

$$\begin{aligned} \mathbf{r}_{\text{img}}(l') &= \mathbf{r}'(l') - 2[\mathbf{r}'(l') \cdot \hat{\mathbf{n}}]\hat{\mathbf{n}} \\ &= h\hat{\mathbf{z}} + l'\hat{\mathbf{l}} - 2h\hat{\mathbf{z}} - 2l'(\hat{\mathbf{l}} \cdot \hat{\mathbf{z}})\hat{\mathbf{z}} = (-h\hat{\mathbf{z}}) + l'\hat{\mathbf{l}}_{\text{img}} \ , \end{aligned} \tag{5.50}$$

[8] The image theory can be found in Section 4.6 on page 150.

with

$$\hat{\mathbf{l}}_{\text{img}} = \hat{\mathbf{l}} - 2(\hat{\mathbf{l}} \cdot \hat{\mathbf{z}})\hat{\mathbf{z}} \ . \tag{5.51}$$

From (4.102) and (5.4) the image current distribution becomes

$$\mathbf{J}_{\text{img}} = -[\mathbf{J} - 2(\mathbf{J} \cdot \hat{\mathbf{z}})\hat{\mathbf{z}}] = -I_0 j(l')\hat{\mathbf{l}}_{\text{img}} \ . \tag{5.52}$$

Therefore, the far-field function of the image dipole is obtained from (5.48) as

$$\mathbf{G}_d(k, -I_0, l, \hat{\mathbf{l}}_{\text{img}}, -h\hat{\mathbf{z}}, \hat{\mathbf{r}}) \ ,$$

and the total far-field function of the dipole above ground plane becomes

$$\mathbf{G}_{\text{dg}}(\hat{\mathbf{r}}) = \mathbf{G}_d(k, I_0, l, \hat{\mathbf{l}}, h\hat{\mathbf{z}}, \hat{\mathbf{r}}) + \mathbf{G}_d(k, -I_0, l, \hat{\mathbf{l}}_{\text{img}}, -h\hat{\mathbf{z}}, \hat{\mathbf{r}}) \ . \tag{5.53}$$

Note that $\mathbf{G}_{\text{dg}}(\hat{\mathbf{r}})$ is a valid field solution only above the ground plane. Under the ground plane $\mathbf{G}_{\text{dg}}(\hat{\mathbf{r}}) = 0$. It is clear that $\mathbf{G}_{\text{dg}}(\hat{\mathbf{r}})$ can be easily computed by using a general subroutine for $\mathbf{G}_d(k, I_0, l, \hat{\mathbf{l}}, \mathbf{r}_0, \hat{\mathbf{r}})$. The above approach is very general and can be extended to other type of antennas above the ground.

We will now restrict ourselves to vertical and horizontal dipoles in order to get simple analytical expressions. This helps us to gain some physical insight into how the ground plane affects the far-field function.

5.1.11 Vertical dipole above ground

The vertical dipole above the ground is shown to the left in Fig. 5.11. Then, we have

$$\hat{\mathbf{l}} = \hat{\mathbf{z}} \ , \qquad \hat{\mathbf{l}}_{\text{img}} = -\hat{\mathbf{z}} \ , \qquad \hat{\mathbf{l}} \cdot \hat{\mathbf{r}} = \cos\theta \ ,$$
$$\hat{\mathbf{l}} - (\hat{\mathbf{l}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} = \hat{\mathbf{z}} - (\hat{\mathbf{z}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} = -\sin\theta\hat{\boldsymbol{\theta}} \ ,$$
$$\hat{\mathbf{l}}_{\text{img}} \cdot \hat{\mathbf{r}} = -\cos\theta \qquad \text{and} \qquad \hat{\mathbf{l}}_{\text{img}} - (\hat{\mathbf{l}}_{\text{img}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} = \sin\theta\hat{\boldsymbol{\theta}} \ .$$

Using these in combination with (5.48) and (5.52), we obtain:

$$\mathbf{G}_{\text{vdp}}(\hat{\mathbf{r}}) = -C_k \eta I_0 \sin\theta \tilde{j}(k\hat{\mathbf{r}} \cdot \hat{\mathbf{z}})[e^{jkh\cos\theta} + e^{-jkh\cos\theta}]\hat{\boldsymbol{\theta}} \ . \tag{5.54}$$

It is also possible to set up (5.54) directly from a study of Fig. 5.10[9]. For the half-wave dipole, by combining the two exponential terms within the brackets and using (5.9), we achieve:

$$\mathbf{G}_{\text{vdp}}(\hat{\mathbf{r}}) = -C_k \eta I_0 \frac{2}{k} \left[\cos\left(\frac{\pi}{2}\cos\theta\right) / \sin\theta\right] 2\cos(kh\cos\theta)\hat{\boldsymbol{\theta}} \ , \tag{5.55}$$

for $\theta < 90^\circ$. We see that the far-field function is given as a product of the far-field function of the vertical half-wave dipole in free space multiplied with a ground plane factor which is $2\cos(kl\cos\theta)$. It is important to be aware that this ground plane factor is different for different orientations of the dipole. Also note that in more complicated antennas the far-field function cannot be factorized in this way. We can only define special ground plane factors when the source and its image has the same or opposite orientations. The source also needs to have a far-field function which is symmetric or antisymmetric around the plane defined by $\theta = \pi/2$, where the direction $\theta = 0$ is normal to the ground plane. These two conditions are

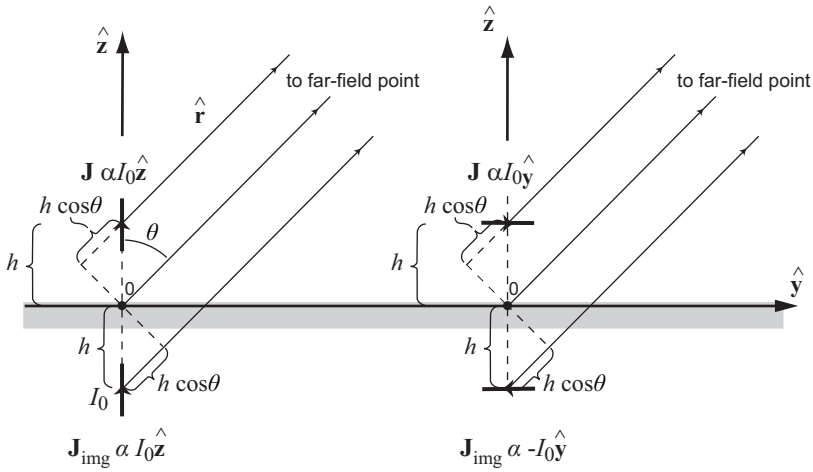


Figure 5.11: Vertical (left) and horizontal (right) dipoles above ground.

very seldom satisfied for other antennas than dipoles. Note that (5.55) is valid in the source region $0^\circ \leq \theta \leq 90^\circ$. In the null-field region below the ground plane $\mathbf{G}_{\rm vdp}(\hat{\mathbf{r}}) = 0$.

The directivity and *radiation resistance* can be determined in the same way as for the isolated dipole. This is most easily done by numerical evaluation of the integrals. It is also possible to use an approach in which the mutual coupling to the image is evaluated. We can get simple analytical results by studying a *short dipole over the ground*. Then, the far-field function is

$$\mathbf{G}_{\rm vdp}(\hat{\mathbf{r}}) = -C_k \eta I_0 (l/2)(\sin\theta) 2\cos(kh\cos\theta)\hat{\boldsymbol{\theta}} \ . \tag{5.56}$$

The power integral for the incremental dipole becomes

$$\begin{aligned} P &= 2\pi \int_0^{\pi/2} |\mathbf{G}_{\rm vdp}(\hat{\mathbf{r}})|^2 \sin\theta d\theta \\ &= 8\pi |C_k \eta I_0 (l/2)|^2 + \int_0^{\pi/2} \cos^2(kh\cos\theta)\sin^3\theta d\theta \ . \end{aligned} \tag{5.57}$$

The latter integral can be evaluated analytically by substituting $u = \cos\theta$ and $du = -\sin\theta d\theta$, and reducing it to some known integrals. The result gives the following directivity

$$D_0 = 2\left[\frac{1}{3} - \frac{\cos(2kh)}{(2kh)^2} + \frac{\sin(2kh)}{(2kh)^3}\right]^{-1} \ , \tag{5.58}$$

and radiation resistance

$$R_{\rm rad} = 2\pi\eta \left(\frac{l}{2\lambda}\right)^2 \left[\frac{1}{3} - \frac{\cos(2kh)}{(2kh)^2} + \frac{\sin(2kh)}{(2kh)^3}\right] \ . \tag{5.59}$$

The radiation pattern of the short dipole above the ground is shown in Fig. 5.12*, and the radiation resistance and directivity in Fig. 5.13*. We see that the radiation pattern has

[9] It is recommended to do this as an exercise.

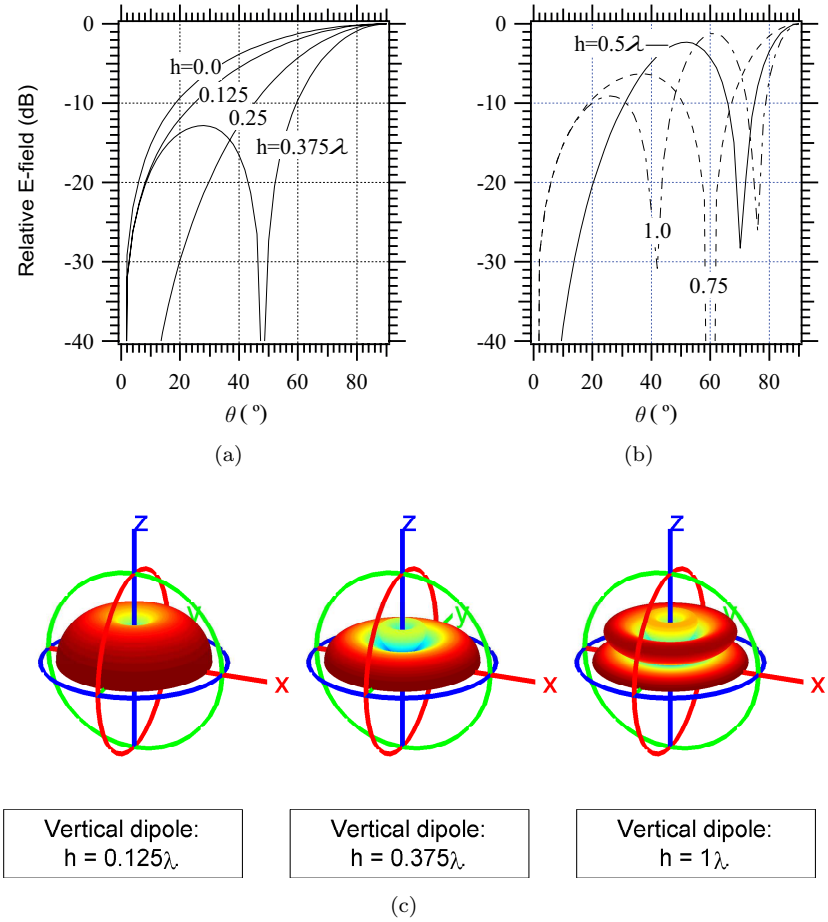


Figure 5.12: *Radiation patterns of a vertical short electric dipole for different heights above an infinite PEC plane. (a) and (b) E-plane pattern for different heights. (c) 3D plots for different heights.

several lobes when $h \geq 0.375\lambda$ due to the ground plane factor. The directivity is seen to be 3 dB larger than for a short dipole in free space when $h = 0$. The reason for this is evident from the formulas in the next section, as a short dipole has a radiation pattern which does not depend on its length. The directivity is 6 dB larger than that of a dipole in free space when $h > 0.5\lambda$. The reason for this is that when the dipole and its image is far away from each other, the far-field function will double its amplitude in the main beam direction, whereas the total radiated power is unaffected by the presence of the image dipole.

5.1.12 Vertical monopole

The far-field function of the vertical monopole has to be the same as that of a vertical dipole of the double length because the total current distribution of the monopole plus its image is

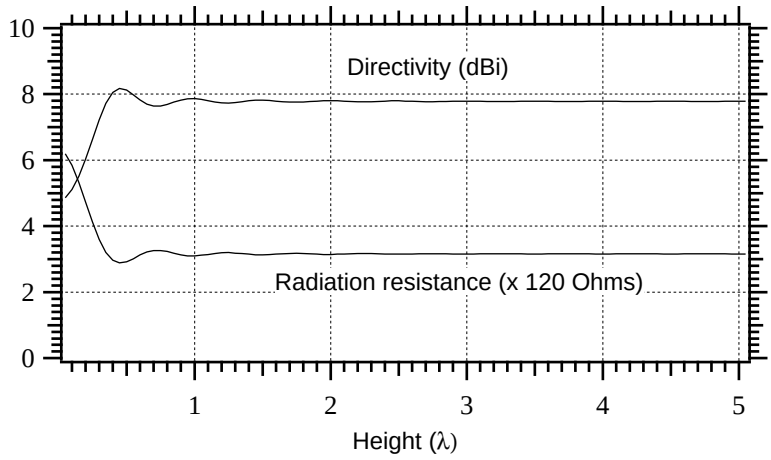


Figure 5.13: ⋆Directivity and radiation resistance of a vertical short electric dipole as a function of its height above an infinite PEC plane. The length of the dipole is $l = \lambda/50$.

the same as for the dipole. Therefore,

$$\mathbf{G}_m(\hat{\mathbf{r}}) = \begin{cases} -C_k \eta I_0 \sin\theta \tilde{j}(k\hat{\mathbf{r}} \cdot \hat{\mathbf{z}})\hat{\boldsymbol{\theta}} & \text{for } \theta < \pi/2 \\ 0 & \text{for } \theta > \pi/2 \end{cases} . \tag{5.60}$$

The power integral becomes

$$P_{\text{mon}} = 2\pi \int_0^{\pi/2} |\mathbf{G}_m(\theta)|^2 \sin\theta d\theta = \frac{1}{2} P_{\text{dp}} , \tag{5.61}$$

which is half that of the half-wave dipole in (5.34). Therefore, the radiation resistance of the monopole is

$$R_{\text{mon}} = \frac{1}{2} R_{\text{dp}} , \tag{5.62}$$

where R_{dp} is the impedance of the half-wave dipole given in (5.36).

5.1.13 Horizontal dipole above ground

The horizontal dipole above the ground is shown to the right in Fig. 5.11. Then we have

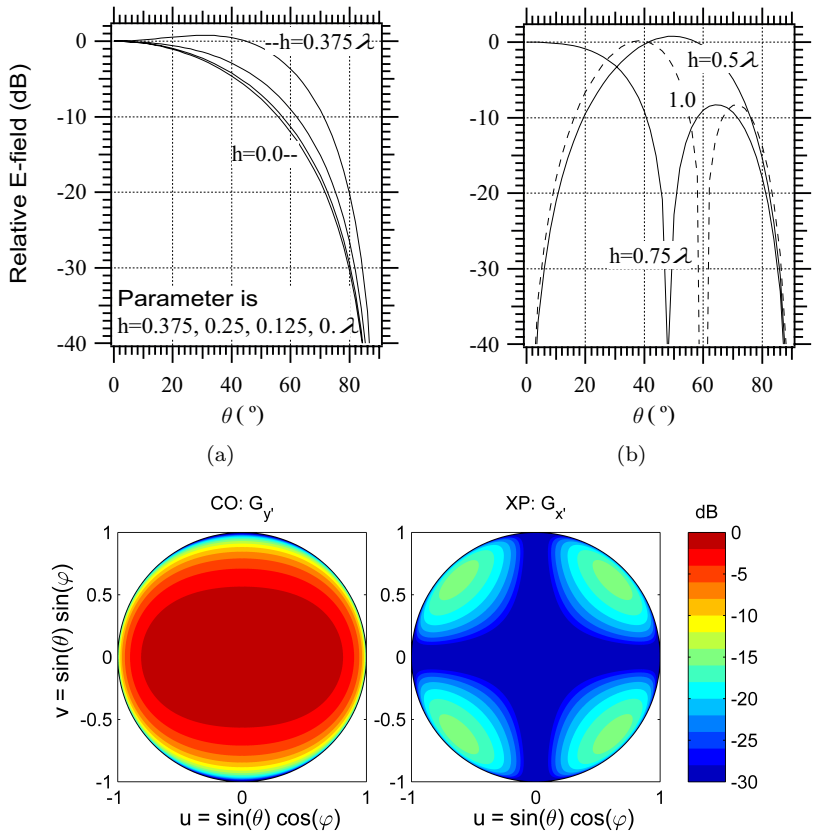
$$\hat{\mathbf{l}}_{\text{img}} = \hat{\mathbf{l}} = \hat{\mathbf{y}} , \qquad \hat{\mathbf{l}} \cdot \hat{\mathbf{r}} = \hat{\mathbf{y}} \cdot \hat{\mathbf{r}} = \sin\theta \sin\varphi ,$$
$$\text{and} \qquad \hat{\mathbf{y}} - (\hat{\mathbf{y}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} = \cos\theta \sin\varphi \hat{\boldsymbol{\theta}} + \cos\varphi \hat{\boldsymbol{\varphi}} .$$

Using these and together with (5.48) and (5.52), we get

$$\mathbf{G}_{\text{hdp}}(\hat{\mathbf{r}}) = C_k \eta I_0 \tilde{j}(k\hat{\mathbf{l}} \cdot \hat{\mathbf{r}})(\cos\theta \sin\varphi \hat{\boldsymbol{\theta}} + \cos\varphi \hat{\boldsymbol{\varphi}})(e^{jkh\cos\theta} - e^{-jkh\cos\theta}) . \tag{5.63}$$

Finally by combining the two exponential terms within the brackets, we get

$$\mathbf{G}_{\text{hdp}}(\hat{\mathbf{r}}) = C_k \eta I_0 (\cos\theta \sin\varphi \hat{\boldsymbol{\theta}} + \cos\varphi \hat{\boldsymbol{\varphi}}) \tilde{j}(k\hat{\mathbf{l}} \cdot \hat{\mathbf{r}}) 2j \sin(kh\cos\theta) . \tag{5.64}$$



(c) Horizontal dipole at $h = 0.25\lambda$.

Figure 5.14: ⋆Radiation patterns of *horizontal* short electric dipole for different heights above an infinite PEC plane. (a) and (b) E-plane ($\phi = 90^\circ$) patterns. (c) Co- and cross-polar patterns for $h = 0.25\lambda$.

We see that the ground plane factor is very different from that of a vertical dipole. Similarly for a short dipole we attain

$$\mathbf{G}_{\mathrm{hdp}}(\hat{\mathbf{r}}) = C_k \eta I_0 (l/2)(\cos\theta \sin\varphi \hat{\boldsymbol{\theta}} + \cos\varphi \hat{\boldsymbol{\varphi}}) 2j \sin(kh\cos\theta) \ . \tag{5.65}$$

The pattern of the latter is plotted in Fig. 5.14 for different heights above the ground⋆. We see that the pattern have several lobes when $h > 0.5\lambda$. The radiation resistance and the directivity in the vertical $\theta = 0^\circ$ direction are calculated by numerical integration and shown in Fig. 5.15⋆. We see that the radiation resistance is zero when $h = 0$. This corresponds to a short-circuited feed gap. When $h = \lambda/4$ the resistance is close to that in free space, and it oscillates around this and approaches it asymptotically when h increases further. The directive gain for $\theta = 0^\circ$ oscillates between minus infinity and 7.8 dB, depending on whether the dipole and its image interfere destructively or constructively in the $\theta = 0^\circ$ direction. 7.8 dB is four times (i.e., 6 dB) more than the directivity of the short dipole in free space.

We have also evaluated the impedance of a horizontal half-wave dipole over the ground by using the self-reaction formula in (5.44), imaging, and the near-field algorithm. The resulting

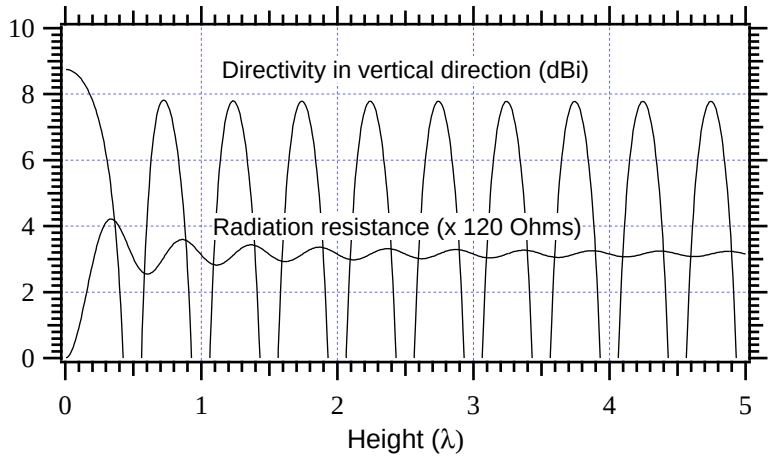


Figure 5.15: ⋆Radiation resistance and directive gain ($\theta = 0°$) of a *horizontal* short electric dipole at height h above an infinite PEC plane.

reflection coefficients with respect to $50\,\Omega$ are shown in Fig. 5.16 for different thicknesses and heights⋆. We see that an electric dipole easily can be matched to $50\,\Omega$ by tuning its length and height. The impedances and admittances show similar characteristics as for a dipole in free space[10].

5.2 Electric loop antenna as vertical magnetic dipole

The *loop antenna* is a circular metal ring fed by a coaxial cable and a balun at some point along the ring. When the circumference is much smaller than λ we may assume that the current is constant around the ring. In addition if we use the thin wire approximation we can express the ring current as (see Fig. 5.17)

$$\mathbf{J}_l(\mathbf{r}') = I_0\hat{\boldsymbol{\varphi}}' \qquad \text{for } \mathbf{r}' = a\hat{\boldsymbol{\rho}}'(\varphi')\ ,\ 0 < \varphi' < 2\pi\ , \tag{5.66}$$

where a is the ring radius, $\boldsymbol{\varphi}'$ is the unit vector in φ'-direction, and $\boldsymbol{\rho}'$ is the radial unit vector in the xy-plane, i.e.,

$$\hat{\boldsymbol{\rho}}'(\varphi') = \cos\varphi'\hat{\mathbf{x}} + \sin\varphi'\hat{\mathbf{y}}\ . \tag{5.67}$$

Now we can use the far-field radiation integral (4.54) with $\mathbf{r}'\cdot\hat{\mathbf{r}} = a\boldsymbol{\rho}'\cdot\hat{\mathbf{r}}$, and we get

$$\mathbf{E} = \frac{1}{r}e^{-jkr}\mathbf{G}_0(\hat{\mathbf{r}})\ , \qquad \mathbf{G}_0(\hat{\mathbf{r}}) = \mathbf{I}_J - (\mathbf{I}_J\cdot\hat{\mathbf{r}})\hat{\mathbf{r}}\ , \tag{5.68}$$

$$\begin{aligned}\mathbf{I}_J &= C_k\int_0^{2\pi}\eta I_0\hat{\boldsymbol{\varphi}}'e^{jka\boldsymbol{\rho}'\cdot\hat{\mathbf{r}}}ad\varphi' \\ &= C_k\eta I_0\int_0^{2\pi}(\cos\varphi'\hat{\mathbf{y}} - \sin\varphi'\hat{\mathbf{x}})e^{jka\boldsymbol{\rho}'\cdot\hat{\mathbf{r}}}ad\varphi'\ ,\end{aligned} \tag{5.69}$$

[10] See Subsection 5.1.8.

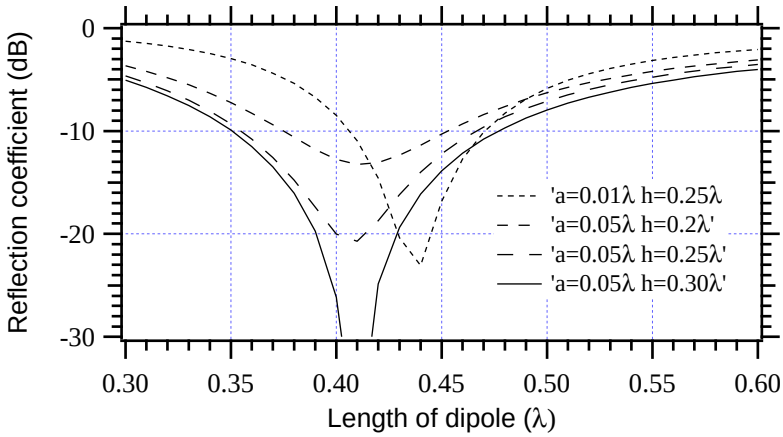


Figure 5.16: *Reflection coefficient of horizontal electric dipole above the ground as a function of dipole length l for different heights h over the ground. The feed line has an impedance of $50\,\Omega$. Segment length in near-field evaluation is 0.005λ.

where we have expressed $\hat{\boldsymbol{\varphi}}'$ in terms of φ' dependent functions and constant vectors, by using $\hat{\boldsymbol{\varphi}}' = \cos\varphi'\hat{\mathbf{y}}' - \sin\varphi'\hat{\mathbf{x}}'$. When $ka \ll 1$ we can expand the phase factor in the integrand according to $e^{jka\boldsymbol{\rho}'\cdot\hat{\mathbf{r}}} \approx 1 + jka\boldsymbol{\rho}'\cdot\hat{\mathbf{r}}$ and get

$$\mathbf{I}_J = C_k\eta I_0 \int_0^{2\pi} [(\cos\varphi' + jka\hat{\boldsymbol{\rho}}'\cdot\hat{\mathbf{r}}\cos\varphi')\hat{\mathbf{y}} - (\sin\varphi' + jka\hat{\boldsymbol{\rho}}'\cdot\hat{\mathbf{r}}\sin\varphi')\hat{\mathbf{x}}]ad\varphi' \ . \tag{5.70}$$

The whole problem has rotational symmetry of type BOR_0, so we may limit the observation point to the xz-plane, in which case $\hat{\mathbf{r}} = \cos\theta\hat{\mathbf{z}} + \sin\theta\hat{\mathbf{x}}$ and $a\hat{\boldsymbol{\rho}}'\cdot\hat{\mathbf{r}} = a\sin\theta\cos\varphi'$. Then,

$$\begin{aligned}\mathbf{I}_J &= C_k\eta I_0 jka^2\sin\theta\left[\int_0^{2\pi}\cos^2\varphi' d\varphi'\hat{\mathbf{y}} - \int_0^{2\pi}\sin\varphi'\cos\varphi' d\varphi'\hat{\mathbf{x}}\right] \\ &= C_k\eta I_0 jk\pi a^2\sin\theta\hat{\mathbf{y}} \ .\end{aligned} \tag{5.71}$$

For an arbitrary φ-plane and for symmetry reasons, we must get the same answer with $\hat{\mathbf{y}}$ replaced by the unit vector $\hat{\boldsymbol{\varphi}}$ normal to the φ-plane. Therefore, we have

$$\mathbf{G}_0(\hat{\mathbf{r}}) = C_k\eta I_0 jk\pi a^2\sin\theta\hat{\boldsymbol{\varphi}} \ . \tag{5.72}$$

We see that the loop antenna has the same far-field function as a magnetic dipole with magnetic moment $M_0 l$ if

$$M_0 l = j\eta I_0 kA \ , \tag{5.73}$$

where $A = \pi a^2$ is the area of the loop.

5.3 Helical antennas

A *helical antenna* consists of a spiral shaped wire with a certain number of turns. The center line of a right-hand turned helical wire can be described by the coordinate vector (see Fig. 5.18)

$$\mathbf{r}'(\varphi') = a\hat{\boldsymbol{\rho}}(\varphi') + (S\varphi'/2\pi)\hat{\mathbf{z}} \qquad \text{for } 0 < \varphi' < N2\pi \ , \tag{5.74}$$

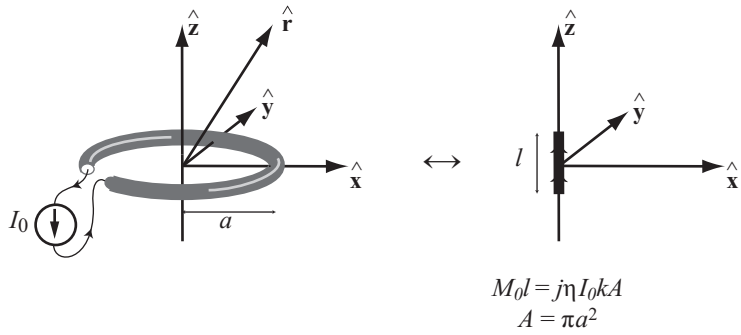


Figure 5.17: Small electric loop antenna (left) and its equivalent magnetic current (right).

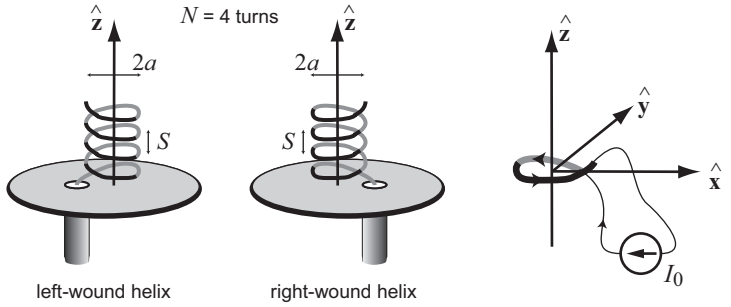


Figure 5.18: Geometry of helical antenna for LHC (left) and RHC (middle) polarizations, and electric current model of one single loop (right).

where a is the radius of the helix, S is the spacing between each turn, and N is the total number of turns. The current distribution along the wire depends strongly on a, S and the total wire length. Thus, it is not possible to find an approximate current distribution which is acceptable enough for analytical work. Therefore, helical antennas are best designed by using previously published empirical data based on a series of measurements, or by computer simulations using codes for wire antennas based on the *Method of Moments*.

However, the following two current distributions on one single loop can be used to describe the two basic modes of operation of the helical antenna, the normal mode and the axial mode. They are

$$\mathbf{J}(\varphi') = I_0 \hat{\boldsymbol{\varphi}}' \,, \qquad \text{when } L_1 \ll \lambda \,, \tag{5.75}$$

$$\mathbf{J}(\varphi') = I_0 e^{-j\varphi'} \hat{\boldsymbol{\varphi}}' \,, \qquad \text{when } L_1 \approx \lambda \,, \tag{5.76}$$

respectively, where L_1 is the length of one turn given by

$$L_1 = \sqrt{(2\pi a)^2 + S^2} \,.$$

The former case corresponds to the case studied in the previous subsection, which gives the same radiation field as a vertical magnetic dipole, except that here we also need to consider the ground plane. This is left to the reader as an exercise. We will therefore find the far-field function of the latter case. This case corresponds to a travelling wave around the loop, which can be realized if the circumference L_1 of a single ring is approximately equal to λ which gives resonance.

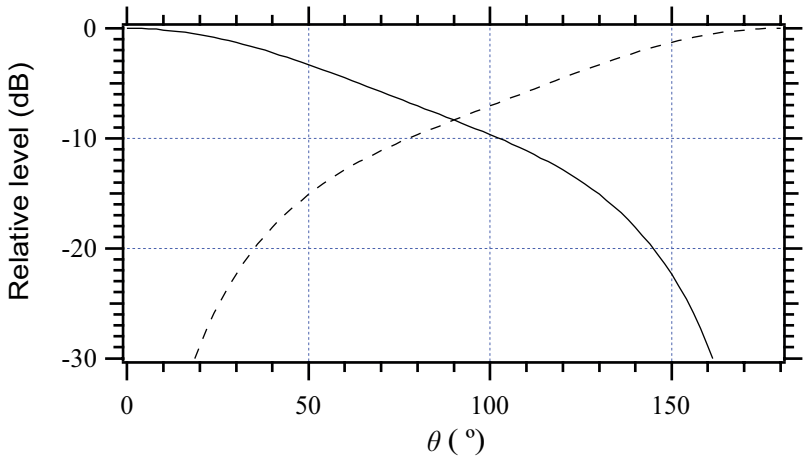


Figure 5.19: *Co- and cross-polar radiation patterns of a single resonant electric current loop of 1λ circumference.

In the same way as in the previous subsection the radiation integral becomes

$$\mathbf{I}_J = C_k \eta I_0 \int_0^{2\pi} e^{-j\varphi'} \hat{\boldsymbol{\varphi}}' e^{jka\boldsymbol{\rho}'\cdot\hat{\mathbf{r}}} a d\varphi' \ . \tag{5.77}$$

By introducing $\varphi' = \cos\varphi'\hat{\mathbf{y}} - \sin\varphi'\hat{\mathbf{x}}$, and $\boldsymbol{\rho}' \cdot \hat{\mathbf{r}} = \sin\theta\cos(\varphi' - \varphi)$, we obtain

$$\mathbf{I}_J = C_k \eta I_0 \int_0^{2\pi} [\cos\varphi' e^{-j\varphi'}\hat{\mathbf{y}} - \sin\varphi' e^{-j\varphi'}\hat{\mathbf{x}}] e^{jka\sin\theta\cos(\varphi'-\varphi)} a d\varphi' \ . \tag{5.78}$$

If we substitute

$$\cos\varphi' = \frac{1}{2}(e^{j\varphi'} + e^{-j\varphi'}) \qquad \text{and} \qquad \sin\varphi' = \frac{1}{2j}(e^{j\varphi'} - e^{-j\varphi'}) \tag{5.79}$$

and use the Bessel function integral representations presented in Appendix E, we achieve

$$\mathbf{I}_J = C_k \eta I_0 \pi a [J_0(ka\sin\theta)(\hat{\mathbf{y}} + j\hat{\mathbf{x}}) - e^{-j2\varphi} J_2(ka\sin\theta)(\hat{\mathbf{y}} - j\hat{\mathbf{x}})] \ , \tag{5.80}$$

where J_0 and J_2 are the Bessel functions of zeroth and second order, respectively. As $J_2(0) = 0$ and $J_0(0) = 1$ is the maximum of J_0, the polarization in (5.80) is clearly recognized as RHC along the positive z-axis and LHC along the negative. This LHC wave will by reflection from a ground plane also become RHC. Thus, a loop on a ground plane with a propagating wave on it gives a circularly polarized wave when the circumference is about one wavelength. Therefore, the helical antenna will radiate circular polarization when $L_1 \approx \lambda$.

The helical antenna is most often located on a small metal disk which acts as a ground plane. The cross-polarization and far-out sidelobes will be less and the symmetry of the main lobe will be better if the ground plane is corrugated with circular grooves [4]. It is also possible to improve the performance by tapering the diameter of the helix to zero at its radiating end.

We have evaluated the co- and cross-polar component of (5.80)*. They are presented in Fig. 5.19. The cross-polar sidelobes would be less if we were able to reduce the diameter of the loop and still keep it resonant.

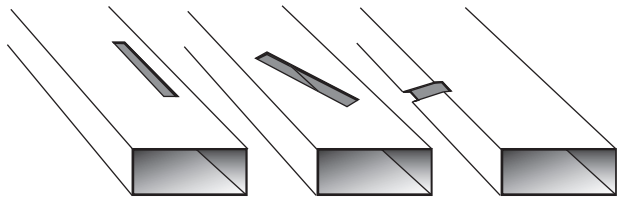


Figure 5.20: Examples of different radiating slots in waveguides.

5.4 Slot antennas

Metal boxes are often used to *shield* electronic equipment in order to avoid radiation of signals (emission), which may disturb other equipments, or reception (susceptibility) of interfering signals from other equipments. Today there exist strong requirements on both emission and susceptibility of electronic equipment in order to ensure *electromagnetic compatibility* (*EMC*). Antenna measurement chambers are also often shielded for the same reasons. It is well known that narrow apertures in shields[11] cause electromagnetic coupling between the inner and outer sides. Therefore, all kind of slots or openings in shields must be avoided.

Slots are also commonly in use as desirable antennas, because they are cheap and accurate to manufacture. In such cases they are almost always fed by waveguides. Indeed, the feed waveguide has often several slots forming a linear array[12]. There may also be several slotted feed waveguides side by side making up a complete two-dimensional array antenna. The slots can be located differently depending on the application (Fig. 5.20). The radiation pattern of a single slot will be derived in the next subsection. The excitation of the slot from the waveguide is rather laborious to include in the analysis, but it can be done very accurately due to the clean geometry. We will limit the analysis to three simple and illustrative cases: a slot excited by a voltage source at its center, a slot excited by a plane incident wave, and a slot excited by the open end of a rectangular waveguide with the same dimensions as the slot. Finally, the excitation of more complicated slots is discussed without detailed analysis.

5.4.1 Field distribution and radiation pattern

We consider a slot of width w and length l in an infinite PEC ground plane. If the ground plane has a finite wall thickness, the slot resembles a short rectangular waveguide. Then, we know that the field inside the slot region can be expressed as a sum of rectangular waveguide modes. If the slot is narrow with $w \ll \lambda$ and with $l \approx \lambda/2$, the basic TE_{10} mode dominates. The higher order modes will be strongly evanescent and cannot propagate through the slot. It is therefore reasonable to approximate the E-field distribution over the slot by the TE_{10} mode of the rectangular waveguide. Thus, for an x-directed slot with the origin of the coordinate system in the center of the slot, the slot field becomes

$$\mathbf{E}_{\mathrm{slt}}(x', y') = E_0 \cos\left(\frac{\pi}{l} x'\right) \hat{\mathbf{y}} \ , \tag{5.81}$$

[11] Such as slots around a removable side of a box, or around the door of a chamber.

[12] See Fig. 10.3, "Examples of resonant, travelling wave and leaky wave linear waveguide slot antenna arrays.," on page 336.

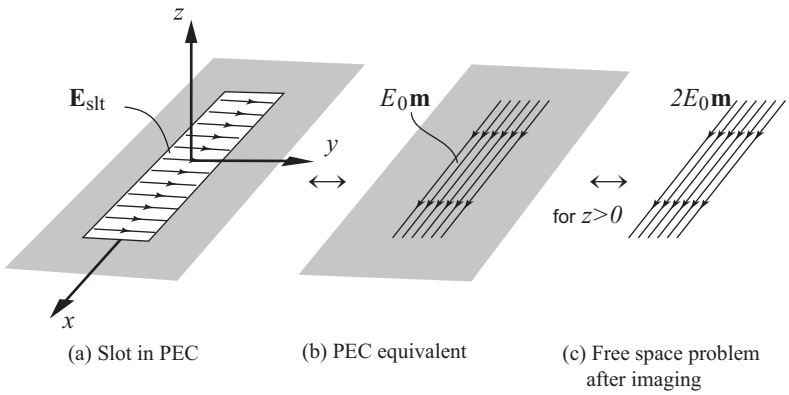


Figure 5.21: E-field distribution in slot (left) in PEC and its equivalent magnetic current (center) in PEC without slot. After imaging the PEC ground plane is removed (right).

for $|x'| < l/2$ and $|y'| < w/2$. This slot field distribution has shown to be a very good approximation even when the wall is infinitely thin. It is also hardly affected by the type of excitation and by the form and shape of the ground plane, provided $w \ll \lambda$ and $l \leq \lambda/2$. We can find the radiation pattern of the slot in the ground plane by using equivalent magnetic sources, as illustrated in Fig. 5.21. We consider the region $z > 0$, replace the E-field at the boundary $z = 0$ by equivalent *magnetic currents* $\mathbf{M} = E_0\mathbf{m} = \mathbf{E}_{\rm slt} \times \hat{\mathbf{z}} = E_0 \cos(\pi x'/l)\hat{\mathbf{x}}$, and close the hole in the PEC[13]. This new equivalent field problem consists of $E_0\mathbf{m}$ on a PEC ground plane with no slot. Therefore, we can now remove the ground plane by imaging. As a result, we can calculate the radiation field in the source region $z > 0$ as the field radiated by a magnetic current $E_0 2\mathbf{m}$ is radiating in free space. We must remember that the solution for $z < 0$ is zero (null-field region)[14]. The radiated E-field is now obtained by using (4.50) and (4.52) to be

$$\mathbf{E}(\mathbf{r}) = \frac{1}{r}e^{-jkr}\mathbf{G}_{\rm slt}(\hat{\mathbf{r}}) \ , \qquad \mathbf{G}_{\rm slt}(\hat{\mathbf{r}}) = 2E_0 w\mathbf{G}_{\rm img}(\hat{\mathbf{r}})\tilde{M}(k\hat{\mathbf{x}}\cdot\hat{\mathbf{r}}) \ , \tag{5.82}$$

$$\text{where} \qquad \mathbf{G}_{\rm img} = C_k(\hat{\mathbf{x}} \times \hat{\mathbf{r}}) \ , \tag{5.83}$$

$$\tilde{M}(k\hat{\mathbf{x}}\cdot\hat{\mathbf{r}}) = \int_{-l/2}^{l/2} \cos(\pi x'/l)e^{jkx'\hat{\mathbf{x}}\cdot\hat{\mathbf{r}}}dx' \ , \tag{5.84}$$

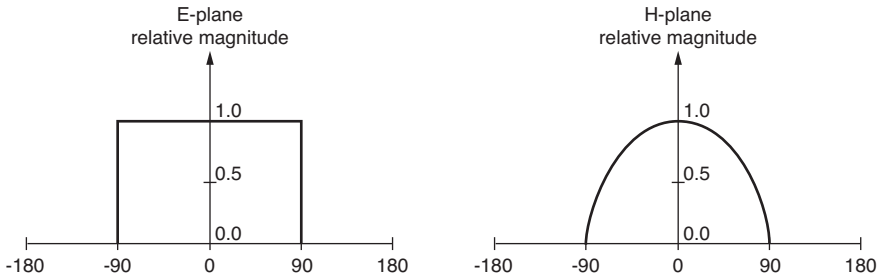
with the incremental source constant $C_k = -jk/4\pi$ as before. In order to obtain the above form, we have assumed that $w \ll \lambda$ so that the y'-integral from $-w/2$ to $w/2$ could be replaced by a multiplication by w. We have also moved the constants E_0 and w outside the radiation integral.

The result is that the *far-field function* of the slot is a product of three factors:

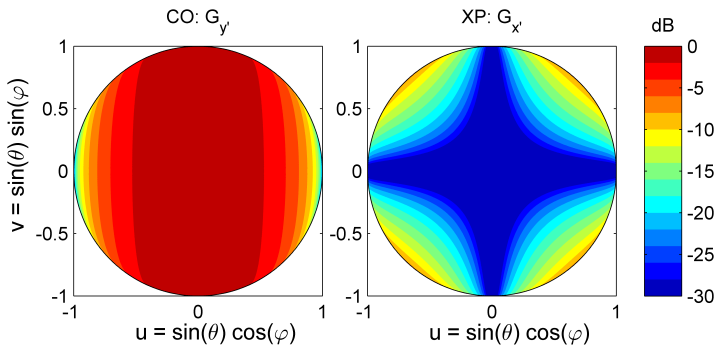
1. the factor $2wE_0$ which is twice the voltage $V_0 = E_0 w$ over the center of the slot,
2. the far-field function $\mathbf{G}_{\rm img}$ of a unit incremental x-directed magnetic current source, and

[13] Note that here we have introduced a normalized magnetic current $\mathbf{m}$ representing the magnetic current distribution when $E_0 = 1$.

[14] See "Imaging" in Section 4.6 on page 150.



(a)



(b)

Figure 5.22: Radiation patterns of half-wave slot in ground plane. (a) E- and H-plane. (b) Co- and cross-polar contour plots.

3. the *Fourier transform* $\tilde{M}(k\hat{\mathbf{x}}\cdot\hat{\mathbf{r}})$ of the magnetic current distribution $M(x') = \cos(\pi x'/l)$. The latter is recognized as a Fourier transform by substituting $k_x = k\hat{\mathbf{x}}\cdot\hat{\mathbf{r}}$.

The above factorization corresponds to that of wire antennas and applies to all straight slots in flat ground planes. When $l = \lambda/2$, the magnetic current distribution is exactly the same as the electric current distribution of an electric half-wave dipole. The reason is $\cos(\pi x'/l) = \sin(k((l/2) - |x'|))$ when $l = \lambda/2$. Therefore, by using $\hat{\mathbf{r}}\cdot\hat{\mathbf{x}} = \sin\theta\cos\varphi$,

$$\tilde{M}(k\hat{\mathbf{r}}\cdot\hat{\mathbf{x}}) = \frac{2}{k}\cos\left(\frac{\pi}{2}\sin\theta\cos\varphi\right)/[1-(\sin\theta\cos\varphi)^2]\ . \tag{5.85}$$

For this slot the xz-plane is the H-plane, and the yz-plane is the E-plane. The radiation patterns are shown in Fig. 5.22. In E-plane $\varphi = 90^\circ$ and $\hat{\mathbf{x}}\times\hat{\mathbf{r}} = -\hat{\boldsymbol{\theta}}$, so the pattern is uniform. In H-plane $\varphi = 0^\circ$ and $\hat{\mathbf{x}}\times\hat{\mathbf{r}} = -\cos\theta\hat{\boldsymbol{\varphi}}$, so the pattern shape is given by $-\tilde{M}(k\hat{\mathbf{r}}\times\hat{\mathbf{x}})\cos\theta$. This is slightly more directive than the incremental magnetic current. The patterns are similar to those of the electric dipole except that the E- and H-planes are interchanged, the field behind $\theta = \pi/2$ is zero due to the ground plane (see Fig. 5.21), and that the first factor is $2E_0 w$ instead of ηI_0. The directivity of the half-wave slot is easily calculated in the same way as for the half-wave dipole. The fields radiate only in the upper half-space, so the power integral becomes

$$P_{\text{slt}} = \frac{1}{2}P_{\text{dp}}\left(\frac{2E_0 w}{\eta I_0}\right)^2\ , \tag{5.86}$$

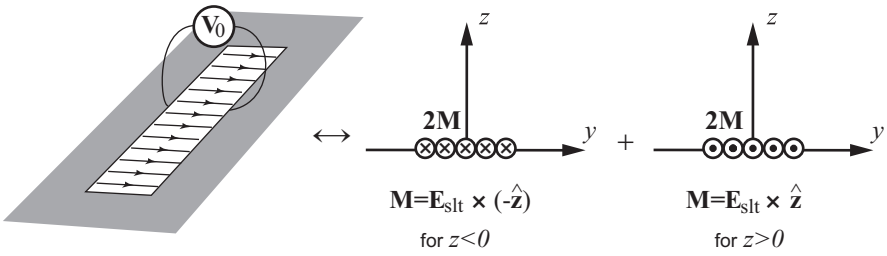


Figure 5.23: Cross-section of slot in infinite PEC plane excited by a voltage source (left), the equivalent problem for $z < 0$ (center) and for $z > 0$ (right).)

where $P_{\rm dp}$ is the power integral of the half-wave electric dipole. Therefore,

$$D_{_0} = \frac{4\pi |G_{\rm slt}|^2_{\rm max}}{P_{\rm slt}} = \frac{2 \cdot 4\pi |G_d|^2_{\rm max}}{P_{\rm dp}} = 3.286 \quad \text{(i.e., } 5.17\,\text{dB)} \ , \tag{5.87}$$

which is 3 dB more directive than a half-wave dipole in free space.

5.4.2 Slot admittance when excited by voltage source

We will now find the radiation resistance (or rather conductance) when the slot is located in a thin metal sheet with free space on both sides, being excited by a lumped voltage source $V_{_0}$ over its center. The impressed voltage creates a slot field which radiates equally into the two half-spaces on both sides of the slot. The slot resembles two opposite sections of a slot line shorted at $x = \pm l/2$. A standing wave is set up on the line such that the electric field in the slot becomes approximately

$$\mathbf{E}_{\rm slt}(x') = -\frac{V_{_0}}{w} \sin\left[k\left(\frac{l}{2} - |x'|\right)\right] / \sin(kl/2)\hat{\mathbf{y}} \ . \tag{5.88}$$

This is equal to the field in Eq. (5.81) when $l = \lambda/2$, and does not give significantly different results for other small lengths either.

The radiation field of this slot which radiates into both sides of the ground plane can be found by treating each side in the way we treated one side in the previous subsection. This means that the field in the region $z > 0$ is identical to that given before, and in the region $z < 0$ we get a similar formula by going through the same derivation using $\mathbf{M}_{z<0} = V_{_0}\mathbf{m}_{z<0} = -\mathbf{E}_{\rm slt}(\hat{\mathbf{z}}) \times \hat{\mathbf{z}}$ instead of $\mathbf{M} = V_{_0}\mathbf{m} = \mathbf{E}_{\rm slt} \times \hat{\mathbf{z}}$, as the normal to the slot is directed along $-\hat{\mathbf{z}}$ on the rear side (see Fig. 5.23). Now by comparison with (5.86) the power integral becomes

$$P_{\rm slt} = P_{\rm dp}\left(\frac{2V_{_0}}{\eta I_{_0}}\right)^2 \ . \tag{5.89}$$

From $P_{\rm rad} = V_{_0}^2/(2R_{\rm slt})$ the *radiation resistance* is found to be

$$R_{\rm slt} = \frac{V_{_0}^2}{2P_{\rm rad}} = \frac{\eta V_{_0}^2}{P_{tot}} = \frac{\eta^2}{4R_{\rm dp}} = 487\,\Omega \ , \tag{5.90}$$

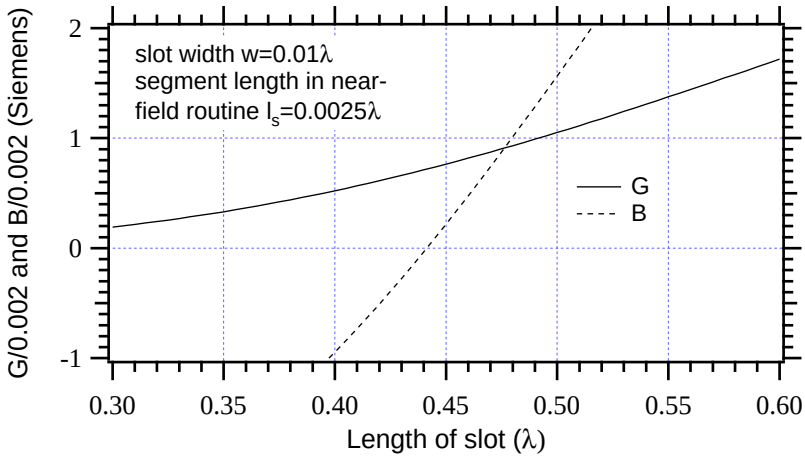


Figure 5.24: ⋆Conductance G and susceptance B of voltage excited slot. The values are presented relative to 0.002 Siemens, i.e., 500 Ω, by reference to Eq. (5.90).

where $R_{\rm dp} = 73\,\Omega$ is the radiation resistance of the electric dipole. We can calculate the complete self-admittance by using the duality relation in (4.56) between the impedance and admittance of electric and magnetic current distributions. Using this, and accounting for the fact that our present magnetic current is excited by a voltage $2V_0$ (and not V_0), for the slot admittance we obtain

$$Y_{\rm slt} = \frac{4}{\eta^2} Z_{\rm dp} \ , \tag{5.91}$$

where $Z_{\rm dp}$ is the impedance of a strip dipole with the same electric current distribution as the E-field distribution of the slot. At resonance, expression (5.91) yields $R_{\rm slt} = \eta^2/(4R_{\rm dp})$ which is the same as (5.90). We have evaluated (5.91), and the result is presented in Fig. 5.24⋆.

If we had evaluated the slot admittance for different slot widths (which must be smaller than the length), we would have seen the following[15]: *A resonant antenna which is excited at a voltage maximum has an input conductance which is not sensitive to variations in the geometry and the analysis model, whereas the susceptance is very sensitive to both such variations. The input impedance shows a typical resonant behavior characterized by a resistance peak at the resonance frequency where the reactance is zero.*

The complex *radiation admittance* can also be calculated by using self-reaction, which corresponds to a *Method of Moments* approach with one basis function and the same test function. We are now dealing with a voltage (or rather equivalent magnetic current) source, so we get the admittance instead of the impedance by normalizing the reaction integral to the slot voltage[16]. The $z > 0$ and $z < 0$ problems give both the same admittance, and, these are in parallel so that the total admittance seen at the terminal of the voltage source must be the sum of them. i.e.,

$$Y_{\rm slt} = \frac{1}{V_0^2}\langle V_0 \mathbf{H}_m(x,y), V_0 \mathbf{m}(x,y)\rangle + \frac{1}{V_o^2}\langle V_0 \mathbf{H}_{mm}(x,y), V_0 \mathbf{m}^m(x,y)\rangle \ . \tag{5.92}$$

[15] This is related to the corresponding conclusion for current excited antennas, see page 182.
[16] According to Section 4.5.4 on page 149.

Here, $\mathbf{H}_m$ and $\mathbf{H}_{mm}$ are the H-fields in the slot calculated by using m for the $z > 0$ problem and $\mathbf{m}^m$ for the ($z < 0$) problem, respectively. The former gives

$$\mathbf{H}_m = \frac{1}{\eta} C_k \int_{-w/2}^{w/2} \int_{-l/2}^{l/2} [2\mathbf{m} C_{N_1} - (2\mathbf{m} \cdot \hat{\mathbf{R}}) \hat{\mathbf{R}} C_{N_2}] \frac{1}{R} e^{-jkR} dx' dy' \ , \tag{5.93}$$

where C_k, C_{N_1}, C_{N_2}, R and $\hat{\mathbf{R}}$ are defined in (4.34) and (4.41)-(4.42). The factor 2 in front of $\mathbf{m}$ comes from imaging. The reaction integral must be evaluated numerically, and this is complicated because the H-field is singular when $x = x'$ or $y = y'$. However, we can avoid these problems by using the near-field algorithm in Section 4.7.2, for magnetic current sources. The result will be the same as using (5.91) if the current distribution used to evaluate $Z_{\rm dp}$ is the same as the field distribution used to evaluate. Note that the two reaction integrals in (5.92) are equal due to symmetry, so that we actually only need to evaluate one of them.

5.4.3 Slot excited by plane wave

We will now study the same slot as in the previous subsection, but when it is excited by an incident plane wave. The main objective is to explain how to calculate the total power transmitted through the slot. In this case we do not know the slot voltage, so we have to determine it from the known amplitude E_i of the incident wave. The equivalent problem for $z < 0$ has now two sources, $\mathbf{M}^m = \mathbf{E}_{\rm slt} \times (-\hat{\mathbf{z}})$ and the plane wave $\mathbf{E}_i = E_i e^{-jkz} \hat{\mathbf{y}}$. The equivalent problem for $z > 0$ has one source $\mathbf{M} = \mathbf{E}_{\rm slt} \times \hat{\mathbf{z}}$. We can find the tangential E-field in the slot by using the boundary condition that the x- and y-components of $\mathbf{H}_{\rm slt}$ at $z = 0$ must be equal for both equivalent problems.

We consider first the equivalent problem of the $z < 0$ region. The incident wave is reflected by the ground plane, so the total H-field at $z = 0$ due to the incident and reflected field becomes

$$\eta \mathbf{H}_i = -2E_i \hat{\mathbf{x}} \ . \tag{5.94}$$

The total E-field is zero at $z = 0$. The H-field $\mathbf{H}_{\rm Mm}$ due to the magnetic current $\mathbf{M}^m$ is given by (5.93). The equivalent problem for $z > 0$ gives the H-field $\mathbf{H}_M$. By investigating (5.93) we find that $\mathbf{H}_{\rm Mm} = -\mathbf{H}_M$ at $z = 0$. Finally, applying the boundary condition, we get

$$[\mathbf{H}_i + \mathbf{H}_{\rm Mm}]_{\rm tan} = [\mathbf{H}_M]_{\rm tan} \ . \tag{5.95}$$

This represents an integral equation with the voltage V_0 of the slot as unknown. We can solve this by using the *Method of Moments* with one basis function. We choose $\cos(\pi x'/l)$ both as basis and weighting functions, i.e., *Galerkin's method*. The cosine distribution is known from (5.81) to represent a plausible approximation to the field. Then, (5.95) gives

$$V_0 = \frac{1}{2} \langle \mathbf{H}_i, \cos(\pi x/l) \hat{\mathbf{x}} \rangle / \langle \mathbf{H}_m, \cos(\pi x/l) \hat{\mathbf{x}} \rangle \ , \tag{5.96}$$

where $\mathbf{H}_m$ now is normalized and given by (5.93). $\mathbf{H}_i$ is the incident H-field at the PEC calculated for the case that the slot is not present. The only part of (5.96) which depends on the incident field is the numerator, which is

$$\langle \mathbf{H}_i, \cos(\pi x/l) \hat{\mathbf{x}} \rangle = \int_{-w/2}^{w/2} \int_{-l/2}^{l/2} \mathbf{H}_i \cdot \hat{\mathbf{x}} \cos(\pi x/l) dx dy \ . \tag{5.97}$$

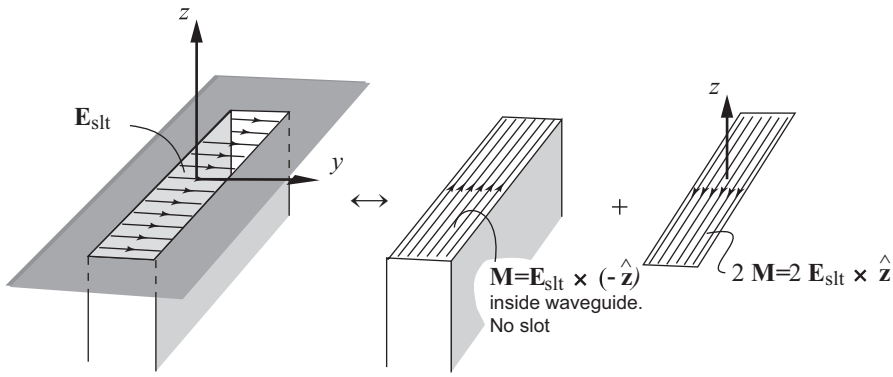


Figure 5.25: Open waveguide radiating from infinite ground plane (left), the equivalent problem of the waveguide region (center), and of the radiating half-space region (right).

This has a maximum when $\mathbf{H}_i$ is parallel with the direction $\hat{\mathbf{x}}$ of the extension of the slot. It also increases with increasing l. Therefore, the excitation voltage V_0 of the slot is strongest when the incident H-field $\mathbf{H}_i$ is parallel with the slot length.

The induced currents in the walls are given by $\mathbf{J} = 2(-\hat{\mathbf{z}}\times\mathbf{H}_i)$. Therefore, we may alternatively express (5.97) as follows: The more the electric current is disturbed by the slots, the stronger is the excitation, V_0, of the slot. This conclusion is general and valid also for other shapes of the metal wall and for any orientation of the slot.

After V_0 has been determined, the amplitude of the scattered field for $z > 0$ can be calculated and thereby also the total radiated power.

5.4.4 Reflection coefficient of open waveguide

The open narrow waveguide radiating from an infinite ground plane and its two equivalent field problems are shown in Fig. 5.25. The difference from the previous problem is that $\mathbf{H}_i$ and $\mathbf{H}_{\mathrm{Mm}}$ are now waveguide field solutions. The procedure is otherwise the same. We will not go into details here.

5.4.5 Slots in waveguide walls

The most commonly used slot antennas are made from slots in the walls of a waveguide. When several slots are located after each other along the same waveguide, the slots in the beginning of the waveguide must be weakly excited, and the slots at the end of the waveguide must be strongly excited in order to ensure that the power leaving the slot array from the beginning and end of it is similar in level.

We study the rectangular waveguide with a TE_{10} mode as an example. Different slot arrangements are shown in Fig. 5.26. There are longitudinal slots, transverse broad wall slots and

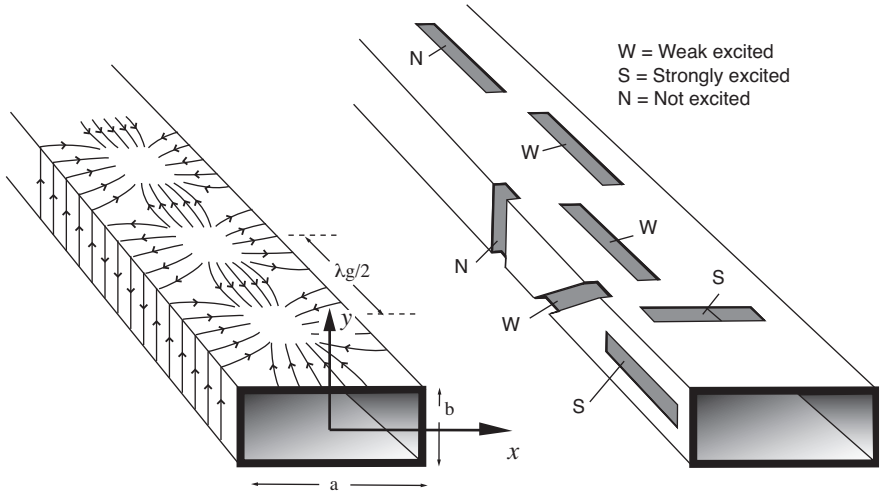


Figure 5.26: Current distribution in rectangular waveguide walls for TE_{10} mode (left), and different weakly or strongly excited slots (right).

transverse tilted narrow wall slots. The electric current distribution in the walls of a waveguide without slots can be found from microwave textbooks with waveguide field solutions. In the broad wall at $y = b/2$ the current is

$$J_x = H_z = -j \sin\left(\frac{\pi x}{a}\right) e^{-j\beta_{10} z} \ , \tag{5.98}$$

$$J_z = H_x = -\frac{-\beta_{10} a}{\pi} \cos\left(\frac{\pi x}{a}\right) e^{-j\beta_{10} z} \ , \tag{5.99}$$

and in the transverse wall at $x = -a/2$

$$J_y = j e^{-j\beta_{10} z} \ . \tag{5.100}$$

In these equations $\beta_{10} = 2\pi/\lambda_g = \sqrt{k^2 - (\pi/a)^2}$ with λ_g the guide wavelength. The wall currents are illustrated in Fig. 5.26.

Longitudinal slots: We see that J_x is zero along the center line $x = 0$. Therefore, a longitudinal slot along the center line $x = 0$ does not disturb the current and hence does not radiate. However, if the longitudinal slot is displaced from the center, the excitation will increase according to $-\sin(\pi x/a)$. Therefore we can tune the size of the excitation by changing the offset from the center line. This is very convenient in slot array design in order to change the excitation along the array. In addition, we can reverse the sign of the excitation by moving the slot to the opposite side of the center line. This is also convenient because it makes it possible to locate radiating slots with $\lambda_g/2$ spacing along the waveguide and still get the same phase of the excitation. This is desirable in order to get a beam normal to the surface of the waveguide without grating-lobes. In a circuit diagram longitudinal slots are represented by a shunt admittance.

Transverse slots: A transverse slot will disturb the z-directed surface current distribution. The broken current is strongest if the slot is located at the center of the broad wall. Such slots are so strongly excited that they cannot be used in large arrays.

Sometimes it is desirable to locate slots in the narrow wall. Entirely transverse and symmetrically located narrow wall slots does not radiate because there are no longitudinal currents in the narrow walls. Therefore, in order to make such slots radiate we need to tilt the slots slightly or introduce asymmetric elements inside the waveguide. Reference [3] is strong on slot antennas.

5.5 Further reading

Dipole antennas are classic and described in details in the books by Kraus [1] and Jasik [2]. The textbook by Elliott is strong on slot antennas [3]. Wire antennas are today normally designed by using wire modeling codes.

5.6 Complementary comments by S. Maci

Some textbooks such as [5] and previous papers make use of the term effective height to describe the directivity of small antennas. The relation between them is explained here. We have already in Section 2.9 introduced the relation between the far-field function $\mathbf{G}(\hat{\mathbf{r}})$ and the complex *effective vector height* $\mathbf{h}_t(\hat{\mathbf{r}})$ in the transmitting mode by

$$\mathbf{h}_t(\hat{\mathbf{r}}) = \frac{4\pi}{I_0 jk\eta}\mathbf{G}(\hat{\mathbf{r}}) \ .$$

In receiving mode, the effective receiving height is defined by $\mathbf{h}_r(\hat{\mathbf{r}})\cdot\mathbf{E}_0 = V_{\rm ar}$ (where $V_{\rm ar}$ is the voltage of the Thevenin equivalent circuit (Fig. 2.22). Reciprocity leads to $\mathbf{h}_t(\hat{\mathbf{r}}) = \mathbf{h}_r(\hat{\mathbf{r}}) = \mathbf{h}(\hat{\mathbf{r}})$. By applying this definition to a dipole with an arbitrary oriented direction $\hat{\mathbf{l}}$, one has

$$\mathbf{h}(\hat{\mathbf{r}}) = -\left[\hat{\mathbf{l}} - (\hat{\mathbf{l}}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}}\right]\int_{-l/2}^{l/2} j(l')e^{jkl'\hat{\mathbf{l}}\cdot\hat{\mathbf{r}}}dl' \ , \tag{5.101}$$

where $j(l')$ is the normalized current along the dipole defined in Section 5.4. The above can be rewritten as $\mathbf{h}(\hat{\mathbf{r}}) = -\left[\hat{\mathbf{l}} - (\hat{\mathbf{l}}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}}\right]\tilde{j}(kl\hat{\mathbf{l}}\cdot\hat{\mathbf{r}})$ where $\tilde{j}(k_l)$ is the Fourier transform of $j(l')$. By assuming a transmission line approximation for the current, $\tilde{j}(kl\hat{\mathbf{l}}\cdot\hat{\mathbf{r}})$ assumes the form in (5.81). When $j(l')$ is positive along the dipole (which happens for $l < \lambda/2$) it follows from (5.101) that the effective vector height takes on a maximum amplitude value $h_{\max} = |\mathbf{h}(\hat{\mathbf{r}})|_{\max}$ for $\hat{\mathbf{l}}\cdot\hat{\mathbf{r}} = 0$, namely

$$h_{\max} = \tilde{j}(0) = \int_{-l/2}^{l/2} j(l')dl' \ . \tag{5.102}$$

Therefore, the maximum effective height is equal to the length of the dipole times the average value of the normalized currents. From there it follows that a short dipole has $h_{\max} = l/2$ and for a resonant dipole $h_{\max} = \lambda/\pi$. Also, there is a relationship between the effective height $h_{\max}$ and the effective area $A_{\rm em}$ [5].

$$A_{\rm em} = \frac{\eta}{4R_{\rm in}}h_{\max}^2 \ , \tag{5.103}$$

where R_{in} is the input resistance of the antenna and is the free-space characteristic impedance. Eq. (5.103) is also generally valid, i.e., not only for dipoles. It expresses the effective area as the square of the effective height weighted by the ratio between the input impedance and the free space impedance. From the general relationship between gain G and effective area, namely $A_{\text{em}} = \frac{\lambda^2}{4\pi} G$, one has

$$G = \frac{\pi\eta}{R_{\text{in}}} \left(\frac{h_{\max}}{\lambda} \right)^2 \qquad \text{and therefore} \qquad G = \frac{\pi\eta}{R_{\text{rad}}} \left(\frac{h_{\max}}{\lambda} \right)^2 , \tag{5.104}$$

where D is the directivity and R_r the radiation resistance. The effective height can also be obtained for a current loop (see (5.72)) as $\mathbf{h}(\hat{\mathbf{r}}) = jkS\hat{\mathbf{r}} \times \hat{\mathbf{n}}$ where S is the surface of the loop and $\hat{\mathbf{n}}$ its normal oriented according to the right hand rule with reference to the direction of the current in the loop. The maximum effective height of a loop is therefore $h_{\max} = kS$. Substituting the latter in (5.104) along with $D = 3/2$, one has, for a circular loop of radius a the following radiation resistance

$$R_{\text{rad}}^{\text{loop}} = \frac{\eta\pi}{6} \left(\frac{2\pi a}{\lambda} \right)^4 . \tag{5.105}$$

It is interesting to compare this expression with the one in (5.22) relevant to the radiation resistance of a short dipole of length l, which also can be found by using (5.4) with $h_{\max} = l/2$. The result is

$$\frac{R_{\text{rad}}^{\text{loop}}}{R_{\text{rad}}^{\text{dip}}} = \frac{(2\pi a/\lambda)^4}{(l/\lambda)^2} . \tag{5.106}$$

If the length of the dipole is equal to the perimeter of the loop ($l = 2\pi a$), this ratio becomes equal to $(l/\lambda)^2 \ll 1$. This means that the radiation resistance is much higher of the dipole than of the loop. This also implies that it is much easier to match a short dipole to $50\,\Omega$ than a loop of the same conductor length.

5.7 Exercises to Chapter 5

1. **Crossed dipoles above ground**: Consider the same two dipoles as in Exercise 4.3, but assume now that both of them are located a distance $h = \lambda/4$ above a ground plane that coincides with the xy-plane. Derive the radiation field and the expression for the co- and cross-polar radiation patterns. What is now the level of the first cross-polar sidelobe relative to the co-polar maximum?
2. **Scattering from matched dipole**: Consider a plane wave with E-field amplitude E_0 incident on a half-wave dipole which is terminated ideally by a conjugate matched load. Find the expression for the received power and an expression for the induced current on the half-wave dipole. Find also an expression for the far-field scattered by the induced current. What is the ratio between the scattered and received power? Explain.
3. **Directivities of two-element dipole array**: Consider two short dipoles which are located parallel with each other with a spacing $s \ll \lambda$. Write down the expression for the far-field function when both dipoles are excited with the same amplitude and phase. What is the directivity compared to a single short dipole in free space? Locate the same two dipoles co-linearly (i.e, along the same line) with a spacing s, and with the same in-phase excitation. Use the result of Section 5.1.11 to find an expression for the directivity. Compared with a simple dipole, how much more is the directivity for a long spacing s (in dB)? Explain the difference from Fig. 5.13.

4. **Infinitesimal loop over ground plane**: Derive the far-field function of a horizontal infinitesimal loop antenna when it is located a height h over a metal ground plane. Sketch the pattern when $h = \lambda/4$ and $h = \lambda/2$.
5. **Infinitesimal loop in ground plane**: Consider two coaxial cables which are connected to two close and small holes in a metal ground plane. The shield of each cable has metal contact with the ground plane, and the center conductor passes through the hole. The two center conductors are connected on the upper side of the ground plane to form a half circular loop. Derive the expression for the far-field function.
6. **Resonant loop**: Consider a resonant horizontal loop antenna which is RHC polarized in free space. Derive the far-field function by assuming $a \ll \lambda$ when evaluating the integral in (4.77). Find the co-polar RHC and cross-polar LHC radiation patterns and sketch them.
7. **Resonant loop over ground plane**: Consider the resonant loop in the previous exercise. Derive the far-field function, find the co- and cross-polar radiation patterns and sketch them when the loop is located at heights $h = \lambda/4$ and $h = \lambda/2$ over a PEC ground plane.
8. **Longitudinal slots**: Consider two longitudinal slots spaced a distance $\lambda_g/2$ along a rectangular waveguide where, λ_g is the guide wavelength. Assume that both slots are excited with equal amplitude, and the phase of the excitation of the first slot is 0.

   a) Derive an expression for the far-field function when both slots are located on the same side of the center line of the waveguide. Where is the main lobe direction?

   b) Derive the expression for the far-field function when the slots are located on opposite sides of the center line of the waveguide. Where is now the main lobe direction?

5.8 References

[1] J.D. Kraus, *Antennas*, McGraw-Hill, New York, 1950.
[2] H. Jasik, *Antenna Engineering Handbook*, McGraw-Hill, New York. 1961 (New edition 1984).
[3] R.S. Elliott, *Antenna Theory and Design*, Prentice-Hall 1981.
[4] Z. Ying, P.-S. Kildal, "Improvements of dipole, helix, spiral, microstrip patch and aperture antennas with ground planes by using corrugated soft surfaces", *IEE Proceedings of Microwave, Antennas and Propagation*, Vol. 143, No. 3, pp. 244-248, June 1996.
[5] G. Franceschetti, *Electromagnetics: Theory, Techniques, and Engineering Paradigms*, Plenum Press, New York, 1997.

Chapter 6

Microstrip antennas and spectral domain methods

Microstrip antennas are made of patterns like strips, patches and slots which are etched out in metal films on dielectric plates (substrates). These plates are most often grounded by a metal film on the opposite side. Microstrip antennas are popular because they are cheap to manufacture and in particular to mass produce. They are alternatives to dipole and slot antennas in many applications. The microstrip antenna is shorter than half a wavelength, depending on the permittivity of the dielectric substrate. Microstrip antennas are often used in linear or planar arrays.

During the two decades 1985-2005 a lot of research was directed towards developing the spectral domain theories for analysis of microstrip antennas, and several commercial software packages are now available.

Microstrip antennas can be excited in several ways as shown in Fig. 6.1. They can be *probe-fed* by a coaxial probe coming up through the ground plane. This feeding is normally used when the dielectric substrate is thin, giving only a few percent bandwidth. Another alternative is to feed the patch directly from a *microstrip line*. Then, the feed network and the patch radiators can be etched out on the same substrate, and there is no need for soldering and making holes in the substrate. This reduces manufacturing cost, but the bandwidth is still narrow. Microstrip antennas have significant ohmic losses due to the substrate. There will also be undesired radiation from the feed lines which may destroy the polarization purity and increase the sidelobe level. The *aperture-coupled* patch overcomes the disadvantages of the probe- and line-fed patches. In this case there is a metal plane between the patches and the feed lines which prevents the latter from radiating, and the patches are excited through apertures in this plane. The bandwidth can be as large as 30 % if the upper substrate is thick (close to $\lambda/4$) and made of low permittivity material. This will also reduce the ohmic losses in the substrate. The upper substrate may even be removed and the patch instead suspended by a thin dielectric film. A problem with the aperture- coupled patch is a significant back radiation from the slot itself, so an additional rear shield may be needed. With a rear shield the slot may couple to parallel plate modes between the two metal planes, which may cause problems with undesired resonances in the feed line layer.

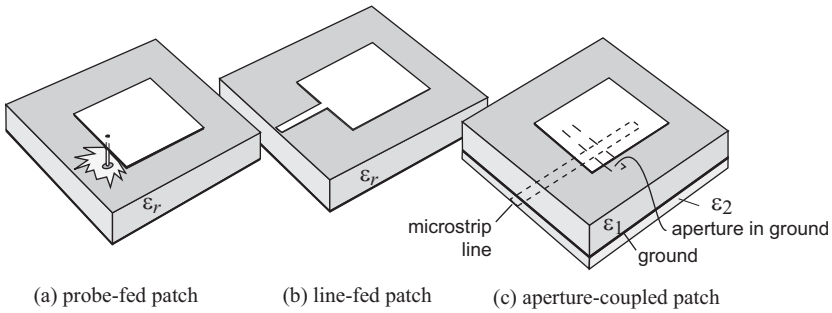


Figure 6.1: Examples of excitation methods for microstrip patch antennas. Probe-fed line-fed and aperture-coupled.

The radiating patches can have all kinds of shapes. They can be rectangular or in circular shape, and with slots in them, parasitic strips, truncated corners, tuning stubs etc. If the patch and its excitation has two planes of symmetry, the antenna will radiate linear polarization, independent of from which side it is fed. We may use asymmetric patches or excitations to create *circular polarization* from one single feed point. The patch can also be provided with two different feed lines or probes in order to radiate dual polarization. Broadband circular polarization can be obtained by locating four patches in a quadratic subarray and excite them sequentially with 90° phase shifts. There may be two or more patches stacked on top of each other in order to create dual- or multi-band performance. There exist an enormous number of possible configurations with different characteristics and many of them have already been investigated.

We will in this book describe two different analysis methods for microstrip antennas, the transmission line model and the spectral domain method. The former is simple and valid for rectangular patches on thin substrates. The latter is very accurate for all geometries provided the ground plane and the substrate are large enough that the effects of their finite widths[1] can be neglected. We will only present numerical results by using the transmission line model, as the spectral domain approach requires faster computational tools than MATLAB. There exist several commercial CAD programs for design of microstrip antennas based on the Method of Moments and the spectral domain approach.

6.1 Transmission line model for rectangular patch

We will first present the simple *transmission line model* for a rectangular patch. In this model the patch is considered as a short microstrip transmission line which is open at both ends, except at the feed point where a narrower microstrip feed line or a coaxial probe is connected. A microstrip transmission line is often described *approximately* by a planar-waveguide model with perfectly conducting magnetic walls (Fig. 6.2). The width w of the microstrip line and the relative permittivity ε_r of the substrate are then replaced by an *effective width* w_{eff} and an *effective relative permittivity* ε_{eff}, so that the *characteristic impedance* Z_k, propagation constant $\beta = 2\pi/\lambda_g$ and *guide wavelength* λ_g can be calculated by using the planar- waveguide

[1] These are called the edge effects.

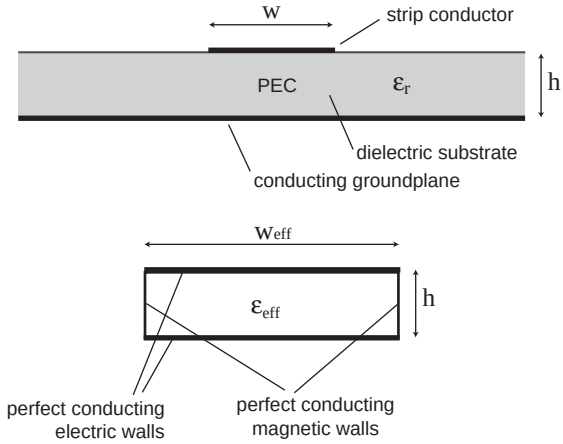


Figure 6.2: Planar-waveguide model for a microstrip line. (a) Cross section of microstrip line. (b) Cross section of planar-waveguide model with magnetic side walls.

formulas

$$Z_k = \frac{\eta}{\sqrt{\varepsilon_{\text{eff}}}} \cdot \frac{h}{w_{\text{eff}}} \ , \qquad \beta = \sqrt{\varepsilon_{\text{eff}}} k \qquad \text{and} \qquad \lambda_g = \frac{\lambda}{\sqrt{\varepsilon_{\text{eff}}}} \ , \tag{6.1}$$

where $\eta = 377\,\Omega$ is the free space wave impedance, and $k = 2\pi/\lambda$ is the wavenumber in free space. The formulas for the effective width and permittivity are based on some early and initially quasi-static approximations [1] which are improved by empirical curve-fitting [3]. The formulas are today most easily found in Chapter 10 in [4]. For infinitely thin microstrip lines they are:

$$w_{\text{eff}} = \frac{2\pi h}{\ln\left\{\frac{hF}{w} + \sqrt{1 + \left(\frac{2h}{w}\right)^2}\right\}} \ , \tag{6.2}$$

$$\text{with} \qquad F = 6 + (2\pi - 6) e^{\left(-\frac{4\pi^2}{3}\right)\left(\frac{h}{w}\right)^{3/4}}$$

$$\text{and} \qquad \varepsilon_{\text{eff}} = \frac{(\varepsilon_r + 1)}{2} + \frac{(\varepsilon_r - 1)}{2\sqrt{1 + (10h/w)}} \ , \tag{6.3}$$

where h is the substrate height, w is the patch width and ε_r is the relative permittivity of the substrate. The latter effective permittivity formula is somewhat simplified compared to that given in [4], but is still accurate for dimensions that give useful impedance values. The characteristic impedance and the relative guide wavelength are plotted in Fig. 6.3[2].

The open ends of the transmission line cause strong reflections, so the area under the patch resembles a cavity. This is resonant when the patch length l is about $\lambda_g/2$. Maximum radiation appears when the patch is resonant. The patch length l is in reality slightly shorter than $\lambda_g/2$ at resonance, due to the fringing fields at the edges of the patch. The larger the height of the substrate is, the stronger these fields are. It is convenient to define an equivalent length at resonance by $l_{\text{eq}} = l + 2\Delta l = \lambda_g/2$. For $l_{\text{eq}} = \lambda_g/2$, the fields under the patch can be

[2] There exist MATLAB code for all figures of which the caption start with ⋆.

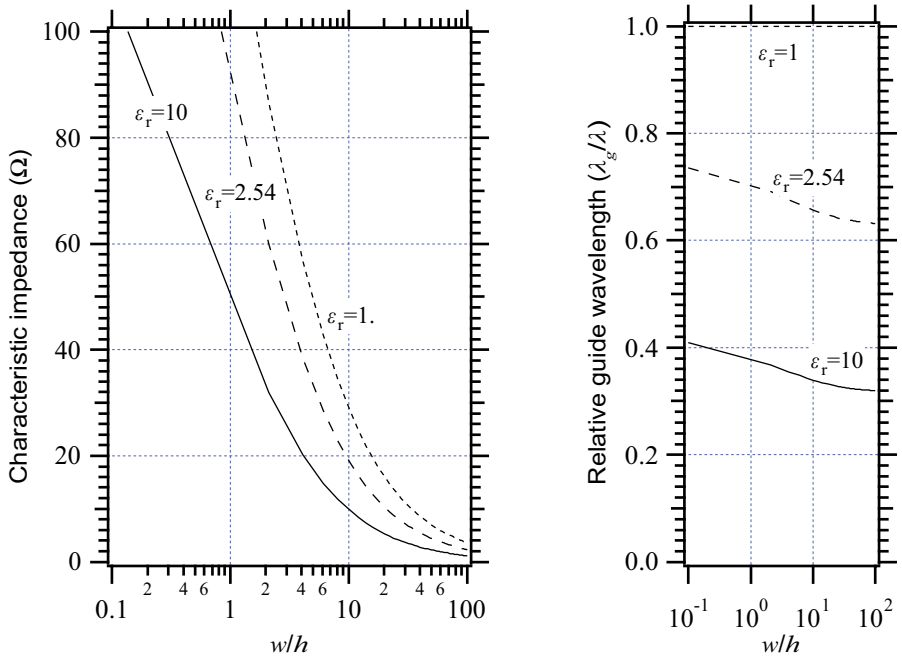


Figure 6.3: *Characteristics of microstrip line as a function of w/h. (a) Characteristic impedance. (b) Relative guide wavelength λ_g/λ.

approximated by those of the resonant microstrip line, i.e., by the sum of one forward and one backward propagating wave with unity reflection coefficient at the ends, i.e., at $y' = 0$ and $y' = l_{\text{eq}}$. This gives

$$\begin{aligned} \mathbf{E} &= E_0 e^{-j\beta y'}\hat{\mathbf{z}} + E_0 e^{-j\beta l_{\text{eq}}} e^{j\beta(y'-l_{\text{eq}})}\hat{\mathbf{z}} \ , \\ (\eta_0/\sqrt{\varepsilon_{\text{eff}}})\mathbf{H} &= E_0 e^{-j\beta y'}\hat{\mathbf{x}} - E_0 e^{-j\beta l_{\text{eq}}} e^{j\beta(y'-l_{\text{eq}})}\hat{\mathbf{x}} \ , \end{aligned}$$

where the first term on the right side represents the forward wave, and the second the backward wave. We see that $\mathbf{E}\cdot\hat{\mathbf{z}} = \pm 2E_0$ at the open ends of $y' = 0$ and $y' = l_{\text{eq}}$ since by definition $\beta l_{\text{eq}} = \pi$. By combining the two terms we get for the fields under the patch

$$\begin{aligned} \mathbf{E} &= \quad 2E_0 \cos(\beta y')\hat{\mathbf{z}} \ , \\ (\eta_0/\sqrt{\varepsilon_{\text{eff}}})\mathbf{H} &= -j2E_0 \sin(\beta y')\hat{\mathbf{x}} \ , \end{aligned} \tag{6.4}$$

for $0 < y' < l_{\text{eq}}$ and $-w_{\text{eff}}/2 < x < w_{\text{eff}}/2$. This cavity field is a valid approximation of the actual field, in spite of the fact that it apparently predicts zero H-field at the open ends of the line. In reality the H-field will have a finite value there in order to give non-zero radiated power. Furthermore, to be more accurate, expression (6.4) also provides non-zero H-field in the aperture if we evaluate the H-field at the physical edges of the patch, i.e., at $y' = \Delta l$ and $y' = l_{\text{eq}} - \Delta l$.

6.1.1 Radiation pattern by two-slot model

We may generate a convenient model for the radiation pattern by replacing the open walls of the above cavity by equivalent currents and thereafter filling the cavity with metal. This gives a magnetic current density around the patch aperture as shown in Fig. 6.4 with $\mathbf{M} = \mathbf{E} \times \hat{\mathbf{n}}$, where $\hat{\mathbf{n}} = \pm\hat{\mathbf{y}}$ at the walls at $y' = l_{\text{eq}}$ and $y' = 0$, and at the walls at $x' = \pm w/2$.We may multiply $\mathbf{M}$ with the height h of the aperture to get line currents $\mathbf{M}_l$. Then,

$$\mathbf{M}_l^{\pm} = 2E_0 h\hat{\mathbf{x}} \ , \qquad \text{at } y' = 0 \text{ and } y' = l_{\text{eq}} \ , \tag{6.5}$$

for $-w/2 < x' < w/2$ and

$$\mathbf{M}_l^{\pm} = \pm 2E_0 h \cos(\beta y')\hat{\mathbf{y}} \ , \qquad \text{at } x' = \pm w/2 \ , \tag{6.6}$$

for $0 < y' < l_{\text{eq}}$. The latter magnetic currents in $\hat{\mathbf{y}}$ direction will not radiate much because they change direction along the edge and have opposite directions at the two opposite edges[3]. The former x-directed currents add constructively and represent a plausible model for the radiation pattern of a patch on a thin substrate. When the substrate is very thin we may even neglect the substrate and remove the ground plane by imaging, which gives the following far-field function for $z > 0$

$$\mathbf{G}(\hat{\mathbf{r}}) = C_k(\hat{\mathbf{x}} \times \hat{\mathbf{r}})I_M\left[1 + e^{jkl_{\text{eq}}\hat{\mathbf{y}}\cdot\hat{\mathbf{r}}}\right] \ , \tag{6.7}$$

$$\text{with} \qquad I_M = \int_{-w_{\text{eff}}/2}^{w_{\text{eff}}/2} 4E_0 h e^{jkx'(\hat{\mathbf{x}}\cdot\hat{\mathbf{r}})} dx' \ . \tag{6.8}$$

Thus, a microstrip antenna on a thin substrate radiates in the same way as two in-phase slots (equivalent magnetic currents) in an infinite ground plane. The field distribution is uniform along the slot in contrast to the cosine distribution of the slots treated in Chapter 4. However, all these slots are short, so the different field distributions do not cause significant differences between the radiation patterns in H-plane. The E-plane radiation pattern is shown in Fig. 6.5*. The equivalent length of a patch, which is $l_{\text{eq}} = \lambda_g/2$, is significantly smaller than $\lambda/2$ for practical substrates for which $\varepsilon_r > 2.1$. The two slots modeling the microstrip antenna are so close that their combined E-plane pattern shows a significant level of radiation along the substrate, except when the substrate is missing (suspended patch) or has a permittivity close to unity (foam). Then, the spacing between the two slots is approximately half the free space wavelength, and we get a null in the direction along the substrate. Thus, in most cases a microstrip element radiates strongly along the ground plane in E-plane in the same way as a waveguide slot antenna does.

6.1.2 Impedance by transmission line model

We will now use the transmission line model to determine the impedance of the line-fed patch. The equivalent circuit is shown in Fig. 6.6. The G and B represent the radiation conductance and capacitive susceptance of each of the magnetic currents (i.e., slots), $Y_c = 1/Z_c$ is the characteristic admittance of the microstrip line formed by the patch (see (6.1)), and Z_{in} is the input impedance. The current sources $Y_{12}V_1$ and $Y_{21}V_2$ are due to the external mutual

[3] This is due to destructive interference.

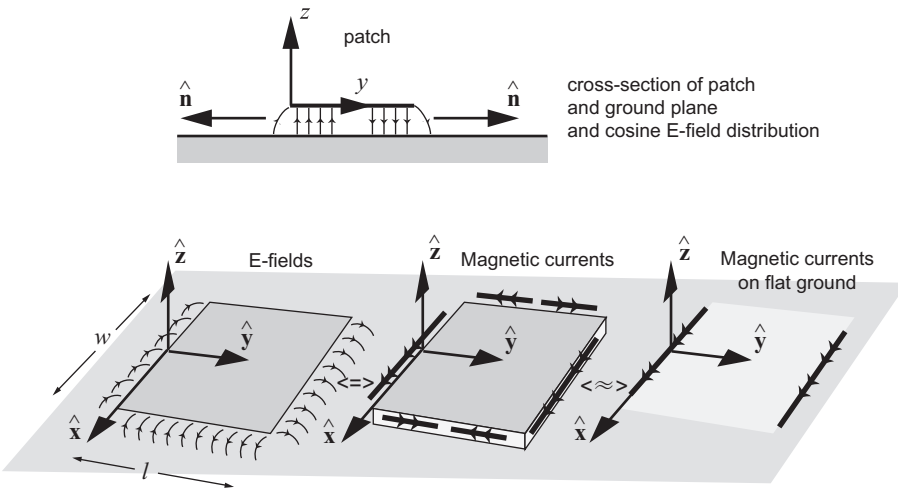


Figure 6.4: Illustration of E-field under the patch as obtained by the transmission line model (upper and lower left), equivalent magnetic currents around the patch (middle), and simplified radiation field model (right).

coupling between the slots. The values of G and B can be calculated from the self-reaction integral of a single radiating slot, and the mutual admittances from the mutual reaction between them[4]. The G can also be calculated by integrating the radiation pattern of one slot, and B can be found from the empirical formula for an equivalent length reduction Δl according to [5]:

$$\begin{aligned} & B = \beta \Delta l Y_c \ , \\ \text{with} \qquad & \Delta l = h \cdot 0.412 \left(\frac{\varepsilon_e + 0.300}{\varepsilon_e - 0.258} \right) \left(\frac{(w/h) + 0.262}{(w/h) + 0.813} \right) \ . \end{aligned} \tag{6.9}$$

We will instead calculate the real part of the overall radiation conductance $G_{\text{in}} = \Re(1/Z_{\text{in}})$ directly at resonance by using the far-field function in (6.7) to evaluate the total radiated power P_{rad} and thereafter find G_{in} from

$$P_{\text{rad}} = \frac{1}{2} G_{\text{in}} V_{_1}^2 \ . \tag{6.10}$$

The result for the radiation resistance $R_{\text{in}} = 1/G_{\text{in}}$ is shown in Fig. 6.7a*. The corresponding resonant length calculated from

$$l_r = \frac{\lambda_g}{2} - 2\Delta l \ , \tag{6.11}$$

by using (6.9), is shown in Fig. 6.7b*. The radiation resistance obtained by integrating the radiation pattern is very accurate. For most of the measured cases in [6] it gives a better value than a rigorous Method of Moments solution of the patch [7]-[8]. The resonance frequency is less accurate.

Let us now simplify and approximate the equivalent circuit in Fig. 6.6 by using

$$G = \frac{1}{2} G_{\text{in}} \ , \qquad B = \beta \Delta l Y_c \qquad \text{and} \qquad Y_{_{12}} = 0 \ . \tag{6.12}$$

[4] See Section 4.5.4 on page 149.

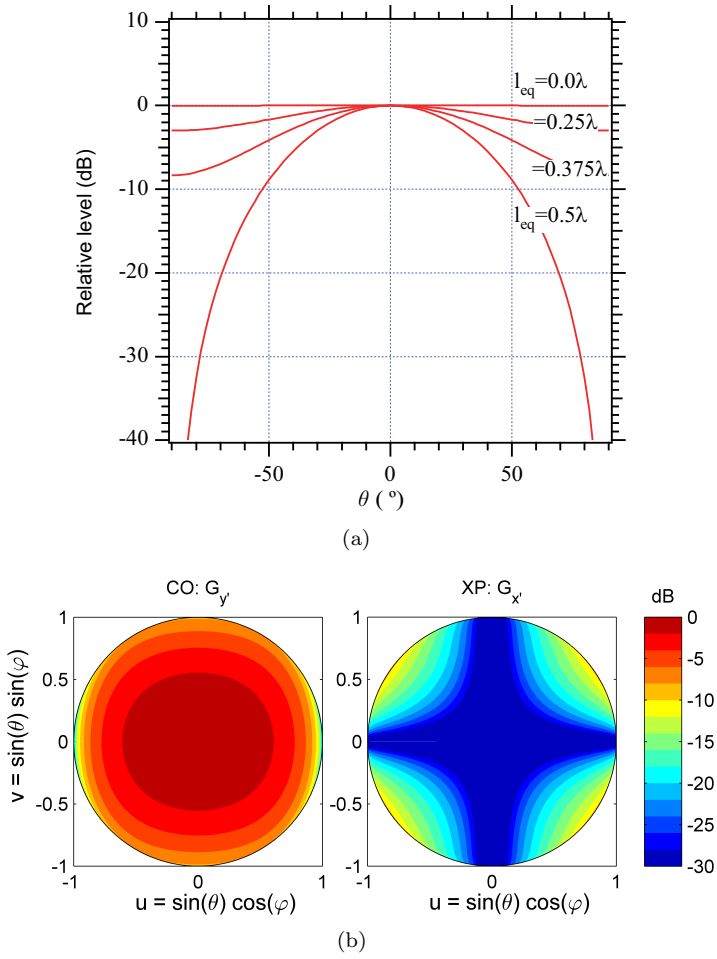


Figure 6.5: *Radiation patterns of patch antennas with different lengths based on the very thin substrate model. λ is the free space wavelength. (a) In E-plane for different l_{eq}. (b) Co- and cross-polar contour patterns for $l_{\text{eq}} = 0.375\lambda$. This corresponds to $l_{\text{eq}} = \lambda_g/2$ when $\varepsilon_e = 1.78$.

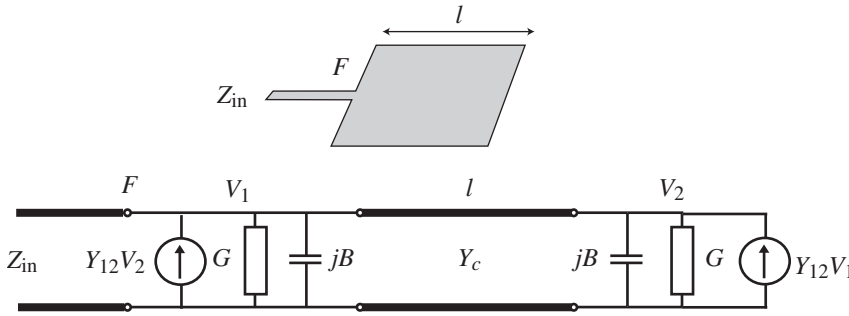


Figure 6.6: Line-fed patch and its equivalent circuit.

The first relation is valid at resonance[5], if we consider G to include even the real part of $Y_{12}V_1$ (or $Y_{21}V_2$). In reality the approximation $Y_{12} = 0$ corresponds to $\Im(Y_{12}) = 0$. We can now make a MATLAB program of the equivalent circuit, if we also make use of the impedance transformation formulas in Section 2.6.5. Fig. 6.8 shows the input impedance calculated in this way for a few different dimensions*. The MATLAB program can be used for parametric studies of variations in permittivity, substrate thickness and patch width. The method is only valid for thin substrates for which $h \ll w$. Still, it gives a good physical picture of the radiation mechanism.

The results are approximately valid also for *probe-fed patches*, if the probe is attached to the edge of the patch. We can also easily modify the equivalent circuit to account for a probe which is attached somewhere else along the center line of the patch. This is left as an exercise. See also Subsection 6.2.2. It is also possible to extend the equivalent circuit to the cases when the patch is excited for circular or dual linear polarizations, which is left as an exercise.

6.2 Self-reaction model for patch impedance

We have already introduced an equivalent circuit of the rectangular patch antenna. We now show how the patch impedance can be found directly by a numerical procedure. This procedure is readily extended to a complete Method of Moments solution for the currents on the patch and excitation probe. We consider the line-fed patch in Fig. 6.9a. We choose to model the line excitation as an ideal voltage source connected to the edge of the patch at the same point where the microstrip line is connected. We further illustrate how to calculate the impedance of the patch at this feed point. Having this, we can find the reflection coefficient on the microstrip line by using the characteristic impedance of the line. Here we limit the analysis to a single rectangular patch for which the current distribution is known quite accurately. We treat the line-fed patch, but the model for a probe-fed patch is almost identical, which will be explained at the end of the section.

The feed line can be included in the analysis by connecting the ideal excitation voltage to the feed line at its input terminal instead of the edge of the patch. The calculation approach will then be similar to what we explain below, where we solve the resulting integral equation by the *Method of Moments* in order to find the current distribution on the patch. However, if the feed line is included in the Method of Moments, we also need to include basis functions for the expansion of the current on the line.

Let us assume that the width of the microstrip line is narrow compared to the patch width, and that the line provides an excitation voltage at the center of the edge of the patch. We assume that this voltage is provided by an E-field $-E_0\hat{\mathbf{z}}$ between the edge and the ground, where $V_0 = E_0 h$ with h the height of the patch over the ground (Fig. 6.9b). Let us now introduce an imaginary infinitely thin vertical cylindrical surface with circular cross section between the ground plane and the excitation point on the patch. The tangential E-field at this surface is $E_0\hat{\mathbf{z}}$, so we may use the PEC equivalent to replace the imaginary surface by a vertical PEC probe with a magnetic current tube $\mathbf{M} = -E_0\hat{\mathbf{z}}\times\hat{\boldsymbol{\rho}} = -E_0\hat{\boldsymbol{\varphi}}$ around it (Fig. 6.9c). We have now generated a new field problem with a probe, that is completely equivalent to

[5] This is when the two slot voltages are equal, i.e., $V_1 = V_2$.

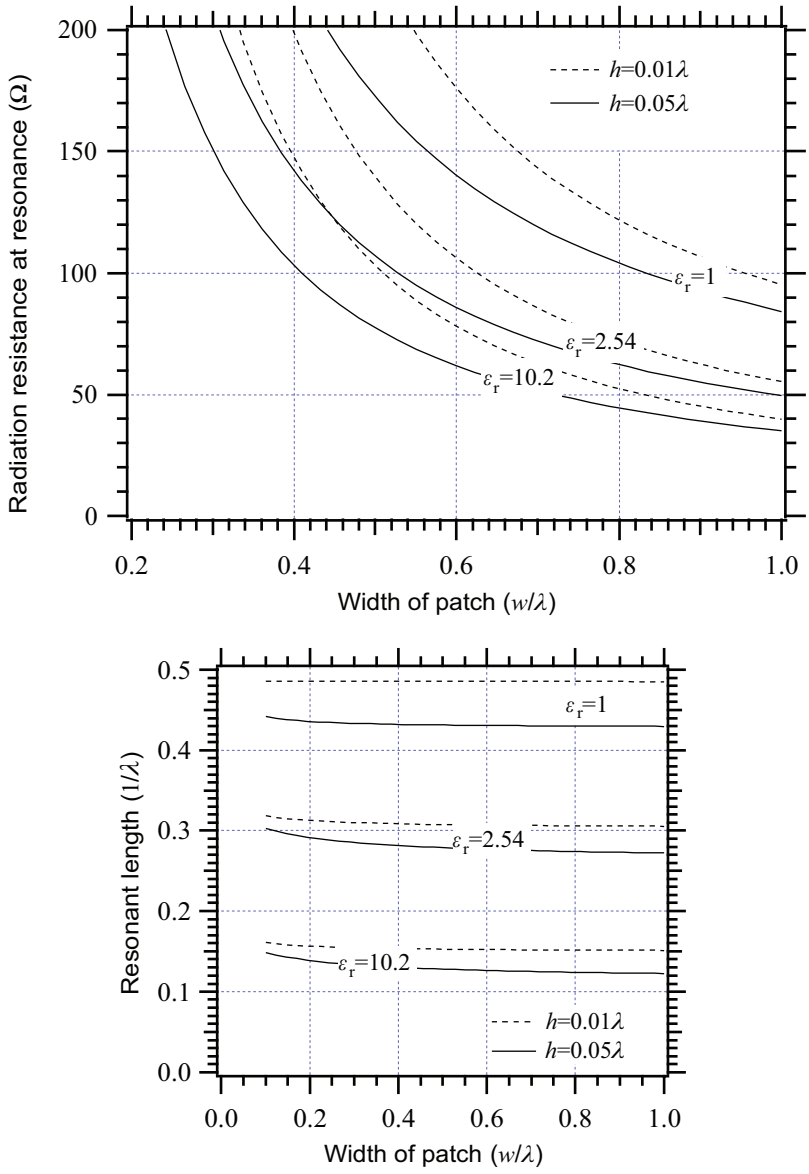


Figure 6.7: *Design parameters for rectangular patch antenna on substrates with different relative permittivity ε_r. (a) Radiation resistance at the resonance frequency. (b) Length of patch at resonance.

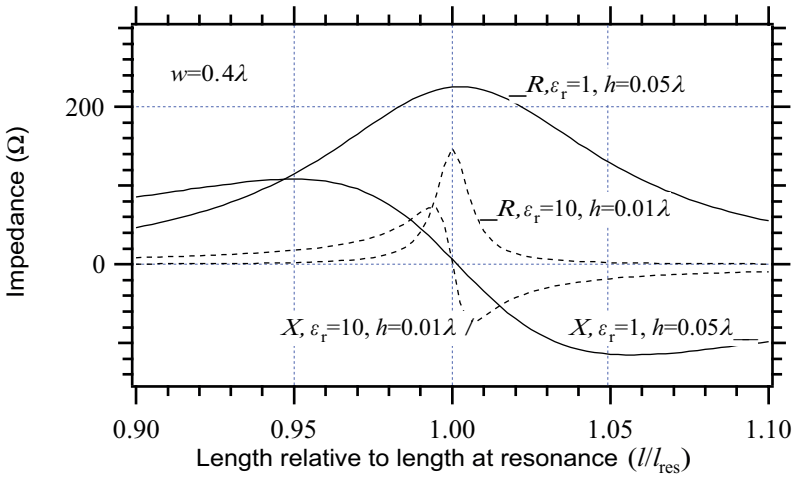


Figure 6.8: *Resistive (R) and reactive (X) part of input impedance of line-fed rectangular microstrip patch antenna for two patches on different substrates The patch widths are in both cases $w = 0.4\lambda$. R and X are plotted as a function of the patch length l relative to its resonant length.

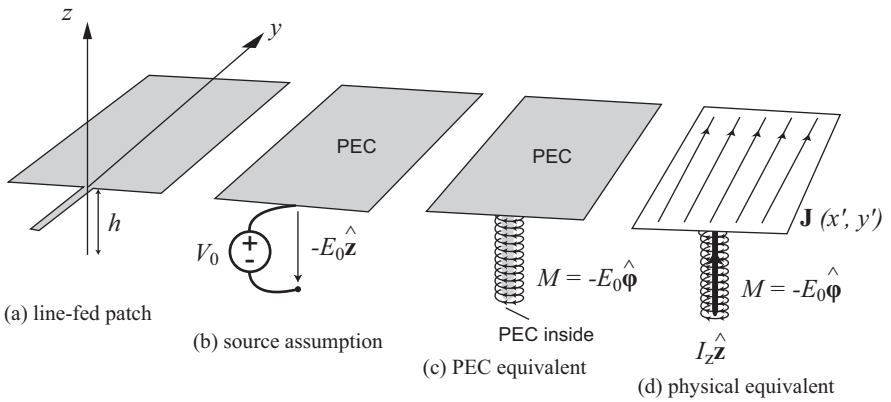


Figure 6.9: Development of the electromagnetic analysis model of a line-fed patch which is located a height h over ground.

the original problem. Nevertheless, both are suffering from the initial approximation that the E-field is $\mathbf{E} = -E_0\hat{\mathbf{z}}$ between the feed point and the ground. The probe has been introduced with the sole purpose of creating a convenient analysis model, and for this reason we refer to it as the analysis probe. The impedance of the patch can be calculated through V_0/I_z with I_z the current on the vertical analysis probe (Fig. 6.9d).

6.2.1 Expansion of current distribution and Method of Moments

We will now approximate the current distribution on the patch and the equivalent probe by basis functions. From the discussion in the previous subsection we may expect the current distribution on the patch to be $\mathbf{J}(x', y') = I_p \sin(\pi y'/l)\hat{\mathbf{y}}$. The vertical analysis probe is short and connected to the patch, so we approximate the probe current by one rectangular basis function, i.e., $I(z)\hat{\mathbf{z}} = I_z\hat{\mathbf{z}}$. These two current distribution choices do not satisfy current continuity at $y' = 0$ where the probe is connected. But, pulse basis functions always give discontinuities in the current representation, so we accept this. All in all, the model is widely accepted and has proven to give good results.

We can now formulate the boundary conditions. The total E-field (due to $\mathbf{M}$, I_z and I_p) is zero at the patch and at the surface of the probe, i.e., $\mathbf{E}_M(\mathbf{r}) + \mathbf{E}_{Iz}(\mathbf{r}) + \mathbf{E}_{Ip}(\mathbf{r}) = 0$. Let us separate the constants V_0, I_z and I_p from the E-fields and express the above equation as

$$V_0\mathbf{E}_m(\mathbf{r}) + I_z\mathbf{E}_{i_z}(\mathbf{r}) + I_p\mathbf{E}_{i_p}(\mathbf{r}) = 0 \ , \tag{6.13}$$

where $\mathbf{E}_m(\mathbf{r})$ is the E-field due to the normalized magnetic current tube $\mathbf{m} = -(1/h)\hat{\boldsymbol{\varphi}}$, $\mathbf{E}_{i_z}$ is the E-field due to a unit electric current on the probe, and $\mathbf{E}_{i_p}$ is due to a normalized electric current distribution $\sin(\pi y'/l)\hat{\mathbf{y}}$ on the patch.

Eq. (6.13) represents an integral equation which we need to solve for the two unknowns I_p and I_z in terms of V_0. We have already expanded the currents in basis functions, so we choose to test them by using $\sin(\pi y'/l)\hat{\mathbf{y}}$ as weighting function over the patch and a constant $\hat{\mathbf{z}}$ as weighting function over the probe (Galerkin's method). Consequently, we arrive at

$$V_0\langle\mathbf{E}_m(\mathbf{r}), \sin(\pi y'/l)\hat{\mathbf{y}}\rangle + I_z\langle\mathbf{E}_{i_z}(\mathbf{r}), \sin(\pi y'/l)\hat{\mathbf{y}}\rangle + I_p\langle\mathbf{E}_{i_p}(\mathbf{r}), \sin(\pi y'/l)\hat{\mathbf{y}}\rangle = 0 \ , \tag{6.14}$$

where $\langle\mathbf{E}, \mathbf{J}\rangle$ denote reaction integrals over the current distributions on the patch, as defined in Section 4.5.1, and

$$V_0\langle\mathbf{E}_m(\mathbf{r}), \hat{\mathbf{z}}\rangle + I_z\langle\mathbf{E}_{i_z}(\mathbf{r}), \hat{\mathbf{z}}\rangle + I_p\langle\mathbf{E}_{i_p}(\mathbf{r}), \hat{\mathbf{z}}\rangle = 0 \ , \tag{6.15}$$

where the reactions integrals are taken along the probe. We highly prefer to avoid field calculation due to sources along the z-direction, because in such cases the field matching at the planar substrate boundaries become complicated. Therefore, we use reciprocity to do the following replacements in (6.14)

$$\begin{aligned} &\langle\mathbf{E}_m(\mathbf{r}), \sin(\pi y'/l)\hat{\mathbf{y}}\rangle = \langle\mathbf{H}_{i_p}(\mathbf{r}), \mathbf{m}\rangle = \langle\mathbf{H}_{i_p}(\mathbf{r}), -\frac{1}{\eta}\hat{\boldsymbol{\varphi}}\rangle = 0 \ , \\ &\langle\mathbf{E}_{i_z}(\mathbf{r}), \sin(\pi y'/l)\hat{\mathbf{y}}\rangle = \langle\mathbf{E}_{i_p}(\mathbf{r}), \hat{\mathbf{z}}\rangle \ , \end{aligned} \tag{6.16}$$

where $\mathbf{H}_{i_p}(\mathbf{r})$ is the field due to the electric source $\sin(\pi y'/l)\hat{\mathbf{y}}$ evaluated at the surface of the magnetic current tube. By letting the diameter of the probe tend to zero, the reaction

between $\mathbf{H}_{i_p}(\mathbf{r})$ and $\hat{\boldsymbol{\varphi}}$ around the cylindrical surface of the analysis probe is zero. Using (6.16) we obtain the following two equations for determining I_p and I_z as a function of V_0:

$$\begin{aligned} C_{11} I_p + C_{12} I_z &= C_{13} \\ C_{21} I_p + C_{22} I_z &= C_{23} \end{aligned} \tag{6.17}$$

with

$$\begin{aligned} C_{11} &= \langle \mathbf{E}_{i_p}(\mathbf{r}), \sin(\pi y'/l)\hat{\mathbf{y}} \rangle & C_{21} &= \langle \mathbf{E}_{i_p}(\mathbf{r}), \hat{\mathbf{z}} \rangle \\ C_{12} &= \langle \mathbf{E}_{i_p}(\mathbf{r}), \hat{\mathbf{z}} \rangle & C_{22} &= \langle \mathbf{E}_{i_z}(\mathbf{r}), \hat{\mathbf{z}} \rangle \\ C_{13} &= 0 & C_{23} &= -V_0 \langle \mathbf{E}_m(\mathbf{r}), \hat{\mathbf{z}} \rangle \ . \end{aligned} \tag{6.18}$$

In order to calculate C_{23} we need to evaluate the E-field due to a magnetic current cylinder at the axis of the cylinder and integrate this along the length of the cylinder. The magnetic current cylinder was originally constructed to make the vertical electric field zero inside the cylinder when it was located in a uniform electric field $E_0\hat{\mathbf{z}}$. Therefore, the field inside the magnetic cylinder when there is no external field must be $V_0\mathbf{E}_m(\mathbf{r}) = -E_0\hat{\mathbf{z}}$, so

$$V_0 \langle \mathbf{E}_m(\mathbf{r}), \hat{\mathbf{z}} \rangle = -V_0 \ . \tag{6.19}$$

We can calculate I_p and I_z from (6.17) and arrive at finally the patch impedance formula given in the next subsection.

6.2.2 Impedance of line-fed patches

The formula for the impedance of the *line-fed* patch which results from the derivations in the previous subsection is

$$Z_{\text{pch}} = \frac{V_0}{I_z} = -\frac{\langle \mathbf{E}_{i_p}(\mathbf{r}), \hat{\mathbf{z}} \rangle^2}{\langle \mathbf{E}_{i_p}(\mathbf{r}), \sin(\pi y'/l)\hat{\mathbf{y}} \rangle} + \langle \mathbf{E}_{i_z}(\mathbf{r}), \hat{\mathbf{z}} \rangle \ . \tag{6.20}$$

This can be evaluated numerically if the Green's function of the grounded substrate is known, see Section 6.3. The evaluations are rather laborious, in particular if the source extends in z-direction as in $\langle \mathbf{E}_{i_z}(\mathbf{r}), \hat{\mathbf{z}} \rangle$, which represents the self-impedance of the analysis probe.

We may simplify (6.20) by introducing

$$Z_{\text{dp}} = -\langle \mathbf{E}_{i_p}(\mathbf{r}), \sin(\pi y'/l)\hat{\mathbf{y}} \rangle \ , \tag{6.21}$$

because the normalized self-reaction of the patch current is the "dipole over ground" impedance we would see across a feed gap at the middle of the patch[6]. This can be seen comparing the present derivations with those for the impedance of electric dipoles. If we also make use of the fact that the E-field under the patch varies sinusoidally according to (6.4), we can introduce the approximation

$$\langle \mathbf{E}_{i_p}(\mathbf{r}), \hat{\mathbf{z}} \rangle = 2E_0 h / I_p = V_0 / I_p = Z_c \ , \tag{6.22}$$

where h is the height of the patch above the ground and $Z_c = 1/Y_c$ is the characteristic impedance in (6.1) of the microstrip line forming the patch. We can simplify (6.20) even further by neglecting the self-reaction of the equivalent or actual feed probe (in the case of

[6] See Section 5.1.7 on page 178 and Section 5.1.13 on page 187.

a line-fed or probe-fed patch, respectively). Then, we finally get the much simpler equation

$$Z_{\rm pch} = \frac{V_0}{I_z} = \frac{Z_c^2}{Z_{\rm dp}} \ . \tag{6.23}$$

Thus, the impedance of an edge-fed patch is proportional to the inverse of the impedance of a center-fed flat dipole of the same width and on the same substrate if the current distributions of the two cases are the same. In order to feed the patch like a dipole we must make a feed gap or slot across the patch at its center, and feed it across the gap with a balun[7].

6.2.3 Impedance of probe-fed patches

The equation for the *probe-fed* patch will be exactly the same, except that the voltage V_0 is applied between the center conductor and the shield in the coaxial opening in the ground plane (Fig. 6.10). Therefore, the magnetic ring current will be distributed over the plane coaxial opening instead of around the probe which was the case for the line-fed patch, and we will need to numerically evaluate the distribution of $\mathbf{E}_m(\mathbf{r})$ (due to the magnetic ring current) along the probe and thereafter the reaction $\langle \mathbf{E}_m(\mathbf{r}), \hat{\mathbf{z}} \rangle$. This reaction is now different from unity, which was for the line-fed patch case (see (6.19)). Taking this into account, and solving (6.17), we arrive at the following general impedance formula for the probe-fed patch:

$$Z_{\rm pch} = \left\{ -\frac{\langle \mathbf{E}_{i_p}(\mathbf{r}), \hat{\mathbf{z}} \rangle^2}{\langle \mathbf{E}_{i_p}(\mathbf{r}), \sin(\pi y'/l)\hat{\mathbf{y}} \rangle} + \langle \mathbf{E}_{i_z}(\mathbf{r}), \hat{\mathbf{z}} \rangle \right\} \cdot \frac{1}{\langle \mathbf{E}_m(\mathbf{r}), \hat{\mathbf{z}} \rangle} \ . \tag{6.24}$$

If we move the probe to another location $y_{\rm prb}$ along the patch, (6.22) will change according to the variation of the patch current at the point where the probe is connected, i.e., according to

$$\langle \mathbf{E}_{i_p}(\mathbf{r}), \hat{\mathbf{z}} \rangle = Z_c \cos(\beta y_{\rm prb}) \ . \tag{6.25}$$

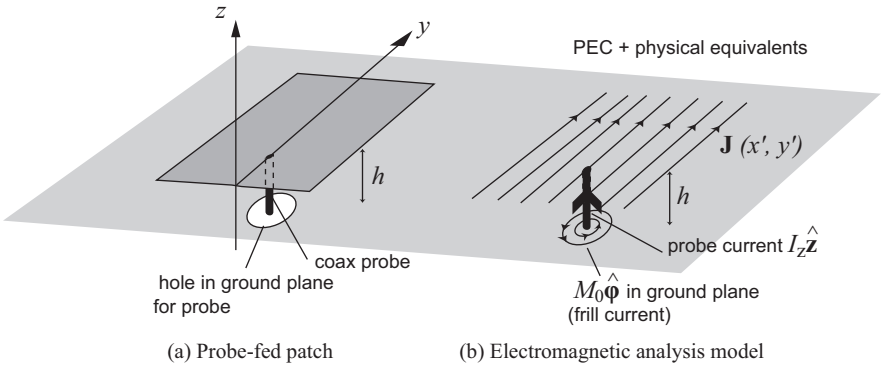
If we do similar approximations for the probe-fed patch rendering (6.23) for the line-fed patch, we achieve a simplified impedance formula

$$Z_{\rm pch} = \frac{\{Z_c \cos(\beta y_{\rm prb})\}^2}{Z_{\rm dp}} \ . \tag{6.26}$$

Thus, the patch impedance can be tuned by varying the position of the probe. The impedance is zero when the probe is located at the center of the patch for which $y_{\rm prb} = l_{\rm eq}/2 = \lambda_g/4$. In order to see a finite impedance $Z_{\rm dp}$ at the center, we must provide a feed gap with help of a balun between the coaxial opening and the two "dipole" arms[8].

If we limit the analysis to a so-called suspended patch it is possible to provide results from (6.23) and (6.26) without performing a complete spectral domain analysis. This means that the patch is suspended on a thin film or similar (or by the feed probe) at a given height h over the ground plane, in a way that there is an air gap between the patch and the ground plane. Then, the relative permittivity of the "substrate" in the above models is unity, and we can numerically evaluate the self-reaction of the sinusoidal patch current by using the near-field algorithm in Section 4.7.2, together with imaging.

[7] For more information see Section 5.1 on page 167.
[8] See the discussion after (6.23).



(a) Probe-fed patch (b) Electromagnetic analysis model

Figure 6.10: Probe-fed patch and its electromagnetic analysis model.

6.3 Spectral domain methods

The reaction formulas for the patch impedance in the previous section are valid for patches on dielectric substrates. However, we then need to evaluate the field in the presence of the dielectric material. Any multi-layer substrate can be included in the analysis by using the spectral domain method. In this method the effect of the finite lateral extents of the substrate and the ground plane are neglected. The spectral domain method is commonly referred to as a "full-wave" method because all waves (i.e., the whole spectrum of waves) are included without approximations.

Here we briefly explain the spectral domain method by using the current sheet model as described in [11]. This model is not so common but it has the advantages of a simple mathematical formulation and physically interpretable spectral domain solutions (Fig. 6.11).

The spectral domain method is strongly related to the theory of plane apertures in the sense that the three-dimensional (3D) source distributions (such as e.g., the electric currents on the patch) are Fourier transformed in two dimensions. After the Fourier transformation we interpret the spectral domain sources in the spatial domain as current sheets. Then we solve the multi-layer problem as a harmonic one-dimensional (1D) field problem, i.e., a field problem with known harmonic variations along the uniform directions of the structure and with boundary conditions applied only in the direction normal to the structure layers. Thus, the spectral domain method can also be referred to as an approach by which the multi-layer field problem is solved by using a spectrum of 1D solutions, i.e., by a spectrum of plane waves.

We illustrate the spectral domain method by using a rectangular patch which is fed by a discrete voltage source V_0 at the edge and is located on a grounded single-layer substrate with low permittivity ($\varepsilon_r \approx 1$). Nevertheless, the method is readily extendable to multiple layers of actual substrates, aperture coupling, and other patch shapes.

6.3.1 3D field problem

We use the physical equivalent to replace the rectangular metal patch by an induced electric current $\mathbf{J}(x', y')$ (Fig. 6.11a). Then, we may express the field radiated by $\mathbf{J}(x', y')$ as

$$\mathbf{E}_S(x, y, z) = \iint_S \mathbf{J}(x', y') \cdot \overline{\mathbf{G}}(x - x', y - y', z) dx' dy' \ , \tag{6.27}$$

where $\overline{\mathbf{G}}(x, y, z)$ is the *Green's function* of the layered grounded substrate, i.e., the E-field due to an incremental current source located at $x' = 0$, $y' = 0$ at the surface $z = 0$ of the substrate. The following boundary condition must be satisfied at the surface S of the patch:

$$[\mathbf{E}_S(\mathbf{r}) + \mathbf{E}_i(\mathbf{r})]_{\tan} = 0 \ , \tag{6.28}$$

where $\mathbf{E}_i(\mathbf{r})$ is the incident E-field due to the excitation, i.e., in our case the discrete voltage source V_0. If $\overline{\mathbf{G}}(x, y, z)$ is known, expressions in (6.28) with (6.27) define an integral equation for determination of the current distribution $\mathbf{J}(x', y')$. This is commonly solved by the Method of Moments. We will now illustrate how to determine $\overline{\mathbf{G}}(x, y, z)$ by Fourier transformation techniques.

6.3.2 Harmonic 1D field problem

We introduce the two-dimensional Fourier transform of the current distribution, as,

$$\tilde{\mathbf{J}}(k_x, k_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{J}(x', y') e^{jk_x x'} e^{jk_y y'} dx' dy' \ . \tag{6.29}$$

The inverse transform

$$\mathbf{J}(x', y') = \left(\frac{1}{2\pi}\right)^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{\mathbf{J}}(k_x, k_y) e^{-jk_x x'} e^{-jk_y y'} dk_x dk_y \tag{6.30}$$

can be interpreted as a superposition of a spectrum of differential sources of the form

$$\Delta\tilde{\mathbf{J}}(k_x, k_y) = \left(\frac{1}{2\pi}\right)^2 \tilde{\mathbf{J}}(k_x, k_y) e^{-jk_x x'} e^{-jk_y y'} \Delta k_x \Delta k_y \ . \tag{6.31}$$

For each value of k_x and k_y, these sources can be interpreted as current sheets located at $z = 0$, with an infinite extent in the xy-plane and with a given uniform harmonic variation in the x- and y-directions.

In this way the 3D field problem has been reduced to a superposition of *harmonic 1D field problems*, where each harmonic 1D field problem consists of a harmonic current sheet on top of a planar grounded substrate (Fig. 6.11b). The solutions to the harmonic 1D field problems can be constructed in terms of plane waves. The fields in all layers of the structure must have the same harmonic x- and y-variations as the current sheet in order to satisfy the boundary conditions at each material interface. We may express the solution of the harmonic 1D field problem as

$$\tilde{\mathbf{E}}_S(k_x, k_y, z) e^{-jk_x x} e^{-jk_y y} = \tilde{\mathbf{J}}(k_x, k_y) \cdot \tilde{\overline{\mathbf{G}}}(k_x, k_y, z) e^{-jk_x x} e^{-jk_y y} \ , \tag{6.32}$$

where $\tilde{\overline{\mathbf{G}}}(k_x, k_y, z)$ is the Green's functions of the harmonic 1D field problem. From the convolution theorem, in the theory of Fourier transforms, this is related to the *Green's function* in (6.27) by

$$\overline{\mathbf{G}}(x, y, z) = \left(\frac{1}{2\pi}\right)^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{\overline{\mathbf{G}}}(k_x, k_y, z) e^{-jk_x x} e^{-jk_y y} dk_x dk_y \ . \tag{6.33}$$

The 3D E-field is normally determined by the *Method of Moments* as follows. $\mathbf{J}(x, y)$ is expanded in basis functions $\mathbf{B}_n(x, y)$ according to

$$\mathbf{J}(x, y) = \sum b_n \mathbf{B}_n(x, y) \ , \tag{6.34}$$

where b_n are the unknown coefficients to be determined. Each basis function is Fourier transformed to a current sheet, and the corresponding harmonic 1D E-field is found from (6.32) by using the known expressions for the Green's function. This E-field is tested over the patch by using the same testing functions as basis functions (i.e., *Galerkin's method*), and the result for each basis function is inverse transformed to the space domain. Thereafter, the unknown coefficients of the basis function expansions are solved by applying the boundary condition in (6.28). When $\mathbf{J}(x, y)$ is found, $\tilde{\mathbf{E}}(x, y, z)$ can be expressed in terms of $\tilde{\mathbf{J}}(k_x, k_y)$ by using (6.32). Afterwards, we obtain the 3D field solution by the inverse Fourier transform

$$\mathbf{E}(x, y, z) = \left(\frac{1}{2\pi}\right)^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{\mathbf{E}}(k_x, k_y, z) e^{-jk_x x} e^{-jk_y y} dk_x dk_y \ . \tag{6.35}$$

The integrands of both (6.33) and (6.35) have simple analytic forms only for single layer substrates (see next subsection). Even for this simplest single layer case, the inverse transform is complicated to numerically evaluate, because the integration boundaries extend to infinity.

6.3.3 Green's function of harmonic 1D field problem

The E-field due to an electric current sheet located at $z = 0$ in free space (so that $k = 2\pi/\lambda$ with λ the wavelength in free space) is given by

$$\tilde{\mathbf{E}}_0 e^{-jk_x x} e^{-jk_y y} = \begin{cases} -\frac{k}{2k_z}[\eta\tilde{\mathbf{J}} - (\eta\tilde{\mathbf{J}}\cdot\hat{\mathbf{k}}^p)\hat{\mathbf{k}}^p] e^{-jk(\hat{\mathbf{k}}^p\cdot\mathbf{r})} & \text{for } z > 0 \\ -\frac{k}{2k_z}[\eta\tilde{\mathbf{J}} - (\eta\tilde{\mathbf{J}}\cdot\hat{\mathbf{k}}^m)\hat{\mathbf{k}}^m] e^{-jk(\hat{\mathbf{k}}^m\cdot\mathbf{r})} & \text{for } z < 0 \end{cases} \ , \tag{6.36}$$

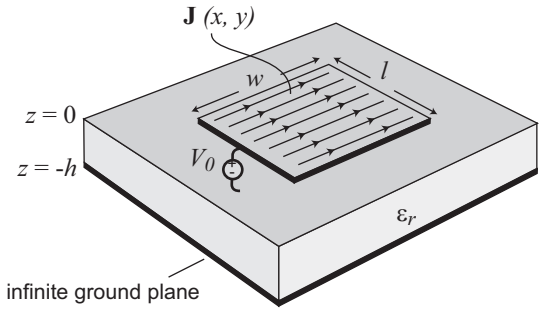
$$\text{with} \quad \tilde{\mathbf{E}}_0 = \tilde{\mathbf{E}}_0(k_x, k_y, z, \tilde{\mathbf{J}}) \ , \quad \tilde{\mathbf{J}} = \tilde{\mathbf{J}}(k_x, k_y) \ , \quad \mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}} \ ,$$

$$\hat{\mathbf{k}}^p = (k_x\hat{\mathbf{x}} + k_y\hat{\mathbf{y}} + k_z\hat{\mathbf{z}})/k \quad \text{and} \quad \hat{\mathbf{k}}^m = (k_x\hat{\mathbf{x}} + k_y\hat{\mathbf{y}} - k_z\hat{\mathbf{z}})/k \ ,$$

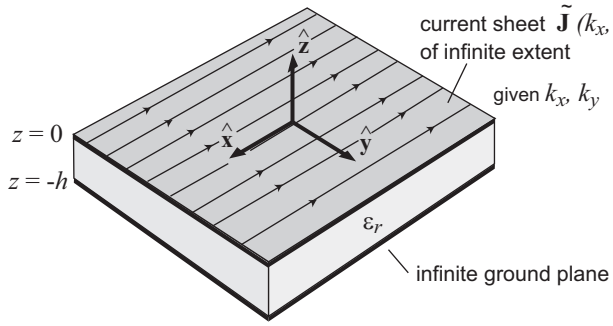
$$k_z = \sqrt{k^2 - k_x^2 - k_y^2} \quad \text{when} \quad k_x^2 + k_y^2 < k^2$$

$$\text{and} \quad k_z = -j\sqrt{k_x^2 + k_y^2 - k^2} \quad \text{when} \quad k_x^2 + k_y^2 > k^2 \ .$$

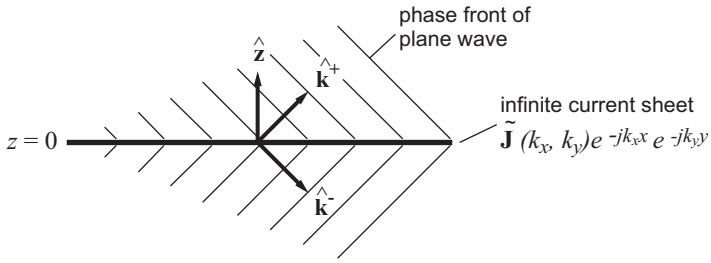
Eq. (6.36) represents the Green's function of the harmonic 1D field problem if we let $\tilde{\mathbf{J}}(k_x, k_y)$ be a unit current sheet. The current sheet corresponds to a point source in the 1D field problem for determining the z-variation. The z-component of the E-field in (6.36) has a discontinuity across the current sheet. The corresponding H-field has a discontinuity equal



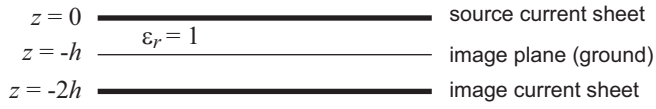
(a) Original 3D field problem with unknown electric current **J**(x, y).



(b) Harmonic 1D spectral domain problem illustrated in spatial domain with current sheet source.



(c) Radiation from current sheet in free space (i.e. Green's function of harmonic 1D field problem).



(d) Imaging of current sheet in metal ground plane. The field solution is valid in the source region $z > -h$, and it is zero below $z = -h$.

Figure 6.11: Illustration of spectral domain analysis of microstrip patch antenna.

to $\tilde{\mathbf{J}}$ in its tangential component, in order to satisfy the boundary condition at the current sheet, see (4.14) in Section 4.1.2.

The E-field in (6.36) is expressed in a form which is easily interpretable. We see that the solution represents two plane waves propagating away from the current sheet: one in the direction $\hat{\mathbf{k}}^p$ in the region $z > 0$ above the sheet[9], and the other in the direction $\hat{\mathbf{k}}^m$ in the region $z < 0$ below the sheet[10]. See Fig. 6.11c.

When the current sheet is located above a ground plane which is at $z = -h$, the E-field becomes (Fig. 6.11d)

$$\begin{aligned}&\tilde{\mathbf{E}}(k_x, k_y, z)e^{-jk_x x}e^{-jk_y y}\\ &= \left(\tilde{\mathbf{E}}_0(k_x, k_y, z, \tilde{\mathbf{J}})e^{-jk_x x}e^{-jk_y y} + \tilde{\mathbf{E}}_0(k_x, k_y, z+2h, \tilde{\mathbf{J}}_{\text{img}})e^{-jk_x x}e^{-jk_y y}\right) ,\end{aligned} \tag{6.37}$$

where the first term is obtained from (6.36) for $z > -h$ (i.e., above the ground plane), and the second is due to the image current sheet. It is obtained from (6.36) by using the argument $z + 2h$ instead of z and replacing $\tilde{\mathbf{J}}$ by the image current $\tilde{\mathbf{J}}_{\text{img}} = -\mathbf{J}$. The spectral domain solution for an actual *multi-layer* substrate can also be expressed in terms of plane waves, but the expressions are too complex to be included here. A method for calculating the spectral domain solution is described in [11].

6.3.4 Numerical implementation

The numerical implementation of the formulas in Subsections 6.3.1 to 6.3.3 is complicated. The reason for the problems is that the boundaries of the inverse integrals in (6.33) and (6.35) extend to infinity. When $k_x^2 + k_y^2 > k^2$, there will be singularities in the field solutions, and these can be interpreted as surface waves inside the substrate and needs special numerical treatment.

Several commercial and semi-commercial computer programs exist for analysis of microstrip antennas by the spectral domain approach.

6.4 Further reading

The number of papers on microstrip antenna theory and design is so large that it is impossible to give a list here. A few important books and articles are listed below in references [4] and [12]-[28]. The dual-slot model for the radiation impedance can be found in [9].

6.5 Complementary comments by S. Maci

In a conventional *Galerkin's method*, the entries of the impedance matrix of co-planar basis functions (i.e., basis function on the same substrate layer) are calculated through the convolution between basis functions $\mathbf{B}_n(\mathbf{r}) = \mathbf{F}_n(\mathbf{r})$ and the Green's functions $\overline{\mathbf{G}}$ of the problem[11],

[9] This is indicated with superscript p = plus.
[10] This is indicated with superscript m = minus.
[11] MoM is treated in Section 4.7 and we use the same notation here.

followed by an integral testing with the test function $\mathbf{W}_m(\mathbf{r}) = \mathbf{F}_m(\mathbf{r})$

$$Z_{nm} = \iint \mathbf{F}_m(\mathbf{r}) \cdot \left[\iint \overline{\mathbf{G}}(\mathbf{r} - \mathbf{r}') \cdot \mathbf{F}_n(\mathbf{r}') dS' \right] dS \ , \qquad n,\ m = 0,\ 1,\ 2, \ldots \ , \tag{6.38}$$

where $\mathbf{r}(x, y)$ denotes the space domain variable. The impedance representation (6.38) is valid only if the Green's function presents a shift-invariant form, namely when no boundary conditions are imposed in the xy-plane. This representation consists of 4 interlaced integrals. In multilayer dielectric problems with interfaces in the xy-plane, the Green's function is known in closed form only in the spectral domain. Thus, calculating directly (6.38) requires the numerical evaluation of the space Green's function $\overline{\mathbf{G}}(\mathbf{r})$ through inverse Fourier transformation of the spectral domain Green's function $\tilde{\overline{\mathbf{G}}}(\mathbf{k})$, namely

$$\overline{\mathbf{G}}(\mathbf{r}) = \frac{1}{(2\pi)^2} \iint \tilde{\overline{\mathbf{G}}}(\mathbf{k}) e^{-j\mathbf{k}\cdot\mathbf{r}} dk_x dk_y \ , \tag{6.39}$$

where $\mathbf{k}(k_x, k_y)$ represents the spectral variable associate with $\mathbf{r}(x, y)$. Eq. (6.39) can be expressed as a spectral domain *Sommerfeld integral* [29], [31] of the following kind

$$G(\mathbf{r}) \equiv G_0(\rho) = \int_0^\infty G_0(k_\rho) J_0(k_\rho \rho) k_\rho dk_\rho \ , \tag{6.40}$$

where ρ is the radial distance from the point source. These type of integrals are also encountered in many other EM problems.

Whenever the basis and test functions are very small in terms of wavelengths, using e.g., the Rao Wilton Glisson (RWG) basis functions [31], there is needed only a few points within the basis function domains to evaluate the integral (6.38). However, since the dyadic Green's function of the fields is hyper-singular at the source point, the *Mixed Potential Integral Equation* (*MPIE*) approach have to be used for treating the self-reaction [32][12].

The spectral Domain approach described in this book, can also be seen as an alternative way to formulate the matrix entries in conventional Galerkin MoM. In fact, the Eq. (6.38) can also be seen as an inverse Fourier transform of the product of the Fourier transforms of F and G in (6.38) [33]-[36], i.e.,

$$Z_{nm} = \iint \tilde{\mathbf{F}}_m(\mathbf{k}) \cdot \tilde{\overline{\mathbf{G}}}(\mathbf{k}) \cdot \tilde{\mathbf{F}}_n(-\mathbf{k}) e^{-j\mathbf{k}\cdot\Delta\mathbf{r}} dk_x dk_y \ , \tag{6.41}$$

where $\tilde{\mathbf{F}}_m(\mathbf{k})$ denote Fourier Transforms of $\tilde{\mathbf{F}}_m(\mathbf{r})$ and $\Delta\mathbf{r} = (\Delta x, \Delta y)$ is the distance between the origins of the reference systems in which the Fourier transforms of the basis and test functions are calculated. The importance of (6.39) in formulating MoM for patch antennas is due to the fact that the spectral Green's function is known in analytic form for single-layer substrates.

Despite the fact that (6.39) requires only 2 integrals, the space formulation in (6.38) is today more used that (6.39) in commercial software for planar antennas. The reason is essentially that it can treat structures of more arbitrary planar geometry, which renders the space domain approach more flexible. However, spectral domain is still used for simple shapes since it is very fast and easy to formulate. In the following, the advantages and disadvantages of the two approaches when applied to patch antenna problems are discussed.

The advantages of spectral domain SD are:

[12] See Section 4.8.

- There is no need to pre-calculate the Green's function, whose spectrum $\tilde{\overline{\mathbf{G}}}(\mathbf{k})$ is available in closed form.
- It is simple to implement due to the direct use of EFIE (Electric Field Integral Equation) without need of MPIE.
- It is possible to use entire domain basis functions for simple shapes with reduction of number of unknown.
- It is possible to obtain simple expressions for the mutual coupling between patches by using single basis functions shaped as a modal solution for the patch current.

The limitation or difficulties of SD are:

- It is not applicable when the Green's function is not shift-invariant, like for instance when treating vertical (z-directed) metallization.
- The basis functions should have closed form Fourier-spectra, otherwise there will be large computational effort for the numerical evaluation of the Fourier transforms.
- Small-domain basis functions like RWG are useful for complex geometries, but they contain a large spectral bandwidth so the integration of (6.39) converges very slowly.
- The representation in (6.39) oscillates fast and hence converges slowly when the separation between basis and test functions are large.

The advantages of the space domain are:

- It is convenient to use the small domain basis functions because the reaction integral converges fast.
- It is possible to re-use space-domain numerical codes developed for free-space problems.

The difficulties of the space domain approaches are:

- There is a need for preprocessing for calculation of the spatial domain Green's functions.
- There is a need for extracting space singularities, but these methods are available from using MPIE for conventional space domain MoM.

We note that the main difficulty of the space-domain approach is concerning with the pre-calculation of the mixed potentials Green's function.

- *Numerical integration*: Traditionally, integration along the real axis combined with pole extraction techniques and averaging methods has been employed, leading to very efficient algorithms [37]. Other methods consists of the extraction of the asymptotic value for k_ρ:

$$G_0(\rho) = \int_0^\infty \left[G_0(k_\rho) - G_\infty(k_\rho)\right] J_0(k_\rho\rho) k_\rho dk_\rho + \int_0^\infty G_\infty(k_\rho) J_0(k_\rho\rho) k_\rho dk_\rho \ . \quad (6.42)$$

  The regularizing function $G_\infty(k_\rho)$ can be chosen as the limit of $G(k_\rho)$ for small ω, ensuring that the last integral can evaluated in closed form. The latter represents the quasi-static value of the Green's function. The regularized first integral in (6.42) converges rapidly, and its convergence can be improved by deforming the real-k_ρ integration into a complex-variable contour [38].

Other related methods are listed below:

- *Complex exponential expansion*: The Green's function is expanded in terms of complex

exponentials by the Generalized Pencil of Function (GPOF) method [39]-[41]

$$G(\rho) \approx \int_0^\infty \left[\sum_{i=1}^{N} b_i \frac{e^{-j\alpha_i \sqrt{k^2-k_\rho^2}}}{\sqrt{k^2-k_\rho^2}} \right] J_0(k_\rho \rho) k_\rho dk_\rho \ ,$$

where α_i are complex coefficient with $\Im\alpha_i < 0$. The closed form can be obtained from integrating it term by term by using the Sommerfeld identity,

$$\frac{e^{-jk\sqrt{\alpha^2+\rho^2}}}{\sqrt{\alpha^2+\rho^2}} = \frac{1}{2\pi j} \int_0^\infty e^{-j\alpha\sqrt{k^2-k_\rho^2}} \frac{J_0(k_\rho\rho)}{\sqrt{k^2-k_\rho^2}} k_\rho dk_\rho \ .$$

This leads to a rapidly converging summation of spherical waves originating from a complex point (phase center). More details on these complex point-source are given in Section 7.8.

- *Complex-pole expansions:* The spectral Green's function is expanded in complex-pole terms through an algorithm called Vect-Fit [42] and [44];

$$G(\rho) \approx \int_0^\infty \left[\sum_{i=1}^{N} \frac{R_n}{k_\rho^2 - \beta_n^2} \right] J_0(k_\rho \rho) k_\rho dk_\rho \ .$$

The closed form can be obtained by integrating it term by term using the identity

$$H_0^{(2)}(\beta\rho) = \frac{2j}{\pi} \int_0^\infty \frac{1}{k_\rho^2 - \beta^2} J_0(k_\rho\rho) k_\rho dk_\rho \ .$$

The final result is given in terms of cylindrical waves propagating with complex wavenumbers β_n. This method is often applied in combination with GPOF, thus leading to a combination of complex point sources waves and cylindrical waves.

- *Uniform asymptotic evaluation*: This method makes use of a deformation of the integration contour into a contour called "Steepest Descent Path" [45]-[46]. This method is valid when the observer is one wavelengths from the source point. It has the advantage to extract physical contributions in the form

$$G(\rho) \approx G_{\rm sp}(\rho) + G_{\rm sw}(\rho) \ ,$$

where $G_{\rm sp}(\rho)$ is the space wave contribution, and $G_{\rm sw}(\rho)$ is the surface wave contribution. When inserted into the reaction integral in (6.38) and extended to an entire patch area, the asymptotic formula provides the surface wave contribution of the coupling, which is the dominant one for large separation and small losses.

The SD method is also often used to study *periodic* multilayer printed structures excited by a phased field of type $e^{j\mathbf{r}\cdot\mathbf{k}_0}$. Applications are found in the analysis of planar patch arrays, *Frequency Selective Surfaces* (*FSS*), and printed metasurfaces. For periodic structures the Floquet theorem [47] ensures that the field can be expanded in terms of plane waves $e^{j\mathbf{r}\cdot\mathbf{k}_{pq}}$, with

$$\mathbf{k}_{pq} = \mathbf{k}_0 + \frac{2\pi}{d_x} p\hat{\mathbf{x}} + \frac{2\pi}{d_y} q\hat{\mathbf{y}} \ .$$

The numerical dimension of the problem is reduced to a single periodic cell. In fact, the unknown current $J_0(\mathbf{r})$ in the cell centered at $\mathbf{r} = \mathbf{r}_0$ and the current $J_l(\mathbf{r})$ centered at $\mathbf{r} = \mathbf{r}_l$ satisfies the phase-shift condition

$$J_l(\mathbf{r}) = J_0(\mathbf{r}) e^{-j(\mathbf{r}_l - \mathbf{r}_0)\cdot \mathbf{k}_0} \ .$$

The SD MoM can be adapted to this periodic case by using the Poisson Summation Formula [48]; thus, reducing the MoM matrix entries to a sampling of the integral in (6.39) at the spectral points $\mathbf{k}_{pq}$; i.e,

$$Z_{nm} = \sum_{pq} \tilde{\mathbf{F}}_m(\mathbf{k}_{pq}) \cdot \tilde{\overline{\mathbf{G}}}(\mathbf{k}_{pq}) \cdot \tilde{\mathbf{F}}_n(-\mathbf{k}_{pq}) \ . \tag{6.43}$$

The above expression represents the reaction integral between the test currents and the field radiated by the infinite periodic distribution of basis function currents centered at points

$$\mathbf{r}_{nm}^{st} = sd_x\hat{\mathbf{x}} + td_y\hat{\mathbf{y}} + \mathbf{r}_{nm} \ , \qquad s,t = 0, \pm 1, \pm 2, ... \ ,$$

and phased by $e^{j\mathbf{r}_{st}\cdot\mathbf{k}_0}$. Several techniques have been presented in the literature to speed up this series, see [49]-[52].

There are two practical cases that should be distinguished in such type of formalism. The first one is relevant to plane wave incidence (often used for FSS or receiving arrays). In this case $\mathbf{k}_0$ is imposed by the excitation, and the objective is to find the surface currents on the metallic elements, the scattered field or the voltages at the array input ports. The second case is relevant to the determination of the modes supported by the structures. In the latter, the excitation is not imposed and $\mathbf{k}_0$ is the unknown of the problem. The dispersion equations of the surface or leaky modes $\mathbf{k}_0 = \mathbf{k}_0(\omega)$ can for such cases be found by setting the determinant of the MoM matrix equal to zero. This makes it also possible to find the bandgaps of the periodic structure, namely frequency bands where surface wave propagation is forbidden [48]. The microstrip antennas are well suited for being realized on curved geometries, see [53]-[55].

6.6 Exercises

1. **Transmission line model**: Find the total radiated power of an incremental magnetic current on a ground plane, and use this to derive an expression for the conductance G of the edge slot when the width w of the slot is very small. Compare them with the values in Fig. 6.7.
2. **Radiation pattern**: An expression for the far-field function of rectangular microstrip patch antennas is given in Section 6.1.1. It is derived under the assumption of a very thin substrate. Write out this expression as an analytic formula in terms of θ and φ in a coordinate system with z-axis normal to the patch. Sketch the radiation patterns in the E- and H-planes when the relative permittivity of the substrate is 2.54. What is the level in dB of the radiation along the ground plane relative to the level at broadside?
3. **Probe-excitation in transmission line model (with Matlab)**: The equivalent circuit in Fig. 6.6 can also be used for probe-excited patches, if the probe is connected to the edge of the patch. Modify this equivalent circuit to account for a probe which is attached to the interior of the patch. This is often done in order to tune the impedance. Implement this change in the MATLAB code for Fig. 6.8 and study how the input impedance varies with distance between the edge and the probe. Does the variation agree with (6.26)?

4. **Circular polarization with transmission line model (with Matlab)**: We can excite a single rectangular patch for radiation of circular polarization in the following way. The patch must be nearly quadratic, so that the resonance frequencies of the two orthogonal transmission lines formed by the patch will be nearly equal. We attach the probe to the patch unsymmetrically, e.g., along the diagonal of the patch, in a way that both the two orthogonal resonances are excited. Then, the patch will radiate circular polarization if the two orthogonal lengths of the patch are adjusted in a way that the impedances of the two orthogonal modes, seen at the probe, are in quadrature. Try to find the equivalent circuit for such a patch. Develop a MATLAB program based on this equivalent circuit, and try to tune the two lengths in a way that the radiated polarization will be circular.
5. See the exercises about arrays of microstrip antennas in Chapter 10.

6.7 References

[1] H.A. Wheeler, "Transmission line properties of parallel strips separated by a dielectric sheet", *IEEE Transactioins on Microwave Theory and Techniques*, Vol. MTT-13, No. 3, March 1965.

[2] E.O. Hammerstad and F. Bekkadal,*Microstrip Handbook*, ELAB Report STF44 A74169, Trondheim, Norway, February 1975.

[3] E.O. Hammerstad, I. Melhus, O. Jensen and F. Bekkadal, *Simulation of microwave components*, ELAB Report STF44F80127, Trondheim, Norway, December 1980.

[4] J.R. James and P.S. Hall (Eds.), *Handbook of Microstrip Antennas*, London, Peter Peregrinus (IEE), 1989.

[5] A.G. Derneryd, "Linearly polarized microstrip antennas", *IEEE Transactions on Antennas and Propagation*, Vol. AP-24, pp. 846-851, November 1976.

[6] D.H. Schaubert, D.M. Pozar and A. Adrian, "Effect of microstrip antenna substrate thickness and permittivity: Comparison of theories with experiment", *IEEE Transactions on Antennas and Propagation*, Vol. 37, No. 6, pp. 677-682, June 1989.

[7] D.M. Pozar, "Input impedance and mutual coupling of rectangular microstrip antenna", *IEEE Transactions on Antennas and Propagation*, Vol. AP-30, pp. 1191-1196, November 1982.

[8] D.M. Pozar and S.M. Voda, "A rigorous analysis of a microstripline fed patch antenna", *IEEE Transactions on Antennas and Propagation*, Vol. AP-35, pp. 1343-1350, December 1987.

[9] E. Lier, "Improved formulas for input impedance of coax-fed microstrip patch antennas", IEE Proceedings Pt H, Vol. 129, No. 4, pp. 161-164, August 1982.

[10] W.F. Richards, Y.T. Lo and D.D. Harrison, "An improved theory for microstrip antennas and applications", *IEEE Transactions on Antennas and Propagation,* Vol. AP-29, pp. 38-46, January 1981.

[11] Z. Sipus, P.-S. Kildal, R. Leijon and M. Johansson, "An algorithm for calculating Green's functions of planar, circular cylindrical and spherical multilayer substrates", *Applied Computational Electromagnetics Society Journal*, Vol. 13, No. 3, pp. 243-254, November 1998.

[12] D.M. Pozar and D.H. Schaubert (Ed.), *Microstrip Antennas: The Analysis and Design of Microstrip Antennas and Arrays*, IEEE Press, 1995.

[13] K.R. Carver and J.W. Mink, "Microstrip antenna technology", *IEEE Transactions on Antennas and Propagation*, Vol. AP-29, No. 1, pp. 2-24, January 1981.

[14] J.P. Daniel et al., "Research on planar antennas and arrays: Structures ", *IEEE Antennas Propagat.*, Magazine, Vol. 35, No. 1, pp. 14-38, February 1993.

[15] D.M. Pozar, and J.R. James, "A review of CAD for microstrip antennas and arrays", pp. 51-56 in [12].

[16] I.J. Bahl, and P. Bhartia,*Microstrip Antennas*, Canton, Mass., Artech House, 1980.

[17] P. Bhartia, K. Rao and R. Tomar, *Millimeter-Wave Microstrip and Printed Circuit Antennas*, Canton, Mass., Artech House, 1991.

[18] K.C. Gupta, and A. Benalla (Eds.), *Microstrip Antenna Design*, Canton, Mass., Artech House, 1988.

[19] P.S. Hall, "Review of practical issues in microstrip antenna design", *Dig. 1990 Journees Internationales de Nice sur les Antennes*, JINA'90, pp. 266-273, November 1990.

[20] J.R. James, "What's new in antennas", *IEEE Antennas and Propagat.*, Vol. 32, No. 1, pp. 6-18, February 1990.

[21] J.R. James, P.S. Hall, and C. Wood, *Microstrip Antenna Theory and Design*, London, Peter Peregrinus (IEE), 1981.

[22] R.J. Mailloux, J.F. McIlvenna, and N.P. Kernweis, "Microstrip array technology", *IEEE Transactions on Antennas and Propagation*, Vol. AP-29, pp. 25-37, January 1981.

[23] D.M. Pozar, "Microstrip antennas", *Electromagnetics*, Vol. 12, pp. 381-401, 1992.

[24] D.M. Pozar, "Radiation and scattering from a microstrip patch on a uniaxial substrate", *IEEE Transactions on Antennas and Propagation*, Vol. 35, pp. 613-621, June 1987.

[25] T. Itoh and W. Menzel, "A full-wave analysis method for open microstrip structures", *IEEE Transactions on Antennas and Propagation,* Vol. 29, pp. 63-68, January 1981.

[26] L. Vegni, R. Cicchetti and P. Capece, "Spectral dyadic Green's function formulation for planar integrated structures", *IEEE Transactions of Antennas and Propagation*, Vol. 36, No. 8, pp. 1057-1065, August 1988.

[27] J.-F. Zürcher, F.E. Gardiol, *Broadband Patch Antennas*, Artech House, 1995.

[28] F.J. Demuynck, G.A.E. Vandenbosch and A.R. Van de Capelle, "The expansion wave concept - Part 1: Efficient calculation of spatial Green's functions in a stratified dielectric medium", *IEEE Transactions of Antennas and Propagation,* Vol. 46, No. 3, pp. 397-406, March 1998.

[29] Mosig, J.R.; Gardiol, F.E., "General integral equation formulation for microstrip antennas and scatterers", *IEE Proceedings H*, Vol. 132, Issue 7, December 1985.

[30] Mosig, J.R.; Gardiol, F.E., "Analytical and numerical techniques in the Green's function treatment of microstrip antennas and scatteres", *IEE Proceedings H (Microwaves, Optics and Antennas)*, Vol. 130, Issue 2, March 1983.

[31] S. Rao, D. Wilton, and A. Glisson, "Electromagnetic scattering by surfaces of arbitrary shape", *IEEE Transactions of Antennas and Propagation*, Vol. 30, No. 3, pp. 409–418, May 1982.

[32] Michalski, K.A.; Mosig, J.R., "Multilayered media Green's functions in integral equation formulations", *Antennas and Propagation, IEEE Transactions on* , Vol. 45, No. 3, pp. 508,519, March 1997.

[33] Pozar, David M., "Microstrip antennas", *Proceedings of the IEEE*, Vol. 80, No. 1, pp. 79,91, January 1992.

[34] D. Pozar, "Input impedance and mutual coupling of rectangular microstrip an- tennas", *IEEE Transactions of Antennas and Propagation*, Vol. 30, No. 6, pp. 1191–1196, November 1982.

[35] Das, N.K.; Pozar, David M., "A Generalized Spectral-Domain Green's Function for Multilayer Dielectric Substrates with Application to Multilayer Transmission Lines", *Microwave Theory and Techniques, IEEE Transactions on*, Vol. 35, No. 3, pp. 326,335, March 1987.

[36] Pozar, David M., "A reciprocity method of analysis for printed slot and slot-coupled microstrip antennas", *Antennas and Propagation, IEEE Transactions on*, Vol. 34, No. 12, pp. 1439,1446, December 1986.

[37] J.R. Mosig, "Integral Equation Techniques", T. Itoh, Ed. New York: Wiley, 1989, *Numerical techniques for microwave and millimeter-wave passive structures*, Ch. 3, pp. 133–213.

[38] P. Gay-Balmaz and J. R. Mosig, "Three-dimensional planar radiating structures in stratified media", *International Journal of Microwave and Millimeter-Wave Computer-Aided Engineering*, Vol. 7, pp. 330–343, 1997.

[39] Mengtao Yuan; Sarkar, T.K.; Salazar-Palma, M., "A direct discrete complex image method from the closed-form Green's functions in multilayered media", *Microwave Theory and Techniques, IEEE Transactions on*, Vol. 54, No. 3, pp. 1025,1032, March 2006.

[40] Feng Ling; Jian-Ming Jin, "Discrete complex image method for Green's functions of general multilayer media", *Microwave and Guided Wave Letters, IEEE*, Vol. 10, pp. 400-402, Oct. 2000.

[41] Alparslan, A.; Aksun, M.I.; Michalski, K.A., "Closed-Form Green's Functions in Planar Layered Media for All Ranges and Materials", *Microwave Theory and Techniques, IEEE Transactions on*, Vol. 58, No. 3, pp. 602,613, March 2010.

[42] Okhmatovski, Vladimir I., and Andreas C. Cangellaris, "Evaluation of layered media Green's functions via rational function fitting", *Microwave and Wireless Components Letters, IEEE*, 14.1 (2004): 22-24.

[43] Kourkoulos, Vassilis N., and Andreas C. Cangellaris, "Accurate approximation of Green's functions in planar stratified media in terms of a finite sum of spherical and cylindrical waves", *Antennas and Propagation, IEEE Transactions on*, 54.5 (2006): 1568-1576.

[44] R. Boix, F. Mesa, and F. Medina, "Application of total least squares to the derivation of closed-form Green's functions for planar layered media", *Microwave Theory and Techniques, IEEE Transactions on*, Vol. 55, No. 2, pp. 268–280, February 2007.

[45] Barkeshli, Sina, and P.H. Pathak, "AOn the dyadic Green's function for a planar multilayered dielectric/magnetic media", *Microwave Theory and Techniques, IEEE Transactions on*, 40.1 (1992): 128-142.

[46] Marin, Miguel, Sina Barkeshli, and Prabhakar H. Pathak, "Efficient analysis of planar microstrip geometries using a closed-form asymptotic representation of the grounded dielectric slab Green's function", *Microwave Theory and Techniques, IEEE Transactions on*, 37.4 (1989): 669-679.

[47] G. Floquet, "Sur les équations différentielles linéaires à coefficients périodiques", Annales de l'École Normale Supérieure 12: 47–88, 1883.

[48] S. Maci, A. Cucini, "FSS-based EBG surfaces", book Chapter in *Electromagnetic metamaterial: physics and engineering explorations*, Edited by R. Ziolkowski and N. Engheta, Wiley Interscience, USA, pp. 57-77, 2006.

[49] F. Capolino, D.R. Wilton, and W.A. Johnson, "Efficient computation of the 2- D Green's function for 1-D periodic structures using the Ewald method", *IEEE Transactions of Antennas and Propagation*, Vol. 53, No. 9, pp. 2977–2984, September 2005.

[50] M.G. Silveirinha and C.A. Fernandes, "A new acceleration technique with exponential convergence rate to evaluate periodic Green functions", *IEEE Transactions of Antennas and Propagation*, Vol. 53, No. 1, pp. 347–355, January 2005.

[51] I. Stevanovic, J.R. Mosig, "Periodic Green's function for skewed 3-D lattices using the Ewald transformation", *Microwave and Optical Technology Letters*, Vol. 49, pp. 1353– 1357, June 2007.

[52] G. Valerio, P. Baccarelli, P. Burghignoli, and A. Galli, "Comparative analysis of acceleration techniques for 2-D and 3-D Green's functions in periodic structures along one and two directions", *IEEE Transactions of Antennas and Propagation*, Vol. 55, No. 6, pp. 1630–1643, June 2007.

[53] N. Herscovici, Z. Sipus, P.-S. Kildal, "The cylindrical omnidirectional patch antenna", *IEEE Transactions of Antennas and Propagation*, Vol. 49, No. 12, pp. 1746-1753, December 2001.

[54] S. Raffaelli, Z. Sipus, P.-S. Kildal, "Analysis and measurements of conformal patch array antennas on multilayer circular cylinder", *IEEE Transactions on Antennas and Propagation*, Vol. 53, No. 3, pp. 1105-1113, March 2005.

[55] Z. Sipus, N. Burum, S. Skokic, and P.-S. Kildal, "Analysis of spherical arrays of microstrip antennas using moment method in spectral domain", *IET Microwaves, Antennas and Propagation*, Vol. 153, No. 6, pp. 533-543, December 2006.

Chapter 7

Radiation from apertures

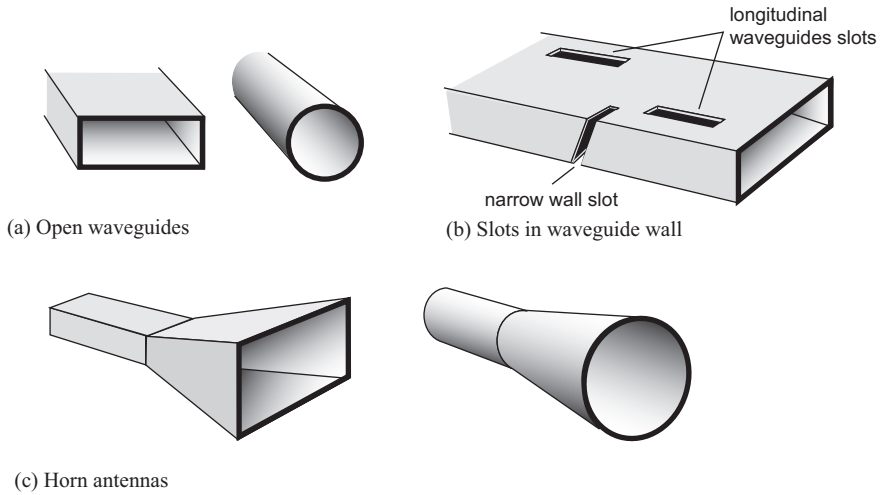
We have already analyzed radiation from small apertures (slots) in Chapter 4. In this chapter we treat apertures more systematically and generalize the theory to all kinds of apertures, as appearing in different antennas. The actual aperture antenna is radiating from openings in metal surfaces. Examples are open-ended rectangular and circular waveguides, slots in waveguide walls, and horn antennas, as shown in Fig. 7.1. The reflector and lens antennas in Fig. 7.2 are normally also called aperture antennas, although they do not radiate through physical openings or apertures. Instead, the structures look more like complementary apertures. However, the field problem to be solved is the same as that of actual aperture antennas, in the following sense: on a surface enclosing the antenna the fields are to a good approximation confined within a limited region defined by the rim of the mechanical structure. These fields are referred to as the aperture field. The surface containing the aperture field can be plane or curved. Sometimes it is advantageous to let it coincide with the desired or expected wavefront of the aperture field.

Antenna arrays are also often said to have apertures (Fig. 7.3). In an array of open waveguides the aperture is defined by the whole radiating surface of the antenna. For an array of more concentrated elements, such as dipoles, the array aperture is defined in a similar way by considering its rim to be located one half element spacing outside the center of the edge elements.

We will first explain how to calculate radiation from apertures in PECs (Section 7.1) and thereafter we will show how to use the *Huygens equivalent* to analyze both actual and complementary apertures (Section 7.2). The theory is thereafter confined to plane apertures (Section 7.3) and applied to apertures of rectangular (Section 7.4) and circular (Section 7.5) shape. Also, some microstrip antennas can be treated approximately as aperture antennas[1].

Finally, we will consider a Gaussian aperture distribution (see Section 7.6). This represents a very important analytical form of the aperture distribution, because both the near- and far-field radiation integrals have analytical solutions, and the resulting near-fields and far-fields take Gaussian shapes as well. Indeed, the *Gaussian beam* can be transformed analytically by the same analytic expression from the aperture to any field point in the near or far-field

[1] For more information see Section 6.1.1 on page 209.

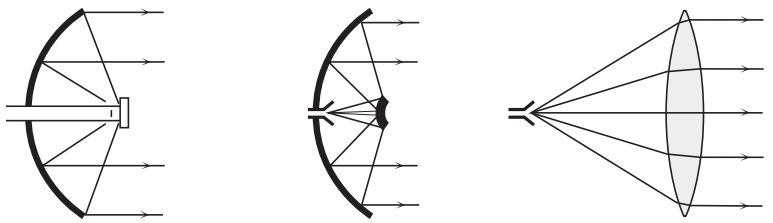


(a) Open waveguides

(b) Slots in waveguide wall

(c) Horn antennas

Figure 7.1: Examples of antennas with actual apertures.

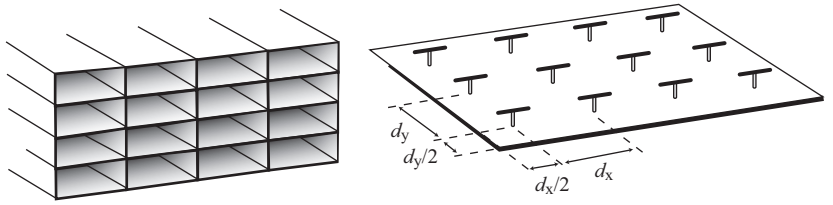


(a) Primary-fed reflector

(b) Dual-reflector antenna

(c) Lens antenna

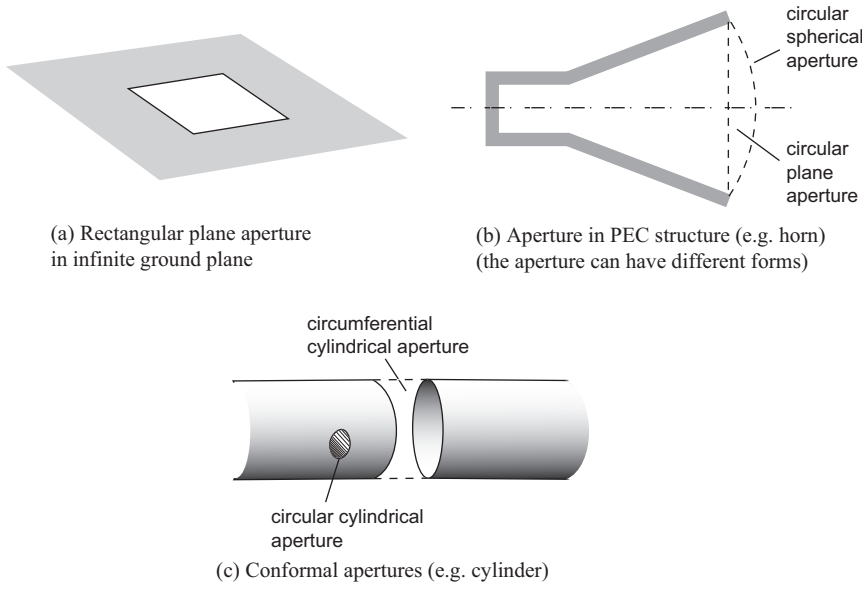
Figure 7.2: Examples of antenna arrays with apertures.



(a) Array of open waveguides

(b) Any planar array
(excited by linear or constant phase)

Figure 7.3: Examples of antennas with complementary apertures.



(a) Rectangular plane aperture in infinite ground plane

(b) Aperture in PEC structure (e.g. horn) (the aperture can have different forms)

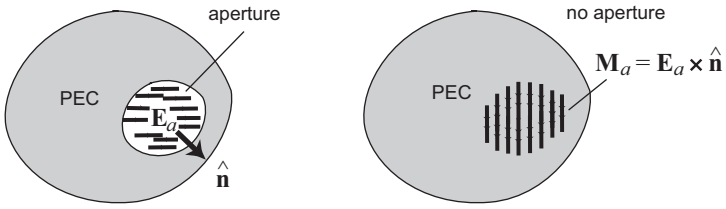
(c) Conformal apertures (e.g. cylinder)

Figure 7.4: Illustration of different actual apertures in PECs and convenient choices of aperture surfaces.

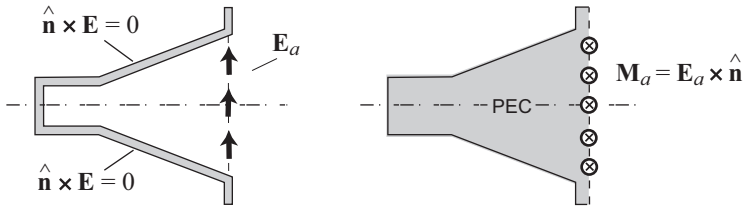
region. This property makes the Gaussian beam very useful for studying general diffraction phenomena in the near-field and far-field regions.

7.1 Apertures in PECs

The actual aperture antennas radiate from holes or openings in metal surfaces. The term aperture is used both to denote the hole itself and the virtual surface over which we describe the aperture fields. For convenience this virtual surface is chosen in a way that simplifies the analysis. This often means that it is chosen to take the form that the metal surface would have without the hole, in such a way that the aperture surface becomes conformal with this metal surface. Some actual apertures are shown in Fig. 7.4. The first is a rectangular plane aperture in a large ground plane. The second is a rotationally symmetric horn antenna. In this case we may select either a plane aperture surface, or a spherical one with its centre of curvature in the apex of the horn. Both these apertures have the same circular rim. The spherical aperture follows the expected phase-front of the aperture field, which is preferable. The plane aperture is advantageous in analytical work, since some radiation integrals for the plane case can be solved analytically, or they can be simplified to a universal form which can be easily scaled with antenna size. The two apertures in Fig. 7.4c are openings in a circular cylinder; In this case, choosing a cylindrical aperture surface conformal with the circular cylinder facilitates the analysis.



(a) Original general problem with aperture and its PEC equivalent without aperture



(b) Horn antenna with aperture and flange, and its PEC equivalent without aperture

Figure 7.5: The PEC equivalent of actual aperture antenna of arbitrary type and of horn antenna.

7.1.1 PECs of arbitrary shape

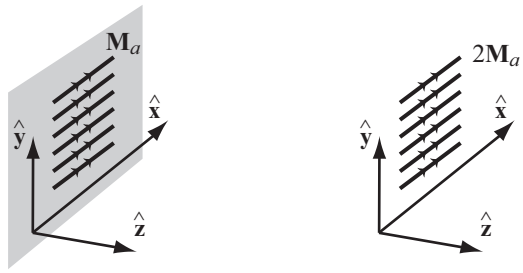
After the aperture surface has been defined, the metal parts are assumed to be PECs, and the aperture theory is developed by using the *PEC equivalent* introduced in Section 4.3.1, as follows (see Fig. 7.5a):

1. Assume an approximate tangential E-field $\mathbf{E}_a$ over the aperture. This assumption is often based on an investigation of the dominant waveguide mode that can propagate in the inner structure leading to the aperture.
2. Introduce the equivalent magnetic current $\mathbf{M}_a = \mathbf{E}_a \times \hat{\mathbf{n}}_a$, where $\hat{\mathbf{n}}_a$ is the outward directed normal to the aperture surface.
3. Fill the volume inside the aperture with PEC.
4. Calculate the induced electric currents $\mathbf{J}_S$ on the PECs caused by the sources $\mathbf{M}_a$. This can generally be done, e.g., by the *Method of Moments*.
5. Calculate the radiation fields from $\mathbf{M}_a$ and $\mathbf{J}_S$ in free space.

This method is very accurate if the assumed form of $\mathbf{E}_a$ is accurate. The calculated fields are valid everywhere except inside the surface of the PEC in the equivalent problem, such as, e.g., inside the horn antenna in Fig. 7.5b. Note that we can improve both the accuracy and the complexity of the results by moving the aperture from the opening of the horn to the opening of the waveguide feeding it, and accordingly perform step 4 on the whole horn flare. This approach closely resembles the Method of Moments approach explained in Section 8.1.6.

7.1.2 Infinite PEC planes

It is very laborious to do the calculations in step 4 of the previous subsection if the outer structure is complex, such as for horn antennas. However, for apertures in large ground



(a) The equivalent problem with $\mathbf{M}_a = \mathbf{E}_a \times \hat{\mathbf{n}}$

(b) Resulting problem after imaging: $2\,\mathbf{M}_a$ radiating in free space. Solution valid for $z > 0$.

Figure 7.6: The PEC equivalent of an aperture in an infinite ground plane and the resulting feed problem after imaging.

planes we may simplify both steps 4 and 5 considerably by using imaging, as follows (see Fig. 7.6):

1. Assume that the PEC ground plane is infinite in extent and remove it by using imaging, corresponding to replacing $\mathbf{M}_a$ on the PEC by $2\mathbf{M}_a$ radiating in free space.
2. Calculate the radiation from $2\mathbf{M}_a$ in free space.

Note that this solution is only valid on the lit side of the plane PEC, i.e., on the side where the magnetic sources are. On the shadow side the fields of the PEC equivalent are zero. In the real problem the latter fields will be given from the form of the propagating modes inside the antenna, that generates the aperture field.

The radiation field is now obtained directly from (4.48), (4.50) and (4.52) to be

$$\mathbf{E}(\hat{\mathbf{r}}) = \frac{1}{r} e^{-jkr} \mathbf{G}(\hat{\mathbf{r}}) \; ; \qquad \mathbf{G}(\hat{\mathbf{r}}) = 2C_k \mathbf{I}_{_M}(\hat{\mathbf{r}}) \times \hat{\mathbf{r}} \; , \tag{7.1}$$

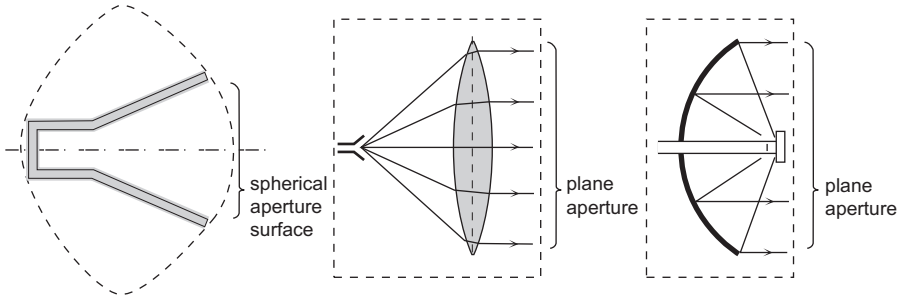
$$\text{where} \qquad \mathbf{I}_{_M}(\hat{\mathbf{r}}) = \iint\limits_A \mathbf{M}_a(\mathbf{r}') e^{jk\mathbf{r}'\cdot\hat{\mathbf{r}}} dS \; ; \qquad \mathbf{M}_a(\mathbf{r}') = \mathbf{E}_a(\mathbf{r}') \times \hat{\mathbf{n}}_a \; , \tag{7.2}$$

with the incremental source constant is $C_k = -jk/(4\pi)$ as before.

7.2 Virtual apertures in free space

By using the equivalence principal we can calculate the radiation field from known tangential E- and H-fields over any actual or virtual surface in space. This fact may be used to construct radiation field solutions originating from a convenient virtual surface around the antenna, over which we are able to find an acceptable approximation for the fields.

For a lens (or *reflector*) *antenna* we may choose a virtual plane aperture surface in front of the lens (reflector), enclosing the lens (reflector) and the feed in a virtual cylinder, see Fig. 7.7b(c). The tangential field components can be assumed to be zero everywhere over the virtual cylinder, except inside the circular plane aperture defined by the rim of the lens (reflector). In this area, we may find a good approximation for the fields by using ray

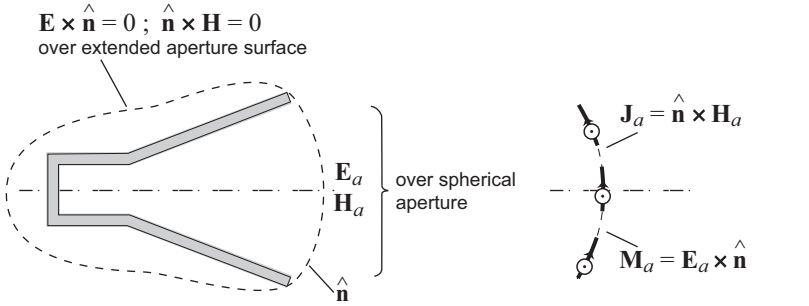


(a) Free space aperture of horn

(b) Free space aperture of lens

(c) Free space aperture of reflector

Figure 7.7: Three antennas with different choices of virtual closed surfaces S with free space apertures.



(a) Actual problem with main assumptions

(b) Equivalent calculation model

Figure 7.8: Horn antenna with aperture surface coinciding with the wavefront of the aperture field.

techniques. In the calculation of the radiation field the aperture can then be regarded as an aperture in a virtual closed surface, i.e., an aperture in free space.

We may also use the free space aperture approach to analyze some actual aperture antennas, such as *horn antennas*. Then, we may avoid the complicated Method of Moments approach described before[2]. We may do this in several ways. One way is similar to the case of the lens and reflector apertures, using a virtual plane free space aperture in front of the horn. We may also define a spherical aperture surface as shown in Fig. 7.8a. The rest of the closed surface may be chosen arbitrarily, as we can simply neglect the contribution from it by assuming zero fields there.

7.2.1 Free space and Huygens equivalents

When we have chosen the aperture surface, the aperture theory is developed by using the *free space equivalent* described in Section 4.3.2, as follows:

[2] This can be found in step 4 in Section 7.1.1 on page 234.

1. Assume approximate forms of the tangential E-field $\mathbf{E}_a$ and H-field $\mathbf{H}_a$ over the aperture in the virtual surface S. This assumption is normally based on a study of the fields that propagate from a feed point inside an enclosed surface to the aperture.
2. Assume that the tangential components of both $\mathbf{E}_a$ and $\mathbf{H}_a$ are zero elsewhere over the closed virtual surface.
3. Introduce the equivalent magnetic current $\mathbf{M}_a = \mathbf{E}_a \times \mathbf{n}_a$ and electric current $\mathbf{J}_a = \hat{\mathbf{n}}_a \times \mathbf{H}_a$ over the aperture, where $\hat{\mathbf{n}}_a$ is the outgoing unit vector normal to the aperture.
4. Calculate the radiation field from $\mathbf{J}_a$ and $\mathbf{M}_a$ in free space.

The calculated fields are valid everywhere except inside the virtual closed surface, which can be chosen arbitrary, yet preferably close to the antenna. The accuracy of the calculated radiation fields depends on the antenna type. For reflectors, lenses and arrays the approach is only accurate close to the main beam, and very inaccurate behind the aperture plane. For horn antennas with large apertures the approach is almost as accurate as the PEC aperture model, and it is much easier to use.

Sometimes we may assume that the E- and H-fields in the aperture are related by the free space wave impedance η. This is in particular a good approximation if the aperture surface coincides with the wavefront of the aperture fields and the aperture diameter is large in terms of wavelengths. Then we have

$$\eta\mathbf{H}_a = \hat{\mathbf{n}}_a \times \mathbf{E}_a \ , \qquad \text{and} \qquad \mathbf{E}_a = -\hat{\mathbf{n}}_a \times \eta\mathbf{H}_a \ .$$

This is the basic assumption of the *Huygens equivalent*, so we refer to it as the Huygens approximation. This gives

$$\eta\mathbf{J}_a = \hat{\mathbf{n}}_a \times \eta\mathbf{H}_a = -\mathbf{E}_a \ , \tag{7.3}$$

$$\text{so that} \qquad \mathbf{M}_a = \mathbf{E}_a \times \hat{\mathbf{n}}_a = -\eta\mathbf{J}_a \times \hat{\mathbf{n}}_a \ . \tag{7.4}$$

We see that $\mathbf{J}_a$ and $\mathbf{M}_a$ together become the Huygens source in Section 4.4.3 with $\hat{\mathbf{n}}$ equal to $\hat{\mathbf{n}}_a$ in (7.4). Therefore, the radiation field can now be expressed by using the free space formulas as follows

$$\mathbf{E}(\mathbf{r}) = \iint_S \mathbf{G}_{_H}(\eta\mathbf{J}_a, \hat{\mathbf{n}}_a, \hat{\mathbf{r}}) \frac{1}{R} e^{-jkr} dS' \ , \tag{7.5}$$

$$\mathbf{G}_{_H}(\eta\mathbf{J}_a, \hat{\mathbf{n}}_a, \hat{\mathbf{r}}) = C_k[\eta\mathbf{J}_a - (\eta\mathbf{J}_a \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} - (\eta\mathbf{J}_a \times \hat{\mathbf{n}}_a) \times \hat{\mathbf{r}}] \ , \tag{7.6}$$

$$\text{where} \qquad \mathbf{J}_a = \mathbf{J}_a(\hat{\mathbf{r}}) \ , \qquad R = |\mathbf{r} - \mathbf{r}'| \ , \qquad C_k = -jk/4\pi$$

and where $\mathbf{G}_{_H}(\eta\mathbf{J}_a, \hat{\mathbf{n}}_a, \hat{\mathbf{r}})$ is the far-field function of the unit Huygens source. These expressions are valid both in the radiating near-field and far-field regions. The simplified far-field expressions become

$$\mathbf{E}(\mathbf{r}) = \frac{1}{r} e^{-jkr} \mathbf{G}(\hat{\mathbf{r}}) \ ; \qquad \mathbf{G}(\hat{\mathbf{r}}) = \iint_S \mathbf{G}_{_H}(\eta\mathbf{J}_a, \hat{\mathbf{n}}_a, \hat{\mathbf{r}}) e^{jk(\mathbf{r}', \hat{\mathbf{r}})} dS' \ . \tag{7.7}$$

Note that the aperture's normal vector varies over the integration surface S, i.e., $\hat{\mathbf{n}}_a = \hat{\mathbf{n}}_a(\mathbf{r}')$, except when the aperture surface is plane. Expresion (7.5) - (7.7) are well suited for numerical integration.

7.2.2 Plane apertures

If the aperture is plane, $\hat{\mathbf{n}}_a$ is constant independent of the source coordinates. Therefore, the far-field function may be expressed as

$$\mathbf{G}(\hat{\mathbf{r}}) = C_k[\mathbf{I}_J - (\mathbf{I}_J \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} - (\mathbf{I}_J \times \hat{\mathbf{n}}_a) \times \hat{\mathbf{r}}] \ , \tag{7.8}$$

where $\mathbf{I}_J$ is the radiation integral

$$\mathbf{I}_J = \iint\limits_A \eta \mathbf{J}_a(\mathbf{r}') e^{jk(\mathbf{r}'\cdot\hat{\mathbf{r}})} dS \tag{7.9}$$

and $C_k = -jk/(4\pi)$. $\mathbf{G}(\hat{\mathbf{r}})$ in (7.8) may be expressed more compactly by introducing the dyadic Green's function $\overline{\mathbf{G}}_H(\hat{\mathbf{r}})$ of the Huygens source. Then,

$$\mathbf{G}(\hat{\mathbf{r}}) = \mathbf{I}_J \cdot \overline{\mathbf{G}}_H(\hat{\mathbf{r}}) \ , \tag{7.10}$$

where the far-field function $\mathbf{G}_H(\eta\mathbf{J}_a, \hat{\mathbf{n}}_a, \hat{\mathbf{r}})$ in (7.6) for the Huygens source is expressed in terms of the dyadic Green's function $\overline{\mathbf{G}}_H(\hat{\mathbf{r}})$ as

$$\mathbf{G}_H(\eta\mathbf{J}_a, \hat{\mathbf{n}}_a, \hat{\mathbf{r}}) = \mathbf{G}_H(\hat{\mathbf{l}}\eta J_a, \hat{\mathbf{n}}_a, \hat{\mathbf{r}}) = \eta J_a \hat{\mathbf{l}} \cdot \overline{\mathbf{G}}_H(\hat{\mathbf{r}}) \tag{7.11}$$

where $\eta\mathbf{J}_a = \eta J_a \hat{\mathbf{l}}$.

From numerical standpoint, expressions in (7.8) - (7.9) have no particular advantages compared to the more general expressions in (7.5) - (7.6). From analytical point of view, expressions in (7.8) - (7.9) are clearly advantageous as we can reduce the radiation integral to a Fourier transform.

7.3 Apertures in xy-plane

The equations in Sections 7.1 and 7.2 are quite general and desirable for making general computer programs. We will now simplify the expressions for the case that the aperture is in the xy-plane, in order to make the physical interpretation simpler, and in order to see the difference between the PEC and the free space apertures more clearly. Moreover, the aperture integral reduces to a double Fourier transform. This is more easily interpretable than the original integral. It may also have an analytic solution, and it is also faster to evaluate numerically. In this section we assume that the aperture field is given as

$$\mathbf{E}_a(\mathbf{r}') = E_{a_x}(x', y')\hat{\mathbf{x}} + E_{a_y}(x', y')\hat{\mathbf{y}} \ ; \qquad \mathbf{H}_a = \frac{1}{\eta_a}\hat{\mathbf{z}} \times \mathbf{E}_a \ , \tag{7.12}$$

with the source coordinates $\mathbf{r}' = x'\hat{\mathbf{x}} + y'\hat{\mathbf{y}}$ over the aperture A in Fig. 7.9. This equation for the H-field is based on one of the following two assumptions: either that there is one single waveguide mode present in the aperture where η_a is the aperture impedance, or, that the aperture is so large in terms of wavelength that $\eta_a = \eta = 377\,\Omega$ the free space wave impedance. In what follows we restrict the analysis to the latter so that the equivalent electric current becomes

$$\eta\mathbf{J}_a = \hat{\mathbf{z}} \times \eta\mathbf{H}_a = -\mathbf{E}_a(\mathbf{r}') \ . \tag{7.13}$$

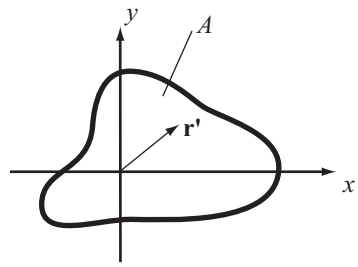


Figure 7.9: Plane aperture with area A in xy-plane.

Now the electric current radiation integral in (7.9) becomes

$$\mathbf{I}_{J}(\hat{\mathbf{r}}) = \tilde{E}_{a_x}(k_x, k_y)(-\hat{\mathbf{x}}) + \tilde{E}_{a_y}(k_x, k_y)(-\hat{\mathbf{y}}) \ , \tag{7.14}$$

where

$$\tilde{E}_{a_x}(k_x, k_y) = \iint_A E_{a_x}(x', y') e^{j(k_x x' + k_y y')} dx' dy' \tag{7.15}$$

$$\tilde{E}_{a_y}(k_x, k_y) = \iint_A E_{a_y}(x', y') e^{j(k_x x' + k_y y')} dx' dy' \tag{7.16}$$

with $k_x = k\hat{\mathbf{x}} \cdot \hat{\mathbf{r}}$ and $k_y = k\hat{\mathbf{y}} \cdot \hat{\mathbf{r}}$. Correspondingly the magnetic current radiation integral in (7.2) becomes

$$\mathbf{I}_{M}(\hat{\mathbf{r}}) = \tilde{E}_{a_y}(k_x, k_y)(\hat{\mathbf{y}} \times \hat{\mathbf{z}}) + \tilde{E}_{a_x}(k_x, k_y)(\hat{\mathbf{x}} \times \hat{\mathbf{z}}) \ . \tag{7.17}$$

We see that the radiation integrals have reduced to scalar two-dimensional *Fourier transforms*. We will in the next two subsections find the form of the far-field functions for this plane *aperture* case, for both the PEC and Huygens apertures, and we will compare the resulting formulas.

7.3.1 PEC aperture and its incremental element factor

The *far-field function* of the *PEC aperture* is now obtained from (7.1) and (7.17) to be

$$\begin{aligned} \mathbf{G}(\hat{\mathbf{r}}) = {} & 2C_k[(\hat{\mathbf{x}} \times \hat{\mathbf{z}}) \times \hat{\mathbf{r}}]\tilde{E}_{a_x}(k\hat{\mathbf{x}} \cdot \hat{\mathbf{r}}, k\hat{\mathbf{y}} \cdot \hat{\mathbf{r}}) \\ & + 2C_k[(\hat{\mathbf{y}} \times \hat{\mathbf{z}}) \times \hat{\mathbf{r}}]\tilde{E}_{a_y}(k\hat{\mathbf{x}} \cdot \hat{\mathbf{r}}, k\hat{\mathbf{y}} \cdot \hat{\mathbf{r}}) \ . \end{aligned} \tag{7.18}$$

We see that each of the two contributions to the *far-field function* is written as a product of two factors. The first factor is the far-field function of the incremental magnetic current, and the second is the radiation integral over the E-field in the aperture. The latter is expressed as a two-dimensional Fourier transform. Thus, the far-field function of a plane aperture is obtained by Fourier transformation of the aperture field multiplied this by the element factor due to the incremental magnetic current.

We may conveniently introduce the forms of the radiation field of the incremental magnetic current source in polar coordinates as given in the equations and substitute $u = \hat{\mathbf{x}} \cdot \hat{\mathbf{r}} =$

$\sin\theta\cos\varphi$ and $v = \hat{\mathbf{y}}\cdot\hat{\mathbf{r}} = \sin\theta\sin\varphi$, where the *uv-coordinates* were introduced in Section 2.3.6. Then, we obtain,

$$\begin{aligned}\mathbf{G}(\theta,\varphi) = &-2C_k(\cos\varphi\hat{\boldsymbol{\theta}} - \cos\theta\sin\varphi\hat{\boldsymbol{\varphi}})\tilde{E}_{a_x}(k\sin\theta\cos\varphi, k\sin\theta\sin\varphi)\\ &-2C_k(\sin\varphi\hat{\boldsymbol{\theta}} + \cos\theta\cos\varphi\hat{\boldsymbol{\varphi}})\tilde{E}_{ay}(k\sin\theta\cos\varphi, k\sin\theta\sin\varphi)\ .\end{aligned} \tag{7.19}$$

7.3.2 Free space aperture and its incremental element factor

Similarly for the *free space aperture*, by using (7.8) and (7.14) we obtain

$$\begin{aligned}\mathbf{G}(\hat{\mathbf{r}}) = &-C_k[\hat{\mathbf{x}} - (\hat{\mathbf{x}}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - (\hat{\mathbf{x}}\times\hat{\mathbf{z}})\times\hat{\mathbf{r}}]\tilde{E}_{a_x}(k\hat{\mathbf{x}}\cdot\hat{\mathbf{r}}, k\hat{\mathbf{y}}\cdot\hat{\mathbf{r}})\\ &-C_k[\hat{\mathbf{y}} - (\hat{\mathbf{y}}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - (\hat{\mathbf{y}}\times\hat{\mathbf{z}})\times\hat{\mathbf{r}}]\tilde{E}_{a_y}(k\hat{\mathbf{x}}\cdot\hat{\mathbf{r}}, k\hat{\mathbf{y}}\cdot\hat{\mathbf{r}})\ ,\end{aligned} \tag{7.20}$$

where $\tilde{E}_{a_x}$ and $\tilde{E}_{a_y}$ are the same as before. Both terms of this *far-field function* are also a product of two factors, like in (7.18). However, the incremental element factors in front of $\tilde{E}_{a_x}$ and $\tilde{E}_{a_y}$ are different. By studying Section 4.4, we recognize them as the free space far-field functions of the incremental Huygens source polarized in negative x- and negative y-directions, respectively. Substitution of the expressions for them in polar coordinates gives

$$\begin{aligned}\mathbf{G}(\theta,\varphi) = &-2C_k\cos^2(\theta/2)[\cos\varphi\hat{\boldsymbol{\theta}} - \sin\varphi\hat{\boldsymbol{\varphi}}]\tilde{E}_{a_x}(k\sin\theta\cos\varphi, k\sin\theta\sin\varphi)\\ &-2C_k\cos^2(\theta/2)[\sin\varphi\hat{\boldsymbol{\theta}} + \cos\varphi\hat{\boldsymbol{\varphi}}]\tilde{E}_{a_y}(k\sin\theta\cos\varphi, k\sin\theta\sin\varphi)\ .\end{aligned} \tag{7.21}$$

By comparing, we see that the only difference between the far-field functions of the PEC and free space aperture models is due to the difference between the far-field functions of the incremental magnetic current source and the Huygens source, respectively.

7.3.3 Power integration over aperture and maximum directivity

The total radiated power can be calculated from the power integral defined in equation (2.65) by using

$$\mathbf{G}_{\mathrm{co}}(\theta,\varphi) = \mathbf{G}(\hat{\mathbf{r}})\cdot\hat{\mathbf{co}}^*(\theta,\varphi)\ , \tag{7.22}$$

$$\mathbf{G}_{\mathrm{xp}}(\theta,\varphi) = \mathbf{G}(\hat{\mathbf{r}})\cdot\hat{\mathbf{xp}}^*(\theta,\varphi)\ , \tag{7.23}$$

where $\hat{\mathbf{co}}^*$ and $\hat{\mathbf{xp}}^*$ are the unit vectors defining the co- and cross-polar radiation fields. The power integral must be evaluated numerically. This can be difficult and time-consuming when the aperture is large, because both the co- and cross-polar fields will have a lot of sidelobes and may vary with φ.

However, when the Huygens approximation $\eta\mathbf{H}_a = \hat{\mathbf{z}}\times\mathbf{E}_a$ is used, the *total radiated power* P_{rad} must be equal to the power passing through the aperture surface, so we may write

$$P_{\mathrm{rad}} = \iint_A \frac{1}{2}(\mathbf{E}_a\times\mathbf{H}_a)\cdot\hat{\mathbf{n}}dS = \frac{1}{2\eta}\iint_A |\mathbf{E}_a|^2 dS \tag{7.24}$$

and the so-called *power integral* defined in Section 2.3.8 becomes $P = 2\eta P_{\mathrm{rad}}$. This aperture integral is much easier to evaluate than the radiation intensity integral. Also it can often be

evaluated analytically. The directive gain of the aperture becomes

$$D_{\rm co}(\theta,\varphi) = \frac{4\pi|\mathbf{G}_{\rm co}(\theta,\varphi)|^2}{P} \ . \tag{7.25}$$

The directivity at $\theta = 0^\circ$ can be obtained from this by using $\mathbf{G}(\hat{\mathbf{r}})$ in (7.18) for the PEC case, or (7.20) for the free space case. The result is the same for both cases and becomes

$$D_0 = \frac{4\pi}{\lambda^2}\left|\iint\limits_A E_{a_{\rm co}}(x',y')dx'dy'\right|^2 / \iint\limits_A \left[|E_{a_{\rm co}}(x',y')|^2 + |E_{a_{\rm xp}}(x',y')|^2\right] dx'dy' \ , \tag{7.26}$$

where $E_{a_{\rm co}} = \mathbf{E}_a \cdot \hat{\mathbf{co}}^*(0,0)$ and $E_{a_{\rm xp}} = \mathbf{E}_a \cdot \hat{\mathbf{xp}}^*(0,0)$. Let us now introduce the mathematical inequality

$$\left|\iint\limits_A f(x,y)g(x,y)dxdy\right|^2 \le \iint\limits_A |f(x,y)|^2 dxdy \iint\limits_A |g(x,y)|^2 dxdy \ ,$$

which results from Schwarz inequality [1, Section 16.35]. If we use this with $f(x,y) = E_{a_{\rm co}}(x,y)$ and $g(x,y) = 1$, we find that

$$\left|\iint\limits_A E_{a_{\rm co}}(x',y')dx'dy'\right|^2 \le A \iint\limits_A |E_{a_{\rm co}}(x',y')|^2 dx'dy' \ ,$$

where A is the area of the plane aperture over which $\mathbf{E}_a(x',y')$ is defined. From this and (7.26) it follows that the maximum value of D_0 appears when $E_{a_{\rm xp}} = 0$ and $E_{a_{\rm co}} = \textit{constant}$ over A. This maximum is

$$D_{\max} = \frac{4\pi}{\lambda^2} A \ . \tag{7.27}$$

This is referred to as the *maximum available directivity* or gain of the antenna, and we used it already in Section 2.5.2 to define the *aperture efficiency*. Previously, in Section 5.1.5, we showed that the directivity of a short dipole and the maximum effective area on reception is related by the same formula. The latter is actually generally valid for all antennas.

We see from the above that the aperture power integrals are identically the same for PEC apertures and free space apertures, whereas the radiation intensity integrals are not identical as the radiation patterns are not equal. However, for large apertures the values of the two radiation intensity integrals asymptotically approach each other and the value of the aperture power integral. The most accurate result for the power integral is always obtained by integrating the far-field functions, in particular if the aperture diameter is small.

For convenience in the analysis to follow in Sections 7.4 and 7.5 we will omit the incremental element factors, and consider only the Fourier transforms.

7.4 Rectangular plane aperture

Rectangular horn antennas and waveguide slots have normally plane rectangular apertures. For such cases the aperture distribution can be ideally linearly polarized in the same direction

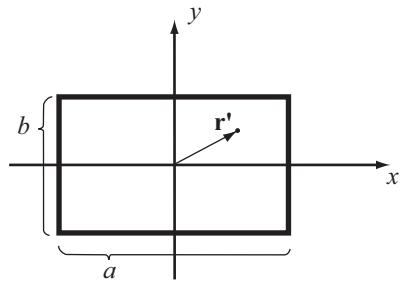


Figure 7.10: Rectangular plane aperture in the xy-plane.

over the whole aperture (if excited for linear polarization). In addition, the distribution is often separable in x' and y', according to

$$\mathbf{E}_{a_y}(x', y') = A(x')B(y')\hat{\mathbf{y}} \qquad \text{for } |x'| < a/2,\ |y'| < b/2\ , \tag{7.28}$$

and zero for $|x'| > a/2$ and $|y'| > b/2$, where we have assumed y-polarization.

The far-field function is given by (7.18) or (7.20) depending on whether the aperture is in a ground plane or in free space. The integrand of $\tilde{E}_{a_y}$ is separable, so we can write

$$\tilde{E}_{a_y}(k\hat{\mathbf{x}}\cdot\hat{\mathbf{r}}, k\hat{\mathbf{y}}\cdot\hat{\mathbf{r}}) = \tilde{A}(k\hat{\mathbf{x}}\cdot\hat{\mathbf{r}})\tilde{B}(k\hat{\mathbf{y}}\cdot\hat{\mathbf{r}})\ , \tag{7.29}$$

where $\tilde{A}$ and $\tilde{B}$ are the one-dimensional Fourier transforms of the aperture distributions $A(x')$ and $B(y')$, respectively, over the aperture in Fig. 7.10. That is,

$$\tilde{A}(k_x) = \int_{-\infty}^{\infty} A(x')e^{jk_x x'}dx'; \qquad \tilde{B}(k_y) = \int_{-\infty}^{\infty} B(y')e^{jk_y y'}dy'\ . \tag{7.30}$$

Thus, for a rectangular aperture distribution the radiation pattern will be determined mainly by the product of the Fourier transforms of the aperture distributions in the x- and y-directions. In addition there is the incremental element factor of the magnetic dipole or Huygens source, for the PEC and free space apertures, respectively.

7.4.1 E- and H-plane patterns

The E-plane patterns become

$$G_{_E}(\theta) = -2C_k\tilde{A}(0)\tilde{B}(k\sin\theta)\hat{\boldsymbol{\theta}}\ , \tag{7.31}$$

$$G_{_E}(\theta) = -2C_k\cos^2(\theta/2)\tilde{A}(0)\tilde{B}(k\sin\theta)\hat{\boldsymbol{\theta}}\ , \tag{7.32}$$

for the PEC and free space apertures, respectively, and the H-plane patterns become

$$G_{_H}(0) = -2C_k\cos\theta\tilde{A}(k\sin\theta)\tilde{B}(0)\hat{\boldsymbol{\varphi}}\ , \tag{7.33}$$

$$G_{_H}(\theta) = -2C_k\cos^2(\theta/2)\tilde{A}(k\sin\theta)\tilde{B}(0)\hat{\boldsymbol{\varphi}}\ , \tag{7.34}$$

for the PEC and free space apertures, respectively. We see that the patterns of the PEC and free space apertures are equal, except for the effect of the incremental source, i.e., the free space Green's function.

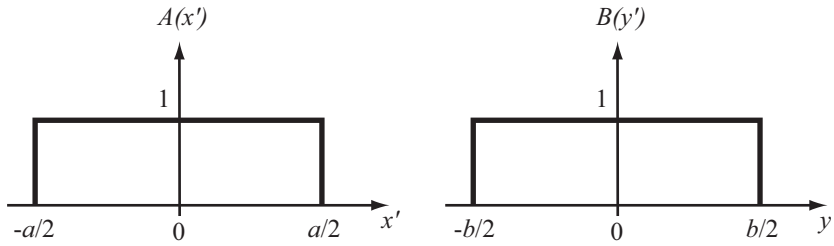


Figure 7.11: Uniform aperture distribution in rectangular aperture.

7.4.2 Directivity and aperture efficiency

For the *rectangular aperture* the aperture *power integral* becomes

$$P = 2\eta P_{\text{rad}} \approx \int_{-a/2}^{a/2} |A(x')|^2 dx' \int_{-b/2}^{b/2} |B(y')|^2 dy' \ . \tag{7.35}$$

By which the directivity becomes separable according to

$$D_{_0} = \frac{4\pi}{\lambda^2} a_{\text{eff}} b_{\text{eff}} \ , \tag{7.36}$$

where
$$a_{\text{eff}} = \left| \int_{-a/2}^{a/2} A(x')dx' \right|^2 / \left\{ \int_{-a/2}^{a/2} |A(x')|^2 dx' \right\} \ , \tag{7.37}$$

$$b_{\text{eff}} = \left| \int_{-b/2}^{b/2} B(y')dy' \right|^2 / \left\{ \int_{-b/2}^{b/2} |B(y')|^2 dy' \right\} \ , \tag{7.38}$$

are effective aperture diameters. In the same way it can be shown that $a_{\text{eff}} \leq a$ and $b_{\text{eff}} \leq b$[3]. Therefore, we may introduce the aperture area $A = ab$ and write

$$D_{_0} = \frac{4\pi}{\lambda^2} e_{\text{ap}} A \ ; \qquad e_{\text{ap}} = e_x e_y \ , \tag{7.39}$$

where $e_{\text{ap}} \leq 1$ is the *aperture efficiency*. This is separable in two terms as shown, where

$$e_x = a_{\text{eff}}/a \ ; \qquad e_y = b_{\text{eff}}/b \ . \tag{7.40}$$

7.4.3 Uniform aperture distribution

The maximum aperture efficiency $e_{\text{ap}} = 1$ and the corresponding maximum directivity appear when the *aperture distribution* is uniform in both planes as shown in Fig. 7.11, i.e.,

$$\begin{aligned} A(x') &= 1 \qquad \text{for } |x'| < a/2 \ , \\ B(y') &= 1 \qquad \text{for } |y'| < b/2 \ , \end{aligned} \tag{7.41}$$

[3] This achieved by Schwarz inequality as used in Section 7.3.3.

Table 7.1: Radiation characteristics of uniformly illuminated rectangular aperture in the principal planes when $b \gg \lambda$.

Topic	**Location**	**Relative level**
First null	$\theta = \arcsin(\lambda/b)$	$-\infty$ dB
First sidelobe	$\theta = \arcsin(1.5\lambda/b)$	-13.2 dB
3 dB half beamwidth	$\theta = \arcsin(0.445\lambda/b)$	-3.0 dB

and zero elsewhere. In this case the Fourier transforms of the aperture field become sinc function, resulting in

$$\tilde{A}(k\sin\theta\cos\varphi) = \int_{-a/2}^{a/2} e^{jkx'(\hat{\mathbf{x}}\cdot\hat{\mathbf{r}})}dx' = a\frac{\sin(ka\sin\theta\cos\varphi/2)}{ka\sin\theta\cos\varphi/2}\ , \tag{7.42}$$

$$\tilde{B}(k\sin\theta\sin\varphi) = \int_{-b/2}^{b/2} e^{jky'(\hat{\mathbf{y}}\cdot\hat{\mathbf{r}})}dy' = b\frac{\sin(kb\sin\theta\sin\varphi/2)}{kb\sin\theta\sin\varphi/2}\ . \tag{7.43}$$

The characteristics of the radiation pattern in E-plane are found by studying (7.31) - (7.32) with $\tilde{B}$ given in (7.43). We see that the location of the nulls in the pattern is determined by $kb\sin\theta/2 = i\pi$ where $i = 1, 2, ...$ The sidelobe maxima are determined by $kb\sin\theta/2 = \pi/2 + i\pi$ where $i = 1, 2, ...$ The results are summarized in Table 7.1. The characteristics are the same in H-plane. The sidelobe levels in the table are only correct when the incremental source factor does not contribute significantly. This condition holds when $b \gg \lambda$ for the E-plane and $a \gg \lambda$ for the H-plane.

Eq. (7.43) represents a universal radiation pattern for all uniformly illuminated apertures with diameters b when plotted against $(b/\lambda)\sin\theta$, which is done in Fig. 7.12[4]. The corresponding radiation pattern when $B(y') = \cos(\pi y'/b)$ for $|y'| < b/2$ is plotted for comparison. The latter *aperture distribution*, which corresponds to the H-plane distribution of an open-ended waveguide, gives broader main beam and lower sidelobes.

The total radiation patterns including incremental source factor in (7.31) - (7.34) are plotted in Fig. 7.13* for $a = b = 8\lambda$. We see that for the PEC aperture the magnetic current distribution in the aperture makes the sidelobes in E-plane higher than in H-plane. Note that we cannot calculate any radiation pattern behind the aperture plane, i.e., for $|\theta| > 90°$, which by definition is a null-field region when using the present PEC equivalent. For a free space aperture the Huygens source makes the patterns equal in E- and H-planes, and the formulas can be used even behind the aperture plane.

Fig. 7.14 shows a contour plot*. We see that the sidelobes are located in the principal planes with very low levels in between them. This is typical for rectangular apertures and a result of the separability of the aperture distribution.

[4] There exist MATLAB code for all figures of which the caption start with *.

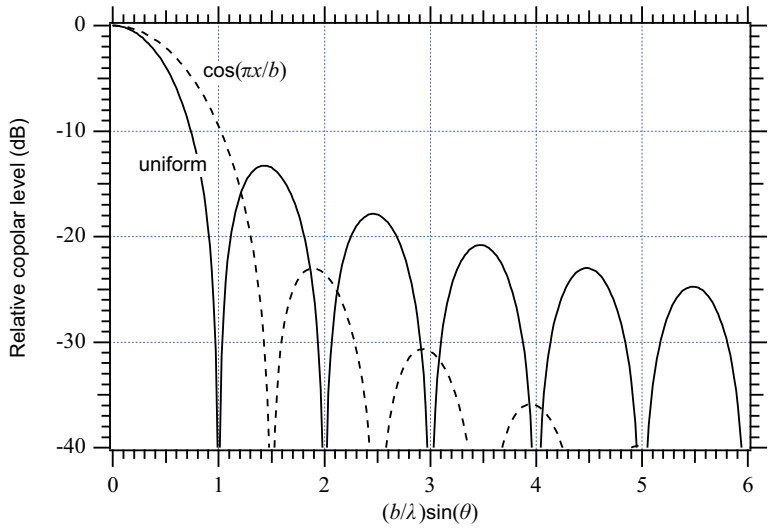


Figure 7.12: *Universal radiation pattern of rectangular aperture versus width for uniform and cosine shaped aperture distributions.

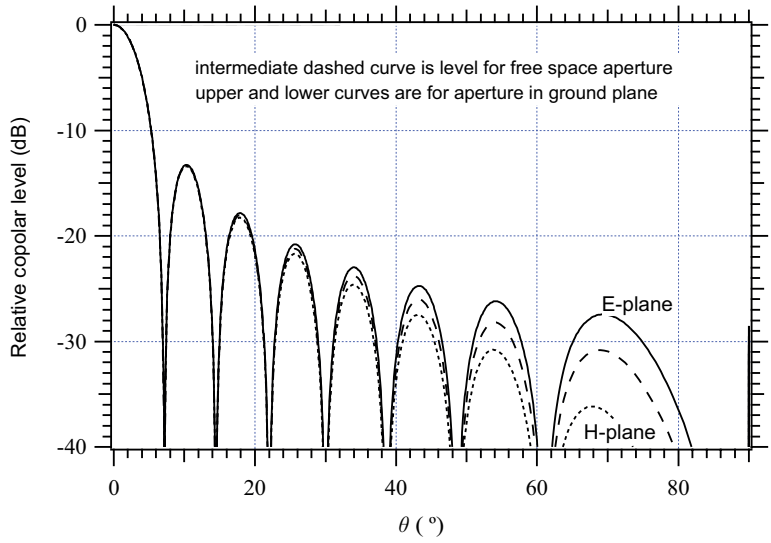


Figure 7.13: *Examples of radiation patterns for uniform field distribution over a square aperture in an infinite ground plane (in E- and H-planes) and in free space (in both planes). The aperture diameter is 8λ.

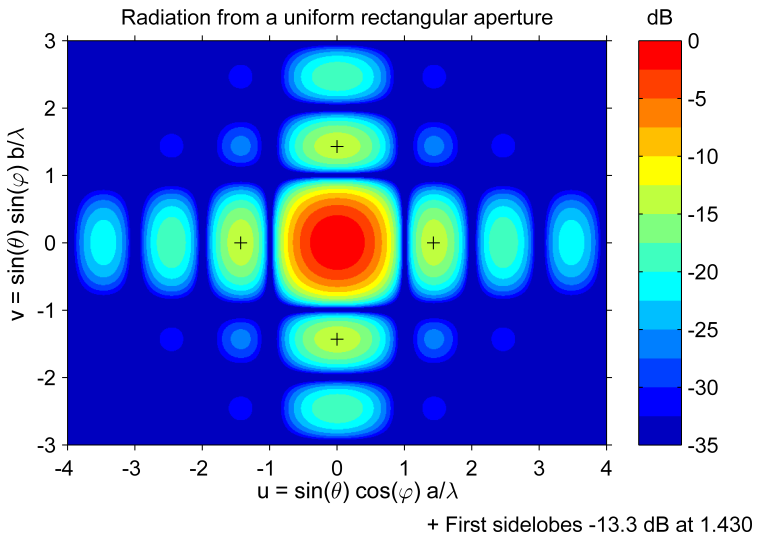


Figure 7.14: *Universal contour plot of the radiation pattern of a uniform field distribution over a rectangular aperture. 5 dB contours down to −35 dB. The aperture diameters are a and b in the xz- and yz-planes, respectively. The maximum level in the center of the main beam is 0 dB.

7.5 Circular aperture with BOR1 excitation

Conical horn antennas and many reflector antennas have circular apertures, as shown in Fig. 7.15. When they are excited for a y-polarized BOR_1 type radiation, the aperture fields are of the form

$$\mathbf{E}_a(\rho',\varphi') = [E_{_E}(\rho')\sin\varphi'\hat{\boldsymbol{\rho}}' + E_{_H}(\rho')\cos\varphi'\hat{\boldsymbol{\varphi}}'] \qquad \text{for } \rho' \le d/2\ , \tag{7.44}$$

where d is the aperture diameter, $\hat{\boldsymbol{\rho}}'$ is the unit vector in the radial ρ'-direction with $\mathbf{r}' = \rho'\hat{\boldsymbol{\rho}}' = \rho'(\cos\varphi'\hat{\mathbf{x}} + \sin\varphi'\hat{\mathbf{y}})$, and $E_{_E}(\rho')$ and $E_{_H}(\rho')$ are the aperture distributions in the E- and H-planes, respectively. By introducing $\hat{\mathbf{x}} = \cos\varphi'\hat{\boldsymbol{\rho}}' - \sin\varphi'\hat{\boldsymbol{\varphi}}'$ and $\hat{\mathbf{y}} = \sin\varphi'\hat{\boldsymbol{\rho}}' + \cos\varphi'\hat{\boldsymbol{\varphi}}'$, this can be written

$$\mathbf{E}_a(\rho',\varphi') = E_{a_y}(\rho',\varphi')\hat{\mathbf{y}} + E_{a_x}(\rho',\varphi')\hat{\mathbf{x}}\ , \tag{7.45}$$

$$\text{where} \qquad E_{a_y}(\rho',\varphi') = E_{\mathrm{co}_{45^\circ}}(\rho') - E_{\mathrm{xp}_{45^\circ}}(\rho')\cos(2\varphi') \tag{7.46}$$

$$\text{and} \qquad E_{a_x}(\rho',\varphi') = E_{\mathrm{xp}_{45^\circ}}(\rho')\sin(2\varphi')\ , \tag{7.47}$$

$$\text{with} \qquad E_{\mathrm{co}_{45^\circ}}(\rho') = \frac{1}{2}[E_{_E}(\rho') + E_{_H}(\rho')] \tag{7.48}$$

$$\text{and} \qquad E_{\mathrm{xp}_{45^\circ}}(\rho') = \frac{1}{2}[E_{_E}(\rho') - E_{_H}(\rho')] \tag{7.49}$$

are the co- and cross-polar aperture fields in the 45°-plane. These BOR_1 *relations* for circular apertures in cylindrical coordinates can be derived in the same way as the corresponding relations in the spherical (θ,φ) coordinate system in equations (2.83)-(2.87) in Section 2.4.2.

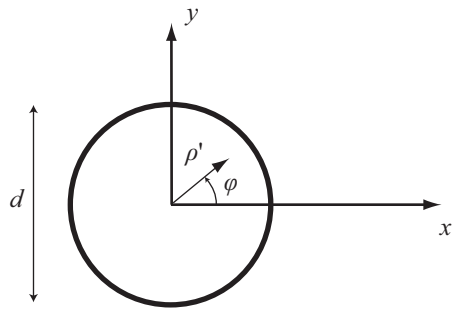


Figure 7.15: Circular plane aperture in xy-plane

7.5.1 Aperture field and far-field function

If the aperture is large, the radiation field is given by (7.18) or (7.20) with

$$\tilde{E}_{a_y}(k\hat{\mathbf{x}}\cdot\hat{\mathbf{r}}, k\hat{\mathbf{y}}\cdot\hat{\mathbf{r}}) = \int_0^{d/2}\int_0^{2\pi} E_{a_y}(\rho',\varphi') e^{jk\rho'\hat{\boldsymbol{\rho}}'\cdot\hat{\mathbf{r}}} d\varphi'\rho' d\rho' \ , \tag{7.50}$$

$$\tilde{E}_{a_x}(k\hat{\mathbf{x}}\cdot\hat{\mathbf{r}}, k\hat{\mathbf{y}}\cdot\hat{\mathbf{r}}) = \int_0^{d/2}\int_0^{2\pi} E_{a_x}(\rho',\varphi') e^{jk\rho'\hat{\boldsymbol{\rho}}'\cdot\hat{\mathbf{r}}} d\varphi'\rho' d\rho' \ , \tag{7.51}$$

obtained from (7.15) - (7.16). The φ'-integrals can be evaluated analytically by introducing the x- and y-components of $\hat{\boldsymbol{\rho}}'$ and $\hat{\mathbf{r}}$, which give

$$\hat{\boldsymbol{\rho}}'\cdot\hat{\mathbf{r}} = \sin\theta(\cos\varphi'\cos\varphi + \sin\varphi'\sin\varphi) = \sin\theta\cos(\varphi-\varphi') \tag{7.52}$$

and using (see Appendix E)

$$\begin{aligned}
\int_0^{2\pi} e^{jk\rho'\sin\theta\cos(\varphi-\varphi')} d\varphi' &= 2\pi J_0(k\rho'\sin\theta) \ , \\
\int_0^{2\pi} \cos(2\varphi') e^{jk\rho'\sin\theta\cos(\varphi-\varphi')} d\varphi' &= -2\pi\cos(2\varphi) J_2(k\rho'\sin\theta) \ , \\
\int_0^{2\pi} \sin(2\varphi') e^{jk\rho'\sin\theta\cos(\varphi'-\varphi)} d\varphi' &= -2\pi\sin(2\varphi) J_2(k\rho'\sin\theta) \ ,
\end{aligned}$$

with J_0 and J_2 the zeroth and second order Bessel functions, respectively. Then, for the free space aperture we finally obtain

$$\tilde{E}_{a_y}(k\hat{\mathbf{x}}\cdot\hat{\mathbf{r}}, k\hat{\mathbf{y}}\cdot\hat{\mathbf{r}}) = \tilde{E}_{a_y}(\theta,\varphi) = \tilde{E}_{\mathrm{co}_{45^\circ}}(\theta) - \tilde{E}_{\mathrm{xp}_{45^\circ}}(\theta)\cos(2\varphi) \ , \tag{7.53}$$

$$\tilde{E}_{a_x}(k\hat{\mathbf{x}}\cdot\hat{\mathbf{r}}, k\hat{\mathbf{y}}\cdot\hat{\mathbf{r}}) = \tilde{E}_{a_x}(\theta,\varphi) = \tilde{E}_{\mathrm{xp}_{45^\circ}}(\theta)\sin(2\varphi) \ , \tag{7.54}$$

where
$$\tilde{E}_{\mathrm{co}_{45^\circ}}(\theta) = 2\pi\int_0^{d/2} E_{\mathrm{co}_{45^\circ}}(\rho') J_0(k\rho'\sin\theta)\rho' d\rho' \ , \tag{7.55}$$

$$\tilde{E}_{\mathrm{xp}_{45^\circ}}(\theta) = -2\pi\int_0^{d/2} E_{\mathrm{xp}_{45^\circ}}(\rho') J_2(k\rho'\sin\theta)\rho' d\rho' \ . \tag{7.56}$$

If we introduce a normalized radius variable $\overline{\rho} = \rho'/d$ in the above two expressions, we get them in the *universal form*

$$\tilde{E}_{\mathrm{co}_{45^\circ}}(\theta) = 2\pi d^2 \int_0^{1/2} \overline{E_{\mathrm{co}_{45^\circ}}}(\overline{\rho}) J_0(k\overline{\rho} d \sin\theta) \overline{\rho} d\overline{\rho} \ , \tag{7.57}$$

$$\tilde{E}_{\mathrm{xp}_{45^\circ}}(\theta) = -2\pi d^2 \int_0^{1/2} \overline{E_{\mathrm{xp}_{45^\circ}}}(\overline{\rho}) J_2(k\overline{\rho} d \sin\theta) \overline{\rho} d\overline{\rho} \ , \tag{7.58}$$

where $\overline{E_{\mathrm{co}_{45^\circ}}}(\overline{\rho})$ and $\overline{E_{\mathrm{xp}_{45^\circ}}}(\overline{\rho})$ are the aperture distributions as a function of the normalized aperture radius. The universal form allows us to calculate general radiation patterns and study them for different types of aperture distributions. These pattern shapes become independent of the aperture diameter d if they are plotted as a function of $kd\sin\theta$. The total far-field functions of the PEC and free space apertures are found by using (7.18) and (7.20), respectively, together with (7.53) - (7.56). We find that the co-polar far-field function of the free space aperture in the 45°-plane is proportional to

$$G_{\mathrm{co}_{45^\circ}}(\theta) = \tilde{E}_{\mathrm{co}_{45^\circ}}(\theta) \cos^2(\theta/2)$$

and the cross-polar to

$$G_{\mathrm{xp}_{45^\circ}}(\theta) = \tilde{E}_{\mathrm{xp}_{45^\circ}}(\theta) \cos^2(\theta/2) \ .$$

7.5.2 Uniform aperture distribution

When the *aperture distribution* is entirely co-polar with constant amplitude and phase we have

$$E_{\mathrm{co}_{45^\circ}} = 1 \qquad \text{for } \rho \le d/2 \tag{7.59}$$

and $E_{\mathrm{xp}_{45^\circ}}(\rho) = 0$. Then, we get

$$\tilde{E}_{\mathrm{co}_{45^\circ}}(\theta) = \int_0^{d/2} J_0(k\rho\sin\theta)\rho d\rho \ . \tag{7.60}$$

If we substitute $t = k\rho\sin\theta$ and use

$$\int_0^T J_0(t) t dt = T J_1(T) \ , \tag{7.61}$$

we get the co-polar radiation pattern

$$\tilde{E}_{\mathrm{co}_{45^\circ}}(\theta) = \frac{2\pi}{(k\sin\theta)^2} \int_0^{kd\sin\theta/2} J_0(t) t dt = 2\pi(d/2)^2 \frac{J_1(kd\sin\theta/2)}{kd\sin\theta/2} \ . \tag{7.62}$$

This has a maximum of $\pi(d/2)^2$ for $\theta = 0$. The directivity becomes

$$D_0 = \frac{4\pi}{\lambda^2}\pi(d/2)^2 = \left(\frac{\pi d}{\lambda}\right)^2 \ . \tag{7.63}$$

The *universal radiation pattern* is plotted in dB in Fig. 7.16 together with that of a circular aperture having a $J_0(k_{01} 2s'/d)$ distribution, where $k_{01} = 2.405$ makes the aperture distribution zero at $\rho' = d/2$ ⋆. The latter aperture distribution will be treated in details in Section 8.8. The location and relative level of the first null, the 3 dB beamwidth, and the first sidelobe, are shown in Table 7.2. This can be compared to Table 7.1 for rectangular apertures. The sidelobes are circular lobes around the main beam as seen in the contour plot in Fig. 7.17⋆, to be compared to Fig. 7.14 for rectangular apertures.

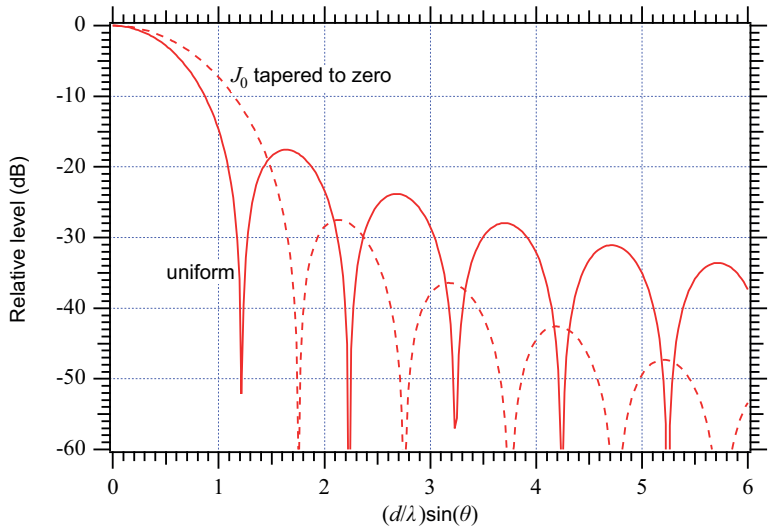


Figure 7.16: *Universal radiation patterns of circular aperture with uniform and tapered aperture distributions. The tapered aperture distribution is the zeroth order Bessel function with a null at the rim.

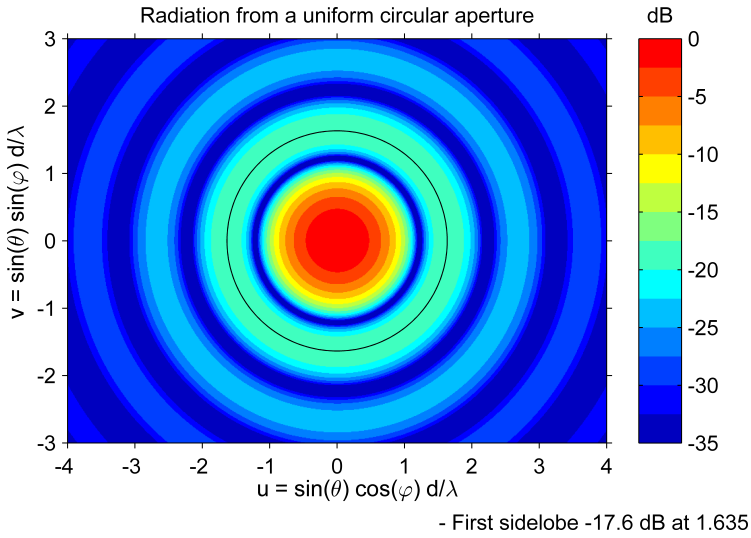


Figure 7.17: *Universal contour radiation pattern of circular aperture with uniform aperture distribution. 5 dB contours down to −35 dB.

Table 7.2: Radiation characteristics of uniformly illuminated circular aperture in the principal planes when $d \gg \lambda$.

Topic	**Location**	**Relative level**
First null	$\theta = \arcsin(1.2\lambda/d)$	$-\infty$ dB
First sidelobe	$\theta = \arcsin(1.65\lambda/d)$	-17.6 dB
3 dB half beamwidth	$\theta = \arcsin(0.51\lambda/d)$	-3.0 dB

7.5.3 Gaussian aperture distribution

The *Gaussian aperture distribution* is of interest because its far-field function has Gaussian shape as well, see Fig. 7.18. We consider a plane free space aperture extending to infinity, with a y-polarized aperture distribution of the form

$$\mathbf{E}_a(\rho) = e^{-(\rho/\rho_a)^2}\hat{\mathbf{y}} = e^{-(x'/\rho_a)^2-(y'/\rho_a)^2}\hat{\mathbf{y}} \ , \tag{7.64}$$

ρ_a is referred to as the 8.7 dB radius or the $1/e$ radius because the field at $\rho = \rho_a$ is $1/e$ or 8.7 dB less relative to the field at $\rho = 0$.

The extent of the aperture is infinite, so we may choose either to analyze it as a rectangular or circular aperture. We choose the former because the integrations are simpler. The aperture integral function becomes separable as in (7.28), so

$$\tilde{E}_{a_y}(k\hat{\mathbf{x}}\cdot\hat{\mathbf{r}}, k\hat{\mathbf{y}}\cdot\hat{\mathbf{r}}) = \tilde{A}(k\hat{\mathbf{x}}\cdot\hat{\mathbf{r}})\tilde{B}(k\hat{\mathbf{y}}\cdot\hat{\mathbf{r}}) \ , \tag{7.65}$$

$$\text{with} \qquad \tilde{A}(k\hat{\mathbf{x}}\cdot\hat{\mathbf{r}}) = \int_{-\infty}^{\infty} e^{-(x'/\rho_a)^2} e^{jkx'\hat{\mathbf{x}}\cdot\hat{\mathbf{r}}} dx' \tag{7.66}$$

and correspondingly for $\tilde{B}(k\hat{\mathbf{y}}\cdot\hat{\mathbf{r}})$. The exponent of the integrand can be simplified as follows

$$\left(\frac{x'}{\rho_a}\right)^2 - jkx(\hat{\mathbf{x}}\cdot\hat{\mathbf{r}}) = \left[\frac{x'}{\rho_a} - \frac{1}{2}jk\rho_a(\hat{\mathbf{x}}\cdot\hat{\mathbf{r}})\right]^2 + \left(\frac{1}{2}k\rho_a\hat{\mathbf{x}}\cdot\hat{\mathbf{r}}\right)^2 \ .$$

Further, we may use

$$\int_{-\infty}^{\infty} e^{-(t-t_0)^2/\rho_a^2} dt = \sqrt{\pi}\rho_a$$

to obtain

$$\begin{aligned} \tilde{A}(k\hat{\mathbf{x}}\cdot\hat{\mathbf{r}}) &= \sqrt{\pi\rho_a^2} e^{-\left(\frac{1}{2}k\rho_a\hat{\mathbf{x}}\cdot\hat{\mathbf{r}}\right)^2} \ , \\ \tilde{B}(k\hat{\mathbf{y}}\cdot\hat{\mathbf{r}}) &= \sqrt{\pi\rho_a^2} e^{-\left(\frac{1}{2}k\rho_a\hat{\mathbf{y}}\cdot\hat{\mathbf{r}}\right)^2} \ . \end{aligned} \tag{7.67}$$

If we introduce

$$(\hat{\mathbf{x}}\cdot\hat{\mathbf{r}})^2 + (\hat{\mathbf{y}}\cdot\hat{\mathbf{r}})^2 = (\sin\theta\cos\varphi)^2 + (\sin\theta\sin\varphi)^2 = \sin^2\theta \ ,$$

the aperture Fourier transform factor becomes

$$\tilde{E}(k\hat{\mathbf{x}}\cdot\hat{\mathbf{r}}, k\hat{\mathbf{y}}\cdot\hat{\mathbf{r}}) = \tilde{E}(\theta) = \pi\rho_a^2 e^{-\sin^2\theta/\sin^2\theta_a} \ ; \qquad \sin\theta_a = 2/(k\rho_a) \ , \tag{7.68}$$

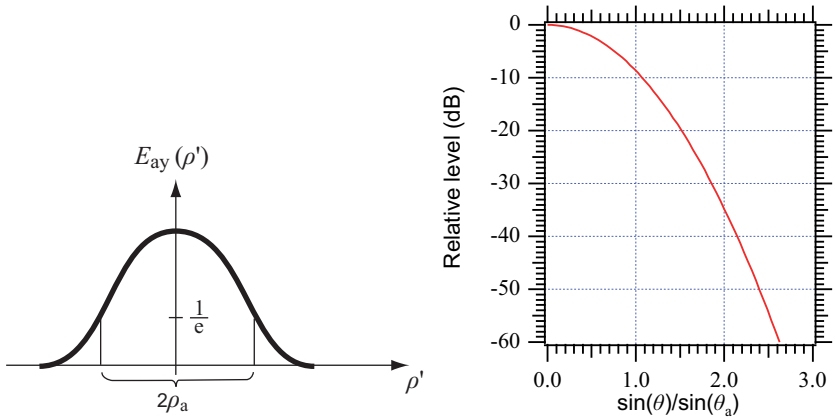


Figure 7.18: *The Gaussian aperture distribution (left) and its radiation pattern (right).

where θ_a is the 8.7 dB half-beamwidth. The total far-field function is found by using this in (7.18) or (7.20) for the PEC or free space apertures, respectively.The power integral over the aperture becomes

$$P = 2\eta P_{\text{rad}} = \iint_{-\infty}^{\infty} |\mathbf{E}_a|^2 dx'dy' = 2\pi \int_0^{\infty} e^{-2(\rho'/\rho_a)^2} \rho' d\rho' = \frac{\pi}{2}\rho_a^2 \ , \tag{7.69}$$

which was obtained by using the known integral

$$\int xe^{-x^2} dx = -\frac{1}{2}e^{-x^2} + C \ .$$

This gives the directivity

$$D_0 = \frac{4\pi|\mathbf{G}(\hat{\mathbf{z}}) \cdot \hat{\mathbf{y}}|^2}{2\eta P_{\text{rad}}} = \frac{4\pi}{\lambda^2} 2\pi\rho_a^2 = (\sqrt{2}k\rho_a)^2 = \frac{8}{\sin^2\theta_a} \ . \tag{7.70}$$

We see that the effective aperture area becomes $A_{\text{eff}} = D_0\lambda^2/4\pi = 2\pi\rho_a^2$. The radius of this effective aperture is equal to $\sqrt{2}$ times the 8.7 dB radius ρ_a. The radiation pattern in dB decreases uniformly with θ according to

$$20\log|G(\theta)/G(0)| = 20\log|E(\theta)/E(0)| = (\sin\theta/\sin\theta_a)^2 \cdot (-8.7\,\text{dB}) \ ,$$

where we have neglected the θ-variation of the incremental element factor. Thus, the Gaussian aperture distribution has a radiation pattern in dB which decreases quadratically with increasing $\sin\theta$. This property makes the Gaussian beam a plausible approximation to most main beams of pencil-beam antennas. The ideal Gaussian radiation pattern has no sidelobes. If the aperture is finite, the Gaussian aperture distribution is truncated, and sidelobes will appear in the radiation pattern.

7.5.4 Tapered aperture distributions

In practice the aperture distribution will be tapered to a finite low value at the aperture rim. Generally, a tapered aperture distribution will reduce both the sidelobes and the aperture

efficiency, compared to that of a uniformly illuminated aperture. Therefore, in antennas where the aperture distribution can be chosen, e.g., in dual-reflector antennas and large arrays, it is often chosen as a compromise between high efficiency and low sidelobes. As an example we consider the truncated Gaussian aperture distribution, which over a circular aperture has the form

$$\mathbf{E}_a(\rho) = \begin{cases} e^{-(\rho/\rho_a)^2}\hat{\mathbf{y}} & \text{for } \rho < d/2 \\ 0 & \text{for } \rho > d/2 \end{cases} . \tag{7.71}$$

The *aperture illumination taper* is defined by $-20\log|\mathbf{E}_a(d/2)/\mathbf{E}_a(0)|$. Fig. 7.19 shows the universal radiation patterns for different aperture illumination tapers, and the aperture efficiency as a function of the aperture taper*. The results have been produced by numerical integration of the radiation integral in (7.57).

7.6 Gaussian beam

The purpose of this section is to study the Gaussian aperture distribution in more details by including a spherical phase-front in the aperture. This will give us the equations used to transform a *Gaussian beam* from the near-field to the far-field.

In order to find the near-field of the Gaussian aperture distribution we need to evaluate the near-field radiation integral in equations (4.37) and (4.39) in Section 4.2.1. These integrals are for the Huygens equivalent of the form (for $kR \gg 1$)

$$\mathbf{E}(\mathbf{r}) = C_k \iint_A [\eta\mathbf{J}_a - (\eta\mathbf{J}_a \cdot \hat{\mathbf{R}})\hat{\mathbf{R}} - (\eta\mathbf{J}_a \times \hat{\mathbf{z}}) \times \hat{\mathbf{R}}] \frac{e^{-jkR}}{R} dx'dy' , \tag{7.72}$$

where $\hat{\mathbf{R}} = \mathbf{R}/R$ with

$$\mathbf{R} = \mathbf{r} - \mathbf{r}' = (x - x')\hat{\mathbf{x}} + (y - y')\hat{\mathbf{y}} + z\hat{\mathbf{z}} , \qquad R = \sqrt{(x - x')^2 + (y - y')^2 + z^2} .$$

Let us consider observation points close to the axis, so that the *paraxial approximation* may be used, i.e.,

$$\hat{\mathbf{R}} \approx \hat{\mathbf{z}} \qquad \text{and} \qquad R = z \tag{7.73}$$

in amplitude factors and

$$R \approx z + \frac{1}{2z}[(x - x')^2 + (y - y')^2] \tag{7.74}$$

in phase terms. Then,

$$\mathbf{E}(\mathbf{r}) \approx -2C_k \frac{1}{z} e^{-jkz} \iint_A \mathbf{E}_a(x', y') e^{-jk\frac{1}{2z}[(x-x')^2+(y-y')^2]} dx'dy' . \tag{7.75}$$

In order to generalize it we assume that the Gaussian aperture field has a spherical wavefront with curvature C_a, i.e., (see Fig. 7.20)

$$\mathbf{E}_a(x, y) = E_0 e^{-\rho'^2/\rho_a^2} e^{-jk\frac{1}{2}C_a\rho'^2}\hat{\mathbf{y}} , \tag{7.76}$$

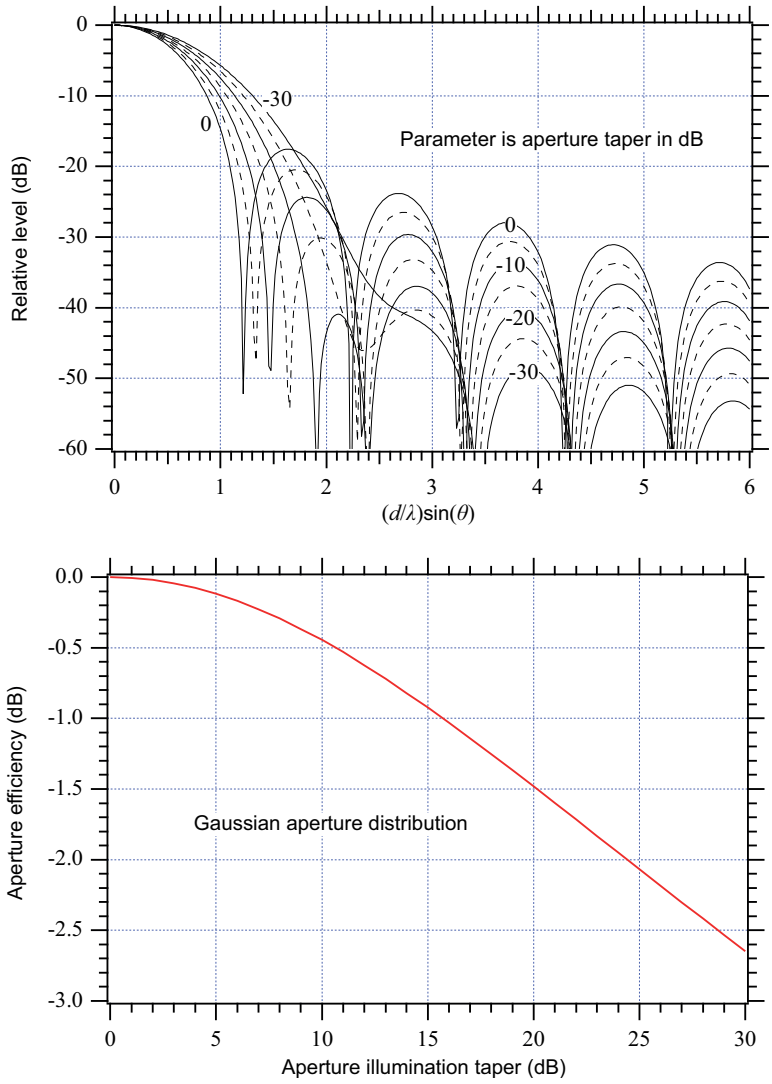


Figure 7.19: *Universal radiation pattern (upper) and aperture efficiency (lower) of truncated Gaussian aperture distribution.

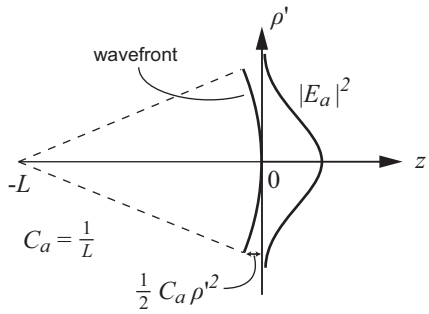


Figure 7.20: Gaussian aperture distribution and its phase center.

where E_0 is a constant and ρ_a is the 8.7 dB radius in the aperture. The quadratic phase term is seen to represent a spherical phase-front from the following expansion of R, which is valid under the paraxial approximation,

$$R = \sqrt{L^2 + \rho'^2} \approx L + [\rho'^2/(2L)] \qquad \text{for } \rho' \ll L \ , \tag{7.77}$$

where $z = -L$ is the center of the sphere, so the wavefront curvature is $C_a = 1/L$. The aperture integral is separable in x' and y', and the former of these factors becomes

$$I_x = \int_{-\infty}^{\infty} e^{-(x'/\rho_a)^2 - jkx'^2/(2L) - jk(x-x')^2/(2z)} dx' \ . \tag{7.78}$$

The y' integral gets a similar form. It is rather laborious to solve (7.78) analytically, yet, it is possible. The result is simple and given in the next subsection. It can for all z be expressed in the same Gaussian form as the original aperture distribution. Therefore, the resulting expression valid for all z is referred to as a Gaussian beam. It is also possible to show that the Gaussian beam is a field solution in free space under the paraxial approximation by directly inserting the results below into Maxwell's equations.

7.6.1 Gaussian near-field

For all z the Gaussian beam can be expressed by

$$\mathbf{E}(\mathbf{r}) = E_0 \frac{\rho_a}{\rho(z)} e^{-\rho^2/[\rho(z)]^2} e^{-jk\frac{1}{2}C(z)\rho^2} e^{j\phi(z)} e^{-jkz} \hat{\mathbf{y}} \ , \tag{7.79}$$

where $\mathbf{r} = \rho\hat{\boldsymbol{\rho}} + z\hat{\mathbf{z}}$ and

$$\rho(z) = \sqrt{(\rho_{\rm dif}(z))^2 + (\rho_{\rm GO}(z))^2} \tag{7.80}$$

is the *8.7 dB radius* of the beam at $r = z$,

$$C(z) = \frac{1}{(L+z)} \left[1 + \frac{L}{z} \left(\frac{\rho_{\rm dif}(z)}{\rho(z)} \right)^2 \right] \tag{7.81}$$

is the *wavefront curvature* at $r = z$, and

$$\phi(z) = \frac{\pi}{2} - \arctan(\rho_{\rm GO}(z)/\rho_{\rm dif}(z)) \tag{7.82}$$

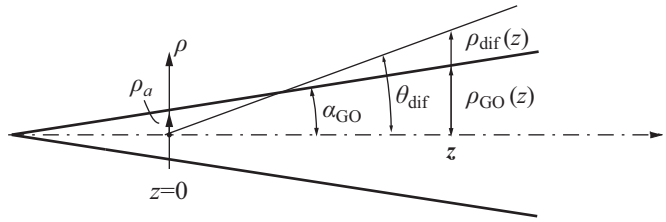


Figure 7.21: Cross section of GO cone and diffraction cone.

is called the *phase slippage* and represents the phase in the center of the beam at $r = z$ relative to the phase at $z = 0$, when the propagation phase $-kz$ has been removed. In these equations

$$\rho_{\rm dif}(z) = 2z/(k\rho_a) \tag{7.83}$$

is the *diffraction cone radius*, and

$$\rho_{\rm GO}(z) = (z + L)\rho_a/L = (1 + C_a z)\rho_a \tag{7.84}$$

is the *Geometrical Optics* (GO) *cone radius*. Both these expressions are explained in the next paragraph.

$\rho_{\rm dif}$ defines a conical surface with its vertex in the aperture at $z = 0$. The cross section is shown in Fig. 7.21. The cone angle is inversely proportional to the wavelength divided by the aperture radius ρ_a, in the same way as the width of the far-field radiation patterns based on *aperture diffraction*, see Section 1.3.2. Examples are the radiation patterns of aperture distributions with constant phase, shown in Fig. 7.12, 7.16 and 7.19. Therefore, we refer to $\rho_{\rm dif}(z)$ as the *diffraction cone radius.* Similarly, $\rho_{\rm GO}(z)$ defines a cone with its vertex in the point $z = -L$ and with the cone angle proportional to ρ_a/L. This corresponds to a GO continuation of the aperture field[5]. Therefore, we refer to $\rho_{\rm GO}(z)$ as the *GO cone radius.* We will discuss these two contributions to the beam radius in more details in Section 7.6.4 and Section 7.6.5.

The Gaussian beam formulas have been implemented in a MATLAB code which can be used to produce numerical versions of the curves in Fig. 7.22.

7.6.2 Phase center of Gaussian beam

It is possible to find the phase center of the Gaussian beam at any z location by using the wavefront curvature $C(z)$. The phase center is, according to Section 2.3.7, the center of curvature of the wavefront. The radius of curvature is given by $R_{\rm pc}(z) = 1/C(z)$, so therefore the near-field *phase center* location for the wavefront at a distance z from the aperture becomes

$$z_{\rm pc} = z - R_{\rm pc}(z) = z - \frac{1}{C(z)} \ . \tag{7.85}$$

Using the above equations we get

$$z_{\rm pc} = -L(\rho_{\rm GO}(z))^2 / \left[(\rho(z))^2 + \frac{L}{z}(\rho_{\rm dif}(z))^2\right] \approx -L(\rho_{\rm GO}(z))^2/(\rho(z))^2 \ . \tag{7.86}$$

[5] See also Section 1.3.2.

The latter simplest formula for $z_{\rm pc}$ is actually quite accurate and very useful. It tells us that the phase centre is at the aperture $z = 0$ when the diffraction cone radius $\rho_{\rm dif}(z)$ dominates over the GO cone radius $\rho_{\rm GO}(z)$, and at $z = -L$ when $\rho_{\rm GO}(z) \gg \rho_{\rm dif}(z)$. Both of the two cone radii $\rho_{\rm dif}(z)$ and $\rho_{\rm GO}(z)$ vary linearly with distance. The former diffraction cone is zero in the aperture and increases linearly with z, whereas the latter GO cone is zero at $z = -L$ which is the centre of curvature of the phase front at $z = 0$. It varies also linearly with increasing z, but the slope depends on the sign of L, i.e., on whether the phase-front is concave (L negative) or convex (L positive) in the aperture plane. See also Fig. 7.22 and the following subsections.

7.6.3 Gaussian far-field

In order to find the expression for the far-field we introduce the 8.7 dB beamwidth

$$\theta_{_0} = \lim_{z\to\infty}\left(\frac{\rho(z)}{z}\right) = \sqrt{\theta_{\rm dif}^2 + \theta_{\rm GO}^2}\ , \tag{7.87}$$

$$\text{where} \qquad \theta_{\rm dif} = 2/(k\rho_a) \qquad \text{and} \qquad \theta_{\rm GO} = \rho_a/L \tag{7.88}$$

(both in radians) are the beamwidth contributions due to aperture diffraction and GO transmission, respectively. These choices of $\theta_{_0}$, $\theta_{\rm dif}$ and $\theta_{\rm GO}$ in radians will make the result more valid and usable for broad beams than the alternative choice based on

$$\tan\theta_{_0} = \lim_{z\to\infty}\left(\frac{\rho(z)}{z}\right) = \sqrt{\tan^2\theta_{\rm dif} + \tan^2\theta_{\rm GO}}\ .$$

By foregoing choices the formula for $\theta_{\rm GO}$ corresponds to measuring ρ_a as the length along a curved wavefront and not radially along a straight line. The *phase center* is located at

$$z_{\rm pc} = -L(\theta_{\rm GO}/\theta_{_0})^2\ .$$

Furthermore, when z approaches infinity we may use

$$r = \sqrt{z^2+\rho^2} \approx z + \frac{1}{2z}\rho^2 \approx z + \frac{1}{2}z\theta^2$$

$$\text{and} \qquad C(z) = 1/(z - z_{\rm pc}) \approx (1/z) + (z_{\rm pc}/z^2)\ ,$$

to obtain

$$z + \frac{1}{2}C(z)\rho^2 \approx z + \frac{1}{2}C(z)z^2\theta^2 \approx z + \frac{1}{2}z\theta^2 + \frac{1}{2}z_{\rm pc}\theta^2 \approx r + \frac{1}{2}z_{\rm pc}\theta^2\ ,$$

which is valid for small θs. This gives the far-field

$$\mathbf{E}(\mathbf{r}) = \frac{1}{r}e^{-jkr}\mathbf{G}(\hat{\mathbf{r}}) \tag{7.89}$$

and the far-field function

$$\mathbf{G}(\hat{\mathbf{r}}) = E_{_0}\frac{\rho_a}{\theta_{_0}}e^{-(\theta/\theta_{_0})^2}e^{-j\frac{1}{2}kz_{\rm pc}\theta^2}e^{j\phi_\infty}\hat{\mathbf{y}}\ , \tag{7.90}$$

$$\text{with} \qquad \phi_\infty = -\arctan(\theta_{\rm dif}/\theta_{\rm GO})\ . \tag{7.91}$$

The $\mathbf{G}(\hat{\mathbf{r}})$ in (7.90) and that resulting from (7.68) should be equal for $L \to \infty$. Although they are not the same, the differences are small for small θ and are explained by the paraxial approximation used in the derivation of (7.90). Note that $\theta_0 = 2/(k\rho_a)$ when $L \to \infty$. It is clear that the beamwidth has increased compared to that in (7.68), due to the spherical phase in the aperture, with the factor

$$q = \frac{(\theta_0)_{\text{with phase error}}}{(\theta_0)_{\text{without phase error}}} = \sqrt{\frac{\theta_{\text{dif}}^2 + \theta_{\text{GO}}^2}{\theta_{\text{dif}}^2}} = \sqrt{1 + (\Delta\phi)^2} \ , \qquad \begin{array}{c} \text{where} \\ \Delta\phi = (k\rho_a^2)/(2L) \end{array} \tag{7.92}$$

is the phase error at $\rho = \rho_a$ in the aperture. The corresponding directivity reduction is $1/q^2$. This can also be found by comparing $\mathbf{G}(\hat{\mathbf{r}})$ resulting from (7.68) with $\mathbf{G}(\hat{\mathbf{r}})$ in (7.90) as follows. The power integrals in the aperture are equal if $E_0 = 1$ and $\rho_0 = \rho_a$. Therefore, the ratio between the directivities (i.e., the *phase efficiency*) is

$$\begin{aligned} \frac{(D_0)_{\text{with phase errors}}}{(D_0)_{\text{without errors}}} &= \left(\frac{\rho_a/\theta_0}{2C_k\pi\rho_a^2}\right)^2 = \left(\frac{2}{k\rho_a\theta_0}\right)^2 = \frac{\theta_{\text{dif}}^2}{\theta_0^2} \\ &= \frac{1}{q^2} \approx 1 - \Delta\phi^2 \ , \end{aligned} \tag{7.93}$$

where the expansion $(1+x)^{-1} \approx 1 - x$ for small x was used to get the final expression. This equation is actually quite general. Phase errors in the aperture cause a gain reduction and a broadening of the beam. For arbitrary but small phase errors, it can be found that the gain reduction is given by the same formula $1 - \Delta\phi^2$ with $\Delta\phi^2$ the mean square phase error around the mean phase error, both averaged over the aperture, with the amplitude of the aperture distribution used as weighting factor in the averaging process, see Eq. (9.91).

7.6.4 Aperture diffraction by constant phase aperture

When there is a *plane phase-front* in the aperture, i.e., $L = \infty$, the 8.7 dB beam radius varies from ρ_a in the aperture to

$$\rho(z) = \sqrt{(\rho_{\text{dif}}(z))^2 + \rho_a^2} \ ; \qquad \rho_{\text{dif}}(z) = \frac{2z}{k\rho_a} \ , \tag{7.94}$$

at a distance z from the aperture, see Fig. 7.22*. This result is qualitatively valid for all aperture distributions with constant phase. The radius of the beam is approximately equal to the aperture width ρ_a in the near-field region $0 < z < \pi\rho_a^2/\lambda \approx D^2/\lambda$ where $D \approx 2\rho_a$. This is a result of GO transmission of a wave with plane phase-front. In the far-field region the beamwidth starts to increase and approaches a linear increase with z, in a way that the angular extent becomes constant and is given approximately by $\theta_0 = \lambda/(\pi\rho_a) \approx \lambda/D$. This effect is commonly referred to as *aperture diffraction*.

7.6.5 GO radiation from aperture with strongly curved wavefront

When the phase-front of the aperture field curves so much that the phase error at $\rho = \rho_a$ is much larger than 60°, i.e.,

$$\phi(\rho_a) - \phi(0) = k\rho_a^2/(2L) \gg 1 \ , \qquad \text{i.e.,} \qquad L \ll \frac{1}{2}k\rho_a^2 \ , \tag{7.95}$$

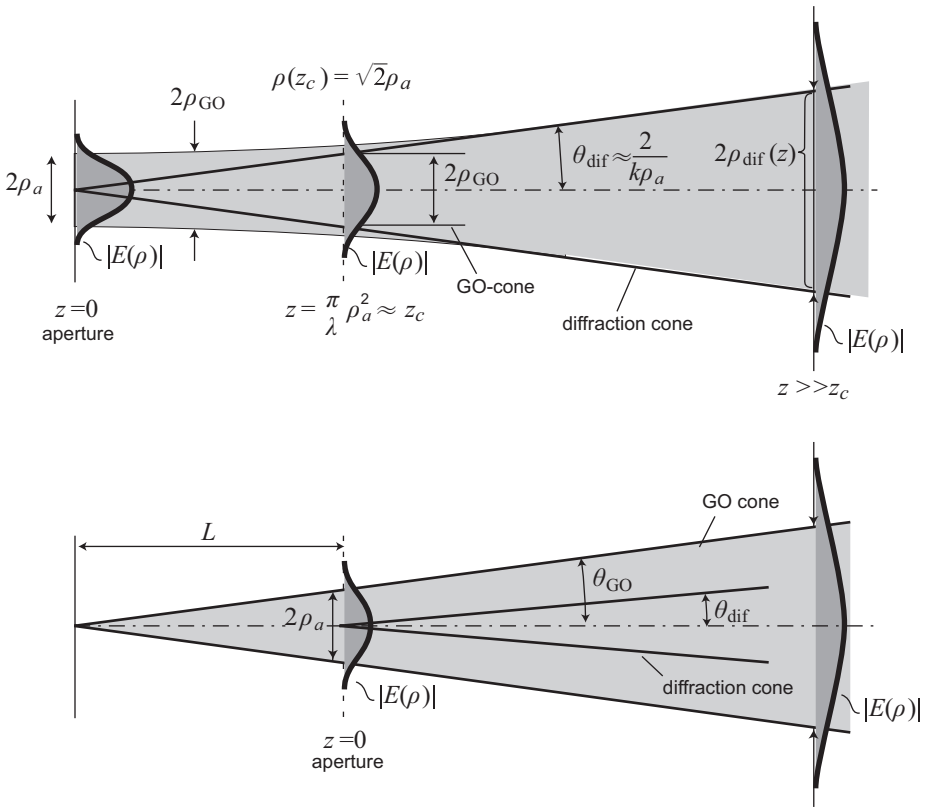


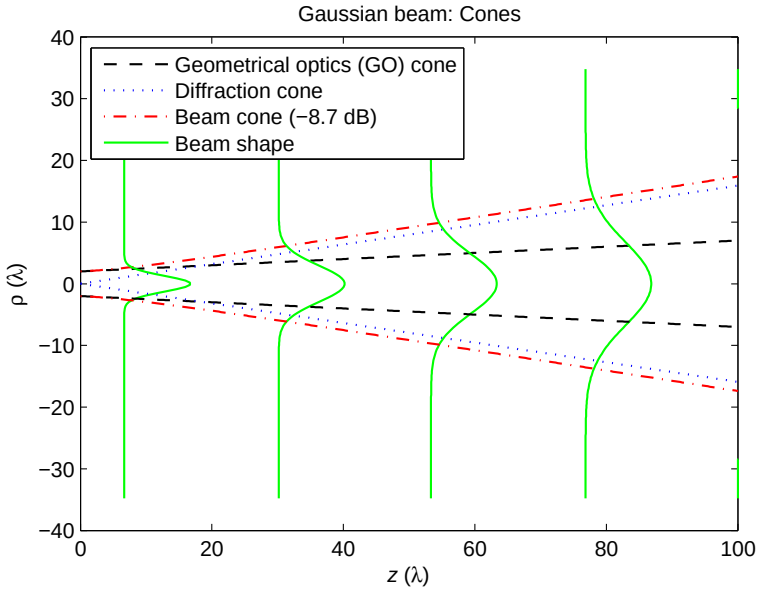
Figure 7.22: ⋆ 8.7 dB beamwidth of Gaussian beam as a function of distance from the aperture. (a) Constant phase aperture. (b) Spherical phase aperture.

we see that the beam radius changes with increasing z according to

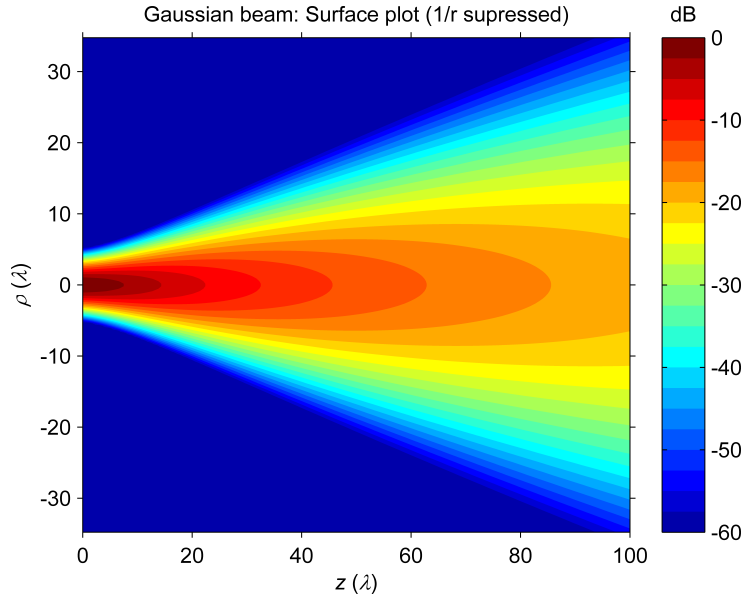
$$\rho(z) \approx \rho_{GO}(z) = (z + L)\rho_a/L \ . \tag{7.96}$$

Thus, the beam diverges linearly with increasing z already from the aperture plane. The angular beamwidth is constant and equal to $\theta_0 = \rho_a/L$. The phase errors in the aperture are spherical. They may, e.g., originate from the far-field of an antenna located at $z = -L$, as the far-field always has a spherical phase-front. If the phase errors are sufficiently large the beam simply continues to propagate through the aperture in the same way as it did to come to the aperture.

The equations in Subsection 7.6.3 are approximately valid also for *broad beams* resulting from strongly curved phase-fronts in the aperture, provided ρ_a is measured along the curved phase-front and not in the aperture plane. The GO transmission dominates in this case for which the beam has the same angular extent at all radial distances from the spherical aperture, when measured from the center of curvature of the wavefronts. See Fig. 7.22 and Fig. 7.23⋆.



(a)



(b)

Figure 7.23: *Illustration of Gaussian beam: (a) Plots of GO cone, diffraction cone and 8.7 dB beamwidth. (b) Contour plot of the same case.

7.6.6 Alternative expressions for Gaussian beam parameters

The Gaussian near-field beam as expressed by the equations (7.79)-(7.84) are easily interpretable by introducing the diffraction and GO cones. However, it is also possible to express the beam parameters conveniently by means of the beam waist $\rho_{\min}$ and the beam waist location $z_{\min}$, see Fig. 7.24. Then, the equations become simpler. The *beam waist* is the narrowest 8.7 dB radius $\rho(z_{\min}) = \rho_{\min}$ along the beam. This can be found by requiring that $\partial\rho(z)/\partial z = 0$ with $\rho(z)$ given by Eq. (7.80). This has the solution

$$z_{\min} = -L\theta_{\text{GO}}^2/(\theta_{\text{dif}}^2 + \theta_{\text{GO}}^2) \ , \tag{7.97}$$

where θ_{dif} and θ_{GO} are the same as in (7.88). This location is for positive L behind the aperture, where the field solutions are not valid. Still, $z_{\min}$ defines where the beam waist would have been located if the Gaussian beam had been continued for negative z. When L is negative, $z_{\min}$ is in front of the aperture. After $z_{\min}$ has been calculated, the beam waist $\rho_{\min} = \rho(z_{\min})$ is easily evaluated from (7.80).

Now we may introduce the confocal distance

$$z_c = \pi\rho_{\min}^2/\lambda \ , \tag{7.98}$$

which corresponds to the far-field distance D^2/λ of an aperture with diameter $D = \sqrt{\pi}\rho_{\min}$. Then, (7.80)-(7.82) simplify to

$$\begin{aligned} \rho(z) &= \rho_{\min}\sqrt{1 + \left(\frac{z - z_{\min}}{z_c}\right)^2} \ , \\ C(z) &= \frac{(z - z_{\min})}{(z - z_{\min})^2 + z_c^2} \ , \\ \phi(z) &= \arctan\left(\frac{z - z_{\min}}{z_c}\right) + \arctan\left(\frac{z_{\min}}{z_c}\right) \ . \end{aligned} \tag{7.99}$$

Therefore, if we first calculate the location and radius of the beam waist from the aperture field parameters ρ_a and L, we can use the simple formulas in (7.99) to calculate the beam radius and center of curvature at any location z.

7.7 Complementary comments

Beams are often used to expand the field radiated by an aperture. A number of different type of beams have been introduced in the literature, the most popular of which are listed below

- *Gaussian Beams* (*GB*)
- Higher order Gaussian-Laguerre Beams (GLB) beams
- Higher order Gaussian-Hermite Beams (GHB) beams
- Bessel Beams (BB)
- Gaussian-Ray Basis Functions (GRBF)
- *Complex Source Points* (*CSP*)
- Complex Conical Beams (CCB)

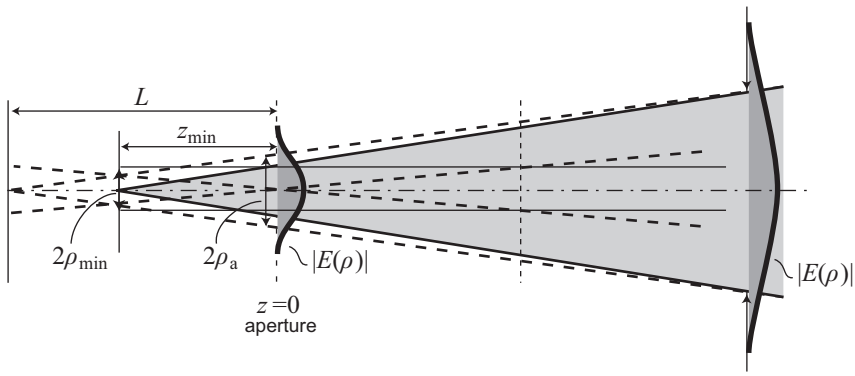


Figure 7.24: Continuation of Gaussian beam for negative z and illustration of location of beam waist. The dotted lines show the GO and diffraction cones when drawn from the original aperture at $z = 0$, and the solid lines show the GO and diffraction cones when drawn from the beam waist at $z = z_{\min}$.

GB [3]-[4], i.e., those treated in the previous section of this chapter, are the most commonly used, especially for representing fields radiated from conical corrugated horns and from reflectors. GB satisfies Maxwell's equations only in the paraxial region, namely close to the axis of propagation, but they may sometimes in practice work well to describe fields outside this region as well. GLB or GHB [5]-[6] deal with expansions around a preferred axis of propagation with the higher order terms representing the off-axis variations. They have the advantage of constituting an orthogonal set, and they are often used as basis for a mode matching technique. GLB and GHB are more suitable for describing rectangular and circular apertures, respectively. Their descriptive capability relaxes but not removes the restriction to the paraxial region, and this produces slow convergence when describing far out-of axis lobes.

In Fourier Optics, the field is often represented by a discrete spectrum of Bessel Beams (BB) [7]. BB are quite similar to GHB, but in spite of the latter they do not diffract as they propagate, thus exhibiting a non-physical behavior in the far-field region. However, they are useful in near-field and laser beam representations.

In contrast to the previously described GB plus higher order beams, which can provide a *global* description from an aperture, Gabor-type (or phase-space) expansions [8] and [11] provide a *local* far field (aperture field spectrum) description. The field is expanded using a lattice of beams that emerge from a set of points in the aperture plane and propagate from each point in a lattice of directions. The beam amplitudes are determined by the local radiation properties (local spectrum) of the aperture near the lattice spectral points. This beam representation can be viewed as a "local" expansion of the aperture field spectrum. It is seen that for off-axis observations, the localized nature of the spectral elements implies more rapid convergence than with global expansions [12].

In the Gabor-based Gaussian beam expansion, the basis set is complete. This poses a restriction on the choice of the spatial and spectral resolutions. The Gabor-frame scheme in [8]-[9] (termed there "windowed Fourier transform frame") relaxes this restriction by using overcomplete sets of GB's, thus enabling the user to choose the spatial and spectral resolutions so that they best fit the local properties of the source distribution. Furthermore, the flexi-

bility gained by using the over-complete frame expansion allows an efficient representation of ultrawide band (UWB) fields [10]-[11].

GRBF [13]-[14] are used to locally match the far field as in the Gabor expansion; however, they have more degrees of freedom. This is obtained by introducing an empirical extra parameter to a conventional Gaussian beam, in order to manually control beam width at a given distance from the source.

An alternative to the global expansion is represented by *Complex Source Points* (*CSP*) expansion [15]-[17]. A spherical wave of type e^{-jkR}/R with $R = \sqrt{(\mathbf{r}-\mathbf{r}')\cdot(\mathbf{r}-\mathbf{r}'}$ respects the wave equation also when the center of phase is formally displaced in a complex space $\mathbf{r}' = \mathbf{r}_0 - j\mathbf{b}$. The waveform obtained by this complex displacement is called CSP. After the analytical continuation, the complex distance R is a multivalued function which respect to the radiation conditions if $\Re(R) \geq 0$. R can be equal to zero, and therefore the field is singular on a disc of radius $|\mathbf{b}|$ orthogonal to $\mathbf{b}$ and centered in $\mathbf{r}_0$. In the half space $(\mathbf{r}-\mathbf{r}_0)\cdot\mathbf{b} > 0$ the CSP has a behavior quite similar to a GB propagating along $\mathbf{b}$ with waist $|\mathbf{b}|$. CSP differs from a GB only far from axis and close to the disk of singularity (where the GB is regular). We emphasize that the GB is an approximate paraxial solution of the wave equation, while the CPS is an exact solution of the wave equation.

CSP have been used for expanding space domain Green's functions for layered dielectric media [18] and [21]. To this end, the spectral Green's function is represented in terms of exponentials by using GPOF method, thus leading to the CSP expansion by using the *Sommerfeld integral*. A similar technique is used to regularize the Kernel of 3D integral equation [20].

A complex source representation of an arbitrary source field inside a spherical region is based on the complex Huygens' principle. This was first proposed in [21], where an exact CPS expansion for arbitrary fields is presented in which the beams are launched from a single point in space and their coefficients are determined from the radiated field on a sphere in real space. In [20] the CPS expansion is obtained first by constructing equivalent currents on a spherical surface using the spherical wave expansion of the original sources, thereafter by extending the surface in complex space to obtain a continuous equivalent distribution of CPS, and finally by properly sampling the continuous distribution to obtain a discrete expansion.

The Conical beams introduced in [21] present a selective concentration of energy around the surface of a cone, and their representation rigorously satisfies the wave equation. An aperture field representation of these beams is generated in a natural way starting from the spectral-domain radiation integral, by expanding the electric field spectrum in a Fourier series, and by approximating the obtained Fourier series coefficients by a sum of complex exponentials using the generalized pencil-of-function method. This transforms the radiation integral to a simpler form which can be evaluated analytically.

7.8 Exercises to Chapter 7

1. **Quadratic apertures in PEC and free space**: Consider two quadratic radiating apertures, one located in a PEC ground plane, and the other located in an aperture plane in free space.

Assume that the aperture fields in both cases are uniformly distributed over the aperture, i.e.,

$$\mathbf{E}_a(x,y) = E_0\hat{\mathbf{y}} \ , \qquad \mathbf{H}_a(x,y) = -\frac{1}{\eta}E_0\hat{\mathbf{x}} \ ,$$

for $|x| < a/2$ and $|y| < a/2$ where η is the free space wave impedance and $\hat{\mathbf{y}}$ is the direction of the polarization. Write down the expressions for the far-field functions for both apertures. Find the expressions for the co- and cross-polar far-field functions for the two cases for a desired linear y-polarization. Assume that $a = 3\lambda$.What are the levels in dB for the relative cross-polar field at $\theta = 10^\circ$ and $\theta = 20^\circ$ in the $\varphi = 45^\circ$ plane for each of the two cases? Find the levels in dBi for the relative sidelobe envelopes of the co-polar field at $\theta = 90^\circ$ in the $\varphi = 0^\circ$ and 90° planes for both apertures.

2. **Cross-polarization for PEC apertures**: Consider the PEC aperture in exercise 7.1. Derive an expression for the location of the first sidelobe in the $\varphi = 45^\circ$ plane, and find an expression for the relative level of it in dB. How does the level vary with increasing a.

3. **Cross-polarization due to aperture field**: Consider the same two apertures as in exercise 7.1, with the same desired polarization. Assume that the aperture field is now

$$\mathbf{E}_a = E_{\rm co}\hat{\mathbf{y}} + E_{\rm xp}\hat{\mathbf{x}} \ , \qquad \mathbf{H}_a = -\frac{1}{\eta}E_{\rm co}\hat{\mathbf{x}} + \frac{1}{\eta}E_{\rm xp}\hat{\mathbf{y}} \ .$$

Assume that $E_{\rm xp} = 0.005E_{\rm co}$ and $E_{\rm co} = const$, and derive the expressions for the co- and cross-polar radiation patterns of both apertures. What is the relative cross-polar level of the radiation pattern for $\theta = 0^\circ$? What are the cross-polar levels at $\theta = 10^\circ$ and $\theta = 20^\circ$ in the $\varphi = 45^\circ$ plane for the two cases, relative to the co-polar level in the same direction? What is the axial ratio of the resulting elliptical polarization at the same angles?

4. **Circular apertures in PEC and free space**: Repeat exercise 7.1 for a circular aperture of diameter $d = 3\lambda$ and uniform co-polar aperture distribution.

5. **Location of beam waist**: Derive the formula for the beam waist $W_{\rm min}$ and its location $z_{\rm min}$ relative to an aperture with a given wavefront with center of curvature at $z = -L$.

6. **Focussing of Gaussian beam**: Consider a Gaussian plane wave (propagating in negative z-direction) with 8.7 dB width $D/2$ which hits a parabolic reflector with focal length F. After reflection from the paraboloid (located at $z = 0$) the Gaussian beam gets a wavefront curvature $1/F$. Where would you expect the beam waist to be located? Compare this with the actual expression for the location of the beam waist. How does the beam waist vary with F/D for a large D? Calculate the beam waist when $F = D = 10\lambda$ and when $F = D = 20\lambda$.

7. See the exercises about arrays of open waveguides and small aperture in Chapter 10.

7.9 References

[1] W. Rudin, *Principles of mathematical analysis*, McGraw-Hill, 1964.

[2] P.F. Goldsmith, *Quasioptical Systems: Gaussian Beam Quasioptical Propagation and Applications*, IEEE Press/Chapman & Hall Publishers, 1998.

[3] G.A. Deschamps, "The Gaussian beam as a bundle of complex rays", *Electronics Letters*, Vol. 7, No. 23, pp. 684–685, 1971.

[4] N.J. McEwan and P.F. Goldsmith, "Gaussian beam techniques for illuminating reflector antennas", *Antennas and Propagation, IEEE Transactions on* , Vol. 37, No. 3, pp. 297–304, March 1989.

[5] W.A. Imbriale and D.J. Hoppe, "Recent trends in the analysis of quasioptical systems", presented at the Millenium Conf. on Antennas Propag., Davos, Switzerland, 2000.

[6] S. Withington, J.A. Murphy, and K.G. Isaak, "Representation of mirrors in beam waveguides as inclined phase-transforming surfaces", *Infrared Physics & Technology*, Vol. 36, No. 3, pp. 723–734, April 1995.

[7] H.E. Hernàndez-Figueroa, M. Zamboni-Rached, and E. Recami, Eds., *Localized Waves*. Hoboken, NJ: Wiley-Interscience, 2008, Ch. 6.

[8] B.Z. Steinberg, E. Heyman, and L.B. Felsen, "Phase space methods for radiation from large apertures", *Radio Science*, Vol. 26, pp. 219–227, 1991.

[9] A. Shlivinski, E. Heyman, A. Boag, and C. Letrou, "A phase-space beam summation formulation for ultrawideband radiation: A multiband scheme", *IEEE Transactions on Antennas and Propagation*, Vol. 52, No. 8, pp. 2042–2056, August 2005.

[10] A. Shlivinski, E. Heyman, and A. Boag, "A phase-space beam summation formulation for ultrawideband radiation—Part II: A multiband scheme", *IEEE Transactions on Antennas and Propagation*, Vol. 53, No. 3, pp. 948–957, March 2005.

[11] A. Shlivinski, E. Heyman, and A. Boag, "A pulsed beam summation formulation for short pulse radiation based on windowed radon transform (WRT) frames", *IEEE Transactions on Antennas and Propagation*, Vol. 53, No. 9, pp. 3030–3048, September 2005.

[12] J.M. Arnold, "Phase-space localization and discrete representation of wave fields", *Journal of The Optical Society of America A*, Vol. 12, No. 1, pp. 111–123, January 1995.

[13] H.-T. Chou, P.H. Pathak, and R.J. Burkholder, "Application of Gaussian-ray basis functions for the rapid analysis of electromagnetic radiation from reflector antennas", *Microwaves, Antennas and Propagation, IEE Proceedings*, Vol. 150, pp. 177–183, 2003.

[14] H.-T. Chou and P.H. Pathak, "Uniform asymptotic solution for electromagnetic reflection and diffraction of an arbitrary Gaussian beam by a smooth surface with an edge", *Radio Science*, Vol. 32, No. 4, pp. 1319–1336, July/August 1997.

[15] L.B. Felsen, "Complex-source-point solutions of the field equations and their relation to the propagation and scattering of Gaussian beams", *Proc. Symp. Math*, Vol. 18, pp. 39–56, 1975.

[16] Y. Dezhong, "Complex source representation of time harmonic radiation from a plane aperture", *IEEE Transactions on Antennas and Propagation*, Vol. 43, No. 7, pp. 720–723, July 1995.

[17] A. Polemi, G. Carluccio, M. Albani, A. Toccafondi, and S. Maci, "Incremental theory of diffraction for complex point source illumination", *Radio Science*, Vol. 42, 2007.

[18] Y.L. Chow, J.J. Yang, D.G. Fang, and G.E. Howard, "A closed form spatial Green function for the microstrip substrate", *Microwave Theory and Techniques, IEEE Transactions on*, Vol. 39, No. 3, pp. 588–592, March 1991.

[19] J. He, T. Yu, N. Geng, and L. Carin, "Method of moments analysis of electromagnetic scattering from a general three-dimensional dielectric target embedded in a multilayered medium", *Radio Science*, Vol. 35, No. 2, pp. 305–313, 2000.

[20] F. Vipiana, A. Polemi, S. Maci, and G. Vecchi, "A mesh-adapted closed-form regular kernel for 3D singular integral equations", *IEEE Transactions on Antennas and Propagation*, Vol. 56, No. 6, pp. 1687–1698, June 2008.

[21] Norris, A. N., and T. B. Hansen (1997), "Exact complex source representations of time-harmonic radiation", *Wave Motion*, 25, pp. 127–141.

[22] E. Martini, S. Maci, "A closed-form conversion from spherical-wave- to complex-point-source-expansion", *Radio Science*, Vol. 46, RS0E22, doi:10.1029/2011RS004665, 2011.

[23] S. Skokic, M. Casaletti, S Maci, S Sorensen, "Complex Conical Beams for Aperture Field Representations", *IEEE Transactions on Antennas and Propagation*, ISSN: 0018-926X, No. 58, February 2011.

Chapter 8

Horn antennas

Horn antennas are often used at microwave frequencies. Their sidelobes and directivity can be easily and accurately predicted. They often have low sidelobes and can have relative bandwidths up to 1.8. Furthermore, they are normally easy to manufacture in small or moderate sizes. However, if the frequency is low or the required directivity is large, they become large and heavy compared to, e.g., reflector antennas. Horn antennas are commonly used as standard gain horns and as feeds for reflector antennas. They find also application as elements in array antennas. Some examples of horn antennas are shown in Fig. 8.1.

The horn antenna is easily fed from a rectangular, quadratic, or circular waveguide. It is most often designed to have a straight flare from the waveguide opening to the aperture, through which the modal field in the waveguide is transformed to the corresponding field in the aperture. In this chapter we first explain the different methods used to analyze horn antennas. Then, we study horns which are flared only in E-plane or H-plane, referred to as E-plane and H-plane sector horns, respectively. Thereafter, we treat the pyramidal horn, which is flared in both planes, and the conical horn with a circular aperture. We also describe corrugated pyramidal and conical horns. In all cases we show how the aperture field can be constructed from the appropriate waveguide mode.

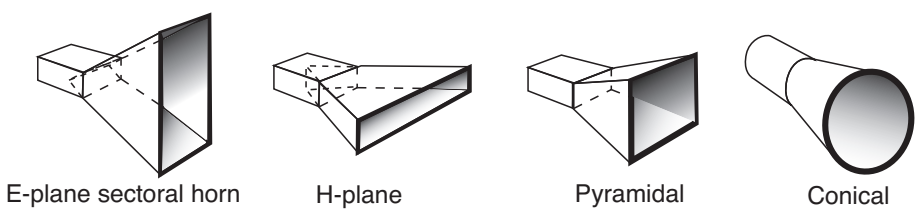


Figure 8.1: Examples of horn antennas.

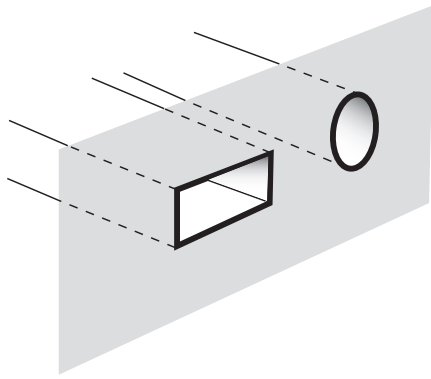


Figure 8.2: Open waveguides.

8.1 Calculation methods

Here we describe the different analytical and numerical methods that are used to model horn antennas. The first four approaches in Subsections 8.1.1 - 8.1.4 are based on approximating the aperture field by using the basic propagating mode inside the horn, and thereafter using the PEC or Huygens equivalent to calculate the far-field function. The approach in Subsection 8.1.5 is based on a numerical calculation of the aperture field using the mode matching technique, and thereafter using the PEC or free space equivalent. This approach can be refined with a Method of Moments (MoM) solution for the outer structure. Finally, in Subsection 8.1.6 we describe how the Method of Moments can be used to calculate the physical currents along the axial cross-section of the complete horn antenna when it is excited from the input waveguide. It is also possible to design horn antennas by using other solution methods for modeling the complete horn such as the *Finite Element Method* (FEM) or the *Finite Difference Time Domain* technique (FDTD).

8.1.1 Cylindrical waveguide plane aperture approach

The simplest waveguide-fed antenna is just an open-ended waveguide (Fig. 8.2). This may be analyzed by assuming that the aperture field is given by the fields of the excited single mode propagating in the waveguide. For waveguides with rectangular and circular cross-sections these modal fields can be found in textbooks on field theory. When the E-and H-fields in the aperture have been found, we can use the PEC or free space equivalents. The Huygens equivalent cannot be used, because the aperture impedance will be very different from the free space impedance due to the small aperture width. The free space equivalent will not be accurate either, because the currents induced on the outer waveguide walls contribute significantly to the radiation, and these currents are neglected in the free space equivalent. The *PEC equivalent* is accurately and easily applied if the aperture resides in a large ground plane.

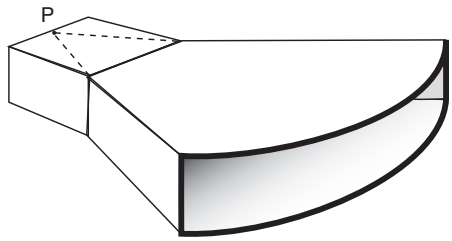


Figure 8.3: Sector horn antenna with circular cylindrical aperture surface.

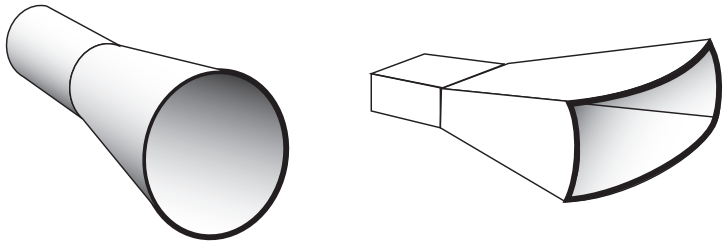


Figure 8.4: The conical horn antenna (left) and the spherical sector horn antenna (right).

8.1.2 Radial cylindrical waveguide approach

There exist analytical modal solutions for the fields in sector waveguides. They are expressed in terms of *cylindrical waves* propagating in the radial direction and can be found in textbooks on field theory. The fields of the basic mode can be used to approximate the aperture field of sector horn antennas (Fig. 8.3). The approximation is acceptable if the two broad walls of the sector horns have a curved circular rim at the aperture which center of curvature P is in the apex of the sector. The free space equivalent should be used.

8.1.3 Conical and spherical sector waveguide approach

There exist also analytical solutions for the modal fields in conical and spherical waveguides. The conical waveguide has a rotationally symmetric wall which coincides with a θ-cone in a coordinate system which is located with its z-axis along the symmetry axis and its origin in the apex of the cone formed by the walls. The spherical waveguide has four walls which coincide with two θ-cones and two φ-planes in a coordinate system which is located with the xy-plane cutting through the center of the horn between the two conical walls. The modes of both structures can be expressed in terms of *spherical wave* functions, and they can be used to approximate the aperture field of both conical horn antennas and spherical sector horn antennas (Fig. 8.4). The resulting aperture field and the Huygens equivalent give quite accurate results, in particular when a spherical aperture surface is used.

8.1.4 Flared cylindrical waveguide approach

The horn geometries in Subsections 8.1.2 and 8.1.3 can also be analyzed by approximating the aperture field distribution over the cylindrical or spherical aperture surface, in a way that it becomes equal to that of a corresponding cylindrical waveguide over its flat aperture surface. This waveguide must have the same rectangular or circular cross-section as the rim of the actual aperture. This approximation is quite good for horns with small flare angles. Once the approximate aperture field has been found, the Huygens equivalent can be used. The advantage of the flared cylindrical waveguide approach is that the analytic form of the aperture field is much simpler than the spherical wave forms otherwise needed. In the latter case we may even need to use the wave functions of non-integer order, which are normally not included in mathematical software libraries for special functions.

The cylindrical or spherical aperture surface can even be approximated by a plane surface if the aperture field is projected on it by using the paraxial approximation. In this way the spherical phase-front becomes a quadratic phase variation over the plane aperture. The advantage of this approach is that it is possible to solve the radiation integral analytically. This approach is quite accurate for small flare angles, but the resulting analytic expressions for the far-field function are so complex that they hardly present any improvement compared to a numerical solution of the original radiation integral. We will still use this approach in Sections 8.2 to 8.8 in combination with numerical integration because it makes it possible to obtain and plot universal radiation patterns which can easily be scaled to different horn dimensions.

8.1.5 Mode matching approach

It is nowadays very common to analyze flared horns with rectangular and circular cross-sections by using mode matching. By this approach the geometry along the whole length of the horn is divided in cascaded sections of cylindrical waveguides of circular or rectangular cross-sections. The cross-section of each section is determined by the actual cross-sectional shape of the horn, and the length of each section must be small enough to model the actual profile with sufficient accuracy (Fig. 8.5). The field within each rectangular or circular section is expressed as a finite series of rectangular or circular cylindrical waveguide modes with initially unknown coefficients. The coefficients are determined by matching the fields at each interface. This is done successively from the aperture to the feed waveguide where a given basic waveguide mode excites the horn. The radiating horn aperture can be analyzed by using a free space equivalent for each mode in the waveguide section closest to the aperture.

The mode matching approach is known to be extremely accurate both for the return loss and the radiation pattern. If it is combined with an accurate Method of Moments solution of the exterior horn structure by using the PEC equivalent, the radiation pattern will be very accurate also in the backward direction.

The whole rim of the horn needs to be located in a plane coinciding with the virtual wall of the last cascaded section. The mode matching approach is well suited for analyzing *corrugated conical* or *pyramidal horns*. The calculations need to be carefully checked with respect to convergence by increasing the number of cascaded sections and the number of modes in each section. In particular, many modes are needed if the flare angle is large.

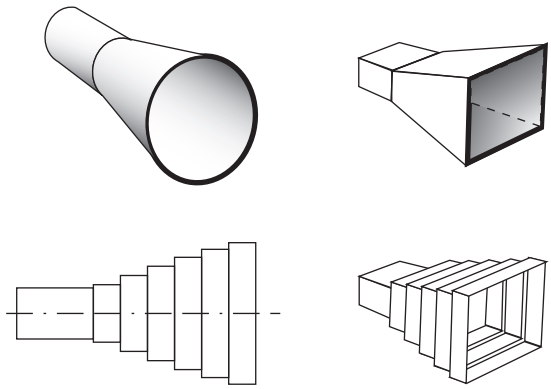


Figure 8.5: Modeling a conical (left) and pyramidal (right) horn by cascaded circular and rectangular cylindrical waveguide sections, respectively.

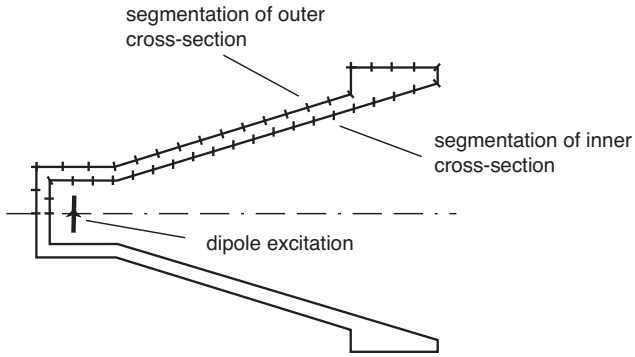


Figure 8.6: Segmentation of the outer and inner cross-section of a rotationally symmetric horn antenna for solution by the Method of Moments.

8.1.6 Method of Moments approach

A complete numerical model of a horn can also be obtained by exciting it with an infinitesimal dipole inside the waveguide and calculating the induced currents on the whole surrounding structure by the *Method of Moments* (MoM), see Fig. 8.6. This is feasible for rotationally symmetric horn antennas with BOR_0 or BOR_1 excitations, as MoM software for *Bodies of Revolution* (BOR) is available (Fig. 8.7). The computation time is long and the storage requirement is large, so the approach is most suitable for small horns. In this approach there are problems with unphysical field resonances at specific frequencies, which cause unreliable results. These problems are most severe for large horn antennas. However, it is not difficult to filter out the invalid results from the usable ones.

Rectangular horn antennas can also be modelled by a complete MoM approach, using software for three-dimensional structures (3D). This approach is even more time-consuming than the BOR MoM approach. Nevertheless, it has been used and plausible results have been reported.

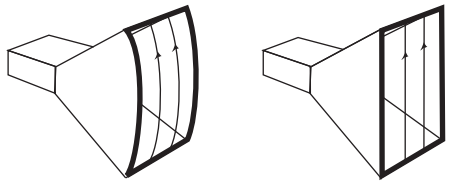


Figure 8.7: The E-plane sector horn antenna with cylindrical aperture (left) and with plane aperture (right). The direction of the E-field lines is marked.

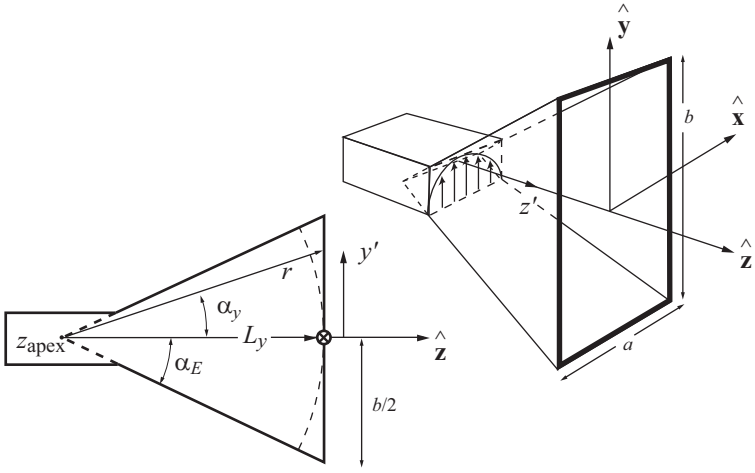


Figure 8.8: Geometry of E-plane sector horn.

8.2 E-plane sector horn

The E-plane sector horn is shown in Figs. 8.7 and 8.8. It is excited by a rectangular waveguide with a TE_{10} mode. The width in H-plane is the same as the waveguide width a, and the aperture width in E-plane is b with flare angle $\alpha_{_E}$. The rim of the horn may be straight so that the aperture surface becomes plane. Alternatively, the rim of the two broad walls may be curved to follow the cylindrical wavefront at the opening of the horn as shown to the left in Fig. 8.7.

8.2.1 Flared cylindrical waveguide approach

In order to find the radiation field we must assume a known aperture field. We may approximate it by using the flared cylindrical waveguide approach described in Subsection 8.1.4, i.e., as a TE_{10} rectangular waveguide mode with a cylindrical phase-front. This is done as follows. Inside a rectangular waveguide the TE_{10} mode has the transverse field distribution

$$\mathbf{E}_w(x', y') = E_{_0} \cos(\pi x'/a)\hat{\mathbf{y}} \qquad \text{for } |x'| < a/2 \ , \ |y'| < b/2 \ . \tag{8.1}$$

In the sector waveguide formed by the horn, the wavefront will be cylindrical rather than plane. Then, we may assume the same field distribution over the cylindrical wavefront, with

y' replaced by the path length $l' = \alpha_y L_y$ for $\alpha_y \leq \alpha_{_E}$, where $\tan\alpha_{_E} = b/(2L_y)$ with L_y the radius of curvature of the cylindrical wavefront in the aperture, i.e., the flare length of the horn. Furthermore, the field in the sector waveguide will have to decay according to $1/\sqrt{R_y}$ away from the apex in order to conserve the power, where $R_y = \sqrt{y'y' + (z'+L_y)(z'+L_y)}$, and $z' = -L_y$ is the location of the apex. Thus, the field in the sector waveguide can be approximated by

$$\mathbf{E}(x', \alpha_y, z') = E_0 \cos(\pi x'/a)[e^{-jkR_y}/\sqrt{R_y}]\hat{\boldsymbol{\alpha}}_y \qquad \text{for } |x'| < a/2 \ , \ |\alpha_y| < \alpha_{_E} \ , \tag{8.2}$$

with

$$\hat{\boldsymbol{\alpha}}_y = \cos\alpha_y \hat{\mathbf{y}} - \sin\alpha_y \hat{\mathbf{z}} \ , \qquad \alpha_y = \arctan(y'/L_y) \ .$$

The field in the cylindrical aperture defined by $R_y = L_y$ becomes

$$\mathbf{E}_a(x', \alpha) = E_0' \cos(\pi x'/a)\hat{\boldsymbol{\alpha}}_y \qquad \text{for } |x'| < a/2 \ , \ |\alpha_y| < \alpha_{_E} \ , \tag{8.3}$$

with $E_0' = E_0 e^{-jkL_y}/\sqrt{L_y}$. The cylindrical aperture surface is described by $R_y = L_y$ or equivalently by the coordinate vector

$$\begin{aligned} \mathbf{r}_a &= x'\hat{\mathbf{x}} + y'\hat{\mathbf{y}} + (\sqrt{L_y^2 - (y')^2} - L_y)\hat{\mathbf{z}} \\ &= x'\hat{\mathbf{x}} + L_y \sin\alpha_y \hat{\mathbf{y}} + L_y(\cos\alpha_y - 1)\hat{\mathbf{z}} \ , \end{aligned}$$

when the origin of the coordinate system is taken to be in the center of the aperture. The surface normal is seen to be $\hat{\mathbf{n}}_a = \sin\alpha_y \hat{\mathbf{y}} + \cos\alpha_y \hat{\mathbf{z}}$. We can now use the Huygens free space equivalent as described in Section 7.2.1. By using (7.6) and (7.7) with (8.3), the far-field function becomes

$$\mathbf{G}(\mathbf{r}) = -\int_{-\alpha_{_E}}^{\alpha_{_E}} \int_{-a/2}^{a/2} E_0' \cos(\pi x'/a)\mathbf{G}_{_H}(\hat{\boldsymbol{\alpha}}_y, \hat{\mathbf{n}}_a, \hat{\mathbf{r}}) e^{jk\mathbf{r}_a \cdot \hat{\mathbf{r}}} dx' L_y d\alpha_y \ , \tag{8.4}$$

where

$$\mathbf{G}_{_H}(\hat{\boldsymbol{\alpha}}_y, \hat{\mathbf{n}}_a, \hat{\mathbf{r}}) = C_k[\hat{\boldsymbol{\alpha}}_y - (\hat{\boldsymbol{\alpha}}_y \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} - (\hat{\boldsymbol{\alpha}}_y \times \hat{\mathbf{n}}_a) \times \hat{\mathbf{r}}] \tag{8.5}$$

is the far-field function of a unit Huygens source. This integral can be evaluated accurately by numerical integration.

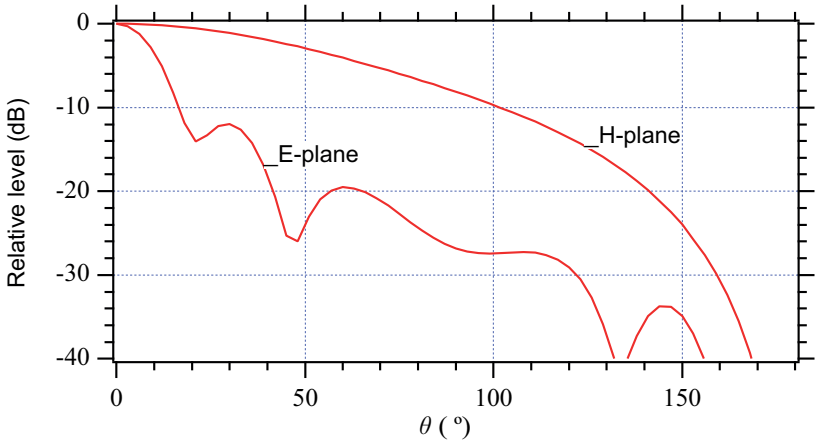
8.2.2 Paraxial approximation for plane aperture field

Sometimes we would want to perform the integration over the plane projection of the cylindrical aperture, as explained before in Section 8.1.4. The field in this plane aperture is obtained by using (8.2) with $z' = 0$ and $R_y = \sqrt{y'y' + L_y L_y}$. We can simplify the expression by assuming that $\alpha \ll 1$ and using the *paraxial approximation* for which

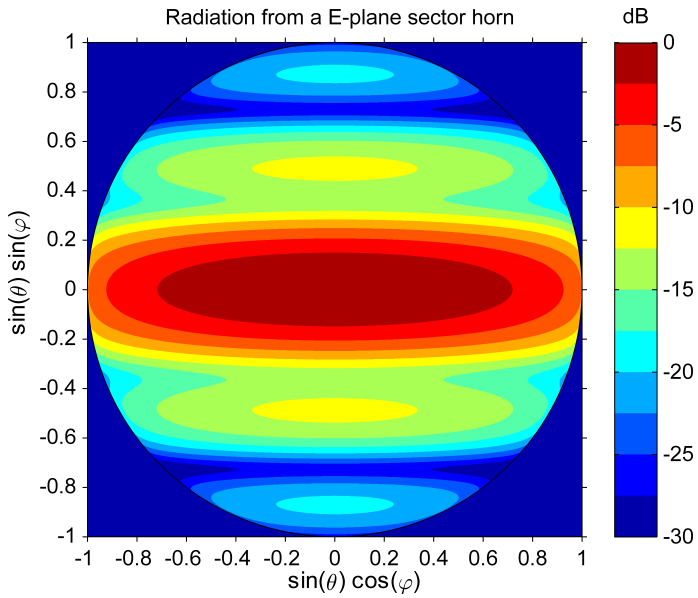
$$R_y = \sqrt{y'^2 + L_y^2} \approx L_y \qquad \text{and} \qquad \hat{\mathbf{n}}_a = \hat{\mathbf{z}} \tag{8.6}$$

in amplitude expressions, and

$$R_y = \sqrt{L_y^2 + y'^2} \approx L_y + y'^2/(2L_y) \tag{8.7}$$



(a)



(b)

Figure 8.9: *Radiation patterns of an E-plane sector horn for $a = 0.5\lambda$, $b = 2.75\lambda$ and $L_y = 6\lambda$. (a) E- and H-plane patterns. (b) Contour plot.

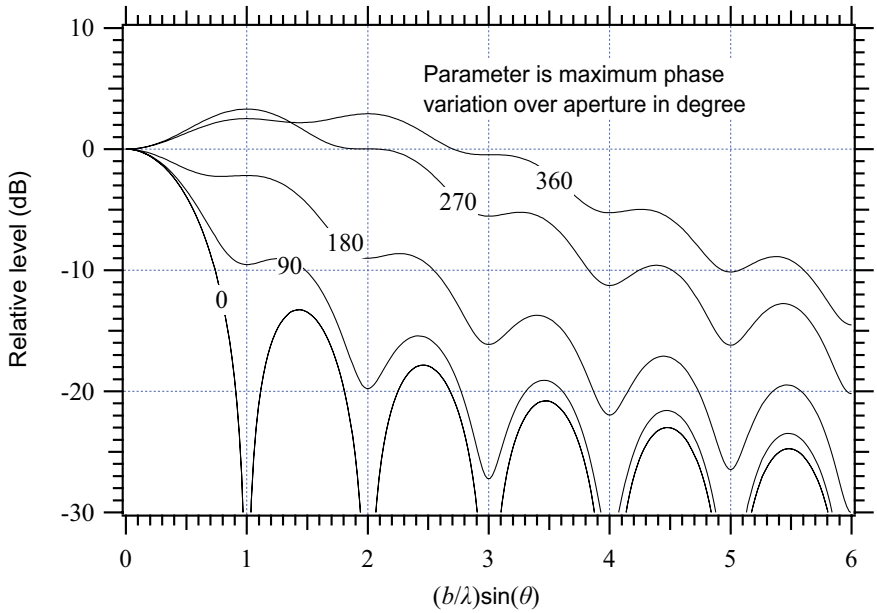


Figure 8.10: *Universal E-plane patterns for E-plane sector horn.

in phase expressions. This gives

$$\mathbf{E}(x',y') \approx E_0' \cos(\pi x'/a) e^{-jky'^2/(2L_y)} \hat{\mathbf{y}} \qquad \text{for } |x'| < a/2\ ,\ |y'| < b/2\ , \tag{8.8}$$

$$\text{with} \qquad E_0' = E_0 e^{-jkL_y}/\sqrt{L_y}\ .$$

The latter is a constant which we may choose arbitrarily, e.g., $E_0' = 1$.

We can now use the theory presented in Section 7.3.2, for the plane free space apertures, and equation (7.21) to get the far-field function

$$\mathbf{G}(\theta,\varphi) = -2C_k \cos^2(\theta/2)[\sin\varphi\hat{\boldsymbol{\theta}} + \cos\varphi\hat{\boldsymbol{\varphi}}]\tilde{A}(k\sin\theta\cos\varphi)\tilde{B}(k\sin\theta\sin\varphi)\ , \tag{8.9}$$

where

$$\tilde{A}(k_x) = \int_{-a/2}^{a/2} \cos(\pi x'/a) e^{jk_x x'} dx' = -\frac{2\pi a\cos(k_x a/2)}{(k_x a)^2 - \pi^2} \tag{8.10}$$

$$\tilde{B}(k_y) = \int_{-b/2}^{b/2} e^{-jky'^2/(2L_y)} e^{jk_y y'} dy' \tag{8.11}$$

and $k_x = k\sin\theta\cos\varphi$, $k_y = k\sin\theta\sin\varphi$. We can transform (8.11) to a universal integral by introducing $u = 2y'/b$ and the maximum phase variation over the aperture

$$\Phi_{\max} = k(b/2)^2/(2L_y) = kb^2/(8L_y)\ .$$

Then,

$$\tilde{B}(k_y) = F(k_y, b, \Phi_{\max}) = \frac{b}{2}\int_{-1}^{1} e^{-ju^2\Phi_{\max}} e^{jk_y bu/2} du\ . \tag{8.12}$$

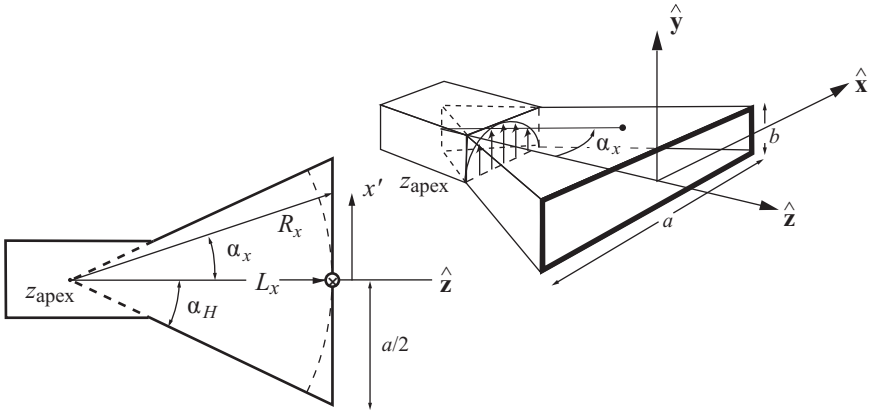


Figure 8.11: Geometry of H-plane sector horn.

The analytic solution of (8.10) can be found in integral handbooks. The integral in (8.12) is most conveniently evaluated numerically. An analytic solution is also available, but it is so complicated that it cannot provide any extra physical insight. Therefore, the numerical evaluation of (8.12) is chosen herein.

8.2.3 Radiation patterns

The radiation patterns in the E- and H-planes for an example of an E-plane sector horn is shown in Fig. 8.9. The first sidelobes of the E-plane have grown into the main beam to form undesirable shoulders. The shoulders will reduce to sidelobes about 13.2 dB less the main beam level when L_y increases. This is due to the uniform aperture illumination in E-plane. The *universal E-plane pattern* represented by $\tilde{B}(k_y)$ is plotted in Fig. 8.10 as a function of $b\sin\theta/\lambda$ for different choices of the maximum *phase error* $\Phi_{\max}$[1]. The actual E-plane pattern is obtained by multiplying the values in Fig. 8.9 with the Huygens source pattern $\cos^2(\theta/2)$. The E-plane sector horn is rarely used due to the high shoulders and sidelobes in E-plane.

8.3 H-plane sector horn

The H-plane sector horn is shown in Fig. 8.11. This is also fed by a TE_{10} rectangular waveguide mode. The H-plane cross-section is flared with an angle α_H, and the E-plane width is equal to the waveguide height b.

8.3.1 Flared cylindrical waveguide approach

Following the flared cylindrical waveguide approach and using the same arguments as in Subsection 8.2.1, we now get an approximate cylindrical wave in the sector guide of the form

[1] There exist MATLAB code for all figures of which the caption start with ⋆.

$$\mathbf{E}(\alpha_x, y', z') = E_0 \cos(\pi\alpha_x/(2\alpha_H))[e^{-jkR_x}/\sqrt{R_x}]\hat{\mathbf{y}} \tag{8.13}$$

for $|\alpha_x| \le \alpha_H$, $|y'| < b/2$, with,

$$\alpha_x = \arctan(x'/L_x) \qquad \text{and} \qquad R_x = \sqrt{(x')^2 + (z' + L_x)^2} \ .$$

The aperture surface is described by

$$\begin{aligned}\mathbf{r}_a &= x'\hat{\mathbf{x}} + y'\hat{\mathbf{y}} + \left(\sqrt{L_x^2 + x'^2} - L_x\right)\hat{\mathbf{z}} \\ &= L_x \sin\alpha_x \hat{\mathbf{x}} + y'\hat{\mathbf{y}} + L_x(\cos\alpha_x - 1)\hat{\mathbf{z}} \ ,\end{aligned}$$

with the surface normal vector $\hat{\mathbf{n}}_a = \sin\alpha_x\hat{\mathbf{x}} + \cos\alpha_x\hat{\mathbf{z}}$. The far-field function is found in the same way as in Subsection 8.2.1 to be

$$\mathbf{G}(\hat{\mathbf{r}}) = -\int_{-\alpha_H}^{\alpha_H}\int_{-b/2}^{b/2} E_0' \cos\left(\frac{\pi\alpha_x}{2\alpha_H}\right)\mathbf{G}_H(\hat{\mathbf{y}}_a, \hat{\mathbf{n}}_a, \hat{\mathbf{r}})e^{jk\mathbf{r}_a\cdot\hat{\mathbf{r}}}dy' L_x d\alpha_x \ , \tag{8.14}$$

with

$$\mathbf{G}_H(\hat{\mathbf{y}}, \hat{\mathbf{n}}_a, \hat{\mathbf{r}}) = C_k[\hat{\mathbf{y}} - (\hat{\mathbf{y}}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - (\hat{\mathbf{y}}\times\hat{\mathbf{n}}_a)\hat{\mathbf{r}}] \tag{8.15}$$

being the radiation field of the unit y-polarized Huygens source. We can accurately evaluate (8.14) with (8.15) by numerical integration, but we will as well proceed with the paraxial approximation in order to get the radiation integral in a universal form.

8.3.2 Paraxial approximation for plane aperture field

Using the *paraxial approximation* we get the following aperture field, in the same way as in Subsection 8.2.2,

$$\mathbf{E}_a(x', y') = E_0' \cos(\pi x'/a)e^{-jkx'^2/(2L_x)}\hat{\mathbf{y}} \tag{8.16}$$

for $|x'| < a/2$ and $|y'| < b/2$. The far-field function becomes

$$\mathbf{G}(\theta, \varphi) = -2C_k \cos^2(\theta/2)[\sin\varphi\hat{\boldsymbol{\theta}} + \cos\varphi\hat{\boldsymbol{\varphi}}]\tilde{A}(k\sin\theta\cos\varphi)\tilde{B}(k\sin\theta\sin\varphi) \ , \tag{8.17}$$

$$\tilde{A}(k_x) = \int_{-a/2}^{a/2} \cos(\pi x'/a)e^{-jkx'^2/(2L_x)}e^{jk_x x'}dx' \ , \tag{8.18}$$

$$\tilde{B}(k_y) = \int_{-b/2}^{b/2} e^{jk_y y'}dy' = b\frac{\sin(k_y b/2)}{k_y b/2} \ . \tag{8.19}$$

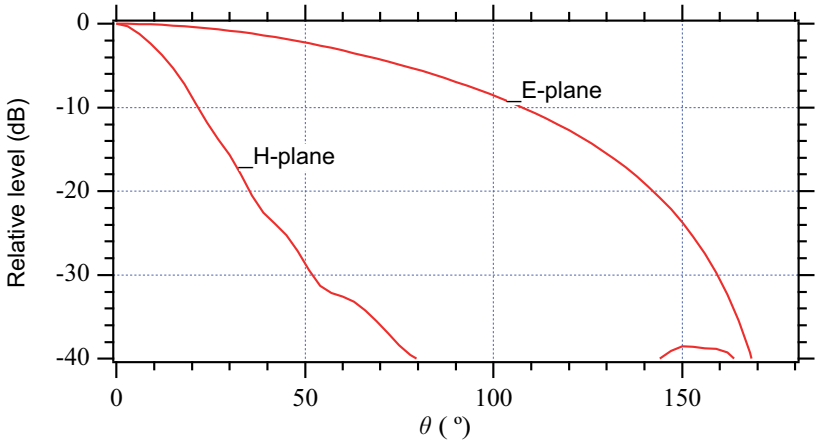
We can transform (8.18) to a universal integral by introducing $u = 2x'/a$ and the maximum phase variation $\Phi_{\max}$ over the aperture:

$$\Phi_{\max} = k(a/2)^2/(2L_x) = ka^2/(8L_x) \ . \tag{8.20}$$

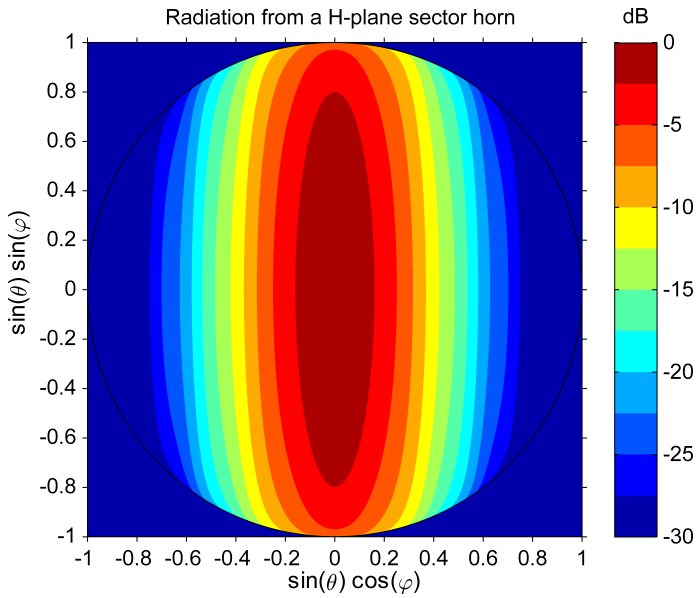
Then,

$$\tilde{A}(k_x) = F(k_x, a, \Phi_{\max}) = \frac{a}{2}\int_{-1}^{1} \cos(\pi u/2)e^{-ju^2\Phi_{\max}}e^{jk_x a u/2}du \ . \tag{8.21}$$

This Fourier transform can be evaluated numerically and general results be presented as a function of $a\sin\theta/\lambda$ for certain choices of the maximum phase variation $\Phi_{\max}$. Eq. (8.21) can even be evaluated analytically, but the solution has a very complicated form, so we will not present it here. It is most conveniently evaluated numerically.



(a)



(b)

Figure 8.12: *Radiation patterns of an H-plane sector horn for $a = 5.5\lambda$, $b = 0.25\lambda$ and $L_x = 6\lambda$. (a) E- and H-plane patterns. (b) Contour plot.

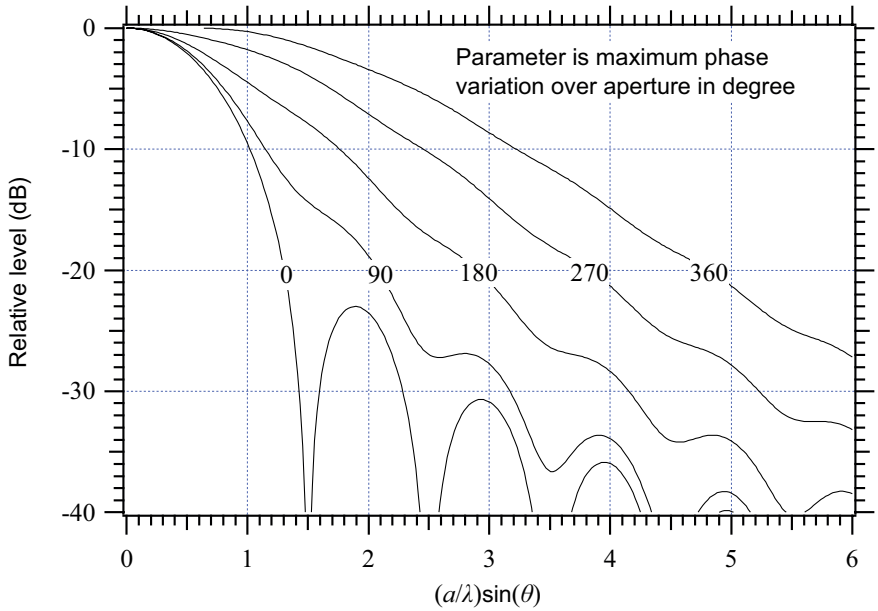


Figure 8.13: *Universal H-plane patterns for H-plane sector and pyramidal horns.

8.3.3 Radiation patterns

The E- and H-plane radiation patterns of an example of a H-plane sector horn are shown in Fig. 8.12*. We see that there are no sidelobes and shoulders. For larger horns there will be sidelobes in H-plane, but the highest first lobe will always be more than 22 dB below the main beam maximum. The *universal H-plane pattern* based on (8.21) is shown in Fig. 8.13 with $\Phi_{\max}$ as parameter*. This must be multiplied with the Huygens source pattern $\cos^2(\theta/2)$ in order to be complete. We see that the sidelobes are less than those for the E-plane sector horn. The E-plane pattern in Fig. 8.12 is very similar to the H-plane pattern in Fig. 8.9 even though the aperture distributions are different. The reason is that the aperture diameters in the two planes are so small that the dominant contribution to the shapes of the radiation patterns are the Huygens factor $\cos^2(\theta/2)$ which is the same in both cases.

8.4 Pyramidal horn

The pyramidal horn is shown in Fig. 8.14. It is flared both in the E- and H-planes, and the flare angles and the apexes of the flares may generally be different in the two planes. If the apexes coincide we may assume a propagating spherical mode inside the horn. Under the same approximation as used in Sections 8.2 and 8.3 this may be expressed by

$$\mathbf{E}(x', y', z') = E'_0 \cos\left(\frac{\pi\alpha_x}{2\alpha_H}\right) \frac{1}{R} e^{-jkR} \hat{\boldsymbol{\alpha}}_y \tag{8.22}$$

for $|\alpha_x| < \alpha_H$ and $|\alpha_y| < \alpha_E$, with

$$R = \sqrt{x'^2 + y'^2 + (z' + L)^2} \ .$$

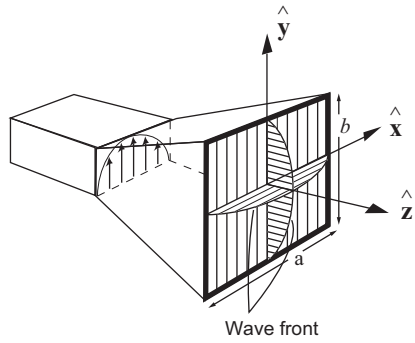


Figure 8.14: Geometry of pyramidal horn.

The spherical aperture is described by

$$\begin{aligned}\mathbf{r}_a(x',y') &= x'\hat{\mathbf{x}} + y'\hat{\mathbf{y}} + (\sqrt{L^2 - x'^2 - y'^2} - L)\hat{\mathbf{z}} \\ &= L\sin\alpha_x\hat{\mathbf{x}} + L\sin\alpha_y\hat{\mathbf{y}} + (L\sqrt{1-\sin^2\alpha_x - \sin^2\alpha_y} - L)\hat{\mathbf{z}}\ ,\end{aligned}$$

with the surface normal vector $\hat{\mathbf{n}}_a = \sin\alpha_x\hat{\mathbf{x}} + \sin\alpha_y\hat{\mathbf{y}} + \sqrt{1-\sin^2\alpha_x - \sin^2\alpha_y}\hat{\mathbf{z}}$. In the same way as before the far-field function becomes, but now by integration over the spherical phase-front in the aperture,

$$\mathbf{G}(\hat{\mathbf{r}}) = -\int_{-\alpha_H}^{\alpha_H}\int_{-\alpha_E}^{\alpha_E} E_0' \cos\left(\frac{\pi\alpha_x}{2\alpha_H}\right)\mathbf{G}_H(\hat{\boldsymbol{\alpha}}_y, \hat{\mathbf{n}}_a, \hat{\mathbf{r}})e^{jk\mathbf{r}_a\cdot\hat{\mathbf{r}}}L^2 d\alpha_x d\alpha_y\ , \tag{8.23}$$

with $\mathbf{G}_H$ the same as in (8.5) and $\hat{\mathbf{n}}_a = \sin\alpha_x\hat{\mathbf{x}} + \sin\alpha_y\hat{\mathbf{y}} + \cos\alpha_x\cos\alpha_y\hat{\mathbf{z}}$, the normal vector to the doubly curved aperture surface.

We get now in the same way as in Subsections 8.2.1 and 8.3.1 by using the *paraxial approximation* the following aperture field

$$\mathbf{E}_a(x',y') = E_0'\cos(\pi x'/a)e^{-jk[x'^2+y'^2]/(2L)}\hat{\mathbf{y}} \tag{8.24}$$

for $|x'| < a/2$ and $|y'| < b/2$ where L is the distance from the aperture plane to the apex of the horn. We can generalize this to different apex locations L_x and L_y for the xz- and yz-planes of the horn as follows

$$\mathbf{E}_a(x',y') = E_0'\cos(\pi x'/a)e^{-jk[x'^2/(2L_x)+y'^2/(2L_y)]}\hat{\mathbf{y}} \tag{8.25}$$

for $|x'| < a/2$ and $|y'| < b/2$. Further, in the same way as before, we obtain the far-field function

$$\mathbf{G}(\hat{\mathbf{r}}) = -2C_k\cos^2(\theta/2)[\sin\varphi\hat{\boldsymbol{\theta}} + \cos\varphi\hat{\boldsymbol{\varphi}}]\tilde{A}(k\sin\theta\cos\varphi)\tilde{B}(k\sin\theta\sin\varphi)\ , \tag{8.26}$$

with $\tilde{A}(k_x)$ the same as in (8.18) and $\tilde{B}(k_x)$ the same as in (8.11). This means that the E-plane pattern is equal to that of the E-plane sector horn, and the H-plane pattern is equal to that of the H-plane sector horn. Therefore, the universal E- and H-plane patterns in Fig. 8.9 and 8.13 apply also to the pyramidal horn. These two universal patterns are for $\phi_{\max} = 90°$ combined

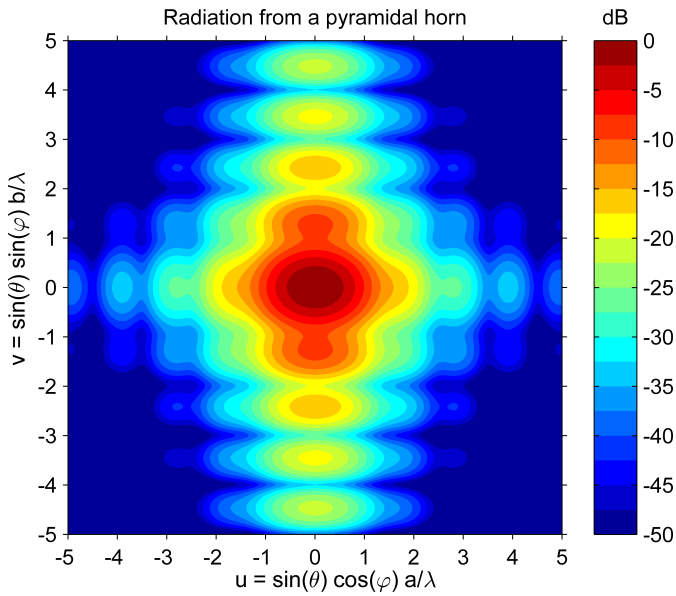


Figure 8.15: *Universal contour radiation pattern of a pyramidal horn when the maximum phase variation over the aperture is 90° in both planes. 5 dB contours down to −45 dB.

to a universal contour plot in Fig. 8.15*. We see that the highest sidelobes appear in the two principal planes, which we already know from the treatment of rectangular apertures in Section 7.4.

There exists also a MATLAB code for calculation of the directivity of a pyramidal horn antenna with given dimensions of the feed waveguide and the aperture. Pyramidal horns are often used as *standard gain horns*.

8.5 Corrugated surfaces

We have seen in Sections 8.2, 8.3 and 8.4 that the E-plane radiation patterns of horn antennas have high sidelobes or shoulders in the main beam. This is undesirable, and we shall discuss how this can be improved by corrugating the E-plane walls of the horn (i.e., the walls that are parallel with the H-field and normal to the E-field in the aperture).

Fig. 8.16 shows the cross-section of a surface which is corrugated with grooves orthogonal to the plane of incidence. We explain below how this surface behaves for plane wave incidence. The coordinate system used in the explanations are given in Fig. 8.16.

8.5.1 Principle of operation in H-plane

First we consider the polarization for which the E-field is orthogonal to the plane of incidence (Fig. 8.16a). This corresponds to the H-plane in antennas with two planes of symmetry. The E-field is parallel with the grooves and the ridges. If the grooves are narrower than $\lambda/2$, the

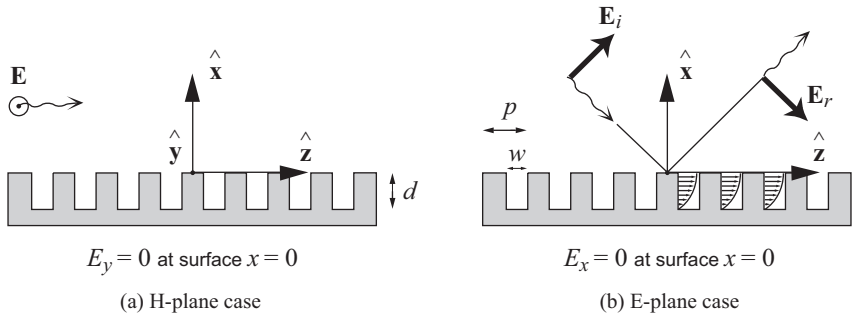


Figure 8.16: Cross-section of corrugated surface and its coordinate system.

E-field cannot penetrate into the grooves because the corresponding parallel plate waveguide formed by the walls in each groove has a cut-off frequency when the groove width is $w = \lambda/2$. Therefore, the E-field will instead be short-circuited by the ridges to become zero at the surface, as for a smooth PEC. The induced electric current at the surface will be $\mathbf{J} = \hat{\mathbf{x}} \times \mathbf{H}$ (from (4.14)) where $\mathbf{H}$ is the total H-field. We consider the H-plane, so the y-component of $\mathbf{H}$ is zero. Therefore, $\mathbf{J}$ will have a y-component only, and this can float in the surface along the ridges without being effected by the grooves. So, the corrugated surface works as a smooth PEC in H-plane.

8.5.2 Principle of operation in E-plane

In the E-plane the E-field has no y-component, and the H-field is oriented in y-direction and has a non-zero value at the surface (Fig. 8.16b). The induced current $\mathbf{J} = \hat{\mathbf{x}} \times \mathbf{H}$ will be z-directed and will try to float down into the grooves. For this polarization the parallel plate waveguide formed by the grooves has no cut-off, so the short-circuits at the bottom of the grooves will be transformed to open-circuited lines at the surface if the groove depth $d = \lambda/4$. Therefore, the z-directed current will have to be zero at the surface, and this in turn means that the corresponding y-component of the H-field will be zero, so the corrugations have created an artificial magnetic conducting surface in E-plane. From studying the parallel-plate transmission line formed by the grooves, when the groove depth d is different from $\lambda/4$, we can show that the ratio between E_z and H_y averaged over the surface becomes

$$Z_z = -E_z/H_y = j\left(\frac{w}{p}\right)\eta\tan(k_g d)\ , \tag{8.27}$$

where η is the wave impedance inside the corrugations (most often free space), and k_g is the guide wavenumber in the corrugations for propagation in the direction normal to the corrugated interface (for straight air-filled corrugations in a plane surface $k_g = k$). Eq. (8.27) is valid when the corrugation period is $p < 0.5\lambda$, preferable $p < 0.25\lambda$. The condition $p < 0.5\lambda$ is also necessary in order to avoid grating-lobes and multiple beams at the surface. The relation in (8.27) is commonly referred to as the surface impedance.

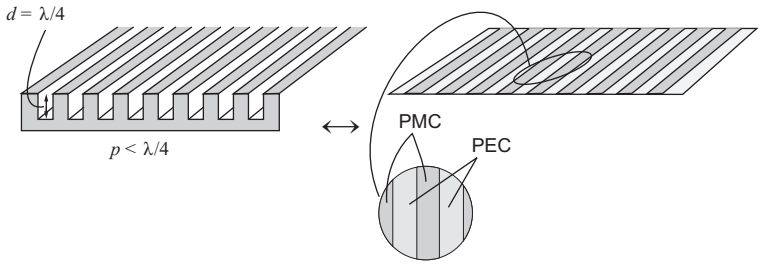


Figure 8.17: A transversely corrugated soft surface (left) and its PEC/PMC strip model (right).

8.5.3 Impedance boundary condition

We may replace the corrugated surface by a thin sheet of equivalent electric and magnetic surface currents and a null-field region behind it. From the discussion in Subsection 8.5.2 and from (4.12) the boundary condition in E-plane at this surface becomes

$$(\hat{\mathbf{x}} \times \mathbf{E})_y = -\mathbf{M}_y \ , \tag{8.28}$$

$$(\hat{\mathbf{x}} \times \mathbf{H})_z = (\hat{\mathbf{x}} \times \mathbf{H}_y) \cdot \hat{\mathbf{z}} = J_z = M_y/Z_z \ , \tag{8.29}$$

where M_y is the induced equivalent magnetic current and $J_z = M_y/Z_z$ is the induced electric current at the surface. In H-plane we have

$$(\hat{\mathbf{x}} \times \mathbf{E})_z = 0 \ , \tag{8.30}$$

$$(\hat{\mathbf{x}} \times \mathbf{H})_y = J_y \ , \tag{8.31}$$

where J_y is the induced electric current floating along the corrugation ridges. The surface impedance is infinite when $k_g d = \pi/2$, i.e., $d = \lambda_g/4$, so then,

$$(\hat{\mathbf{x}} \times \mathbf{H})_z = H_y = 0 \ , \tag{8.32}$$

$$(\hat{\mathbf{x}} \times \mathbf{E})_z = E_y = 0 \ , \tag{8.33}$$

which corresponds to the boundary condition at a surface made of parallel thin strips of PEC and PMC oriented in y-direction (Fig. 8.17).

The impedance boundary conditions in (8.28) - (8.31) and (8.32) - (8.33) can be used for accurate analysis of wave propagation along corrugated surfaces, provided wave propagation is orthogonal to the corrugations, i.e., in y-direction in Fig. 8.16. When the wave propagation has a component along the direction of the corrugations, the impedance boundary condition is not accurate.

8.5.4 Corrugations as soft surface

There are some nearly equivalent boundary conditions that are easier to use than the impedance boundary condition for constructing approximate field solutions and for interpreting field behavior. They are the soft and hard boundary conditions introduced in Section 4.1.4. By using Maxwell's equations it is possible to show that $H_y = 0$ in (8.32) corresponds to

$E_x = 0$. Therefore, by using also (8.33) we see that the transversely corrugated surface is soft regardless of polarisation because

$$E_y = E_x = 0 \tag{8.34}$$

and for this reason we refer to it as a soft surface. It is also possible, yet much more complicated, to realize polarization independent hard surfaces by means of corrugations [1]. Different realizations of soft surfaces and their bandwidth are summarized in [2].

8.5.5 Bandwidth and surface waves

The transversely corrugated soft surface is known to have a wide bandwidth. This means that if it is long enough, the vertical x-component of the E-field will be zero or close to zero at the surface even if the surface impedance is not infinity. The usable bandwidth is within the range where

$$\pi/2 < k_g d < \pi \ ,$$

which for air-filled straight corrugations means that

$$\lambda/4 < d < \lambda/2 \ .$$

This corresponds to a relative bandwidth of $2:1$. In practice the usable relative bandwidth is smaller, but can be up to 1.8. The best soft E-plane performance is obtained when $k_g d = \pi/2$ or $d \approx \lambda/4$.

The corrugated soft surface does not work when the corrugations are shallower than $\lambda/4$, because then *surface wave* solutions may exist. Surface waves are waves propagating along the surface with an exponential decay normal to it. These waves destroy the soft boundary condition completely. The same types of surface waves may be present when

$$\lambda/2 < d < 3\lambda/4 \ .$$

In practice, the depth may be a bit shallower than $\lambda/4$ and still give the soft boundary condition.

8.6 Corrugated pyramidal horn

This horn is similar to the pyramidal horn with smooth PEC walls, except that the two E-plane walls are corrugated to obtain a polarization independent soft surface as explained in Section 8.5. We can construct the aperture field by using the facts that $E_y = 0$ at the smooth PEC walls, and $E_y = E_x = 0$ at the corrugated *soft surface* wall. For y-polarization this gives a field solution inside the rectangular waveguide of the form

$$\mathbf{E}_w(x', y') = \cos(\pi x'/a)\cos(\pi y'/b)\hat{\mathbf{y}} \ . \tag{8.35}$$

To an acceptable approximation the same field solution will also be present inside a rectangular waveguide with all four walls corrugated. The advantage of corrugating all walls is that the desired $\cos(\pi x'/a)\cos(\pi y'/b)$ aperture distribution is obtained for any polarization

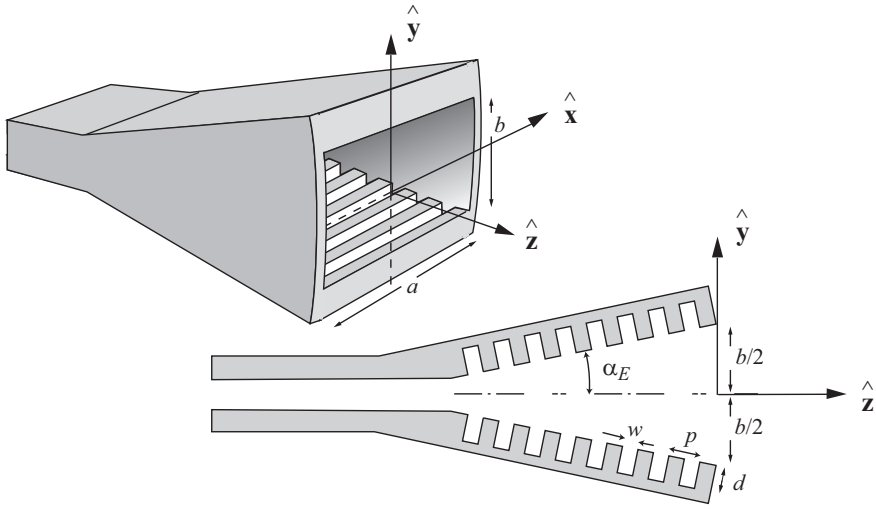


Figure 8.18: Example of corrugated pyramidal horn and its cross section.

of the fields within the waveguide. If we now use the same assumptions as for the previous rectangular horn cases, we may construct the aperture field to be

$$\mathbf{E}(x', y') = \cos(\pi x'/a)\cos(\pi y'/b)e^{-jk[x'^2/(2L_x)+y'^2/(2L_y)]}\hat{\mathbf{y}} \ . \tag{8.36}$$

If the corrugations are straight as shown in Fig. 8.18, the flare angle α in H-planes, must be very small in order to avoid generation of higher order modes in the horn. This can be avoided by curving the corrugations to follow the shape of the wavefronts inside the horn.

The two aperture integrals are obtained in the same way as before. We achieve the same type of universal radiation pattern in both E- and H-planes, i.e., the nice universal H-plane pattern in Fig. 8.13. The grooves do not need to make an angle of 90° to the surface, but they have a larger bandwidth if they do.

8.7 Smooth conical horn

The conical horn is shown in Fig. 8.19. It is rotationally symmetric with flare angle α. We will construct the aperture field by using the flared cylindrical waveguide approach like we did in Sections 8.2 and 8.3. We only consider BOR_1 excitation by a TE_{11} *circular cylindrical waveguide mode*.

The field solution of this mode can be found in text books on elementary field theory. For linear y-polarization (see Eq. (7.44)) it is

$$E_{E} = E_\rho = \frac{1}{\rho'} J_1(\kappa_1 \rho'/a)\sin\varphi' \ , \tag{8.37}$$

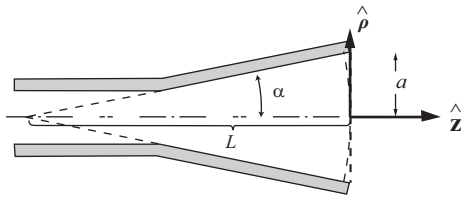


Figure 8.19: Cross section of conical horn antenna.

$$\begin{aligned} E_{H} = E_{\varphi} &= \frac{\partial}{\partial \rho'}[J_1(\kappa_1 \rho'/a)]\cos\varphi' \\ &= \left[-\frac{\kappa_1}{a}J_2(\kappa_1\rho'/a) + \frac{1}{\rho'}J_1(\kappa_1\rho'/a)\right]\cos\varphi' \end{aligned} \tag{8.38}$$

for $\rho' \leq a$, where $\kappa_1 = 1.841$, J_1 is the Bessel function of first kind and first order, and J_2 is the Bessel function of first kind and second order. κ_1 is determined from the requirement $E_\varphi = 0$ at $\rho = a$. We can now use the BOR_1 relations in Section 7.5 to express this as

$$E_{a_y}(\rho', \varphi') = E_{\mathrm{co}_{45^\circ}}(\rho') - E_{\mathrm{xp}_{45^\circ}}(\rho')\cos(2\varphi') \ , \tag{8.39}$$

$$E_{a_x}(\rho', \varphi') = E_{\mathrm{xp}_{45^\circ}}(\rho')\sin(2\varphi') \ , \tag{8.40}$$

where

$$E_{\mathrm{co}_{45^\circ}}(\rho') = \left[\frac{1}{\rho'}J_1(\kappa_1\rho'/a) + \frac{\kappa_1}{2a}J_2(\kappa_1\rho'/a)\right] = \frac{\kappa_1}{2a}J_0(\kappa_1\rho'/a) \ , \tag{8.41}$$

$$E_{\mathrm{xp}_{45^\circ}}(\rho') = \frac{\kappa_1}{2a}J_2(\kappa_1\rho'/a) \ , \tag{8.42}$$

in which J_0 is the zeroth order Bessel function of the first kind.

We can now use the same approach as in Section 8.2 to construct the aperture field on the spherical surface with the center of curvature in the apex of the horn. Doing this we arrive at

$$\mathbf{E}_a(\theta', \varphi') = E_{a_y}(\theta', \varphi')\hat{\mathbf{co}}(\theta', \varphi') + E_{a_x}(\theta', \varphi')\hat{\mathbf{xp}}(\theta', \varphi') \tag{8.43}$$

where

$$E_{a_y}(\theta', \varphi') = E_{\mathrm{co}_{45^\circ}}(\theta') - E_{\mathrm{xp}_{45^\circ}}(\theta')\cos(2\varphi') \ , \tag{8.44}$$

$$E_{a_x}(\theta', \varphi') = E_{\mathrm{xp}_{45^\circ}}(\theta')\sin(2\varphi') \ , \tag{8.45}$$

with

$$E_{\mathrm{co}_{45^\circ}}(\theta') = J_0(\kappa_1\theta'/\alpha) \ , \tag{8.46}$$

$$E_{\mathrm{xp}_{45^\circ}}(\theta') = J_2(\kappa_1\theta'/\alpha) \ , \tag{8.47}$$

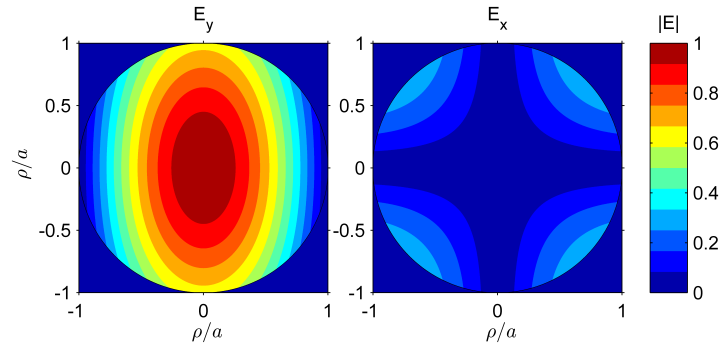
for $\theta' \leq \alpha$ where θ' is in radians. Here we have omitted the constant $(\kappa_1/2a)$ and have used the co- and cross-polar unit vectors defined by (2.55) - (2.56) in Section 2.3.5. The far-field function is obtained by using (7.7) in Section 7.2.1 over the spherical cap, with $\eta\mathbf{J}_a = -\mathbf{E}_a(\theta', \varphi')$.

We can transform the field to the plane aperture by using the *paraxial approximation*. In the same way as before this yields

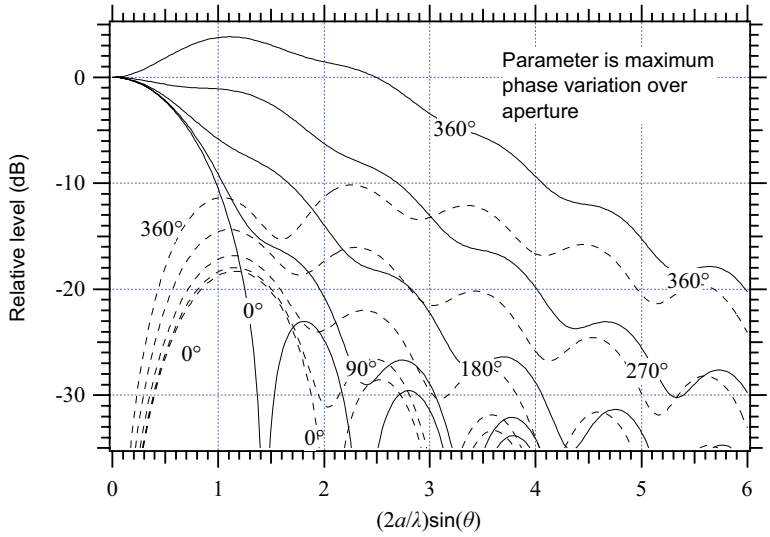
$$E_a(\rho', \varphi') = [E_{a_y}(\rho', \varphi')\hat{\mathbf{y}} + E_{a_x}(\rho', \varphi')\hat{\mathbf{x}}]e^{-jk\rho'^2/(2L)} \ , \tag{8.48}$$

with

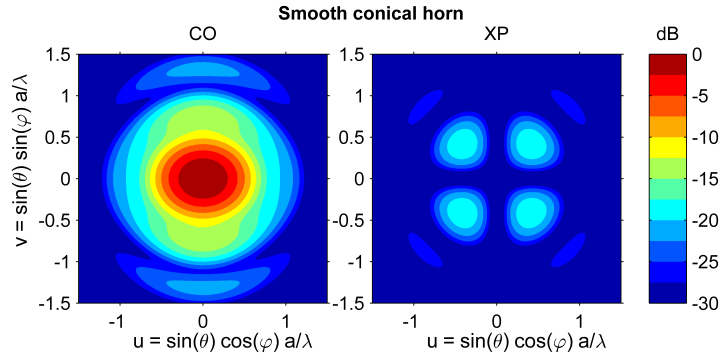
$$E_{a_y}(\rho', \varphi') = E_{\mathrm{co}_{45^\circ}}(\rho') - E_{\mathrm{xp}_{45^\circ}}(\rho')\cos(2\varphi') \ , \tag{8.49}$$



(a)



(b)



(c)

Figure 8.20: *Co- and cross-polar aperture-field and radiation patterns of a smooth conical horn. (a) Contour plot of aperture field. (b) Universal radiation patterns in 45°-plane. The parameter is the maximum phase variation $\phi_{\max}$ over the aperture. (c) Contour plots of universal radiation pattern for $\phi_{\max} = 90°$.

$$E_{a_x}(\rho',\varphi') = E_{\mathrm{xp}_{45^\circ}}(\rho')\sin(2\varphi') \ , \tag{8.50}$$

where
$$E_{\mathrm{co}_{45^\circ}}(\rho') = J_0(\kappa_1\rho'/a) \ , \tag{8.51}$$

$$E_{\mathrm{xp}_{45^\circ}}(\rho') = J_2(\kappa_1\rho'/a) \ . \tag{8.52}$$

We have now a plane aperture and we can use the results in (7.53) to (7.56) in Subsection 7.5.1. Doing so, we obtain

$$\mathbf{G}(\hat{\mathbf{r}}) = 2C_k\cos^2(\theta/2)\cdot\{[\tilde{E}_{\mathrm{co}_{45^\circ}}(\theta) - \tilde{E}_{\mathrm{xp}_{45^\circ}}(\theta)\cos 2\varphi]\hat{\mathbf{co}} + \tilde{E}_{\mathrm{xp}_{45^\circ}}(\theta)\sin 2\varphi\hat{\mathbf{xp}}\} \tag{8.53}$$

with $\hat{\mathbf{co}}$ and $\hat{\mathbf{xp}}$ given by (2.55) and (2.56) and

$$\tilde{E}_{\mathrm{co}_{45^\circ}}(\theta) = 2\pi\int_0^a J_0(\kappa_1\rho'/a)e^{-jk\rho'^2/(2L)}J_0(k\rho'\sin\theta)\rho' d\rho' \ , \tag{8.54}$$

$$\tilde{E}_{\mathrm{xp}_{45^\circ}}(\theta) = -2\pi\int_0^a J_2(\kappa_1\rho'/a)e^{-jk\rho'^2/(2L)}J_2(k\rho'\sin\theta)\rho' d\rho' \ . \tag{8.55}$$

Universal patterns are obtained by substituting $u = \rho'/a$ and $\Phi_{\max} = ka^2/(2L)$, the maximum phase variation over the aperture. This gives

$$\tilde{E}_{\mathrm{co}_{45^\circ}}(\theta) = 2\pi a^2\int_0^1 J_0(\kappa_1 u)e^{-j\Phi_{\max}u^2}J_0(kau\sin\theta)u du \ , \tag{8.56}$$

$$\tilde{E}_{\mathrm{xp}_{45^\circ}}(\theta) = -2\pi a^2\int_0^1 J_2(\kappa_1 u)e^{-j\Phi_{\max}u^2}J_2(kau\sin\theta)u du \ . \tag{8.57}$$

The *universal radiation patterns* in the 45°-plane as given in (8.56) and (8.57) are shown in Fig. 8.20(b)*, and the *uv*-contour plot of it in Fig. 8.20(c)*. The far-field function must in this case be multiplied by $\cos^2(\theta/2)$ in order to get the correct pattern including the Huygens source. But this represents a minor correction in the vicinity of the main beam (for large horns with narrow beams). We see that the cross-polar sidelobe in the 45°-plane is very large and that the main beam gets a shoulder or a dip in the center for large phase errors. Therefore, conical horns with smooth walls are most often designed with phase variations $\phi_{\max}$ smaller than 45°. The cross-polar sidelobe can be reduced and the performance for large $\phi_{\max}$ improved by means of corrugations, as described in the next section. Similar performance as with a corrugated horn, but over a much more limited bandwidth, can be obtained by using a dual mode horn (having smooth walls as well) [3]. This is a horn which is designed to contain both the TE_{11} and TM_{11} cylindrical waveguide modes in the aperture, including the quadratic phase.

8.8 Corrugated soft conical horn

The conical horn antenna has a high cross-polar sidelobe in the 45°-plane corresponding to differences in the E- and H-plane radiation patterns. The reason for this is that the metal cone is a PEC which has a soft boundary condition in H-plane and a hard boundary condition in E-plane. We know from Sections 8.5 and 8.6, that this can be improved by providing the horn with corrugated walls as shown in Fig. 8.21. The field distribution inside a corrugated circular waveguide, when the frequency is within the range where the corrugation represent a soft surface, can be found to be approximately

$$\mathbf{E}(\rho') = J_0(\kappa_{01}\rho'/a)\hat{\mathbf{y}} \ , \tag{8.58}$$

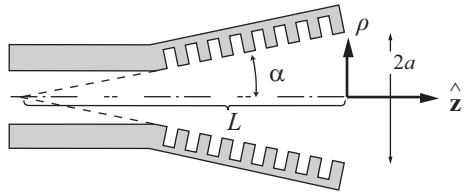


Figure 8.21: Cross section of example of conical corrugated horn antenna.

where $\kappa_{01} = 2.405$ is the value of x for the first null of $J_0(x)$. This aperture distribution is rotationally symmetric, zero at the rim, and there is no cross-polarization as shown in Fig. 8.22(a). These characteristics are due to the *soft surface* created by the corrugations. We can now use this to find an approximate aperture distribution for the conical corrugated horn by using the same procedure and paraxial approximation as in Section 8.7. The result is

$$\mathbf{E}(\rho') = J_0(\kappa_{01}\rho'/a)e^{-jk\rho'^2/(2L)}\hat{\mathbf{y}} \ , \tag{8.59}$$

where L is the length of the flare of the horn measured from the apex. The far-field function is obtained by using (8.53) with $E_{\mathrm{xp}_{45^\circ}}(\theta) = 0$ and

$$\tilde{E}_{\mathrm{co}_{45^\circ}}(\theta) = 2\pi \int_0^a J_0(\kappa_{01}\rho'/a)e^{-jk\rho_1'^2/(2L)} J_0(k\rho' \sin\theta)\rho' d\rho' \ . \tag{8.60}$$

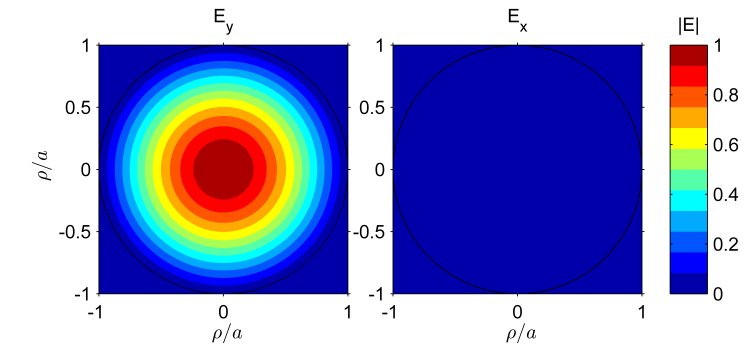
This is almost the same formula as for the co-polar pattern of the smooth conical horn, but κ_{01} in (8.60) is different from κ_1 in the smooth horn formula.

The *universal pattern* is obtained by substituting $u = \rho'/a$ and $\Phi_{\max} = ka^2/(2L)$ the maximum phase variation over the aperture. Then,

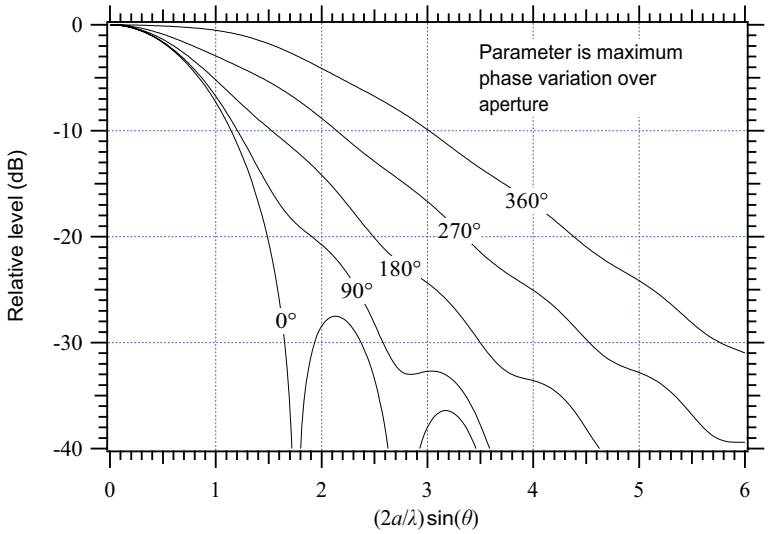
$$\tilde{E}_{\mathrm{co}_{45^\circ}}(\theta) = 2\pi a^2 \int_0^1 J_0(\kappa_{01}u)e^{-j\Phi_{\max}u^2} J_0(kau \sin\theta)u du \ . \tag{8.61}$$

This is plotted in Fig. 8.22(b) and 8.22(c)*. We see that the sidelobes are lower than for the smooth wall horn (in Fig. 8.20), and they vanish when the phase variation over the aperture increases. In the present simplified theory the cross-polar sidelobes vanish. In practice, there will always be cross-polar sidelobes present, but in most designs they are more than 30 dB smaller than the main beam maximum.

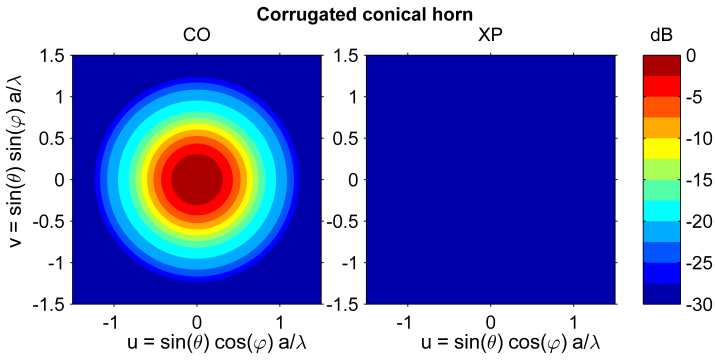
It is important to note that the field distribution in (8.59) is an approximation. The actual propagating mode in the horn is a hybrid mode consisting of both a TE_{10} and a TM_{10} part which both have to be present in order to satisfy the boundary conditions at the corrugated wall. However, at the frequency for which $k_g d = \pi/2$, the *hybrid mode* has the approximate field distribution in (8.59) with zero cross-polarization, which is referred to as a balanced hybrid HE_{11} mode. This balanced field distribution is approximately valid for the co-polar radiation field over the whole interval for which $\pi/2 < k_g d < 0.9\frac{1}{4}$, whereas the cross-polar fields changes over this band.



(a)



(b)



(c)

Figure 8.22: *Co- and cross-polar aperture-fields and radiation patterns of conical corrugated (soft) horn antenna. (a) Contour plot of aperture field. (b) Universal radiation patterns in 45°-plane. The parameter is the maximum phase variation $\phi_{\max}$ over the aperture. (c) Contour plots of universal radiation pattern for $\phi_{\max} = 90°$. The cross-polar level is vanishing with this simplified theoretical model, and it can in practice also be below -30 dB over significant bandwidth.

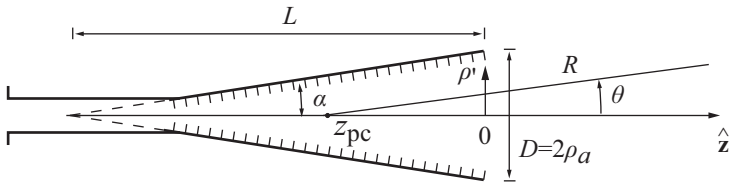


Figure 8.23: Corrugated conical horn antenna with far-field phase center $z_{\rm pc}$.

8.9 Modeling corrugated horn with Gaussian beams

The corrugated horn has a radiation pattern which has an approximate Gaussian shape. This may be used to obtain some general design curves that are very convenient to use for initial horn designs. These are based on the *Gaussian beam* formulas in Section 7.6 and will be formulated in terms of the dimensions shown in Fig. 8.23.

The aperture distribution of the linearly y-polarized *corrugated conical horn* was in Section 8.8 found to be

$$\mathbf{E}_a(\rho) = J_0(\kappa_{01}\rho'/a)e^{-jk\rho'^2/(2L)}\hat{\mathbf{y}} \qquad \text{for } \rho' \le a \ . \tag{8.62}$$

Let us approximate this as a Gaussian aperture distribution of the form

$$\mathbf{E}_a(\rho') \approx Ke^{-\rho'^2/\rho_a^2}e^{-jk\rho'^2/(2L)}\hat{\mathbf{y}} \qquad \text{for } 0 < \rho' < \infty \ . \tag{8.63}$$

8.9.1 Choosing the Gaussian beam parameters

We choose K and ρ_a in (8.63) such that the power integrals in the aperture are equal and that the directivities are equal when $L \to \infty$. The Gaussian distribution in (8.63) extends to infinity, so the conservation of power requires

$$\int_0^a [J_0(\kappa_{01}\rho'/a)]^2\rho' d\rho' = K^2\int_0^\infty e^{-2\rho'^2/\rho_a^2}\rho' d\rho' \ , \tag{8.64}$$

which becomes

$$[J_1(\kappa_{01})]^2a^2/2 = K^2\rho_a^2/4 \ . \tag{8.65}$$

In order to give equal directivities in the far-field for the case that $L \to \infty$, the following equation must be satisfied[2]

$$\int_0^a J_0(\kappa_{01}\rho'/a)\rho' d\rho' = \int_0^\infty Ke^{-\rho'^2/\rho_a^2}\rho' d\rho' \ , \tag{8.66}$$

which gives

$$a^2J_1(\kappa_{01})/\kappa_{01} = K\rho_a^2/2 \ . \tag{8.67}$$

If we combine (8.65) and (8.67) we get

$$\rho_a = \sqrt{2}(a/\kappa_{01}) \approx 0.59a \ . \tag{8.68}$$

[2] See Section 7.5 on page 246.

The Gaussian beam formulas become more applicable for broad beams if we measure ρ_a and a along a wavefront, instead of normal to the z-axis. This means that (8.68) becomes $\rho_a/L = 0.59\alpha$ where α is the half flare angle of the horn.

8.9.2 Radiation field

The radiation field of the horn is now described by the formulas in Section 7.6.4 with $\rho_a = 0.59a$ and $C_a = 1/L$. This means that the far-field function is given by

$$\mathbf{G}(\hat{\mathbf{r}}) = K\frac{\rho_a}{\theta_0}e^{-(\theta/\theta_0)^2}e^{-jkz_{\rm pc}\theta^2}e^{j\phi_\infty}\hat{\mathbf{y}} \ , \tag{8.69}$$

with (all angles in radians)

$$\theta_0 = \sqrt{\theta_{\rm dif}^2 + \theta_{\rm GO}^2} \ , \tag{8.70}$$

$$\theta_{\rm dif} = 2/(k\rho_a) \ , \tag{8.71}$$

$$\theta_{\rm GO} = \rho_a/L \ , \tag{8.72}$$

$$z_{\rm pc} = -L(\theta_{\rm GO}/\theta_0)^2 \ , \tag{8.73}$$

$$\phi_\infty = -\arctan(\theta_{\rm dif}/\theta_{\rm GO}) \ . \tag{8.74}$$

These equations give the 8.7 dB beamwidth θ_0 and the *phase center* location $z_{\rm pc}$ as a function of the radius a of the horn aperture and the flare length L of the horn, when $z = 0$ in the center of the aperture. We refer to the discussions in Sections 7.6.4 and 7.6.5, for interpreting $\theta_{\rm dif}$ as the beamwidth due to aperture diffraction and $\theta_{\rm GO}$ as that due to GO transmission.

We shall now describe three different choices of horn dimensions which we shall refer to as flare angle-controlled horns, aperture-controlled horns and maximum gain horns.

8.9.3 Flareangle-controlled horn

From Eqs. (8.70)- (8.72) we see that the beamwidth θ_0 is independent of the frequency if the beamwidth contribution $\theta_{\rm GO}$ due to GO transmission is much larger than the contribution of $\theta_{\rm dif}$ due to aperture diffraction, i.e.,

$$\theta_{\rm GO} \gg \theta_{\rm dif} \ , \qquad \rho_a/L \gg 2/(k\rho_a) \ ,$$

or equivalently if

$$L \ll \frac{k}{2}\rho_a^2 = \pi\rho_a^2/\lambda = 1.1a^2/\lambda \ .$$

This condition corresponds to that the phase variation over the aperture is large, satisfying

$$\phi_{\rm max} \gg \pi \ , \tag{8.75}$$

because

$$\phi_{\rm max} = ka^2/(2L) \gg a^2/\rho_a^2 = (\kappa_{01}/\sqrt{2})^2 = 2.9 \approx \pi \ . \tag{8.76}$$

In this case we achieve (in radians)

$$\theta_0 = \rho_a/L = 0.59a/L \qquad \text{and} \qquad z_{\rm pc} = -L \ . \tag{8.77}$$

We can refer to this case as an ideally flareangle-controlled mode of operation, as the beamwidth θ_0 is proportional to the half flareangle of the horn, which under the paraxial approximation, equals a/L radians. The phase center is located in the apex of the horn as $z_{\rm pc} = -L$ for all observation points. When Eq. (8.75) is not ideally satisfied, we will still refer to the horn as flareangle-controlled. We may call it slightly flareangle-controlled if

$$\pi < \phi_{\max} < 2\pi \tag{8.78}$$

and strongly flareangle controlled when $\phi_{\max} > 2\pi$.

8.9.4 Aperture-controlled horn

When the beamwidth contribution $\theta_{\rm dif}$ due to aperture diffraction is much larger than $\theta_{\rm GO}$ due to GO transmission, i.e., when

$$\theta_{\rm dif} \gg \theta_{\rm GO} \ ; \qquad L \gg 1.1a^2/\lambda$$

the phase variation over the aperture is small, satisfying

$$\phi_{\max} \ll \pi \ . \tag{8.79}$$

Then, in the limit $\phi_{\max} \to 0$, we see that (in radians)

$$\theta_0 = 2/(k\rho_a) = 0.54\lambda/a \ , \qquad z_{\rm pc} = 0 \ . \tag{8.80}$$

In this case the beamwidth θ_0 varies linearly with the wavelength and is inversely proportional to the frequency. It is determined completely by diffraction from the aperture plane at $z = 0$. We can therefore refer to the horn as ideally aperture-controlled. The beam radius $\rho(z)$ varies from $\rho = \rho_a$ in the aperture to $\rho(z) = z\theta_0$ (with θ_0 in radians) when $z \gg \pi\rho_a^2/\lambda$ (see Section 7.6.4). The phase center varies from $z_{\rm pc} = -L$ in the close near-field to $z_{\rm pc} = 0$ (i.e., at the aperture plane) for observation points in the far-field.

8.9.5 Maximum gain horn

It is possible to differentiate Eq. (8.70) with respect to a and find the aperture radius a which gives the narrowest beam, i.e., the highest directivity for a given constant length L of the horn. This appears when

$$\begin{aligned} &\theta_{\rm GO} = \theta_{\rm dif} \ ; \qquad L = 1.1a^2/\lambda \ , \\ &\text{for which} \qquad \phi_{\max} \approx \pi \ . \end{aligned} \tag{8.81}$$

For this case

$$\theta_0 \sqrt{(4/kL)} = 0.8\sqrt{(\lambda/L)} \ , \qquad z_{\rm pc} = -L/2 \ . \tag{8.82}$$

We refer to this case as a maximum gain horn. Its phase center is seen to be located halfway between the aperture and the apex. If we use (7.70), we see that the maximum directivity for a given length L of the horn can be expressed as

$$D_0 = 8/\theta_0^2 = 2kL = 4\pi L/\lambda \ . \tag{8.83}$$

with θ_0 in radians. The maximum gain horn is therefore the shortest horn that can give a certain specified directivity, and the length of it can be obtained from Eq. (8.83). Its far-field phase center is halfway between the aperture and the apex of the horn. By using Eqs. (8.83) and (8.69) we obtain a very simple equation to determine L, if a certain taper $T = 20\log|E(\theta)/E(0)|$ is specified at an angle θ;

$$L/\lambda = -0.07T/\theta^2 \ , \tag{8.84}$$

with θ in radians. It is also possible to design corrugated horns for providing as small beam radius as possible at a certain distance z_N from the aperture. The dimensions of such horns are determined by the equation

$$(L + z_N)\rho_a/L = 2z_N/k\rho_a \ . \tag{8.85}$$

Then, at the distance z_N from the aperture, the phase center of the beam is located at $z_{\rm pc} = -L/2$.

8.9.6 Design curves

We have plotted some design curves based on Eqs. (8.70) - (8.73) in Fig. 8.24 - 8.26*. If we want to design the horn so as to give a prescribed illumination taper at a certain angle θ_e, we can use Fig. 8.24 to find the required directivity of the horn. If we want -20 dB taper at $\theta_e = 10^\circ$, we read a directivity of $G_0 = 28$ dBi. This number can then be used in Fig. 8.25 to determine the aperture diameter and flareangle of the horn. We see that with $D = 11\lambda$ and $\alpha = 5^\circ$ we have a slightly aperture-controlled horn, but this will be $L = 63\lambda$ long. The shortest horn (maximum gain horn) having the same directivity has $D = 13\lambda$ and $\alpha \approx 8^\circ$, i.e., a length of $L = 46\lambda$. A flareangle-controlled horn will have $\alpha \approx 10^\circ$ and $D \approx 23\lambda$, for which $L = 65\lambda$. Finally, we can find the phase center location in Fig. 8.26.

The Gaussian beam model for the radiation pattern of a corrugated horn antenna may be used in general work on reflector antennas where an analytical feed pattern is needed. The model was used here to define aperture-controlled and flareangle-controlled horns. These can also be referred to as narrow-band and wide-band horns, respectively. The shortest horn giving a specified directivity is called a maximum gain horn. This also represents the boundary between flareangle-controlled and aperture-controlled horns at which the maximum phase variation over the aperture is $\phi_{\max} \approx \pi$.

8.9.7 Example: Design of dual band horn

We will now use the Gaussian beam formulas for corrugated conical horn antennas in order to determine the minimum size a horn can have in order to work at both 8.5 GHz (X-band) and 2.3 GHz (S-band) according to some specifications to be given. The dual-band operation is possible by using dual-depth corrugations, i.e., every second of the corrugations in the horn are quarter wavelength deep at S-band and every second a quarter wavelength deep at X-band. We will here only consider how to determine the major dimensions such as the diameter and length of the horn. The horn is intended to feed an offset paraboloid for nearly equal performance at S- and X-bands. To do this, we require that the feed pattern have a taper of between 12.5 dB and 15 dB for $\theta = 45^\circ$ at both frequencies. Furthermore, we

require that the phase center locations at S-band and X-band be closer than $\lambda_S/4$, where λ_S is the wavelength at S-band, in order to avoid a significant reduction in directivity due to defocussing of the paraboloidal reflector.

SOLUTION:

From Fig. 8.24 we see that we need a directivity of $D_0 \approx 13.2\,\text{dB}$ in order to obtain $-15\,\text{dB}$ taper at $\theta_e = 45°$. From the same figure we see that this corresponds to an $8.69\,\text{dB}$ half beamwidth of $\theta_0 = 34°$. We can calculate this more accurately by using

$$20\log|G(\theta)/G(0)| = -(\theta/\theta_0)^2 \cdot 8.69\,\text{dB} \ .$$

Inserting $15\,\text{dB}$ taper at $\theta_e = 45°$ gives $-15\,\text{dB} = -(45°/\theta_0)^2 8.69\,\text{dB}$. Thus, the $8.7\,\text{dB}$ width which provides $15\,\text{dB}$ taper for $\theta = 45°$ is

$$\theta_0 = 45°\sqrt{8.69/15} = 34° \ .$$

We can obtain identical beamwidths at both S- and X-band by using an ideal flareangle-controlled horn, i.e., a large horn with flareangle

$$\alpha = \theta_0/0.59 = 34°/0.59 = 58° \ .$$

This value can also be read from Fig. 8.25 by using $D_0 \approx 13.2\,\text{dB}$. The aperture diameter must be at least 10λ, which is very large.

We are satisfied with $-12.5\,\text{dB}$ taper for $\theta = 45°$ at X-band. A $-15\,\text{dB}$ taper corresponds to a gain of $13.2\,\text{dB}$, and $-12.5\,\text{dB}$ taper corresponds to about $11.2\,\text{dB}$ (see Fig. 8.24). Thus, we can accept $2\,\text{dB}$ less horn gain at S-band compared to X-band.

We have already shown that a flareangle-controlled horn should have $\alpha = 58°$ to produce $D_0 \approx 13.2\,\text{dB}$. Let us study the curve for $\alpha = 60°$ flareangle in Fig. 8.25 which is very close to the design value of 58°. We see that the directivity is larger than $11.2\,\text{dB}$ when $D > 2.4\lambda$. Thus, if we choose $D = 2.4\lambda_S$ at S-band, we obtain

$$D = 2.4\lambda_S \cdot \frac{8.5\,\text{GHz}}{2.3\,\text{GHz}} = 8.9\lambda_X \ ,$$

at X-band, for which the $\alpha = 60°$ curve in Fig. 8.25 gives $D_0 = 13.2\,\text{dB}$. The wavelength at S-band is

$$\lambda_S = \frac{300}{2.3}\ \text{mm} = 130\,\text{mm} \ ,$$

so the physical dimensions of the horn becomes

$$D = 2.4 \cdot 130\,\text{mm} = 320\,\text{mm} \ ,$$

$$L = D/(2\tan\alpha) = 90\,\text{mm} \ .$$

The phase center locations of this horn can be read from Fig. 8.26 or calculated by using (8.73). At S-band we get $z_{\text{pc}} = -0.65L$ and at X-band $z_{\text{pc}} = -0.97L$. The spacing between these two locations is $\Delta z_{\text{pc}} = 0.32L$. Thus,

$$\Delta z_{\text{pc}}/\lambda_S = 0.32 \cdot 90\,\text{mm}/130\,\text{mm} = 0.2215 \ .$$

This satisfies the requirement of $\Delta z_{\text{pc}} < \lambda_S/4$, so the horn satisfies the desired specification.

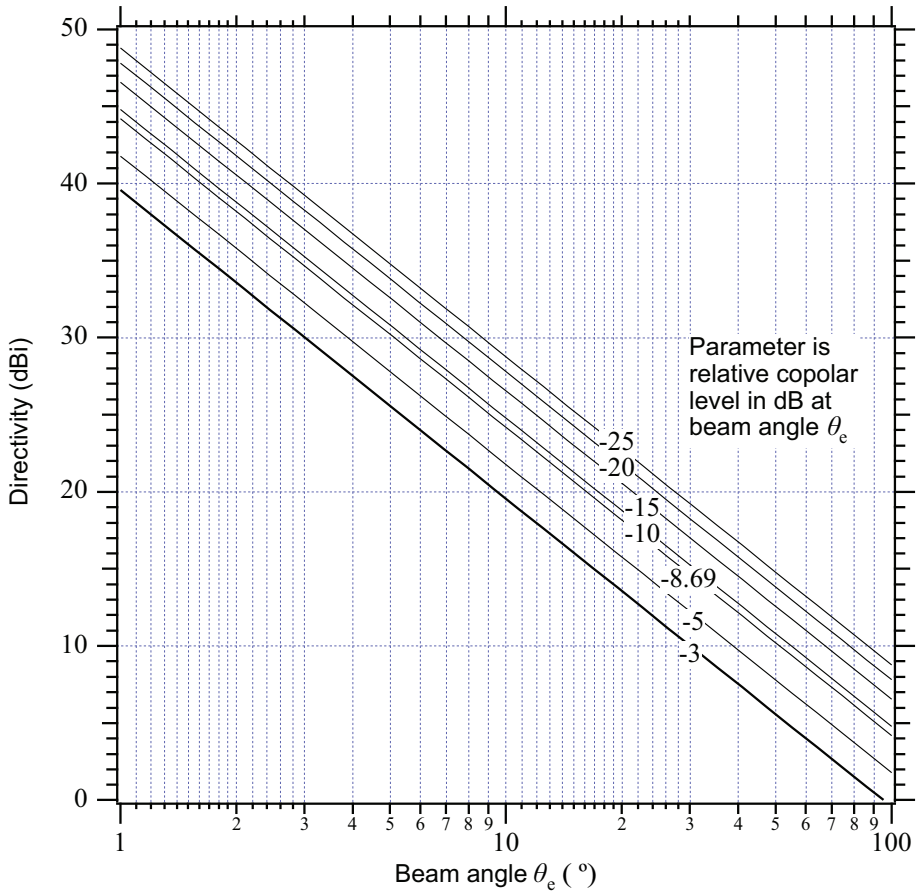


Figure 8.24: *Directivity as a function of the beam angle θ_e at which a Gaussian beam pattern is a certain level in dB smaller than the level at the center of the main lobe.

8.10 Other types of horn antennas

The corrugated conical horn was first described in 1966 [4], [5] and [6], and since then it has developed to have special features and for a large number of applications. We will here only refer to the basic paper by Thomas [7], the extensive book of Olver, Clarricoats et al. [8]. In addition, the compact corrugated horn in [9] is worth mentioning. It has constant beam width suitable for prime-focus paraboloids, low cross-polarization and low return loss over almost an octave bandwidths. Corrugated horns can also be provided with a lens to make them more compact [10], or they can be profiled to a shape that is more compact then their original conical shapes. The Gaussian beam model of corrugated horns presented in this chapter was first published in [12].

The corrugated surface is the basis of the concept of soft and hard surfaces[3], and the concept itself grew out from a desire to realize horns with uniform aperture distribution, so-called hard horns [13]. The concept of soft and hard surfaces has inspired to other ways of realizing

[3] The soft and hard boundary conditions are described in Section 4.1.4.

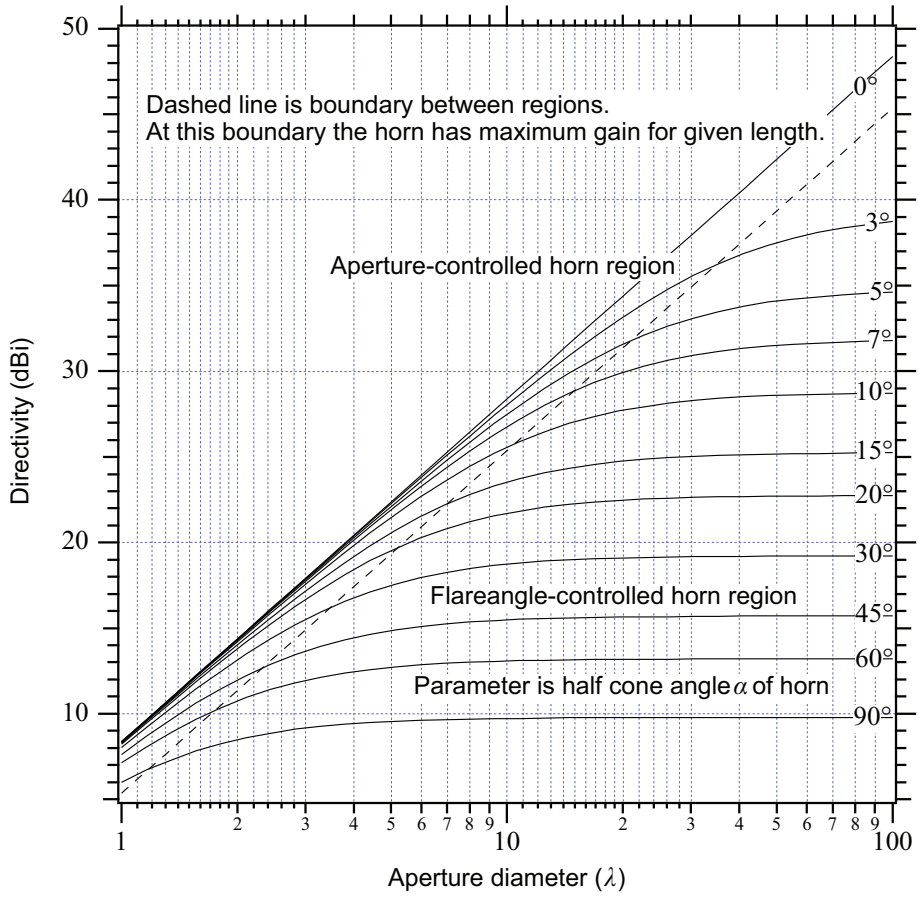


Figure 8.25: *Directivity of a corrugated horn as a function of the aperture diameter D for different flareangles α. The dashed line represents the boundary between the aperture-controlled (left) and flareangle-controlled (right) horns, at which the horn has its maximum gain for a given length.

soft horns than by corrugations, such as the strip-loaded horn [14] and the metamaterial horn [15]-[16]. The latter makes use of a metal-wired texture and is actually almost a hard horns. The boundary condition at the wall is the same in E- and H-plane, not entirely hard but with much better aperture efficiency than a soft horn. Some more soft and hard horn antennas are reviewed in [17].

Several studies have been made on hard horns with uniform aperture distribution. However, they cannot be realized with large sufficient bandwidth for practical applications. Still, if the application has two narrow bands a hard waveguide section may be used to improve smooth wall horn designs [18].

The quad-ridge horn is a very wideband coaxially-fed horn-type antenna. It was originally developed for use in test ranges and for EMC applications of almost decade bandwidth [19]. However, it is also possible to use them over a smaller but still very wide frequency band to feed reflectors [20].

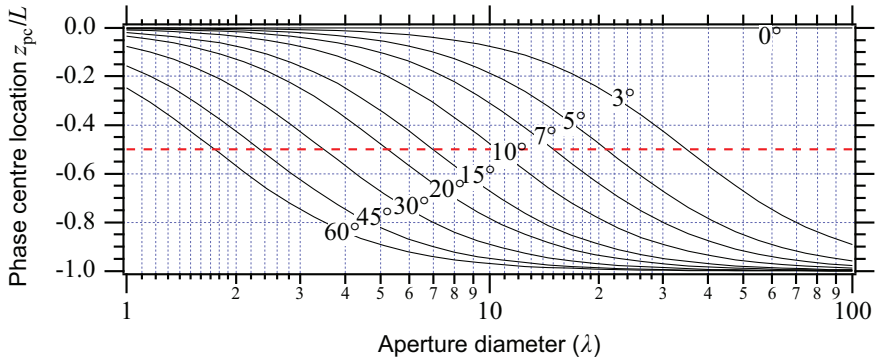


Figure 8.26: *Location of far-field phase center of corrugated horn. The dashed line represents the boundary between the aperture-controlled (upper) and flareangle-controlled (lower) horns.

8.11 Exercises to Chapter 8

1. **Pyramidal horn**: Consider a y-polarized large pyramidal horn antenna with a rectangular aperture.

   a) Use the theory for apertures in free space to find the far-field function of it, when we assume that the phase is constant on the aperture. Use the Huygens equivalent. What is the cross-polar radiation pattern?

   b) We excite the same horn for circular polarization in a way that there is no cross-polarization in the center of the horn aperture. What is the maximum relative cross-polar level in the aperture (i.e., relative to the co-polar maximum)? Use the Huygens equivalent to derive an expression for the radiation field. Find the expressions for the co- and cross-polar radiation patterns in the two principal planes when the aperture is quadratic. What is the relative cross-polar level in dB in these two planes for $\sin\theta = \lambda/(2a)$ where $2a$ is the aperture width?

2. **Conical corrugated horn**: Use the Gaussian beam formulas for horn antennas to design a conical corrugated horn with a 10 dB half-beamwidth of 15° at 12 GHz. The half-beamwidth is only allowed to vary by ±2 % over a ±5 % bandwidth. Try to keep the dimensions as small as possible.

3. **Dual-band corrugated horn**: We shall now use the Gaussian beam formulas for corrugated conical horn antennas in order to determine the *minimum size* a horn can have in order to work at both 8.5 GHz (X-band) and 2.3 GHz (S-band) according to some specifications to be given. The dual-band operation is possible by using dual-depth corrugations, but we will here only consider how to determine the major dimensions such as the diameter and length of the horn. The horn is intended to feed an offset paraboloid for nearly equal performance at S- and X-bands. This corresponds to requiring that the feed pattern has a taper of between 12.5 dB and 15 dB for $\theta = 45^\circ$ at both frequencies. Furthermore, the phase center locations at S-band and X-band must be closer than $\lambda_S/5$, where λ_S is the wavelength at S-band, in order to avoid a reduction in directivity due to defocussing of the paraboloidal reflector.

   a) Determine the 8.7 dB width of the Gaussian beam which provides 15 dB taper for $\theta = 45^\circ$.

   b) If we have no requirement on the size of the horn, we can obtain identical beamwidths and phase center positions at S- and X-bands for a certain choice of dimensions. What kind of horn is this, and what is the flare angle?

   c) Use the flare angle determined in step *b)*, and determine the minimum length of the horn which satisfies the requirements on the taper and phase center positions.

4. See the exercises about arrays of horn antennas in Chapter 10.

8.12 References

[1] P.-S. Kildal, "Artificially soft and hard surfaces in electromagnetics", *IEEE Transactions on Antennas and Propagation*, Vol. 38, pp. 1537-1544, October 1990.

[2] Z. Ying, P.-S. Kildal and A. Kishk, "Study of different realizations and calculation models for soft surfaces by using vertical monopole on soft disk as test bed", *IEEE Transactions on Antennas and Propagation*, Vol AP-44, No. 11, pp. 1474-1481, November 1996.

[3] P.D. Potter, "A new horn antenna with suppressed sidelobes and equal beamwidth", *Microwave J.*, Vol. VI, pp. 71-78, June 1963.

[4] A.J. Simmons and A.F. Kay, "The scalar feed - A high performance feed for large paraboloidal reflectors", in *Inst. Elec. Eng. Conf. Publ.* 21, June 1966, pp. 213-217.

[5] H.C. Minnett and B. MacA. Thomas, "A method of synthesizing radiation patterns with circular symmetry", *IEEE Transactions on Antennas and Propagation*, Vol. AP-14, pp. 654-656, September 1966.

[6] V.H. Rumsey, "Horn antennas with uniform power patterns around their axis", *IEEE Transactions on Antennas and Propagation*, Vol. AP-14, pp. 656-658, September 1966.

[7] B.M. Thomas, "Design of Corrugated Conical Horns," *IEEE Transactions on Antennas and Propagation*, Vol. AP-26, No. 2, March 1978.

[8] A.D. Olver and P.J.B. Clarricoats, A.A. Kishk, and L. Shafai, *Microwave Horns and Feeds*, IEEE Press/IEE, 1994.

[9] Z. Ying, A.A. Kishk, and P.-S. Kildal, "Broadband compact horn feed for prime-focus reflectors", *Electronics Letters*, Vol. 31, No. 14, pp. 1114-1115, July 1995.

[10] P.-S. Kildal, K. Jakobsen and K. Sudhakar Rao, "Meniscus lens-corrected corrugated horn: An efficient feed for a Cassegrain antenna", IEE Proc., MOA, Part H, No. 6, pp. 390-394, December 1984.

[11] P.-S. Kildal and K. Jakobsen, "Scalar horn with shaped lens improves Cassegrain efficiency", *IEEE Transactions on Antennas and Propagation*, Vol. 32, No. 10, pp. 1094-1100, October 1984.

[12] P.-S. Kildal, "A Gaussian beam model for aperture-controlled and flareangle- controlled corrugated horn antennas," *IEE Proceedings*, Part H, MOA, Vol. 135, No. 4, pp. 237-240, August 1988.

[13] E. Lier and P.-S. Kildal, "Soft and hard horn antennas", *IEEE Transactions on Antennas and Propagation*, Vol. 36, No. 8, pp. 1152-1157, August 1988.

[14] E. Lier and T. Schaug-Pettersen, "The strip-loaded hybrid-mode feed horn", *IEEE Transactions on Antennas and Propagation*, Vol. AP-35, pp. 1086-1089, September 1987.

[15] C.P. Scarborough, Q. Wu, D.H. Werner, E. Lier, Robert K. Shaw, and Bonnie G. Martin, "Demonstration of an Octave-Bandwidth Negligible-Loss Metamaterial Horn Antenna for Satellite Applications", *IEEE Transactions on Antennas and Propagation*, Vol. 61, No. 3, pp. 1081-1088, March 2013.

[16] E. Lier, D.H. Werner, C.P. Scarborough, Q. Wu, J.A. Bossard, "An octave-bandwidth negligible-loss radiofrequency metamaterial", *Nature Materials* 10, pp. 216-222, 2011.

[17] E. Lier, "Review of Soft and Hard Horn Antennas, Including Metamaterial-Based Hybrid-Mode Horns", *IEEE Antennas and Propagation Magazine*, Vol. 52, No. 2, April 2010.

[18] O. Sotoudeh, P.-S. Kildal, P. Ingvarson, and S.P. Skobelev, "Single- and dual- band multimode hard horn antennas with partly corrugated walls", *IEEE Transactions on Antennas and Propagation*, Vol. 54, No. 2, pt. 1, pp. 330-339, February 2006.

[19] V. Rodriguez, "An open-boundary quad-ridged guide horn antenna for use as a source in antenna pattern measurement anechoic chambers", *IEEE Antennas and Propagation Magazine*, Vol. 48, No. 2, pp. 157-169, 2006.

[20] A. Akgiray, S. Weinreb, W. A. Imbriale, C. Beaudoin, "Circular Quadruple-Ridged Flared Horn Achieving Near-Constant Beamwidth Over Multioctave Bandwidth: Design and Measurements", *IEEE Transactions on Antennas and Propagation*, Vol. 61, No. 3, pp. 1099-1108, March 2013.

Chapter 9

Reflector antennas

Reflector antennas are widely used in communication, radar and radio astronomy. Their sizes vary from small 0.3 m diameter reflectors for millimeter-wave communication links to large ground stations for satellite communications or radio telescopes with diameters of several tens of meters. The largest reflector antenna in the world is the radio telescope in Arecibo which has a spherical reflector with a diameter of 300 m. Reflector antennas have often a fixed main beam direction, but can also be steered by mechanical displacement of the feed or by rotation and tilting of the whole antenna.

Reflector antennas can have many different forms. Normally they consist of one or more reflectors which are designed to collimate an incident plane wave by reflection and transmission via each reflector to a focal point at a convenient location. The first reflector on reception is the largest and is called the main reflector. The next reflector is called the subreflector. If there are three or more reflectors, the last ones may either be called extra subreflectors, or feed reflectors. The feed reflectors may make up a beam waveguide. The feed antenna is located in the focal point. This is normally a horn antenna or a dipole above a ground plane. On transmit, the feed antenna illuminates the reflector system such that a desired field distribution is generated in the aperture plane in front of the main reflector. If the antenna has a pencil beam, the desired aperture distribution always has constant phase.

The reflectors may also be fed by an array of feeds, whereby we can get multiple simultaneous main beams (*multi-beam antenna*). The feed array is in this case often called a feed cluster. It is also possible to generate *contoured beams*, either by having one feed and shaping the reflector, or by using a feed array, or both. Some examples of different reflector antennas are illustrated in Fig. 9.1. The first two are primary-fed paraboloidal reflectors with an open waveguide feed and a self supported rear-radiating waveguide feed respectively. For comparison the left antenna in Fig. 1.1 shows a paraboloid with a self-supported dipole-disk feed. The third antenna in Fig. 9.1 is an offset parabolic reflector with a small conical horn feed. The fourth is a rotationally symmetric dual-reflector antenna with a corrugated horn feed. If the subreflector has convex shape as shown, it is called a Cassegrain type subreflector. The fifth antenna is an offset dual-reflector with a concave (Gregorian type) subreflector and a corrugated horn feed. The sixth antenna is a parabolic cylinder antenna fed by a linear array of crossed dipoles. The linear feed array is also called the line feed. The seventh antenna is the Arecibo spherical reflector antenna with a dual-reflector feed which can be

moved mechanically to steer the beam.

We will in Section 9.1 describe how a reflecting surface of any known shape can be analyzed by integration of a high frequency approximation of the induced currents on the surface, and by aperture integration using the formulas in Chapter 5. Thereafter, we will show how to use these methods to analyze a primary-fed paraboloid (Section 9.2) and a rotationally symmetric Cassegrain antenna (Section 9.3). In both these cases the antennas will be characterized in terms of different subefficiencies when the feed is a BOR_1 antenna (Section 9.4).

9.1 General reflector antenna theory

We will in this section introduce Geometrical Optics (GO), Physical Optics (PO) integration and aperture integration by describing how to analyze an arbitrary single metal reflector which is illuminated by an antenna with a known far-field function (See Fig. 9.2). The approaches are general and extendable to multi-reflector systems.

9.1.1 General description of reflector and feed

The location and orientation of the feed is defined by the location $\mathbf{r}_f$ of the origin of its coordinate system (i.e., phase reference point) and the directions $\hat{\mathbf{x}}_f$, $\hat{\mathbf{y}}_f$ and $\hat{\mathbf{z}}_f$ of its axes, all expressed in their components in a global coordinate system.

The reflector surface is defined by the position vector

$$\mathbf{r}_{_S}(u,v) = x_{_S}(u,v)\hat{\mathbf{x}} + y_{_S}(u,v)\hat{\mathbf{y}} + z_{_S}(u,v)\hat{\mathbf{z}} \ , \tag{9.1}$$

when u and v vary between their boundaries, that define the reflector rim. We choose the reflector coordinate system to coincide with the global coordinate system. A rotationally symmetric reflector is defined by

$$\mathbf{r}_{_S}(\rho,\varphi) = \rho\cos\varphi\hat{\mathbf{x}} + \rho\sin\varphi\hat{\mathbf{y}} + z(\rho)\hat{\mathbf{z}} \ , \tag{9.2}$$

for $0 < \rho < D/2$, where D is the reflector diameter.

We will in the analysis to follow need the *surface normal* $\hat{\mathbf{n}}$. This can be calculated from the two tangent vectors

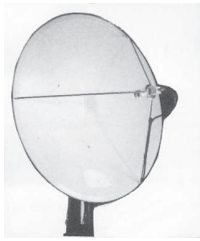
$$\mathbf{t}_u = \frac{\partial}{\partial u}\mathbf{r}_{_S}(u,v) = \hat{\mathbf{x}}\frac{\partial}{\partial u}(x_{_S}(u,v)) + \hat{\mathbf{y}}\frac{\partial}{\partial u}(y_{_S}(u,v)) + \hat{\mathbf{z}}\frac{\partial}{\partial u}(z_{_S}(u,v)) \ ,$$

$$\mathbf{t}_v = \frac{\partial}{\partial v}\mathbf{r}_{_S}(u,v) = \hat{\mathbf{x}}\frac{\partial}{\partial v}(x_{_S}(u,v)) + \hat{\mathbf{y}}\frac{\partial}{\partial v}(y_{_S}(u,v)) + \hat{\mathbf{z}}\frac{\partial}{\partial v}(z_{_S}(u,v)) \ ,$$

by using

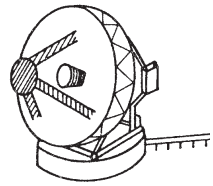
$$\hat{\mathbf{n}} = \mathbf{t}_u \times \mathbf{t}_v / |\mathbf{t}_u \times \mathbf{t}_v| \ . \tag{9.3}$$

We must check that $\hat{\mathbf{n}}$ is directed towards the lit side of the reflector, and if it is not we reverse its sign.



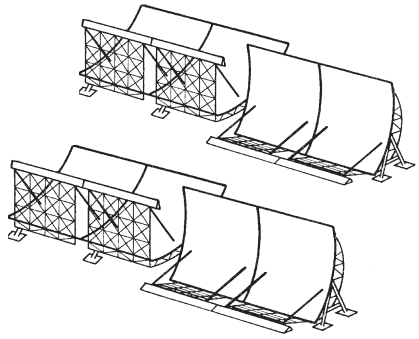
Primary-fed paraboloids.
Feed supported by struts (left) and self supported (right)

Offset primary-fed
parabolic reflector

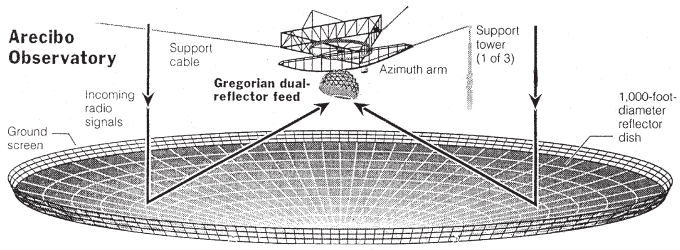


Symmetrical Cassegrain

Offset Gregorian
dual-reflector antennas



EISCAT VHF parabolic cylinder antenna



Arecibo spherical reflector with dual-reflector feed (inside enclosure)

Figure 9.1: Examples of reflector antennas.

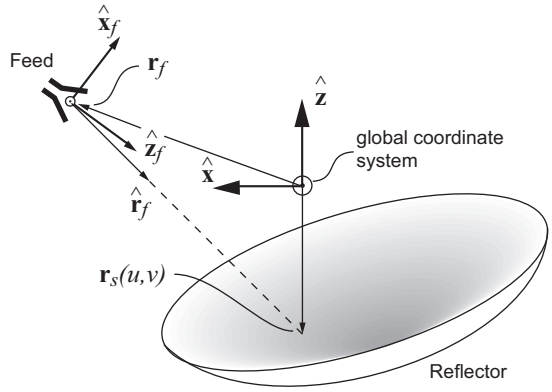


Figure 9.2: Feed and arbitrary reflector with their coordinate systems.

9.1.2 Incident field on reflector

We assume here that the reflector is in the far-field of the feed. Then, the radiation field of the feed at an arbitrary observation point $\mathbf{r}$ can be expressed as (see Eq. (2.39) in Section 2.3.3)

$$\mathbf{E}_f(\mathbf{r}) = \frac{1}{r_f} e^{-jkr_f} \mathbf{G}_f(\hat{\mathbf{r}}_f) \ , \tag{9.4}$$

where $\mathbf{G}_f(\hat{\mathbf{r}}_f)$ is the far-field function and $\hat{\mathbf{r}}_f = (\mathbf{r} - \mathbf{r}_f)/r_f$ with $r_f = |\mathbf{r} - \mathbf{r}_f|$ and $\mathbf{r}_f$ the phase reference point. The incident field on the reflector is

$$\mathbf{E}_i(\mathbf{r}_{_S}) = \mathbf{E}_f(\mathbf{r}_{_S}) \ , \tag{9.5}$$

where $\mathbf{E}_f(\mathbf{r}_{_S})$ is given by (9.4) with $r_f = |\mathbf{r}_{_S}(u,v) - \mathbf{r}_f|$ and $\hat{\mathbf{r}}_f = (\mathbf{r}_{_S}(u,v) - \mathbf{r}_f)/r_f$. The incident H-field on the reflector is

$$\mathbf{H}_i = \frac{1}{\eta} \hat{\mathbf{r}}_f \times \mathbf{E}_i \ . \tag{9.6}$$

9.1.3 Reflected GO field

Geometrical Optics (GO) is an approximate method to determine electromagnetic fields. It is asymptotically correct for high frequencies. In GO, all fields in free space propagate geometrically along straight lines, referred to as rays, and these rays are reflected at material interfaces by the classical reflection law. In order to analyze a multi-reflector system with GO we need to consider both reflection and transmission of GO fields. We will here only present the equation needed to find the propagation direction of the reflected field and its complex amplitude and direction at the reflected point. This is sufficient in order to analyze single-reflector systems.

The incident field on the reflector, as defined in (9.4), is a spherical wave. A field incident at any specific point $\mathbf{r}_s(u_r, v_r)$ can be interpreted as a field propagating along a GO ray defined by (see Fig. 9.3)

$$\mathbf{r}(s) = \mathbf{r}_f + s\hat{\mathbf{s}}_i \ ; \qquad \hat{\mathbf{s}}_i = \hat{\mathbf{r}}_f \ , \tag{9.7}$$

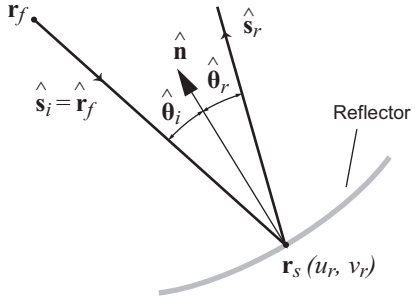


Figure 9.3: Illustration of GO reflection.

where $\hat{\mathbf{s}}_i$ is the direction of the ray and s is a parameter measuring distance along the ray and being zero at the start point $\mathbf{r}_f$ of the ray. The ray is reflected at $\mathbf{r}_s(u_r, v_r)$, and the direction of the reflected ray can be found by using the reflection law, which in vector notation becomes

$$\hat{\mathbf{s}}_r = \hat{\mathbf{s}}_i - 2(\hat{\mathbf{n}} \cdot \hat{\mathbf{s}}_i)\hat{\mathbf{n}} \ . \tag{9.8}$$

This corresponds to $\hat{\mathbf{n}} \cdot \hat{\mathbf{s}}_r = -\hat{\mathbf{n}} \cdot \hat{\mathbf{s}}_i$ which means that $\cos\theta_r = \cos\theta_i$ and $\theta_r = \theta_i$, i.e., the angle of reflection is equal to the angle of incidence. The operation in (9.8) can be expressed verbally as changing the sign of the $\hat{\mathbf{n}}$ component of $\hat{\mathbf{s}}_i$, because we subtract the $\hat{\mathbf{n}}$ component twice.

The reflected E-field $\mathbf{E}_r$ must be orthogonal to $\hat{\mathbf{s}}_r$ in order to be a GO field, and, the total E-field $\mathbf{E}_{\text{tot}} = \mathbf{E}_i + \mathbf{E}_r$ at $\mathbf{r}_s(u_r, v_r)$ must satisfy $\mathbf{E}_{\text{tot}} \times \hat{\mathbf{n}} = 0$, i.e., the tangential E-field being zero, if the reflector is a good conductor. These two conditions are satisfied when

$$\mathbf{E}_r = -\mathbf{E}_i + 2(\mathbf{E}_i \cdot \hat{\mathbf{n}})\hat{\mathbf{n}} \ . \tag{9.9}$$

The corresponding reflected H-field becomes

$$\mathbf{H}_r = \mathbf{H}_i - 2(\mathbf{H}_i \cdot \hat{\mathbf{n}})\hat{\mathbf{n}} = \frac{1}{\eta}\hat{\mathbf{s}}_r \times \mathbf{E}_r \ . \tag{9.10}$$

The reflected field propagates along the reflected ray

$$\mathbf{r}(s) = \mathbf{r}_s(u_r, v_r) + s\hat{\mathbf{s}}_r \ . \tag{9.11}$$

The equation for propagating the reflected GO field along this ray is given by a general GO transmission formula, which is more complicated than Eq. (9.4). It depends on the two principal curvatures of the reflector surface at $\mathbf{r}_s(u_r, v_r)$. The equation will not be introduced here. We will only treat the case for which the $\hat{\mathbf{s}}_r$ of all reflected rays are parallel, i.e.,

$$\hat{\mathbf{s}}_r = \hat{\mathbf{z}} \tag{9.12}$$

in all reflection points. Then, the GO transmission gives no variation of the field amplitude along the ray, rather only a phase progression according to

$$\mathbf{E}_r(\mathbf{r}_s + s\hat{\mathbf{s}}_r) = \mathbf{E}_r(\mathbf{r}_s)e^{-jks} \ . \tag{9.13}$$

The E- and H-fields at a point $\mathbf{r}(s)$ on a GO ray in free space are always orthogonal to the direction $\hat{\mathbf{s}}$ of the ray, and related by

$$\mathbf{H}(\mathbf{r}(s)) = \frac{1}{\eta}\hat{\mathbf{s}} \times \mathbf{E}(\mathbf{r}(s)) \ . \tag{9.14}$$

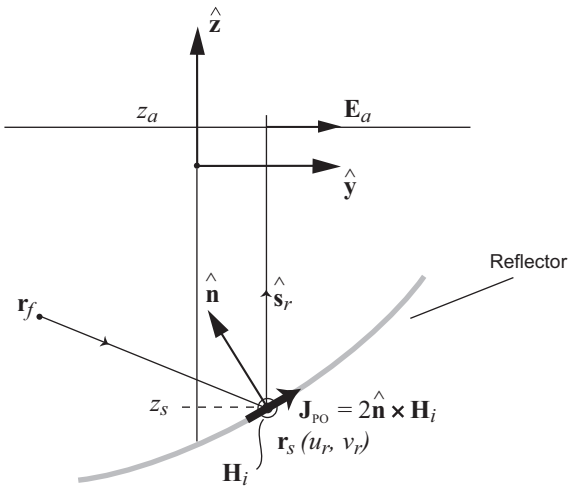


Figure 9.4: Illustration of PO currents and aperture field.

9.1.4 PO integration

In order to determine the radiation (scattering) from the reflector we must find the induced electric currents $\mathbf{J}_s(\mathbf{r}_s)$ on it (Fig. 9.4). This is generally very laborious, but we may approximate it by using GO as follows. If the scattered field at $\mathbf{r}_s$ is approximated by the GO reflected field. By using (9.10) the total H-field becomes

$$\mathbf{H}_{\text{tot}}(\mathbf{r}_S) = \mathbf{H}_i + \mathbf{H}_r = 2\mathbf{H}_i - 2(\mathbf{H}_i \cdot \hat{\mathbf{n}})\hat{\mathbf{n}} \ . \tag{9.15}$$

The induced electric currents at a PEC reflector are from equation (4.14) given as $\mathbf{J} = \hat{\mathbf{n}} \times \mathbf{H}$. Using this with $\mathbf{H}_{\text{tot}}$ approximated by (9.15) we get

$$\mathbf{J}_{\text{PO}}(\mathbf{r}_s) = 2\hat{\mathbf{n}} \times \mathbf{H}_i \ . \tag{9.16}$$

This is commonly referred to as the *Physical Optics* (PO) approximation for the current distribution. We see that the induced PO currents are found from the surface normal vector and the incident H-field.

The radiation field of the total antenna, including both the feed ($\mathbf{E}_f(\mathbf{r})$) and the scattered field $\mathbf{E}_s(\mathbf{r})$ from the reflector, in a point $\mathbf{r}$ in the far-field of the reflector, is now ($r = |\mathbf{r}|$)

$$\mathbf{E}_t(\mathbf{r}) = \mathbf{E}_f(\mathbf{r}) + \mathbf{E}_s(\mathbf{r}) = \frac{1}{r} e^{-jkr} \mathbf{G}_t(\hat{\mathbf{r}}) \ , \tag{9.17}$$

where the far-field function of the total antenna is

$$\mathbf{G}_t(\hat{\mathbf{r}}) = \mathbf{G}_f(\hat{\mathbf{r}}) e^{jk(\mathbf{r}_f \cdot \hat{\mathbf{r}})} + \mathbf{G}_{\text{PO}}(\hat{\mathbf{r}}) \ . \tag{9.18}$$

Here, the phase factor following $\mathbf{G}_f(\hat{\mathbf{r}})$ is found by using (2.52)[1] with $\mathbf{r}_f = \mathbf{r}_A$, and where the contribution from the PO currents on the reflector is found by using (4.49) and (4.51)[2];

[1] See Section 2.3.4.
[2] See Section 4.2.2.

$$\mathbf{G}_{\text{PO}}(\hat{\mathbf{r}}) = \mathbf{I}_J - (\mathbf{I}_J \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} \; ; \qquad \mathbf{I}_J = C_k \iint\limits_{S'} \eta \mathbf{J}_{\text{PO}}(\mathbf{r}_s) e^{jk(\mathbf{r}_s \cdot \hat{\mathbf{r}})} dS \; , \tag{9.19}$$

with $C_k = -jk/(4\pi)$ as before. The phase reference point of $\mathbf{G}_t(\hat{\mathbf{r}})$ is the origin of the global coordinate system. The total field is always found by summing up the contributions from both actual and induced sources, i.e., from the feed (giving $\mathbf{G}_f(\hat{\mathbf{r}})$) and the induced currents on the reflector (giving $\mathbf{G}_{\text{PO}}(\hat{\mathbf{r}})$). $\mathbf{I}_J$ is the PO far-field integral. The PO integration method is very accurate for large reflector antennas, i.e., when the diameter $D \gg \lambda$.

9.1.5 Aperture integration

Let us assume that all the rays reflected from the reflector are parallel with $\hat{\mathbf{z}}$. Then, we may define an aperture plane with normal $\hat{\mathbf{z}}$ in front of the reflector, and the aperture field can be found by using the GO formula in (9.13) as (Fig. 9.4)

$$\mathbf{E}_a = \mathbf{E}_r(\mathbf{r}_s) e^{-jk(z_a - z_s)} \; , \tag{9.20}$$

where z_a is the location of the aperture plane and $z_s = z_s(u_r, v_r)$ is the z-coordinate of the reflection point $\mathbf{r}_s = \mathbf{r}_s(u_r, v_r)$ at the reflector surface. The H-field in the aperture is

$$\mathbf{H}_a = \frac{1}{\eta} \hat{\mathbf{z}} \times \mathbf{E}_a \; . \tag{9.21}$$

We have now a plane aperture in free space, and we can calculate the radiation field from it by using the Huygens equivalent. This gives for the y-polarized component of the aperture field (see Eq. (7.20))

$$\mathbf{G}_a(\hat{\mathbf{r}}) = -C_k[(\hat{\mathbf{y}} - (\hat{\mathbf{y}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}) - (\hat{\mathbf{y}} \times \hat{\mathbf{z}}) \times \hat{\mathbf{r}}] \tilde{E}_{a_y}(k(\hat{\mathbf{x}} \cdot \hat{\mathbf{r}}), k(\hat{\mathbf{y}} \cdot \hat{\mathbf{r}})) e^{jkz_a} \; , \tag{9.22}$$

where C_k is as before and

$$\tilde{E}_{a_y}(k_x, k_y) = \iint\limits_A E_{a_y}(x', y') e^{j(k_x x' + k_y y')} dx' dy' \; , \tag{9.23}$$

with $E_{a_y}(x', y') = \mathbf{E}_a \cdot \hat{\mathbf{y}}$, $k_x = k(\hat{\mathbf{x}} \cdot \hat{\mathbf{r}})$ and $k_y = k(\hat{\mathbf{y}} \cdot \hat{\mathbf{r}})$. The x-polarized component is calculated correspondingly. The approach above can only be used when all the rays reflected from the reflector are parallel, i.e., $\hat{\mathbf{s}}_r = \hat{\mathbf{z}}$. This is approximately the case in most reflector antennas. If $\hat{\mathbf{s}}_r \approx \hat{\mathbf{z}}$, we may use the paraxial approximation $\hat{\mathbf{s}}_r = \hat{\mathbf{z}}$, and still calculate the aperture field in the way shown above. This is a very good approximation when the observation direction $\hat{\mathbf{r}} = \hat{\mathbf{z}}$, since then the phase of the integrand becomes equal to the phase of the integrand of the PO integral. This is evident from (9.24) in the following subsection. The approximation is acceptable because the accuracy of the calculation of the far-field function is more sensitive to the correct phase in the aperture than the correct amplitude.

9.1.6 Aperture integration by projection of the PO integral

Aperture integration is convenient compared to *PO integration* because the aperture integral has the form of a two-dimensional Fourier transform. However, the aperture fields can only

be defined well when the rays reflected by the main reflector are parallel so that (9.20) can be applied. This is rarely exactly true, and it is never true when the feed is not located at the focal point of the reflector. Therefore, we shall here introduce a more general aperture field distribution, obtained by projection of the PO integral onto an arbitrarily chosen aperture plane (Fig. 9.5). Let us write the PO far-field function in (9.19) as

$$\mathbf{G}_{\text{PO}}(\mathbf{r}) = C_k \iint_S [\eta \mathbf{J}_{\text{PO}}(\mathbf{r}_s) - (\eta \mathbf{J}_{\text{PO}}(\mathbf{r}_s) \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}] e^{jk(\mathbf{r}_s \cdot \hat{\mathbf{r}})} dS \ . \tag{9.24}$$

Let us now define an aperture in the $x'y'$-plane normal to the $\hat{\mathbf{z}}'$ axis of an arbitrary primed coordinate system[3]. We introduce a phase and amplitude projection $\mathbf{J}_a$ of $\mathbf{J}_{\text{PO}}$ in the aperture plane A by

$$\mathbf{J}_a dA = (\mathbf{J}_{\text{PO}} - (\mathbf{J}_{\text{PO}} \cdot \hat{\mathbf{z}}')\hat{\mathbf{z}}') e^{-jk(z'_a - z'_s)} dS \ , \tag{9.25}$$

where z'_a and z'_s are the z'-locations of the aperture and the reflection points, respectively, in the primed coordinate system. Moreover,

$$dA = (\hat{\mathbf{n}} \cdot \hat{\mathbf{z}}')dS \tag{9.26}$$

is the projection of the infinitesimal surface element in Fig. 9.5 on to the primed aperture plane. Then, the integrand in (9.24) can be rearranged as follows

$$\begin{aligned}
\mathbf{I}_{\text{PO}} &= (\mathbf{J}_{\text{PO}} - (\mathbf{J}_{\text{PO}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}) e^{jk(\mathbf{r}_s \cdot \hat{\mathbf{r}})} dS \\
&= \{\mathbf{J}_{\text{PO}} - (\mathbf{J}_{\text{PO}} \cdot \hat{\mathbf{z}}')\hat{\mathbf{z}}' + (\mathbf{J}_{\text{PO}} \cdot \hat{\mathbf{z}}')\hat{\mathbf{z}}' - (\mathbf{J}_{\text{PO}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}\} e^{jk(\mathbf{r}_s \cdot \hat{\mathbf{r}})} \\
&= \left\{ \mathbf{J}_a e^{jk(z'_a - z'_s)} dA + (\mathbf{J}_{\text{PO}} \cdot \hat{\mathbf{z}}')\hat{\mathbf{z}}' - (\mathbf{J}_{\text{PO}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} dS \right\} e^{jk(\mathbf{r}_s \cdot \hat{\mathbf{r}})} \ .
\end{aligned}$$

By using $(\mathbf{J}_a \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} e^{jk(z'_a - z'_s)} dA = (\mathbf{J}_{\text{PO}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} dS - (\hat{\mathbf{J}_{\text{PO}}} \cdot \hat{\mathbf{z}}')(\hat{\mathbf{z}}' \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} dS$, we arrive at

$$\begin{aligned}
\mathbf{I}_{\text{PO}} &= e^{jk(\mathbf{r}_s \cdot \hat{\mathbf{r}})} \cdot \left\{ [\mathbf{J}_a - (\mathbf{J}_a \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}] e^{jk(z'_a - z'_s)} dA - (\mathbf{J}_{\text{PO}} \cdot \hat{\mathbf{z}}')(\hat{\mathbf{z}}' \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} + (\mathbf{J}_{\text{PO}} \cdot \hat{\mathbf{z}}')\hat{\mathbf{z}}' \right\} dS \\
&\approx [\mathbf{J}_a - (\mathbf{J}_a \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}] e^{jk(x'_s \hat{\mathbf{x}}' + y'_s \hat{\mathbf{y}}') \cdot \hat{\mathbf{r}}} e^{jkz'_a} dA \ .
\end{aligned}$$

The last approximation is obtained by using

$$(\mathbf{J}_{\text{PO}} \cdot \hat{\mathbf{z}}')\hat{\mathbf{z}}' - (\mathbf{J}_{\text{PO}} \cdot \hat{\mathbf{z}}')(\hat{\mathbf{z}}' \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} \approx 0 \ , \qquad \text{and}$$

$$\begin{aligned}
\mathbf{r}_s \cdot \hat{\mathbf{r}} + (z'_a - z'_s) &= (x'_s \hat{\mathbf{x}}' + y'_s \hat{\mathbf{y}}') \cdot \hat{\mathbf{r}} + z'_s((\hat{\mathbf{z}}' \cdot \hat{\mathbf{r}}) - 1) + z'_a \\
&\approx (x'_s \hat{\mathbf{x}}' + y'_s \hat{\mathbf{y}}') \cdot \hat{\mathbf{r}} + z'_a \ ,
\end{aligned}$$

when $|\hat{\mathbf{r}} - \hat{\mathbf{z}}'| \ll 1$.

In this way the PO integral of the far-field function has been approximated by an aperture integral which has the form of a Fourier transform, i.e.,

$$\mathbf{G}_A = C_k(\eta \tilde{\mathbf{J}}_a - (\eta \tilde{\mathbf{J}}_a \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}) e^{jkz'_a} \ ; \qquad \tilde{\mathbf{J}}_a = \iint_A \mathbf{J}_a e^{jk(x'_s \hat{\mathbf{x}} + y'_s \hat{\mathbf{y}}) \cdot \hat{\mathbf{r}}} dx' dy' \ . \tag{9.27}$$

The results $\mathbf{G}_{\text{PO}}(\hat{\mathbf{r}})$ (obtained by PO) and $\mathbf{G}_A(\hat{\mathbf{r}})$ (obtained by aperture integration) from the above are identical for $\hat{\mathbf{r}} = \hat{\mathbf{z}}'$, i.e., when we observe normal to the aperture plane. Therefore,

[3] The aperture in Section 9.1.5 was defined by $\hat{\mathbf{z}}' = \hat{\mathbf{z}}$ and by the location $\mathbf{r}_0 = z_a \hat{\mathbf{z}}$ of its origin in the global coordinate system.

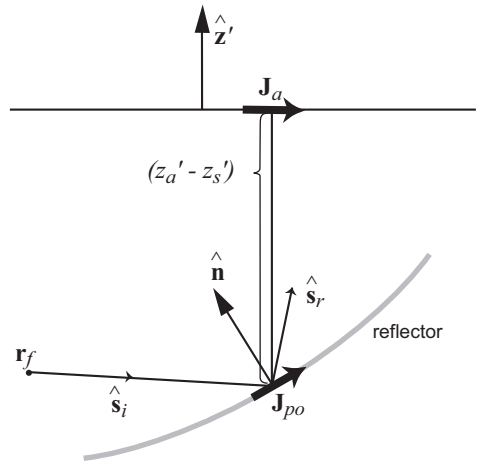


Figure 9.5: Projection of the PO currents into an aperture.

$\mathbf{G}_A(\hat{\mathbf{r}})$ can be used as an acceptable approximation for $\mathbf{G}_{\rm PO}(\hat{\mathbf{r}})$ for small observation angles θ', i.e., when $|\mathbf{r} - \hat{\mathbf{z}}'| \ll 1$.

The aperture integral $\mathbf{G}_A(\hat{\mathbf{r}})$ obtained by projection of the PO integral can be used for all choices of directions $\hat{\mathbf{z}}'$ even if the reflected GO fields from the reflector are not parallel with the $\hat{\mathbf{z}}'$ axis. However, the accuracy degrades more rapidly away from the $\hat{\mathbf{z}}'$ direction (i.e., with increasing $|\hat{\mathbf{r}} - \hat{\mathbf{z}}'|$) if the direction of the reflected rays deviate considerably from $\hat{\mathbf{z}}'$ (i.e., if $|\hat{\mathbf{s}}_r - \hat{\mathbf{z}}'|$ is large). When $\hat{\mathbf{r}} = \hat{\mathbf{s}}_r = \hat{\mathbf{z}}'$, $\mathbf{G}_A(\hat{\mathbf{r}})$ becomes identical to the conventional aperture integral $\mathbf{G}_a(\hat{\mathbf{r}})$ of the GO aperture fields presented in (9.22), except for the incremental element factor which for $\mathbf{G}_a(\hat{\mathbf{r}})$ is that of a Huygens source and for $\mathbf{G}_A(\hat{\mathbf{r}})$ is an electric current.

9.2 The paraboloidal reflector

We now apply the formulations in Section 9.1 to a rotationally symmetric *paraboloidal reflector* (Fig. 9.6), and present the formulas for this case. The paraboloid with *focal point* in the origin is defined by

$$\mathbf{r}_s(\rho,\varphi) = \rho\hat{\boldsymbol{\rho}} + z(\rho)\hat{\mathbf{z}} \; ; \qquad \hat{\boldsymbol{\rho}} = \cos\varphi\hat{\mathbf{x}} + \sin\varphi\hat{\mathbf{y}} \; , \tag{9.28}$$

$$\text{with} \qquad z(\rho) = -F + \rho^2/(4F) \qquad \text{for } 0 \le \rho \le D/2 \; , \tag{9.29}$$

where F is the focal length and D is the diameter of the reflector. An alternative parametric description giving identical shape is

$$\mathbf{r}_s(\theta_f,\varphi) = r(\theta_f)\hat{\mathbf{r}} \; ; \qquad \hat{\mathbf{r}} = \sin\theta_f\hat{\boldsymbol{\rho}} - \cos\theta_f\hat{\mathbf{z}} \; , \tag{9.30}$$

$$\text{with} \qquad r(\theta_f) = F/\cos^2(\theta_f/2) = 2F/(1+\cos\theta_f) \qquad \text{for } 0 < \theta_f < \theta_0 \; , \tag{9.31}$$

where θ_0 is the subtended half-angle of the paraboloid. The polar-angle θ_f of the feed is measured relative to the negative z-axis of the global coordinate system. The primary-fed paraboloid is normally deep with $70^\circ < \theta_0 < 90^\circ$.

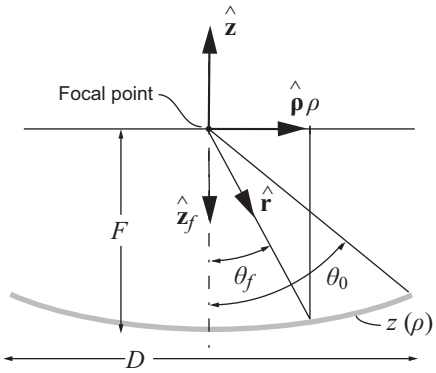


Figure 9.6: Geometry of rationally symmetric paraboloid.

Two independent parameters are needed in order to specify the paraboloid. We can use any two of the three parameters F, D and θ_0 (or any two combinations of them). A common choice is D and F/D, but we will here prefer D and θ_0 since θ_0 gives a more direct measure of the required beamwidth of the feed pattern than the F/D. The following formulas relate ρ and D to θ_f and θ_0:

$$\rho = r(\theta_f)\sin\theta_f = F\sin\theta_f/\cos^2(\theta_f/2) = 2F\tan(\theta_f/2)\ , \tag{9.32}$$

$$d\rho/d\theta_f = F/\cos^2(\theta_f/2) = r(\theta_f)\ , \tag{9.33}$$

$$D = 4F\tan(\theta_0/2)\ . \tag{9.34}$$

9.2.1 Surface normal, incident and reflected ray

We calculate the surface normal vector as explained in Subsection 9.1.1. This gives

$$\mathbf{t}_\varphi = \frac{\partial}{\partial\varphi}\mathbf{r}_s(\theta_f,\varphi) = r(\theta_f)\frac{\partial}{\partial\varphi}\hat{\mathbf{r}} = r(\theta_f)\sin\theta_f\frac{\partial}{\partial\varphi}\hat{\boldsymbol{\rho}} = r(\theta_f)\sin\theta_f\hat{\boldsymbol{\varphi}}\ , \tag{9.35}$$

$$\begin{aligned}\mathbf{t}_{\theta_f} &= \frac{\partial}{\partial\theta_f}\mathbf{r}_s(\theta_f,\varphi) = \hat{\mathbf{r}}\frac{\partial}{\partial\theta_f}r(\theta_f) + r(\theta_f)\frac{\partial}{\partial\theta_f}\hat{\mathbf{r}}\\ &= F\cos^{-3}(\theta_f/2)\sin(\theta_f/2)\hat{\mathbf{r}} + r(\theta_f)\hat{\boldsymbol{\theta}}_f\\ &= F\cos^{-3}(\theta_f/2)[\sin(\theta_f/2)\hat{\mathbf{r}} + \cos(\theta_f/2)\hat{\boldsymbol{\theta}}_f]\ .\end{aligned} \tag{9.36}$$

If we now use $\hat{\boldsymbol{\varphi}}\times\hat{\mathbf{r}} = -\hat{\boldsymbol{\theta}}_f$ and $\hat{\boldsymbol{\varphi}}\times\hat{\boldsymbol{\theta}}_f = \hat{\mathbf{r}}$ we obtain

$$\hat{\mathbf{n}}' = (\mathbf{t}_\varphi\times\mathbf{t}_{\theta_f})/|\mathbf{t}_\varphi\times\mathbf{t}_{\theta_f}| = -\sin(\theta_f/2)\hat{\boldsymbol{\theta}}_f + \cos(\theta_f//2)\hat{\mathbf{r}}\ . \tag{9.37}$$

This $\hat{\mathbf{n}}'$ is directed into the shadow side, so we invert it to achieve

$$\hat{\mathbf{n}} = \sin(\theta_f/2)\hat{\boldsymbol{\theta}}_f - \cos(\theta_f/2)\hat{\mathbf{r}}\ . \tag{9.38}$$

We locate the feed with its phase reference point in the focal point. Then, the direction of the incident ray is

$$\hat{\mathbf{s}}_i = \hat{\mathbf{r}}\ . \tag{9.39}$$

The direction of the reflected ray becomes

$$\hat{\mathbf{s}}_r = \hat{\mathbf{s}}_i - 2(\hat{\mathbf{n}} \cdot \hat{\mathbf{s}}_i)\hat{\mathbf{n}} = \hat{\mathbf{z}} \ , \tag{9.40}$$

by using $\hat{\mathbf{n}} \cdot \hat{\mathbf{s}}_i = -\cos(\theta_f/2)$ and

$$\hat{\mathbf{r}} + 2\cos(\theta_f/2)\hat{\mathbf{n}} = \hat{\mathbf{r}} + \sin\theta_f \hat{\boldsymbol{\theta}}_f - 2\cos^2(\theta_f/2)\hat{\mathbf{r}} = -\cos\theta_f \hat{\mathbf{r}} + \sin\theta_f \hat{\boldsymbol{\theta}}_f = \hat{\mathbf{z}} \ .$$

Thus, all the rays reflected from the paraboloid are parallel with the z-axis when the feed is located with its phase reference point in the focus.

9.2.2 Aperture field

If the feed is of BOR_1 type and polarized in $\hat{\mathbf{y}}_f$-direction, its far-field function is

$$\mathbf{G}_f(\mathbf{r}) = \mathbf{G}_f(\theta_f, \varphi_f) = G_{_E}(\theta_f)\sin\varphi_f \hat{\boldsymbol{\theta}}_f + G_{_H}(\theta_f)\cos\varphi_f \hat{\boldsymbol{\varphi}}_f \ , \tag{9.41}$$

where the index refers to the coordinate system of the feed, and $G_{_E}(\theta_f)$ and $G_{_H}(\theta_f)$ are the E- and H-plane far-field functions, respectively. θ_f, φ_f, $\hat{\boldsymbol{\theta}}_f$ and $\hat{\boldsymbol{\varphi}}_f$ can be calculated from $\mathbf{r}$, $\hat{\mathbf{x}}_f$, $\hat{\mathbf{y}}_f$ and $\hat{\mathbf{z}}_f$ by standard formulas. We locate the feed such that its phase reference point coincides with the focal point of the paraboloid. Then, the incident field on the reflector is found from (9.4), (9.5) and (9.31) to be

$$\begin{aligned} \mathbf{E}_i &= \frac{1}{r(\theta_f)} e^{-jkr(\theta_f)} [G_{_E}(\theta_f)\sin\varphi_f \hat{\boldsymbol{\theta}}_f + G_{_H}(\theta_f)\cos\varphi_f \hat{\boldsymbol{\varphi}}_f] \\ &= \frac{1}{r(\theta_f)} e^{-jkr(\theta_f)} [G_{_E}(\theta_f)\sin\varphi \hat{\boldsymbol{\theta}}_f - G_{_H}(\theta_f)\cos\varphi \hat{\boldsymbol{\varphi}}] \ , \end{aligned} \tag{9.42}$$

where in the last line, by using that $\hat{\mathbf{x}}_f = -\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}_f = \hat{\mathbf{y}}$, so that $\varphi_f = \pi - \varphi$ and $\hat{\boldsymbol{\varphi}}_f = -\hat{\boldsymbol{\varphi}}$, we have introduced the azimuth angle φ of the global coordinate system . For $\theta_f < \theta_0$, the reflected field becomes

$$\mathbf{E}_r = -\mathbf{E}_i + 2(\mathbf{E}_i \cdot \hat{\mathbf{n}})\hat{\mathbf{n}} = -\frac{1}{r(\theta_f)} e^{-jkr(\theta_f)} [G_{_E}(\theta_f)\sin\varphi \hat{\boldsymbol{\rho}} + G_{_H}(\theta_f)\cos\varphi \hat{\boldsymbol{\varphi}}] \ , \tag{9.43}$$

because (see Fig. 9.7)

$$\hat{\boldsymbol{\theta}}_f - 2(\hat{\boldsymbol{\theta}}_f \cdot \hat{\mathbf{n}})\hat{\mathbf{n}} = \hat{\boldsymbol{\rho}} \ , \qquad \hat{\boldsymbol{\varphi}} - 2(\hat{\boldsymbol{\varphi}} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}} = \hat{\boldsymbol{\varphi}} \ . \tag{9.44}$$

Finally, the aperture field at $z_a = 0$ becomes

$$\mathbf{E}_a(\rho, \varphi) = E_{_E}(\rho)\sin\varphi \hat{\boldsymbol{\rho}} + E_{_H}(\rho)\cos\varphi \hat{\boldsymbol{\varphi}} \tag{9.45}$$

for $\rho < D/2$, with

$$\left.\begin{matrix} E_{_E}(\rho) \\ E_{_H}(\rho) \end{matrix}\right\} = -\frac{1}{F}\cos^2(\theta_f/2) \begin{pmatrix} G_{_E}(\theta_f) \\ G_{_H}(\theta_f) \end{pmatrix} e^{-j2kF} \ , \tag{9.46}$$

where the relation between ρ and θ_f is given in (9.32). The phase term $-2kF$ is a result of the total path length

$$r(\theta_f) + r(\theta_f)\cos\theta_f = kr(\theta_f)(1 + \cos\theta_f) = 2kF \ . \tag{9.47}$$

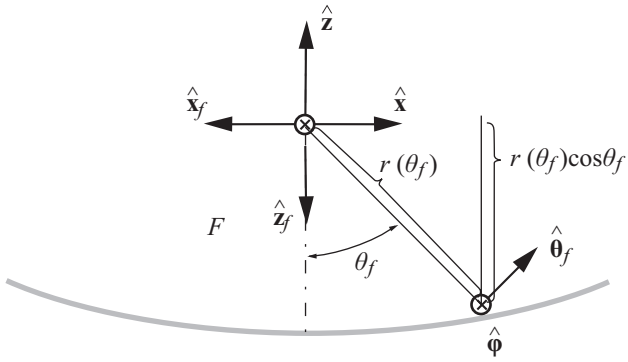


Figure 9.7: Reflection at paraboloidal surface.

This is readily seen by using (9.31). The feed can also be described by its co- and cross-polar patterns $G_{\mathrm{co}_{45^\circ}}(\theta_f)$ and $G_{\mathrm{xp}_{45^\circ}}(\theta_f)$ in the 45°-plane[4]. Using these the aperture fields become[5]

$$\mathbf{E}_a(\rho,\varphi) = [E_{\mathrm{co}_{45^\circ}}(\rho) - E_{\mathrm{xp}_{45^\circ}}(\rho)\cos 2\varphi]\hat{\mathbf{y}} - E_{\mathrm{xp}_{45^\circ}}(\rho)\sin 2\varphi\hat{\mathbf{x}} \tag{9.48}$$

for $\rho < D/2$, with

$$\left.\begin{matrix} E_{\mathrm{co}_{45^\circ}}(\rho) \\ E_{\mathrm{xp}_{45^\circ}}(\rho) \end{matrix}\right\} = -\frac{1}{F}\cos^2(\theta_f/2)\begin{pmatrix} G_{\mathrm{co}_{45^\circ}}(\theta_f) \\ G_{\mathrm{xp}_{45^\circ}}(\theta_f) \end{pmatrix} e^{-j2kF} \ . \tag{9.49}$$

The cross-polar part in (9.48) varies as $\sin 2\varphi$ which gives four cross-polar sidelobes with maximum in the $\varphi = \pm 45^\circ$-planes, as for the feed pattern. We see also that the aperture field has constant phase if the far-field function of the feed has constant phase. This never appears exactly in reality, but the most constant phase appears when the phase reference point is at the phase center of the feed. In addition, the feed must be located relative to the reflector such that its phase center is located at the focal point of the paraboloid. This appears automatically if the phase reference point coincides with the phase center. Actually all the above derivations were done with the initial assumption that the phase reference point coincides with the focal point. Finally, we see that the illumination of the aperture rim relative to the center of the aperture is

$$\begin{aligned} &-20\log|E_{\mathrm{co}_{45^\circ}}(D/2)/E_{\mathrm{co}_{45^\circ}}(0)| = \\ &= -20\log|\cos^2(\theta_0/2)G_{\mathrm{co}_{45^\circ}}(\theta_0)/G_{\mathrm{co}_{45^\circ}}(0)| \ . \end{aligned} \tag{9.50}$$

This value is commonly referred to as the *aperture illumination taper*. The *feed illumination taper* is correspondingly

$$-20\log|G_{\mathrm{co}_{45^\circ}}(\theta_0)/G_{\mathrm{co}_{45^\circ}}(0)| \ . \tag{9.51}$$

The difference between these two is due to the space attenuation between the feed and the reflector, which has a $\cos^2(\theta/2)$ variation with θ.

[4] See the BOR_1 relations in Eqs. (2.84)-(2.87) in Section 2.4.2.
[5] See also Eq. (7.44) to (7.49) in Section 7.5.1.

9.2.3 Typical radiation pattern of paraboloidal reflector

The far field pattern can be computed most accurately by PO integration as explained in Section 9.1.4. It is then important to add the contribution from the far field function of the feed, according to (9.17). If we do not do this, there will be a large erroneous radiation in the region behind the reflector. The reason is that the PO integral over the reflector and the direct radiation from the feed cancel each other vectorialy in the shadow region behind the reflector, because the PO integration provides the scattered field from the reflector[6].

The far field pattern from aperture integration is less accurate, but more convenient for analytical work and for interpretations. When we have found the aperture field we can directly use the aperture field integration formulas for free space apertures in Chapter 7. It is important to note that the aperture field is an approximation of the total field caused by the far field of the feed being reflected by the reflector surface, in contrast to the PO integral that only represents the scattered field of the reflector[7]. Therefore, when we include the direct contribution to the total far field from the far field of the feed, we must in this method remove the part of the feed pattern that is blocked by the reflector. This is illustrated in Fig. 9.8. The part of the feed pattern that radiates outside the reflector is referred to as spillover, and represents both a loss in aperture efficiency and increased sidelobes. By comparing the right graphs of Fig. 9.8(b) and 9.8(c) we can clearly see the contribution to the total sidelobes due to direct feed radiation. These so-called spillover lobes appears close to $\theta = 180^\circ - \theta_0$. There is a sharp so-called shadow boundary at the edge of the reflector at $\theta = 180^\circ - \theta_0$, providing a shadow behind the reflector. This will in reality be diffuse due to edge diffraction, which can be accounted for by extending the ray tracing theory with edge diffraction theory. This causes the spillover lobes to be lower than shown in Fig. 9.8. The free space aperture integration formula (using the Huygens equivalent) provides also a far field in the shadow region behind the reflector. This is inaccurate but still usable in theoretical works. The PO integration approach provides much more correct far field patterns than the ones presented here by aperture integration and addition of feed spillover. Still, the present approach is still very useful for interpretation, and, it is numerically faster.

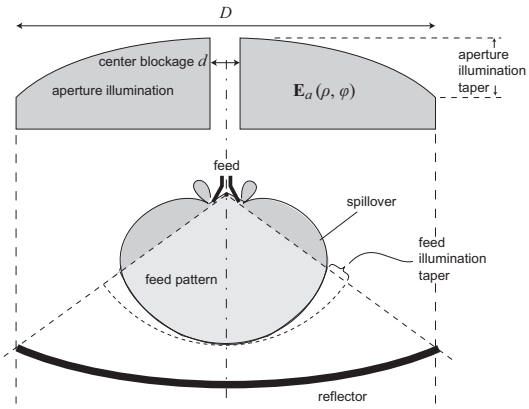
Fig. 9.8(a) illustrates a center blockage of the aperture field, due to the feed (or the subreflector in a rotationally symmetric dual-reflector antenna). This effect will be treated in more detail in Section 9.4.6. We see the effect on the sidelobes by comparing the near-in sidelobes in the two graphs in Fig. 9.8(c). The center blockage causes an increased first sidelobe, and the increase is larger the stronger the feed illumination taper is. We see that every second sidelobe get increased level due to the center blockage.

9.2.4 Directivity, feed efficiency and spillover

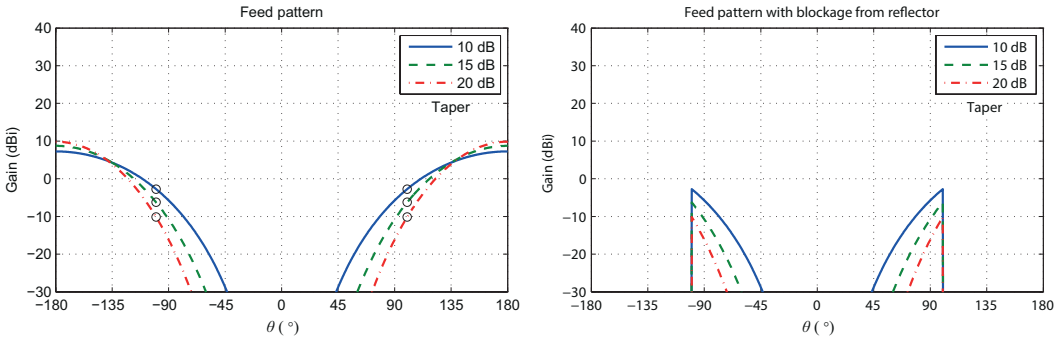
The total radiated power can be found by integrating the radiation intensity of the feed. The corresponding *power integral* P becomes (see the definition in (2.93))

$$P = 2\pi \int_0^\pi [|G_{\mathrm{co}_{45^\circ}}(\theta_f)|^2 + |G_{\mathrm{xp}_{45^\circ}}(\theta_f)|^2] \sin\theta_f d\theta_f \ . \tag{9.52}$$

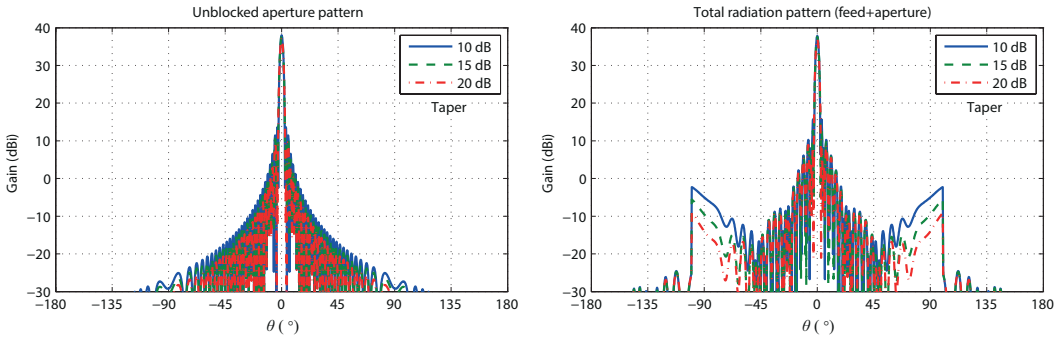
[6] See the definition of scattering in Section 1.3.1.
[7] See the definition of scattering in Section 1.3.1.



(a)



(b)



(c)

Figure 9.8: *Effects of spillover, center blockage and taper in symmetrical paraboloid. (a) Illustration of feed pattern and aperture distribution in cross section of rotationally symmetric paraboloidal reflector with diameter D. (b) and (c) Contributions to co-polar far field patterns of paraboloid in 45°-plane when $D = 30\lambda$, $d = 3\lambda$ and $\theta_0 = 80°$ for different feed illumination tapers. Illustrated for $\cos^n(\theta/2)$ feed pattern, and by integration over reflector aperture. All contributions are presented in the output coordinate system of the reflector with vertical z-axis. Therefore, the maximum of the feed pattern appears at $\pm 180°$.

The far-field function of the reflector can be calculated from the aperture field in (9.48) by using the equations in Section 7.5.1. We will study this far-field function for $\theta = 0$, in order to derive expressions for the aperture efficiency. We then have

$$\mathbf{G}(0,0) = -jC_k 2\tilde{E}_{a_y}(0,0)\hat{\mathbf{y}} \ , \tag{9.53}$$

where $C_k = -jk/4\pi$ and (if we use (9.34) and suppress e^{-j2kF})

$$\begin{aligned} \tilde{E}_{a_y}(0,0) &= 2\pi \int_0^{D/2} E_{\mathrm{co}_{45^\circ}}(\rho)\rho d\rho \\ &= \pi D \cot\left(\frac{\theta_0}{2}\right) \int_0^{\theta_0} G_{\mathrm{co}_{45^\circ}}(\theta_f) \tan\left(\frac{\theta_f}{2}\right) d\theta_f \ . \end{aligned} \tag{9.54}$$

In what follows, we use both the two expressions for the aperture integral in (9.54). The former is an integral over the aperture distribution, whereas the equivalent latter expression is written as an integral over the feed pattern, see the illustration in Fig. 9.8. The former expression is generally valid for all reflector types, whereas the latter expression can only be used for feeds illuminating paraboloidal reflectors.

The *directivity* becomes[8]

$$D_0 = \frac{4\pi|\mathbf{G}(0,0)\cdot\hat{\mathbf{co}}^*|^2}{P} = \frac{4\pi}{\lambda^2} e_{\mathrm{ap}} \pi \left(\frac{D}{2}\right)^2 = e_{\mathrm{ap}} \left(\frac{\pi D}{\lambda}\right)^2 \ , \tag{9.55}$$

where the *aperture efficiency* is

$$e_{\mathrm{ap}} = |\tilde{E}_{a_y}(0,0)|^2/(PA) \tag{9.56}$$

with $A = \pi(D/2)^2$, as the aperture area. This can be expressed as

$$e_{\mathrm{ap}} = \frac{4\pi \cot^2(\theta_0/2) \left| \int_0^{\theta_0} G_{\mathrm{co}_{45^\circ}}(\theta_f) \tan(\theta_f/2) d\theta_f \right|^2}{2\pi \int_0^{\pi} [|G_{\mathrm{co}_{45^\circ}}(\theta_f)|^2 + |G_{\mathrm{xp}_{45^\circ}}(\theta_f)|^2] \sin\theta_f d\theta_f} \ . \tag{9.57}$$

The effects of tolerances and the blockage caused by the feed and its support legs have not been included in this version of e_{ap}. It is a function of the radiation characteristics of the feed, and it is therefore often referred to as the *feed efficiency*. It is also a function of the depth θ_0 of the paraboloid. We can alternatively express the aperture efficiency in terms of the integrals of the aperture distribution, according to

$$e_{\mathrm{ap}} = e_{\mathrm{sp}} \left(\frac{\left(\frac{1}{A}\right) \left| 2\pi \int_0^{D/2} E_{\mathrm{co}_{45^\circ}}(\rho)\rho d\rho \right|^2}{2\pi \int_0^{D/2} [|E_{\mathrm{co}_{45^\circ}}(\rho)|^2 + |E_{\mathrm{xp}_{45^\circ}}(\rho)|^2]\rho d\rho} \right) \ , \tag{9.58}$$

where

$$e_{\mathrm{sp}} = \frac{2\pi \int_0^{\theta_0} [|G_{\mathrm{co}_{45^\circ}}(\theta_f)|^2 + |G_{\mathrm{xp}_{45^\circ}}(\theta_f)|^2] \sin\theta_f d\theta_f}{2\pi \int_0^{\pi} [|G_{\mathrm{co}_{45^\circ}}(\theta_f)|^2 + |G_{\mathrm{xp}_{45^\circ}}(\theta_f)|^2] \sin\theta_f d\theta_f} \tag{9.59}$$

is the *spillover efficiency*. The spillover efficiency represents the power within the subtended angle θ_0 (i.e., the power hitting the main reflector) relative to the total power radiated by

[8] See Section 2.5.2.

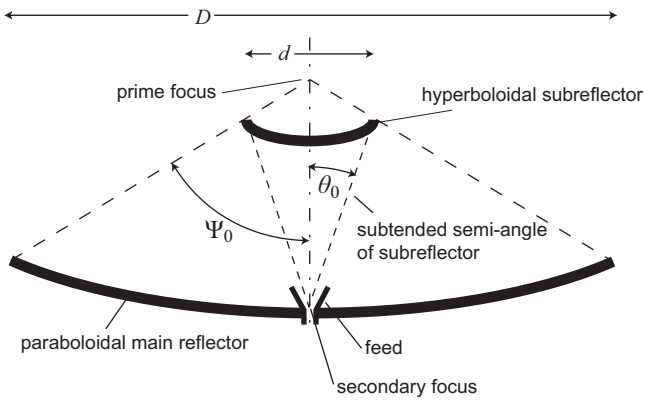


Figure 9.9: Example of Cassegrain dual-reflector antenna

the feed. The total power in the aperture is equal to the total power reflected from the main reflector.

We will in Section 9.4 factorize e_{ap} in different subefficiencies, characterizing different properties of the feed pattern. We will also give numerical values for the case that $G_{\text{co}_{45^\circ}}(\theta_f) = \cos^n(\theta_f/2)$.

9.3 The Cassegrain antenna

The classical *Cassegrain antenna* consists of a paraboloidal main reflector and a hyperboloidal subreflector. We need four independent parameters in order to uniquely describe the geometry of the Cassegrain (Fig. 9.9). The most convenient ones are the main reflector diameter D, the subreflector diameter d, the subtended half-angle θ_0 of the subreflector, and the subtended half-angle Ψ_0 of the main reflector. The shape of the main reflector is given by (9.31) by replacing θ_f and θ_0 by Ψ and Ψ_0. We will not give the equation for the subreflector here. The paraboloidal main reflector has a primary focal point in the origin of the global coordinate system, and the dual-reflector system has a secondary focal point, whose location can be calculated from d, Ψ_0 and θ_0. We will not give the formula here. In a Cassegrain antenna θ_0 is normally small ($5^\circ < \theta_0 < 30^\circ$) and Ψ_0 is large ($70^\circ < \Psi_0 < 90^\circ$).

9.3.1 Aperture field and efficiency

At high frequencies the *Cassegrain* can be analyzed approximately by aperture integration. We assume that the feed is located with its phase reference point in the secondary focus and has a far-field function of the form given in (9.41). Then, the aperture field is found by using GO ray tracing from the feed via the subreflector and the main reflector to the aperture. The result can be shown to be

$$\mathbf{E}_a(\rho,\varphi) = [E_{\text{co}_{45^\circ}}(\rho) - E_{\text{xp}_{45^\circ}}(\rho)\cos 2\varphi]\hat{\mathbf{y}} - E_{\text{xp}_{45^\circ}}(\rho)\sin 2\varphi\hat{\mathbf{x}}\ , \tag{9.60}$$

with

$$\left.\begin{matrix} E_{\mathrm{co}_{45^\circ}}(\rho) \\ E_{\mathrm{xp}_{45^\circ}}(\rho) \end{matrix}\right\} = \frac{1}{F_{\mathrm{eq}}} \cos^2(\theta_f/2) \begin{pmatrix} G_{\mathrm{co}_{45^\circ}}(\theta_f) \\ G_{\mathrm{xp}_{45^\circ}}(\theta_f) \end{pmatrix} e^{-jkL_{\mathrm{tot}}} \tag{9.61}$$

for $\theta_f < \theta_0$ where $F_{\mathrm{eq}} = D/(4\tan(\theta_0/2))$ is called the *equivalent focal length*. Further, L_{tot} is the total path length from the secondary focus via the two reflectors to the aperture, e.g., measured along the central ray, and $\rho = 2F_{\mathrm{eq}}\tan(\theta_f/2)$. From the form of the aperture field it is clear that the aperture efficiency of the Cassegrain has to be given by the same formula (9.57) as the aperture efficiency of the primary fed paraboloid.

We will in the next section factorize the aperture efficiency in different subefficiencies. In addition, we will add center blockage to it. The center blockage effect in the Cassegrain antenna is caused by the subreflector and in the primary-fed paraboloid by the feed.

9.4 Subefficiencies of paraboloids and Cassegrains

When characterizing a reflector antenna system with a pencil beam, it is quite common to separate the aperture efficiency in different subefficiencies. In this section we show how this can be done for rotationally symmetric paraboloids and Cassegrain antennas. The principle is general and can also apply to offset and multi-reflector systems. Nevertheless, the formulas for the different subefficiencies are different. The feed is assumed to be of BOR_1 type.

9.4.1 Spillover, polarization, illumination and phase efficiencies

The first approximation of the efficiency of a paraboloid or Cassegrain antenna was already given by (9.57), and is often referred to as the *feed efficiency*. This can be factorized in several contributions [1], according to

$$e_{\mathrm{ap}} = e_{\mathrm{sp}} e_{\mathrm{pol}} e_{\mathrm{ill}} e_{\phi} \ , \tag{9.62}$$

where each factor is described below.

The first efficiency is the *spillover efficiency* e_{sp}. This is the power within the subtended angle θ_0 (i.e., the power hitting the reflector) relative to the total power radiated by the feed. The equation for it was already given in (9.59). The relative spillover power is given by $1 - e_{\mathrm{sp}}$. This should be reduced as much as possible in order to improve the directivity. In addition, it is a major contributor to the antenna noise temperature, in particular if the spillover radiation hits the hot ground[9]. In a paraboloid or Cassegrain antenna the spillover efficiency is typically between $-0.05\,\mathrm{dB}$ and $-0.5\,\mathrm{dB}$, depending on the illumination taper of the aperture and the quality of the feed pattern.

The next subefficiency e_{pol} is the *polarization sidelobe efficiency*. This is the power of the co-polar field relative to the total power, both within θ_0. For circular polarization this becomes

[9] See Subsection 2.5.8.

$$
\begin{aligned}
e_{\text{pol}} &= \int_0^{D/2} |E_{\text{co}_{45^\circ}}(\rho)|^2 \rho d\rho \Big/ \int_0^{D/2} [|E_{\text{co}_{45^\circ}}(\rho)|^2 + |E_{\text{xp}_{45^\circ}}(\rho)|^2] \rho d\rho \\
&= \int_0^{\theta_0} |G_{\text{co}_{45^\circ}}(\theta_f)|^2 \sin\theta_f d\theta_f \Big/ \int_0^{\theta_0} [|G_{\text{co}_{45^\circ}}(\theta_f)|^2 + |G_{\text{xp}_{45^\circ}}(\theta_f)|^2] \sin\theta_f d\theta_f \ .
\end{aligned}
\tag{9.63}
$$

In this section we are considering ideal excitations for which the radiation field of the feed has zero cross-polarization on axis. Therefore, this polarization sidelobe efficiency is a measure of the power lost in cross-polar sidelobes within θ_0. This polarization sidelobe efficiency must not be mixed up with the polarization efficiency in Section 2.3.11, which accounts for a non-ideal excitation of the antenna giving a cross-polar level on axis.

The polarization sidelobe efficiency for linear polarization gets a slightly different form than that for circular polarization in (9.63). The reason is that the BOR_1 type far-field function for linear polarization has a cross-polar part with $\sin(2\varphi)$ variation, giving sidelobes in the 45°-planes only, whereas it has ring-shaped sidelobes around the symmetry axis for circular polarization. The polarization sidelobe efficiency e_x for linear polarization is always the highest and can be expressed in terms of e_{pol} for circular polarization, according to

$$
e_x = 1 - \frac{1}{2}(1 - e_{\text{pol}}) \ . \tag{9.64}
$$

It is normally not needed to differentiate between the two forms of the polarization efficiency. It is sufficient to use (9.63) for both polarizations, and consider it as a measure of the reduction in efficiency due to phase and amplitude differences between the E- and H-plane far-field functions. In most reflector antennas the polarization sidelobe efficiency is very high, typically better than $-0.1\,\text{dB}$.

By using (9.63) as the polarization efficiency, we achieve the following *illumination efficiency*

$$
\begin{aligned}
e_{\text{ill}} &= \frac{\left| 2\pi \int_0^{D/2} |E_{\text{co}_{45^\circ}}(\rho)| \rho d\rho \right|^2}{A 2\pi \int_0^{D/2} |E_{\text{co}_{45^\circ}}(\rho)|^2 \rho d\rho} \\
&= 2\cot^2\left(\frac{\theta_0}{2}\right) \frac{\left[\int_0^{\theta_0} |G_{\text{co}_{45^\circ}}(\theta_f)| \tan\left(\frac{\theta_f}{2}\right) d\theta_f \right]^2}{\int_0^{\theta_0} |G_{\text{co}_{45^\circ}}(\theta_f)|^2 \sin\theta_f d\theta_f} \ ,
\end{aligned}
\tag{9.65}
$$

with $A = \pi(D/2)^2$. The illumination efficiency, e_{ill}, becomes unity for a uniform aperture illumination, i.e., when $|E_{\text{co}_{45^\circ}}(\rho)| = const$ corresponding to $|G_{\text{co}_{45^\circ}}(\theta_f)| = const/\cos^2(\theta_f/2)$. The illumination efficiency is in a practical antenna typically between $-0.4\,\text{dB}$ and $-1.5\,\text{dB}$, for illumination tapers varying between 10 dB and 20 dB.

The remaining subefficiency is due to phase errors in the co-polar radiation field $G_{\text{co}_{45^\circ}}(\theta_f)$. This is the *phase efficiency*

$$
e_\phi = \frac{\left| \int_0^{D/2} E_{\text{co}_{45^\circ}}(\rho) \rho d\rho \right|^2}{\left[\int_0^{D/2} |E_{\text{co}_{45^\circ}}(\rho)| \rho d\rho \right]^2} = \frac{\left| \int_0^{\theta_0} G_{\text{co}_{45^\circ}}(\theta_f) \tan\left(\frac{\theta_f}{2}\right) d\theta_f \right|^2}{\left[\int_0^{\theta_0} |G_{\text{co}_{45^\circ}}(\theta_f)| \tan\left(\frac{\theta_f}{2}\right) d\theta_f \right]^2} \ . \tag{9.66}
$$

The phase efficiency is the only subefficiency which depends on the location of the phase reference point of the feed, i.e., the location of the feed relative to the focal point of the reflector. This fact can be used to uniquely define a *phase center* for the feed, corresponding to the feed location which maximizes the phase efficiency. A formula for calculation of the phase center will be derived from this definition to be presented in Section 9.4.3. When the feed is located with its phase center in the focal point of the reflector, the phase efficiency is normally very high, typically better than $-0.1\,\mathrm{dB}$. If the co-polar far-field function has constant phase, the phase efficiency is $0.0\,\mathrm{dB}$.

9.4.2 Example: $\cos^n(\theta_f/2)$ feed

In order to illustrate the significance of the different subefficiencies it is convenient to use a theoretical radiation pattern as an example. Let us therefore assume that

$$G_{\mathrm{co}_{45^\circ}}(\theta_f) = \cos^n(\theta_f/2)\ , \qquad G_{\mathrm{xp}_{45^\circ}}(\theta_f) = 0\ , \tag{9.67}$$

which describes the shape of the main lobe of the most common feed patterns by choosing n such that, e.g., the $10\,\mathrm{dB}$ tapers of (9.67) and of the actual feed pattern are equal. Thus, n can be calculated from

$$n = |A_0|_{\mathrm{dB}}/(-20\log(\cos(\theta_0/2)))\ ,$$

where $|A_0|_{\mathrm{dB}} = -20\log(|A_0|)$ is the amplitude taper A_0 in dB of the actual feed pattern at θ_0, i.e.,

$$A_0 = G_{\mathrm{co}_{45^\circ}}(\theta_0)/G_{\mathrm{co}_{45^\circ}}(0)\ . \tag{9.68}$$

This $\cos^n(\theta_f/2)$ pattern is very similar to a Gaussian beam within the $10\,\mathrm{dB}$ beamwidth, but it has the advantage over the Gaussian beam that the feed efficiency integral can be solved analytically. By using (9.57), this becomes

$$e_f = 4\cot^2(\theta_0/2)[1-\cos^n(\theta_0/2)]^2(n+1)/n^2\ . \tag{9.69}$$

Eq. (9.69) is shown in Fig. 9.10, as a function of the *feed illumination taper* $|A_0|_{\mathrm{dB}}$ [10]. It is also common to consider the efficiency as a function of the *aperture illumination taper* defined in (9.50) which is stronger than the feed taper because the space attenuation increases with θ. For our $\cos^n(\theta_f/2)$ feed pattern the aperture taper becomes

$$A_{\mathrm{ap}} = A_0\cos^2(\theta_0/2)\ .$$

We see in Fig. 9.10 that higher feed efficiencies are available in Cassegrain systems (small θ_0) than in primary-fed paraboloids (large θ_0). The highest feed efficiency is about $-0.9\,\mathrm{dB}$, and appears for $10\,\mathrm{dB}$ feed illumination taper in a Cassegrain antenna. Higher efficiencies are available with improved feed designs[11].

The spillover efficiency for the same feed pattern becomes

$$e_{\mathrm{sp}} = 1-\cos^{2n+2}(\theta_0/2) \approx 1-(A_0)^2\ , \tag{9.70}$$

where the approximation is valid for $\theta_0 < 30^\circ$. The most accurate formula is plotted as a function of the feed illumination taper in Fig. 9.10. Thus, in a Cassegrain with $-10\,\mathrm{dB}$

[10] There exist MATLAB code for all figures of which the caption start with ⋆.

[11] Shaped reflectors can also be used.

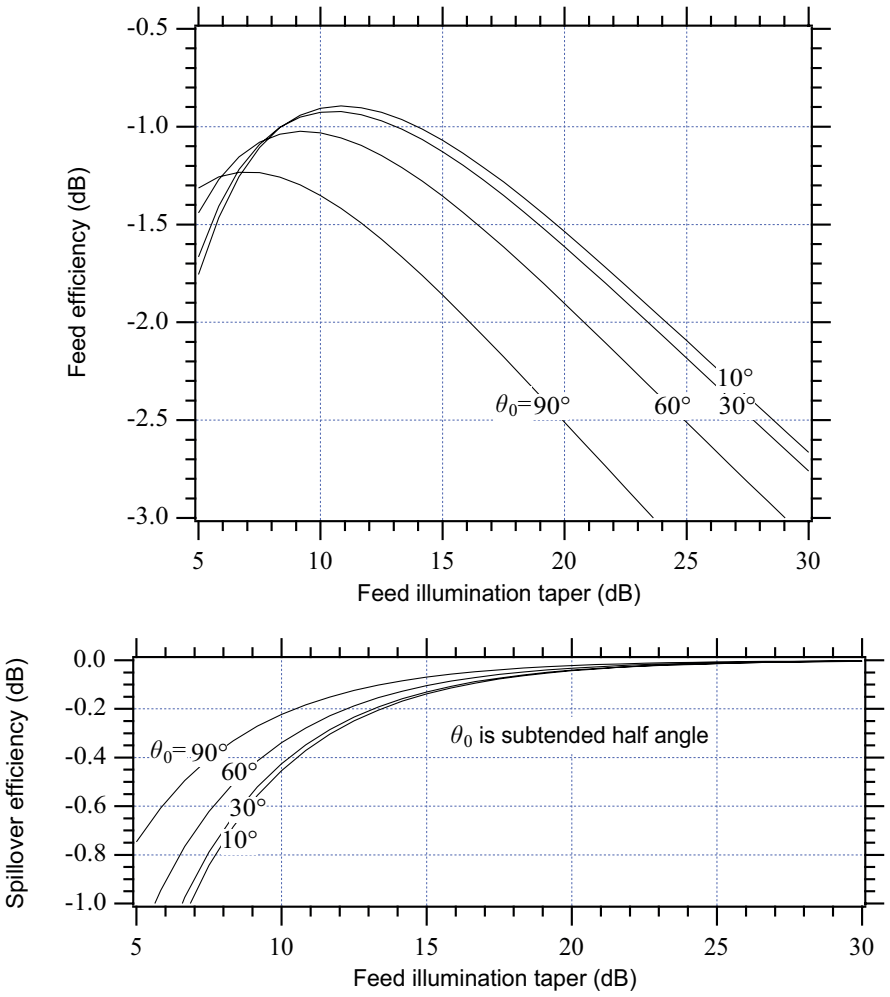


Figure 9.10: *Feed efficiency and spillover efficiency for the theoretical feed pattern $\cos^n(\theta/2)$ when it is used to feed a paraboloid or Cassegrain antenna.

illumination taper the spillover efficiency is about 90 % whereas it is 99 % if the taper is 20 dB. The illumination efficiency can be found by dividing (9.69) with (9.70). It becomes almost equal to the spillover efficiency when the taper is 10 dB.

The cross-polar feed pattern $G_{\mathrm{xp}_{45^\circ}}(\theta_f)$ is half the difference between the E-and H-plane patterns. It is very sensitive to phase differences between them. If we assume that the E- and H-plane patterns have ideal phase centers which are separated by 2δ, and that the patterns otherwise are equal, we may write

$$G_{_E}(\theta_f) = \cos^n(\theta_f/2)e^{j2k\delta\sin^2(\theta_f/2)} \ , \tag{9.71}$$

$$G_{_H}(\theta_f) = \cos^n(\theta_f/2)e^{-j2k\delta\sin^2(\theta_f/2)} \ , \tag{9.72}$$

where $k = 2\pi/\lambda$ is the wavenumber. This gives

$$G_{\mathrm{co}_{45^\circ}}(\theta_f) = \cos^n(\theta_f/2)\cos(2k\delta\sin^2(\theta_f/2)) \ , \tag{9.73}$$

$$G_{\mathrm{xp}_{45^\circ}}(\theta_f) = j\cos^n(\theta_f/2)\cos(2k\delta\sin^2(\theta_f/2)) \ . \tag{9.74}$$

We can get some simple and useful results if we approximate $\cos^{2n}(\theta_f/2)$ by

$$|G_{_H}(\theta_f)|^2 = |G_{_E}(\theta_f)|^2 = 1 - (1 - A_{_0}^2)(\theta_f/\theta_{_0})^2 \qquad \text{for } \theta_f < \theta_{_0} \ , \tag{9.75}$$

where $A_{_0} = |G_{_E}(\theta_{_0})/G_{_E}(0)|$ is the feed illumination taper. Then, when $\theta_{_0}$ is small and $2k\delta\sin^2(\theta_{_0}/2) \ll \pi/2$, it is possible to derive analytical expressions for the maximum of the cross-polar pattern $G_{\mathrm{xp}_{45^\circ}}(\theta_f)$. The maximum is found to appear for

$$\theta_m = \theta_{_0}\sqrt{(2/[3(1 - A_{_0}^2)])} \ , \tag{9.76}$$

for which the co-polar level is

$$|G_{\mathrm{co}_{45^\circ}}(\theta_m)/G_{\mathrm{co}_{45^\circ}}(0)|^2 = 1/3 \quad (\text{i.e., } -5\,\text{dB}) \ , \tag{9.77}$$

and the maximum cross-polar sidelobe level becomes

$$|G_{\mathrm{xp}_{45^\circ}}(\theta_m)/G_{\mathrm{co}_{45^\circ}}(0)|^2 \approx (\Delta\phi_{_0})^2/27 \ , \tag{9.78}$$

where $\Delta\phi_{_0}$ is the difference in radians between the phases of the E- and H-plane radiation patterns for $\theta_f = \theta_{_0}$, i.e., $\Delta\phi_{_0} = 4k\delta\sin^2(\theta_{_0}/2)$ from (9.71) and (9.72). This means that the maximum of the cross-polar sidelobe appears where the co-polar field is about 5 dB lower, if the cross-polarization is caused by different phase centers in E- and H-planes. Also, from (9.78), we see that in order to get a cross-polar sidelobe which is more than 30 dB lower than the co-polar maximum, we must require that the difference $\Delta\phi_{_0}$ between the E- and H-plane phases is less than about 10° for $\theta_f = \theta_{_0}$.

An approximate and illustrative formula for the polarization efficiency can also be derived when $|G_{_E}(\theta_f)|^2$ and $|G_{_H}(\theta_f)|^2$ are given by (9.75). With the same assumptions which were used to obtain (9.76)-(9.78), we obtain

$$e_{\mathrm{pol}} \approx 1 - (\Delta\phi_{_0})^2/24 \ , \tag{9.79}$$

for illumination tapers of 10 dB or more, and

$$e_{\mathrm{pol}} \approx 1 - (\Delta\phi_{_0})^2/12 \ , \tag{9.80}$$

for a uniform illumination. We see that the polarization efficiency is higher than 99 % if the maximum of the cross-polar sidelobe is more than 20 dB lower than the main lobe. In fact, by comparing (9.78) and (9.79) we see that for a tapered illumination we have the following approximate relation between the polarization efficiency $e_{\rm pol}$ and the maximum cross-polar sidelobe

$$e_{\rm pol} \approx 1 - |G_{{\rm xp}_{45^\circ}}(\theta_m)/G_{{\rm co}_{45^\circ}}(0)|^2 \ . \tag{9.81}$$

Therefore, if the maximum cross-polar sidelobe is $20\log|G_{{\rm xp}_{45^\circ}}(\theta_m)/G_{{\rm co}_{45^\circ}}(0)|^2 = -20\,{\rm dB}$, the polarization efficiency is about 99 %, i.e., $-0.05\,{\rm dB}$. Thus, the polarization efficiency is very high for normal cross-polar sidelobe levels.

9.4.3 Phase center

The *phase center* is the location of the phase reference point which minimizes the phase variations of the co-polar far-field function $G_{{\rm co}_{45^\circ}}(\theta_f)$ over a specified angular region. A simple formula to calculate the phase center was given in Section 2.3.7. This is valid if the phase of $G_{{\rm co}_{45^\circ}}(\theta_f)$ is nearly constant when referred to the phase center. However, in practice the phase of $G_{{\rm co}_{45^\circ}}(\theta_f)$ will never be constant, so we introduce an exact definition of it, as the phase reference point which maximizes the phase efficiency in (9.66). This definition is valid for feeds in paraboloids and Cassegrain antennas, and we will use it to derive an expression for calculation of its location [2]. First, we write η_ϕ in a form which makes it more convenient for maximization. We introduce

$$G_{{\rm co}_{45^\circ}}(\theta) = |G_{{\rm co}_{45^\circ}}(\theta)|e^{j\phi(\theta)} \ ; \qquad G_{{\rm co}_{45^\circ}\delta}(\theta) = |G_{{\rm co}_{45^\circ}}(\theta)|e^{j\phi_\delta(\theta)} \ , \tag{9.82}$$

$$\text{where} \qquad \phi_\delta(\theta) = \phi(\theta) - k\delta\cos\theta \tag{9.83}$$

is the phase function when the phase reference point is moved to $z = \delta$ (See (2.51)). We assume small phase errors, i.e.,

$$\phi_\delta(\theta) - \phi_\delta(0) \ll \frac{\pi}{2} \tag{9.84}$$

and get by Taylor expansion of the phase factor

$$e_{\phi_\delta} = 1 - \overline{(\phi_\delta^2)} + (\overline{\phi_\delta})^2 \ , \tag{9.85}$$

where $\overline{\phi_\delta}$ and $\overline{(\phi_\delta^2)}$ are the mean and the mean-square phase errors, respectively, given by

$$\overline{\phi_\delta} = \int_0^{\theta_0} w(\theta)[\phi_\delta(\theta) - \phi_\delta(0)]d\theta \Big/ \int_0^{\theta_0} w(\theta)d\theta \ , \tag{9.86}$$

$$\overline{(\phi_\delta^2)} = \int_0^{\theta_0} w(\theta)[\phi_\delta(\theta) - \phi_\delta(0)]^2 d\theta \Big/ \int_0^{\theta_0} w(\theta)d\theta \ , \tag{9.87}$$

where the weight is

$$w(\theta) = |G_{{\rm co}_{45^\circ}}(\theta)|\tan(\theta/2) \ . \tag{9.88}$$

An alternative and equal expression is

$$e_{\phi\delta} = 1 - \overline{(\phi_\delta - \overline{\phi_\delta})^2} \ , \tag{9.89}$$

$$\text{where} \qquad \overline{(\phi_\delta - \overline{\phi_\delta})^2} = \int_0^{\theta_0} w(\theta)[\phi_\delta(\theta) - \overline{\phi_\delta}]^2 d\theta / \int_0^{\theta_0} w(\theta) d\theta \tag{9.90}$$

is the mean-square phase error around the mean phase error.

By substituting (9.83) into (9.86) and (9.87), we can write (9.85) as a function of $k\delta$ as

$$e_{\phi\delta} = c - 2bk\delta - a(k\delta)^2 \ , \tag{9.91}$$

where

$$c = 1 - \frac{I_{\mathrm{w}\phi 2}}{I_w} + \left(\frac{I_{\mathrm{w}\delta}}{I_w}\right)^2 \ , \qquad b = \frac{I_{\mathrm{w}\phi\mathrm{c}}}{I_w} - \frac{I_{\mathrm{w}\phi} I_{\mathrm{wc}}}{I_w^2} \ , \qquad a = \frac{I_{\mathrm{wc}2}}{I_w} - \left(\frac{I_{\mathrm{wc}}}{I_w}\right)^2 \ ,$$

$$\text{with} \qquad I_w = \int_0^{\theta_0} w(\theta) d\theta \ , \qquad I_{\mathrm{wc}} = \int_0^{\theta_0} w(\theta)[\cos\theta - 1] d\theta \ ,$$

$$I_{\mathrm{wc}2} = \int_0^{\theta_0} w(\theta)[\cos\theta - 1]^2 d\theta \ , \qquad I_{\mathrm{w}\phi} = \int_0^{\theta_0} w(\theta)[\phi(\theta) - \phi(0)] d\theta \ ,$$

$$I_{\mathrm{w}\phi 2} = \int_0^{\theta_0} w(\theta)[\phi(\theta) - \phi(0)]^2 d\theta \qquad \text{and} \qquad I_{\mathrm{w}\phi\mathrm{c}} = \int_0^{\theta_0} w(\theta)[\cos\theta - 1][\phi(\theta) - \phi(0)] d\theta \ .$$

Eq. (9.91) has a simple dependence on δ and can easily be maximized. The result $\delta = \delta_0$, which we define as the phase center, becomes

$$k\delta_0 = \frac{b}{a} = \frac{I_w I_{\mathrm{w}\phi\mathrm{c}} - I_{\mathrm{w}\phi} I_{\mathrm{wc}}}{I_{\mathrm{wc}2} I_w - (I_{\mathrm{wc}})^2} \ . \tag{9.92}$$

The phase center, as defined here, is a function of the subtended angle θ_0. The corresponding maximum *phase efficiency* is

$$(e_{\phi\delta})_{\max} = c + \frac{b^2}{a} \ . \tag{9.93}$$

The above formulas assume small phase errors. If they are not small, we may first calculate an approximate phase center by using (2.63), transform the phase reference point to this location, and thereafter use the above formulas.

9.4.4 Axial displacements of feed

For small axial displacements of the feed around the phase center δ_0, we can rewrite (9.91) in the following form

$$e_{\phi\delta} = (e_{\phi\delta})_{\max} - a[k(\delta - \delta_0)]^2 \ . \tag{9.94}$$

This equation can be used to find the required axial tolerances of the feed position. Let us define the tolerance $\pm\Delta\delta$ as the displacement, causing an efficiency reduction of 0.1 dB, corresponding to $e_{\phi\delta} = 0.977(e_{\phi\delta})_{\max}$. This gives from (9.94)

$$k\Delta\delta = 0.151\sqrt{\frac{1}{a}(e_{\phi\delta})_{\max}} \ . \tag{9.95}$$

From (9.91) we find that it is independent of the phase pattern $\phi(\theta)$. It is therefore illustrative to elaborate (9.94) and (9.95) with an example. We consider a theoretical feed of the form

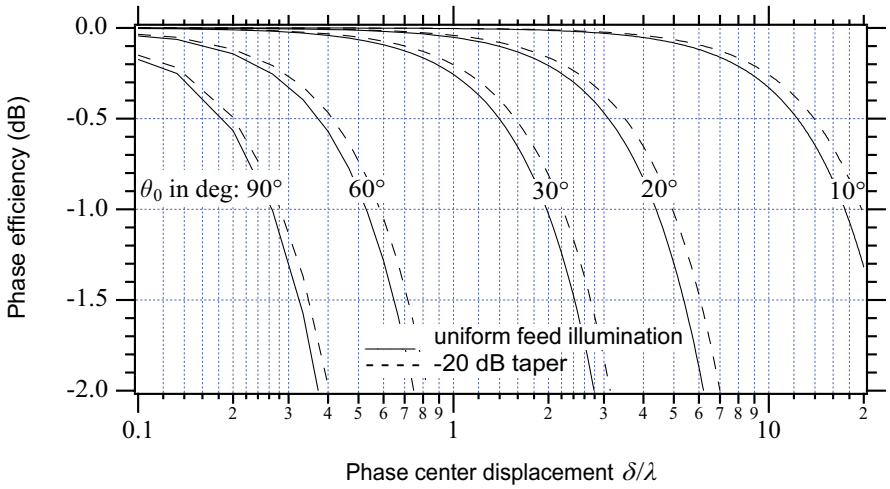


Figure 9.11: ⋆Efficiency losses due to defocussing of paraboloid or conventional Cassegrain. θ_0 is subtended half angle of reflector system, and δ is spacing between focal point of reflector system and phase center of feed.

$\cos^n(\theta/2)$, evaluate a numerically, and present results as function of the feed illumination taper $\cos^n(\theta_0/2)$ for $(e_{\phi\delta})_{\max} = 1$ in Fig. 9.11 and Fig. 9.12⋆. We see that $\Delta\delta$ depends strongly on the subtended angle θ_0, and is almost independent of the feed taper. In Cassegrain antennas ($\theta_0 < 30°$) we can displace the feed by several wavelengths before any defocussing appears, whereas in deep primary-fed paraboloids ($\theta_0 \approx 90°$) the feed needs to be located with its phase center within fractions of a wavelength from the focal point.

The analytical results are rather complicated expressions, and are therefore not given here. A useful analytic approximation (except for $\theta_0 \approx 90°$) for the phase center tolerance is

$$\Delta\delta/\lambda = 0.045/\sin^2(\theta_0/2) \ . \tag{9.96}$$

9.4.5 Surface tolerances

The aperture efficiency is also reduced due to the finite surface accuracy of the reflector. This causes phase errors, and the corresponding efficiency reduction can be accounted for by using (9.85) with (see Fig. 9.13)

$$\phi(\theta,\varphi) = (1+\cos\theta)k\Delta z(\theta,\varphi) \approx 2k\Delta z(\theta,\varphi) \ , \tag{9.97}$$

with $\Delta z(\theta,\varphi)$ is the surface error. We see that (9.85) becomes

$$e_{\rm tot} = 1 - (2k\Delta z_{\rm rms})^2 \ ; \qquad \Delta z_{\rm rms} = \sqrt{\overline{(\Delta z)^2} - (\overline{\Delta z})^2} \tag{9.98}$$

is the weighted *Root Mean-Square* (RMS) of the surface error around the mean surface error $\overline{\Delta z}$. For less than 0.3 dB loss we need a RMS surface accuracy better than $\Delta z_{\rm rms} = \lambda/50$.

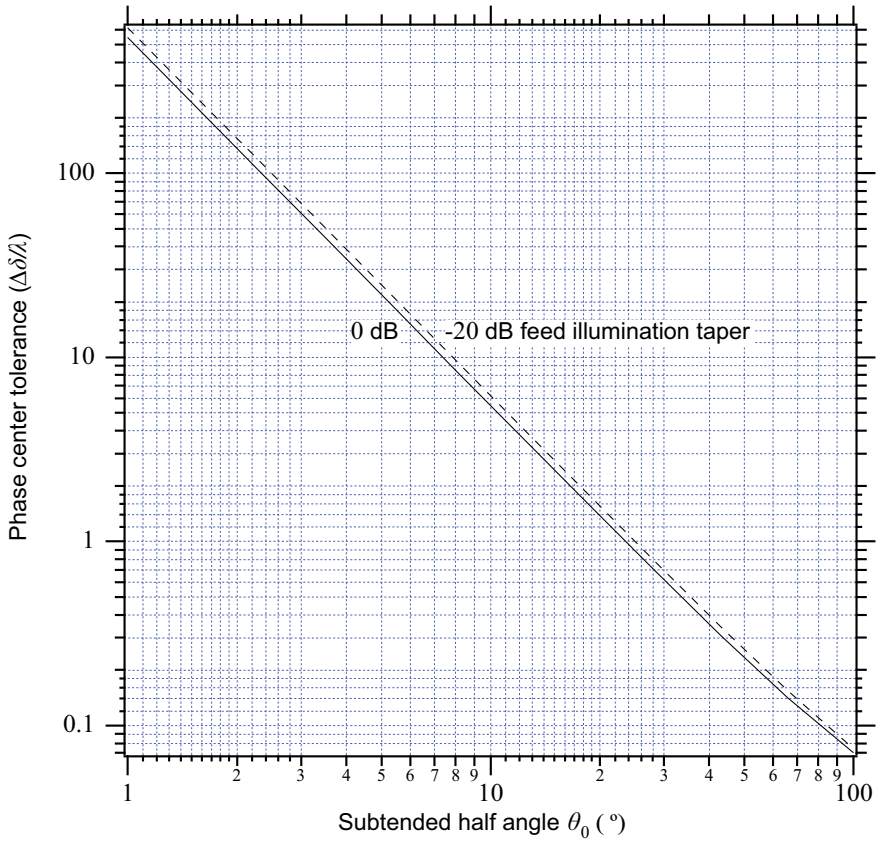


Figure 9.12: *Maximum defocussing $\Delta\delta$ of prime-focus paraboloid or conventional Cassegrain, corresponding to 0.1 dB reduction in phase efficiency $\eta_{\phi\delta}$ for feed illumination tapers of 0 dB and -20 dB.

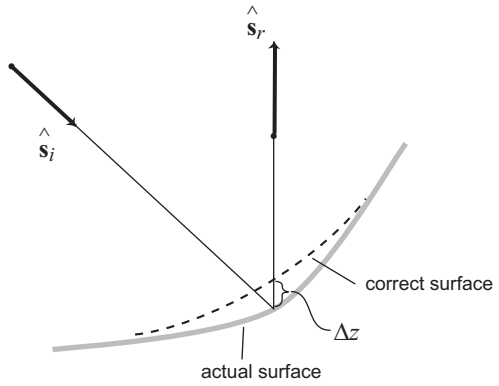


Figure 9.13: Illustration of surface errors.

Some surface errors may increase quadratically towards the reflector rim. If this is the case, the maximum acceptable error at the rim is four times the RMS error. The effects of reflector tolerances on the directivity and radiation pattern was first studied by J. Ruze in the classic paper [3].

9.4.6 Aperture blockage

In rotationally symmetric antennas the feed or the subreflector and their support struts will cause *aperture blockage* (see Fig. 9.8). We will here only treat center blockage from the feed and the subreflector, and refer to [4] for *strut blockage*. The *center blockage* can be accounted for in the efficiency calculations by a *center blockage efficiency*, which is obtained by removing the central region from the aperture integral. By doing this for the total aperture efficiency (see (9.56)) we obtain

$$(e_{\text{ap}})_b = |\tilde{E}_{a_{y_b}}(0,0)|^2/(PA) \ , \tag{9.99}$$

$$\text{with} \qquad \tilde{E}_{a_{y_b}}(0,0) = 2\pi \int_{d/2}^{D/2} E_{\text{co}_{45^\circ}}(\rho)\rho d\rho \ , \tag{9.100}$$

where d is the blockage diameter. We may express (9.100) as

$$\tilde{E}_{a_{y_b}}(0,0) = \tilde{E}_{a_y}(0,0) - 2\pi \int_{0}^{d/2} E_{\text{co}_{45^\circ}}(\rho)\rho d\rho \ , \tag{9.101}$$

where $\tilde{E}_{a_y}(0,0)$ is the unblocked aperture integral in (9.54). Therefore, we see that the aperture efficiency $(e_{\text{ap}})_b$ with blockage may be expressed as

$$(e_{\text{ap}})_b = e_{\text{ap}}|1 - \Delta_{\text{cb}}|^2 \ , \tag{9.102}$$

$$\text{where} \qquad \Delta_{\text{cb}} = \int_0^{d/2} E_{\text{co}_{45^\circ}}(\rho)\rho d\rho / \int_0^{D/2} E_{\text{co}_{45^\circ}}(\rho)\rho d\rho \ . \tag{9.103}$$

If we assume that $d \ll D$ and $E_{\text{co}_{45^\circ}}(\rho) \approx \textit{constant}$ for $0 < \rho < d/2$, and if we introduce (9.54), we get

$$\Delta_{\text{cb}} = C_b(d/D)^2 \ , \tag{9.104}$$

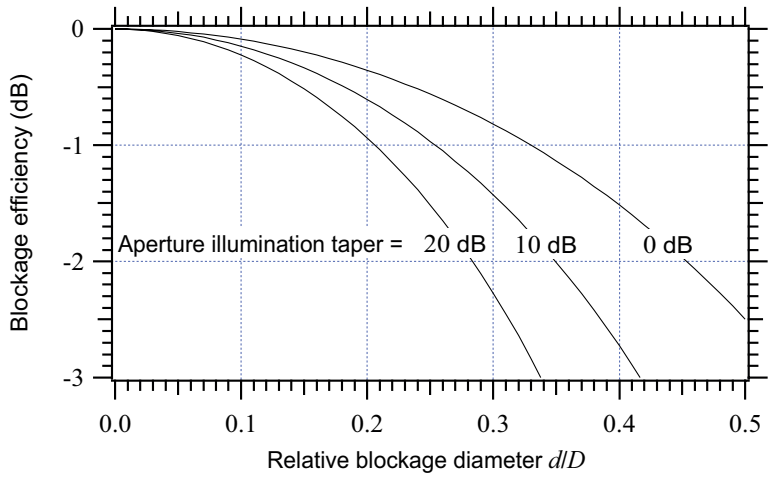


Figure 9.14: ⋆Aperture blockage efficiency $|1-\Delta_{\rm cb}|^2$ as a function of subreflector over main reflector diameter d/D.

$$\text{with} \qquad C_b = \frac{E_{\rm co_{45^\circ}}(0)}{2\int_0^1 E_{\rm co_{45^\circ}}(uD/2)udu} = \frac{\tan^2(\theta_0/2)G_{\rm co_{45^\circ}}(0)}{\int_0^{\theta_0} G_{\rm co_{45^\circ}}(\theta_f)\tan(\theta_f/2)d\theta_f} \ . \tag{9.105}$$

If we use the same $\cos^n(\theta_f/2)$ feed pattern as before, the blockage constant becomes

$$C_b = \frac{\tan^2(\theta_0/2)\cdot n}{2(1-\cos^n(\theta_0/2))} \ . \tag{9.106}$$

The aperture blockage efficiency $|1-\Delta_{\rm cb}|^2$ is evaluated from the above formulas and plotted in Fig. 9.14⋆. The aperture taper is provided by a feed pattern of $\cos^n(\theta_f/2)$ shape by choosing the value of n appropriately. A common choice of d/D in a Cassegrain antenna is $d/D = 0.1$. Then, we see that the blockage efficiency varies between -0.15 dB and -0.3 dB for aperture tapers between 10 dB and 20 dB.

If d/D is larger than 0.1, the above equations will not be accurate. The reason is that the blocked power becomes significant. This is neglected in the above analysis. But in reality, it will be reflected by the blocking element (the feed or the subreflector) back towards the main reflector where it is reradiated and may cause *multiple reflections* and standing waves between the blocking element and the main reflector. These multiple reflections will effect the radiation on axis very differently at different frequencies. It is, however, possible to use them constructively to improve an antenna [5].

Aperture blockage will also cause increased sidelobes of the antenna. These can be calculated by subtracting the far-field function of the blocked region from the far-field function of the unblocked aperture, in a similar way as for the efficiency reduction.

9.4.7 Edge diffraction efficiency

The field incident on the main reflector of the Cassegrain is reflected by GO from the subreflector. GO is only valid when the reflectors are large in terms of wavelengths. The finite

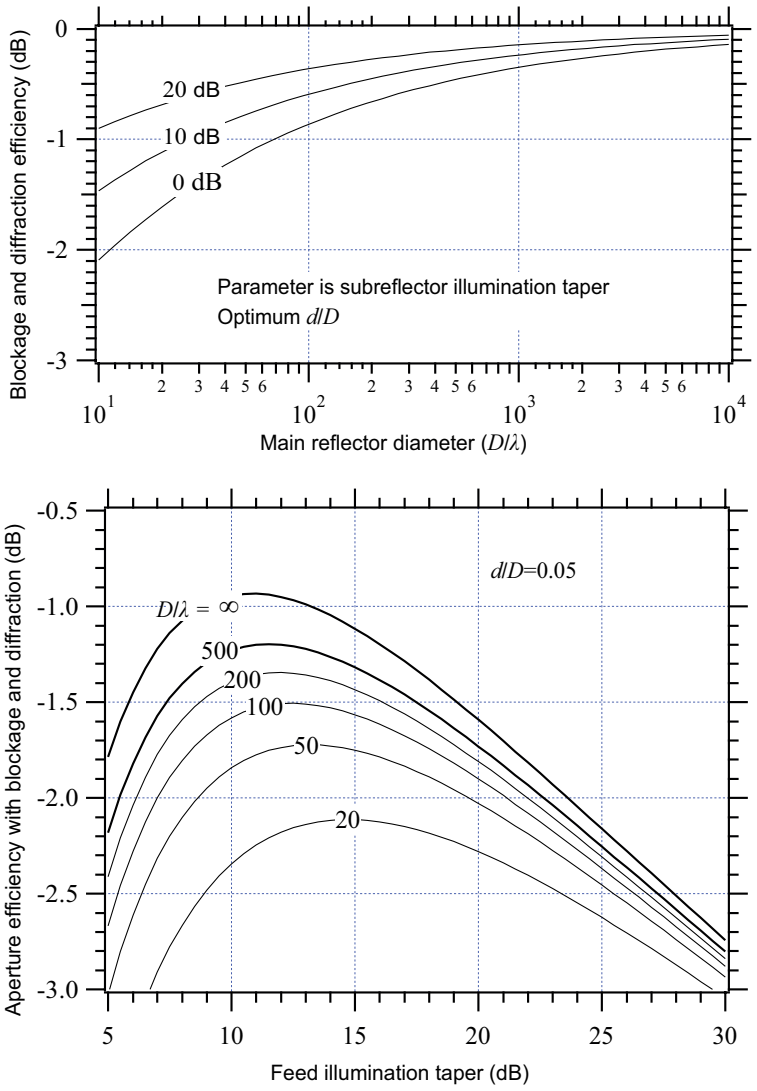


Figure 9.15: *Losses due to subreflector blockage and edge diffraction in Cassegrain. (a) Blockage and diffraction efficiency as a function of main reflector diameter for different illumination tapers A_0 in dB for the optimum choice of d/D. (b) Total aperture efficiency for $d/D = 0.05$ with the main reflector diameter D as parameter as a function of the feed illumination taper in dB.

diameter of the subreflector will therefore cause *edge diffraction* losses. In order to keep these losses as small as possible, the subreflector must have a diameter $d > 10\lambda/\sin(\Psi_0)$, where Ψ_0 is the subtended half angle of the paraboloid, as defined in Fig. 9.9. This can easily be seen from the formulas in the following text.

We will give an asymptotic formula for a combined *edge diffraction and blockage efficiency* $e_{\mathrm{b+d}}$ [6] which is derived by using the uniform geometry of diffraction (see reference [29] in Chapter 1 of this textbook). The total aperture efficiency is

$$e_{\mathrm{ap_{tot}}} = e_{\mathrm{ap}} e_{\mathrm{b+d}} \ , \tag{9.107}$$

where e_{ap} is the previous aperture efficiency in (9.62) and

$$e_{\mathrm{b+d}} = |1 - \Delta_{\mathrm{cb}} - \Delta_d|^2 \tag{9.108}$$

where the blockage term Δ_{cb} is the same as in (9.104), and the diffraction term

$$\Delta_d = (1-j) C_d \sqrt{\frac{\lambda}{d \sin \Psi_0}} \sqrt{1 - \frac{d}{D}} A_0 \ , \tag{9.109}$$

with D the main reflector diameter, d the subreflector diameter, A_0 the aperture illumination taper (amplitude), and

$$C_d = \frac{1}{\pi} \cos^2(\theta_0/2) C_b \ . \tag{9.110}$$

Some results based on these formulas are presented in Fig. 9.15*. It is possible to differentiate (9.108) with respect to d/D in order to determine the d/D ratio which maximizes the efficiency. This optimum d/D is used to generate the results in Fig. 9.15a.The result of such a differentiation gives $(d/D) \approx 0.05$ for most practical antenna sizes. For small main reflector diameters it may be necessary to increase this in order to illuminate the subreflector properly with as small spillover as possible, but we should not use $(d/D) > 0.10$. Then, multiple scattering effects may dominate over blockage and diffraction losses.

9.4.8 Example: Corrugated feed horn for Cassegrain antenna

We consider a classical Cassegrain antenna with 20 m diameter for operation at 8 GHz with an expected aperture efficiency of -1 dB (80 %) when aperture blockage is not accounted for. The subtended half angle seen from the focal point towards the subreflector is 15°.

a) Determine the expected directivity when blockage is not accounted for.

b) Use the expected aperture efficiency to determine the length and diameter of the shortest conical corrugated horn which can feed the Cassegrain.

c) Find where to locate the horn relative to the focal point of the Cassegrain.

SOLUTION:

a) The directivity becomes ($\lambda = 30/8 = 3.75$ cm):

$$(D_0)_{\mathrm{dBi}} = e_{\mathrm{ap}} \left(\frac{\pi D}{\lambda} \right)^2 = 0.8 \left(\frac{\pi \cdot 2000}{3.75} \right)^2 = 63.51 \, \mathrm{dBi} \ .$$

b) For a Cassegrain θ_0 is normally small, so that the aperture illumination taper becomes equal to the feed illumination taper. Then, to get -1 dB efficiency we need a taper of about 11 dB (see Fig. 9.10). The aperture taper is for quite small subtended angles equal to the feed taper. The shortest possible horn which can provide this is a maximum gain horn, see Section 8.9.5. We can use (8.84) to determine the length L of the horn, with $T = -11$ dB and $\theta = 15^\circ = 0.2618$ rad, i.e.,

$$L = -0.07\frac{(-11)}{(0.2618)^2}\lambda = 11\lambda = 41\,\text{cm}.$$

The horn diameter is determined from $L = 1.1a^2/\lambda$, i.e.,

$$D_{\text{hrn}} = 2a = 2\sqrt{\frac{\lambda L}{1.1}} = 2\sqrt{10}\lambda = 6.3\lambda = 23.6\,\text{cm}\ .$$

c) The phase center of a maximum gain horn is located at $Z_{\text{pc}} = -L/2$ from the aperture. This must coincide with the focal point of the Cassegrain.

9.5 Other reflector shapes

Modern reflector antennas have often surfaces that are shaped in order to provide a desired radiation pattern. They may be shaped to maximize the gain, reduce the sidelobes in particular regions or to generate a contoured beam. The art of designing *shaped reflectors* is referred to as *reflector synthesis*.

The original reflector synthesis is based on solving differential equations to obtain a desired aperture distribution, which is given from requirements on the directivity and first sidelobes of the far-field function. The aperture distribution should e.g. be uniform with constant phase to get the highest directivity[12]. The original reflector synthesis approach was to use *Geometrical Optics* (*GO*) reflections on the subreflector and main reflector, and formulate differential equations for two rotationally symmetric reflector surfaces [7]. This was used to maximize gain in the large ground stations for satellite communications used in the 1980s. The synthesis method was extended to asymmetric offset geometries in [8]-[9] in order to reduce the sidelobes due to the center blockage, and maintain the higher aperture efficiency and lower cross-polar sidelobes otherwise being present. A completely new approach to the reflector synthesis was introduced in [10]-[11] in connection with the design of a dual-reflector feed for the 300 m diameter spherical reflector of the radio-telescope in Arecibo [12]. The method made use of the laws in [13] that governs GO mapping of the far-field of the feed onto the aperture, and the modern GO ray tracing formulations [14]. This made it possible to avoid differential equations, and instead achieve the reflector shapes directly after the main geometry (relative locations and diameters of the reflectors) have been specified. The solution is done by solving repeatedly linear equations when progressing stepwise in circles from the central ray to the rim of the reflectors.

Multi-reflector antennas can be analyzed both by GO ray tracing and PO integration. The PO approach is more accurate, but takes long time. The results for the dual-reflector feed of

[12] See Section 7.5.

the Arecibo tri-reflector are compared in [12]. The GO analysis can be extended with UTD[13]. This becomes too complicated when there are more than two reflectors. However, it can be conveniently done by using a UTD-approach by which the width of a transition region is traced via several reflectors [15], and finally the edge diffraction efficiency can be evaluated in the aperture [16]. This approach is fast and shows good agreement with *PO integration* approach [17]. This approach is also easily applied to offset dual-reflector systems [18], and they correspond for the rotationally symmetric case to the edge diffraction efficiency result in Section 9.4.7.

Reflector synthesis can also include edge diffraction [19], normally by PO integration which is very flexible [20]. Reflector antennas for satellite applications with contoured or multiple beams can also be achieved by synthesizing the excitation of an array of feeds [20]. The feed can also be realized as a dense array, in which case we call it a Focal Plane Array (FPA). The FPA described in [21] is used to form a sector beam that can illuminate the subreflector of a Cassegrain system in such a way that the sensitivity of it improves. At the same time multiple beams can be achieved. It is also possible to synthesize sector beams for this purpose by using a lens in front of the horn aperture [22].

Reflector antennas can also be realized as *parabolic cylinders* with a linear array feed [23]. Then, diffraction efficiencies similar to those in [6] are observed due to diffraction from the ends of the line feed [24].

9.6 Prime-focus feeds

Corrugated horns[14] are normally very flexible to use as feeds for dual-reflector antennas, because the can be designed for any beam width with high performance in terms of bandwidth, cross-polarization and sidelobes. They are heavy though. Corrugated horns are also used as prime-focus feeds, and they can provide wide beams with constant width over almost octave bandwidth [25]. Some more feeds and their design principles are overviewed in [26]. We will briefly summarize them here.

The *hat feed* is a self-supported rear-radiating corrugated horn type feed. The design was originally proposed in [27] for satellite-TV reception. The first successful commercial application was for radio links [28]. The hat feed has a ring-shaped phase center, so the optimum reflector is a ring-focus paraboloid [29]. The bandwidth was originally very small, but it has finally reached 33 % which is very good for such a primary feed [30]. The hat feeds has been mass produced in more than one million copies for use in radio links between 5 and 38 GHz.

Dipole feeds are practical at frequencies below 5 GHz, but the usability can be stretched to higher frequencies. They can also be self-supported like the hat feed but requires a small ground plane. The E- and H-plane patterns will be very different, but they can be improved by a beam-forming ring [31]. They can also be used in small resonant reflector antennas, in which the directivity is improved by making use of constructive resonances between the feed and the parabolic reflector [32]. The E- and H-plane patterns of dipoles above ground plane can also be made equal by locating two dipoles in parallel (in eleven configuration). This is

[13] UTD is defined in Section 1.5.
[14] Corrugated horns are treated in Section 8.9.

used in the so called eleven antenna [33] in which also the two dipoles are realized as cascaded log-periodic folded dipoles. The eleven antenna is a very wideband feed with constant beam width 11 dBi directivity over more than decade bandwidth. The solution in [33] works between 2 and 13 GHz. The quad-ridge horn can also be used as a wideband feed, see reference [19] in Chapter 8. The reference list finally contains three books on reflector antennas and physical optics that has been important contributions [34]-[36].

9.7 Exercises to Chapter 9

1. **Reflector antenna for satellite-TV**: We shall now design a reflector antenna for satellite-TV reception at 12 GHz.

   a) We have a paraboloidal reflector and do not know its F/D or subtended half-angle θ_0. Show how you can determine F/D and θ_0 by measuring the diameter D and depth Δz of the paraboloid. Use $D = 90\,\text{cm}$ and $\Delta z = 15\,\text{cm}$.

   b) Use the results in Subsection 9.4.2 to determine the optimum taper of the $\cos^n(\theta/2)$ feed pattern that maximizes the aperture efficiency of the paraboloid. What is the taper of the aperture distribution?

   c) How large will the aperture of a TE_{11} mode conical horn feed be in order to provide this illumination taper?

   d) The actual feed pattern has sidelobes outside the first null in the pattern. The sidelobes are measured to be 18 dB lower than the main lobe. Assumed that this feed pattern can be modeled as

   $$G(\theta) = \begin{cases} \cos^n(\theta/2) & 0 < \theta < \theta_a \\ a & \theta_a < \theta < \pi \end{cases},$$

   where $20\log\cos^n(\theta_a/2) = 20\log|a| = -21\,\text{dB}$. In this way the -18 dB sidelobes are modeled as a uniform level of -21 dB from the -21 dB level of the main lobe and out to 180°. How much will these sidelobes reduce the spillover efficiency (in dB), and how much will the feed efficiency decrease? (Assume $\theta_0 < \theta_a$)

2. **Design of Cassegrain antenna**: Consider now a Cassegrain antenna with subtended half-angle of the subreflector of $\theta_0 = 15°$. We want to feed it with a circular corrugated horn antenna at 12 GHz.

   a) Assume that the horn has a $\cos^n(\theta/2)$ feed pattern and find the feed efficiency when the illumination taper is 20 dB. This strong taper is chosen to achieve low sidelobes.

   b) Use the universal radiation pattern for conical corrugated horns to determine the diameter of the horn? Is this horn aperture-controlled or flare angle-controlled? What is the length of the horn, if we require that the maximum phase error in the plane aperture should be less than 45°?

   c) Use design curves for the Gaussian beam model of corrugated horns to determine the length of the shortest horn that can provide the 20 dB taper at 15°?

   d) There is a dielectric lens across the aperture of the horn in order to make the phase front plane. However, this lens increases the sidelobe level of the horn due to reflections in the lens. Assume that the sidelobe level is uniform at -25 dB relative to the main lobe maximum, from the -25 dB point of the $\cos^n(\theta/2)$ pattern and out to 180°. Use the same form of the feed pattern as in Exercise 1-d, with $20\log\cos^n(\theta_a/2) = 20\log|a| = -25\,\text{dB}$. How much will now the aperture efficiency decrease due to the added spillover?

   e) The Cassegrain antenna has a diameter of about 10 m. What are the approximate relative sidelobe levels (relative to the main beam maximum) due to feed spillover, and in which angular region do they appear?

f) The phase center of the co-polar radiation pattern of the feed is measured when the feed is mounted on the measurement table with the z-position of its aperture over the rotation axis of the measurement table. The measured phase is proportional to θ_f^2 with a $\phi(\theta_0) - \phi(0) = -360^\circ$ where $\theta_0 = 15^\circ$. Where is the location of the phase center of the feed? Use the Gaussian beam approximation of the horn.

3. **Aperture efficiency formula**: Derive the expression for the aperture efficiency in (9.69) and the blockage constant in (9.106), for the theoretical $\cos^n(\theta/2)$ feed pattern.

9.8 References

[1] P.-S. Kildal, "Factorization of the feed efficiency of paraboloids and Cassegrain antennas", *IEEE Transactions on Antennas and Propagation*, Vol. AP-33, No. 8, pp. 903-908, August 1985.

[2] P.-S. Kildal, "Combined E- and H-plane phase-centers of antenna feeds", *IEEE Transactions on Antennas and Propagation*, Vol. AP-31, No. 1, pp. 199-202, January 1983,

[3] J. Ruze, "Antenna tolerance theory - A review", *Proceedings of the IEEE*, Vol. 54, pp. 633-640, April 1966.

[4] P.-S. Kildal, E. Olsen and J.A. Aas, "Losses, sidelobes and cross-polarization caused by feed-support struts in reflector antennas; Design curves", *IEEE Transactions of Antennas and Propagation*, Vol. AP-36, No. 2, pp. 182-190, February 1988.

[5] P.-S. Kildal, "A small dipole-fed resonant reflector antenna with high efficiency, low cross-polarization, and low sidelobes", *IEEE Transactions of Antennas Propagation*, Vol. AP-33, No. 12, pp. 1386-1391, December 1985.

[6] P.-S. Kildal, "The effects of subreflector diffraction on the aperture efficiency of a conventional Cassegrain antenna - An analytical approach", *IEEE Transactions of Antennas and Propagation*, Vol. AP-31, No. 6, pp. 903-909, November 1983.

[7] V. Galindo, "Design of dual reflector antennas with arbitrary phase and amplitude distributions", *IEEE Transactions on Antennas and Propagation*, Vol. AP-12, pp. 402-408, July 1964.

[8] G. Bjøntegaard, T. Pettersen, "An offset dual-reflector antenna shaped from near field measurements of the feed horn: Theoretical calculations and measurements", *IEEE Transactions on Antennas and Propagation*, Vol. AP-31, pp. 973-977, November 1983.

[9] V. Galindo-Israel, R. Mittra, "Synthesis of offset dual shaped subreflector antennas for control of Cassegrain aperture distributions", *IEEE Transactions on Antennas and Propagation*, Vol. 32, No. 1, pp. 86-92, January 1984.

[10] P.-S. Kildal, "Synthesis of multireflector antennas by kinematic and dynamic ray tracing", *IEEE Transactions on Antennas and Propagation*, Vol. 38, No. 10, pp. 1587-1599, October 1990.

[11] P-S. Kildal, "A new approach to the synthesis of reflector antennas", *Radio Science*, Vol. 26, No. 2, pp. 619-623, March-April 1991.

[12] P.-S. Kildal, L. Baker, T. Hagfors, "The Arecibo upgrading: Electrical design and expected performance of the dual-reflector feed system", *Proceedings of IEEE*, Vol. 82, No. 5, pp. 714-724, May 1994.

[13] P.-S. Kildal, "Laws of geometrical optics mapping in multi-reflector antennas with application to elliptical apertures", *IEE Proceedings*, Part H, MOA, No. 6, pp. 445-453, December 1989. (See also Erratum published in Vol. 137, Pt. H, No. 2, pp. 152, April 1990.)

[14] P.-S. Kildal, "Analysis of numerically specified multireflector antennas by kinematic and dynamic ray tracing", *IEEE Transactions on Antennas and Propagation*, Vol. 38, No. 10, pp. 1600-1606, October 1990.

[15] P.-S. Kildal and J. Stamnes, "Asymptotic transition region theory for edge diffraction. Part 1: Tracing transition regions via reflectors", *IEEE Transactions on Antennas and Propagation*, Vol. 38, No. 9, pp. 1350-1358, September 1990.

[16] P.-S. Kildal, "Asymptotic transition region theory for edge diffraction. Part 2: Calculation of diffraction losses in multi-reflector antennas", *IEEE Transactions on Antennas and Propagation*, Vol. 38, No. 9, pp. 1359-1365, September 1990.

[17] P.-S. Kildal, S.A. Skyttemyr, "Diffraction analysis of a proposed dual-reflector feed for the spherical reflector antenna of the Arecibo Observatory", *Radio Science*, Vol. 24, No. 5, pp. 601-617, September-October 1989.

[18] D.I.L. de Villiers, R. Lehmensiek, "Analytical Evaluation of the Efficiency Improvement of Shaped over Classical Offset Dual-Reflector Antennas Including Sub-Reflector Diffraction", *IEEE Antennas and Wireless Propagation Letters*, Vol. 10, pp. 947-950, 2011.

[19] D. Dah-Weih Duan, Y. Rahmat-Samii, "A generalized diffraction synthesis technique for high performance reflector antennas", *IEEE Transactions on Antennas and Propagation*, Vol. 43, No. 1, pp. 27–40, January 1995.

[20] S.K. Rao, "Parametric Design and Analysis of i Multiple-Beam Reflector Antennas for Satellite Communications", *IEEE Antennas and Propagation Magazine*, Vol. 45, No. 4, August 2003.

[21] M.V. Ivashina, M. Ng Mou Kehn, P.-S. Kildal and R. Maaskant, "Decoupling efficiency of a wideband Vivaldi focal plane array feeding a reflector antenna", *IEEE Transactions on Antennas and Propagation*, Vol. 57, No. 2, pp. 373-382, February 2009.

[22] P.-S. Kildal and K. Jakobsen, "Scalar horn with shaped lens improves Cassegrain efficiency", *IEEE Transactions on Antennas and Propagation*, Vol. 32, No. 10, pp. 1094-1100, October 1984.

[23] P.-S. Kildal, "Radiation characteristics of the EISCAT VHF parabolic cylinder antenna", *IEEE Transactions on Antennas and Propagation*, Vol. AP-32, No. 6, pp. 541-552, June 1984.

[24] P.-S. Kildal, "Diffraction corrections to the cylindrical wave radiated by a linear array feed of a cylindrical reflector antenna", *IEEE Transactions on Antennas and Propagation*, Vol. 32, No. 10, pp. 1111-1116, October 1984.

[25] Z. Ying, P.-S. Kildal, and A. Kishk, "A broadband compact horn feed for prime-focus reflectors", *Electronics Letters*, Vol. 31, 14, pp. 1114,1115, 6th July 1995.

[26] P.-S. Kildal, "Some Reflector and Feed Antenna Inventions that Made a Difference - Fundamentals and examples from radio telescopes, satellite communications and radio links", *IEEE Africon 2013*, Mauritius, 9-12 September 2013.

[27] P.-S. Kildal, "The hat feed: a dual-mode rear-radiating waveguide antenna having low cross polarization", *IEEE Transactions on Antennas and Propagation*, Vol. AP-35, pp. 1010-16, 1987.

[28] J. Hansen, A.A. Kishk, P.-S. Kildal, and O. Dahlsjo, "High performance reflector hat antenna with very low sidelobes for radio-link applications", *IEEE Antennas and Propagation Symposium*, New York, NY, USA, pp. 893-896, 1995.

[29] J. Yang and P.-S. Kildal, "Calculation of ring-shaped phase centers of feeds for ring-focus paraboloids", in IEEE Transactions on Antennas and Propagation, Vol. 48, pp. 524-8, 2000.

[30] M. Denstedt, T. Ostling, Y. Jian, and P.-S. Kildal, "Tripling bandwidth of hat feed by genetic algorithm optimization", in IEEE Antennas and Propagation Symposium, Piscataway, NJ, USA, pp. 2197-200, 2008.

[31] P.-S. Kildal and S.A. Skyttemyr, "Dipole-disk antenna with beamforming ring", *IEEE Transactions on Antennas and Propagation*, Vol. AP-30, No. 4, pp. 529-534, July 1982.

[32] P.-S. Kildal, "A small dipole-fed resonant reflector antenna with high efficiency, low cross polarization and low sidelobes", *IEEE Transactions on Antennas and Propagation*, Vol. AP-33, No. 12, pp. 1386-1391, December 1985.

[33] J. Yang, M. Pantaleev, P.-S. Kildal, B. Klein, Y. Karandikar, L. Helldner, N. Wadefalk, C. Beaudoin, "Cryogenic 2-13 GHz Eleven Feed for Reflector Antennas in Future Wideband Radio Telescopes", *IEEE Transactions on Antennas and Propagation*, Vol. 59, No. 6, pp. 1918-1934.

[34] A.W. Love (Ed.), *Reflector Antennas*, IEEE Reprints Series, IEEE Press, 1978.

[35] L. Diaz, and T. Milligan, *Antenna Engineering Using Physical Optics*, Artech House, 1996.

[36] R. Levy, *Structural Engineering of Microwave Antennas: For Electrical, Mechanical and Civil Engineers*, IEEE Press, 1996.

Chapter 10

Array antennas

Until now we have only considered single antennas or at the maximum two neighboring antennas or an antenna and its image in a ground plane. In the present chapter, we show how to analyze several interfering antennas, i.e., an array antenna. The chapter is mainly devoted to *linear arrays*.

If antennas with large directivity are needed, they must have apertures which are large in terms of wavelengths. High gain antennas can conveniently be realized as reflector antennas, but arrays are often preferred because they can be flat and thereby occupy less space than a reflector. Another advantage of an array is that the main beam direction can be rapidly changed by electronic steering of the phase of each element in the array. It is also possible to control the shape of the beam by steering the phase. The best radiation pattern control is obtained by steering both the phase and amplitude of each element (Fig. 10.1). Even multiple beams can be created. Full flexibility is obtained in a so-called signal processing antenna with digital beam-forming capability. The elements of an array may be located with a spacing which is different for different positions in the array, but the most common is to use a uniform spacing. This book only treats arrays with uniform element spacing.

The disadvantage of the array solution is that it is normally much more expensive to de-

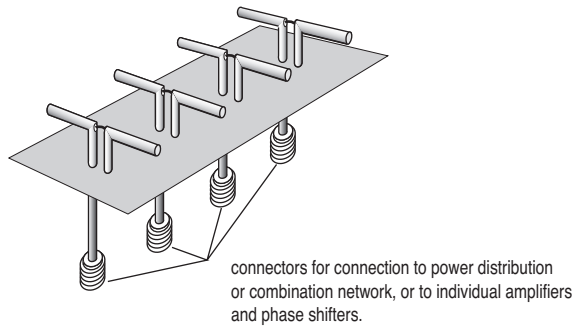


Figure 10.1: Linear array of dipoles where each element can be fed individually with different amplitude and phase.

sign and manufacture than a reflector antenna, in particular if fast steering of the beam or other forms of beam-control are required. It has also normally narrower bandwidth than a reflector.

Arrays can be linear, planar or conformal. In linear arrays the elements are located along a straight line, in planar arrays they are distributed over a plane surface, and in conformal arrays they are distributed over and conform to a single or double curved surface. Other configurations are also possible. A planar array can often be considered as a linear array of linear arrays, so the theory of linear arrays is valid also for planar arrays.

An array antenna can be designed for *full scan* or *limited scan*. Full scan means normally $\pm 60^\circ$ or more from broadside, whereas limited scan can be anything smaller than this. The full scan capability is difficult to obtain, and requires that the elements are located as close as half a wavelength from each other.

The elements of an array may be of any type. The most popular elements are dipoles, waveguide slots, microstrip patches, open waveguides or horns. Big reflector antennas may also be located in an array, such as in certain radio telescopes. Then, the reflector antenna elements will be movable relative to each other, so that the ambiguity due to multiple main lobes (also called grating-lobes) can be removed by repeated measurements for different antenna spacings. Such array systems are referred to as interferometers.

The elements of an array may be fed individually from a *power distribution network* (or *beam-forming network*) (Fig. 10.2, right). The elements of a linear array can also be fed in series from the same transmission line, such as a waveguide or a microstrip line (Fig. 10.2, left). In this latter case, we distinguish between three different types of arrays (Fig. 10.3); the resonant array, the travelling wave array and the leaky wave array:

a) In the resonant array the transmission line is short-circuited at the end such that a standing wave is formed along the transmission line. The elements of the array are all equal, weakly coupled to the line, and located at each standing wave peak. The *resonant array* is easy to design and the main beam is always broadside.The return loss at the input of the transmission line can be tuned to a low value by a matching network (e.g., irises or screws) at the input to the transmission line, but the bandwidth of the return loss becomes narrow.

b) In the *travelling wave array* the transmission line is terminated with a matched dummy load at its end in order to avoid the standing wave. The first elements are weakly excited whereas the last elements are strongly excited, in order to minimize the power absorbed in the dummy load. The travelling wave array has better return loss bandwidth than the resonant array. In addition the main beam can be designed to point in any direction. However, the pointing direction will change with frequency.

c) Several elements along a transmission line may also be designed as a *leaky wave antenna*. In this case the element spacing is smaller than a half wavelength. The elements work as a periodic perturbation of the transmission line and cause a change in the propagation constant along the line.

The present analysis will be based on the so-called *isolated element approach* [1, p. 424]. This means that the *shape* of the electric or magnetic current distribution of each antenna element is independent of how the element is excited, i.e., the current distribution function is the same when the element is excited via mutual coupling as when it is excited by a terminal

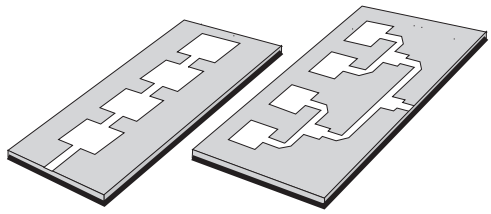


Figure 10.2: Series-fed (left) and parallel-fed (right) linear microstrip patch arrays.

voltage or current. This is true only for single mode antenna elements such as half-wave dipoles, slots, patches and open basic mode waveguides.

In these antennas, the shape of the current distributions or the aperture field distributions do not depend on the surroundings.

10.1 Linear array of equispaced elements

Consider N equispaced and equal antenna elements which are oriented in the same direction and located with their individual phase reference points at

$$\mathbf{r}_n = \mathbf{r}_c + a_n\hat{\mathbf{a}} = \mathbf{r}_c + \left(n - \frac{N+1}{2}\right) d_a\hat{\mathbf{a}} \qquad \text{for } n = 1, 2, ..., N \ , \tag{10.1}$$

where $\mathbf{r}_c = (\mathbf{r}_{_N} + \mathbf{r}_{_1})/2$ is the geometrical center of the array, $\hat{\mathbf{a}}$ is a unit vector defining the direction of the array axis, and d_a is the element spacing (Fig. 10.4). We choose to define the total length L of the array by extending it to a $d_a/2$ length outside the first and the last element, so that

$$L = Nd_a \ . \tag{10.2}$$

The far-field function of element number n at a point $\mathbf{r}$ is

$$\mathbf{E}_n(\mathbf{r}) = \frac{1}{r}e^{-jkr}\mathbf{G}(\hat{\mathbf{r}})e^{jk\mathbf{r}_n\cdot\hat{\mathbf{r}}} \ , \tag{10.3}$$

where $\mathbf{G}(\hat{\mathbf{r}})$ is the far-field function when referred to the phase reference point $\mathbf{r}_n$ of the element, and where $\mathbf{G}(\hat{\mathbf{r}})e^{jk\mathbf{r}_n\cdot\hat{\mathbf{r}}}$ is the far-field function when referred to the origin of the coordinate system of the whole array antenna. The latter is easily obtained by using Eq. (2.52) to move the phase reference point from $\mathbf{r}_n$ to the origin, i.e., by using $\mathbf{r}_{_A} = \mathbf{r}_n$.

10.1.1 Array factor as an element-by-element sum

From (10.1)-(10.3) by using superposition the far-field function of the whole array becomes

$$\mathbf{G}_{_A}(\hat{\mathbf{r}}) = \sum_{n=1}^{N} A_n e^{j\Phi_n}\mathbf{G}(\hat{\mathbf{r}})e^{jk\mathbf{r}_n\cdot\hat{\mathbf{r}}} \ , \tag{10.4}$$

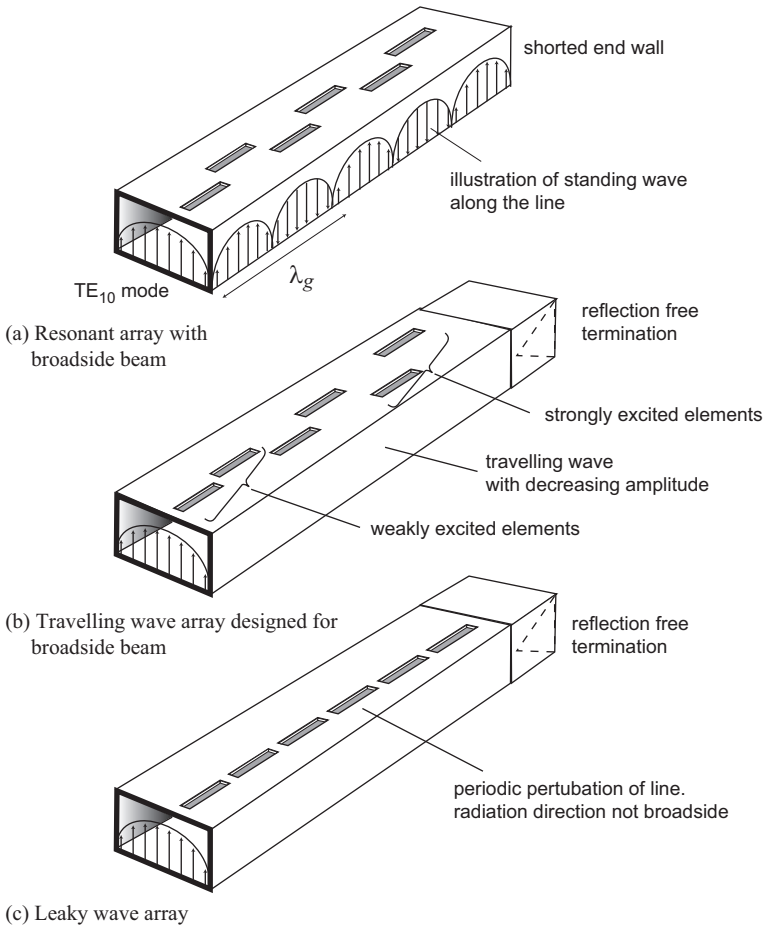


Figure 10.3: Examples of linear waveguide slot array: resonant, travelling wave and leaky wave types.

where A_n is the amplitude and Φ_n is the phase of the current or voltage excitation of element number n. If we assume that all elements are identical and have the same far-field function, this can be written more conveniently as [1]

$$\mathbf{G}_{A}(\hat{\mathbf{r}}) = \mathbf{G}(\hat{\mathbf{r}})\mathsf{AF}(\hat{\mathbf{r}}) \ , \tag{10.5}$$

where

$$\mathsf{AF}(\hat{\mathbf{r}}) = \sum_{n=1}^{N} A_n e^{j\Phi_n} e^{jk\mathbf{r}_n\cdot\hat{\mathbf{r}}} \tag{10.6}$$

is referred to as the *array factor* being expressed as an *element-by-element sum*. Thus, the far-field function of an array of equal and co-oriented elements is the product of the element

[1] This formulation is correct if $\mathbf{G}(\hat{\mathbf{r}})$ is the embedded far-field function of an element, and all elements have identical embedded far-field functions. In practice this is the case for elements of large regular arrays, except for the outer two rows of elements.

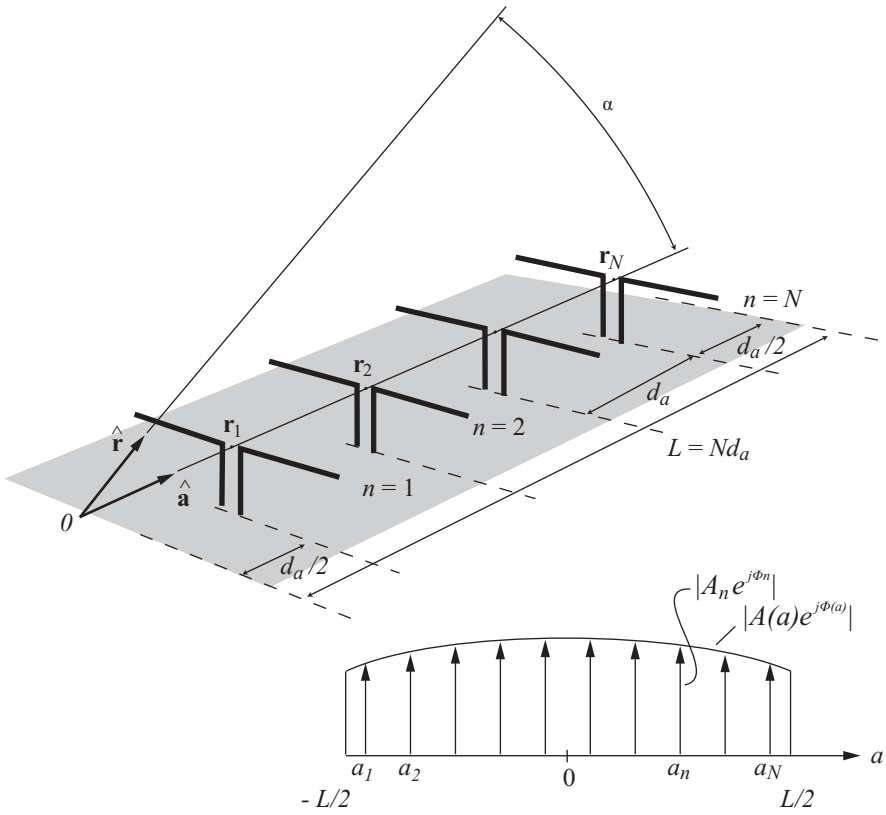


Figure 10.4: Geometry of equispaced linear array (upper), and illustration of element excitations and the excitation distribution (lower).

factor (i.e., the far-field function of the element) and an array factor. The element-by-element sum can readily be evaluated for small arrays[2], but for large arrays is becomes time-consuming and gives also little understanding of how arrays work. Therefore, we will also introduce the infinite grating-lobe sum, which is better in this respect, see Subsection 10.1.3.

10.1.2 Array factor for uniform amplitude and linear phase

We assume now that all elements have the same amplitude excitation, i.e., $A_n = A$. Furthermore, we assume that the phase excitation (in radians) progresses linearly along the array according to

$$\Phi_n = \Phi_c - \left(n - \frac{N+1}{2}\right) k_\Phi d_a \qquad \text{with } k_\Phi = -\Delta\Phi/d_a \ , \tag{10.7}$$

where $\Delta\Phi$ in radians is the phase difference between neighboring elements, and Φ_c is a constant phase offset[3]. It is important always to choose $\Delta\Phi$ in the domain $-\pi < \Delta\Phi < \pi$,

[2] There exist MATLAB code for all figures of which the caption start with $\star$.

[3] This corresponds to the phase at the geometrical center $\mathbf{r}_c$ of the array.

otherwise the main beam will appear as a grating-lobe in the equations below. The k_Φ represents a propagation constant for the phase progression along the array. The *array factor* becomes

$$\mathsf{AF}(\hat{\mathbf{r}}) = Ae^{j(\Phi_c + k\mathbf{r}_c\cdot\hat{\mathbf{r}})}\sum_{n=1}^{N} e^{j\left(n-\frac{N+1}{2}\right)\Delta\Psi} \ , \tag{10.8}$$

where we have $\Delta\Psi = (k_a - k_\Phi)d_a$ with

$$k_a = k\hat{\mathbf{a}}\cdot\hat{\mathbf{r}} = k\cos\alpha \tag{10.9}$$

the product of k and the projection of the observation direction $\hat{\mathbf{r}}$ along the direction $\hat{\mathbf{a}}$ of the array. The α is the angle between $\hat{\mathbf{r}}$ and $\hat{\mathbf{a}}$ and defines a cone around $\hat{\mathbf{a}}$. The sum expression is a geometrical series and can be summed by using a standard formula. The result is

$$\sum_{n=1}^{N} e^{j\left(n-\frac{N+1}{2}\right)\Delta\Psi} = e^{-j\left(\frac{N-1}{2}\right)\Delta\Psi}\frac{(e^{jN\Delta\Psi}-1)}{(e^{j\Delta\Psi}-1)} = \frac{\sin(N\Delta\Psi/2)}{\sin(\Delta\Psi/2)} \ . \tag{10.10}$$

This finally gives

$$\mathsf{AF}(\hat{\mathbf{r}}) = A\cdot N\frac{\sin(N(k_a-k_\Phi)d_a/2)}{N\sin((k_a-k_\Phi)d_a/2)}e^{jk\mathbf{r}_c\cdot\hat{\mathbf{r}}}e^{j\Phi_c} \ . \tag{10.11}$$

The array factor of an array with length $L = Nd_a$ and $\Delta\Phi = 0$ is plotted in Fig. 10.5* for different number of elements N. For large N, it approaches the main lobe and near-in sidelobe characteristics of the sinc funtion[4]

$$\sin(N(k_a-k_\Phi)d_a/2)/(N(k_a-k_\Phi)d_a/2) \ .$$

Eq. (10.11) will be studied carefully in Subsections 10.1.3 to 10.1.8. However, we will first introduce an alternative expression for the array factor in which we make use of Fourier transforms, in the same way as we did when analyzing straight wire antennas and plane apertures.

10.1.3 Array factor as a grating-lobe sum

The element-by-element sum expression for the array factor in (10.6) is valid for any amplitude A_n and phase ϕ_n excitation of the array. We will now derive an alternative sum expression for the array factor, which also is valid for any A_n and ϕ_n. The alternative expression will be referred to as an infinite grating-lobe sum and is obtained by a Fourier transform of a smooth and continuous excitation distribution $A(a)$ which represent the distribution of A_n along the array. In order to do this we define the element locations by $\mathbf{r}(a_n) = \mathbf{r}_n$ where $\mathbf{r}(a)$ is a continuous function over the interval $-L/2 < a < L/2$ defined by (see Fig. 10.4)

$$\mathbf{r}(a) = \mathbf{r}_c + a\hat{\mathbf{a}} \ . \tag{10.12}$$

The continuous variable a takes on the following values at the element locations:

$$a_n = -(L/2) + \left(n-\frac{1}{2}\right)d_a \qquad \text{for } n = 1, 2, ..., N \ . \tag{10.13}$$

[4] See Section 7.4.3.

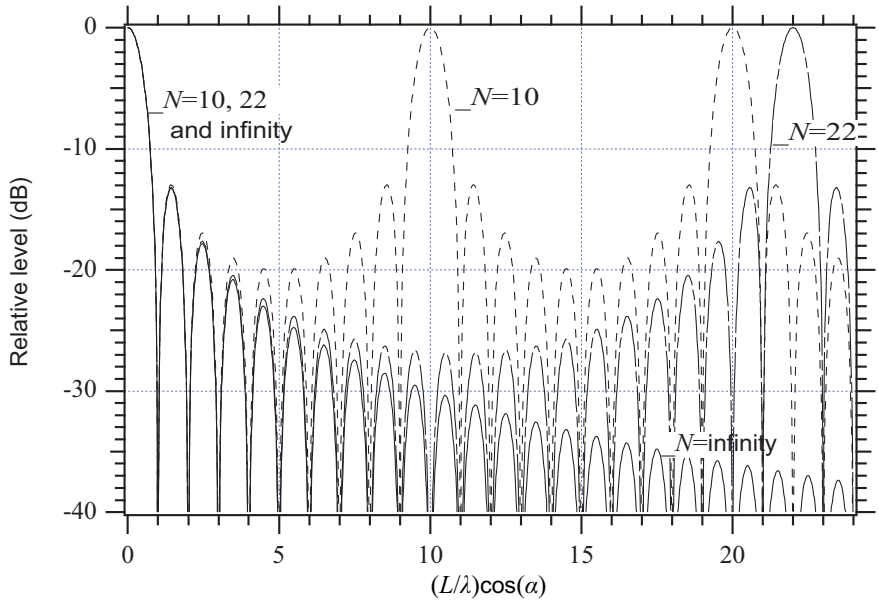


Figure 10.5: *Universal radiation pattern representing the array factor for a *uniformly excited linear array* of length $L = Nd_a$. The parameter is the number of elements N. The solid curve represent an infinite number of elements distributed along L, corresponding to a uniform continuous aperture distribution.

We see that $\mathbf{r}(-L/2)$ and $\mathbf{r}(L/2)$ define the ends of the array (see Fig. 10.4). We then define $A(a)$ to be smooth and continuous within $|a| < L/2$ and to take on the values

$$A(a) = \begin{cases} A_n & \text{for } a = a_n \ , \ n = 1, 2, ..., N \\ 0 & \text{for } |a| > L/2 \end{cases} \quad . \tag{10.14}$$

Similarly, we introduce $\Phi(a_n) = \Phi_n$ in (10.7) with

$$\Phi(a) = \Phi_c - k_\Phi a \ . \tag{10.15}$$

Eq. (10.15) describes a pure linear phase variation along the array, which may look like a restriction. However, if the actual phase variation is more complex, we may include the remaining phase variation in $A(a)$ by defining A_n as complex constants. By using the above equations and the sampling properties of the delta function, we can write the general array factor in (10.6) in terms of $A(a)$ as follows

$$\mathsf{AF}(\hat{\mathbf{r}}) = e^{jk\hat{\mathbf{r}}\cdot\mathbf{r}_c} e^{j\Phi_c} \sum_{n=1}^{N} \left\{ \int_{-\infty}^{\infty} A(a) e^{-jk_\Phi a} \delta(a - a_n) e^{jk_a a} da \right\} \ . \tag{10.16}$$

We now interchange the integral and summation signs and obtain

$$\mathsf{AF}(\hat{\mathbf{r}}) = e^{jk\hat{\mathbf{r}}\cdot\mathbf{r}_c} e^{j\Phi_c} \int_{-\infty}^{\infty} A(a) e^{j(k_a - k_\Phi)a} \sum_{n=-\infty}^{\infty} \delta(a - a_n) da \ . \tag{10.17}$$

We have here changed the summation boundaries from $(1, N)$ to $(-\infty, \infty)$. This is allowed because $A(a)$ is zero for $a < -L/2$ and $a > L/2$, which corresponds to $n < 1$ and $n > N$,

respectively. The sum of delta functions in (10.17) forms a periodic function with period d_a, so we can expand it in a Fourier series with exponential terms of the form $e^{-jp2(\pi/d_a)a}$ for $p = -\infty, ..., 0, ..., \infty$. All the coefficients of the Fourier series become 1 when N is odd (i.e., when the center r_c coincides with an element), and every second coefficient become $+1$ and -1 when N is even (i.e., when the center r_c is halfway between two elements). Therefore, the following formula containing N is valid for both these cases (this is called the Poisson summation formula);

$$\sum_{n=-\infty}^{\infty} \delta(a - a_n) = \frac{1}{d_a} \sum_{p=-\infty}^{\infty} (-1)^{p(N-1)} e^{-jp(2\pi/d_a)a} \ . \tag{10.18}$$

Inserting this into (10.17) and interchanging the summation and integral signs again give

$$\mathsf{AF}(\hat{\mathbf{r}}) = e^{jk\hat{\mathbf{r}}\cdot\mathbf{r}_c} e^{j\Phi_c} \frac{1}{d_a} \sum_{p=-\infty}^{\infty} (-1)^{p(N-1)} \tilde{A}\left(k_a - k_\Phi + p\frac{2\pi}{d_a}\right) \ , \tag{10.19}$$

where $k_a = k\cos\alpha$, and where

$$\tilde{A}(k_a) = \int_{-\infty}^{\infty} A(a) e^{jk_a a} da = \int_{-L/2}^{L/2} A(a) e^{jk_a a} da \ , \tag{10.20}$$

is the Fourier transform of the amplitude excitation distribution $A(a)$ which is zero outside the interval $-L/2 < a < L/2$.

Thus, by (10.19) and (10.20) we have expressed the array factor in terms of the Fourier transform of the continuous amplitude excitation distribution. This is related to the results of previous chapters as follows:

1. The far-field function of the straight wire[5] was found to be the product of the Fourier transform of the current distribution and the far-field function of the incremental electric current.

2. The far-field function of the rectangular aperture[6] was found to be the product of the Fourier transform of the E-field aperture distribution and the far-field function of either an incremental magnetic current or a Huygens source.

3. Presently, we have shown that the far-field function of a linear array is a product of the far-field function of the array element and a factor which is a sum of displaced Fourier transforms of the amplitude excitation distribution, each one centered around an interference maximum which we[7] will refer to as a grating-lobe.

Thus, the present periodic excitation distribution has caused an infinite sum of displaced Fourier transforms, each one describing the shape of the array factor around a direction corresponding to $k_a = k_\Phi - p(2\pi/d_a)$. The alternative sum expression for the array factor in (10.19) may be named an *infinite grating-lobe sum*. For infinite arrays it corresponds to what is called a sum over *Floquet modes*. The formula in (10.19) has been obtained from (10.6) without approximations, but it is only valid for a linearly progressive phase excitation. For instance, when $A(a)$ is uniform, (10.19) gives exactly the same result as (10.11), but the two

[5] It is given in Section 5.1.3.
[6] See Sections 7.3 and 7.4.
[7] For more information see Section 10.1.5.

formulations have completely different form. The advantage of (10.19) compared to (10.11) is that (10.19) is valid for any amplitude excitation distribution. The advantage with the infinite grating-lobe sum in (10.19) compared with the element-by-element sum in (10.6) is that Eq. (10.19) is much faster to evaluate for large arrays, i.e., when the number of array elements is large. The infinite sum in (10.19) converges very fast for large arrays (large N). Often only the dominant $p = 0$ term or the two terms for which $p = 0$ and $p = 1$ (or $p = -1$) need to be included. In the next sections we will study both (10.11) and (10.19).

The excitation distribution will most often be a real function with some taper at the ends $a = \pm L/2$. If so, the Fourier transform $\tilde{A}(k_a)$ will have a maximum for $k_a = 0$. If $A(a) = 1$, we achieve the sinc function

$$\tilde{A}(k_a) = L\frac{\sin(k_a L/2)}{k_a L/2} \ , \tag{10.21}$$

which is the same as the curve marked $N = \infty$ in Fig. 10.5. The characteristics of this radiation pattern is given in Section 7.4.3. For tapered excitations $\tilde{A}(k_a)$ will have lower sidelobe levels. An example of a tapered excitation is given in Section 7.5.3.

10.1.4 Steered main lobe

If we only include the $p = 0$ term in (10.19), the array factor has the form

$$\mathsf{AF}(\hat{\mathbf{r}}) \approx e^{jk\mathbf{r}_c\cdot\hat{\mathbf{r}}}e^{j\Phi_c}\frac{1}{d_a}\tilde{A}(k\cos\alpha - k_\Phi) \ .$$

This has a maximum when $k_a = k\cos\alpha = k_\Phi$, i.e., when

$$\hat{\mathbf{r}}\cdot\hat{\mathbf{a}} = \cos\alpha_0 = k_\Phi/k = -\Delta\Phi/(kd_a) \ , \tag{10.22}$$

where α_0 defines a cone around the $\hat{\mathbf{a}}$ axis. If we align the array with the z-axis so that $\hat{\mathbf{a}} = \hat{\mathbf{z}}$, we see that this corresponds to

$$\cos\theta_0 = k_\Phi/k = -\Delta\Phi/(kd_a) \ , \tag{10.23}$$

where $\Delta\Phi$ in radians is the phase progression from element to element, see (10.7). The special array factor in (10.11) has a maximum for the same angle.

The lobe around this $p = 0$ direction is the *main lobe* or main beam. We see that we can steer the main lobe by changing the phase progression $\Delta\Phi$ along the array. When $\Delta\Phi = 0$ we have $\alpha_0 = 90°$. This is referred to as a *broadside array* because it radiates normal to the array axis, see the left drawing in Fig. 10.6. When $\Delta\Phi = -kd_a = -2\pi d_a/\lambda$, we have $\alpha_0 = 0°$. This is referred to as an *endfire array*. It radiates along the array axis as shown in the right drawing in Fig. 10.6. A condition for endfire radiation is that the element pattern is not zero in that direction.

According to (10.5), the φ-variation of the radiation field on the α_0 cone will be determined by the far-field function of the elements, i.e., the element factor.

10.1.5 Graphical representation of array factor

The array factors in both (10.11) and (10.19) are functions of $k_a = k\hat{\mathbf{r}}\cdot\hat{\mathbf{a}}$. The k_a is constant on a cone around the array axis defined by $\hat{\mathbf{r}}\cdot\hat{\mathbf{a}} = \cos\alpha$ where α is the angle between the

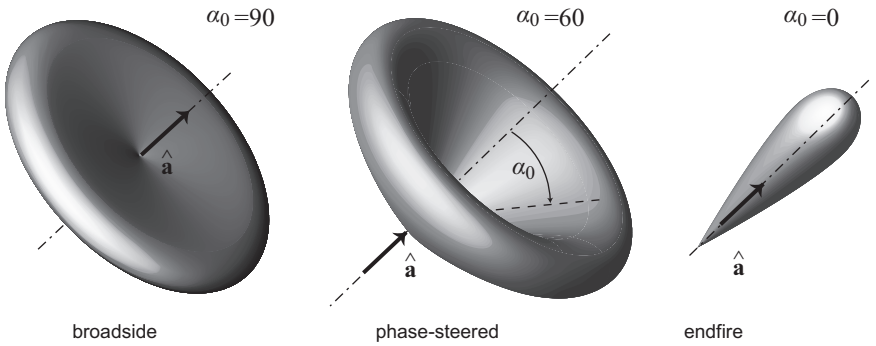


Figure 10.6: Illustration of main beam cone for broadside array $\alpha_0 = 90°$, phase-steered array with $\alpha_0 = 60°$, and endfire array $\alpha_0 = 0°$.

vectors $\hat{\mathbf{a}}$ and $\hat{\mathbf{r}}$. If we locate the array along the z-axis we get $\hat{\mathbf{a}} = \hat{\mathbf{z}}$ and $\cos\alpha = \cos\theta$, with θ the polar angle in the spherical coordinate system. If we locate the array along the x-axis (or y-axis) we get $\hat{\mathbf{a}} = \hat{\mathbf{x}}$ (or $\hat{\mathbf{a}} = \hat{\mathbf{y}}$) and $\cos\alpha = \sin\theta\cos\varphi$ (or $\cos\alpha = \sin\theta\sin\varphi$).

It is convenient to represent the array factor in a diagram as a function of k_a or preferably as a function of $\cos\alpha$ (Fig. 10.7a*). The functions are most conveniently normalized to a maximum of unity when $\alpha = \alpha_0$. The curves can conveniently be extended outside $-1 < \cos\alpha < 1$ in order to pick up the first maxima of $|\mathsf{AF}(\cos\alpha - \cos\alpha_0)|$ outside these boundaries. However, it is important to note that only $|\cos\alpha| < 1$ represents visible directions in the real physical space. The region for which $|\cos\alpha| > 1$ is invisible. The location of the main lobe of the array factor within $|\cos\alpha| < 1$ varies with the choice of α_0, i.e., the main beam direction. The array factor needs to be multiplied with the element factor (see (10.5) and Fig. 10.7b) in order to obtain the total far-field function. This can conveniently be done by plotting the normalized element factor in the same figure as the array factor (Fig. 10.7c and d). When α_0 varies, the array factor is simply translated along the α-axis. The shape of the array factor does not change with α_0 when plotted as a function of $\cos\alpha$. However, it is very important to be aware that the array factor lobe widths in degrees change with α_0. The lobes are much wider in degrees near endfire ($\alpha = 0°$ and $\alpha = 180°$) than near broadside.

10.1.6 Grating-lobes

Let us study the two versions of the array factor in (10.11) and (10.19). The array factor in (10.11) has its main lobe maximum at $\alpha = \alpha_0$. In addition there are maxima when

$$(k_a - k_\Phi)d_a = -p2\pi \qquad \text{for } p = \pm1, \pm2, \ldots . \tag{10.24}$$

The m'th term in (10.19) has a maximum for the same k_a, i.e., when $p = m$. Eq. (10.24) can also be written as

$$\cos\alpha_p = \cos\alpha_0 - p\frac{\lambda}{d_a} . \tag{10.25}$$

The directions α_ps define grating-lobe cones around the array axis. They are normally undesired and can be avoided by choosing the element spacing sufficiently small that the

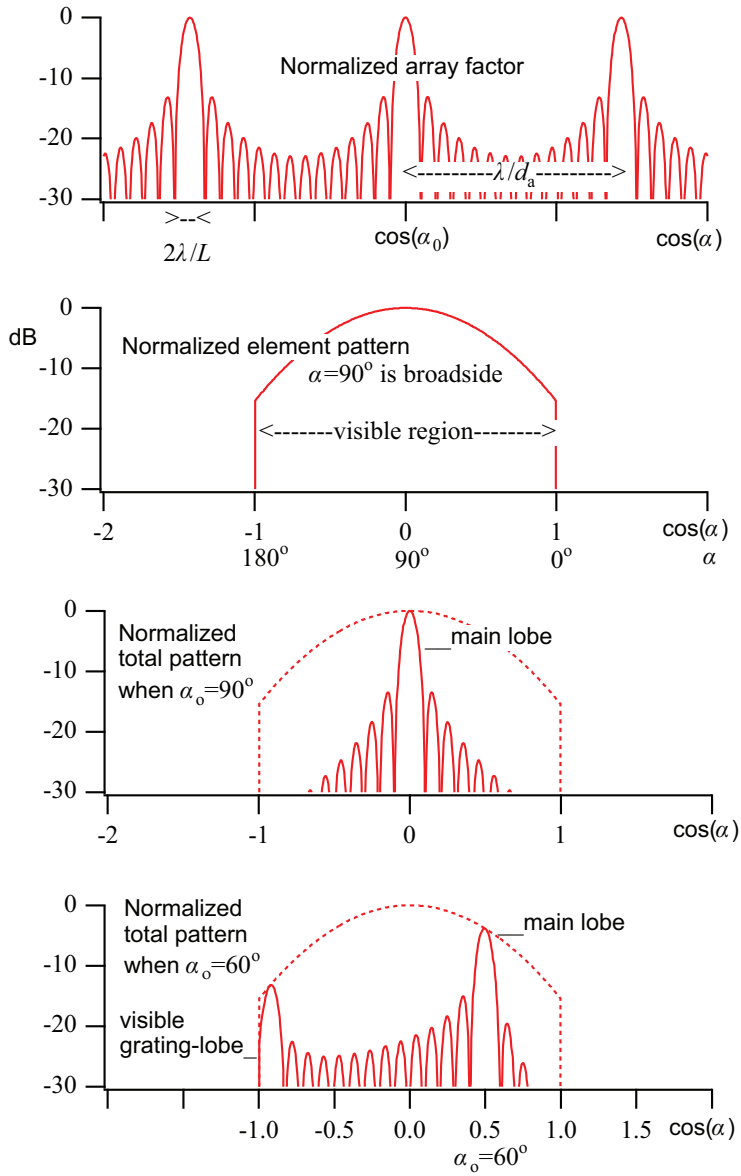


Figure 10.7: *Different contributions to the radiation pattern of a linear array in a plane through the array axis for $d_a \approx 0.7\lambda$ and $L \approx 10\lambda$. (a) Normalized array factor as a function of $\cos\alpha$. (b) Normalized radiation pattern of the element. (c) Combined radiation pattern when the main beam direction is at broadside (i.e., $\alpha_0 = 90°$. (d) Combined radiation pattern when the main beam is at 30° from broadside (i.e., $\alpha_0 = 60°$).

grating-lobes appear in the invisible region $|\cos\alpha_p| > 1$ for all desired α_0. By using (10.25) this *requirement for nonradiating grating-lobes* becomes

$$d_a \leq \frac{\lambda}{1 + |\cos\alpha_0| + (\lambda/L)} \ , \tag{10.26}$$

where L is the length of the array. To obtain this we have assumed that the half-width of the grating-lobe is λ/L which corresponds to the case of a uniform excitation distribution $A(a) = 1$. The condition is easily seen by studying Fig. 10.7d. We see from (10.26) that the elements must be located closer than 1λ in order to avoid grating-lobes in a long broadside array. For a long endfire array the requirement is closer than 0.5λ.

Grating-lobe problems can also be avoided by suppressing them with the element pattern.

10.1.7 Sidelobes

The near-in sidelobes of an array is mainly determined by the array factor, i.e., by the sidelobes of the Fourier transform $\tilde{A}(k_a)$ of the excitation function. The levels of the near-in sidelobes of the array are for most scan angles α_0 not affected by the element pattern. The reason is that the array normally is scanned over regions where the element pattern is flat or nearly flat. The desired element pattern for maximum beam steering is indeed nearly constant within

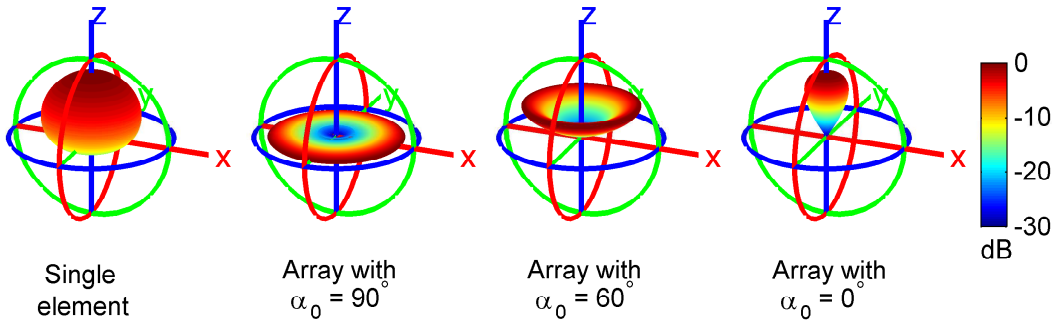
$$\alpha_{\min} < \alpha < \alpha_{\max} \ , \tag{10.27}$$

where $\alpha_{\min}$ and $\alpha_{\max}$ define the limits of the steering. In addition, the element spacing d_a should be small enough to avoid grating-lobes inside the main beam of the element pattern.

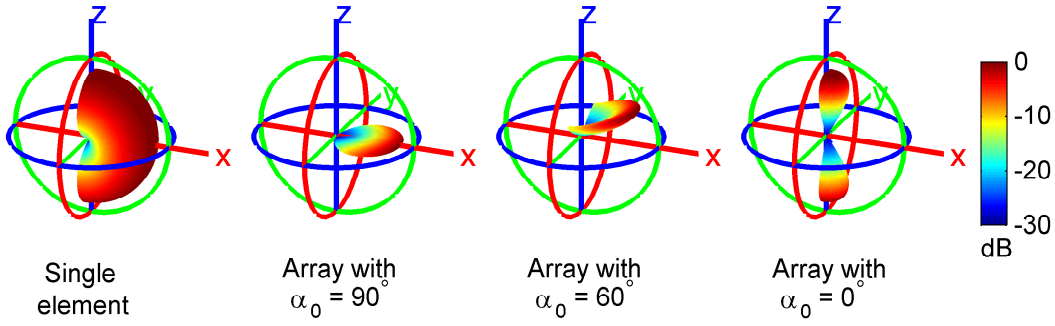
Some 3D patterns of theoretical linear phased arrays including the element factor are shown in Fig. 10.8, for $\alpha_0 = 90^\circ$ corresponding to broadside radiation (left figure), for $\alpha_0 = 60^\circ$, and for $\alpha_0 = 0^\circ$ corresponding to endfire radiation. There is assumed to be 10 elements with spacing of 0.5λ. The upper Fig. 10.8a shows the radiation patterns of the element and the linear array when the elements are Huygens sources pointing in z-direction. The Huygens source elements are omnidirectional around their pointing direction. Therefore, the linear array pattern is rotationally symmetric. When the array is phased to endfire direction ($\alpha_0 = 0^\circ$), there should be a grating-lobe at $\alpha_0 = 180^\circ$. However, this is completely suppressed by the null of the element pattern in this direction.

The elements of the radiation patterns in Fig. 10.8b are y-directed z-polarized small slots in an infinite ground plane, corresponding to incremental y-directed magnetic currents. We see that now there is a grating-lobe at $\alpha_0 = 180^\circ$ that is equally strong as the main lobe when $\alpha_0 = 0^\circ$, because the element pattern is uniform in xz-plane. There is no radiation behind the ground plane.

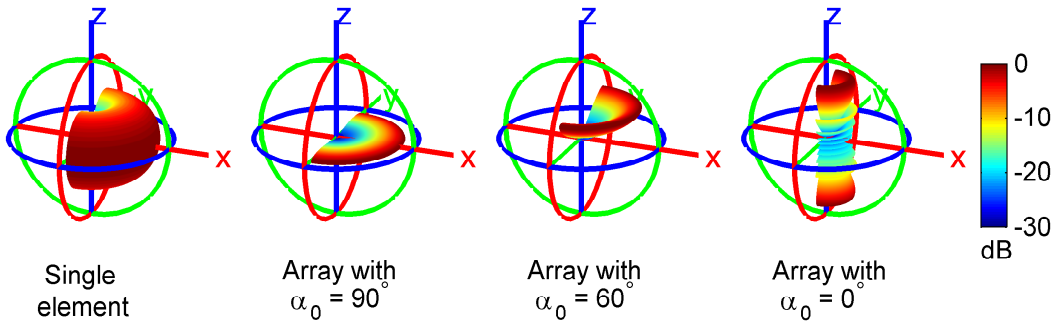
The elements of the radiation patterns in Fig. 10.8c are z-directed y-polarized small slots in an infinite ground plane, corresponding to incremental z-directed magnetic currents. We see that now there is no grating-lobe at $\alpha_0 = 180^\circ$ when $\alpha_0 = 0^\circ$, and there is no main lobe either, both because the element pattern is zero along both the positive and negative z-axis. There is no radiation behind the ground plane.



(a) The elements are z-directed Huygens sources.



(b) The elements are y-directed z-polarized small slots in an infinite ground plane coinciding with the yz-plane, i.e., y-directed magnetic currents.



(c) The elements are z-directed y-polarized small slots in an infinite ground plane coinciding with the yz-plane, i.e., z-directed magnetic currents.

Figure 10.8: *3D radiation patterns (total far field) of three different linear phased arrays. The three linear arrays have different elements as explained in (a), (b) and (c). The radiation patterns of the elements are shown to the left, and the radiation patterns of the arrays are shown to the right for three different main beam directions. There are 10 elements in each array, and the spacing between them are 0.5λ.

10.1.8 Directivity of long linear array

To calculate the directivity of the array we may conveniently locate $\hat{\mathbf{a}}$ along $\hat{\mathbf{z}}$. Then, the far-field function becomes

$$\mathbf{G}_{A}(\theta,\varphi) = \mathbf{G}(\theta,\varphi)\mathsf{AF}(\theta) \ , \tag{10.28}$$

where $\mathbf{G}(\theta,\varphi)$ is the element factor and $\mathsf{AF}(\theta)$ is the array factor. The *power integral* defined in Section 2.3.8 becomes

$$P = \int_0^{2\pi}\int_0^{\pi} |\mathbf{G}(\theta,\varphi)|^2 |\mathsf{AF}(\theta)|^2 \sin\theta d\theta d\varphi \ , \tag{10.29}$$

$$\text{where} \qquad |\mathbf{G}(\theta,\varphi)|^2 = |\mathbf{G}(\theta,\varphi)\cdot\hat{\boldsymbol{\theta}}|^2 + |\mathbf{G}(\theta,\varphi)\cdot\hat{\boldsymbol{\varphi}}|^2 \ . \tag{10.30}$$

For long arrays, $\mathsf{AF}(\theta)$ is much more directive than $\mathbf{G}(\theta,\varphi)$, so it takes the properties of a sampling delta function. Therefore, we may approximate (10.29) by using (10.19) to arrive at:

$$P = \sum_p \int_0^{2\pi} |\mathbf{G}(\theta_p,\varphi)|^2 d\varphi \left(\frac{1}{d_a}\right)^2 \int_0^{\pi} |\tilde{A}(\cos\theta - \cos\theta_p)|^2 \sin\theta d\theta \ , \tag{10.31}$$

where the sum is taken over all values of p for which θ_p is in visible space. The latter can be done without significant errors when $A(k_a)$ is a narrow function such as when $L \gg \lambda$, provided the main lobe of $(\cos\theta - \cos\theta_p)$ is completely in the visible region, i.e., typically

$$1 - |\cos\theta_p| > \lambda/L \ . \tag{10.32}$$

This means that the present analysis is not valid when the main or grating-lobes appear at endfire directions.

Let us define the following integral

$$P_\theta = \left(\frac{1}{d_a}\right)^2 \int_0^{\pi} |\tilde{A}(\cos\theta - \cos\theta_p)|^2 \sin\theta d\theta \ . \tag{10.33}$$

This can be simplified by substituting k_a with $k\cos\theta$ and dk_a with $-k\sin\theta d\theta$ and extending the integration boundaries of the k_a integral to $\pm\infty$. Then,

$$P_\theta = \left(\frac{1}{d_a}\right)^2 \frac{1}{k} \int_{-\infty}^{\infty} |\tilde{A}(k_a)|^2 dk_a \ , \tag{10.34}$$

and (10.31) simplifies to

$$P \approx P_\theta \sum_p P_\varphi(\theta_p) \qquad \text{with} \qquad P_\varphi(\theta_p) = \int_0^{2\pi} |\mathbf{G}(\theta_p,\varphi)|^2 d\varphi \ , \tag{10.35}$$

where the sum is still taken over all values of p for which θ_p is in visible space. We may now use Parseval's theorem for Fourier transforms

$$\int_{-\infty}^{\infty} |A(a)|^2 da = \frac{1}{2\pi}\int_{-\infty}^{\infty} |\tilde{A}(k_a)|^2 dk_a \ , \tag{10.36}$$

to finally obtain

$$P_\theta = \frac{\lambda}{d_a}\frac{1}{d_a}\int_{-L/2}^{L/2} |A(a)|^2 da \ . \tag{10.37}$$

Thus, the longitudinal part of the power integral can be evaluated by integrating the square of the excitation distribution $A(a)$ along the array. This corresponds to what we did when we introduced the power integral over the rectangular aperture in the aperture power integral in Eq. (7.35)[8]. However, here we consider an integral over a one-dimensional line distribution instead of two-dimensional aperture distribution. We can now express the directivity as a product of three factors, according to

$$D_0 = 4\pi \left| G_{\rm co}(\theta_0, \varphi_0) \frac{1}{d_a} \tilde{A}(0) \right|^2 / P = D_\theta D_\varphi e_{\rm grt} \ . \tag{10.38}$$

Here,

$$D_\theta = 2 \left| \frac{1}{d_a} \tilde{A}(0) \right|^2 / P_\theta = \frac{2}{\lambda} \left| \int_{-L/2}^{L/2} A(a) da \right|^2 / \int_{-L/2}^{L/2} |A(a)|^2 da \tag{10.39}$$

is the *directivity in the longitudinal φ_0-plane* containing the main lobe maximum,

$$D_\varphi = \frac{2\pi |G_0(\theta_0, \varphi_0)|^2}{P_\varphi(\theta_0)} = 2\pi |G_{\rm co}(\theta_0, \varphi_0)|^2 / \int_0^{2\pi} |\mathbf{G}(\theta_0, \varphi)|^2 d\varphi \tag{10.40}$$

is the *directivity on the main lobe cone θ_0*, and

$$e_{\rm grt} = P_\varphi(\theta_0) / \sum_p P_\varphi(\theta_p) \tag{10.41}$$

is a *grating efficiency* which represents the power loss due to grating-lobes.

When the elements have an omnidirectional far-field function around the array axis, D_φ does not change when the beam is steered from broadside ($\theta_0 = 90°$) towards endfire ($\theta_0 = 0°$). The longitudinal directivity is seen to have a maximum of $2L/\lambda$ when the excitation distribution is uniform. This maximum is independent of θ_0.This is logical since the main beam cone occupies a ring-shaped solid angle of $2\pi \sin\theta_0 \cdot \theta_{3\,\rm dB}({\rm rad})^2$ where $\theta_{3\,\rm dB}$ is the double 3 dB beamwidth of the ring-shaped main lobe. The total solid angle of the main beam cone is independent of θ_0 for narrow beams because $\theta_{3\,\rm dB} \propto 1/\sin\theta_0$ which can be seen as follows. The beam width $\Delta(\cos\theta)$ does not vary when plotted against $\cos\theta$, whereas this corresponds to $|\Delta\theta|$ proportional to $1/(\sin\theta)$ when plotted against θ, because $\Delta(\cos\theta) = -\sin\theta \Delta\theta$.

The grating efficiency contains a sum over all radiating grating-lobes. This can be easily estimated from knowing the direction θ_p of the grating-lobe and the level of $|\mathbf{G}(\theta_p, \varphi_0)|$ in the principal plane φ_0. When there is only one grating-lobe, the *grating efficiency* becomes approximately

$$e_{\rm grt} = \frac{|\mathbf{G}(\theta_0, \varphi_0)|^2}{|\mathbf{G}(\theta_0, \varphi_0)|^2 + |\mathbf{G}(\theta_{\pm 1}, \varphi_0)|^2} \ , \tag{10.42}$$

where $\theta_{\pm 1}$ is the direction of the first grating-lobe which corresponds either to $p = -1$ or $p = +1$. To obtain this simplified formula we have assumed that the shapes of $|\mathbf{G}(\theta_0, \varphi)|^2$ and $|\mathbf{G}(\theta_{\pm 1}, \varphi)|^2$ are almost equal when plotted as a function of φ. This is in particular true if the element patterns are omnidirectional, i.e., that they have no variations with φ.

[8] See Section 7.4.

10.1.9 Directivity of endfire array

The *endfire array* needs special care because part of the main lobe vanishes due to the invisible region. In this case the radiation integral of the $m = 0$ term in (10.19) becomes ($\cos\theta_0 = 1$)

$$P = \int_0^{2\pi} \int_0^{\pi} |\mathbf{G}(\theta_0, \varphi_0)|^2 \left(\frac{1}{d_a}\right)^2 |\tilde{A}(\cos\theta - 1)|^2 \sin\theta d\theta d\varphi \ . \tag{10.43}$$

The elements used in endfire arrays will always have a broad lobe near $\theta_0 = 0$, so we may write

$$P = |\mathbf{G}(0,0)|^2 2\pi \left(\frac{1}{d_a}\right)^2 \int_{-1}^{1} |\tilde{A}(\cos\theta - 1)| d(\cos\theta) \ . \tag{10.44}$$

We see that in this case the integral contains only the half of the main lobe of $|\tilde{A}(k_a - k_\varphi)|$ which is in the visible region. Therefore, this time we get (compare with the derivations in the previous subsection)

$$P = 2\pi |\mathbf{G}(0,0)|^2 \frac{1}{2d_a^2} \int_{-L/2}^{L/2} |A(a)|^2 da \ . \tag{10.45}$$

Finally, the *directivity* for the endfire case becomes

$$\begin{aligned} D &= 4\pi \left| \mathbf{G}(0,0) \frac{1}{d_a} \tilde{A}(0) \right|^2 / P \\ &= \frac{4}{\lambda} \left| \int_{-L/2}^{L/2} A(a) da \right|^2 / \int_{-L/2}^{L/2} |A(a)|^2 da \ , \end{aligned} \tag{10.46}$$

which is twice that of a broadside array of elements that are omnidirectional in φ. In practice, endfire arrays cannot be designed with grating-lobes present, so the grating efficiency will be unity. The maximum value that the endfire directivity can take is therefore

$$D_{\max} = 4L/\lambda \ , \tag{10.47}$$

which appears when the excitation is uniform, i.e., $A(a) = 1$ within $-L/2 < a < L/2$.

10.1.10 Example: Linear array of waveguide apertures

Consider an infinite ground plane with a linear array of rectangular waveguide apertures. Each aperture has a length $l = 0.7\lambda$ and width $w \ll \lambda$. We want to design the array for scan in H-plane. The excitation is uniform with linearly progressive phase.

a) Write the expression for the radiation field function $\mathbf{G}(\theta, \varphi)$ of one single element when we locate the coordinate system with its z-axis normal to the aperture and its y-axis in the direction of the polarization.

b) What are the approximate radiation patterns of the element in the E- and H-planes when we assume that l is small?

c) Determine the length of the array when we require that the 3 dB half beamwidth in the plane of scan should be smaller than or equal to 0.2° at broadside.

d) How many elements are needed when we require that the grating-lobe maximum shall not radiate in visible space when the main beam is at broadside?

e) Choose the minimum possible element distance. How far from broadside can we scan the beam when we require that the grating-lobe maximum shall not radiate in visible space?

f) Calculate the directivity in dBi as a function of scan angle when there is no grating-lobes.

SOLUTION:

a) The radiation field function of the waveguide aperture is:

$$\begin{aligned}\mathbf{G}(\theta,\varphi) &= C(\hat{\mathbf{x}}\times\hat{\mathbf{r}})\tilde{M}(k(\hat{\mathbf{x}}\cdot\hat{\mathbf{r}})) \\ &= C(\sin\varphi\hat{\boldsymbol{\theta}}+\cos\theta\cos\varphi\hat{\boldsymbol{\varphi}})\tilde{M}(k\sin\theta\cos\varphi)\ ,\end{aligned}$$

where C is a constant and $\tilde{M}(k_x)=\int_{-l/2}^{l/2}\cos(\pi x'/l)e^{jk_x x'}dx'$.

b) The E-plane pattern ($\varphi=90^\circ$) is omnidirectional, and the H-plane pattern goes approximately as $\cos\theta$.

c) The radiation field function of the array is given by $\mathbf{G}_A(\theta,\varphi)=\mathbf{G}(\theta,\varphi)\mathsf{AF}(\hat{\mathbf{r}})$ where the array factor is

$$\mathsf{AF}(\hat{\mathbf{r}})=\frac{1}{d_a}\sum_{m=-\infty}^{\infty}(-1)^{m(n-1)}\tilde{A}\left(k_a-k_\varphi+m\frac{2\pi}{d_a}\right)\ .$$

If we assume that there is no grating-lobes, we have

$$\mathsf{AF}(\hat{\mathbf{r}})\approx\frac{1}{d_a}\tilde{A}(k_a-k_\varphi)=\frac{1}{d_a}\tilde{A}(k_a)\ ,$$

where the latter form is valid when the main beam radiates broadside. The 3 dB width of $\tilde{A}(k_a)$ is[9] $\theta_{\rm dB}=\arcsin(0.445\lambda/L)=0.2^\circ$. Thus, we need an array length

$$L=0.445\lambda/\sin(0.2^\circ)=127.5\lambda\ .$$

d) The requirement for no radiating grating-lobe maxima when the main lobe radiates in the broadside direction (i.e., $\alpha_0=90^\circ$) is

$$d_a\le\frac{\lambda}{1+|\cos\alpha_0|}=\lambda\ .$$

Thus, we need at least $N=L/\lambda=127.5$, i.e., 128 elements.

e) The minimum possible element spacing is equal to the element size $d_a=0.7\lambda$. We can use

$$d_a=\frac{\lambda}{1+|\cos\alpha_0|}=0.7\lambda$$

to calculate

$$|\cos\alpha_0|=\frac{\lambda}{d_a}-1=0.429\qquad\text{and}\qquad\alpha_0=\arccos(0.429)=64.6\ .$$

[9] See Table 7.1 in Section 7.4.3.

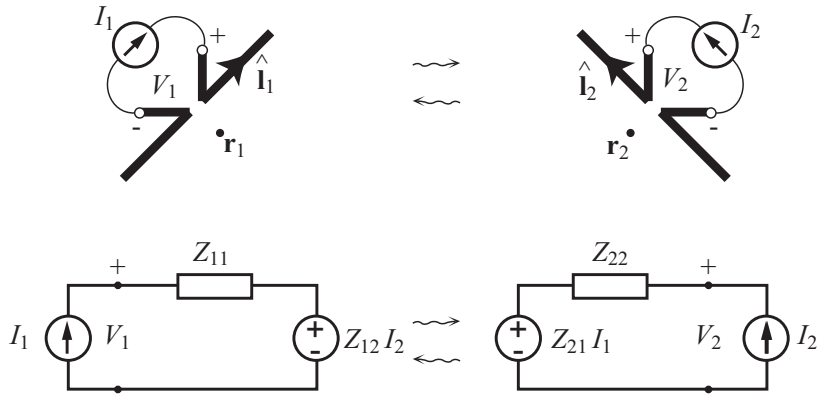


Figure 10.9: Two dipoles located at arbitrary locations and with arbitrary orientations (upper), and their equivalent circuits (lower).

Thus, we can steer to $(90^\circ - 64.6^\circ) = 25.4^\circ$ from broadside before grating-lobe maxima appears. In this case, we do not need to use the slightly stricter condition which includes the length L of the array, since the H-plane pattern has a null at both endfire directions, and therefore will suppress the grating-lobes at endfire.

f) The directivity is given by $D_0 = D_\theta D_\varphi e_{\rm grt}$. When there is no radiating grating-lobes $e_{\rm grt} = 1$. A uniform excitation gives a longitudinal directivity $D_\theta = 2L/\lambda = 24.07\,\text{dB}$. The slot is omnidirectional in the upper half-space on a cone around the x-axis. If now φ_x is measured around the x-axis, we obtain the directivity on the main lobe cone:

$$D_{\varphi \rm x} = 2\pi / \left\{ \int_0^\pi d\varphi \right\} = 2 \quad \text{(i.e., } 3.01\,\text{dB)} \; .$$

Thus, the directivity of the linear array becomes $(D_0)_{\rm dBi} = (24.07 + 3.01)\,\text{dBi} = 27.08\,\text{dBi}$. This is independent of scan angle a long as there is no radiating grating-lobes.

10.2 Scan Impedance or Admittance

Each separate array element is characterized by its self-impedance (or admittance) in free space. However, when the element is located in the array environment its apparent input impedance changes due to mutual coupling to the other elements, in particular to the closest ones. The input impedance of an element, when all the array elements are excited for the desired radiation pattern, is called the *scan impedance* or the *active impedance*. As the name suggests this impedance varies with the scan angle when the main beam is phase-steered. In the following subsections, we will show how to calculate the scan impedance for an array of half-wave dipoles. The scan impedance is traditionally referred to as the active impedance, but the preferred name is now scan impedance to avoid confusion with active antennas, i.e., antennas with active integrated components.

10.2.1 Mutual impedance between two dipoles

We consider two half-wave dipoles: one located at $\mathbf{r}_1$ and oriented in the direction $\hat{\mathbf{l}}_1$, and the other located at $\mathbf{r}_2$ and oriented in the direction $\hat{\mathbf{l}}_2$, see Fig. 10.9. Both are modeled by the sinusoidal dipole current of a half-wave dipole[10], i.e.,

$$\mathbf{J}_1(l_1) = I_1 \cos(\pi l_1/\lambda)\hat{\mathbf{l}}_1 \qquad \text{for } |l_1| < \lambda/4$$

for dipole number 1 and

$$\mathbf{J}_2(l_2) = I_2 \cos(\pi l_2/\lambda)\hat{\mathbf{l}}_2 \qquad \text{for } |l_2| < \lambda/4$$

for dipole number 2. Both dipoles have a feed gap at their centers where the port current is I_1 and I_2, respectively. The voltages at the two ports are correspondingly V_1 and V_2.

Let us now consider a current source I_2 at port 2. This causes an E-field $\mathbf{E}_{12}$ at dipole 1, and a voltage V_{12} is induced over its port. Next we consider a current source I_1 at port 1 which induces a voltage V_{21} at port 2. From reciprocity we know that the reaction between V_{12} (due to I_2) and I_1 must be equal to the reaction between V_{21} (due to I_1) and I_2 [11]. This means that

$$-V_{12}I_1 = -V_{21}I_2 \ . \tag{10.48}$$

Also, in the same way the reaction between $\mathbf{E}_{12}$ (due to I_2) and $\mathbf{J}_1(l_1)$ must be equal to the reaction between V_{21} (due to $\mathbf{J}_1(l_1)$) and I_2, i.e.,

$$\langle \mathbf{E}_{12}, \mathbf{J}_1(l_1)\rangle = -V_{21}I_2 \ . \tag{10.49}$$

By using (10.48) and (10.49), we can calculate V_{12} as follows

$$V_{12} = -\frac{1}{I_1}\langle \mathbf{E}_{12}, \mathbf{J}_1(l_1)\rangle \ . \tag{10.50}$$

$\mathbf{E}_{12}$ is proportional to I_2, so we may introduce a *mutual impedance* of the form

$$Z_{12} = \frac{V_{12}}{I_2} = -\frac{1}{I_1 I_2}\langle \mathbf{E}_{12}, \mathbf{J}_1(l_1)\rangle \ . \tag{10.51}$$

The equivalent circuits are shown in Fig. 10.9 as well. They are obtained from Section 2.6.1. The mutual impedances are included as voltage sources which are proportional to the current of the opposite dipole. The actual expressions which must be evaluated are

$$\frac{\mathbf{E}_{12}(\mathbf{r})}{I_2} = C_k \int_{-\lambda/4}^{\lambda/4} \eta \cos\left(\frac{\pi I_2'}{\lambda}\right) [\hat{\mathbf{l}}_2 C_{N_1} - (\hat{\mathbf{l}}_2 \cdot \hat{\mathbf{R}})\hat{\mathbf{R}} C_{N_2}] \frac{1}{R} e^{-jkR} dl_2' \ , \tag{10.52}$$

where C_k, C_{N_1} and C_{N_2} are given in equation (4.41)-(4.42) and $R = |\mathbf{r} - \mathbf{r}'|$, with

$$\begin{aligned} \mathbf{r} &= \mathbf{r}_1 + l\,\hat{\mathbf{l}}_1 \qquad && \text{for } -\lambda/4 < l < \lambda/4 \ , \\ \mathbf{r}' &= \mathbf{r}_2 + l'\hat{\mathbf{l}}_2 \qquad && \text{for } -\lambda/4 < l' < \lambda/4 \ . \end{aligned} \tag{10.53}$$

[10] See Section 5.1.2.
[11] See Section 4.5.

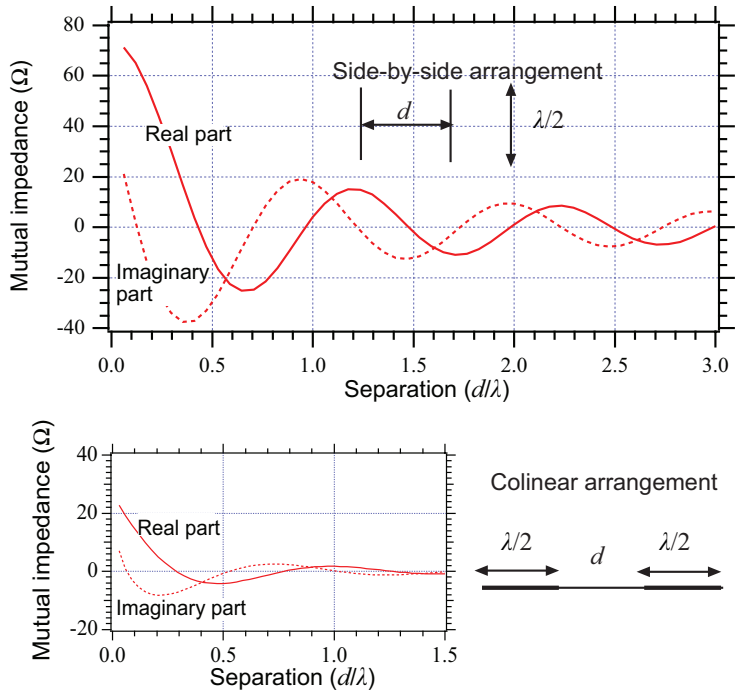


Figure 10.10: *Mutual impedances between two half-wave dipoles for side-by-side (upper) and colinear arrangements (lower).

The reaction integral becomes

$$\frac{1}{I_2 I_1}\langle \mathbf{E}_{21}(\mathbf{r}), \mathbf{J}_1(l_1)\rangle = \int_{-L_1/2}^{L_1/2} (\mathbf{E}_{12}(\mathbf{r})/I_2)\cdot \cos(\pi l_1/\lambda)\hat{\mathbf{l}}_1 dl_1 \ . \tag{10.54}$$

The mutual impedance is much easier to calculate than the self-impedance because there is no problem with singularities in the Green's function. The mutual impedance for two half-wave dipoles is evaluated from the above, and is plotted in Fig. 10.10 for side-by-side and colinear orientations*. We could also have evaluated it by using the near-field algorithm presented in Section 4.7.1.

10.2.2 Scan impedance (active impedance)

The equivalent circuit of dipole number m in an array of dipoles is shown in Fig. 10.11. The total voltage at its port will be

$$V_m = \sum_{n=1}^{N} Z_{\mathrm{mn}} I_n \ , \tag{10.55}$$

where Z_{mn} is the mutual impedance between dipoles m and n when $n \neq m$, and the self-impedance of dipole m when $n = m$.

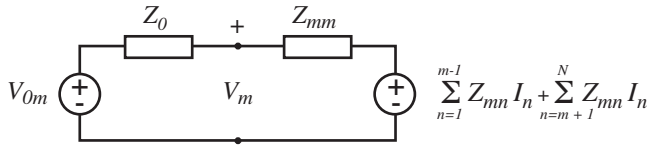


Figure 10.11: Equivalent circuit of dipole number m in a linear array of dipoles.

When phase steering the array, we excite the array factor in (10.6) by the voltages

$$V_{0_m} = A_m e^{-jka_m \sin(\alpha_0)+j\varphi_c} \qquad \text{for } m = 1, 2, ..., N \ , \tag{10.56}$$

at the ports of the N elements where α_0 is the main lobe direction defined by $\sin\alpha_0 = k_\phi/k$. If we model the physical excitation circuit by its Thevenin equivalent, the current on element m needs to be calculated from the circuit equation

$$V_{0_m} = Z_0 I_m + \sum_{n=1}^{N} Z_{\rm mn} I_n \ , \tag{10.57}$$

where Z_0 is the source impedance. This defines N linear equations with the N unknowns $I_1, I_2, ..., I_N$ that can be readily solved. When the unknown currents have been found we obtain the apparent impedance of each element m to be

$$Z_{m_{\rm scn}} = \sum_{n=1}^{N} Z_{\rm mn} \left(\frac{I_n}{I_m} \right) \ . \tag{10.58}$$

The mutual impedances are strongest for the closest elements, so the boundaries of the sum expression may often be truncated to the three to five neighboring elements on each side. The solutions for I_m can be expressed as

$$I_m = \frac{V_{0_m}}{Z_0 + Z_{m_{\rm scn}}} \ . \tag{10.59}$$

$Z_{m_{\rm scn}}$ is called the *scan impedance*, because it depends on the scan angle α_0. The traditional name is *active impedance*. This is the apparent impedance seen at the terminal of element m when all the elements of the array are excited, i.e., when all elements are active. The scan impedance is in principle different for each element. However, if the amplitude excitation varies slowly over the array, and if the phase variation is linear as in (10.56), the scan impedance will be the same for all elements. However, the three to four elements that are located closest to each of the two ends of the array may have a different scan impedance.

There exist special techniques for calculating the scan impedance of infinite arrays, by using so-called Floquet mode expansions of the field solutions. The infinite array impedance can also be measured in an infinite *array simulator*, which can be realized by locating one or more array elements inside a rectangular waveguide. Imaging in the walls of the waveguide create the "infinite" array environment.

10.2.3 Scan blindness

The scan impedance changes with the direction α_0 of the main lobe. If we match the scan impedance to the source (or transmission line) impedance at broadside ($\alpha_0 = 90°$), we will

get mismatch and corresponding power loss when the beam is scanned away from broadside. At some angle the scan impedance may change very rapidly creating a large mismatch. This effect is referred to as *scan blindness*.

Scan blindness can have two reasons: it may be due to the sudden appearance of a grating-lobe in the visible region of the array factor $\mathsf{AF}(\cos\alpha)$, or it may be due to the sudden and large excitation of a surface wave in the antenna structure.

a) The first case appears when the element pattern is very broad, such as, e.g., in the E-planes of waveguide slot and microstrip antennas. The first radiating grating-lobe has its maximum along the array when (see (10.25) and (10.26))

$$|\cos\alpha_{_0}| = (\lambda/d_a) - 1 \ . \tag{10.60}$$

Therefore, scan blindness appears when $\alpha_{_0}$ approaches and exceeds this value. Then, all array elements will be phased for radiation along the array, so that the mutual coupling voltages of all elements add in phase and cause a large change in the scan impedance and thereby a large mismatch of the elementss. We also get a large and sudden change in the directivity, because the grating efficiency $\eta_{\rm grt}$ decreases. The angle in (10.60) cannot be overpassed during the scan and represents a limitation of the scanning range. This blindness never appears if $d_a < \lambda/2$ since then $|\cos\alpha_{_0}| > 1$.

b) The second case appears when the structure upon which the array elements are located can support surface waves. This may, e.g., be the grounded substrate of a microstrip antenna. The substrate may support a surface wave with wavenumber $k_{\rm sw} > k$. The scan blindness appears when the main beam is scanned to

$$k|\cos\alpha_{_0}| - \left(\frac{2\pi}{d_a}\right) = -k_{\rm sw} \ , \tag{10.61}$$

where $\lambda_{\rm sw} = 2\pi/k_{sw}$ is the wavelength of the surface wave. Since $(\lambda/\lambda_{\rm sw}) > 1$, the blindness angle $\alpha_{_0}$ can be located in the visible range also if $(\lambda/d_a) > 2$. This blindness corresponds to the excitation of a grating lobe appearing as a surface wave in the structure, i.e., being in invisible space. It is not radiating, but propagates inside the structure. It couples to the elements and changes their scan impedances strongly.

10.2.4 Active, scan and embedded element patterns

The scan characteristics of the array can be measured directly by exciting and phasing all elements. However, it is also possible to measure the scan characteristics by exciting one single element in the middle of the array. All the other elements must then be present and terminated by the same impedance which the voltage sources will have when all the elements are excited. The measured far-field pattern of this array with one element excited (as a function of the polar angle in the principal φ-plane) will then represent the variation of the directivity (of the complete array when all the elements are excited) as a function of the scan angle in the same φ-plane. Such scan element patterns measured in the array environment show sudden dips in directions with scan blindness, see Fig. 10.12. Traditionally the *scan element pattern* is called *active element pattern*, so most previous literature use the latter term. The term scan element pattern is also used in a recent textbook.

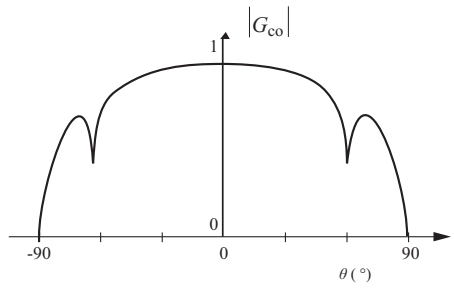


Figure 10.12: Example of scan element pattern with blindness dip at $\theta = 60°$.

However, the scan (or active) element pattern is nowadays even more descriptively referred to as the embedded element pattern. This is the far-field function of one element in the array when all the other elements are present and terminated. The term is originating from [4]. However, the main concept that the embedded element has a very low efficiency in a dense array originates from 1964 [5] and explains the paradox that the gain of a dense array always is smaller than the sum of the gains of each isolated element. This so-called *embedded element efficiency* is a concept that quantifies the effects of mutual coupling in one single efficiency value and thereby provides a unifying concept [6]. Is used also to characterize MIMO arrays [7]. The embedded element pattern is time-consuming to compute because the mutual couplings between all elements must be included. The low efficiency associated with a single embedded element in a dense array is not known in previous literature using the active element terminology.

The embedded element efficiency is treated in detail in Chapters 3 and 11.

10.3 Planar arrays of equispaced elements

The *planar array* consists of elements that are located in a planar two-dimensional grid. It can be looked upon as a linear array of linear arrays. Therefore, the features present in linear arrays will also appear in planar arrays. The planar array has the capability of phase steering the main beam to any direction in space if properly designed. It is also more versatile than the linear array in other ways, e.g., the main beam can be shaped to any form by controlling the amplitudes and phases of all elements. The elements of a planar array can be located in many types of grids. The most common ones are the *rectangular grid* and the *triangular grid* (Fig. 10.13). Our treatment is limited to the rectangular grid, even though the triangular grid has better scan characteristics.

10.3.1 Array factor as an element-by-element sum

A rectangular grid in the xy-plane is shown in Fig. 10.14. The element spacing is d_x in x-direction and d_y in y-direction. Then, the element locations are described by

$$\mathbf{r}_{\mathrm{mn}} = \mathbf{r}_c + \left(m - \frac{M+1}{2}\right) d_x \hat{\mathbf{x}} + \left(n - \frac{N+1}{2}\right) d_y \hat{\mathbf{y}} \ , \tag{10.62}$$

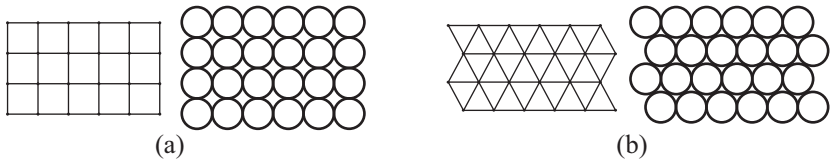


Figure 10.13: Element locations in array antennas with (a) quadratic and (b) triangular grids. The elements shown have circular apertures. The triangular grid is very compact.

for $n = 1, 2, ..N$, $m = 1, 2, ..M$ where $\mathbf{r}_c$ is the geometrical center of the array. The far-field function of the array becomes

$$\mathbf{G}_{A}(\hat{\mathbf{r}}) = \mathbf{G}(\hat{\mathbf{r}})\mathsf{AF}(\hat{\mathbf{r}}) \ , \tag{10.63}$$

$$\mathsf{AF}(\hat{\mathbf{r}}) = \sum_{n=1}^{N}\sum_{m=1}^{M} A_{\mathrm{mn}} e^{j\Phi_{\mathrm{mn}}} e^{jk\mathbf{r}_{\mathrm{mn}}\cdot\hat{\mathbf{r}}} \ , \tag{10.64}$$

where $\mathbf{G}(\hat{\mathbf{r}})$ is the far-field function of the embedded element, and $\mathsf{AF}(\hat{\mathbf{r}})$ is the array factor written as an *element-by-element sum* with $A_{\mathrm{mn}}e^{j\Phi_{\mathrm{mn}}}$ the excitation of element mn. The element-by-element sum in (10.64) is convenient only for arrays with few elements. For large arrays the approach in the next section is preferable.

10.3.2 Array factor as a grating-lobe sum

In the same way as for the linear arrays we introduce an excitation distribution $A(x, y)$. This is smooth and continuous for $|x| < L_x/2$ and $|y| < L_y/2$ with $L_x = Md_x$ and $L_y = Nd_y$. Further $A(x, y)$ is zero when $|x| > L_x/2$ or $|y| > L_y/2$, and takes on the values

$$A(x, y) = A_{\mathrm{mn}} \tag{10.65}$$

at all $\mathbf{r}_{\mathrm{mn}}$ inside the array. Similarly, we define a smooth and continuous phase function $\Phi(x, y)$. We assume this to vary linearly with x and y according to

$$\Phi(x, y) = \Phi_c - k_{\Phi_x}x - k_{\Phi_y}y \ , \tag{10.66}$$

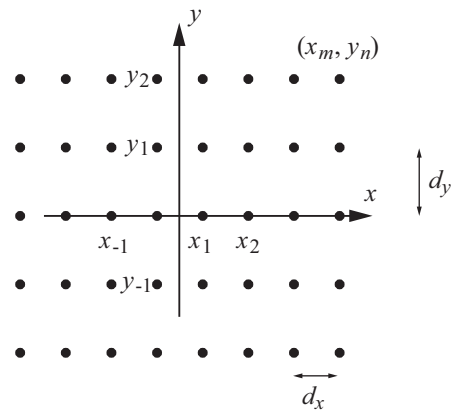
where $k_{\Phi_x}(k_{\Phi_y})$ is the "propagation" constant of the phase excitation in x-direction (y-direction), defined by

$$\begin{aligned} k_{\Phi_x} &= -\Delta\Phi_x/d_x \qquad \text{with } -\pi < \Delta\Phi_x < \pi \ , \\ (k_{\Phi_y} &= -\Delta\Phi_y/d_y \qquad \text{with } -\pi < \Delta\Phi_y < \pi) \ , \end{aligned} \tag{10.67}$$

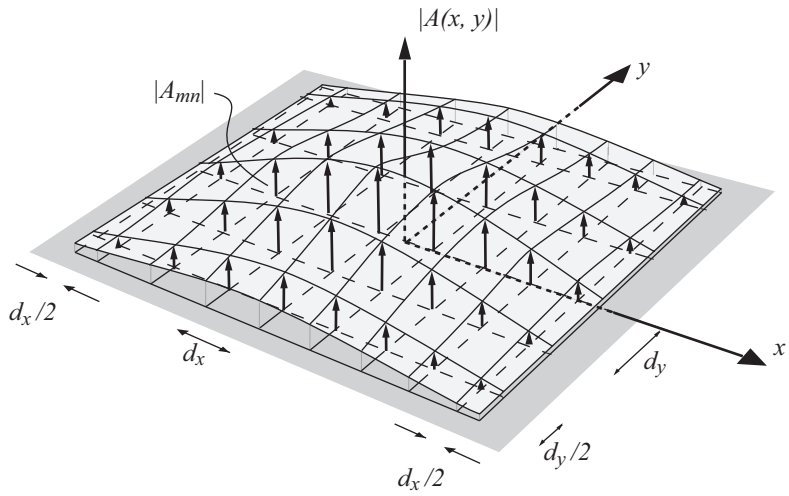
with $\Delta\Phi_x(\Delta\Phi_y)$ the phase difference between neighboring elements in x-direction (y-direction). Of convenience we choose both $\Delta\Phi_x$ and $\Delta\Phi_y$ in the interval $-\pi$ to π, in the same way as we did it for the linear array.

We can now express the array factor as a double integral by using the sampling properties of the delta function, according to

$$\begin{aligned} &\mathsf{AF}(\hat{\mathbf{r}}) = \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left\{ \sum_{m=-\infty}^{\infty} \delta(y - md_y) \sum_{m=-\infty}^{\infty} \delta(x - md_x) \right\} \cdot A(x, y) e^{j\Phi(x,y)} e^{jk\mathbf{r}\cdot\hat{\mathbf{r}}} dxdy \ , \end{aligned} \tag{10.68}$$



(a) Lay-out of planar array in rectangular grid.



(b) Illustration of excitation distribution $|A(x, y)|$.

Figure 10.14: Example of array with rectangular aperture and rectangular grid, and its excitation distribution $A(x, y)$.

where $\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}}$. Here, the finite element by element sums have been extended to infinity, which could be done because $A(x, y)$ is zero when $n < 1$, $n > N$, $m < 1$ and $m > M$. The delta series in x-direction is periodic with period d_x, so it can be expanded in a Fourier series according to (see also Eq. (10.18))

$$\sum_{m=-\infty}^{\infty} \delta(x - md_x) = \frac{1}{d_x} \sum_{p=-\infty}^{\infty} (-1)^{p(M-1)} e^{-jp(2\pi/d_x)x} \ . \tag{10.69}$$

We can expand the delta series in y-direction in the same way, giving

$$\begin{aligned} &\sum_{m=-\infty}^{\infty} \delta(x - md_x) \sum_{n=-\infty}^{\infty} \delta(y - nd_y) = \\ &= \frac{1}{d_x d_y} \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} (-1)^{p(M-1)+q(N-1)} e^{-jp(2\pi/d_x)x} e^{-jq(2\pi/d_y)y} \end{aligned} \ . \tag{10.70}$$

Using these expansions in (10.68) and interchanging the integration and summation we finally achieve

$$\mathsf{AF}(\hat{\mathbf{r}}) = e^{j\phi_c} \frac{1}{d_x d_y} \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \tilde{A}\left(k_x - k_{\Phi_x} - p\frac{2\pi}{d_x}, k_y - k_{\Phi_y} - q\frac{2\pi}{d_y}\right) \ , \tag{10.71}$$

$$\text{where} \qquad \tilde{A}(k_x, k_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(x, y) e^{jk_x x} e^{jk_y y} dx dy \tag{10.72}$$

is the two-dimensional Fourier transform of the excitation distribution $A(x, y)$. This expression is very similar to that of the linear array[12]. Thus, we may also here refer to the infinite sum expression in (10.71) as an *infinite grating-lobe (or Floquet mode) sum* version of the array factor. The array factor is a sum of equal contributions, which are displaced relative to each other. If $A(x, y)$ is a real function, $\tilde{A}(k_x, k_y)$ will have a maximum at $\tilde{A}(0, 0)$. Therefore, each of the contributions is centered around its maximum for which $k_{\text{xp}} = k_{\Phi_x} + p(2\pi/d_x)$ and $k_{\text{yq}} = k_{\Phi_y} + q(2\pi/d_y)$. The array is located in the xy-plane. Therefore, the array factor is conveniently represented in terms of

$$k_x/k = \sin\theta\cos\varphi \ , \qquad k_y/k = \sin\theta\sin\varphi \ , \tag{10.73}$$

where k_x/k and k_y/k are the *uv-coordinates* introduced in Section 2.3.6.

10.3.3 Steered main lobe

The first maximum of the array factor appears when $p = q = 0$ and

$$k_{x0} = k_{\Phi_x} \ , \qquad k_{y0} = k_{\Phi_y} \ . \tag{10.74}$$

This means that the maximum appears for the θ_0 and φ_0 given by (see Fig. 10.15)

$$\sin\theta_0 = \sqrt{(k_{\Phi_x})^2 + (k_{\Phi_y})^2}/k \ , \qquad \tan\varphi_0 = k_{\Phi_y}/k_{\Phi_x} \ . \tag{10.75}$$

Therefore, we can also now steer the beam by exciting the elements with a linearly progressive phase. The shape of the main beam is mainly determined by the Fourier transform of the excitation distribution.

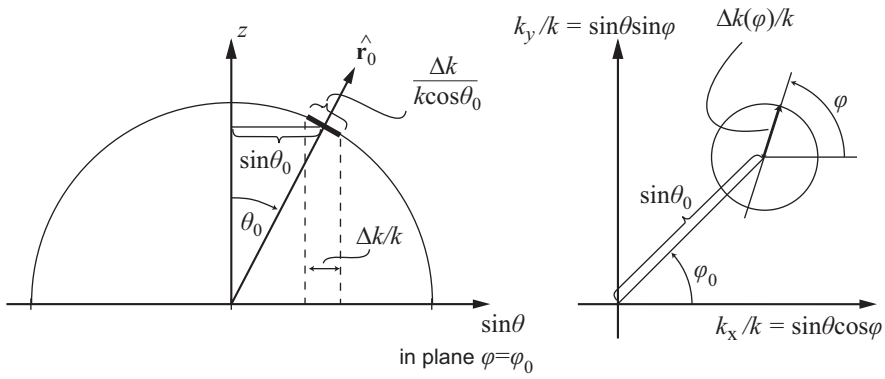


Figure 10.15: Illustration of the main lobe for the case in Fig. 10.16 and its width in different planes.

It is of interest to know the width of the main beam in different planes. The shape of the main beam is determined by $\tilde{A}(k_x - k_{x0}, k_y - k_{y0})$, and the width of this in terms of k_x and k_y is independent of the values of k_{x0} and k_{y0}. Let us define the width in a certain φ-plane through the main lobe to be $\Delta k(\varphi)$. Then, it is possible to show that the corresponding angular width of the beam in radians as seen in a coordinate system with z-axis in the direction θ_0, φ_0 becomes

$$\Delta\theta(\varphi) = \left(\frac{\Delta k(\varphi)}{k}\right)\sqrt{(\sin(\varphi-\varphi_0))^2 + (\cos(\varphi-\varphi_0)/\cos\theta_0)^2}\ . \tag{10.76}$$

This can be seen by studying Fig. 10.15 and Fig. 10.16c.The proof is left as an exercise. We see that the beamwidth broaden as $1/\cos\theta_0$ in the plane of scan φ_0 when the scan angle θ_0 increases. There is no beam broadening in the plane orthogonal to the scan plane.

10.3.4 Grating-lobes

The array factor has also maxima when

$$k_{\mathrm{xp}} = k_{\Phi_x} + p\frac{2\pi}{d_x}\ , \qquad k_{\mathrm{yq}} = k_{\Phi_y} + q\frac{2\pi}{d_y}\ , \tag{10.77}$$

for $p = \pm1, \pm2, ...$ and $q = \pm1, \pm2, ...$, which corresponds to

$$\begin{aligned}\sin\theta_{pq}\cos\varphi_{pq} &= \sin\theta_0\cos\varphi_0 + p\frac{\lambda}{d_x}\ ,\\ \sin\theta_{pq}\sin\varphi_{pq} &= \sin\theta_0\sin\varphi_0 + q\frac{\lambda}{d_y}\ .\end{aligned} \tag{10.78}$$

These are grating-lobes. The locations of the grating-lobes appear with equal spacing on a rectangular grid in a plane with *uv-coordinates* $u = \sin\theta\cos\varphi$ and $v = \sin\theta\sin\varphi$ (see Fig. 10.16). The shape of the main lobe and grating-lobes is determined by $\tilde{A}(k_x - k_{\mathrm{xp}}, k_y - k_{\mathrm{yq}})$ and can be plotted as contours around the directions in (10.78).

[12] See the discussion at the end of Section 10.1.3.

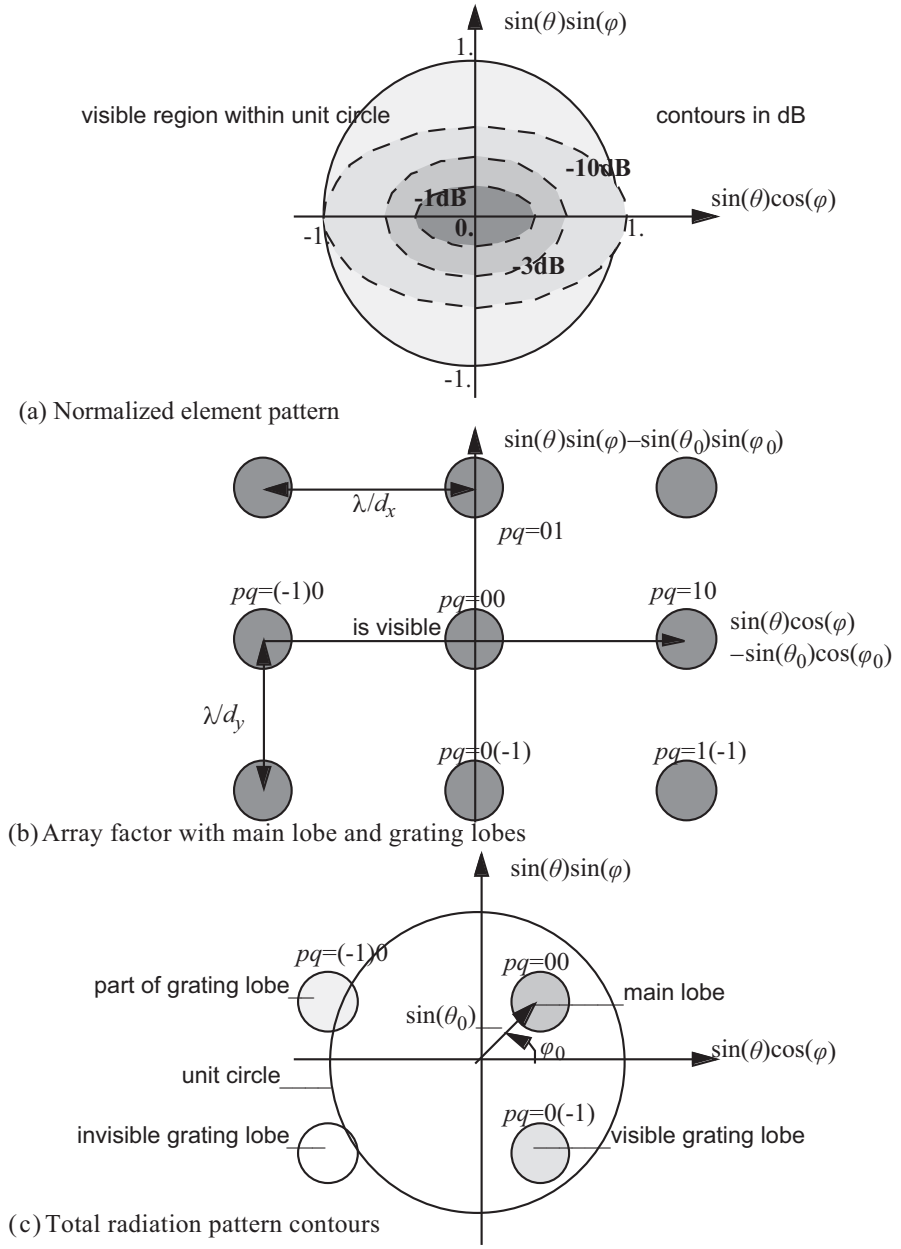


Figure 10.16: Illustration of grating-lobe locations in array with circular aperture and rectangular grid. (a) Element factor. The contours show the level in dB relative to broadside. (b) Array factor. The contour illustrate the circular beamwidth. (c) Combined array and element factors (the grey tone illustrates modification of level due to element pattern).

Let us take a look at the first grating-lobe in the diagonal plane of the unit cell of the array. The diagonal plane of the element geometry is defined by $\varphi = \arctan(d_y/d_x)$, whereas it is clear that the corresponding first *grating-lobes appear in the complementary diagonal planes* defined by $\varphi = \pm\arctan(d_x/d_y)$. This is surprising, and important. The level of all the grating-lobes are determined by the level of the element patterns in their directions relative to the level of the element pattern in the main lobe direction.

In order to avoid visible grating-lobes, we must require that the grating-lobes are located outside the circle in Fig. 10.16, for which $\sin\theta = 1$, which defines the visible region. The width of the grating-lobes are about λ/D, where D is the diameter of the array in the φ-plane of the grating-lobe, so we must typically require that $\sin\theta_{pq} > 1 + (\lambda/D)$ which corresponds to

$$\sqrt{\left(\sin\theta_0\cos\varphi_0 + p\frac{\lambda}{d_x}\right)^2 + \left(\sin\theta_0\sin\varphi_0 + q\frac{\lambda}{d_y}\right)^2} > 1 + \frac{\lambda}{D} \tag{10.79}$$

in order to avoid any radiation of the grating-lobe into visible space becomes. This sets a similar *requirement for non-radiating grating lobes* to the element spacings d_x and d_y, as that in (10.26) for linear arrays. The grating-lobes can also be suppressed by the element factor. This is clearly seen by the contour plots in Fig. 10.17.

10.3.5 Directivity

The power integral of the *planar array* is

$$P = \int_0^{2\pi}\int_0^{\pi/2} |\mathbf{G}_A(\theta,\varphi)|^2 \sin\theta d\theta d\varphi \ , \tag{10.80}$$

where we have assumed that the element factor is zero behind the array, i.e., for $|\theta| > \pi/2$. This is true when the array elements are located on a ground plane or in other ways are unidirectional. There may be some radiation for $\theta > \pi/2$ caused by diffraction from the ground plane edge, but these effects cannot be accounted for by the array factor expressions we have derived. Let us now introduce the array factor in (10.71) into (10.80), and do the following variable substitution,

$$k_\rho = k\sin\theta \ , \qquad dk_\rho = k\cos\theta d\theta \ . \tag{10.81}$$

Then,[13]

$$P = \int_0^{2\pi}\int_0^{k} |\mathbf{G}(\theta,\varphi)|^2 |\mathsf{AF}(k_\rho\cos\varphi, k_\rho\sin\varphi)|^2 \frac{1}{k^2\cos\theta} k_\rho dk_\rho d\varphi \ . \tag{10.82}$$

For arrays having a large number of elements in both x- and y-directions, the array factor will be much narrower than the element factor. Then, the array factor in (10.82) may be considered as a sum of delta functions, so by using (10.71) and (10.82) we may write

$$P = \sum_{pq} |\mathbf{G}(\theta_{pq},\varphi_{pq})|^2 \frac{(2\pi)^2}{k^2\cos\theta_{pq}(d_x,d_y)^2} P_{pq} \ , \tag{10.83}$$

[13] We have not removed the inverse $\cos\theta$ factor by the substitution, because this varies slowly compared to the array factor.

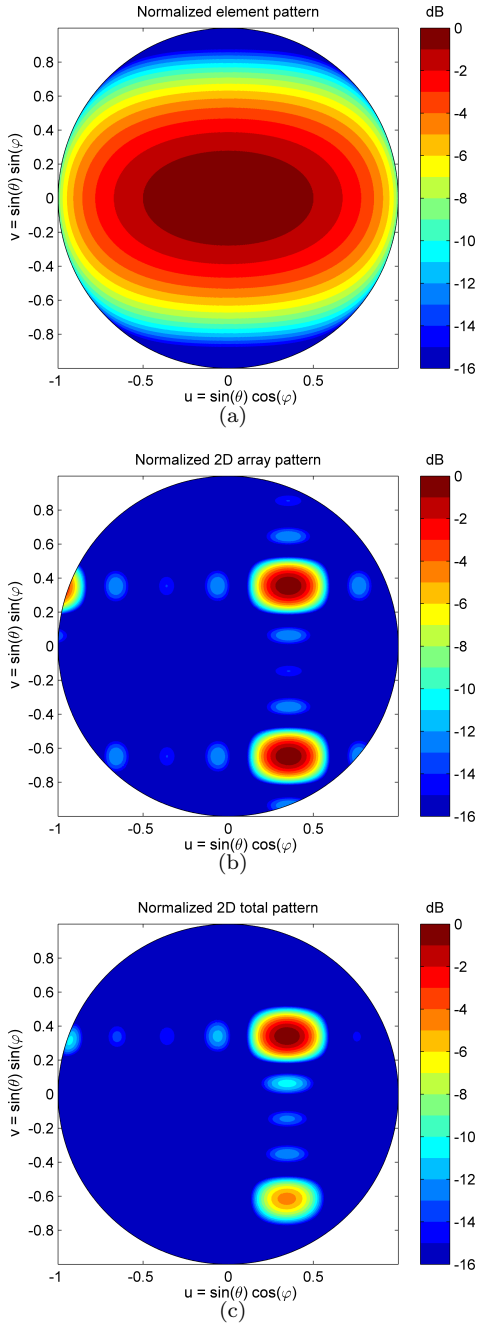


Figure 10.17: *Contour plots of: (a) element factor, (b) array factor and (c) total far field function of the planar array illustrated in Fig. 10.16. $M = 5$, $d_x = 0.7\lambda$, $N = 5$, $d_y = 1\lambda$, $\theta_0 = 30^\circ$ and $\varphi_0 = 45^\circ$.

where the sum is taken over all pqs in the visible region, and

$$P_{pq} = \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^k |\tilde{A}(k_x - k_{\mathrm{xp}}, k_y - k_{\mathrm{yq}})|^2 k_\rho dk_\rho d\varphi \ , \tag{10.84}$$

with k_{xp} and k_{yq} the same as in (10.77). We can alternatively write

$$P_{pq} = \frac{1}{(2\pi)^2} \iint\limits_{\text{visible space}} |\tilde{A}(k_x - k_{\mathrm{xp}}, k_y - k_{\mathrm{yq}})|^2 dk_x dk_y \ . \tag{10.85}$$

We can extend the integration limits to infinity and use Parseval's theorem for Fourier transforms in the same way as when we treated the linear arrays. The result is

$$P_{pq} = \iint\limits_A |A(x,y)|^2 dxdy \tag{10.86}$$

independent of the values of p and q. The latter integral is only taken over the aperture area A because the excitation distribution $A(x,y)$ is zero outside A. The area A is assumed to extend half an element spacing outside the outer elements. Eq. (10.86) is very easy to evaluate, and makes (10.83) a convenient expression for the power integral.

The *directivity* can be separated in several different factors by using (10.63), (10.75), (10.83) and (10.86), according to

$$\begin{aligned} D &= 4\pi(|\mathbf{G}(\theta_0, \varphi_0)\mathsf{AF}(k\sin\theta_0\cos\varphi_0, k\sin\theta_0\sin\varphi_0) \cdot \hat{\mathbf{co}}^*|^2/P) \\ &= e_{\mathrm{grt}}\cos\theta_0 e_{\mathrm{pol}} e_{\mathrm{ill}} D_{\max} \ . \end{aligned} \tag{10.87}$$

The different factors are the *maximum available directivity*

$$D_{\max} = \frac{4\pi}{\lambda^2} A \ , \tag{10.88}$$

where A is the aperture area. the aperture *illumination efficiency*

$$e_{\mathrm{ill}} = \frac{1}{A} \left| \iint\limits_A A(x,y)dxdy \right|^2 / \iint\limits_A |A(x,y)|^2 dxdy \ , \tag{10.89}$$

the *polarization efficiency*,

$$e_{\mathrm{pol}} = |\mathbf{G}(\theta_0, \varphi_0) \cdot \hat{\mathbf{co}}^*|^2 / \left\{ |\mathbf{G}(\theta_0, \varphi_0) \cdot \hat{\mathbf{co}}^*|^2 + |\mathbf{G}(\theta_0, \varphi_0) \cdot \hat{\mathbf{xp}}^*|^2 \right\} \ , \tag{10.90}$$

and the *grating efficiency*,

$$e_{\mathrm{grt}} = |\mathbf{G}(\theta_0, \varphi_0)|^2 / \left\{ \sum_{pq} |\mathbf{G}(\theta_{pq}, \varphi_{pq})|^2 \frac{\cos\theta_0}{\cos\theta_{pq}} \right\} \ . \tag{10.91}$$

The sum in the latter is taken over all visible grating- and main-lobes. This factor therefore represents the power lost in the grating-lobes, and it is very accurate [77].

Thus, the directivity of a planar array decreases with $\cos\theta_0$ when steered an angle θ_0 from broadside. This is due to the projection of the aperture into the observation direction. The aperture illumination efficiency is the same as that of a plane aperture. It is unity for uniform

illumination. The grating efficiency accounts for the power loss due to grating-lobes. Note that all the above formulas are independent of the shape of the aperture, i.e., the contour of $A(x, y)$. The most common shapes are circular, quadratic or rectangular. Some forms of $\tilde{A}(x, y)$ for different apertures and aperture distributions are given in Chapter 5.

The directivity can be evaluated quite straightforwardly by using (10.87) and (10.88). For a rectangular array (i.e., an array with a rectangular aperture) the excitation distribution may often be separable in x- and y. Then, the double integrals in (10.89) are separable in single integrals.

10.3.6 Determination of the realized gain

The realized gain G_{arr} of an array with equispaced equal element can be determined directly from the directivity formula in (10.87), by multiplying it with the total radiation efficiency in (2.98) providing this is the same on all elements. The total radiation efficiency contains the mismatch factor e_r and the radiation efficiency e_{abs}. This gives

$$G_{\text{arr}} = e_r e_{\text{abs}} e_{\text{grt}} \cos\theta_0 e_{\text{pol}} e_{\text{ill}} D_{\max} \ . \tag{10.92}$$

If we know the far-field and efficiency of the embedded element we can always find the correct far-field of the full array by using the most general sum (10.4). Then, the total available power is also given by the sum of the power available on all ports of the array, so we can determine the realized gain of the full array directly from its far-field and the total available power. The formula will be the same independent of how the ports are excited, i.e., independent of the values of A_n and Φ_n in (10.4). This must be stressed that such approach can only be used when the far-fields of the embedded elements are known including the mutual couplings and in particular the embedded element efficiency. There is more about this in Section 11.4, where the realized gain is shown to become very simple when all elements are excited with the same amplitude and phase, i.e.,

$$G_{\text{arr}} = MNG_{\text{emb}} \tag{10.93}$$

where G_{emb} is the realized gain of the embedded element.

10.3.7 Example: Design of planar array

We want to design a planar array of equal pyramidal horn antenna elements (with smooth metal walls). The horns are located side by side as close as possible, and they are excited with the same amplitude and a linearly varying phase. The element spacing is uniform (but can be different) in the two principal planes. Assume that we have a loss of 0.5 dB due to radiation into grating-lobes.

a) Determine the aperture widths of the whole array when we require that the directivity for broadside radiation shall be larger than 40.5 dB, and that the main beamwidths in the two principal planes shall be the same.

b) Determine the element spacing and total number of elements in H-plane, if we require that we shall be able to phase steer the beam to 30° from broadside without having radiating grating-lobes.

c) Assume that the element spacing in E-plane is $d_{_E} = 1.98\lambda$. Determine the direction of the first grating-lobe when the main beam is in the broadside direction.

d) Use a figure in the chapter about horn antennas to determine the level of the grating-lobe in step c) when the horn is very long. Explain.

e) Use the same figure to determine the minimum length of the horn allowed to keep the grating-lobe level more than about 10 dB below the main beam maximum.

f) Use the same figure to estimate how much we can phase steer the beam, of the array in step e), in E-plane if we require that the grating-lobe level should not be higher than -6 dB relative to the main beam maximum. Explain.

g) Determine the directivity of the array in step f).

SOLUTION:

a) The requirement of equal beamwidths in the E- and H-planes will be satisfied by a quadratic array. The directivity is given by (10.87). We have $(e_{\rm grt})_{\rm dB} = -0.5\,{\rm dB}$ and a uniform excitation gives $(e_{\rm ill})_{\rm dB} = 0\,{\rm dB}$. Therefore, the directivity without grating-lobes must be $40.5\,{\rm dB} + 0.5\,{\rm dB} = 41\,{\rm dB}$. From $(D_{_0} = (4\pi/\lambda^2)A = 4\pi(W/\lambda)^2)$ we get the aperture width

$$\frac{W}{\lambda} = \sqrt{D_{_0}/(4\pi)} = \sqrt{10^{4.1}/(4\pi)} = 31.65 \ .$$

b) We phase steer to $\theta_{_0} = 30^\circ$ from broadside in H-plane. We choose $\varphi = 0$ as H-plane, use (10.79) with $q = 0$ and $p = \pm 1$. This gives the following condition on the element spacing in H-plane for avoiding grating-lobes:

$$d_{_H} \leq \frac{\lambda}{1 + \sin\theta_{_0} + \frac{\lambda}{W}} = 0.65\lambda \ .$$

The total number of elements is $N_{_H} = W/d_{_H} = 49$.

c) The direction of the first grating-lobe in E-plane (i.e., $\varphi_{_0} = 90^\circ$) when the main beam radiates broadside is

$$\sin\theta_{_{\pm 1}} = \sin\theta_{_0} \pm \frac{\lambda}{d_{_E}} = \pm\frac{\lambda}{d_{_E}} = \pm 0.5051 \ .$$

This gives $\theta_{_{\pm 1}} = \pm 30.33^\circ$ from broadside.

d) The aperture diameter of the horn is equal to the element spacing, $b = d_{_E}$. The first grating-lobe appears when $b\sin\theta/\lambda = 1.98\sin(30.33^\circ) = 1.0$. In Fig. 8.9 we see that grating-lobe coincides with the first null of the element pattern for a long horn with constant phase over the aperture.

e) From Fig. 8.10 we see that we must keep the phase variation $\Delta\phi = kb^2/(8L_{\rm hrn})$ over the horn aperture smaller than 90° in order to ensure that the first grating-lobe is more than 10 dB below the main beam maximum. Thus, as $b = 2.98\lambda$ we get the minimum horn length

$$L_{\rm hrn} = \frac{kb^2}{8\Delta\phi} = \frac{360\cdot(1.98)^2\lambda}{8\cdot 90} = 1.96\lambda \ .$$

f) We use again Fig. 8.10. The spacing between the grating-lobe and the main lobe is λ/b in $\sin\theta$ space, i.e., 1 in $(b/\lambda)\sin\theta$ space. We can therefore displace an interval of width 1 as much as we can before the level difference between the two ends of it is 6 dB. We find that we can phase steer to $(b/\lambda)\sin\theta_0 = 0.25$, i.e., $\theta_0 = 7.25°$, before the grating-lobe will be more than 6 dB less than the main beam maximum.

g) The directivity varies with main beam direction θ_0 according to

$$D = \eta_{\rm grt} \cdot \cos\theta_0 \cdot D_{\max} \ .$$

The relative level of the grating-lobe of -6 dB corresponds to a relative power of 0.25. The corresponding efficiency reduction becomes by using (10.91) with one radiating grating-lobe in addition to the main lobe:

$$10\log\left(\frac{1}{1+0.25}\right) = -0.96\,\text{dB} \ .$$

The directivity becomes $(41 - 0.035 - 0.96)\,\text{dB} = -40\,\text{dB}$.

10.4 Complementary comments by S. Maci

10.4.1 Embedded element pattern and open-circuit element pattern

The existing publications on array antennas give some alternative descriptions of the far-field function of array antennas than those given in this chapter. These are summarized here.

As underlined in Section 10.2, the array design must take into account that the elements are affected by the mutual coupling to the neighboring elements. The currents on the neighboring elements, produce a reaction on the element itself that modifies both its current and input impedance. An N-element array can therefore be regarded as an N-port device with its own impedance matrix (or, equivalently, admittance or scattering matrix). However, the knowledge of the N port scattering matrix and the far-field function of the isolated element is not sufficient to obtain the far-field function of the embedded element. In other words, the far-field function of the isolated element is in general different from that of the embedded element.

Assume a Thevenin representation, where each port of an array is fed by a unit voltage with an internal series impedance. Next, consider the same array, when only one single element in the array is excited by its voltage source, and all the other elements are passively terminated in matched loads (i.e., the voltage sources are short-circuited). The far-field function obtained in the latter case is the so-called embedded element far-field pattern used in the descriptions in the previous sections. This embedded $|\mathbf{G}_n^{\rm ET}(\hat{\mathbf{r}})|$ of the n-th array element accounts for the power absorbed by the other elements and provides the far-field function of the whole array by

$$\sum_n V_n \mathbf{G}_n^{\rm ET}(\hat{\mathbf{r}}) e^{jk\mathbf{r}_n\cdot\hat{\mathbf{r}}} \ , \tag{10.94}$$

where V_n is the voltage excitation of the n-th element, $\hat{\mathbf{r}}$ is the direction of observation and $\mathbf{r}_n$ is the coordinate of the array element location. This corresponds to (10.4) when taking into account that the far-field functions of each of the embedded elements may be different, and, in particular, the far-field functions for elements near the edge of the array will be different from those in the center. We have also introduced a superscript ET to denote that these far-field functions are for elements that are Embedded and Terminated, to separate then from the far-fields for the open-circuit case below.

Alternatively, the array can be described by the open-circuit far-field function [8]-[10] of an embedded element. Note that this is also a far-field of an embedded element, but it is obtained under a different condition, namely that all the non-excited ports are open-circuited. We choose to call this an open-circuited embedded far field function, to separate it from the embedded far-field function used elsewhere in this book. Thus, an embedded far-field function means by default that all non-excited elements are terminated with the port impedance, whereas we add the explanatory "open-circuit" term if they are open-circuited.

It is clear that an array never operate with open-circuited ports. However, this concept helps to understand the assumptions underlying some often encountered approximations. The open-circuit far-field $\mathbf{G}_n^{\mathrm{OC}}(\hat{\mathbf{r}})$ of a given element in an array is the far-field function obtained when the element of interest is excited with a unit current source, while all the other elements are left open-circuited. It is possible to prove that, once all $\mathbf{G}_n^{\mathrm{OC}}(\hat{\mathbf{r}})$ are known, along with the array impedance (or scattering) matrix $[Z_A]$ (or $[S_A]$), the embedded element far-field functions $\mathbf{G}_n^{\mathrm{ET}}(\hat{\mathbf{r}})$ can be obtained for any set of internal impedances of the generators feeding the array. To this end, define $[Z_L] = diag(Z_L^{(n)})$ as a diagonal matrix whose diagonal elements $Z_L^{(n)}$ contains all the array terminations. Then, we get from [8]-[10]

$$\left[\mathbf{G}_n^{\mathrm{OC}}\right] = ([Z_A] + [Z_L]) \cdot \left[\mathbf{G}_n^{\mathrm{ET}}\right] \ , \tag{10.95}$$

where $[\mathbf{G}]$ means a single column matrix containing the vector far-field function $\mathbf{G}$. Therefore, the embedded far-field function can be computed from the open-circuited one.

If the array is receiving an incident plane wave $\mathbf{E}_i(\hat{\mathbf{r}})$ coming from the direction $-\hat{\mathbf{r}}$, the voltages across the terminals can be obtained from the above by applying reciprocity to the element of interest in its array environment. When the elements are open-circuited or terminated with the load $Z_L^{(n)}$, the voltages across the terminal of the n-th element are

$$V_n^{\mathrm{OC}} = -\frac{2j}{\eta}\mathbf{G}_n^{\mathrm{OC}}(\hat{\mathbf{r}}) \cdot \mathbf{E}_i(\hat{\mathbf{r}}) \qquad \text{or} \qquad V_n^L = -\frac{2j}{\eta} Z_L^{(n)} \mathbf{G}_n^{\mathrm{ET}}(\hat{\mathbf{r}}) \cdot \mathbf{E}_i(\hat{\mathbf{r}}) \ ,$$

respectively. Both are in full agreement with (2.129). From (10.95) one has

$$\left[V^L\right] = [Z_L]\left([Z_A] + [Z_L]\right)^{-1}\left[V^{\mathrm{OC}}\right] \ , \tag{10.96}$$

where $[V] = \{V_n\}_{n=1,\ldots,N}$ denote column vectors containing the received voltages. Eq. (10.96) is consistent with the external characterization of a general N-port network [11]. Self-contained proof of (10.96) is provided in [12], based on reciprocity concepts from [13].

The currents on the elements that are open-circuited are normally weaker than the currents on those that are terminated in matched load. There exist "minimum scattering" antennas [14] which scatter very weakly when they are open-circuited. Only a few antennas types satisfy this condition; among them there are thin dipole antennas, except when they are

placed in very dense arrays (spacings less than about $\lambda/4$). When the minimum scattering approximation holds, an element do not interact with other elements when they are open-circuited, and the open-circuit far-field function becomes very similar to the isolated-element far-field function $\mathbf{G}_n^{\mathrm{ISO}}(\hat{\mathbf{r}})$, i.e.,

$$\mathbf{G}_n^{\mathrm{ISO}}(\hat{\mathbf{r}}) \approx \mathbf{G}_n^{\mathrm{OC}}(\hat{\mathbf{r}}) \ . \tag{10.97}$$

This approximation represents a large simplification in array analysis, because the isolated-element far-field function can often be found analytically, and it is the same for all elements if the elements are identical. Thus, if (10.97) is satisfied, we can determine the complete array performance from the isolated element patterns and the mutual impedance matrix. Otherwise not. Still, in some other cases it may also be possible to do this, such as when there is very small mutual coupling between the elements.

The open-circuit far-field function corresponds to a short-circuited far-field function for slots in ground plane. The treatment is equivalent. Slots can be also viewed as "minimum scattering" antenna, in the sense than when they are short-circuited, the only contribution is the reflection from the smooth infinite ground plane without a slot.

The open-circuit (or short-circuit) far-field function will be equal to the isolated element far-field function if the elements radiate via a single antenna mode, such as thin dipoles (or narrow small slots). Then, the shape of the far-field function of the element will be the same independently of how it is excited, i.e., independent on whether the excitation is via a current on its port or via mutual coupling.

10.4.2 MoM for infinite periodic array through periodic Green's function

For sufficiently large periodic arrays of identical elements [15], infinite-array approaches will give useful insight into the behavior of the elements. While FEM and FDTD [23] have been developed for such simulations, we will here explain how this can be simulated by integral-equation approaches. The main difference with respect to finite-array approaches lies in the use of periodic Green's functions, the convergence of which has been the subject of intensive research.

The infinite array solution can be obtained with the MoM by simply replacing the single source Green's function by an infinite series. This series involves the same type of terms, each term being related to a "copy" of the source in other cells of the array and multiplied by a phase factor $e^{-j\mathbf{r}_{nm}\cdot\mathbf{k}_0}$. This phase factor takes into account the linear phase progression along the array. In other terms, the MoM formulation will have the same formal expression as for normal Green's functions, provided the individual element Green's function is substituted by the periodic Green's function

$$g_a(\mathbf{r}, \mathbf{k}_0) = \sum_{n,m=-\infty}^{\infty} g_0(\mathbf{r} - \mathbf{r}_{mn}) e^{-j\mathbf{r}_{mn}\cdot\mathbf{k}_0} \ , \tag{10.98}$$

Where $\mathbf{r}_{mn} = md_x\hat{\mathbf{x}} + nd_y\hat{\mathbf{y}}$ and $\mathbf{r} = (x, y)$ are the positions of array point source elements. This explicit space-domain summation exhibits a poor convergence. It can be converted in a spectral-domain series by a Poisson summation. For free-space Green's function, this leads

to

$$g_a(\mathbf{r},\mathbf{k}_0)=\frac{1}{d_x d_y}\sum_{p,q=-\infty}^{\infty}\frac{e^{-jk_{\mathrm{zpq}}|z|}}{8\pi^2 jk_{\mathrm{zpq}}}e^{-j\mathbf{r}\cdot(\mathbf{k}_0+\mathbf{k}_\infty)}\ , \tag{10.99}$$

where

$$\mathbf{k}_{pq}=\mathbf{k}_0+\frac{2\pi p}{d_x}\hat{\mathbf{x}}+\frac{2\pi q}{d_y}\hat{\mathbf{y}}\qquad\text{and}\qquad k_{\mathrm{zpq}}=\sqrt{k^2-\mathbf{k}_{pq}\cdot\mathbf{k}_{pq}}\ .$$

Eq. (10.99) is also known as Floquet-Wave (FW) representation of the array Green's function. From the expression of k_{zpq} it can be seen that, when both d_x, and d_y are smaller than $\lambda/2$ and $|\mathbf{k}_0|<k$ only the dominant (0,0 indexed) plane-wave of the series representation is propagating in z direction, while all the other are attenuated exponentially. If indeed $d_{x,y}\in(\lambda/2,\lambda)$, one higher order backward FW will emerge from cut-off and start to propagate when we increase $|\mathbf{k}_0|$. This corresponds the case in which a grating lobe enters into the visible range.

For p and q large, the exponent at the right hand side of (10.99) is of the type $e^{-\sqrt{\mathbf{k}_{pq}\cdot\mathbf{k}_{pq}}|z|}$ which means that the convergence versus p and q is very fast for large $|z|$ and becomes slow for small values of $|z|$. This behavior is analogue to that in a single-mode waveguide, in which only one mode propagates and all the others are in cut-off.

The slow convergence for small $|z|$ is generally referred to as the "on-plane" convergence [19]. Improving the on-plane convergence is important when calculating the mutual coupling of coplanar elements. There are many techniques being used to speed up the convergence. One way is to use formulas for series accelerations, like the Shanks formula and the Levin-T method [20]. Alternative methods are based on a combination of space domain and spectral-domain approaches, like the Ewald method [21]-[24]. Two-dimensional arrays of point sources can also be viewed as a subset of 3D arrays, which led to finding a method with exponential convergence [25]. A detailed review of periodic Green's functions can be found in [26].

10.4.3 MoM for finite periodic array

The infinite array approach, although important in preliminary design, do neither account for any desired amplitude variation of the excitation along the array, nor for edge effects due to the truncation of the infinite size. To overcome this limitation a finite-array analysis [27] can be carried out by the "windowing" method. This method consists of a windowing done on the periodic Green's function in (10.98) within a MoM solution. In other words, in place of using in the integral equation kernel $g_a(\mathbf{r},\mathbf{k}_0)$, one use $g_a(\mathbf{r},\mathbf{k}_0)A(\mathbf{r})$, where $A(\mathbf{r})$ is the amplitude excitation distribution function introduced in Section 10.3.2 that follows the port excitations of the array and is zero outside the array boundary. This leads to equivalent formulations, one in space domain [27]-[29] if (10.98) is used and the other one in spectral domain [30] if (10.99) (or an analogous form for multilayer) is used. We note that, since the windowing function breaks the shift-invariance form of the array Green's function, each element solved by MoM provides different currents and then a different embedded element pattern.

Windowing methods provide good accuracy, especially when the effects of the truncation is moderate, with a numerical complexity that is almost the same as that of the one for solving a single periodic cell of the array. This means for low edge excitation of the array, or for co-linear arrays of dipoles. Therefore, research has been devoted to speed up the calculation of large finite-array or *truncated* periodic Green's functions $g_a(\mathbf{r},\mathbf{k}_0)U(\mathbf{r})$ where $g_a(\mathbf{r},\mathbf{k}_0)$ is represented as in (10.99) and $U(\mathbf{r})$ is 1 inside the array region and 0 outside. This

is done by the Truncated Floquet Waves (TFW) method. In this method, the edge effect is represented by UTD-like diffraction of the Floquet Waves in (10.99) at the array truncation. The diffraction effects can be asymptotically isolated by canonical problems of semi-infinite arrays [31]-[33] or corners arrays [34]-[35]. For arrays of relatively simple elements, like slots or dipoles, asymptotic solutions for semi-infinite arrays enabled the representation of currents resulting from edge effects with the help of basis functions that cover the whole array domain [36]-[38]. The TFW method can also explain the effects of global oscillation of the scan impedance along the array, first observed in [39]. This phenomenon can be explained by interference between the current of the element in infinite arrays and the diffracted Floquet wave contributions coming form the edges.

Today, the approximate windowing methods are not so much used, despite they are extremely fast when used in conjunction with TFW expansions. The reason is the strong evolution of MoM formulations for large problems such as Fast Multiple Methods (FMM) [40]-[41], Adaptive Integral Method (AIM) [42], Adaptive Cross Approximation (ACA) [43]-[44], and incomplete QR methods [45]-[46]. These methods are today able to handle array problems up to dimensions of twenty by twenty wavelengths. Particularly important in this context is the significant development of non-iterative methods, based on a reduction of the effective number of unknowns. These methods consist of aggregating basis functions into relatively small sets defined over every unit cell of the array [47]-[58].

10.5 Practical array antennas

Antenna arrays can be made with almost any types of antenna elements. Many of them are for military radar applications, and then there is not so much published on the electromagnetic design. Common elements in military systems are waveguides [59], waveguide slot antennas [60]-[62]. Array antennas can also be very wideband such as several dense array designs in [63], and the one in [64]. Microstrip antenna arrays are common [65], particularly in base station [66]. They have also the advantage that they can be made conformal [67]-[68]. The production cost can also be made low for helical and spiral arrays [69], and parallel plate slot arrays [70]. At millimeterwaves the microstrip technology gives too much losses, so it can be replaced by other planar technologies such as post-wall waveguides [71], also called Substrate Integrated Waveguides (SIW) [72], and different gap waveguide technologies are also under development for use at such high frequencies [73]-[75]. The laminated waveguide is also attractive [76].

10.6 Exercises

1. **Linear dipole array**: Consider a linear array of 20 transverse dipoles located with an element spacing of 0.5 wavelengths at 2 GHz. The dipoles are series fed from a straight rigid dielectric-filled coaxial line (relative permittivity $\varepsilon_r = 2.5$). The signal is coupled with each dipole via directional couplers in a way that the amplitude excitation is the same of all dipoles. The spacing between the directional couplers is the same as the element spacing.
   a) Determine the main lobe direction and the beamwidth.
   b) Determine the element spacing needed to get broadside radiation.
   c) Use the element spacing in step b) and determine the main beam direction at 2.3 GHz.

2. **Linear waveguide slot array**: Consider a linear array of 10 longitudinal slots in the broad wall of a rectangular waveguide. Each slot is weakly coupled to the waveguide and we assume that all slots are excited by the same forward travelling wave. Further, we assume no reflections in the waveguide. Every second slot is located on opposite sides of the center line of the wall, why? The cut-off frequency of the waveguide is 5 GHz.

   a) Determine a formula for the element spacing which is needed in order to get the main beam in the broadside direction. Evaluate this spacing for a frequency of 7 GHz.

   b) Find the main beam direction when the spacing is that evaluated in step a) and the frequency is 8 GHz.

   c) Assume that all the elements are excited with equal amplitude. This can in practice be done by increasing the offset from the center line along the waveguide. What is the beamwidth between the two first nulls in the radiation pattern in the plane of the waveguide axis. Evaluate it both at 7 GHz and 8 GHz.

   d) Determine the bandwidth of the antenna when we require that the radiation level in the broadside direction $\theta = 90^\circ$ should not vary by more than 1 dB over the frequency band.

3. **Linear microstrip array**: Consider an array of rectangular microstrip antennas on a very thin substrate with high permittivity. The radiation pattern of each single patch can be approximated as that of two magnetic line currents located at each end of the patch.

   a) Assume that the two magnetic line currents are so close that they radiate in the same way as one single incremental magnetic dipole. What is then the radiation pattern in E-plane? Sketch it.

   b) We want to use the above elements in a vertical linear array to get a double 3 dB beamwidth of 5° at broadside in the vertical plane with as few elements as possible. How long does the array need to be?

   c) How many elements do we need in order to avoid grating-lobes when the main lobe points at broadside?

   d) How many elements do we need in order to avoid grating-lobes when the main lobe is steered to 20° from broadside?

   e) What will the relative level of the grating-lobe be if we use the element spacing in step d and scan to 35°? Evaluate the grating efficiency in dB.

4. **Scan impedance of array of two dipoles**: Consider two thin half-wave electric dipoles located side-by-side with a spacing $d = .25\lambda$.

   a) Assume that each separate dipole is resonant in free space. What is the value of the input impedance of dipole 1 when dipole 2 is open-circuited?

   b) Use Fig. 10.10 to find the value of the input impedance of dipole 1 when dipole 2 is short-circuited.

   c) Find also the input impedance of dipole 1 when the dipoles are excited by equal amplitude for broadside radiation. What is the reflection coefficient when we feed both of them by a 75 Ω coaxial line?

   d) Assume that we conjugate match the dipoles when they radiate at broadside. (In practice this can be done by tuning the dipole length to get the scan reactance zero, and using a quarter-wave transformer to transform the scan resistance to become equal to the characteristic impedance of the feed line.) We here choose to model this approximately by an equivalent circuit consisting of a feed line with characteristic impedance equal to the scan resistance at broadside, and by using a tuned reactance in series with the scan impedance. The latter is equal to the negative of the scan reactance at broadside. Find the reflection coefficient when the dipoles are phase-steered to radiate 30° from broadside. Find also the reflection coefficient when they are phase-steered to the opposite side of broadside.

5. **Design of circular planar array**: We want to design an array with a circular aperture by using as few rectangular microstrip elements as possible. The elements are located in a regular

quadratic grid. The array should be phase-steered to 20° from broadside in all directions. The frequency is 10 GHz.

a) Determine the diameter of the aperture both in wavelengths and centimeter when we require that the directivity shall be at least 43 dB over the whole scan range.

b) Determine thereafter the element spacing and the number of elements needed.

c) What is the 3 dB beamwidth at broadside?

d) The beam at 20° has an elliptical cross section. Find the minimum and maximum 3 dB beamwidths. In which planes do they appear?

6. **Design of rectangular planar array**: We want to design an array antenna with a rectangular aperture by using as few elements as possible. The elements are open rectangular waveguides in a large ground plane. The waveguide apertures have width $w \ll \lambda$ and length $l = 0.6\lambda$.

   a) We want the 3 dB half-beamwidth of the array to be 1° in H-plane when the array is excited for broadside radiation. Determine the diameter of the array in H-plane.

   b) We want the directivity at broadside to be 45 dB. Determine the diameter of the array and the 3 dB beamwidth in E-plane.

   c) We want to scan the array out to 60° from broadside in E-plane without grating-lobes. There is no scanning in H-plane. Determine the required number of elements in both E- and H-planes.

   d) What is the directivity and beamwidth when the beam is scanned to 60° from broadside in E-plane?

   e) Use the number of elements in step c) and determine the directivity when the array is scanned to 75°.

7. **Beamwidth of steered planar array**: Derive the expression in equation (10.76) for the variation of the beamwidth with the main beam direction. It is most simply derived in the two planes $\varphi = \varphi_0$ and $\varphi = \varphi_0 + \frac{\pi}{2}$.

8. **Array of rectangular horns**: Consider a large planar array of pyramidal horn antennas with quadratic apertures with $3\lambda \times 3\lambda$ size. The horn is very long, so the phase can be considered constant over the aperture. Assume that the horns are polarized in y-direction. Assume that each isolated element is radiating from an infinite ground plane.

   a) Write down the expression for the far-field function of each element. Determine the 3 dB beamwidths in E- and H-plane. Draw the location of the first null of the element pattern into a $\sin\theta\cos\varphi$, $\sin\theta\sin\varphi$ diagram.

   b) Find the locations of the grating-lobes and draw them into the same diagram.

   c) Determine the relative levels of the grating-lobes when the array radiates in the broadside direction.

   d) Determine the relative levels of the grating-lobes when we phase steer the beam in E-plane to 10° from broadside.

   e) Determine the relative levels of the grating-lobes when we phase steer the beam in H-plane to 10° from broadside.

10.7 References

[1] A. Mortazawi, T. Itoh and J. Harvey, *Active Antennas and Quasi- Optical Arrays*, IEEE Press, 1998.

[2] R.C. Hansen, *Microwave Scanning Antennas: Vol II, Array Theory and Practice*, Academic Press 1964.

[3] N. Amitay, V. Galindo and C.P. Wu, *Theory and Analysis of Phased Array Antennas*, Wiley-Interscience, 1972.

[4] A.C. Ludwig, "Mutual coupling, gain, and directivity of an array of two identical antennas", *IEEE Transactions on Antennas and Propagation*, pp. 837-841, November 1976.

[5] P.W. Hannan , "The element-gain paradox for a phased-array antenna", *IEEE Transactions on Antennas and Propagation*, Vol. AP-12, No. 7, pp. 423-433, July 1964.

[6] W.K. Kahn, "Element efficiency: A unifying concept for array antennas", *IEEE Antennas and Propagation Magazine*, Vol. 49, No. 4, pp. 48-56, August 2007.

[7] P.-S. Kildal and K. Rosengren, "Correlation and capacity of MIMO systems and mutual coupling, radiation efficiency and diversity gain of their antennas: Simulations and measurements in reverberation chamber", *IEEE Communications Magazine*, Vol. 42, No. 12, pp. 102-112, December 2004.

[8] W. Kahn, "Impedance-match and element-pattern constraints for finite arrays", *IEEE Transactions on Antennas and Propagation*, Vol. 25, No. 6, pp. 747–755, November 1977.

[9] J.W. Wallace and M.A. Jensen, "Mutual coupling in MIMO wireless systems: a rigorous network theory analysis," *Wireless Communications, IEEE Transactions on*, Vol. 3, No. 4, pp. 1317–1325, July 2004.

[10] K.F. Warnick and M.A. Jensen, "Effects of Mutual Coupling on Interference MitigationWith a Focal Plane Array", *IEEE Transactions on Antennas and Propagation*, Vol. 53, No. 8, pp. 2490–2498, August 2005.

[11] I. Gupta and A. Ksienski, "Effect of mutual coupling on the performance of adaptive arrays," *IEEE Transactions on Antennas and Propagation*, Vol. 31, No. 5, pp. 785–791, September 1983.

[12] R. Maaskant, "Analysis of large antenna arrays," Ph.D. dissertation, Eindhoven University of Technology, Eindhoven, 7 June 2010.

[13] A.T. de Hoop, "The n-port receiving antenna and its equivalent electrical network", *Philips Res. Repts., Tech. Rep.*, 1975.

[14] A.J. Roscoe and R.A. Perrott, "Large finite array analysis using infinite array data", *IEEE Transactions on Antennas and Propagation*, Vol. 42, No. 7, pp. 983–992, July 1994.

[15] H. Holter and H. Steyskal, "On the size requirement for finite phased-array models," *IEEE Transactions on Antennas and Propagation*, Vol. 50, No. 6, pp. 836–840, June 2002.

[16] H. Holter and H. Steyskal, "Infinite phased-array analysis using FDTD periodic boundary conditions-pulse scanning in oblique directions", *IEEE Transactions on Antennas and Propagation*, Vol. 47, No. 10, pp. 1508–1514, October 1999.

[17] A. Papoulis, *The Fourier Integral and its Applications*, McGraw-Hill , New-York (1962).

[18] L.B. Felsen, and N. Marcuvitz, *Radiation and Scattering of Waves*, Prentice-Hall, Englewood Cliffs, NJ, 1973, also IEEE Press, Piscataway, NJ, 1994.

[19] R.E. Jorgenson and R. Mittra, "Efficient calculation of the free-space periodic Green's function", *IEEE Transactions on Antennas and Propagation*, Vol. 38, No. 5, pp. 633–642, May 1990.

[20] S. Singh and R. Singh, "On the use of Levin's T-transform in accelerating the summation of series representing the free-space periodic Green's functions", *Microwave Theory and Techniques, IEEE Transactions on*, Vol. 41, No. 5, pp. 884–886, May 1993.

[21] P.P. Ewald, "Die berechnung optischer und elektrostatischen gitterpotentiale", *Annalen der Physik*, Vol. 369, No. 3, pp. 253–287, October 1921.

[22] F. Capolino, D.R. Wilton, and W.A. Johnson, "Efficient computation of the 2- D Green's function for 1-D periodic structures using the Ewald method," *IEEE Transactions on Antennas and Propagation*, Vol. 53, No. 9, pp. 2977–2984, September 2005.

[23] I. Stevanovic and J.R. Mosig, "Periodic Green's function for skewed 3-D lattices using the Ewald transformation", *Microwave and Optical Technology Letters*, Vol. 49, pp. 1353– 1357, June 2007.

[24] G. Valerio, P. Baccarelli, P. Burghignoli, and A. Galli, "Comparative analysis of acceleration techniques for 2-D and 3-D Green's functions in periodic structures along one and two directions", *IEEE Transactions on Antennas and Propagation*, Vol. 55, No. 6, pp. 1630–1643, June 2007.

[25] M.G. Silveirinha and C.A. Fernandes, "A new acceleration technique with exponential convergence rate to evaluate periodic Green functions", *IEEE Transactions on Antennas and Propagation*, Vol. 53, No. 1, pp. 347–355, January 2005.

[26] C. Craeye, X. Radu, A. Schuchinsky, and F. Capolino, "Fundamentals of method of moments for metamaterials", in *Handbook of Metamaterials*, F. Capolino, Ed. Taylor and Francis, 2009.

[27] A. Roederer, "Etude des réseaux finis de guides rectangulaires à parois épaisses", *Londe Electrique*, Vol. 51, pp. 854–861, November 1971.

[28] B. Munk and G. Burrell, "Plane-wave expansion for arrays of arbitrarily oriented piecewise linear elements and its application in determining the impedance of a single linear antenna in a lossy half-space", *IEEE Transactions on Antennas and Propagation*, Vol. 27, No. 3, pp. 331–343, May 1979.

[29] A.K. Skrivervik and J.R. Mosig, "Analysis of printed array antennas", *IEEE Transactions on Antennas and Propagation*, Vol. 45, No. 9, pp. 1411–1418, September 1997.

[30] A.K. Skrivervik and J.R. Mosig, "Analysis of finite phase arrays of microstrip patches", *IEEE Transactions on Antennas and Propagation*, Vol. 41, No. 8, pp. 1105–1114, August 1993.

[31] F. Capolino, M. Albani, S. Maci, R. Tiberio, "High-frequency analysis of an array of line sources on a truncated ground-plane", (pubblicazione su rivista, 1998) *IEEE Transactions on Antennas and Propagation*, Vol. 46, pp. 570–578, 1998.

[32] F. Capolino, M. Albani, S. Maci, and L. B. Felsen, "Frequency-domain Green's function for a planar periodic semi-infinite phased array I. Truncated floquet wave formulation", *IEEE Transactions on Antennas and Propagation*, Vol. 48, No. 1, pp. 67–74, January 2000.

[33] F. Capolino, M. Albani, S. Maci, L. Felsen, "Frequency domain Green's function for a planar periodic semi-infinite phased array. Part II: diffracted wave phenomenology", (pubblicazione su rivista, 2000) *IEEE Transactions on Antennas and Propagation*, Vol. 48, pp. 75–85, 2000.

[34] F. Capolino, S. Maci, and L. Felsen, "Asymptotic high-frequency Green's function for a planar phased sectoral array of dipoles", *Radio Science*, Vol. 35, No. 2, pp. 579–593, 2000.

[35] F. Mariottini, F. Capolino, S. Maci, and L. B. Felsen, "Asymptotic high-frequency Green's function for a large rectangular planar periodic phased array of dipoles with weakly tapered excitation in two dimensions", *IEEE Transactions on Antennas and Propagation*, Vol. 53, No. 2, pp. 608–620, February 2005.

[36] A. Neto, S. Maci, G. Vecchi, and M. Sabbadini, "A truncated Floquet wavediffraction method for the full wave analysis of large phased arrays. I. Basic principles and 2-D cases", *IEEE Transactions on Antennas and Propagation*, Vol. 48, No. 4, pp. 594–600, April 2000.

[37] A. Neto, S. Maci, G. Vecchi, and M. Sabbadini, "A truncated Floquet wave diffraction method for the full-wave analysis of large phased arrays .II. Generalization to 3-D cases", *IEEE Transactions on Antennas and Propagation*, Vol. 48, No. 4, pp. 601–611, April 2000.

[38] O. Çivi, P. Pathak, H.-T. Chou, and P. Nepa, "A hybrid uniform geometrical theory of diffraction moment method for efficient analysis of electromagnetic radiation/ scattering from large finite planar arrays", *Radio Science*, Vol. 35, No. 2, pp. 607–620, 2000.

[39] R.C. Hansen and D. Gammon, "A Gibbsian model for finite scanned arrays", *IEEE Transactions on Antennas and Propagation*, Vol. 44, No. 2, pp. 243–248, February 1996.

[40] R. Coifman, V. Rokhlin, and S. Wandzura, "The fast multipole method for the wave equation: a pedestrian prescription", *IEEE Antennas and Propagation Magazine*, Vol. 35, No. 3, pp. 7–12, June 1993.

[41] W.C. Chew, J.-M. Jin, E. Michielssen, and J.M. Song, *Fast and Efficient Algorithms in Computational Electromagnetics*, Norwood, MA: Artech House, 2001.

[42] E. Bleszynski, M. Bleszynski, and T. Jaroszewicz, "AIM: Adaptive integral method for solving large-scale electromagnetic scattering and radiation problems", *Radio Science*, Vol. 31, No. 5, pp. 1225–1251, 1996.

[43] M. Bebendorf, "Approximation of boundary element matrices", *Numerische Mathematik*, Vol. 86, pp. 565–589, October 2000.

[44] M. Bebendorf and S. Rjasanow, "Adaptive low-rank approximation of collocation matrices", *Computing*, Vol. 70, pp. 1–24, March 2003.

[45] N.A. Ozdemir and J.-F. Lee, "A low-rank IE-QR algorithm for matrix compression in volume integral equations", *IEEE Transactions on Magnetics*, Vol. 40, No. 2, pp. 1017–1020, March 2004.

[46] E. Suter and J.R. Mosig, "A subdomain multilevel approach for the efficient MoM analysis of large planar antennas", *Microwave and Optical Technology Letters*, Vol. 26, pp. 270–277, March 2000.

[47] L. Matekovits, G. Vecchi, G. Dassano, and M. Orefice, "Synthetic function analysis of large printed structures: The solution space sampling approach", in Proc. IEEE AP-S Soc. Int. Symp., Boston, pp. 568–571, June 2001.

[48] V.V.S. Prakash and R. Mittra, "Characteristic basis function method: A new technique for efficient solution of method of moments matrix equations", *Microwave and Optical Technology Letters*, Vol. 36, pp. 95–100, January 2003.

[49] S.J. Kwon, K. Du, and R. Mittra, "Characteristic Basis Function Method: a numerically efficient technique for analazing microwave and RF circuits", *Microwave and Optical Technology Letters*, Vol. 38, pp. 444–448, September 2003.

[50] W.B. Lu, T.J. Cui, X.X. Yin, Z.G. Qian, and W. Hong, "Fast algorithms for largescale periodic structures using subentire domain basis functions", *IEEE Transactions on Antennas and Propagation*, Vol. 53, No. 3, pp. 1154–1162, March 2005.

[51] L. Matekovits, V.A. Laza, and G. Vecchi, "Analysis of large complex structures with the synthetic-functions approach", *IEEE Transactions on Antennas and Propagation*, Vol. 55, pp. 2509–2521, September 2007.

[52] C. Craeye and R. Sarkis, "Finite array analysis through combination of macro basis functions and array scanning methods", *Applied Computational Electromagnetics Society Journal* (ACES), Vol. 23, pp. 999–1007, September 2008.

[53] R. Maaskant, R. Mittra, and A. Tijhuis, "Fast Analysis of Large Antenna Arrays Using the Characteristic Basis Function Method and the Adaptive Cross Approximation Algorithm", *IEEE Transactions on Antennas and Propagation*, Vol. 56, No. 11, pp. 3440–3451, November 2008.

[54] L. Matekovits, G. Vecchi, M. Bercigli, and M. Bandinelli, "Synthetic-functions analysis of large aperture-coupled antennas", *IEEE Transactions on Antennas and Propagation*, Vol. 57, pp. 1936–1943, July 2009.

[55] J. Laviada, F. Las-Heras, M. Pino, and R. Mittra, "Solution of electrically large problems with multilevel characteristic basis functions", *IEEE Transactions on Antennas and Propagation*, Vol. 57, No. 10, pp. 3189–3198, October 2009.

[56] D.J. Bekers, S.J.L. van Eijndhoven, and A.G. Tijhuis, "An eigencurrent approach for the analysis of finite antenna arrays", *IEEE Transactions on Antennas and Propagation*, Vol. 57, No. 12, pp. 3772–3782, December 2009.

[57] C. Craeye and D. González-Ovejero, "A review on array mutual coupling analysis", *Radio Science*, Vol. 46, No. RS2012, 2011.

[58] D. González-Ovejero and C. Craeye, "Interpolatory macro basis func- tions analysis of non-periodic arrays", *IEEE Transactions on Antennas and Propagation*, Vol. 59, No. 8, pp. 3117–3122, August 2011.

[59] H. Steyskal, "Analysis of circular waveguide arrays on cylinder", *IEEE Transactions on Antennas and Propagation*, Vol. 25, No. 5, pp. 610-616, September 1977.

[60] R.S. Elliot, "An improved design procedure for small arrays of shunt slots", *IEEE Transactions on Antennas and Propagation*, Vol. AP-31, No. 1, pp. 48-53, January 1983.

[61] L. Josefsson, "Analysis of longitudinal slots in rectangular waveguides", *IEEE Transactions on Antennas and Propagation*, Vol. 35, No. 12, pp. 1351-1357, December 1987.

[62] S. Rengarajan, "Compound radiating slots in a broad wall of a rectangular waveguide", *IEEE Transactions on Antennas and Propagation*, Vol. 37, No. 9, pp. 1116-1123, September 1989.

[63] B.A. Munk, *Finite Antenna Arrays and FSS*, Wiley, 2003.

[64] H. Holter, "Dual-polarized broadband array antenna with BOR-elements, mechanical design and measurements", *IEEE Transactions on Antennas and Propagation*, Vol. 55, No. 2, pp. 305-312, February 2007.

[65] R.J. Mailloux, J.F. McIlvenna, and N.P. Kernweis, "Microstrip array technology", *IEEE Transactions on Antennas and Propagation*, Vol. AP-29, pp. 25-37, January 1981.

[66] B. Lindmark, S. Lundgren, J.R. Sanford, C. Beckman, "Dual-Polarized Array for Signal-Processing Applications in Wireless Communications", *IEEE Transactions on Antennas and Propagation*, Vol. 46, No. 6, pp. 758-763, June 1998.

[67] L. Josefsson, P. Persson, *Conformal Array Antenna Theory and Design*, IEEE Press Series, J. Wiley, 2006.

[68] Z. Sipus, S. Skokic, M. Bosiljevac, N. Burum, "Study of Mutual Coupling Between Circular Stacked-Patch Antennas on a Sphere", *IEEE Transactions on Antennas and Propagation*, Vol. 56, No. 7, pp. 1834-1844, July 2008.

[69] N. Nakano, *Helical and spiral antennas*, Wiley, 1987.

[70] J. Hirokawa, M. Ando, N. Goto, "Waveguid-fed parallel plate slot array antenna", *IEEE Transactions on Antennas and Propagation*, Vol. 40, No. 2, pp. 218-223, February 1992.

[71] J. Hirokawa, M. Ando, "Efficiency of 76-GHz post-wall waveguide-fed parallel-plate slot arrays", *IEEE Transactions on Antennas and Propagation*, Vol. 48, No. 11, pp. 1742-1745, November 2000.

[72] X.-P. Chen, K. Wu, L. Han, F. He,"Low-Cost High Gain Planar Antenna Array for 60-GHz Band Applications", *IEEE Transactions on Antennas and Propagation*, Vol. 58, No. 6, June 2010.

[73] E. Pucci, E. Rajo-Iglesias, J.-L. Vazquez-Roy and P.-S. Kildal, "Planar Dual-Mode Horn Array with Corporate-Feed Network in Inverted Microstrip Gap Waveguide", *IEEE Transactions on Antennas and Propagation*, Vol. 62, Issue 7, pp. 3534-3542, 2014.

[74] S.A. Razavi, P.-S. Kildal, L. Xiang and E. Alfonso, "2 × 2-slot Element for 60 GHz Planar Array Antenna Realized on Two Doubled-sided PCBs Using SIW Cavity and EBG-type Soft Surface fed by Microstrip-Ridge Gap Waveguide", *IEEE Transactions on Antennas and Propagation*, Vol. 62, Issue 9, pp. 4564-4573, 2014.

[75] A.U. Zaman and P.-S. Kildal, "Wideband slot antenna array with single-layer corporate feed-network in ridge gap waveguide technology", *IEEE Transactions on Antennas and Propagation*, Vol. 62, Issue 6, pp. 2992-3001, 2014.

[76] Y. Miura, J. Hirokawa, M. Ando, Y. Shibuya, G. Yoshida, "Double-Layer Full-Corporate-Feed Hollow-Waveguide Slot Array Antenna in the 60-GHz Band", *IEEE Transactions on Antennas and Propagation*, Vol. 59, No. 8, pp. 2844-2851, 2011.

[77] A. Vosoogh and P.-S. Kildal, "Simple Formula for Aperture Efficiency Re-duction Due to Grating Lobes in Arrays," submitted to *IEEE Transactions on Antennas and Propagation*, February 2015.

Chapter 11

Fundamental limitations on directivity, array gain and bandwidth

This chapter deals with the fundamental limitations on directivity, radiation efficiency and gain of antennas with both single and multiple ports, when the antennas are made of lossless materials. The *maximum available directivity* of large antennas is given by the common aperture directivity formula. For small antennas, it is given by the directivity of the *Huygens source*. These two expressions are combined in a heuristic manner to a continuous maximum directivity limitation as a function of the diameter of the smallest sphere that can surround the antenna [1, 2]. The theoretical limitation is plotted together with directivities of antennas known to have large directivities, and the results are seen to be very close to the heuristic limit.

The chapter also deals with fundamental limitations of dense arrays. Dense linear arrays can theoretically be used to provide superdirectivity, but this remains to be proven in practice. Their claimed superdirectivity is often lower than what can be obtained with optimized single-port antennas of the same size. Dense arrays also suffer from severe gain reduction due to strong coupling to the ports of the neighbouring elements. This gain reduction can be characterized in terms of a decoupling efficiency, also referred to as embedded element efficiency, and appears as a reduction of the total radiated power. The fundamental property of this decoupling efficiency is known from classical papers, and it has recently been verified by measurements in a reverberation chamber and by simulations.

Finally the chapter discusses the fundamental bandwidth limitations of small antennas in terms of their so-called radiation Q-factor. This determines bandwidth limitations for the following classes of antennas: *i)* single-resonance-type small antennas, *ii)* wideband gradual-transition-type small antennas, and *iii)* wideband cascaded-resonances-type small antennas.

11.1 Background

Publications on fundamental limitations of small antennas normally study the relation between the antenna size and its radiation Q-factor [3]-[4]. The Q-factor is originally a measure of the quality of a narrow band resonator, and it represents the inverse relative bandwidth of the resonator. For antennas, however, we want normally larger bandwidth so the user-experienced quality is better if we have a larger bandwidth. Thus, it is always understood that when dealing with antennas the lower Q the better. The Q is only valid as a characterization method when Q is large, because it is originally based on a series expansion of frequency variations around the frequency of operation. Therefore, it becomes very questionable if Q can be used at all to characterize antennas for which it is desirable to have low Q, because, the accuracy of the characterization method becomes worse when the antenna is better. Also, the overall antenna bandwidth can easily be increased (i.e., Q can be decreased) by cascading resonance-type antenna elements with small bandwidths (i.e., large Qs). This makes the radiation Q even more questionable as a fundamental inverse measure of antenna bandwidth. Nevertheless, we will here explain with reference to [5]: *i)* How the radiation Q is related to interpretable quantities such as bandwidth-efficiency product of single-resonance-type small antennas. *ii)* How it is related to mismatch factor variations, and tolerance requirements of multiple-resonance-type small antennas. *iii)* How it is related to a radiation cut-off frequency of gradual-transition-type small antennas (related to gradual cut-off of spherical waves).

We will first overview other fundamental limitations that apply to antennas, such as maximum available directivity of single-port antennas and radiation efficiency of multi-port antennas, and thereafter discuss the practical consequences of the limitations determined by Q-factor and size. The main focus will be on small antennas, but the limitations of large antennas are overviewed as well because by mistake these limitations often are applied to small antennas. In particular, the efficiency limitations of multiport antennas for diversity and MIMO systems[1] will be overviewed. Their fundamental limitations can be characterized in terms of the radiation efficiency of each embedded element, i.e., an embedded element efficiency, and the paper will describe how these limitations are related to the fundamental limitations of dense arrays outlined in [6], [7] and [8]. This will be described by an example. The chapter will also discuss the limitations that apply when dense arrays are used to generate superdirectivity [9]-[10].

11.2 Maximum directivity of single-port antennas

11.2.1 Large antennas

The *maximum available directivity* of a large antenna covering a planar area A is well known to be given by

$$D_{0_{\max}} = \frac{4\pi}{\lambda^2} A \ . \tag{11.1}$$

This formula is only valid when the cross section of the area A is large in terms of wavelength. In practice, it applies to areas with diameters typically larger than three wavelengths. The

[1] See Section 3.3.3.

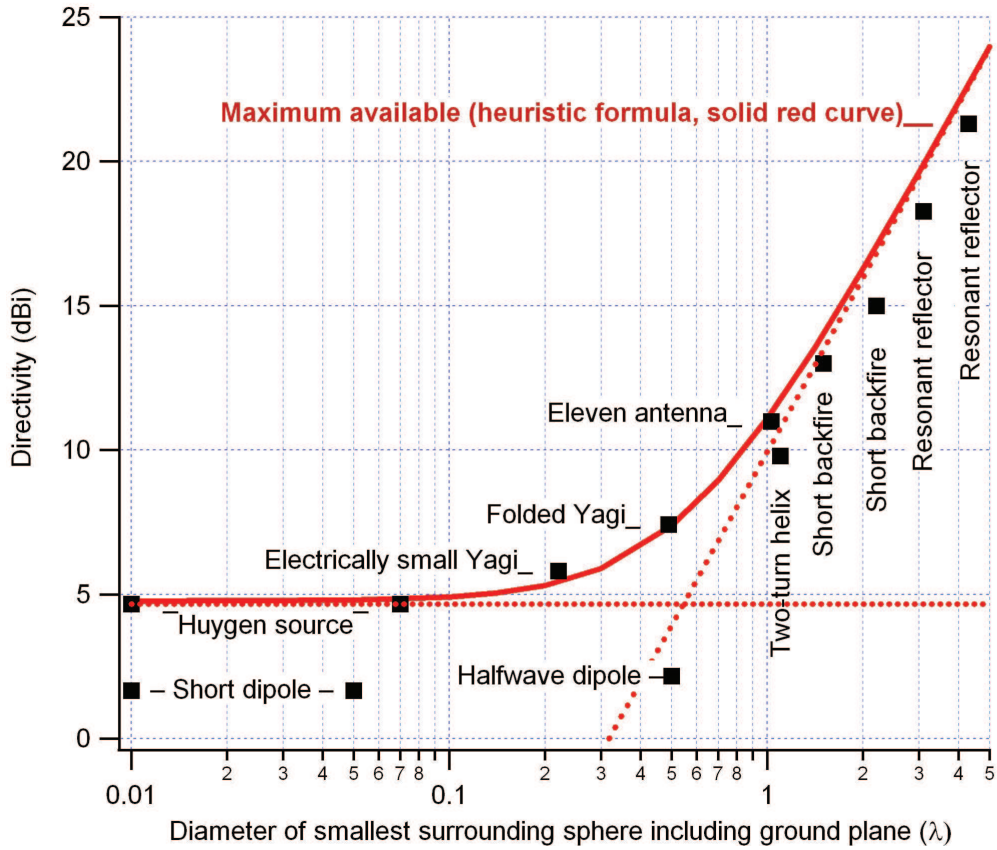


Figure 11.1: Fundamental directivity limitations of large antenna (dotted steep line), small single-port antennas (dotted horizontal line), and heuristically combined to a curve valid for any antenna size (solid line), together with directivities of some theoretical and practical antennas (squares), all as a function of the diameter of the smallest sphere that can enclose the antenna including its ground plane (if any).

formula applies both to large apertures and array antennas. The formula is sometime used also as a directivity limitation of small antennas, but this is wrong as explained in the next section. We will herein use the formula in (11.1) as a limitation also in terms of the size of the minimum sphere that can surround the antenna. If this sphere has a diameter D, we will use the following expression

$$D_{0_{\max}} = \left[\frac{\pi D}{\lambda}\right]^2 , \tag{11.2}$$

which agrees with the maximum directivity of a large circular planar aperture. This approach is justified by reference [11], which shows that the maximum available directivity of sources inside a large sphere of radius R approaches $D_{0_{\max}}$ in (11.2) for large R ($R = D/2$). The result in [11] contains also small R corrections to $D_{0_{\max}}$, but these do not converge to the known 4.8 dBi directivity of the incremental Huygens source, see Subsection 11.2.2.

11.2.2 Small antennas

Short electric dipole antennas have a directivity of $D = 1.5 = 1.8\,\text{dBi}$, independent of the thickness and length of the wire as long as both are much shorter than half a wavelength, see Section 5.1.4. Similarly, small electric loop antennas (representing incrementally small magnetic currents) have 1.8 dBi directivity as well. It is very simple to theoretically construct a small antenna that has 4.8 dBi directivity even in free space without ground plane. This is the so-called *Huygens source* that consists of both a short electric and a short magnetic dipole, see Section 4.4.3. The incremental electric dipole excites the basic spherical TM mode, and the incremental magnetic dipole excites the basic TE mode. It is easy to show that the Huygens source represents the specific combination of the short electric and magnetic dipoles that maximizes the directivity. Such a Huygens source can in practice be constructed from wires by combining an electric dipole antenna and a loop antenna. Thus, the maximum available directivity of small antennas is 4.8 dBi. Many practical small antennas show directivities close to this, such as, e.g., the inverted F antennas. The half-wave dipole in free space has 2.2 dBi directivity, which is 2.6 dB below the 4.8 dBi limit for even smaller antennas.

11.2.3 Heuristic combination valid for any antenna size

The maximum available directivity of antennas that are neither large nor small in terms of wavelengths is not explicitly known. Therefore, we herein present a way of combining the two above-mentioned maximum directivities in a heuristic manner to get a formula that can be used for all antenna sizes. The following power addition formula seems reasonable:

$$\begin{aligned} D_{0_{\max}} &= \left[\frac{\pi D}{\lambda}\right]^2 + 3 \ , \\ \{D_{0_{\max}}\}_{\text{dBi}} &= 10\log\left[\left(\frac{\pi D}{\lambda}\right)^2 + 3\right] \ . \end{aligned} \tag{11.3}$$

The limitations of large antennas, small antennas and this heuristic combination of them are all plotted in Fig. 11.1. The directivities of some directive practical antennas that are between 0.2 and 4λ in diameter are also shown. These are the closely spaced folded Yagi antenna in [12] and the electrically small Yagi in [13], the helix antenna and two short backfire antennas found in [14][2], as well as two resonant reflector antennas with dimensions found in [15][3], and the so-called Eleven antenna in [16][4]. The antennas are designed to obtain as high directivity as possible using numerical methods, and the resulting directivities are very close to the new heuristic limit introduced by (11.3). The electrically small Yagi has even 0.38 dB higher directivity than (11.3), but the reference for it provide only simulated results, so there may be numerical uncertainties. Also, [13] gives lower experimental directivities of similar antennas. Furthermore, there is no theoretical foundation for our heuristic formula in the intermediate range $0.2\lambda < 2a < 2\lambda$, so it could happen that some practical antennas

[2] Backfire antennas make use of strong multiple reflections between one large and one small reflecting plate to increase directivity.

[3] They are making use of weaker multiple reflections between a paraboloidal reflector and its feed to increase the directivity.

[4] The eleven antenna consists in its simplest narrowband solution of two parallel dipoles over a small ground plane.

have slightly higher directivity. All the above examples given in [14]-[16] have been verified by measurements.

11.2.4 Small antennas on large ground planes

Short cavity-backed slot in an infinite ground plane radiates like a magnetic current, but only on one side of the ground plane, and the radiation pattern on the radiating side is not affected by the ground plane, so the directivity is the double of that of a short magnetic current in free space (3 dB higher), i.e., $D_0 = 2 \cdot 1.5 = 3 = 4.8$ dBi. Short vertical monopole on a ground plane has also 4.8 dBi directivity, of the same reason. In practice these directivities are high also for quite small ground planes.

Thus, it is possible to make small antennas on large ground planes that have higher directivities than 4.8 dBi. For example, the directivity of a small horizontal dipole above a large ground plane approaches 8.8 dBi, see, e.g., [4]. However, when we include the size of the ground plane in the definition of the smallest sphere, the directivities of such antennas will normally not come close to the limiting curve in (11.3). An exception is the eleven antenna [7]. This consists in its basic configuration of two parallel dipoles spaced by 0.5λ and located about 0.15λ over a ground plane. It has 11 dBi directivity when the square ground plane is wider than 0.7λ, which gives a minimum sphere radius of 1.03λ. This is only 0.3 dB below the limiting curve.

11.2.5 Planar array antennas

In practice the above formulas apply also to regular antenna arrays when the elements are combined to one port with a passive combination network, but only if the aperture area A of the array is defined by (Section 10.3.2)

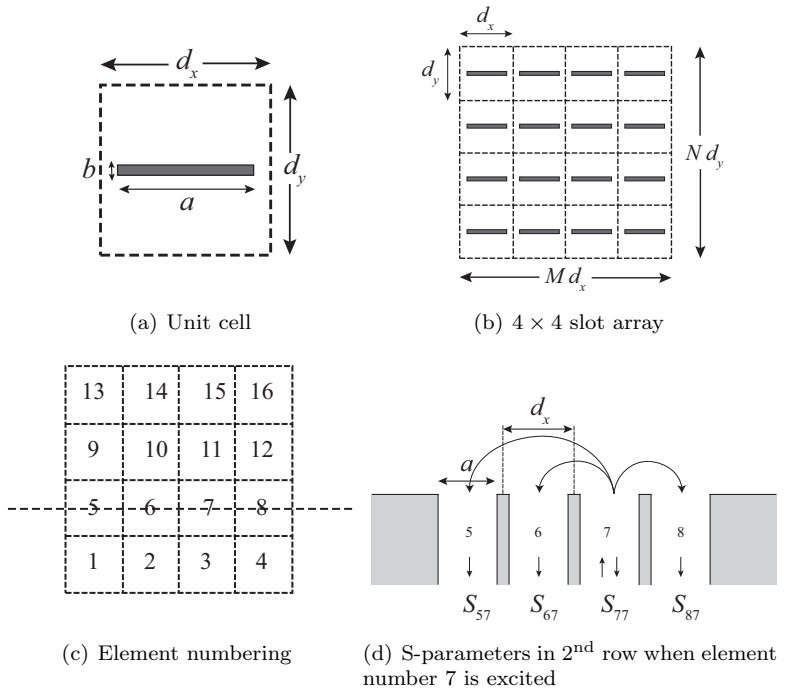
$$A = Md_xNd_y \ , \tag{11.4}$$

where M (N) is the number of elements in x-direction (y-direction), and d_x (d_y) is the element spacing in x-direction (y-direction), and the element spacings are small enough to avoid grating-lobes (normally $d_x < \lambda$ and $d_y < \lambda$ are sufficient for broadside radiating arrays). Very small array antennas that can be located within spheres or diameter smaller than 0.5λ may easily have directivities up to 4.8 dBi, in the same way as small single-port antennas can have this.

Eq. (11.4) is valid for a rectangular planar array, but we can readily define an equivalent diameter of a circular array with the same area, i.e.,

$$D = 2\sqrt{Md_xNd_y/\pi} \ . \tag{11.5}$$

This array is then governed by the directivity limit in (11.3). The definitions in (11.4) are illustrated in Fig. 11.2. The radiating elements are illustrated by the solid lines (dipoles or narrow slots). *Note that the definition of the area and equivalent diameter of the array include the whole unit cell of the edge elements, even if the radiating element itself does not fill the dashed unit cell. This is very important for the maximum available gain formula to be used for arrays.*



(a) Unit cell

(b) 4×4 slot array

(c) Element numbering

(d) S-parameters in 2nd row when element number 7 is excited

Figure 11.2: Example of planar array geometry. (a) Definition of unit cell of an element in a regular planar array. (b) Illustration of the definition of the aperture areas in (11.3) of planar arrays with rectangular aperture shapes. The radiating elements are illustrated by thick solid lines indicating narrow waveguide apertures, and the unit cells of the grid are illustrated as dashed boxes. (c) Illustration of the numbering of the elements used in (11.6) for a 4×4 array. (d) The S-parameters of elements 5, 6, 7 and 8 when element 7 is excited.

11.2.6 Superdirectivity

R. C. Hansen uses the following conservative definition of *superdirectivity* [10]:

"A useful operational definition of antenna superdirectivity (formerly called supergain) is directivity higher than that obtained with the same antenna configuration uniformly excited (constant amplitude and linear phase). Excessive array superdirectivity inflicts major problems in low radiation resistance (hence low efficiency), sensitive excitation and position tolerances, and narrow bandwidth. Superdirectivity applies in principle to arrays of isotropic elements although, of course, actual antenna arrays are composed of nonisotropic elements".

Thus, what is left after these restrictions is merely a concept of theoretical interest. Therefore, superdirective antennas have also only been reported theoretically, and the most "promising" publications are related to endfire linear arrays. However, publications on such theoretical superdirective linear arrays report actually much lower directivity than the maximum available directivity of an optimum more planar antenna structure radiating from within the same minimum surrounding sphere. This can be seen by comparing the limiting curve in Fig. 11.1 with the results in [10, Tables I and II]. There is one exception: According to [10, Table III]

a superdirective short endfire array of 0.5λ length can have 5.66 dBi directivity, if it is made of 9 elements with 0.0625λ element spacing. However, the reported $Q = 1.7 \cdot 10^{14}$ corresponds to a relative bandwidth of $1/Q = 0.6 \cdot 10^{-14}$, which is completely useless. Also, it seems that the *embedded element efficiency* has not been taken into account in the simulations.

Modern computers make it easier to produce simulated results with realistic vector current sources. It can easily be shown that two small dipoles in a closely spaced ($\ll \lambda$) array, fed 180° out-of-phase achieve a directivity of ≈ 5.8 dBi. If the phase of one element is adjusted slightly away from 180°, the array can operate in a superdirectivity mode and achieve a directivity of ≈ 7.2 dBi [17]. These directivities are above the limit of the heuristic formula in (11.3), and this array shows also superdirectivity according to Hansen's definition.

Arrays with such small element spacing will in practice suffer from large coupling losses causing a low *decoupling efficiency*, see Section 11.3. See also [18] where the decoupling efficiency of a single element, referred to as the embedded element efficiency, has been measured for a dense Vivaldi array. In principle, all dense arrays suffer from limitations due to mutual coupling [6]-[8], and this mutual coupling can only be correctly accounted for by properly including source impedances of the exciting sources on all excited ports and dummy loads at not excited ports. Unfortunately, it seems that the existing work on superdirectivity till now does not include source impedances, even though it readily can be included in the otherwise complete formulation in [9]. Such practice can give very erroneous results because without source impedances, it is possible to enforce any excitation on the elements of an array, whereas with source impedances the excitations will be affected by the excitations on neighboring ports so that they cannot be fully controlled. This can be characterized by the general decoupling efficiency in (11.9). In practice it may not at all be possible to excite arrays for superdirectivity with realistic source impedances.

11.3 Embedded element efficiency of arrays

11.3.1 Single-port antennas

The radiation efficiency of single-port small antennas can be large if they are made of low loss materials. However, the losses due to using lossy materials in small antennas may often appear larger than expected when compared to using the same materials in larger antennas. The reason is that small antennas are based on radiation at resonance, which makes the waves bounce back and forth inside the material many times before leaving the antenna.

11.3.2 Multi-port antennas

Small arrays for diversity and MIMO systems may suffer from strong mutual coupling between the antenna elements. They are different from classical arrays in the sense that the phase and amplitude excitations (both in transmitting and receiving mode) of the elements are dynamically adjusted to match the statistical field variations in the environment[5]. Therefore, they cannot be impedance matched for one specific excitation. Each single element are used more or less independently of each other, and therefore they have to be analyzed in their

[5] This is done digitally inside the signal processor, but it corresponds to changing the excitation.

embedded situation, when only one element is excited and the others are terminated with the port impedances. Therefore, there exists a degradation of the embedded element efficiency caused by absorption in the source impedances or loads connected to the other *element ports*. This efficiency was introduced for diversity and MIMO arrays in Section 3.3.3.

The *embedded element efficiencies* needed to characterize MIMO arrays represent a fundamental limitation caused by mutual coupling. The simplest expression for the embedded radiation efficiency of port j in an antenna array with MN elements is[6]

$$(e_{\text{emb}})_j = 1 - \sum_{i=1}^{MN} |S_{ij}|^2 \ , \tag{11.6}$$

where S_{ij} is the S-parameter between ports i and j, respectively. The numbering convention used in this equation is illustrated in Fig. 11.2. Such fundamentally limiting radiation efficiencies are also present in large classical arrays for producing narrow beams[7]. The embedded efficiency becomes very low only if the element spacing is very small, such as in dense arrays for multiple beams. Then, this limitation is known as Steins's limit with reference to [6][8] and is also thoroughly treated in [7].

In a classical array, the embedded element efficiency can be seen as the ratio between the realized gain per element of the array and the directivity of one embedded element [7]. For small element spacings, the embedded element pattern is known to have a $\cos(\theta)$ shape [7][9]. When the radiation intensity varies as $\cos(\theta)$ over half space, the power integral in (2.65) becomes

$$P = 2\pi \int_0^{\pi/2} \cos\theta \sin\theta d\theta = \pi \ .$$

Thus, the corresponding directivity of the embedded element is $D_{\text{emb}} = 4\pi/\pi = 4$, i.e., 6 dBi. The maximum available gain per element of a large array can be found, by using (11.1), giving

$$G_{\text{emb}} = \frac{D_{0_{\max}}}{MN} = 4\pi \frac{d_x}{\lambda} \frac{d_y}{\lambda} \ . \tag{11.7}$$

This means that the embedded element will have a maximum available radiation efficiency of

$$e_{\text{emb}} = \frac{G_{\text{emb}}}{D_{\text{emb}}} = \pi \frac{d_x}{\lambda} \frac{d_y}{\lambda} \tag{11.8}$$

in a dense array, under the assumption that the directivity always is 6 dBi or larger. This result originates from the discussions in Hannan's paper [7], see also [19]. It is an asymptote valid for both d_x and d_y being small. Therefore, we will refer to (11.8) as *Hannan's asymptote*, representing the maximum available embedded element efficiency for a dense array. The term embedded efficiency is justified as long as $e_{\text{emb}} < 1$, corresponding to both d_x and d_y being smaller than $1/\sqrt{\pi}\lambda = 0.56\lambda$, at the same time. The embedded element efficiency has also been measured for a singly-excited element in a dense focal-plane array [18, Fig. 11]. It becomes very small for small element spacings, and is already $e_{\text{emb}} = \pi/4 = 79\,\%$ for planar arrays with 0.5λ element spacing [7]. The embedded element efficiency has contributions

[6] Same as Eq. (3.8).
[7] See Section 11.4.
[8] This name was proposed in [8].
[9] This is due to the $\cos(\theta)$ projection of the aperture area.

both from reflections at the singly-excited port, and from lost power coupled into the ports of neighboring elements. The latter is normally the larger.

When many ports are excited, the embedded element efficiency becomes instead a general *decoupling efficiency*. For an array with MN antenna ports in total, of which a subarray of $(MN)_{\rm exc}$ ($\leq MN$) ports are used to generate the beam, the decoupling efficiency becomes

$$e_{\rm dc} = \frac{P_{\rm rad}}{P_{\rm inc}} = 1 - \frac{P_{\rm dc}}{P_{\rm inc}} = 1 - \left\{ \sum_{i=1}^{MN} |b_i|^2 / \sum_{j=1}^{(MN)_{\rm exc}} |a_j|^2 \right\} ,$$
$$b_i = \sum_{j=1}^{(MN)_{\rm exc}} S_{ij} a_j \ , \tag{11.9}$$

where $P_{\rm rad}$ is the radiated power, $P_{\rm dc}$ is the total power lost in the terminations on the ports[10], and $P_{\rm inc}$ is the sum of the incident powers on all $(MN)_{\rm exc}$ excited ports of the antenna. Further $|a_j|^2$ is the forward power incident on an excited port j, $|b_i|^2$ is the backward power leaving any port i, and S_{ij} is the usual scattering matrix element between the ports i and j when all the other ports are match-terminated. Hence, $MN - (MN)_{\rm exc}$ ports are unexcited. This efficiency plays a major role when characterizing focal plane arrays [18].

11.4 Gain limitations of regular antenna arrays

The embedded element efficiency plays via (11.8) a major gain-limiting role of dense regular array antennas, in the same way as the grating efficiency in (10.42) does for sparse arrays [20]. The following example illustrates this. With regular arrays we mean here an array with elements in a regular grid with element spacings d_x in x-direction and d_y in y-direction, as illustrated in Fig. 11.2.

The example is an array of 32×32 open-ended waveguides, each with aperture dimensions $a = 0.505\lambda$ and $b = 0.067\lambda$ along x- and y-directions, respectively. Thus, $M = 32$ and $N = 32$ in Fig. 11.2. The element spacing is fixed to $d_x = 0.67\lambda$ in x-direction (H-plane) and it varies from 0.1λ to 10λ in y-direction (E-plane). Thus, when the element spacing is 0.1λ, the slots are located very close to each other, and when the element spacing is 10λ they are located very far apart.

The total area of the array is given by (11.4), and the maximum available directivity of this area is given by (11.1). This is the straight solid diagonal line plotted in Fig. 11.3 marked "*maximum available*". This represents of course also the maximum available realized gain.

The realized gain of the whole array and its elements have been found by three different numerical approaches for each d_y, by using a commercial full wave code. The term full wave means that no approximations have been used in the simulation methods, except for the numerical discretization. The three approaches are:

[10] This includes the source impedances on the excited ports.

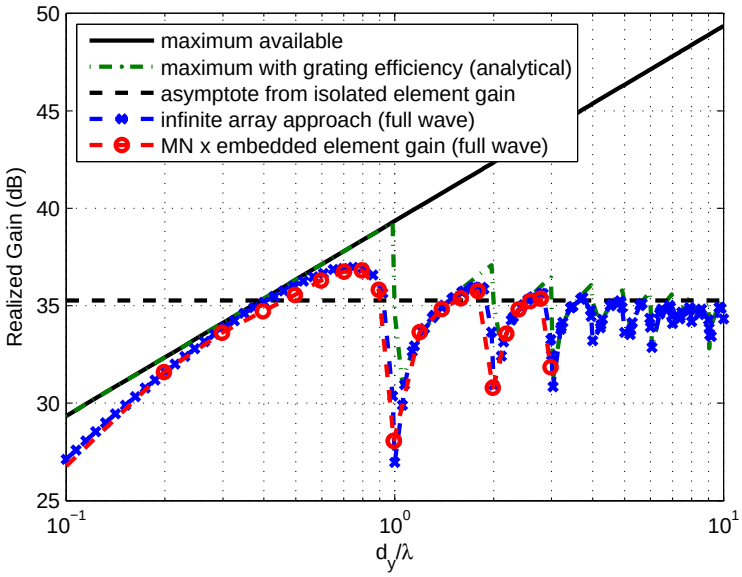


Figure 11.3: Realized gain of 32×32 element regular array of open-ended waveguides in infinite ground plane when the element spacing in H-plane is $d_x = 0.67\lambda$ for different element spacings in E-plane, evaluated by different methods.

a) *Infinite array approach* This means simulating a unit cell of the array with periodic boundary conditions. This corresponds to exciting all waveguide elements with the same amplitude and phase. The actual finiteness of the arrays due to the 32×32 elements is taken into account by a truncation of the infinite array, neglecting so-called edge effects. This means that elements close to the edge of the arrays are assumed to radiate the same way as any other element. The approach is known to be very accurate for large arrays like this, both regarding the input S-parameter (being equal for all elements) and the far-field function of the whole array. The realized gain is plotted as the curve marked "*infinite array approach*" in Fig. 11.3.

b) *Embedded element approach* An element in the center of the array is simulated when (in principle) all the other elements are present and terminated. Our elements are rectangular waveguides, so a termination means that there is an ideally matched load at the end of the waveguide. The results of the simulation are the far-field function, directivity, and realized gain G_{emb} of the embedded element. The realized gain of the total array is

$$G_{\text{arr}} = MNG_{\text{emb}} \ . \tag{11.10}$$

This is plotted as the curved marked with "*MN* $\times$ *embedded element gain*" in Fig. 11.3. The curve is seen to be almost identical to the "*infinite array approach*", which it should be. The discrepancies when d_y is between 0.3λ and 0.8λ are due to the numerical accuracy.

c) *Isolated element approach* This is one open-ended waveguide in an infinite ground plane. The results are its input S-parameter, far-field function, and realized gain G_{iso}. An approximate array gain can be obtained from

$$G_{\text{arr}} = MNG_{\text{iso}} \ . \tag{11.11}$$

The result is shown as the curve called "*asymptote from isolated element gain*" in Fig. 11.3. The directivity of the isolated element can in our case also be found analytically to be 5.2 dBi by using equations (5.82) and (5.85).

Fig. 11.3 shows that the computed realized gain with the infinite array method approaches the isolated element asymptote in a slow and periodic manner for large d_y. The slow convergence is due to all the grating-lobes appearing with periodic intervals when d_y is larger than 1λ[11]. They have a large effect in E-plane because the isolated element pattern of a slot is omnidirectional in E-plane. Each new grating-lobe causes a sudden reduction in gain. Therefore, the graph shows dips that appear with regular intervals when d_y increases, corresponding to the sudden appearance of grating-lobes along the array in E-plane.

The effect on the array gain due to the sudden appearance of grating-lobes can easily be modeled by multiplying the maximum available realized gain by the grating efficiency in (10.42), i.e.,

$$G_{\text{arr}} = \frac{4\pi MNd_xd_y}{\lambda^2} e_{\text{grt}} \ . \tag{11.12}$$

When evaluating this we assume that the far-field function of an isolated slot is uniform in E-plane (see (5.82) and (5.85)), and then the expression becomes completely analytical and can be evaluated very easily. The result is the curve marked "*maximum with grating efficiency*" in Fig. 11.3. We see that this is able to model the periodic variation of realized gain very well, except precisely at d_y/λ for which the grating-lobes suddenly appears, i.e., at each multiple of wavelengths. Therefore, the grating efficiency in (10.42) gives a good understanding of losses in directivity and realized gain due to grating-lobes.

Let us look more carefully at the realized gain for small element spacing d_y/λ, or rather the related embedded element efficiency. Fig. 11.4 shows embedded element efficiencies evaluated in different ways. They are evaluated by Hannan's asymptotic formula in (11.8), and by using the definition of the embedded element efficiency for a lossless multi-port antenna in (11.6). The latter is evaluated numerically "*from all S-parameters*", resulting from the simulations with the embedded element approach in *b)* above. Fig. 11.4 also shows the embedded element efficiency when we correct Hannan's asymptote by the mismatch factor of the fully-excited array, i.e., the mismatch factor when all elements are equally excited. The forth curve is marked "*gain per element minus 6 dB*". This is obtained by taking the result from the infinite array approach, dividing it with MN to get the realized gain per element, and removing the 6 dB directivity of a single embedded element in a dense array. The latter was found when deriving (11.8), and originated from Hannan.

The figure shows that Hannan's asymptote represents the highest efficiency for all d_y/λ. Thus, it is the fundamental limiting factor describing the fact that the embedded element will have low radiation efficiency for dense array. When we correct this with the mismatch factor we get exactly the same result as obtained from the full wave simulation of "*gain per element minus 6 dB*". This shows that it is very easy to correct for Hannan's asymptote to get the actual realized gain in dense regular arrays, but we need to know the S_{11} of the array elements for the all-excited case. Finally, we see that the actual embedded element efficiency evaluated from (11.6) always is lower than the three other results, and approaches them the smaller d_y/λ is. Eq. (11.6) can never be larger than unity (0 dB) by definition. However, the three other curves can be larger than unity when the assumptions $d_x \ll \lambda$ and $d_y \ll \lambda$ for which they are evaluated, are not satisfied. This assumption is implicit also in the full

[11] According to (10.78).

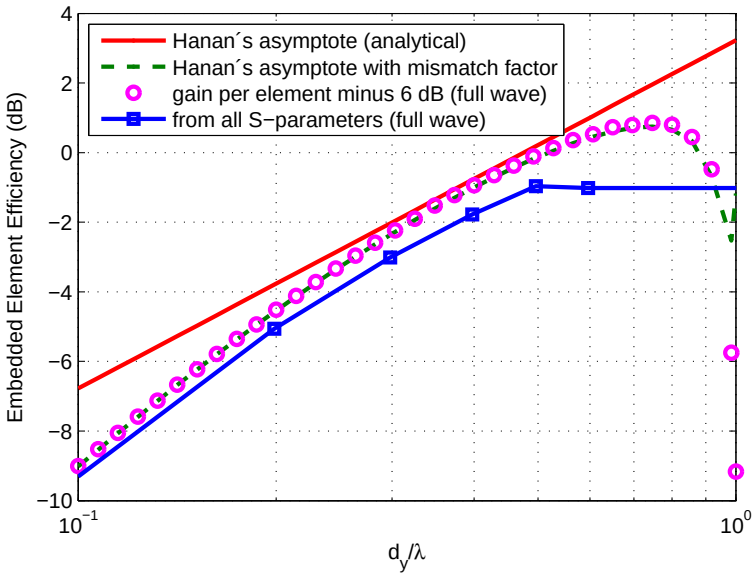


Figure 11.4: Embedded element efficiency of the same array as in Fig. 11.3, evaluated by different accurate and approximate methods.

wave efficiencies when assuming that the directivity of the embedded element is 6 dBi. If the directivity is larger, the computed value of the embedded element efficiency will be lower and satisfy the physical requirement of $e_{\text{emb}} \leq 1$. Thus, it is reasonable to believe that embedded elements in dense arrays always will have directivities equal to or larger than 6 dBi.

The most important conclusion from Fig. 11.4 is that the embedded element efficiency is a major factor contributing to the realized gains of dense array antennas. Unfortunately, this is not so well known, and, e.g., research on superdirectivity has not taken this into account, although the understanding of this fundamental limiting efficiency dates back to Hannan in 1964 [7].

11.5 Bandwidth limitations due to antenna size

The physical limitations of small antennas have always been of concern to antenna engineers, often expressed in terms of questions like "How large must an antenna be in order to radiate efficiently?" or "What is the available bandwidth of a small antenna of a certain size?" The mutual relation between size and bandwidth has been known since the early days of radio wave communications. At that time there were no computers and therefore all theoretical studies were based on mathematical analysis. Similarly, practical designs were entirely based on experiments without any computational support.

The above facts made it very difficult to predict frequency variation and thereby bandwidth before the antenna actually was built. However, it was possible to estimate bandwidth by calculating the so-called *Q-factor* from analytical expressions for both the dissipated power and stored time-averaged energy at the resonance frequency of the antenna [21]-[22]. Such

approach for estimating frequency dependences was known from prediction of quality of circuit resonators [23]-[24], which naturally is defined by

$$Q = f_0/(2\Delta f) \ , \tag{11.13}$$

where f_0 is the resonance frequency and $2\Delta f$ is the full 3 dB bandwidth. It was shown analytically that this could be determined from

$$Q = 2\pi f_0 \frac{W_{\text{ave}}}{P} \ , \tag{11.14}$$

where W_{ave} is the time averaged stored electric and magnetic energy at f_0, and P is the total dissipated power at f_0.

Half-wave dipoles and slots are resonant by nature, and smaller antennas become resonant at least after impedance matching by external components. Therefore, (11.14) was used even to predict Q and relative bandwidth $1/Q$ of small antennas, by using the radiated power as the dissipated power P in (11.14) because it appears as losses in the radiation resistor in the equivalent circuit of the antenna.

It is very difficult to get useful bandwidth out of small antennas, and therefore the practice of evaluating Q-factors lend itself to studying fundamental bandwidth limitations of small antennas. The lowest limiting radiation Q could be evaluated from Maxwell's equations without knowing the exact shapes of the antennas [21]-[22]. When studying fundamental bandwidth limitations of small antennas in terms of the intrinsic radiation Q, it is important to be aware of the fact that even if Q stands for quality, it is actually the inverse of Q and the radiation efficiency that represents the quality when dealing with antennas. Therefore, the Q term itself is misleading. Still, the treatment of fundamental limitations of small antennas in terms of their radiation Q is common practice also today [25]-[29].

The Q-limitation is normally expressed in terms of the radius of the smallest sphere that can enclose the antenna, but recently this has been extended to any limiting shape [29]. Different authors derive different formulas for Q, such as one Q for small antennas radiating circular polarization [30] and another Q for small antennas exciting both TE to r and TM to r spherical modes with arbitrary polarization [27], where r is the radial direction.

The present section aims explaining what this limiting intrinsic radiation Q means in terms of practical bandwidth limitations of three different types of small antennas: single-resonance-type, gradual-transition type, and cascaded-resonances-type. The most important quality measure of a small antenna, besides its relative bandwidth, is its total radiation efficiency, i.e., the total radiated power relative to the maximum power available at its terminal when the antenna is matched to its port impedance. This includes both the mismatch factor and the radiation efficiency due to dissipation losses. The mismatch factor can be improved by impedance matching techniques, whereas the dissipation normally is associated with lossy materials. Small antennas often have low efficiencies and narrow bandwidths [31]. This low efficiency makes the bandwidth larger than the intrinsic limit, in the same way as a loaded resonator has smaller Q than an unloaded one. In the case of losses the intrinsic radiation Q actually becomes a physical limitation on the bandwidth-efficiency product, in agreement with the theory of unloaded and loaded resonators [24]. In principle there are two basic approaches for making antennas radiate well:

a) *The resonance approach*: Small antennas like dipoles and slots have natural electromagnetic resonances when they are half a wavelength long. These resonance frequencies can

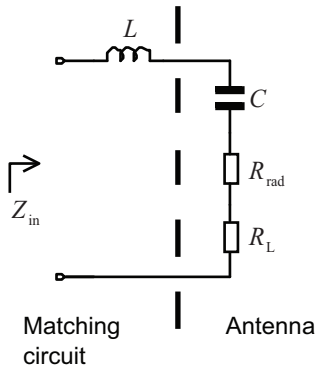


Figure 11.5: Equivalent circuit for a small electric current antenna such as a short dipole with both dissipation losses R_L and radiation losses $R_{\rm rad}$.

be reduced, e.g., by locating materials with high permittivity close to the antenna, but the bandwidth is reduced. Non-resonant small antennas will normally also be made resonant by impedance matching them with capacitors or inductors. Furthermore, many resonant antennas can be cascaded to get a wideband antenna.

b) *The gradual transition approach*: This is based on changing the geometry gradually from that of the feeding waveguide or transmission line to a geometry that makes the waves smoothly transit to free space, such as this is done in, e.g., flared horns (waveguide transition) and Vivaldi tapered slot antennas (slot line transition).

The limitations of resonance-type antennas can be determined by the inverse-Q limitation on the bandwidth-efficiency product, whereas it is not obvious that the limitations of wideband gradual-transition type antennas are determined by this. The gradual-transition-type antennas are limited by the intrinsic radiation Q expression, but that this determines a lowest frequency of operation rather than an efficiency-bandwidth limitation. This lowest frequency of operation is the same as the gradual cut-off frequency of spherical waves known from EM field theory [32], [33]. The size of the surrounding sphere at which this gradual cut-off appears has also been referred to as a radiansphere [34], [35].

Furthermore, it is well known that the bandwidth of resonance-type antennas can be improved by making use of several resonances coupled in cascade, as series or parallel resonances. We will show that ideally any bandwidth can be obtained in such cases, and that the intrinsic radiation Q instead becomes a measure of how many cascaded-resonances is needed to cover a certain frequency band.

The results in this chapter are valid for small antennas, but also for "not-so-small" antennas if only the basic spherical TE or TM to r modes are excited, or both of them.

11.5.1 Intrinsic radiation Q for TE, TM and TE&TM

The intrinsic *minimum radiation* Q of a lossless antenna exciting the basic TM to r spherical mode was first derived by Chu [21], and repeated later in a different way by Collin and

Rothschild [22], to be

$$Q_{\text{TM}} = Q_{\text{TE}} = \frac{1}{(ka)^3} + \frac{1}{ka} \ , \tag{11.15}$$

where a is the radius of the smallest sphere that can surround the antenna and $k = 2\pi/\lambda$ is the wavenumber. This limiting Q is valid for TM case under the assumption given by the equivalent circuit in Fig. 11.5, that the capacitive impedance of the small antenna (of electric current type) is tuned out by an external inductance in series with the capacitance. The Q for TE case is given by the same formula, and can be excited by a magnetic current such as a slot or small circular electric current ring. These have equivalent circuits in the form of parallel resonances, and they need to be tuned out by external capacitors. Later Harrington showed that the Q is a factor 2 lower for circular polarization if two orthogonally polarized TE and TM spherical modes are excited in quadrature [30]. He derived it using the same approximations as Chu did for linear polarization. Recently McLean recalculated the Q directly from the fields for both linear and circular polarization [25]. D.H. Kwon [27] showed that the TE and TM modes also can be combined to a directive linearly polarized pattern to give the same low Q as for Harrington's circularly polarized case. This combination of TE and TM sources actually corresponds to a so-called Huygens source, as given, e.g., in Section 4.4.3. Actually, any combination by which the basic TE and TM modes are excited equally strongly will give the same lower intrinsic radiation Q, i.e.,

$$Q_{\text{int}} = Q_{\text{TE\&TM}} = \frac{1}{2}\left(\frac{1}{(ka)^3} + \frac{2}{ka}\right) \ . \tag{11.16}$$

This can readily be seen by looking into the derivations in both [27] and [30]. The reason is that the TE and TM modes are orthogonal, and thereby the power integrals over each one of them can be added when both are excited.

11.5.2 Single-resonance-type small antennas

We first consider a *single-resonance-type antenna*. At resonance, the equivalent circuit for a small electric current element can be simplified as shown in Fig. 11.5, where we have modelled losses as a resistor $R_{_L}$ in series with the radiation resistance R_{rad}. The antenna is matched at frequency $f_{_0}$ by a lossless inductor with an inductance of $L = 1/\omega_0^2 C$, where $\omega_{_0} = 2\pi f_{_0}$. The total input impedance including the matching inductor becomes

$$Z_{\text{in}} = R_{\text{rad}} + R_{_L} + \frac{1}{j\omega C} + j\frac{\omega}{\omega_0^2 C} \ . \tag{11.17}$$

The equivalent circuit is the same as that of a resonator, so it becomes natural to express Z_{in} in terms of Q and the resonance frequency $f_{_0} = 1/(2\pi\sqrt{LC})$, in the same way as for resonators [23], giving

$$Z_{\text{in}} = (R_{\text{rad}} + R_{_L})(1 + jQ_{_0}\frac{(f - f_{_0})}{f_{_0}}\frac{(f + f_{_0})}{f}) \ , \tag{11.18}$$

where $Q_{_0}$ is the Q at the resonance frequency

$$Q_{_0} = \frac{1}{\omega_{_0}(R_{\text{rad}} + R_{_L})C} \ .$$

We see that there are two contributions to this Q; from radiation losses in $R_{\rm rad}$ and from dissipation in $R_{_L}$. Therefore, it is convenient to separate $Q_{_0}$ into the intrinsic radiation Q of the antenna, denoted $Q_{\rm int}$ and the Q contribution due to dissipative losses, denoted $Q_{_L}$, according to

$$\frac{1}{Q_{_0}} = \frac{1}{Q_{\rm int}} + \frac{1}{Q_{_L}} \; ; \qquad Q_{\rm int} = \frac{1}{\omega_{_0} R_{\rm rad} C} \; , \qquad Q_{_L} = \frac{1}{\omega_{_0} R_{_L} C} \; . \tag{11.19}$$

There are no other approximations in (11.18) and (11.19) other than those represented by assuming $R_{\rm rad}$, $R_{_L}$, L and C constant in Fig. 11.5.

Let us now assume that the antenna is ideally matched at the resonance frequency to a transmission line with characteristic impedance $Z_{_0} = R_{\rm rad} + R_{_L}$. Then, the input reflection coefficient becomes

$$\begin{aligned} r &= \frac{Z_{\rm in} - Z_{_0}}{Z_{\rm in} + Z_{_0}} = \frac{-jQ_{_0}\frac{(f-f_{_0})}{f_{_0}}\frac{(f+f_{_0})}{f}}{2 + jQ_{_0}\frac{(f-f_{_0})}{f_{_0}}\frac{f+f_{_0}}{f}} \; , \\ |r|^2 &\approx \frac{|Q_{_0}\frac{(f-f_{_0})}{f_{_0}}|^2}{1 + |Q_{_0}\frac{(f-f_{_0})}{f_{_0}}|^2} \qquad \text{for } (f - f_{_0}) \ll f_{_0} \; . \end{aligned} \tag{11.20}$$

The total radiation efficiency can be expressed as

$$\begin{aligned} e_{tot} &= e_r e_{\rm rad} \; , \\ e_r = 1 - |r|^2 &\approx \frac{1}{1 + |Q_{_0}\frac{(f-f_{_0})}{f_{_0}}|^2} \qquad \text{for } (f - f_{_0}) \ll f_{_0} \quad , \\ e_{\rm rad} &= \frac{R_{\rm rad}}{R_{\rm rad} + R_{_L}} = \frac{Q_{_0}}{Q_{\rm int}}, \end{aligned} \tag{11.21}$$

where $e_{\rm rad}$ is the radiation efficiency due to dissipation in $R_{_L}$ and e_r is the mismatch factor.

We see from (11.20) and (11.21) that Q determines both the ohmic loss efficiency and the bandwidth of the mismatch factor. It would be convenient to define the bandwidth in terms of a specific value of e_r. We see that when

$$\frac{|f - f_{_0}|}{f_{_0}} = \frac{1}{Q_{_0}} \; , \tag{11.22}$$

the input power reflection coefficient is 0.5 and the mismatch factor $-3\,$dB. This means that the full bandwidth between the $-3\,$dB values becomes

$$\frac{2\Delta f}{f_{_0}} = \frac{2}{Q_{_0}} \; . \tag{11.23}$$

This may look like it does not agree with (11.13), but it does, because the resonator now has an external load equal to the internal load, so the Q degrades by a factor 2. If we introduce the radiation efficiency $e_{\rm rad}$ in (11.21) we get

$$\frac{2\Delta f}{f_{_0}} e_{\rm rad} = \frac{2}{Q_{\rm int}} \; . \tag{11.24}$$

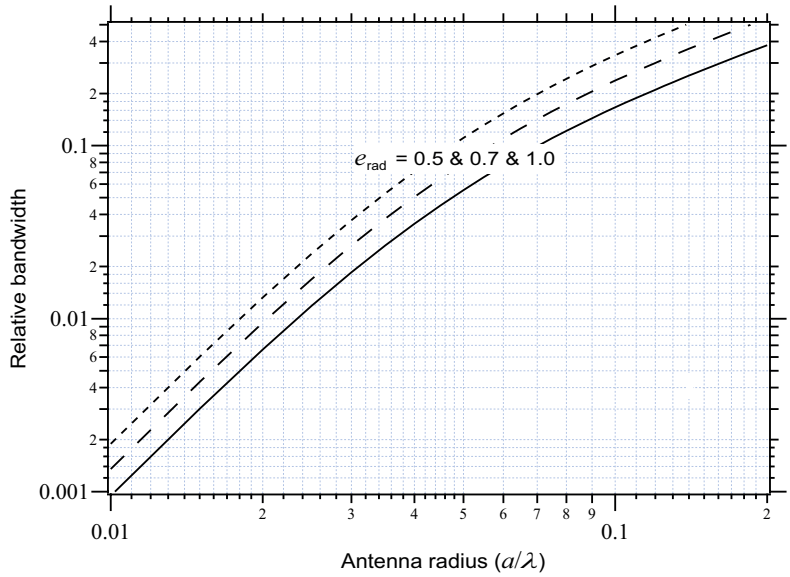


Figure 11.6: The theoretical maximum available relative bandwidth between -3 dB points of a small antenna versus radius of smallest surrounding sphere a for three different radiation efficiencies.

This means that the inverse intrinsic radiation Q, i.e., $1/Q_{\rm int}$, in (11.16) represents a fundamental limitation on the bandwidth-efficiency product of a small single-resonance-type antenna. Therefore, if we reduce the efficiency by introducing losses, the bandwidth will increase in such a way that the bandwidth-efficiency product is constant. *This means that we need to know both the bandwidth and efficiency of a small antenna in order to determine how close it is to the intrinsic size limitation.*

The results are plotted in Fig. 11.6 for three values of the radiation efficiency $e_{\rm rad}$. We see that antennas with low radiation efficiency have the potential of larger bandwidth, and visa versa. The simple approximate limitation formula in (11.24) is valid up to a radius of $a = 0.2\lambda$.

11.5.3 Wideband gradual-transition-type small antennas

Let us now consider the equivalent circuit in Fig. 11.7, which is the same as that in Fig. 11.5 but without any matching circuit. We will assume that this equivalent circuit is valid over such a wide and high frequency band that no matching network is required. In other words, we have been able to impedance match the antenna by gradually adjusting the antenna geometry so that the basic spherical TE or TM mode is excited, or both. Still, in this way we cannot remove the low frequency mismatch represented by the capacitance C, due to the fundamental or intrinsic limitations. Then, if we in the same way as in Section 11.3 use a characteristic impedance of the feed line of $Z_0 = R_{\rm rad} + R_L$, we get the following input

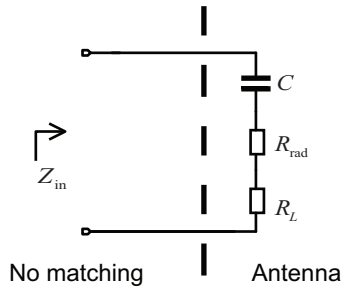


Figure 11.7: Equivalent circuit for a small electric current antenna with a lower gradual radiation cut-off.

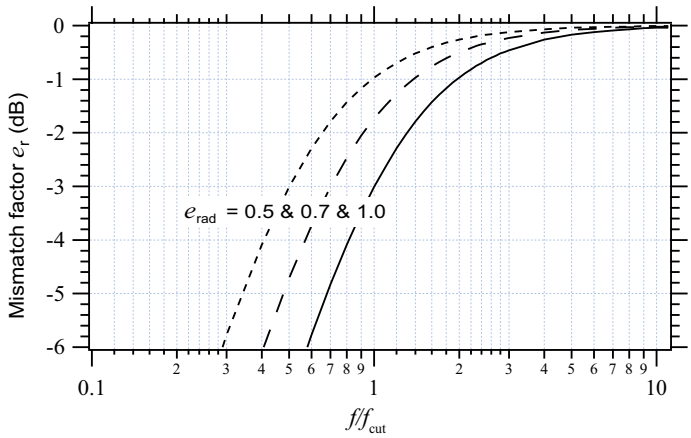


Figure 11.8: Illustration of mismatch factor with gradual radiation cut-off for the antenna in Fig. 11.7. The minimum available gradual radiation cut-off frequency $f_{\rm cut}$ for the lossless curve is determined by (11.26) and related to the intrinsic radiation Q in (11.15).

reflection coefficient and mismatch factor from (11.20) and (11.21)

$$r = \frac{Z_{\rm in} - Z_0}{Z_{\rm in} + Z_0} = \frac{\frac{1}{j\omega C}}{2(R_{\rm rad} + R_L) + \frac{1}{j\omega C}} \ , \qquad |r|^2 = \frac{|\frac{f_0}{2f} Q_0|^2}{1 + |\frac{f_0}{2f} Q_0|^2} \ ,$$
$$e_r = 1 - |r|^2 = \frac{1}{1 + |\frac{f_0}{2f} Q_0|^2} \ . \tag{11.25}$$

The radiation efficiency $e_{\rm rad}$ due to dissipation in R_L is the same as before. Note that we have kept the Q in the expressions, even though we here do not deal with a resonance. This is convenient, because we are dealing with fundamental limitations that are expressed in terms of Q. Also, the intrinsic Qs in (11.15) and (11.16) have the same $1/f$ variation for large frequency as we get from the $Q_0 f_0/f$ variation in (11.25), so the equivalent circuit should be quite good as a fundamental limitation at least for the lossless case.

We see that the frequency f_0 at which $Q_0 = 2$ in (11.25) plays the role of a *gradual radiation cut-off* frequency, at which the most of the available power radiates. To emphasize this

approach, let us introduce an intrinsic radiation cut-off frequency by setting $Q = 2$ in (11.15). This gives

$$ka = 1 \rightarrow \lambda_{\text{cut}} = 2\pi a \ , \qquad f_{\text{cut}} = c/\lambda_{\text{cut}} \ , \tag{11.26}$$

where c is the velocity of light. Using e_{rad} in (11.20) and (11.26), we can express the mismatch factor in (11.25) as

$$e_r = \frac{1}{1 + |e_{\text{rad}} f_{\text{cut}}/f|^2} \ . \tag{11.27}$$

e_r in dB is plotted as a function of relative frequency f/f_{cut} in Fig. 11.8. We see that in the lossless case the mismatch factor is $-3\,\text{dB}$ at the frequency of radiation cut-off. We see also that the apparent gradual cut-off change to lower frequency (i.e., improve) when the losses increase (mismatch factor decrease). The concept of gradual cut-off is known from electromagnetic field theory for spherical modes [33] and is also mentioned in [30]. The gradual cut-off from field equations for spherical TE and TM modes are studied in [33, Sec. 6-4]. Using this together with [33, Eq. (1-68)-(1-70)] we see that we can at a given radius define a gradual cut-off frequency at which the real and imaginary parts of the wave impedance are equal, corresponding to $Q = 1$. This appears from the approximate figure in [33, Fig. 6-6] approximately when $ka = 1$ which is in reasonable agreement with our equations. The discrepancy may be due to the fact that the Q formula in (11.14) is approximate. Smaller discrepancy could be obtained by calculating more accurate Q values using the spherical mode formulas in [33]. Thus, the gradual cut-off of the mismatch factor of gradual transition-type antennas is related to the gradual cut-off of spherical modes.

Practical examples of wideband gradual-transition-type antennas are Vivaldi antennas [42], self-grounded bowtie antennas [43]-[44], dual- and quad-ridge horn antennas [45], and spiral antennas [46].

11.5.4 Cascaded-resonances-type small antennas

In Section 11.3 it was shown that the bandwidth of a small antenna can be increased by introducing losses. This is in most applications not acceptable. A more acceptable way is to make use of multiple resonances that are spaced in frequency in such a way that a continuous frequency band is obtained. A typical example is the log-periodic dipole array [36]. A special log-periodic solution in which the radiation appears orthogonal to the log period is the "eleven antenna", consisting of two half wavelength spaced log-periodic folded dipole arrays located over a ground plane [37]. In this configuration the bandwidth is increased upwards by adding more resonant folded dipole pairs without affecting the antenna volume. The highest frequency limit depends mainly on manufacturing issues as the dipoles and their feed structure becomes very small. In principle, this technique for increasing bandwidth would also be desirable in smaller antennas. Therefore, we look into the limitations that apply to the cascading of several resonances logarithmically displaced in frequency.

We choose to simplify the study to series RLC resonance circuits connected in parallel[12]. The resonance frequencies are scaled by a factor k in a log periodic manner as shown in Fig. 11.9. The Q is equal for all resonators and the resistance is fixed to $50\,\Omega$, which also is

[12] Note we will neglect all kinds of mutual couplings between circuits, but the approach is still useful for understanding how cascaded-resonances-type antennas behave, which we can judge well after having experience with the practical design of several eleven antennas [37].

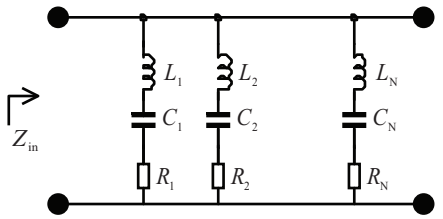


Figure 11.9: Illustration of cascaded parallel RLC resonance circuits log-periodically scaled in frequency to increase the bandwidth over which power is absorbed in the resistors and not reflected at the input. The resonance frequencies are log-periodically scaled with a factor k, and the Q is the same for all resonators.

the system impedance. If the scaling factor is very large the antenna will obviously behave like a multiband antenna with the same number of bands as resonators. However, the scaling factor can be decreased to make a large continuous band as shown in Fig. 11.10. The optimum scaling factor depends on the requirement on the mismatch factor. The spacing between the resonances must be smaller for stricter impedance matching requirements.

The numerical study showed that the mismatch near the edges of the log-periodic array is different from centrally located resonators. Therefore, the increase of bandwidth is smaller than what would be expected, see Fig. 11.10. Only the bandwidth of the largest centrally located band is considered here. A summary of the bandwidths for different numbers of resonators and different scaling factors is found in [1]. The resistance in the resonators needs to be changed when the scaling factor is small. This can be understood by considering two extreme cases. If the scaling factor is large, the bands of different resonators will not affect each other so that the resistance in each resonator should be $50\,\Omega$ for best matching. On the other hand, if the scaling factor equals 1, the resonance frequencies will be the same for all resonators. The resistors will then be connected in parallel, and the optimum value will be $50\,\Omega$ times the number of resonators. For choices of scaling factors between these two extremes, the resistance will be affected mostly by the closest resonators and will need to be increased a bit to enable matching to the $50\,\Omega$ feed line. Such a detailed study goes beyond the scope of this section.

By assuming that each resonator has a relative bandwidth which is inversely proportional to its Q, and by equating the upper and lower frequency limits for two adjacent resonators, the scaling factor can be obtained as:

$$k = \frac{2Q + \xi}{2Q - \xi} \approx 1 + \frac{\xi}{Q} \,, \tag{11.28}$$

where ξ is a constant which specifies the required overlap between the bandwidths of each resonator. It depends on the chosen requirement for the return loss and the smoothness of the mismatch factor curve. The simulations in [1] show that $\xi \approx 3.3$ when a $-3\,\mathrm{dB}$ input reflection coefficient is required, $\xi \approx 2.1$ for $-6\,\mathrm{dB}$ and $\xi \approx 1.6$ for $-10\,\mathrm{dB}$. This means that a good initial value for k can be found from the Q of the resonators, which can then be subject to further optimization. If the scaling factor is increased too much, the bandwidth will split in several subbands and thereby cause degradation in the continuous bandwidth. In practice, the scaling factor should be chosen slightly smaller than the initial value found from (11.28).

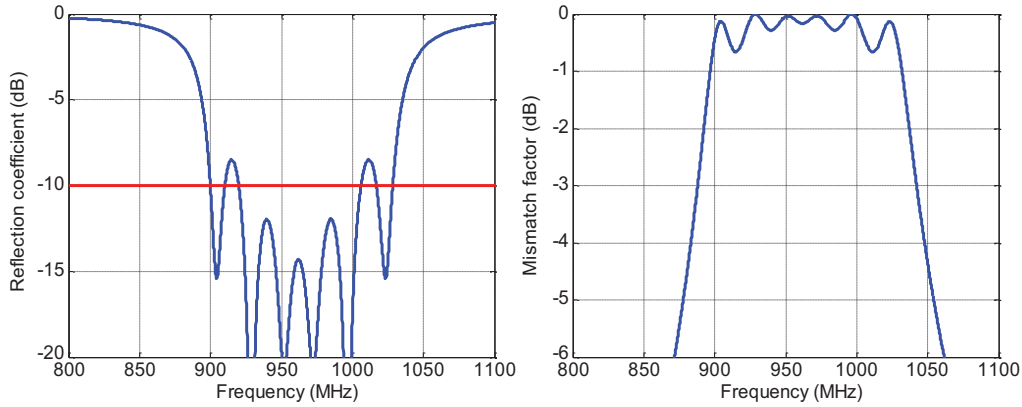


Figure 11.10: Input reflection coefficient and mismatch factor of the cascaded log-periodic resonator array in Fig. 11.9 for six resonators when $k = 1.027$ and $Q = 60$.

The concept of intrinsic Q has also been found to work quite well for estimating limitations on cascaded-resonances-type antennas, when the resonances are spaced far away from each other [38]. By definition Q is only valid for each single resonance. However, Fig. 11.11 shows that the Q of a resonator (and thereby also the intrinsic Q of a single-resonance-type antenna) determines the number of resonances needed to continuously cover a certain frequency band when designing cascaded-resonances-type antennas. Therefore, even if the intrinsic Q is large, corresponding to very small antennas, the overall bandwidth can be made large by such cascading of log-periodically scaled resonant small antennas. $Q = 30$ corresponds to a diameter of the smallest surrounding sphere of 0.1 wavelengths, and $Q = 45$ corresponds to 0.09 wavelengths.

The intrinsic Q will also determine the slope of the return loss and the mismatch factor at the lower and upper ends of the operational frequency band and the ripples within the band, as seen in Fig. 11.10. A larger Q corresponds to larger slope and faster ripples.

11.5.5 Summary

The intrinsic radiation Q is a fundamental limitation on the bandwidth-efficiency product of a single-resonance-type small antenna, more specifically on the achievable relative $-3\,\mathrm{dB}$ bandwidth of the mismatch factor times the radiation efficiency. $Q = 2$ corresponds to the lowest available gradual cut-off frequency, $f_{\mathrm{cut}} = c/(2\pi a)$, of a gradual-transition-type antenna, and this is the same as the value of the gradual cut-off of basic TE and TM spherical modes at which the stored energy is equal to the radiated energy. The intrinsic radiation Q determines the number of resonances needed to cover a certain frequency band with a cascaded-resonances-type small antenna. All these three cases are summarized in Fig. 11.11. The antenna size is constant in the figure, given in terms of the f_{cut}. Thus, the figure illustrates the difference in maximum available bandwidth for single- and multiple-resonance type and gradual-transition type antennas with the same physical size, when they operate at different frequencies. Notice that the mismatch factor is $-3\,\mathrm{dB}$ for the gradual-cut off case at $f = f_{\mathrm{cut}}$.

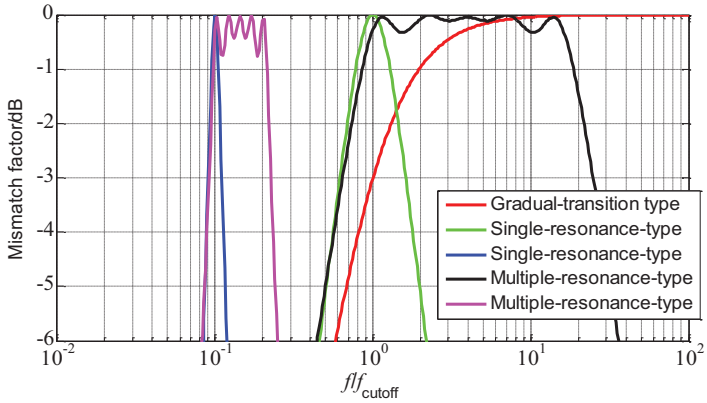


Figure 11.11: Mismatch factor of different hypothetical antennas of equal physical size in terms of the radius a of the smallest surrounding sphere, representing the maximum available bandwidth. The antennas are: A single resonant antenna resonant at $f/f_{\text{cut}} = 1$, i.e., $Q = 2$. A single resonant antenna resonant at $f/f_0 = 0.1$, i.e., $Q = 10$. A gradual transition antenna with the lowest possible cut-off, $Q = 2$. A cascaded resonance-type antenna with five resonances, $Q = 2$ for each resonance and $k = 2$. And, a cascaded resonance-type antenna with five resonances, $Q = 10$, $k = 1.2$.

Note that the Q is determined by (11.13) if we know the bandwidth between the $-3\,\text{dB}$ points of the mismatch factor, and the relation to the size and radiation efficiency is given by (11.16) and (11.24). The mismatch factor varies quadratically with frequency around the resonance f_0. Therefore, if we want the bandwidth B corresponding to Q at another dB level $(e_B)_{\text{dB}}$ of the mismatch factor e_B, we can use the relation

$$\frac{\Delta f}{B} = \sqrt{\frac{-3\,\text{dB}}{(e_B)_{\text{dB}}}}\,.$$

11.6 Complementary comments by S. Maci

Some authors define the quality factor by (11.14) rather than by (11.13), but its physical background is better illustrated by (11.14). There is also an alternative definition using either $Q = 2\omega_0 W_e/P$ or $Q = 2\omega_0 W_m/P$ for $W_e > W_m$ or $W_m > W_e$, respectively, where W_e (W_m) is the average stored electric (magnetic) energy and P is the total dissipated power of the resonator at the tuning frequency ω_0. When dealing with antennas, the dissipated power is assumed to be the sum of ohmic losses in the antenna and radiated power. If this is replaced by the radiated power only, we get the minimum bound on Q, i.e., the intrinsic Q. Indeed, since $W_e = W_m$, the total energy is $2W_e = 2W_m$ at a tuning frequency ω_0.

The Chu limit in (11.15) can be found by using two assumptions. First, by assuming that the field outside the minimum sphere is equal to the dominant spherical TM_{01} (TE_{01}) mode, i.e., the mode excited by an elementary electric (magnetic) dipole. Elementary dipoles are classical names of what herein is called incremental dipoles. This indeed implies the minimum reactive energy around the antenna. A further reduction of the intrinsic Q by a factor 2 (see (11.12)) is obtained by exciting TM_{01} and TE_{02} modes simultaneously. This becomes a

Huygens source, but only for a specific phase and amplitude balance between the two sources. Second, by using that the energy inside the minimum sphere is zero. This assumption gives the minimum possible Q. Therefore, in order to arrive as close as possible to the Chu limit one can design an antenna which radiates almost zero field inside the minimum sphere, for instance by synthesizing equivalent currents on a small spherical surface [39].

The maximum directivity of an antenna is in [11] derived by a spherical wave expansion, giving rise to the simple formula

$$D_{0_{\max}} \approx N^2 + N \ ,$$

where N is the maximum radial index of significant spherical harmonics radiated by the antenna. For any source, this number is related to the minimum sphere. In fact, it can be seen that the n-indexed Hankel function (which is the radial constituent of the spherical harmonics) starts to decrease rapidly when its index n surpasses the argument kr. Therefore, the n-indexed harmonics radiated by any source of radius a with $n > ka$, contributes only to the reactive energy very close to the minimum sphere and they are not significant after moving just a fraction of wavelengths away from the surface of the minimum sphere. This gives

$$D_{0_{\max}} \approx (ka)^2 + ka \ ,$$

The above expression is accurate for $ka > 3$. Note that the dominant term of this expression when the radius a is large, coincides with the one in (11.2).

The minimum sphere of a source is also related to the concept of degrees of freedom (DoF) of the field. This number can be defined as the sufficient and non-redundant number of samples of an electromagnetic field on a surface S around the sources, from which one can fully reconstruct the field within a predetermined error margin. A general way to determine the DoF is based on the evaluation of the significant singular values of the radiation operator which maps the currents on a source boundary onto the field radiated on the observation domain. Since the singular values decrease exponentially after a certain critical value is reached, the number of DoF is practically independent of the required precision [40]-[41]. If the observer is at a certain small distance from the sphere, the DoF is equal to the number of significant TE/TM spherical harmonics radiated by any source inside a minimum sphere. By analogy with the Shannon theorem, the number of DoF can be seen as two times the surface of the minimum sphere in terms of half-wavelengths square; namely $N_{\text{DoF}} = 2\Sigma/(\lambda/2)^2$, where Σ is the surface of the minimum sphere, and the factor 2 arises from the need of representing the fields by two orthogonal polarizations. Comparing the last formula with the one of maximum directivity, it is also seen that the maximum directivity is approximately half the number of DoF of the field in the full space.

11.7 References

[1] P.-S. Kildal and D. Nyberg, "Fundamental directivity and efficiency limitations of small antennas with one or more ports", *The Second European Conference on Antennas and Propagation (EuCAP 2007)*, Edinburgh, 11-16 November 2007.

[2] P.-S. Kildal and S.R. Best, "Further investigations of fundamental directivity limitations of small antennas with and without ground planes",*IEEE International Symposium on Antennas and Propagation (IEEE AP-S)*, San Diego, July 2008.

[3] L.J. Chu, "Physical limitations of omni-directional antennas", *Journal of Applied Physics*, Vol. 199, pp. 1163-1175, 1948.

[4] R.E. Collin, S. Rothschild. "Evaluation of antenna Q",*IEEE Transactions on Antennas and Propagation*, Vol. 12, No. 1, pp. 23-27, 1964.

[5] D. Nyberg, P.-S. Kildal, J. Carlsson, "Effects of intrinsic radiation Q on mismatch factor of three types of small antennas: single-resonance, gradual-transition and cascaded-resonance types", *IET Microwaves, Antennas and Propagation*, Vol. 4, No. 1, pp. 83-90, January 2010.

[6] S. Stein, "On Cross Coupling in Multiple-Beam Antennas", *IRE Transactions on Antennas and Propagation*, Vol. AP-10, No. 5, pp. 548-557, September 1962.

[7] P.W. Hannan, "The element-gain paradox for a phased-array antenna", *IEEE Transactions on Antennas and Propagation*, Vol. AP-12, No. 7, pp. 423-433, July 1964.

[8] J.F. Johansson, "Stein's limit for finite and infinite multi-beam arrays", *IEEE AP-S International Symposium*, Digest, Vol. 2, pp. 713-716, 26-30 June 1989.

[9] R. F. Harrington, "Antenna excitation for maximum gain", *IEEE Trans. Antennas Propagat.*, Vol. AP-13, No. 6, pp. 896-903, November 1965.

[10] R.C. Hansen, "Fundamental limitations in antennas",*Proceedings of the IEEE*, Vol. 69, No. 2, pp. 170-182, February 1981.

[11] R.F. Harrington, "On gain and beamwidth of directional antennas", *IRE Transactions on Antennas and Propagation,* pp. 219-225, July 1958.

[12] S. Lim, H. Ling, "Design of a Closely Spaced, Folded Yagi Antenna", *IEEE Antennas and Wireless Propagation Letters*, Vol. 5, pp. 302-305, 2006.

[13] S. Lim, H. Ling, "Design of electrically small Yagi", *Electronics Letters*, Vol. 43, No. 5, 1st March 2007.

[14] K. Fujimoto, J.R. James, *Mobile Antenna Systems Handbook 2nd Ed.*, Artech House, 2001.

[15] P.-S. Kildal, "A small dipole-fed resonant reflector antenna with high efficiency, low cross polarizaton and low sidelobes", *IEEE Transactions on Antennas and Propagation*, Vol. AP-33, No. 12, pp. 1386-1391, December 1985.

[16] R. Olsson, P.-S. Kildal, S. Weinreb, "The Eleven antenna: a compact low-profile decade bandwidth dual polarized feed for reflector antennas", *IEEE Transactions on Antennas and Propagation*, Vol. 54, No. 2, Pt. 1, pp. 368-375, February 2006.

[17] T.H. O'Donnell, A.D. Yaghjian, E.E. Altshuler, "Frequency Optimization of Parasitic Superdirective Two Element Arrays", *IEEE APS International Symposium*, Honolulu, Hawaii, July 2007.

[18] M.V. Ivashina, M. Ng Mou Kehn, P.-S. Kildal and R. Maaskant, "Decoupling efficiency of a wideband Vivaldi focal plane array feeding a reflector antenna", *IEEE Transactions on Antennas and Propagation*, Vol. 57, No. 2, pp. 373-382, February 2009.

[19] W.K. Kahn, "Element efficiency: A unifying concept for array antennas," *IEEE Antennas and Propagation Magazine*, Vol. 49, No. 4, pp. 48-56, August 2007.

[20] P.-S. Kildal, A. Vosoogh, S. Maci, "Fundamental directivity and efficiency limitations of dense array antennas explaining the element-gain paradox", submitted to IEEE Antennas and Propagation, February 2015.

[21] L.J. Chu. "Physical limitations of omni-directional antennas", *Journal of Applied Physics*, Vol. 19, pp. 1163-1175, 1948.

[22] R.E. Collin, S. Rothschild. "Evaluation of antenna Q",*IEEE Transactions on Antennas and Propagation,* Vol. 12, No. 1, pp. 23-27, 1964.

[23] J.E. Aitken, "Swept-frequency microwave Q factor measurement",*Proceedings of the IEE*, Vol. 123, pp. 855-862, 1976.

[24] L. Malter and G.R. Brewer, "Microwave Q measurements in the presence of series losses", *Journal of Applied Physics*, Vol. 20, pp. 918-925, October 1949.

[25] J.S. McLean, "A re-examination of the fundamental limits on the radiation Q of electrically small antennas", *IEEE Transactions on Antennas and Propagation*, Vol. AP-44, pp. 672-676, 1996.

[26] W. Geyi, "Physical limitations of antenna", *IEEE Transactions on Antennas and Propagation,* Vol. 51, No. 8, pp. 2116-2123, 2003.

[27] D.-H. Kwon, "On the radiation Q and the gain of crosses electric and magnetic dipole moments", *IEEE Transactions on Antennas and Propagation*, Vol. 53, No. 5, pp. 1681-1687, 2005.

[28] A.D. Yaghjian, S.R. Best, "Impedance, bandwidth, and Q of antennas", *IEEE Transactions on Antennas and Propagation*, Vol. 53, Nr. 4, pp. 1298-1324, 2005.

[29] M. Gustafsson, C. Sohl, and G. Kristensson. "Physical limitations on antennas of arbitrary shape", *Proceedings of the Royal Society A*, Vol. 463, pp. 2589-2607, 2007.

[30] R.F. Harrington. "Effect on antenna size on gain, bandwidth, and efficiency", *Journal of Research of the National Bureau of Standards- D: Radio Propagation*, Vol. 64D, No. 4, pp. 1-12, 1960.

[31] H.A. Wheeler, "Fundamental limitations of small antennas",*Proceedings of the IRE*, pp. 1479-1488, December 1947.

[32] R.F. Harrington, "On gain and beamwidth of directional antennas", *IRE Transactions on Antennas and Propagation*, pp. 219-225, July 1958.

[33] R.F. Harrington. *Time-Harmonic Electromagnetic Fields*, Wiley 1961.

[34] H.A. Wheeler. "Small antennas", *IEEE Transactions on Antennas and Propagation*, Vol. AP-23, pp. 462-469, 1975.

[35] H.A. Wheeler. "The radiansphere around a small antenna",*Proceedings of the IRE*, Vol. 47, pp. 1325-1331, 1959.

[36] D.E. Isbell, "Log Periodic Dipole Arrays,", *IRE Transactions on Antennas Propagation*, Vol. AP-7, pp. 329-334, October 1959.

[37] J. Yang, M. Pantaleev, P.-S. Kildal, B. Klein, Y. Karandikar, L. Helldner, N. Wadefalk, C. Beaudoin, "Cryogenic 2-13 GHz Eleven feed for reflector antennas in future wideband radio telescopes", *IEEE Transactions on Antennas and Propagation*, Vol. 59, No. 6, pp. 1918-1934, June 2011.

[38] S.R. Best. "The inverse relationship between quality factor and bandwidth in multiple resonant antennas", *IEEE Antennas and Propagation Society International Symposium*, pp. 623–626, 2006.

[39] S.K. Oleksiy, O. Breinbjerg, A. Yiaghjian, "Electrically Small Magnetic Dipole Antennas with Quality factors approaching the Chu Lower Bound", *IEEE Transactions on Antennas Propagation*, Vol. 58, pp. 1869-1906, June 2008.

[40] O.M. Bucci, and G. Franceschetti, "On the Degrees of Freedom of Scattered Fields", *IEEE Transactions on Antennas Propagation*, Vol. 37, No. 7, pp. 918-926, July 1989.

[41] O.M. Bucci, C. Gennarelli, and C. Savarese, "Representation of Electromagnetic Fields over Arbitrary Surfaces by a Finite and Nonredundant Number of Samples", *IEEE Transactions on Antennas Propagation*, Vol. 46, No. 3, pp. 351-359, March 1998.

[42] E. Gazit, "Improved design of the Vivaldi antenna", *IEE Proceedings H (Microwaves, Antennas and Propagation)*, Vol. 135, No.2, pp. 89-92, April 1988.

[43] J. Yang, A. A. Kishk, "A Novel Low-Profile Compact Directional Ultra-Wideband Antenna: The Self-Grounded Bow-Tie Antenna", *IEEE Transactions on Antennas Propagation*, Vol. 60, No. 3, pp. 1214-1220, March 2012.

[44] H. Raza, A. Hussain, J. Yang, P.-S. Kildal, "Wideband Compact 4-Port Dual Polarized Self-Grounded Bowtie Antenna", *IEEE Transactions on Antennas Propagation*, Vol. 62, No. 9, pp. 4468-4473, Sept. 2014.

[45] V. Rodriguez, "An open-boundary quad-ridged guide horn antenna for use as a source in antenna pattern measurement anechoic chambers", *IEEE Antennas and Propagation Magazine*, Vol. 48, No. 2, pp. 157-169, 2006.

[46] H. Nagano, K. Nogami, S. Arai, H. Mimika, J. Yamauchi, "A spiral antenna backed by a conducting plane reflector" *IEEE Transactions on Antennas Propagation*, Vol. 34, No.6, pp. 791-796, June 1986.

Chapter

Appendices

A Derivation of vector integral forms of E- and H-fields

We will here show how the vector integral forms of the E- and H-fields in Section 4.2 can be derived from the more common vector potentials $\mathbf{A}$ and $\mathbf{F}$. The magnetic vector potential is given as an integral over $\mathbf{J}(\mathbf{r}')$ by (4.22). The H-field can be found from $\mathbf{A}$ by using

$$\mathbf{H}_{A} = \iint_{\rho} d\mathbf{H}_{A} \; ; \qquad d\mathbf{H}_{A} = (\nabla \times (\psi \mathbf{J}))dS' \; . \tag{A.1}$$

Applying a known vector identity we can write (A.1) as

$$d\mathbf{H}_{A} = [(\nabla\psi) \times \mathbf{J} + \psi(\nabla \times \mathbf{J})]dS' \; , \tag{A.2}$$

where the last term is zero because the differentiation is with respect to the unprimed observation coordinates, whereas the current is a function of the primed source coordinates only. Since ψ is a scalar function of R, we get

$$\nabla\psi = \frac{d\psi}{dR}\hat{\mathbf{R}} \; ; \qquad \frac{d\psi}{dR} = -\frac{jkR+1}{R}\psi = -jk\left[1 + \frac{1}{jkR}\right]\psi \; , \tag{A.3}$$

which finally gives

$$d\mathbf{H}_{A} = (\mathbf{J} \times \hat{\mathbf{R}})jkC_{N}\Psi dS' \; ; \qquad C_{N} = \left[1 + \frac{1}{jkR}\right] \; . \tag{A.4}$$

The electric field is given by (4.25) with $\mathbf{F} = 0$. This gives

$$d\mathbf{E}_{A} = -j\omega\mu\psi\mathbf{J}dS' - j\frac{1}{\omega\varepsilon}\nabla[\nabla \cdot (\psi\mathbf{J})]dS' \; . \tag{A.5}$$

By using a known vector identity we can develop the second term further to get

$$\nabla[\nabla \cdot (\psi\mathbf{J})] = \nabla[\mathbf{J} \cdot ((\nabla\psi) + \psi(\nabla\mathbf{J}))] \; . \tag{A.6}$$

The last term is zero because the differentiation is with respect to the observation coordinates whereas the current is a function of the source coordinates only. Further, by using (A.3) we get

$$\nabla[\nabla\cdot(\psi\mathbf{J})]=\nabla\left[\mathbf{J}\cdot\frac{\partial\psi}{\partial R}\hat{\mathbf{R}}\right]=\nabla\left[J_{R}\frac{\partial\psi}{\partial R}\right]=\nabla(J_{R})\frac{\partial\psi}{\partial R}+J_{R}\nabla\left(\frac{\partial\psi}{\partial R}\right)\ , \tag{A.7}$$

where $J_{R}=\mathbf{J}\cdot\hat{\mathbf{R}}$. The gradient to $\mathbf{J}\cdot\hat{\mathbf{R}}$ shall be evaluated by differentiation with respect to the observation point coordinates. Therefore,

$$\nabla(\mathbf{J}\cdot\hat{\mathbf{R}})=\mathbf{J}\cdot(\nabla\hat{\mathbf{R}})=\mathbf{J}\cdot\left(\nabla\frac{1}{R}\mathbf{R}\right)\ . \tag{A.8}$$

By using another known vector identity we get

$$\nabla(\frac{1}{R}\mathbf{R})=\frac{1}{R}\nabla\mathbf{R}+\mathbf{R}\nabla\left(\frac{1}{R}\right)=\frac{1}{R}(\hat{\mathbf{x}}\hat{\mathbf{x}}+\hat{\mathbf{y}}\hat{\mathbf{y}}+\hat{\mathbf{z}}\hat{\mathbf{z}})+\mathbf{R}\left(-\frac{1}{R^{2}}\right)\hat{\mathbf{R}}\ . \tag{A.9}$$

Therefore,

$$\nabla(\mathbf{J}\cdot\hat{\mathbf{R}})=\frac{1}{R}[\mathbf{J}-(\mathbf{J}\cdot\hat{\mathbf{R}})\hat{\mathbf{R}}]\ . \tag{A.10}$$

Furthermore, we need

$$\nabla\frac{\partial\psi}{\partial R}=\frac{\partial^{2}\psi}{\partial R^{2}}\hat{\mathbf{R}}=-k^{2}\left[1+\frac{2}{jkR}-\frac{2}{(kR)^{2}}\right]\psi\hat{\mathbf{R}}\ . \tag{A.11}$$

Inserting (A.10) and (A.11) in (A.7) and using $-j\omega\mu=-jk^{2}/(\omega\varepsilon)$ finally gives

$$d\mathbf{E}_{A}=-j\eta k\{\mathbf{J}C_{N_{1}}-(\mathbf{J}\cdot\hat{\mathbf{R}})\hat{\mathbf{R}}C_{N_{2}}\}\psi dS'\ , \tag{A.12}$$

with

$$C_{N_{1}}=\left[1+\frac{1}{jkR}-\frac{1}{(kR)^{2}}\right]\ ,\qquad C_{N_{2}}=\left[1+\frac{3}{jkR}-\frac{3}{(kR)^{2}}\right]\ .$$

B Useful Series Expansions and Integrals

$$\frac{1}{1+x}\approx 1-x+x^{2}\qquad\text{when } x\ll 1\ , \tag{B.13}$$

$$\frac{1+x}{1-x}\approx 1+2x+2x^{2}\qquad\text{when } x\ll 1\ , \tag{B.14}$$

$$e^{x}\approx 1+x+\frac{1}{2}x^{2}\qquad\text{when } x\ll 1\ , \tag{B.15}$$

$$e^{-jk\Delta}\approx 1-jk\Delta-\frac{1}{2}(k\Delta)^{2}\qquad\text{when } k\Delta\ll 1\ , \tag{B.16}$$

$$\ln(1+x)=x-\frac{1}{2}x^{2}\qquad\text{when } x\ll 1\ , \tag{B.17}$$

$$\log(1+x)=(\log e)\ln(1+x)=\log e\left(x-\frac{1}{2}x^{2}\right)\qquad\text{when } x\ll 1\ , \tag{B.18}$$

$$\sqrt{z^{2}+\Delta^{2}}\approx z+\frac{1}{2}\frac{\Delta^{2}}{z}\qquad\text{when } \Delta^{2}\ll z^{2}\ , \tag{B.19}$$

$$10^x = e^{x \ln 10} \approx (1 + x \ln 10) \qquad \text{when } x \ll 1/\ln 10 = 0.43 \ , \tag{B.20}$$

$$A = 10^{A_{\text{dB}}/10} = e^{A_{\text{dB}}(\ln 10)/10} \approx (1 + A_{\text{dB}} \ln 10/10) \approx (1 + 0.23 A_{\text{dB}}) \ , \quad \text{when } A_{\text{dB}} \ll 4.3 \ , \tag{B.21}$$

$$\int \sin\theta d\theta = \cos\theta + C \ , \tag{B.22}$$

$$\int \cos\theta d\theta = -\sin\theta + C \ , \tag{B.23}$$

$$\int e^{\alpha x} \sin(\beta x + \gamma) dx = e^{\alpha x} \frac{1}{\alpha^2 + \beta^2} [\alpha \sin(\beta x + \gamma) - \beta \cos(\beta x + \gamma)] \ . \tag{B.24}$$

C Coordinate transformations

We often need to transform between rectangular and spherical coordinate systems. The following relations between their unit vectors must then be used:

$$\begin{aligned} \hat{\mathbf{x}} &= \sin\theta\cos\varphi\hat{\mathbf{r}} + \cos\theta\cos\varphi\hat{\boldsymbol{\theta}} - \sin\varphi\hat{\boldsymbol{\varphi}} \ , \\ \hat{\mathbf{y}} &= \sin\theta\sin\varphi\hat{\mathbf{r}} + \cos\theta\sin\varphi\hat{\boldsymbol{\theta}} + \cos\varphi\hat{\boldsymbol{\varphi}} \ , \\ \hat{\mathbf{z}} &= \cos\theta\hat{\mathbf{r}} - \sin\theta\hat{\boldsymbol{\theta}} \ , \end{aligned} \tag{C.25}$$

$$\begin{aligned} \hat{\mathbf{r}} &= \sin\theta\cos\varphi\hat{\mathbf{x}} + \sin\theta\sin\varphi\hat{\mathbf{y}} + \cos\theta\hat{\mathbf{z}} \ , \\ \hat{\boldsymbol{\theta}} &= \cos\theta\cos\varphi\hat{\mathbf{x}} + \cos\theta\sin\varphi\hat{\mathbf{y}} - \sin\theta\hat{\mathbf{z}} \ , \\ \hat{\boldsymbol{\varphi}} &= -\sin\varphi\hat{\mathbf{x}} + \cos\varphi\hat{\mathbf{y}} \ . \end{aligned} \tag{C.26}$$

The following cross-products are useful:

$$\begin{aligned} \hat{\mathbf{x}} \times \hat{\mathbf{r}} &= -\sin\varphi\hat{\boldsymbol{\theta}} - \cos\theta\cos\varphi\hat{\boldsymbol{\varphi}} \ , \\ \hat{\mathbf{y}} \times \hat{\mathbf{r}} &= \cos\varphi\hat{\boldsymbol{\theta}} - \cos\theta\sin\varphi\hat{\boldsymbol{\varphi}} \ , \\ \hat{\mathbf{z}} \times \hat{\mathbf{r}} &= \sin\theta\hat{\boldsymbol{\varphi}} \ . \end{aligned} \tag{C.27}$$

D Useful material parameters

See Table D.1, D.2 and D.3.

Table D.1: Values in free space [1].

Symbol	**Name**	**Value in free space**
η	Free space impedance	$376.7303\,\Omega \approx 120\pi\,\Omega$
c	Velocity of EM waves	2.99792458×10^8 m/sec

Table D.2: Relative permittivity $\varepsilon_r' - j\varepsilon_r'' = \varepsilon_r'[1 - j\tan\delta]$ of some dielectric materials. ε_r'' is called the dielectric dissipation factor, δ the loss angle, and $\tan\delta$ the loss tangent [2].

Frequency [GHz]	**0.1**	**0.1**	**3.**	**3.**	**25.**	**25.**
Material / Parameter	$\boldsymbol{\varepsilon_r'}$	$\boldsymbol{\tan\delta}$	$\boldsymbol{\varepsilon_r'}$	$\boldsymbol{\tan\delta}$	$\boldsymbol{\varepsilon_r'}$	$\boldsymbol{\tan\delta}$
Quarts glass (silicon dioxide)	3.78	.0002	3.78	.00006	3.78	.00025
Polystyrene	2.55	.0001	2.55	.00033	2.54	.0012
Polystyrene foam (0.25% filler)	1.03	-	1.03	.0001	1.03	-
Magnesium titanat (ceramics)	13.9	.0005	13.8	.0017	13.7	.0065
Teflon	2.1	.0002	2.1	.00015	2.08	.0006

Table D.3: Resistivity in Ωcm $\times 10^{-6}$ and skin depth of some conducting materials at room temperature. The conductivity is one over the resistivity.

		Skin depth		
Material	**Resistivity**	**100 MHz**	**1 GHz**	**10 GHz**
Aluminium	2.62	8 μm	2.6 μm	0.8 μm
Copper	1.7241	6.6 μm	2.1 μm	0.7 μm
Gold	2.44	7.9 μm	2.5 μm	0.8 μm
Iron	9.71	15.7 μm	5.0 μm	1.6 μm
Silver	1.62	6.4 μm	2.0 μm	0.6 μm
Steel (different kinds)	13-90	18-48 μm	5.7-15.2 μm	1.8-4.8 μm

E Bessel functions

The following integrals can be expressed in terms of the Bessel functions J_0, J_1 and J_2 of first kind, and where the index denote the zeroth, first and second order:

$$\int_0^{2\pi} e^{ja\cos(\varphi-\varphi_0)}d\varphi = 2\pi J_0(a) \ ,$$

$$\int_0^{2\pi} e^{-j\varphi}e^{ja\cos(\varphi-\varphi_0)}d\varphi = 2\pi j e^{-j\varphi_0}J_1(a) \ ,$$

$$\int_0^{2\pi} e^{-j2\varphi}e^{ja\cos(\varphi-\varphi_0)}d\varphi = -2\pi e^{-j2\varphi_0}J_2(a) \ .$$

F Equivalent circuits of two-port circuits

A general two-port circuit is shown in Fig. F.1. The circuit itself is of no interest to us, the only thing that is interesting is how the circuit responds to external sources and loads. Thus, the two-port is viewed as a "black box" and we deal only with the voltages and currents at the ports.

Generally we can write relations between the voltages and currents in matrix form as:

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \ ,$$

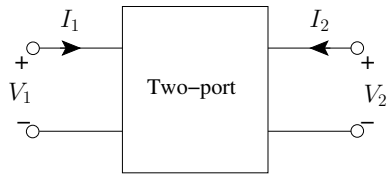


Figure F.1: Port voltages and currents for a two-port circuit.

where X and Y represent any of the voltages or currents (V_1, V_2, I_1 and I_2).

The matrix elements can be determined by using the relation:

$$A_{ij} = \left.\frac{X_i}{Y_j}\right|_{Y_{k\neq j,k=1,2}=0} .$$

It can be shown that we can write down six different matrix relations for the two-port circuit in Fig. F.1. Often only three of the six possible representations are used and consequently we restrict ourselves to these three representations. The representations that we will use are the impedance, admittance and the chain parameters. We will also consider another way to describe two-ports which is not based on the total voltages and currents but rather on the scattering of waves in the ports, i.e., the scattering parameters.

F.1 The impedance parameters for a two-port

The impedance parameters for a two-port are defined as:

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \text{ , i.e., } [V] = [Z][I] .$$

From the matrix relation we obtain the equivalent circuit in Fig. F.2(a).

For a reciprocal two-port (most passive, linear circuits are reciprocal, but not if they contain a magnetised material such as RF circulators) the Z-matrix is symmetrical, i.e., $Z_{12} = Z_{21}$ (the transpose of the Z-matrix is equal to the matrix itself). Thus, for a reciprocal circuit it is sufficient to describe the circuit with only three impedance elements, Fig. F.2(b).

It is convenient to use the Z-matrix representation when two two-port circuits are connected in series, Fig. F.2(c).

For the series connection in Fig. F.2(c) the resulting Z-matrix is given by simple matrix addition, i.e., $[Z] = [Z_A] + [Z_B]$.

F.2 The admittance parameters for a two-port

The admittance parameters for a two-port are defined as:

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \text{ , i.e., } [I] = [Y][V] .$$

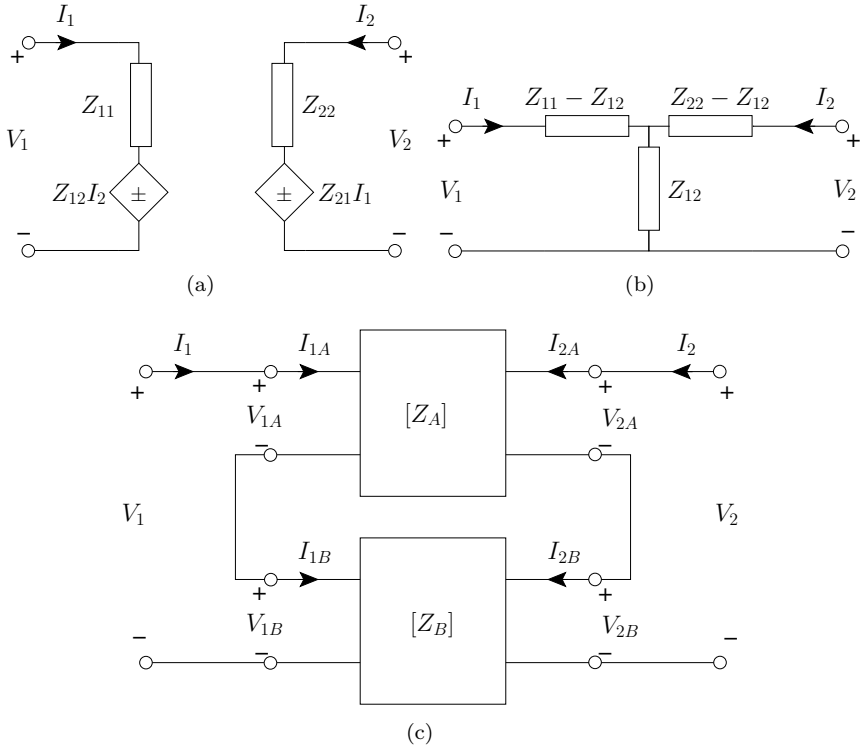


Figure F.2: (a) Equivalent circuit for a two-port described by its Z-matrix. (b) Circuit representation for a reciprocal two-port, T-network. (c) Series connection of two two-ports.

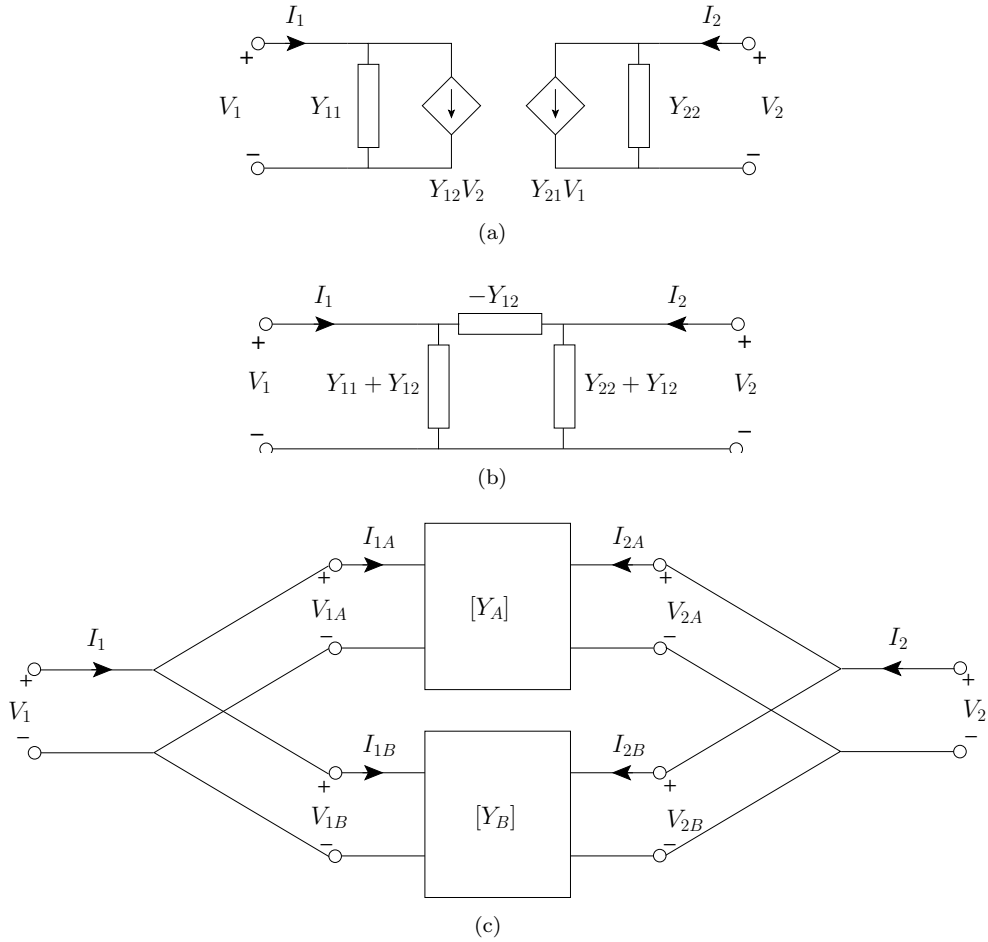


Figure F.3: (a) Equivalent circuit for a two-port described by its Y-matrix. (b) Circuit representation for a reciprocal two-port, π-network. (c) Parallel connection of two two-ports.

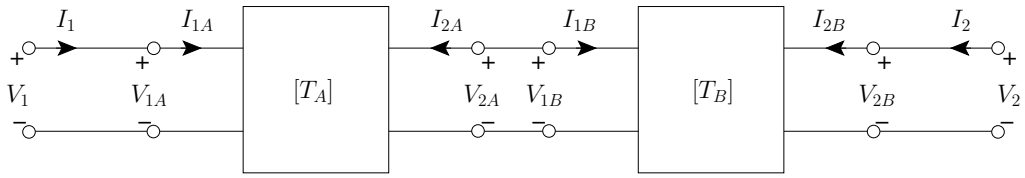


Figure F.4: Cascade connection of two two-ports.

From the matrix relation we obtain the equivalent circuit in Fig. F.3(a).

For a reciprocal two-port the Y-matrix is symmetrical, i.e., $Y_{12} = Y_{21}$. Thus, for a reciprocal circuit it is sufficient to describe the circuit with only three admittance elements, see Fig. F.3(b).

It is convenient to use the Y-matrix representation when two two-port circuits are connected in parallel, see Fig. F.3(c). The resulting Y-matrix is given by simple matrix addition, i.e., $[Y] = [Y_A] + [Y_B]$.

F.3 The chain parameters for a two-port

The chain parameters for a two-port are defined as:

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V_2 \\ -I_2 \end{bmatrix} \text{ , i.e., } [X_1] = [T][X_2] \text{ .}$$

Note the minus sign for the current I_2.

There exists no equivalent circuit representation for the chain parameters.

For a reciprocal two-port the elements in the T-matrix fulfill the following relation: $AD - BC = 1$ (i.e., the determinant equals unity). For the special case of a reciprocal two-port the inverse T-matrix will have a simple form and therefore we can express the output quantities in terms of the input quantities in the following simple way:

$$\begin{bmatrix} V_2 \\ -I_2 \end{bmatrix} = \begin{bmatrix} D & -B \\ -C & A \end{bmatrix} \begin{bmatrix} V_1 \\ I_1 \end{bmatrix} \text{ .}$$

It is convenient to use the T-matrix representation when two two-ports are connected in cascade, Fig. F.4.

For the cascade connection in Fig. F.4 the resulting T-matrix is given by simple matrix multiplication, i.e., $[T] = [T_A][T_B]$.

F.4 The scattering parameters for a two-port

The scattering parameters are not defined in terms of the total voltages and currents shown in Fig. F.1 and consequently not one of the six possible representations mentioned before. The scattering parameters are defined in terms of transmitted and reflected waves at the

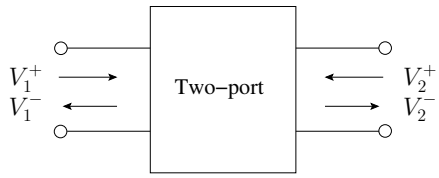


Figure F.5: Transmitted and reflected waves used for the definition of the scattering parameters.

ports, Fig. F.5. The scattering parameters are included here since when measuring the characteristics of a two-port (or N-port), a network analyzer is often used and the analyzer is designed for measuring the scattering parameters.

The scattering parameters for a two-port are defined as:

$$\begin{bmatrix} V_1^- \\ V_2^- \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} V_1^+ \\ V_2^+ \end{bmatrix} \text{ , i.e., } [V^-] = [S][V^+] \text{ .}$$

The total voltage at a port is equal to the sum of the amplitudes of the transmitted and reflected waves, and similarly for the currents, thus:

$$\begin{cases} V_i = V_i^+ + V_i^- \\ I_i = I_i^+ - I_i^- \end{cases} ,$$

where, for a two-port, $i = 1 \ldots 2$ and the minus sign for the current is due to the reference directions. If we assume the impedance in the measuring system connected to the two-port is the same on both sides of the two-port and equal to Z_0, the following relations between current amplitudes and voltage amplitudes are obtained:

$$\begin{cases} I_i^+ = \frac{V_i^+}{Z_0} \\ I_i^- = \frac{V_i^-}{Z_0} \end{cases} \rightarrow I_i = \frac{1}{Z_0}(V_i^+ - V_i^-) \text{ .}$$

Normally the impedance Z_0 is equal to $50\,\Omega$.

Using the above expressions we can now obtain the following relations between the scattering parameters and the impedance parameters:

$$[S] = ([Z] + Z_0[U])^{-1}([Z] - Z_0[U]) \text{ ,}$$
$$[Z] = Z_0([U] + [S])([U] - [S])^{-1} \text{ ,}$$

where $[U]$ is the identity matrix and $(\ldots)^{-1}$ means matrix inversion.

F.5 Extension to N-ports

When dealing with N-ports we are here only interested in N-ports with an equal number of input and output ports, i.e., N-ports with an even number of ports. Thus, we consider N-ports with $2M$ ports, where M is an integer number.

In order to distinguish between the input and output ports we start the numbering with the input ports, i.e., the input ports will have numbers ranging from 1 to M and the output

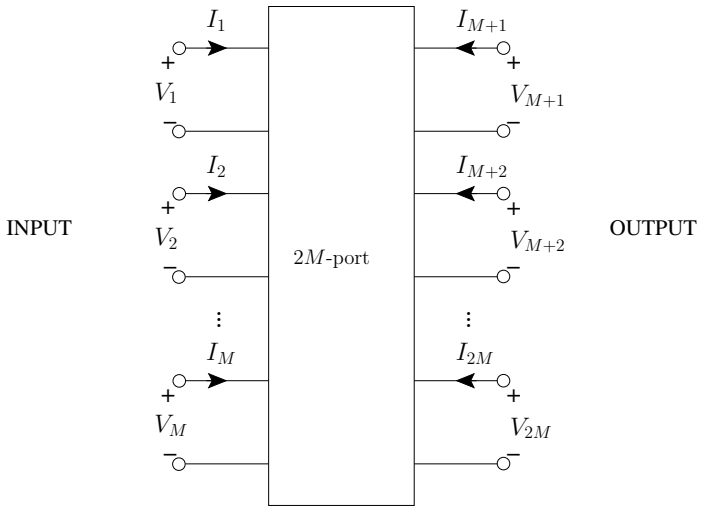


Figure F.6: N-port with 2M ports.

ports numbers from $M+1$ to $2M$. It should be noted that the ports are often connected unsymmetrically, e.g., the lower nodes of the ports are connected to a common point (the ground point).

As for the special case of a two-port we can write down the relations between the port voltages and currents in matrix form, the only difference is that the dimension of the matrix will now be N by N.

$$\begin{bmatrix} X_1 \\ X_2 \\ \cdots \\ X_N \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & A_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ A_{N_1} & A_{N_2} & \cdots & A_{NN} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_N \end{bmatrix} ,$$

where X and Y represent any of the port voltages or currents.

The matrix elements can be determined by using the relation:

$$A_{ij} = \left. \frac{X_i}{Y_j} \right|_{Y_{k \neq j, k=1,\ldots N}=0} .$$

As is clear from the matrix relation above we have many different possibilities to represent an N-port. However, as for the special case of two-ports we also for this case restrict the number of representations to three, the impedance, admittance and the chain parameters.

F.6 The impedance parameters for an N-port

The impedance parameters for an N-port with $2M$ ports are defined as:

$$\begin{bmatrix} V_1 \\ V_2 \\ \dots \\ V_{2M} \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} & \dots & Z_{12M} \\ Z_{21} & Z_{22} & \dots & Z_{22M} \\ \dots & \dots & \dots & \dots \\ Z_{2M1} & Z_{2M2} & \dots & Z_{2M2M} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ \dots \\ I_{2M} \end{bmatrix} ,$$

$$\text{i.e., } \begin{bmatrix} [V_1] \\ [V_2] \end{bmatrix} = \begin{bmatrix} [Z_{11}] & [Z_{12}] \\ [Z_{21}] & [Z_{22}] \end{bmatrix} \begin{bmatrix} [I_1] \\ [I_2] \end{bmatrix} ,$$

$$\text{where } [V_1] = \begin{bmatrix} V_1 & V_2 & \dots & V_M \end{bmatrix}^T , \qquad [I_1] = \begin{bmatrix} I_1 & I_2 & \dots & I_M \end{bmatrix}^T ,$$

$$[V_2] = \begin{bmatrix} V_{M+1} & V_{M+2} & \dots & V_{2M} \end{bmatrix}^T , \qquad [I_2] = \begin{bmatrix} I_{M+1} & I_{M+2} & \dots & I_{2M} \end{bmatrix}^T ,$$

for a reciprocal N-port the Z-matrix is symmetrical, i.e., $Z_{ij} = Z_{ji}$.

The impedance parameters are convenient to use when two N-ports are connected in series. For the series connection the resulting Z-matrix is given by matrix addition, i.e., $[Z] = [Z_A] + [Z_B]$.

F.7 The admittance parameters for an N-port

The admittance parameters for an N-port with $2M$ ports are defined as:

$$\begin{bmatrix} I_1 \\ I_2 \\ \dots \\ I_{2M} \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{12} & \dots & Y_{12M} \\ Y_{21} & Y_{22} & \dots & Y_{22M} \\ \dots & \dots & \dots & \dots \\ Y_{2M1} & Y_{2M2} & \dots & Y_{2M2M} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ \dots \\ V_{2M} \end{bmatrix} ,$$

$$\text{i.e., } \begin{bmatrix} [I_1] \\ [I_2] \end{bmatrix} = \begin{bmatrix} [Y_{11}] & [Y_{12}] \\ [Y_{21}] & [Y_{22}] \end{bmatrix} \begin{bmatrix} [V_1] \\ [V_2] \end{bmatrix} ,$$

$$\text{where } [V_1] = \begin{bmatrix} V_1 & V_2 & \dots & V_M \end{bmatrix}^T , \qquad [I_1] = \begin{bmatrix} I_1 & I_2 & \dots & I_M \end{bmatrix}^T ,$$

$$[V_2] = \begin{bmatrix} V_{M+1} & V_{M+2} & \dots & V_{2M} \end{bmatrix}^T , \qquad [I_2] = \begin{bmatrix} I_{M+1} & I_{M+2} & \dots & I_{2M} \end{bmatrix}^T .$$

For a reciprocal N-port the Y-matrix is symmetrical, i.e., $Y_{ij} = Y_{ji}$.

The admittance parameters are convenient to use when two N-ports are connected in parallel. For the parallel connection the resulting Y-matrix is given by matrix addition, i.e., $[Y] = [Y_A] + [Y_B]$.

F.8 The chain parameters for an N-port

The chain parameters for an N-port with $2M$ ports are defined as:

$$\begin{bmatrix} V_1 \\ V_2 \\ \cdots \\ I_{M-1} \\ I_M \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} & \cdots & \cdots & T_{12M} \\ T_{21} & T_{22} & \cdots & \cdots & T_{22M} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ T_{2M1} & T_{2M2} & \cdots & \cdots & T_{2M2M} \end{bmatrix} \begin{bmatrix} V_{M+1} \\ V_{M+2} \\ \cdots \\ -I_{2M-1} \\ -I_{2M} \end{bmatrix} ,$$

$$\text{i.e., } \begin{bmatrix} [V_1] \\ [I_1] \end{bmatrix} = \begin{bmatrix} [A] & [B] \\ [C] & [D] \end{bmatrix} \begin{bmatrix} [V_2] \\ -[I_2] \end{bmatrix} ,$$

$$\text{where } [V_1] = \begin{bmatrix} V_1 & V_2 & \cdots & V_M \end{bmatrix}^T , \qquad [I_1] = \begin{bmatrix} I_1 & I_2 & \cdots & I_M \end{bmatrix}^T ,$$

$$[V_2] = \begin{bmatrix} V_{M+1} & V_{M+2} & \cdots & V_{2M} \end{bmatrix}^T , \qquad [I_2] = \begin{bmatrix} I_{M+1} & I_{M+2} & \cdots & I_{2M} \end{bmatrix}^T .$$

The chain parameters are convenient to use when two N-ports are connected in cascade. For the cascade connection the resulting T-matrix is given by matrix multiplication, i.e., $[T] = [T_A][T_B]$.

F.9 The scattering parameters for an N-port

The scattering parameters for an N-port with $2M$ ports are defined as:

$$\begin{bmatrix} V_1^- \\ V_2^- \\ \cdots \\ V_{2M}^- \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{12M} \\ S_{21} & S_{22} & \cdots & S_{22M} \\ \cdots & \cdots & \cdots & \cdots \\ S_{2M1} & S_{2M2} & \cdots & S_{2M2M} \end{bmatrix} \begin{bmatrix} V_1^+ \\ V_2^+ \\ \cdots \\ V_{2M}^+ \end{bmatrix} , \text{ i.e., } [V^-] = [S][V^+] .$$

F.10 General relations between matrix representations

Relations between Z & Y

$$[Z] = [Y]^{-1} ,$$
$$[Y] = [Z]^{-1} .$$

In cases when matrix inversion is not possible the circuit cannot be represented in both ways.

Relations between Z & T

$$\begin{bmatrix} [Z_{11}] & [Z_{12}] \\ [Z_{21}] & [Z_{22}] \end{bmatrix} = \begin{bmatrix} [A] \cdot [C]^{-1} & ([A] \cdot [C]^{-1} \cdot [D] - [B]) \\ [C]^{-1} & [C]^{-1} \cdot [D] \end{bmatrix} ,$$
$$\begin{bmatrix} [A] & [B] \\ [C] & [D] \end{bmatrix} = \begin{bmatrix} [Z_{11}] \cdot [Z_{21}]^{-1} & ([Z_{11}] \cdot [Z_{21}]^{-1} \cdot [Z_{22}] - [Z_{12}]) \\ [Z_{21}]^{-1} & [Z_{21}]^{-1} \cdot [Z_{22}] \end{bmatrix} .$$

An equal number of input and output ports are assumed.

Relations between Y & T

$$\begin{bmatrix}[Y_{11}] & [Y_{12}]\\ [Y_{21}] & [Y_{22}]\end{bmatrix} = \begin{bmatrix}[D]\cdot[B]^{-1} & ([C]-[D]\cdot[B]^{-1}\cdot[A])\\ -[B]^{-1} & [B]^{-1}\cdot[A]\end{bmatrix} ,$$
$$\begin{bmatrix}[A] & [B]\\ [C] & [D]\end{bmatrix} = \begin{bmatrix}-[Y_{21}]^{-1}\cdot[Y_{22}] & -[Y_{21}]^{-1}\\ -([Y_{11}]\cdot[Y_{21}]^{-1}\cdot[Y_{22}]-[Y_{21}]) & -[Y_{11}]\cdot[Y_{21}]^{-1}\end{bmatrix} .$$

An equal number of input and output ports are assumed.

Relations between Z & S

$$[Z] = Z_0([U]+[S])([U]-[S])^{-1} ,$$
$$[S] = ([Z]+Z_0[U])^{-1}([Z]-Z_0[U]) ,$$

where $[U]$ is the identity matrix and Z_0 is the impedance in the measuring system (usually $50\,\Omega$).

Relations between Y & S

$$[Y] = \frac{1}{Z_0}([U]-[S])([U]+[S])^{-1} ,$$
$$[S] = ([U]+Z_0[Y])^{-1}([U]-Z_0[Y]) ,$$

where $[U]$ is the identity matrix and Z_0 is the impedance in the measuring system (usually $50\,\Omega$).

G References

[1] *The NIST Reference on Constants, Units, and Uncertainty*, www.nist.gov/cuu, National Institute of Standards and Technology, USA.

[2] *Reference Data for Radio Engineers*, Howard W. Sams & Co. ITT, 1975.

Chapter

About the Author

The author Per-Simon Kildal has his Master degree (1976) and two Doctoral degrees (1982 & 1990) from the Norwegian Institute of Technology (NTH) in Trondheim. The Master degree had the highest grades with notification to The King of Norway. He was during 1979-1989 employed as Research Scientist at ELAB, SINTEF Group in Trondheim, and received as such their annual award in 1994. He became a professor at Chalmers University of Technology, Gothenburg, Sweden in 1989.

Professor Kildal has had many awards and appointments for the IEEE Antennas and Propagation Society (IEEE APS) since he became a member in 1983. He has written more than 150 journal articles or letters, most of them published in IEEE APS journals, resulting in a publication h-index of 41. He has also received prestigious acknowledgements from this society: R.W.P. King Award for the best paper of a young author in 1985, S. A. Schelkunoff Award for the best paper in their Transactions journal in 1991, IEEE APS Fellow in 1995, Distinguished Lecturer during the two periods 1991-94 and 2011-14, and in 2011 he received their highest award, the prestigious IEEE APS Distinguished Achievements Award. Professor Kildal has also been very active in European School of Antennas (ESoA) as an organizer of two one-week short courses on Reflector and Lens Antennas, and Metasurfaces for Antennas, respectively. In 2013 he was Chairman of the 7th EuCAP conference in Gothenburg with 1100 delegates. The owner of EuCAP is the European Association on Antennas and Propagation (EurAAP).

Professor Kildal makes innovative and application-oriented research within three main areas. These are described below.

Reflector antennas and feeds: During 1980-85 Kildal started his carrier at ELAB by developing EM methods and software for designing the $40[m]\times120[m]$ cylindrical reflector and line feeds of the European EISCAT Scientific Association. This was done as a PhD project for Professor Tor Hagfors, Director of EISCAT. Hagfors went back to USA in 1982, and made Kildal responsible for developing the EM methods and software for designing the Gregorian dual-reflector feed of the 300 m diameter radio telescope in Arecibo. This work was done on ELAB projects for Cornell University during 1986-90. During these years at ELAB, Kildal also invented two self-supported reflector-feeds that became industrial products: *i)* The patented dipole-disk with ring that was used as a feed in the INMARSAT terminal of the Norwegian company NERA for

more than 10 years, and *ii)* the patented hat-fed reflector (also called Comhat antenna) that in 1998 was commercialized in Comhat AB for radio links. The Comhat antenna has till now been manufactured in more than 1 million copies mainly for use in the MINI-LINK product of Ericsson AB. Kildal invented a third popular feed in 2003, the so-called eleven antenna. The eleven antenna is a very compact log-periodic dual-dipole feed with more than decade bandwidth. It is commercialized in Gapwaves AB, and is used as a feed in ultra wideband applications, such as radio telescopes for SKA and VLBI2010.

Research on metasurfaces: Kildal introduced in 1988 the concept of "Soft and hard surfaces" in electromagnetics. This is today regarded as the first metamaterials concept. The soft surface is an early forerunner of electromagnetic bandgap surfaces (EBGs), which has been a popular research area since 2000. Kildal used the hard surface to realize invisibility (remove shadowing) during 1990-97. This is nowadays called cloaking, so the hard surface was used to experimentally realize cloaking more than 10 years before cloaking became a popular research topic in 2007. Based on these fundamental surface representations Kildal invented a new gap waveguide technology for mm-waves in 2008. He has now a large 2.5 M€ research grant from European Research Council (ERC) in order to explore the industrial potentials of gap waveguides. The intellectual property rights are gathered in Gapwaves AB.

Research on OTA technology (OTA means Over-The-Air testing of wireless devices): Professor Kildal introduced the reverberation chamber for OTA-testing in 2000. This inward-reflecting chamber is the complete opposite to an anechoic chamber, and it was used to emulate what later has been referred to as Rich Isotropic MultiPath (RIMP). He and his research group researched during 2000-2010 this reverberation chamber into an accurate instrument for OTA testing. It is actually the fastest and most compact OTA test instrument available. Bluetest AB was formed in 2000 in order to commercialize the approach, and in 2009 Bluetest was the first to perform OTA throughput measurements of 5G LTE devices. After this Bluetest has become the world-leading provider of such instruments, and till now they have sold more than 200 test systems worldwide.

List of achievements

IEEE and similar involvements (IEEE APS = IEEE Antennas and Propagation Society, IEEE TAP = IEEE Transactions on Antennas and Propagation):

1985: R.W.P. King Award for best paper by young author in IEEE TAP
1991: S. A. Schelkunoff Award for best paper in IEEE TAP
1995: IEEE Fellow
1991-94: IEEE APS Distinguished Lecturer
2011-14: IEEE APS Distinguished Lecturer
2011: IEEE APS Distinguished Achievements Award (their highest award)
2013: Chairman of EuCAP 2013 in Gothenburg (1100 delegates)

Reflector antennas and feeds:

1980-85: Designed EISCAT 40 m × 120 m cylindrical reflector and line feeds
1986-90: Designed Gregorian feed of ARECIBO radio telescope for Cornell University
1980s: Invented dipole-disk with ring, used by Norwegian company NERA
1986: Invented hat-fed reflector antenna (Comhat antenna)

1998: Founded Comhat AB; till now one million hat antennas produced for Ericsson AB. Comhat antenna is now in LEAX Arkivator Telecom.
2003: Invented UWB "the eleven antenna", a feed for VLBI 2010 radio telescopes

Metamaterials:
1988: Invented "Soft and hard surfaces" - the first metamaterial concept. Soft surface is forerunner of EBG surfaces (popular research topic since 2000).
1990-97: Introduced hard surface to realize invisibility, nowadays called cloaking, more than 10 years before cloaking became a hot research topic in 2007 .
2008: Invented new gap waveguide technology for mm-waves
2013: Received 2.5 M€ research grant from ERC to explore gap waveguides.

OTA technology (OTA means Over-The-Air testing of wireless devices):
2000: Invented OTA tests in reverberation chamber and started Bluetest AB.
2000-12: Researched reverberation chamber to accurate OTA instrument.
2002: Defined and measured effective diversity gain
2005: Measured MIMO capacity and defined embedded element efficiency.
2005: Introduced rich isotropic multipath (RIMP) as reference environment
2009: Bluetest was first to perform OTA measurements of LTE devices.
2011: Introduced theoretical LTE throughput model based on threshold receiver.
2011-14: Bluetest has major breakthrough in OTA market.
2014: 200 Bluetest RTS systems have been sold to mobile phone manufacturers.

Endorsements of the book:

Prof. ***Raj Mittra****, EMC Lab, Penn State and Central Florida Universities:*
"This is one of the most comprehensive works on Antennas as well as on propagation from a design viewpoint. It has its own unique and refreshing ways to present the complex subjects. The author is well known in our field because of the many pioneering contributions he has made to the fields of antennas and propagation. I am pleased to endorse this book very strongly."

Prof. ***Werner Wiesbeck****, Karlsruhe Institute of Technology:*
"I do not know any book in Antenna Engineering that covers theory and system engineering so concise and understandable. It stimulates imagination and vision. This book is a must also for students and engineers of wireless communications and radar sensing."

Prof. Em. ***Ted Simpson****, University of South Carolina, Historian of IEEE APS:*
"In this 2nd edition, Kildal continues to present advanced antenna theory in a way that satisfies the needs of both beginners and professionals. E.g., he defines a Rich Isotropic MultiPath (RIMP) to uniquely characterize antennas in multipath. Extensive lists of references, and comments by Prof. Maci, give added value."

Prof. ***Ahmed A. Kishk****, Concordia University, Montreal, Fellow of IEEE APS:*
"I have used the previous edition in my classroom for several years. The compact formulations make it easy to relate Math to Physics. The book introduces novel concepts that tie traditional theories to modern technology. The defined RIMP environment makes it easier to understand MIMO technology."

Per-Simon Kildal is a professor at Chalmers University of Technology in Gothenburg, Sweden. He has several patents on antenna designs that have been commercialized. He is the inventor of the reverberation chamber used for measuring performance of wireless devices in multipath, commercialized in Bluetest AB. His latest invention is called GAP waveguides; a millimeter-wave technology very useful for electronic packaging. He earned two doctoral degrees from Norwegian Institute of Technology in Trondheim.

Foundations of Antenna Engineering: A Unified Approach for Line-of-Sight and Multipath

This is the first resource that contains a holistic treatment of traditional antennas mounted on masts (Line-of-Sight antenna systems), as well as small antennas used on modern wireless devices such as smart phones being subject to signal variations (fading) due to multipath propagation. The focus is on characterization, as well as describing classical antennas by modern complex vector theory -- Thereby linking together many disciplines such as electromagnetic theory, classical antenna theory, wave propagation, and antenna system performance. Overall, the book represents a rethinking of the way basic antenna theory is presented.

A MATLAB handbook accompanies this book, made available at www.kildal.se. This can be used to provide initial antenna designs before entering into time-consuming numerical simulations.

"This is one of the most comprehensive works on antennas, as well as on propagation, from a design viewpoint. It has its own unique and refreshing ways to present the complex subjects. " - Professor Raj Mittra, EMC Lab, Penn State and Central Florida Universities

"I do not know any antenna book that covers theory and system engineering so concisely. The book is a MUST." - Professor Werner Wiesbeck, Karlsruhe Institute of Technology, Germany.

"I have used the previous edition in my classroom for several years. The book introduces novel concepts that tie traditional theories to modern technology. The defined Rich Isotropic Multipath (RIMP) environment makes it easier to understand MIMO technology." - Prof. Ahmed A. Kishk , Concordia University, Montreal.

Contents Overview: Introduction, History, and Terminology; Characterization of Directive Antennas; Characterization in Multipath; Theory of Radiation from Current Sources; Small Wire and Slot Antennas; Microstrip Antennas and Spectral Domain Methods; Radiation from Apertures; Horn Antennas; Reflector Antennas; Array Antennas; Fundamental Limitations; Appendices.

Prof. Stefano Maci at University of Siena, Director of European School of Antennas, has written comments to all chapters, including extensive reference lists.

KILDAL

ISBN 978-91-637-8514-6

9 789163 785146 >