

LECTURE NOTES
ANTENNA ENGINEERING

by Natalia K. Nikolova

for the course ECE753
Department of Electrical and Computer Engineering
McMaster University

Version 2018

cite as: Natalia K. Nikolova, *Notes on Antenna Engineering 2018*, McMaster University, Canada,
available online at http://www.ece.mcmaster.ca/faculty/nikolova/antenna_dload/current_lectures/

LECTURE 1: Introduction into Antenna Studies

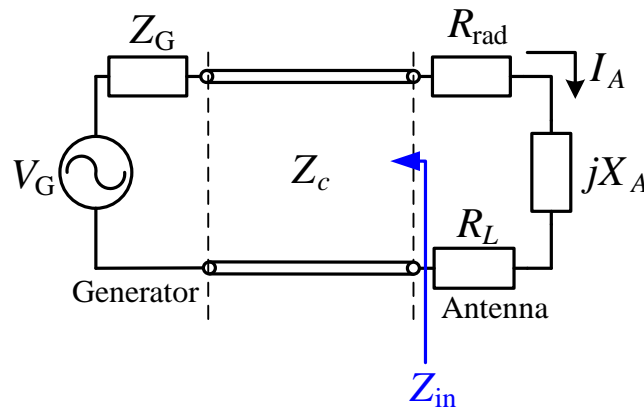
(Definition and circuit theory description. Brief historical notes. General review of antenna geometries and arrangements. Wireless vs. cable communication systems. The radio-frequency spectrum.)

1. Definition and circuit theory description

The antenna (aerial, EM radiator) is a device, which radiates or receives electromagnetic waves.

The antenna is the transition between a guiding device (transmission line, waveguide) and free space (or another usually unbounded medium). Its main purpose is *to convert the energy of a guided wave into the energy of a free-space wave (or vice versa) as efficiently as possible, while at the same time the radiated power has a certain desired pattern of distribution in space*. At lower frequencies, where the length of the transmission line is negligible, we can view the antenna as a **device that converts free-space EM waves into voltage/current signals or vice versa**.

a) transmission-line Thevenin equivalent circuit of a radiating (transmitting) antenna



V_G - voltage-source generator (transmitter)

Z_G - impedance of the generator (transmitter)

Z_c - characteristic impedance of the connecting TL

R_{rad} - radiation resistance (relates to the radiated power as

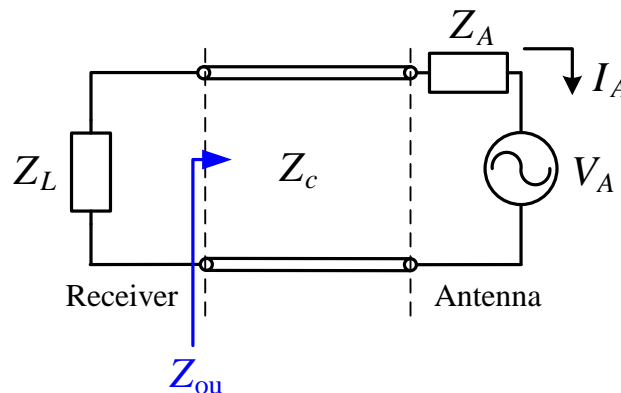
$$P_{\text{rad}} = I_A^2 \cdot R_{\text{rad}})$$

- R_L - loss resistance (related to conduction and dielectric losses)
- jX_A - antenna reactance
- Z_{in} - input impedance of feed network as seen from antenna terminals
- $Z_A = (R_{rad} + R_L) + jX_A$ - antenna impedance

One of the most important issues in antenna design is the matching of the antenna to the transmission line (TL) and the generator ($Z_A = Z_{in}^*$). Matching is often measured in terms of the voltage standing-wave ratio (VSWR). Standing waves must be avoided because they may cause arcing or discharge in the TL in high-power RF systems (radar, broadcasting). But the main benefit of good impedance match (with low VSWR) is the maximum power transfer to/from the antenna.

The resistive/dielectric losses (see R_L) are not desirable either. They decrease the efficiency of the antenna. On the other hand, in special applications such as ultra-wideband (UWB) antennas in imaging and radar, the antenna resistance may be increased intentionally in order to improve the bandwidth and suppress “ringing” in the transmitted or received signals.

b) transmission-line Thevenin equivalent circuit of a receiving antenna

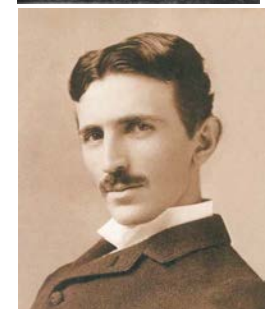


Z_{ou} - output impedance of the antenna-plus-feed network, which serves as a signal generator as seen from the receiver terminals

The antenna is a critical component in a wireless communication system. A good design of the antenna can relax the system requirements and improve its overall performance.

2. Brief historical notes

- James Clerk Maxwell formulates the mathematical model of electromagnetism (classical electrodynamics), “*A Treatise on Electricity and Magnetism*”, 1873. He shows that light is an electromagnetic (EM) wave, and that all EM waves propagate through space with the same speed, the speed of light.
- Heinrich Rudolph Hertz demonstrates in 1886 the first wireless EM wave system: a $\lambda/2$ -dipole is excited with a spark; it radiates predominantly at $\lambda \approx 8$ m; a spark appears in the gap of a receiving loop some 20 m away. In 1890, he publishes his memoirs on electrodynamics, replacing all potentials by field strengths.¹
- May 7, 1895, a telegraph communication link is demonstrated by the Russian scientist, Alexander Popov. A message is sent from a Russian Navy ship 30 miles out in sea, all the way to his lab in St. Petersburg, Russia. This accomplishment is little known today.
- In 1892, Tesla delivers a presentation at the IRE of London about “transmitting intelligence without wires,” and, in 1895, he transmits signals detected 80 km away. His patent on wireless links precedes that of Marconi.
- Guglielmo Marconi sends signals over large distances and successfully commercializes wireless communication systems. In 1901, he performs the first transatlantic transmission from Poldhu in Cornwall, England, to Newfoundland, Canada. He receives the Nobel prize for his work in 1909.



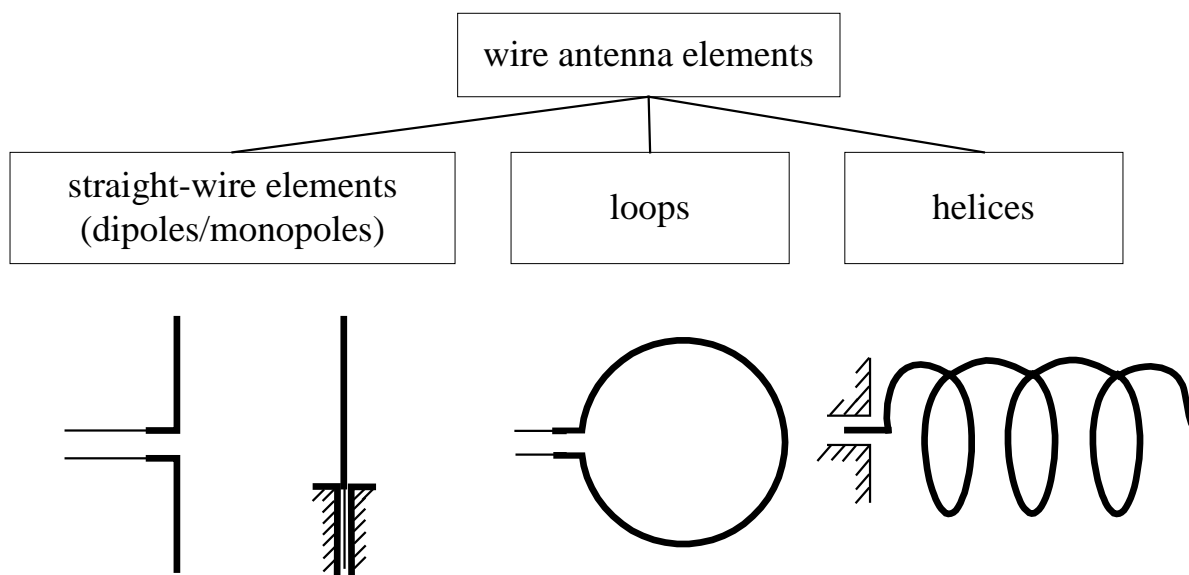
¹ Similar work is done at about the same time by the English scientist Oliver Heaviside.
Nikolova 2018

- The beginning of 20th century (until WW2) marks the boom in wire-antenna technology (dipoles and loops) and in wireless technology as a whole, which is largely due to the invention of the DeForest triode tube, used as a radio-frequency (RF) generator. Radio links are realized up to UHF (about 500 MHz) and over thousands of kilometers.
- WW2 marks a new era in wireless communications and antenna technology. The invention of new microwave generators (magnetrons and klystrons) leads to the development of the microwave antennas such as waveguide apertures, horns, reflectors, etc.

3. General review of antenna geometries and arrangements

3.1. Single-element radiators

A. Wire radiators (single-element)



There is a variety of shapes corresponding to each group. For example, loops can be circular, square, rhombic, etc. Wire antennas are simple to make but their dimensions are commensurable with a wavelength. This limits the frequency range of their applicability. At low frequencies, these antennas become increasingly large. At very high frequencies, they are very small and the parasitics become difficult to control.

B. Aperture antennas (single element)



(Q-par Angus)

(a) pyramidal horn



(b) conical horn



[Radiometer Physics Gmbh]

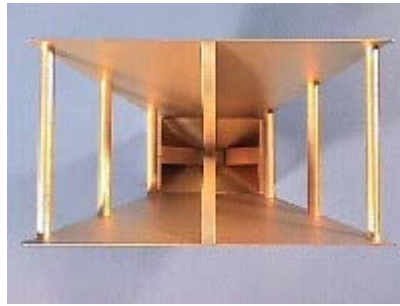
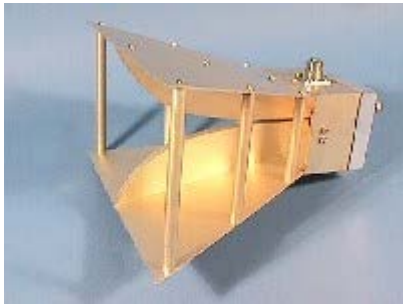
www.quinstar.com



www.labvolt.com

(c) open rectangular waveguides

Aperture antennas were developed before and during WW2 together with waveguide technology. Waveguides were primarily developed to transfer high-power microwave signals (cm wavelengths), generated by high-power sources such as magnetrons and klystrons. These types of antennas are preferable in the frequency range from 1 to 20 GHz.



[Quinstar Technology Inc.]

(d) double-ridge horns (TEM, linear polarization, ultra-wide band)

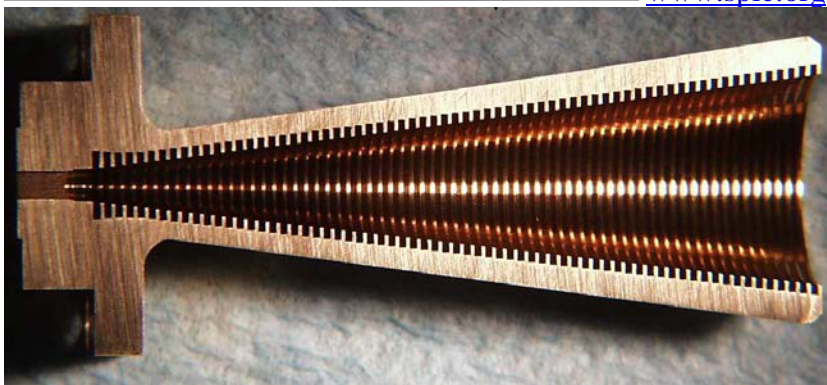


[TMC Design Corp.]

(e) quad-ridge horns (TEM, dual linear polarization allowing for many types of polarization depending on feed , ultra-wide band)



www.spie.org



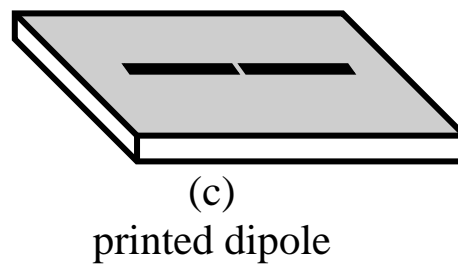
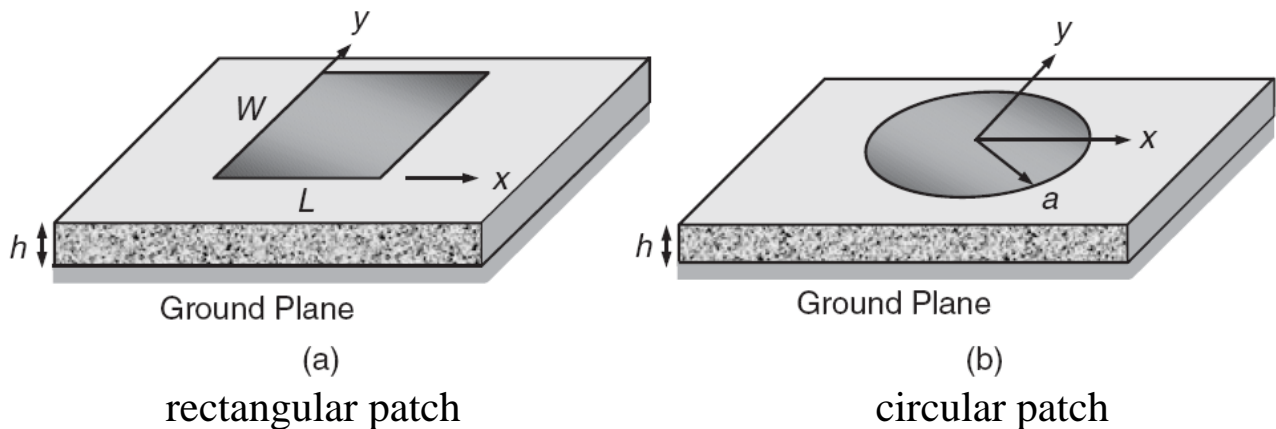
[ZAX Millimeter Wave Corp.]

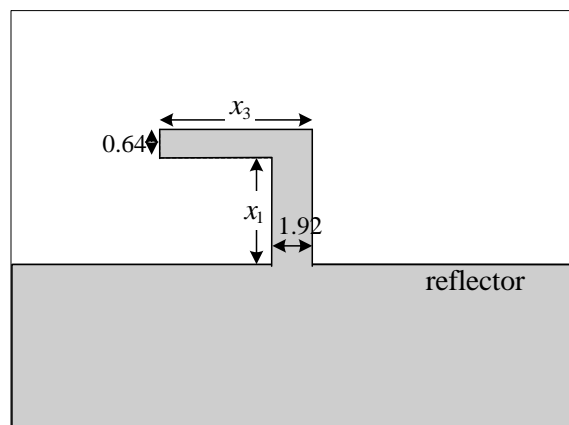
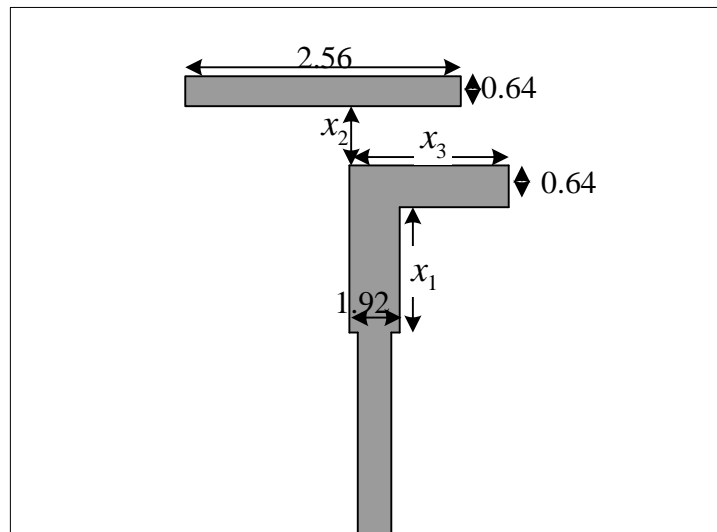
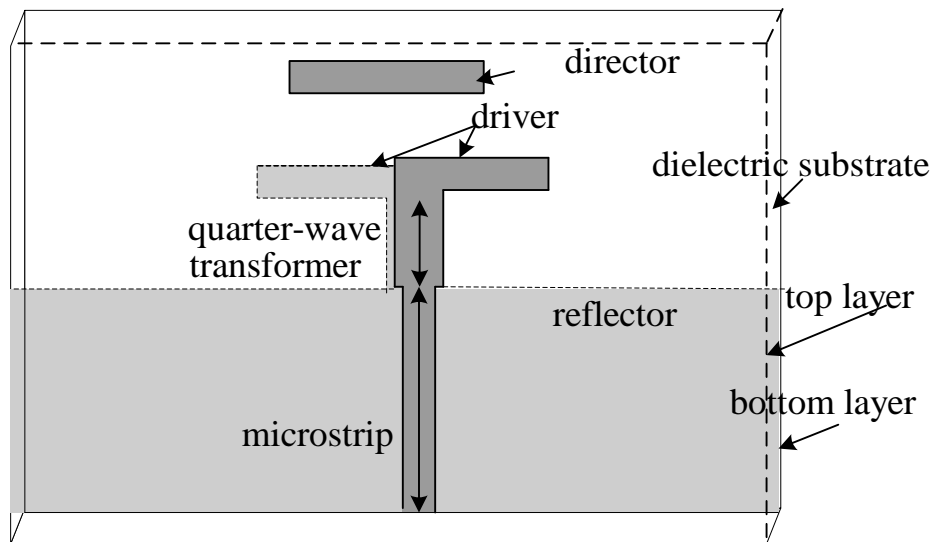
(f) corrugated horns (symmetric patterns, low side lobes, low cross-polarization), often used as primary feeds in reflector antennas

C. Printed antennas

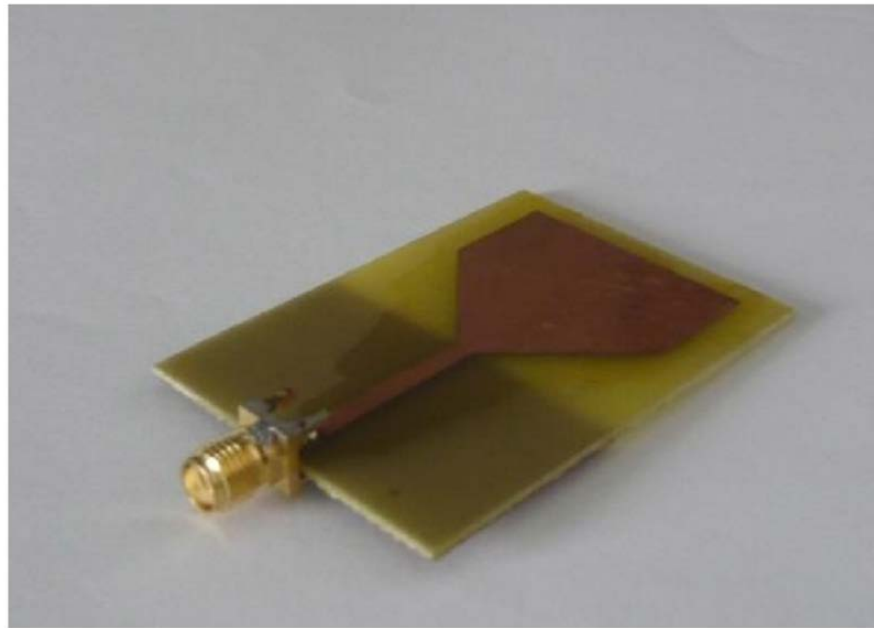
The patch antennas consist of a metallic patch etched on a dielectric substrate, which has a grounded metallic plane at the opposite side. They are developed in the beginning of 1970s. There is a great variety of geometries and ways of excitation. Modern integrated antennas often use multi-layer designs with a feed coupled to the radiator electromagnetically (no galvanic contact).

PRINTED PATCH RADIATORS



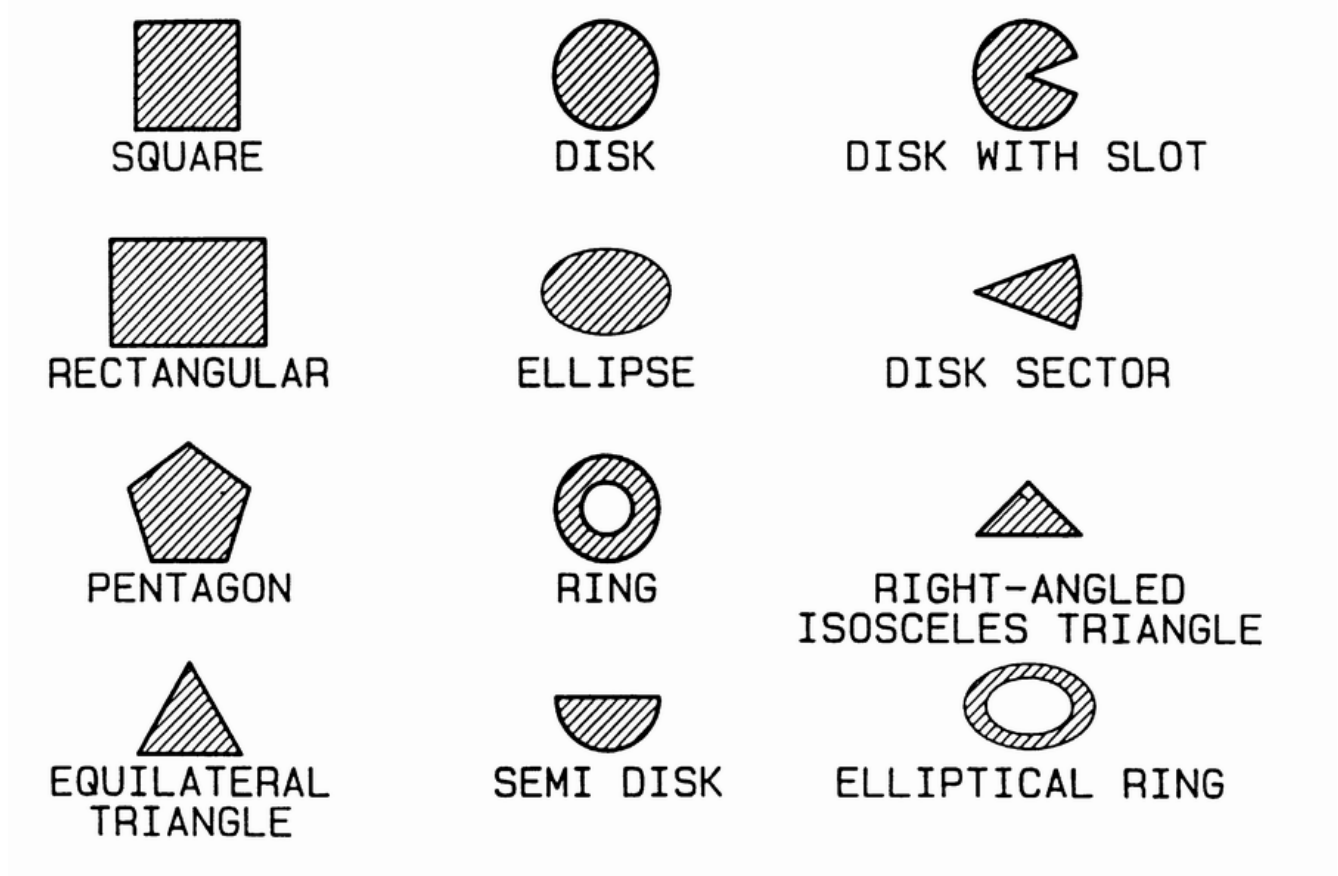


(d)
double-layer printed Yagi with microstrip feed

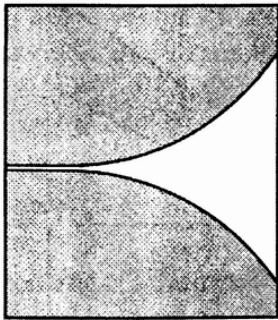


(e)
printed monopole antenna

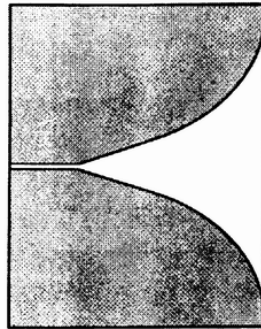
Various shapes used to form a radiating patch:



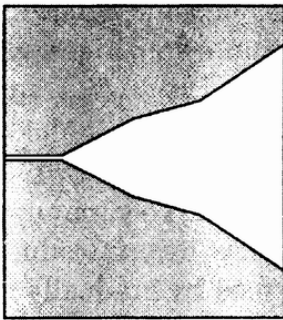
PRINTED SLOT RADIATORS



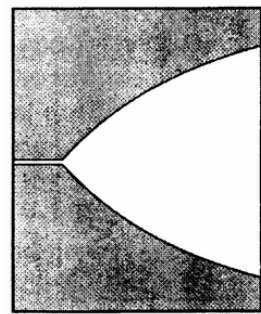
(a)



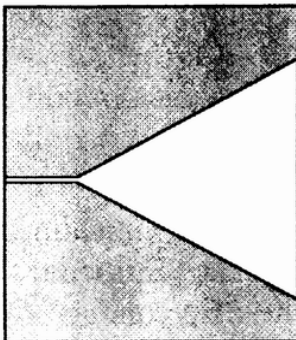
(b)



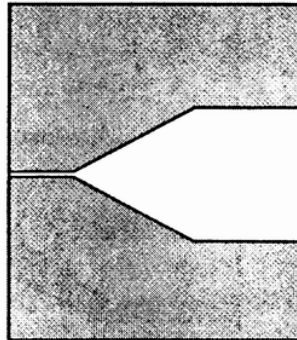
(c)



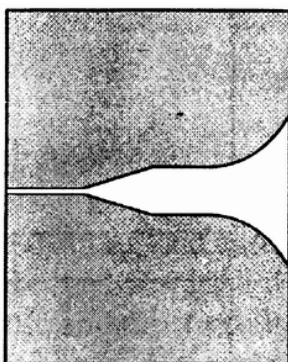
(d)



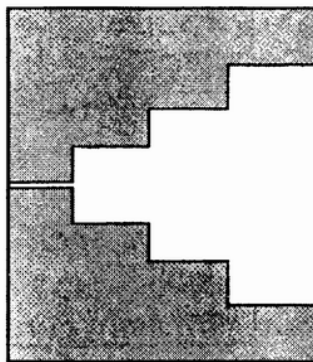
(e)



(f)



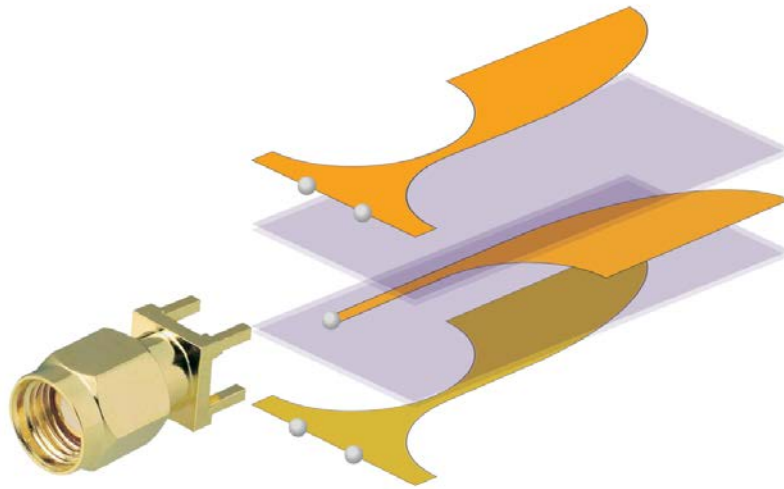
(g)



(h)

Slot antennas were developed in the 1980s and there is still research on new shapes and types of excitation. They are suited for integration with slot-line circuits, which are usually designed to operate at frequencies above 10 GHz. Popular slot antenna in the microwave range is the Vivaldi slot (see a).

Patch and slot antennas share some common features. They are easy and cheap to fabricate. They are easy to mount; they are light and mechanically robust. They have low cross-polarization radiation. Their directivity is not very high. They have relatively high conducting and dielectric losses. These radiators are widely used in patch/slot arrays, which are esp. convenient for use in spacecraft, satellites, missiles, cars and other mobile applications.



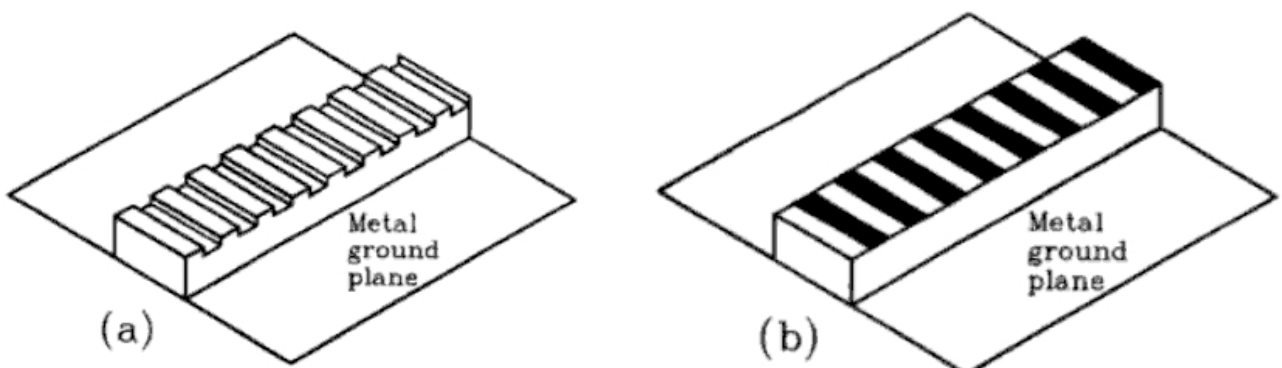
<http://www.radartutorial.eu/06.antennas/Tapered%20Slot%20Antenna.en.html>

(i)

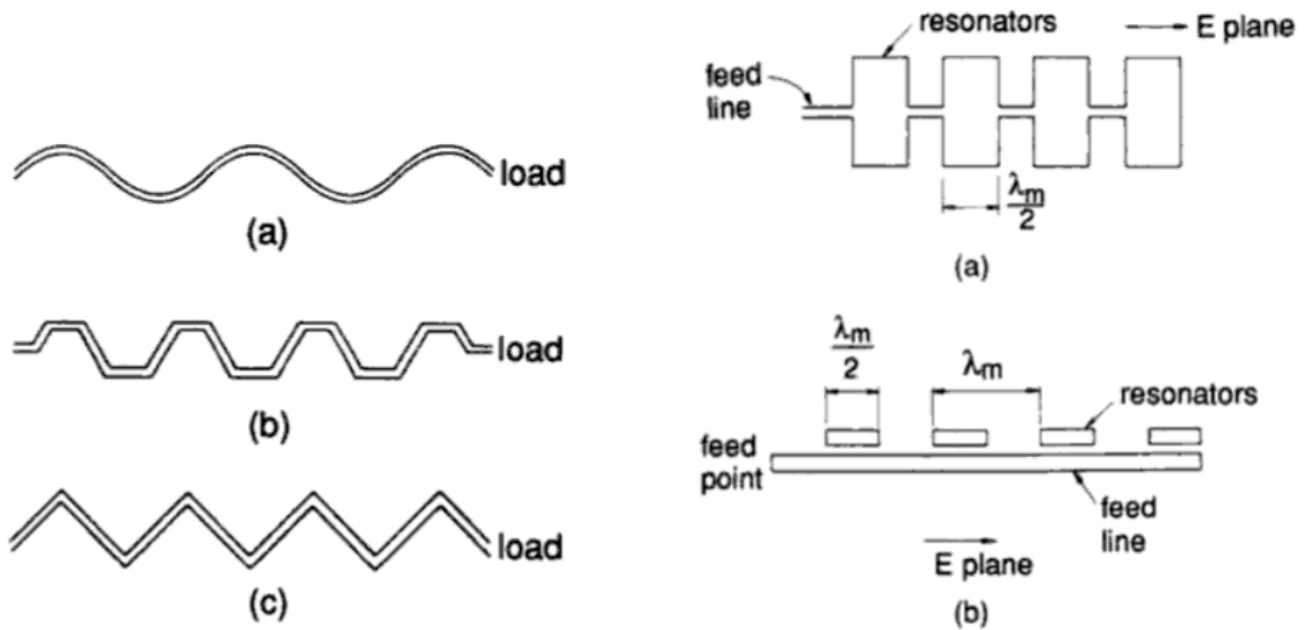
UWB printed tapered slot (Vivaldi) antenna

D. Leaky-wave antennas

These are antennas derived from millimeter-wave (mm-wave) guides, such as dielectric guides, microstrip lines, coplanar and slot lines. They are developed for applications at frequencies above 30 GHz, infrared frequencies included. Periodical discontinuities are introduced at the end of the guide that lead to substantial radiation leakage (radiation from the dielectric surface). These are traveling-wave antennas.



Dielectric-image guides with gratings



Printed leaky-wave antennas

The antennas in the mm-wave band are of big variety and are still the subject of intensive study.

E. Reflector antennas

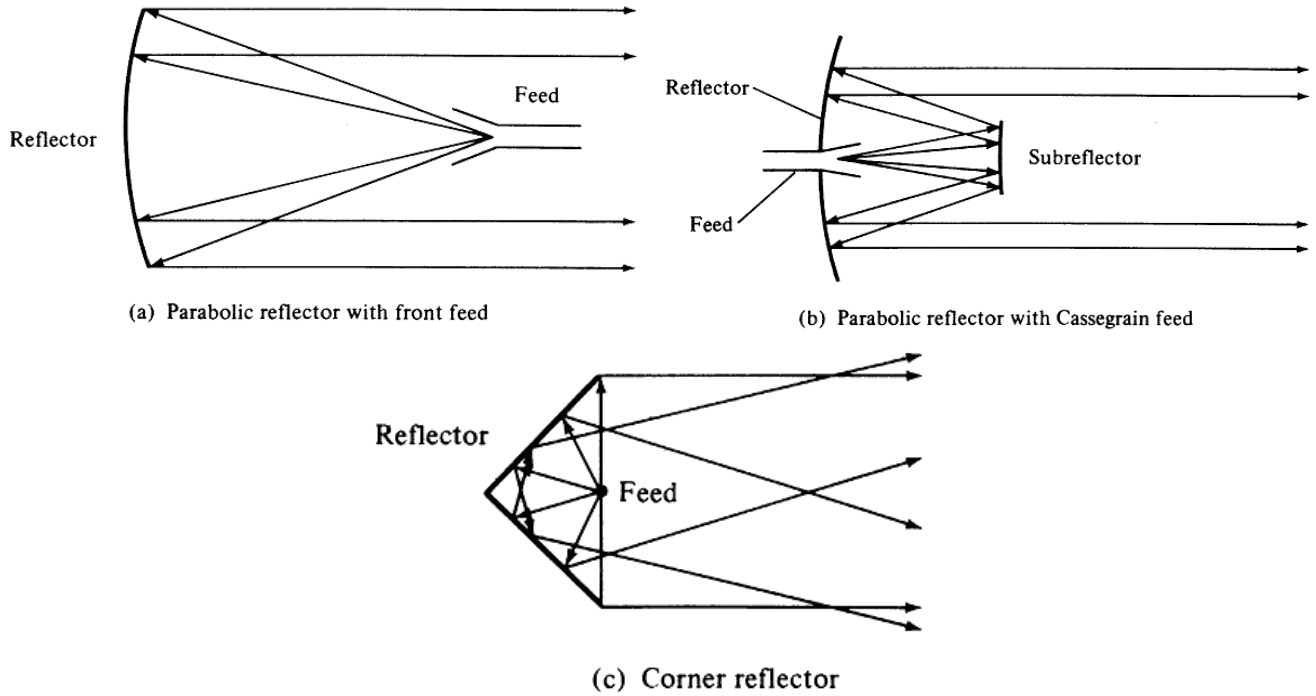
A reflector is used to concentrate the EM energy in a focal point where the receiver or the feed is located. Optical astronomers have long known that a mirror shaped as a parabolic cylinder transforms rays from a line source on its focal line into a bundle of parallel rays. Reflectors are usually parabolic. A parabolic-cylinder reflector was first used for radio waves by Heinrich Hertz in 1888. Sometimes, corner reflectors are used. Reflector antennas have very high gain and directivity. Typical applications include radio telescopes, satellite communications. These antennas are electrically large with their size being on the order of hundreds and thousands of wavelengths. They are not easy to fabricate and in their conventional technology they are rather heavy. It is difficult to make them mechanically robust.

The largest radio telescopes:

- Max Plank Institut für Radioastronomie radio telescope, Effelsberg (Germany), 100-m paraboloidal reflector
- National Astronomy and Ionosphere Center (USA) radio telescope in Arecibo (Puerto Rico), 1000-ft (304.8-m) spherical reflector

- The Green Bank Telescope (the National Radio Astronomy Observatory) – paraboloid of aperture 100 m

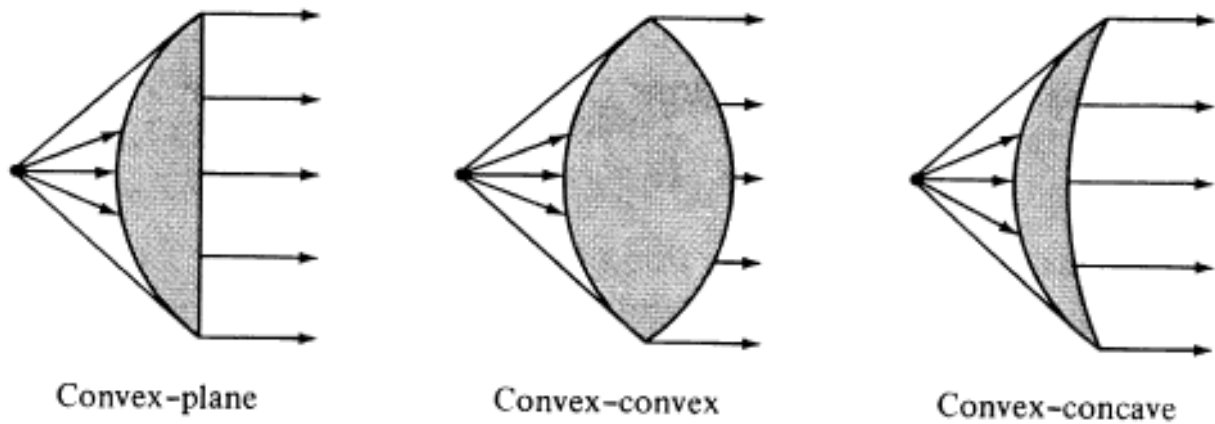
TYPICAL REFLECTORS



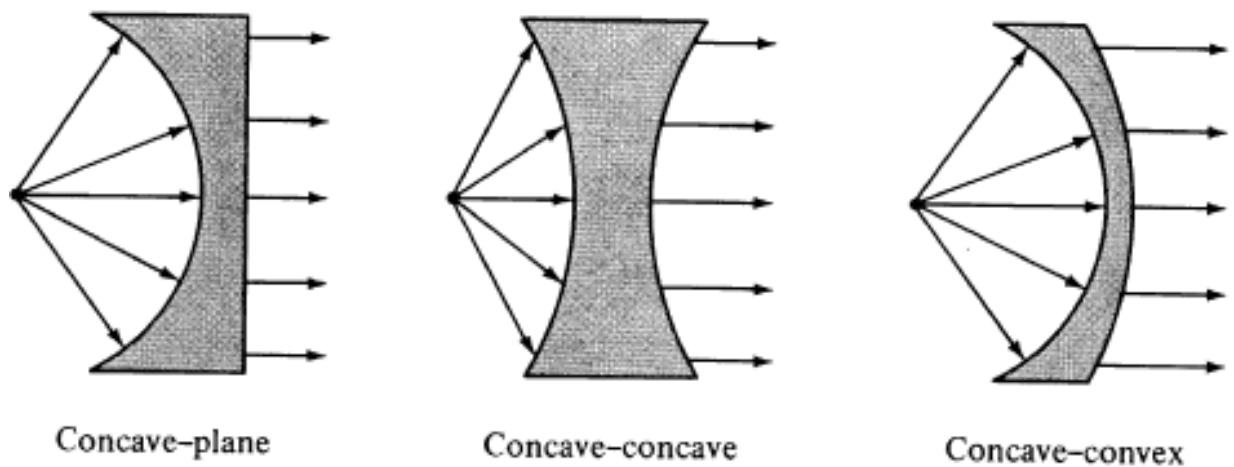
The Radio Telescope of the Arecibo Observatory

F. Lens antennas

Lenses play a similar role to that of reflectors in reflector antennas. They collimate divergent energy into a plane EM wave. Lenses are often preferred to reflectors at higher frequencies ($f > 100$ GHz). They are classified according to their shape and the material they are made of.



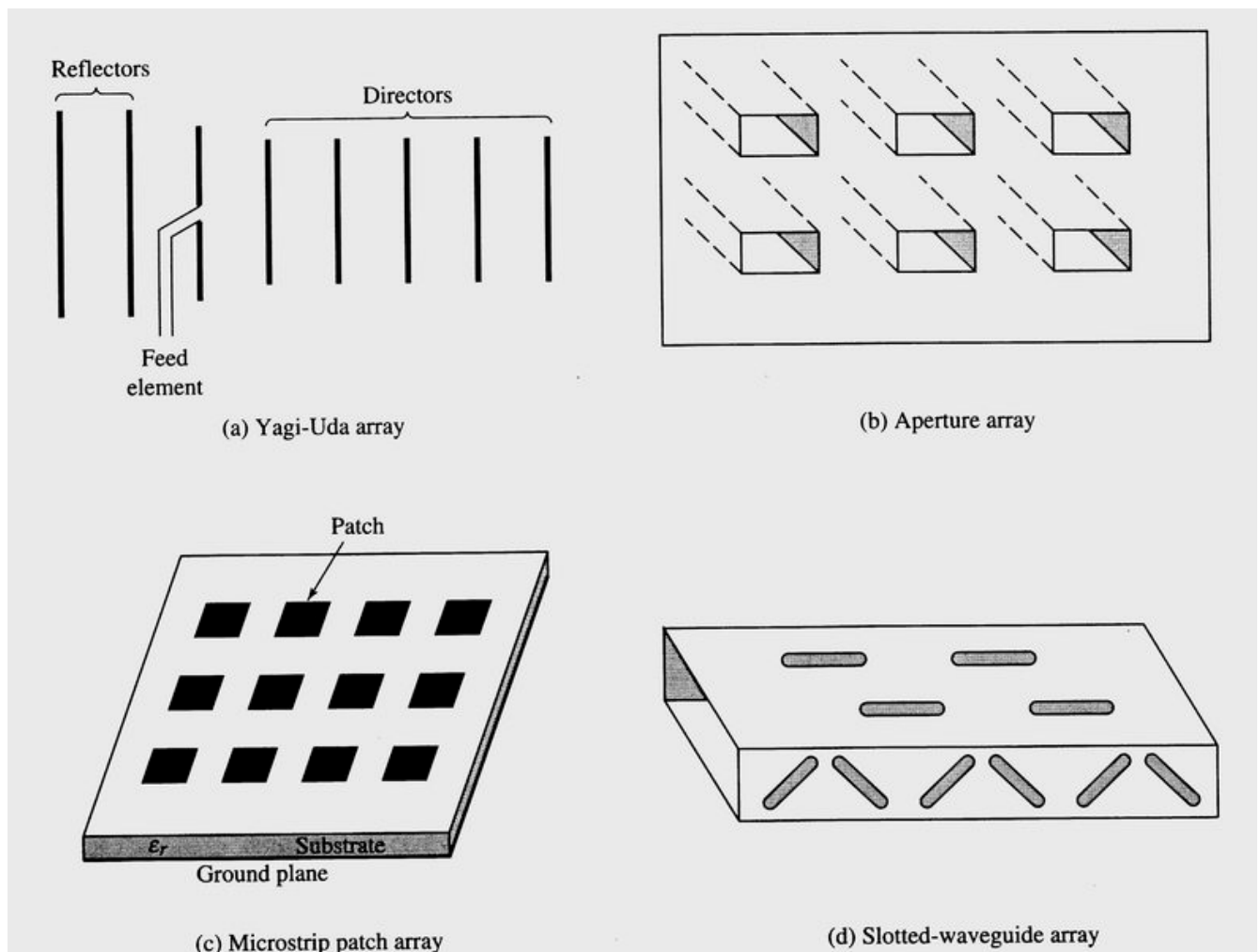
(a) Lens antennas with index of refraction $n > 1$



(b) Lens antennas with index of refraction $n < 1$

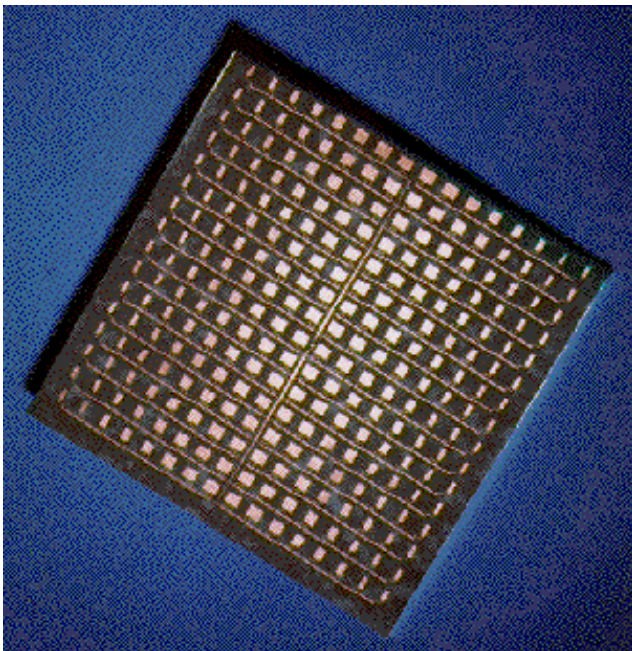
3.2. Antenna arrays

Antenna arrays consist of multiple (usually identical) radiating elements. Arranging the radiating elements in arrays allows for achieving unique radiation characteristics, which cannot be obtained through a single element. The careful choice and control of the phase shift and the amplitude of the signal fed to each element allows for the electronic control of the radiation pattern, i.e., for electronic scanning. Such arrays are called *phased arrays*. The design and the analysis of antenna arrays is a subject of its own and is also related to signal processing and communication theory. Research is ongoing in the subjects of smart antennas, MIMO antennas, tracking antennas, etc. Some commonly met arrays are shown in the figure below.

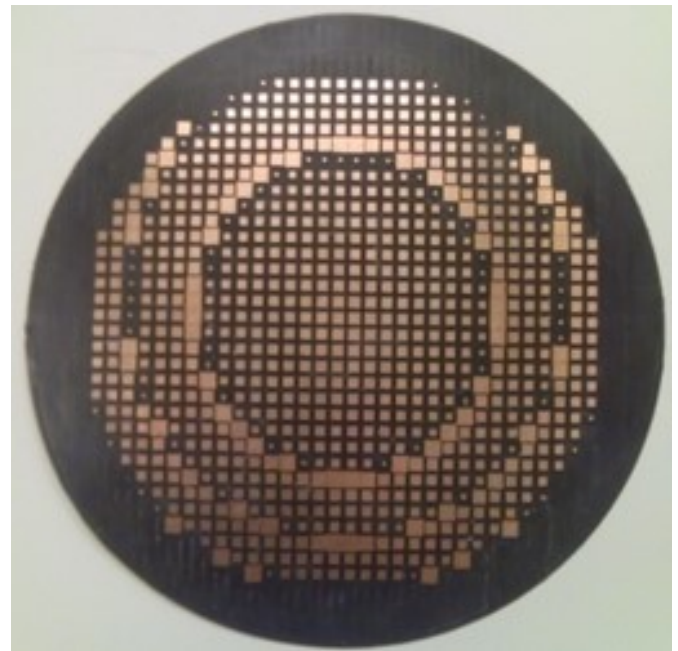




NRAO/ALMA (Atacama Large Millimeter Array): array of radio telescopes



Array of Microstrip Patches



Reflectarray of Printed Elements

4. Wireless vs. cable communication systems

There are two broad categories of communication systems: those that utilize transmission lines as interconnections (*cable or wire systems*), and those that use EM radiation with an antenna at both the transmitting and the receiving end (*wireless systems*).

In areas of high density population, the cable systems are economically preferable, especially when broadband communication is in place. Even for narrow-band communication, such as voice telephony and low-data-rate digital transmission, it is much simpler and cheaper to build wire networks with twisted-pair cables, when many users are to be interconnected. Such lines introduce very little attenuation at low frequencies, e.g., at about 10 kHz the loss is 2-3 dB/km. At higher frequencies, however, the losses increase and so does the signal dispersion. At 10 MHz, a twisted-pair cable has a typical loss value of 7 dB per 100 meters.

At high-frequency carriers for broadband signals (TV transmission and high-data-rate digital transmission), coaxial cables are commonly used. At 1 GHz, the loss of a typical high-quality coaxial cable is around 2 dB per 100 meters (power decreases about 1.6 times). In the USA, the cable loss is rated in dB per 100 feet, so a good coaxial cable has about 0.6 dB/100ft loss.

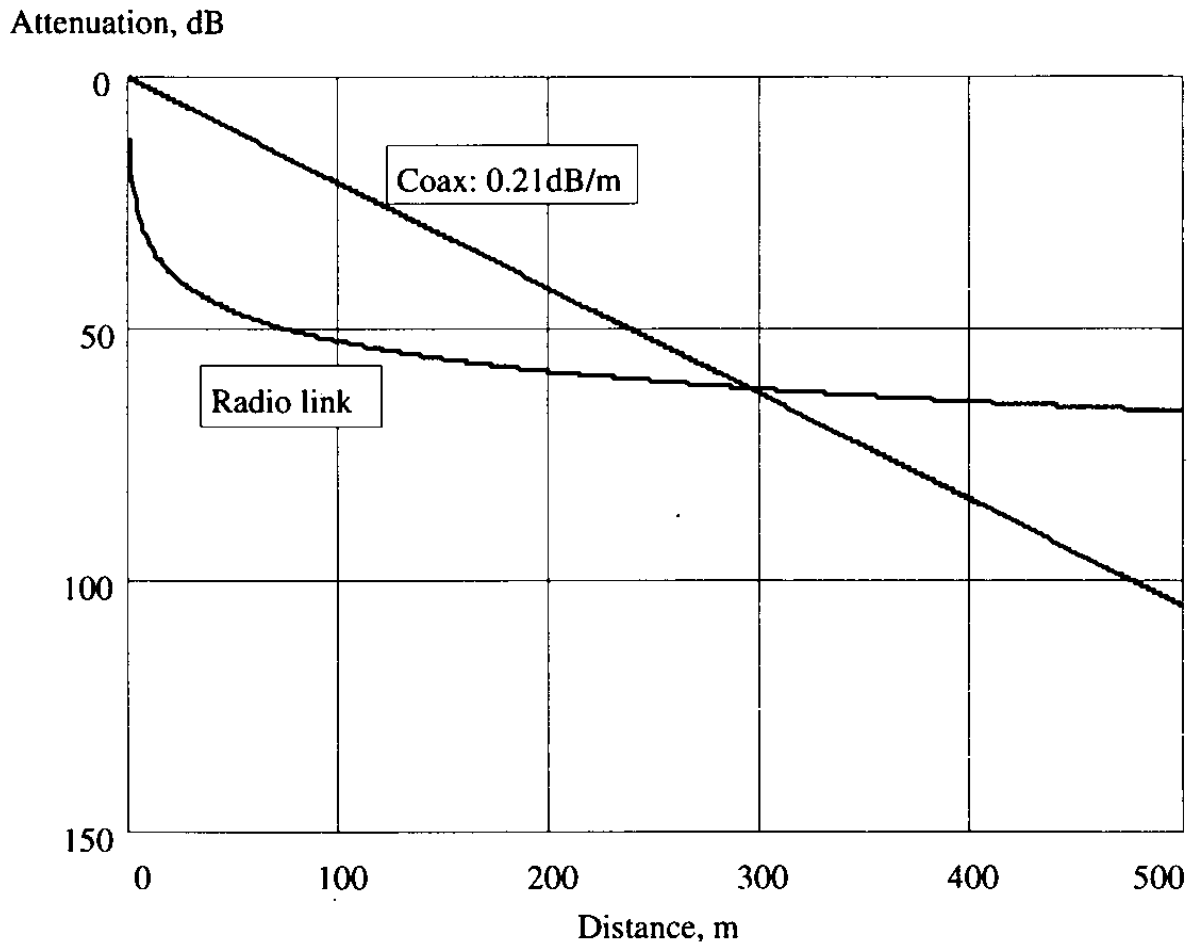
The least distortion and losses are offered by the optical-fiber transmission lines, which operate at three different wavelengths: 850 nm ($\cong 2.3$ dB/km), 1300 nm ($\cong 0.25$ dB/km) and 1550 nm ($\cong 0.25$ dB/km). Optical fibers are relatively expensive and the respective transmitting/receiving equipment is also costly.

Transmission lines provide a measure of security and noise-suppression (coaxial, optical-fiber), but they are not the best option in many cases (long-haul transmission, wide spread over large areas).

A fundamental feature of all transmission lines is the exponential increase of the lost (dissipated) power. Thus, if the loss is 5 dB/km, then a 20-km line will have 100 dB power loss (input power is reduced by a factor of 10^{-10}), a 40-km line will have a 200 dB power loss. This makes it obvious why wireless systems are preferred for long-range communications and in scarcely populated areas.

In most wireless channels, the radiated power per unit area decreases as the inverse square of the distance r between the transmitting and the receiving point. Doubling the distance r would decrease the received power by a factor

of 4 (or 6 dB are added to the loss). Thus, if a particular system has a 100 dB loss at $r = 20$ km, doubling the distance will result in 106 dB loss (as compared to 200 dB loss in a cable system). The comparison between the coax-line losses and free-space attenuation at $f=100$ MHz is given in the figure below.



(Fig. 33 in Siwiak, *Radiowave Propagation and Antennas for Personal Communications*)

Modern personal mobile communications services

- cordless telephony
- cellular telephony
- mobile voice and data (3G and 4G PCS)
- computer network communications: WLANs and Bluetooth
- Wi-Fi and WiMAX networks
- personal satellite communications
- global positioning and navigation systems
- body-centric communication systems (bio-telemetry and bio-sensing)

Besides, there is a variety of special application of wireless technology in

- radar systems (navigation, collision, guidance, defense, missile, etc.)
 - remote-control vehicles (RCV), unmanned aerial vehicle (UAV, aka *drones*)
 - microwave relay links and repeaters
 - satellite systems (TV, telephony, military)
 - radio astronomy
 - biomedical engineering (imaging, hyperthermia)
 - RF identification (RFID)
 - animal (migration) tracking
- etc.

5. The radio-frequency spectrum

Table 1.1: General designation of frequency bands

Frequency band	EM wavelength	Designation	Services
3-30 kHz	100-10 km	Very Low Frequency (VLF)	Navigation, sonar*, submarine
30-300 kHz	10-1 km	Low Frequency (LF)	Radio beacons, navigation
300-3000 kHz	1000-100 m	Medium Frequency (MF)	AM broadcast, maritime/ coast-guard radio
3-30 MHz	100-10 m	High Frequency (HF)	Telephone, telegraph, fax; amateur radio, ship-to-coast and ship-to-aircraft communication
30-300 MHz	10-1 m	Very High Frequency (VHF)	TV, FM broadcast, air traffic control, police, taxicab mobile radio
300-3000 MHz	100-10 cm	Ultrahigh Frequency (UHF)	TV, satellite, radiosonde, radar, cellular (GSM, PCS)
3-30 GHz	10-1 cm	Super high Frequency (SHF)	Airborne radar, microwave links, satellite, land mobile communication
30-300 GHz	10-1 mm	Extremely High Frequency (EHF)	Radar, experimental

Table 2.1: Microwave-band designation

Frequency	Old	New
500-1000 MHz	VHF	C
1-2 GHz	L	D
2-3 GHz	S	E
3-4 GHz	S	F
4-6 GHz	C	G
6-8 GHz	C	H
8-10 GHz	X	I
10-12.4 GHz	X	J
12.4-18 GHz	Ku	J
18-20 GHz	K	J
20-26.5 GHz	K	K
26.5-40 GHz	Ka	K

* Sonar (an acronym for Sound, Navigation and Ranging) is a system for underwater detection and location of objects by acoustical echo. The first sonars, invented during World War I by British, American and French scientists, were used to locate submarines and icebergs. Sonar is an American term dating from World War II.

LECTURE 2: Introduction into the Theory of Radiation

(Maxwell's equations – revision. Power density and Poynting vector – revision. Radiated power – definition. Basic principle of radiation. Vector and scalar potentials – revision. Far fields and vector potentials.)

1. Maxwell's Equations – Revision

(a) the law of induction (Faraday's law):

$$-\nabla \times \mathbf{E} = \frac{\partial \mathbf{B}}{\partial t} + \mathbf{M}^* \quad (2.1)$$

$$\oint_c \mathbf{E} \cdot d\mathbf{c} = -\frac{\partial}{\partial t} \iint_{S_{[c]}} \mathbf{B} \cdot d\mathbf{s} \Leftrightarrow e = -\frac{\partial \Psi}{\partial t} \quad (2.1-i)$$

\mathbf{E} (V/m)	electric field (electric field intensity)
\mathbf{B} (T=Wb/m ²)	magnetic flux density
\mathbf{M} (V/m ²)	magnetic current density*
Ψ (Wb=V·s)	magnetic flux
e (V)	electromotive force

(b) Ampere's law, generalized by Maxwell to include the displacement current $\partial \mathbf{D} / \partial t$:

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \mathbf{J} \quad (2.2)$$

$$\oint_c \mathbf{H} \cdot d\mathbf{c} = \iint_{S_{[c]}} \left(\frac{\partial \mathbf{D}}{\partial t} + \mathbf{J} \right) \cdot d\mathbf{s} \Leftrightarrow I = \oint_c \mathbf{H} \cdot d\mathbf{c} \quad (2.2-i)$$

\mathbf{H} (A/m)	magnetic field (magnetic field intensity)
\mathbf{D} (C/m ²)	electric flux density (electric displacement)
\mathbf{J} (A/m ²)	electric current density
I (A)	electric current

* \mathbf{M} is a fictitious quantity, which renders Maxwell's equations symmetrical and which proves a useful mathematical tool when solving EM boundary value problems applying equivalence theorem.

(c) Gauss' electric law:

$$\nabla \cdot \mathbf{D} = \rho \quad (2.3)$$

$$\oiint_S \mathbf{D} \cdot d\mathbf{s} = \iiint_{V_{[S]}} \rho dv = Q \quad (2.3-i)$$

ρ (C/m³) electric charge density
 Q (C) electric charge

Equation (2.3) follows from equation (2.2) and the continuity relation:

$$\nabla \cdot \mathbf{J} = -\frac{\partial \rho}{\partial t}. \quad (2.4)$$

Hint: Take the divergence of both sides of (2.2).

(d) Gauss' magnetic law:

$$\nabla \cdot \mathbf{B} = \rho_m^{**} \quad (2.5)$$

The equation $\nabla \cdot \mathbf{B} = 0$ follows from equation (2.1), provided that $\mathbf{M} = 0$.

Maxwell's equations alone are insufficient to solve for the four vector quantities: \mathbf{E} , \mathbf{D} , \mathbf{H} , and \mathbf{B} (twelve scalar quantities). Two additional vector equations are needed.

(e) Constitutive relationships

The constitutive relationships describe the properties of matter with respect to electric and magnetic forces.

$$\mathbf{D} = \bar{\epsilon} \cdot \mathbf{E} \quad (2.6)$$

$$\mathbf{B} = \bar{\mu} \cdot \mathbf{H}. \quad (2.7)$$

In an anisotropic medium, the dielectric permittivity and the magnetic permeability are *tensors*. In vacuum, which is isotropic, the permittivity and the permeability are constants (or tensors whose *diagonal elements only* are non-zero and are the same): $\epsilon_0 \approx 8.854187817 \times 10^{-12}$ F/m, $\mu_0 = 4\pi \times 10^{-7}$ H/m. In an isotropic medium, \mathbf{D} and \mathbf{E} are collinear, and so are \mathbf{B} and \mathbf{H} .

The dielectric properties relate to the electric field (electric force). Dielectric materials with relative permittivity $\epsilon_r > 1$ are built of

** ρ_m is a fictitious quantity introduced via the continuity relation $\nabla \cdot \mathbf{M} = -\partial \rho_m / \partial t$. As per experimental evidence, $\nabla \cdot \mathbf{B} = 0$.

atomic/molecular sub-domains, which have the properties of dipoles. In an external electric field, the dipoles tend to orient in such a way that their own fields have a cancellation effect on the external field. The electric force $\mathbf{F}_e = Q\mathbf{E}$ exerted on a test point charge Q_t from a source Q_s in such medium is ϵ_r times weaker than the electric force of the same source in vacuum.

On the contrary, magnetic materials with relative permeability $\mu_r > 1$ are made of sub-domains, which tend to orient in the external magnetic field in such a way that their own magnetic fields align with the external field. The magnetic force $\mathbf{F}_m = Q_t \mathbf{v} \times \mathbf{B}$ exerted on a moving (with velocity \mathbf{v}) test point charge Q_t in such a medium is μ_r times stronger than the force that this same source (e.g. electric currents) would create in vacuum.

We are mostly concerned with isotropic media, i.e., media where the equations $\mathbf{B} = \mu_0 \mu_r \mathbf{H}$ and $\mathbf{D} = \epsilon_0 \epsilon_r \mathbf{E}$ hold.

(f) Time-harmonic field analysis

In harmonic analysis of EM fields, the field phasors are introduced:

$$\begin{aligned} \mathbf{e}(x, y, z, t) &= \text{Re} \left\{ \mathbf{E}(x, y, z) e^{j\omega t} \right\} \\ \mathbf{h}(x, y, z, t) &= \text{Re} \left\{ \mathbf{H}(x, y, z) e^{j\omega t} \right\}. \end{aligned} \quad (2.8)$$

For example, the phasor of $e(x, y, z, t) = E_m(x, y, z) \cos(\omega t + \varphi_E)$ is $E(x, y, z) = E_m e^{j\varphi_E}$. For clarity, from this point on, we will denote time-dependent field quantities with lower-case letters (bold for vectors), while their phasors will be denoted with upper-case letters. Complex-conjugate phasors will be denoted with an asterisk $*$.

The frequency-domain Maxwell equations are obtained from the time-dependent equations using the following correspondences:

$$\begin{aligned} f(x, y, z, t) &\doteq F(x, y, z) \\ \frac{\partial f(x, y, z, t)}{\partial t} &\doteq j\omega F(x, y, z) \\ \frac{\partial f}{\partial \xi} &\doteq \frac{\partial F}{\partial \xi}, \quad \xi = x, y, z. \end{aligned}$$

Thus, Maxwell's equations in phasor form are:

$$\nabla \times \mathbf{H} = j\omega \bar{\epsilon} \mathbf{E} + \mathbf{J}, \quad \bar{\epsilon} = \epsilon' - j(\epsilon'' + \sigma / \omega) \quad (2.9)$$

$$-\nabla \times \mathbf{E} = j\omega \bar{\mu} \mathbf{H} + \mathbf{M}, \quad \bar{\mu} = \mu' - j\mu'' \quad (2.10)$$

These equations include the equivalent (fictitious) magnetic currents \mathbf{M} . The imaginary part of the *complex dielectric permittivity* $\bar{\epsilon}$ describes loss. Often, the dielectric loss is represented by the dielectric loss angle δ_d :

$$\bar{\epsilon} = \epsilon' \left[1 - j \left(\frac{\epsilon''}{\epsilon'} + \frac{\sigma}{\omega \epsilon'} \right) \right] = \epsilon' \left[1 - j \left(\tan \delta_d + \frac{\sigma}{\omega \epsilon'} \right) \right]. \quad (2.11)$$

Similarly, the magnetic loss is described by the imaginary part of the *complex magnetic permeability* $\bar{\mu}$ or by the magnetic loss angle δ_m :

$$\bar{\mu} = \mu' - j\mu'' = \mu' \left(1 - j \frac{\mu''}{\mu'} \right) = \mu' (1 - j \tan \delta_m). \quad (2.12)$$

In antenna theory, we are mostly concerned with *isotropic, homogeneous* and *loss-free* propagation media.

2. Power Density, Poynting Vector, Radiated Power

2.1. Poynting vector – revision

In the time-domain analysis, the Poynting vector is defined as

$$\mathbf{p}(t) = \mathbf{e}(t) \times \mathbf{h}(t), \text{ W/m}^2. \quad (2.13)$$

As follows from Poynting's theorem, \mathbf{p} is a vector representing the density and the direction of the EM power flow. Thus, the total power leaving certain volume V is obtained as

$$\Pi(t) = \oiint_{S_{[V]}} \mathbf{p}(t) \cdot d\mathbf{s}, \text{ W}. \quad (2.14)$$

Since

$$\mathbf{e}(t) = \text{Re} \left\{ \mathbf{E} e^{j\omega t} \right\} = \frac{1}{2} \left(\mathbf{E} e^{j\omega t} + \mathbf{E}^* e^{-j\omega t} \right), \quad (2.15)$$

and

$$\mathbf{h}(t) = \text{Re} \left\{ \mathbf{H} e^{j\omega t} \right\} = \frac{1}{2} \left(\mathbf{H} e^{j\omega t} + \mathbf{H}^* e^{-j\omega t} \right), \quad (2.16)$$

the instantaneous power density appears as

$$\mathbf{p}(t) = \underbrace{\frac{1}{2} \text{Re} \left\{ \mathbf{E} \times \mathbf{H}^* \right\}}_{\mathbf{p}_{av}} + \frac{1}{2} \text{Re} \left\{ \mathbf{E} \times \mathbf{H} \cdot e^{2j\omega t} \right\}. \quad (2.17)$$

The first term in (2.17), \mathbf{p}_{av} , has no time dependence. It is the average value, about which the power flux density fluctuates. It is a vector of unchanging direction showing a constant outflow (positive value) or inflow

(negative value) of EM power. It describes the active (or time-average) power flow,

$$\Pi_{av} = \oiint_{S_{[V]}} \mathbf{p}_{av} \cdot d\mathbf{s}. \quad (2.18)$$

The second term in (2.17) is a vector changing its direction with a double frequency 2ω . It describes power flow, which fluctuates in space (propagates to and fro) without contribution to the overall transport of energy. If there is no phase difference between \mathbf{E} and \mathbf{H} , $\mathbf{p}(t)$ always maintains the same direction (the direction of the **outgoing** wave) although it changes in intensity. This is because the 2nd term in (2.17) never exceeds in magnitude the first term, i.e., \mathbf{p}_{av} . This indicates that the power moves away from the source at every instant of time, with the Poynting vector never directed toward the source.

However, if \mathbf{E} and \mathbf{H} are out of phase ($\Delta\varphi = \varphi_H - \varphi_E \neq 0$), there are time periods during which the Poynting vector does reverse its direction toward the source and when doing so it achieves a maximum value of $0.5 \cdot \text{Im}\{\mathbf{E} \times \mathbf{H}^*\}$.

In fact, the time-dependent Poynting vector can be decomposed into two parts: (i) a real-positive (active) part fluctuating with double frequency about the average value $\mathbf{p}_{av} = 0.5 \cdot \text{Re}\{\mathbf{E} \times \mathbf{H}^*\}$, i.e., swinging between zero and $E_m H_m \cos \Delta\varphi$, and (ii) a double-frequency part of magnitude $0.5 \cdot \text{Im}\{\mathbf{E} \times \mathbf{H}^*\}$, which becomes negative every other quarter-period (reactive). You will prove and illustrate this in your next Assignment.

Definition: The complex Poynting vector is the vector

$$\mathbf{P} = \frac{1}{2} \mathbf{E} \times \mathbf{H}^*, \quad (2.19)$$

whose real part is equal to the average power flux density: $\mathbf{p}_{av} = \text{Re} \mathbf{P}$.

2.2. Radiated power

Definition: Radiated power is the average power radiated by the antenna:

$$\Pi_{rad} = \oiint_{S_{[V]}} \mathbf{p}_{av} \cdot d\mathbf{s} = \oiint_{S_{[V]}} \text{Re} \mathbf{P} \cdot d\mathbf{s} = \frac{1}{2} \oiint_{S_{[V]}} \text{Re}\{\mathbf{E} \times \mathbf{H}^*\} \cdot d\mathbf{s}. \quad (2.20)$$

3. Basic Principle of Radiation

Radiation is produced by accelerated or decelerated charge (time-varying current element).



3.1. Current element

Definition: A current element ($I\Delta l$), $A \times m$, is a filament of length Δl carrying current I . It is sufficiently small to imply constant magnitude of the current along Δl .

The time-varying current element is the elementary source of EM radiation. It has fundamental significance in radiation theory similar to the fundamental concept of a point charge in electrostatics. The field radiated by a complex antenna in a linear medium can be analyzed using the superposition principle after decomposing the antenna into elementary sources, i.e., current elements.

The time-dependent current density vector \mathbf{j} depends on the charge density ρ and its velocity \mathbf{v} as

$$\mathbf{j} = \rho \cdot \mathbf{v}, \quad A / m^2. \quad (2.21)$$

If the current flows along a wire of cross-section ΔS , then the product $\rho_l = \rho \cdot \Delta S$ [C/m] is the charge per unit length (charge line density) along the wire. Thus, for the current $i = \mathbf{j} \cdot \Delta \mathbf{S}$ it follows that

$$i = v \cdot \rho_l, \quad A. \quad (2.22)$$

Then

$$\frac{di}{dt} = \rho_l \frac{dv}{dt} = \rho_l \cdot a, \quad A/s, \quad (2.23)$$

where a (m/s^2) is the acceleration of the charge. The time-derivative of a current element $i\Delta l$ is then proportional to the amount of charge q enclosed in the volume of the current element and to its acceleration:

$$\Delta l \frac{di}{dt} = \Delta l \cdot \rho_l \cdot a = q \cdot a, \quad A \times m/s. \quad (2.24)$$

3.2. Mathematical description of the accelerated charge as a radiation source

It is not immediately obvious from Maxwell's equations that the time-varying current is the source of radiation. A simple transformation of the time-dependent Maxwell equations,

$$\left\{ \begin{array}{l} -\nabla \times \mathbf{e} = \mu \frac{\partial \mathbf{h}}{\partial t} \\ \nabla \times \mathbf{h} = \varepsilon \frac{\partial \mathbf{e}}{\partial t} + \mathbf{j} , \end{array} \right. \quad (2.25)$$

into a single second-order equation either for \mathbf{E} or for the \mathbf{H} field proves this statement. By taking the curl of both sides of the first equation in (2.25) and by making use of the second equation in (2.25), we obtain

$$\nabla \times \nabla \times \mathbf{e} + \mu\varepsilon \frac{\partial^2 \mathbf{e}}{\partial t^2} = -\mu \frac{\partial \mathbf{j}}{\partial t}. \quad (2.26)$$

From (2.26), it is obvious that the time derivative of the electric current is the source for the wave-like vector \mathbf{e} . Time-constant currents do not radiate.

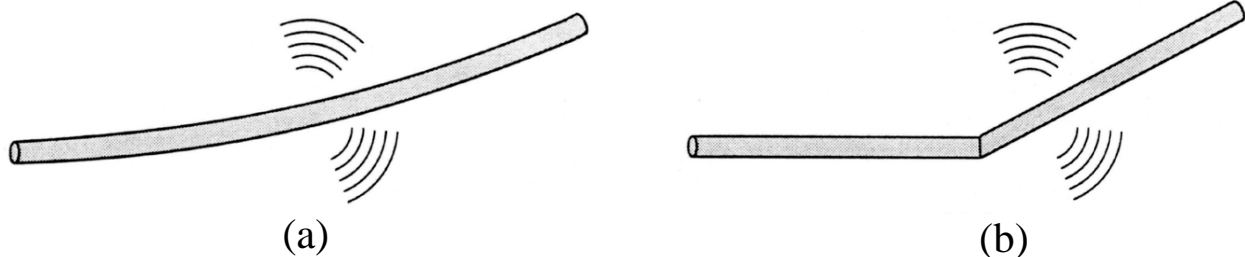
In an analogous way, one can obtain the wave equation for the magnetic field \mathbf{H} and its sources:

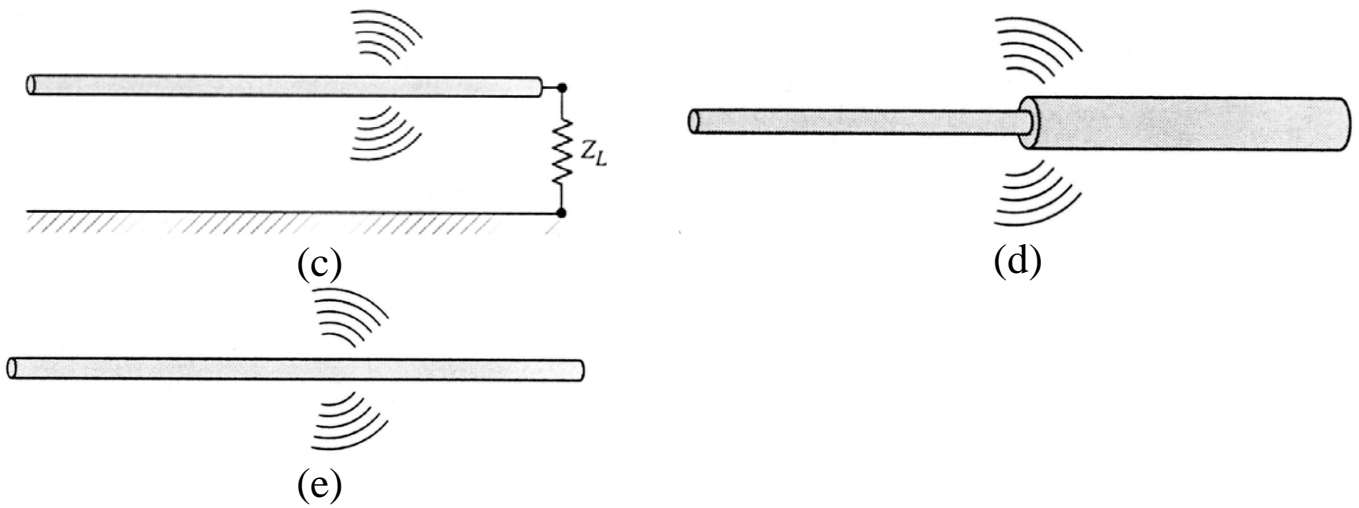
$$\nabla \times \nabla \times \mathbf{h} + \mu\varepsilon \frac{\partial^2 \mathbf{h}}{\partial t^2} = \nabla \times \mathbf{j}. \quad (2.27)$$

Notice that, as follows from (2.27) and (2.25), curl-free currents (e.g., $\mathbf{j} = \nabla \psi$) do not radiate either.

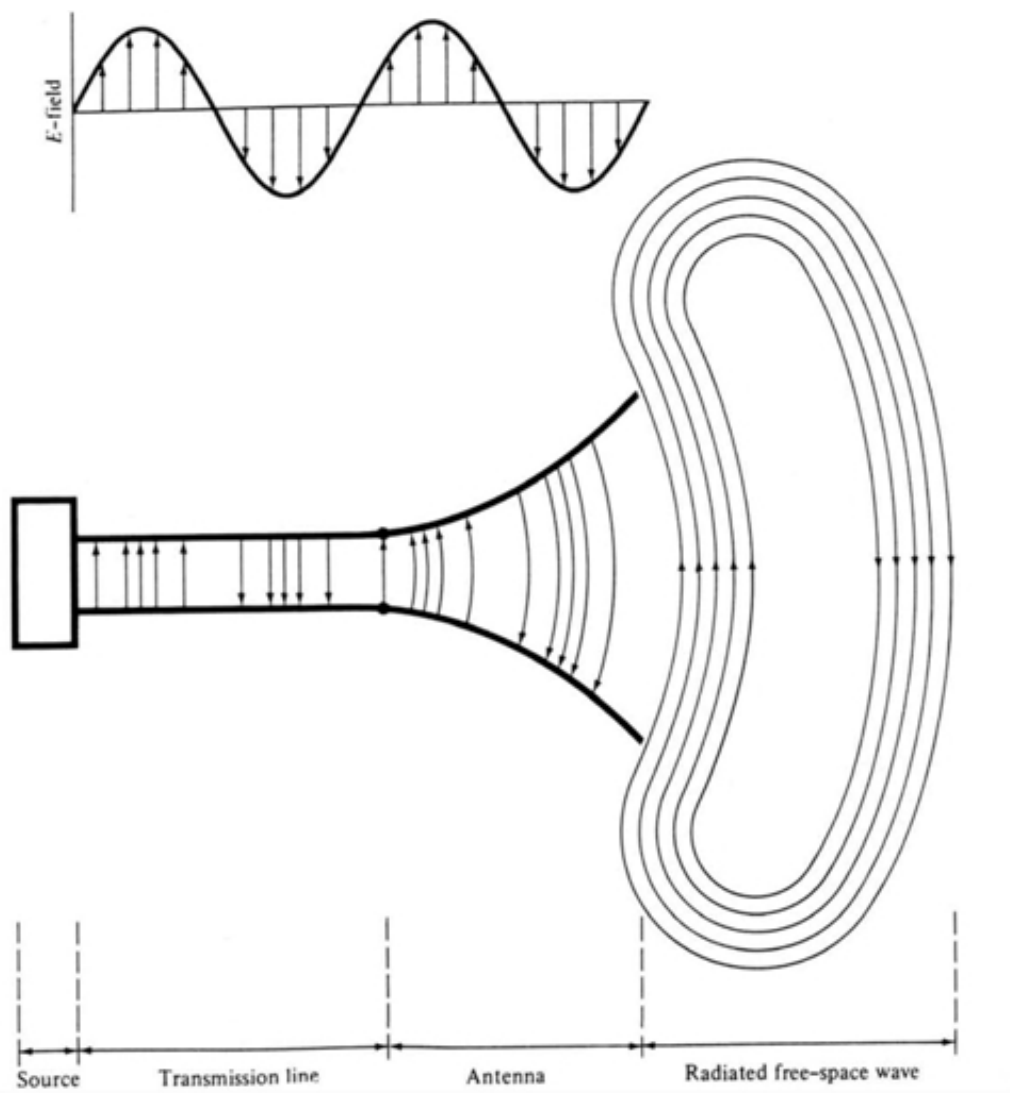
To accelerate/decelerate the charges, one needs sources of electromotive force and/or discontinuities of the medium in which the charges move. Such discontinuities can be bends or open ends of wires, change in the electrical properties of the region, etc. In summary:

- If charge is not moving, current is zero \Rightarrow no radiation.
- If charge is moving with a uniform velocity (DC) \Rightarrow no radiation.
- If charge is accelerated due to electromotive force or due to discontinuities, such as terminations, bends, curvatures \Rightarrow radiation occurs.

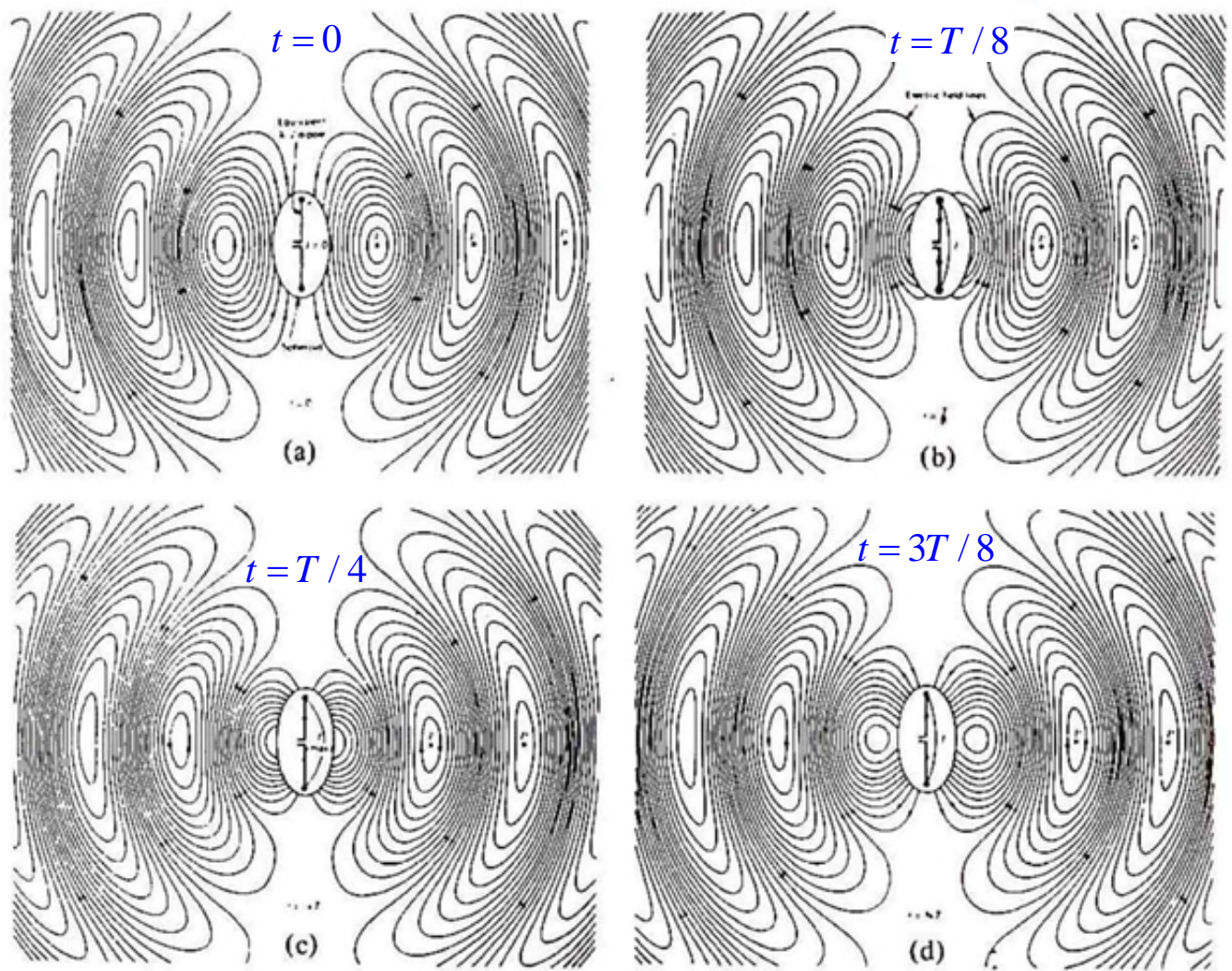




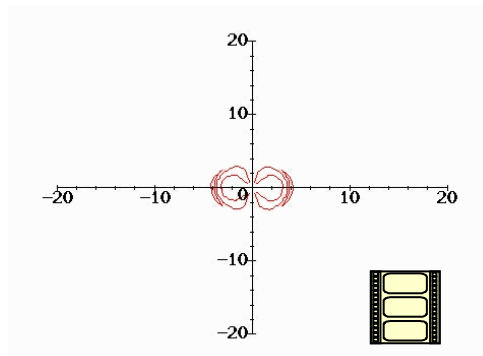
3.1. Intuitive representation of radiation from simple sources



(a) Illustration of the \mathbf{E} -field lines in a transmission (feed) line and at the antenna aperture [Balanis, 3rd ed.]



(b) Snapshots of the \mathbf{E} -field lines around a dipole



[<http://physics.usask.ca/~hirose/ep225/anim.htm>]
 [<http://www.falstad.com/emwave2/fullscreen.html>]

(c) animation of the \mathbf{E} -field lines of a small dipole antenna

4. Radiation Boundary Condition

With few exceptions, antennas are assumed to radiate in open (unbounded) space. This is a crucial factor determining the field behavior. Often, the EM sources (currents and charges on the antenna) are more or less accurately known. These sources are then assumed to radiate in unbounded space and the resulting EM field is determined as integrals over the currents on the antenna. Such problems, where the field sources are known and the resulting field is to be determined are called *analysis (forward, direct)* problems.¹ To ensure the uniqueness of the solution in an unbounded analysis problem, we have to impose the radiation boundary condition (RBC) on the EM field vectors, i.e., for distances far away from the source ($r \rightarrow \infty$),

$$\left| \begin{array}{l} r(\mathbf{E} - \eta \mathbf{H} \times \hat{\mathbf{r}}) \rightarrow 0, \\ r(\mathbf{H} - \frac{1}{\eta} \hat{\mathbf{r}} \times \mathbf{E}) \rightarrow 0 . \end{array} \right. \quad (2.28)$$

The above RBC is known as the *Sommerfeld* vector RBC or the *Silver-Müller* RBC. Here, η is the intrinsic impedance of the medium; $\eta = \sqrt{\mu_0 / \epsilon_0} \approx 377 \Omega$ in vacuum.

The specifics of the antenna problems lead to the introduction of auxiliary vector potential functions, which allow simpler and more compact solutions.

It is customary to perform the EM analysis for the case of time-harmonic fields, i.e., in terms of phasors. This course adheres to this tradition. Therefore, from now on, all field quantities (vectors and scalars) are to be understood as complex phasor quantities, the absolute values of which correspond to the *magnitudes* (not the RMS value!) of the respective sine waves.

5. Vector and Scalar Potentials – Review

In radiation theory, the potential functions are almost exclusively in the form of retarded potentials, i.e., the magnetic vector potential \mathbf{A} and its scalar counterpart Φ form a 4-potential in space-time and they are related through the Lorenz gauge. We next introduce the retarded potentials.

5.1. The magnetic vector potential \mathbf{A}

We first consider only electric sources (\mathbf{J} and ρ , $\nabla \cdot \mathbf{J} = -j\omega\rho$).

¹ The inverse (or design) problem is the problem of finding the sources of a known field.

$$\begin{cases} \nabla \times \mathbf{E} = -j\omega\mu\mathbf{H}, \\ \nabla \times \mathbf{H} = j\omega\varepsilon\mathbf{E} + \mathbf{J}. \end{cases} \quad (2.29)$$

Since $\nabla \cdot \mathbf{B} = 0$, we can assume that

$$\mathbf{B} = \nabla \times \mathbf{A}. \quad (2.30)$$

Substituting (2.30) in (2.29) yields

$$\begin{cases} \mathbf{E} = -j\omega\mathbf{A} - \nabla\Phi, \\ j\omega\varepsilon\mathbf{E} = \nabla \times \left(\frac{1}{\mu} \nabla \times \mathbf{A} \right) - \mathbf{J}. \end{cases} \quad (2.31)$$

From (2.31), a single equation can be written for \mathbf{A} :

$$\nabla \times \nabla \times \mathbf{A} + j\omega\mu\varepsilon(j\omega\mathbf{A} + \nabla\Phi) = \mu\mathbf{J}. \quad (2.32)$$

Here, Φ denotes the electric scalar potential, which plays an essential role in the analysis of electrostatic fields. To uniquely define \mathbf{A} , we need to define not only its curl, but also its divergence. There are no restrictions in defining $\nabla \cdot \mathbf{A}$. Since $\nabla \times \nabla \times = \nabla \nabla \cdot - \nabla^2$, equation (2.32) can be simplified by assuming that

$$\nabla \cdot \mathbf{A} = -j\omega\mu\varepsilon\Phi. \quad (2.33)$$

Equation (2.33) is known as the *Lorenz gauge*. It reduces (2.32) to

$$\nabla^2 \mathbf{A} + \omega^2 \mu\varepsilon \mathbf{A} = -\mu\mathbf{J}. \quad (2.34)$$

If the region is lossless, then μ and ε are real numbers, and (2.34) can be written as

$$\nabla^2 \mathbf{A} + \beta^2 \mathbf{A} = -\mu\mathbf{J}, \quad (2.35)$$

where $\beta = \omega\sqrt{\mu\varepsilon}$ is the *phase constant* of the medium. If the region is lossy, the complex permittivity $\bar{\varepsilon}$ and the complex permeability $\bar{\mu}$ are introduced. Then, (2.34) becomes

$$\nabla^2 \mathbf{A} - \gamma^2 \mathbf{A} = -\mu\mathbf{J}. \quad (2.36)$$

Here, $\gamma = \alpha + j\beta = j\omega\sqrt{\bar{\mu}\bar{\varepsilon}}$ is the *propagation constant* and α is the *attenuation constant*.

5.2. The electric vector potential \mathbf{F}

The magnetic field is a solenoidal field, i.e., $\nabla \cdot \mathbf{B} = 0$, because there are no physically existing magnetic charges. Therefore, there are no physically existing magnetic currents either. However, the fictitious (equivalent) magnetic currents (density is denoted as \mathbf{M}) are a useful tool for antenna analysis when applied with the equivalence principle. These currents are introduced in Maxwell's equations in a manner dual to that of the electric currents \mathbf{J} . Now,

we consider the field due to *magnetic sources only*, i.e., we set $\mathbf{J} = 0$ and $\rho = 0$, and therefore, $\nabla \cdot \mathbf{D} = 0$. Then, the system of Maxwell's equations is

$$\begin{cases} \nabla \times \mathbf{E} = -j\omega\mu\mathbf{H} - \mathbf{M}, \\ \nabla \times \mathbf{H} = j\omega\varepsilon\mathbf{E}. \end{cases} \quad (2.37)$$

Since \mathbf{D} is solenoidal (i.e. $\nabla \cdot \mathbf{D} = 0$), it can be expressed as the curl of a vector, namely the electric vector potential \mathbf{F} :

$$\mathbf{D} = -\nabla \times \mathbf{F}. \quad (2.38)$$

Equation (2.38) is substituted into (2.37). All mathematical transformations are analogous to those made in Section 5.1. Finally, it is shown that a field due to magnetic sources is described by the vector \mathbf{F} alone, where \mathbf{F} satisfies

$$\nabla^2 \mathbf{F} + \omega^2 \mu \varepsilon \mathbf{F} = -\varepsilon \mathbf{M} \quad (2.39)$$

provided that the Lorenz gauge is imposed as

$$\nabla \cdot \mathbf{F} = -j\omega\mu\varepsilon\Psi. \quad (2.40)$$

Here, Ψ is the magnetic scalar potential.

In a linear medium, a field due to both types of sources (magnetic and electric) can be found by superimposing the partial field due to the electric sources only and the one due to the magnetic sources only.

TABLE 2.1: FIELD VECTORS IN TERMS OF VECTOR POTENTIALS

Magnetic vector-potential \mathbf{A} (electric sources only)	Electric vector-potential \mathbf{F} (magnetic sources only)
$\mathbf{B} = \nabla \times \mathbf{A}, \mathbf{H} = \frac{1}{\mu} \nabla \times \mathbf{A}$	$\mathbf{D} = -\nabla \times \mathbf{F}, \mathbf{E} = -\frac{1}{\varepsilon} \nabla \times \mathbf{F}$
$\mathbf{E} = -j\omega\mathbf{A} - \frac{j}{\omega\mu\varepsilon} \nabla \nabla \cdot \mathbf{A}$ or	$\mathbf{H} = -j\omega\mathbf{F} - \frac{j}{\omega\mu\varepsilon} \nabla \nabla \cdot \mathbf{F}$ or
$\mathbf{E} = \frac{1}{j\omega\mu\varepsilon} \nabla \times \nabla \times \mathbf{A} - \frac{\mathbf{J}}{j\omega\varepsilon}$	$\mathbf{H} = \frac{1}{j\omega\mu\varepsilon} \nabla \times \nabla \times \mathbf{F} - \frac{\mathbf{M}}{j\omega\mu}$

6. Retarded Potentials – Review

Retarded potential is a term usually used to denote the solution of the inhomogeneous Helmholtz' equation (in the frequency domain) or that of the inhomogeneous wave equation (in the time domain) in an unbounded region.

Consider the z -directed electric current density $\mathbf{J} = \hat{\mathbf{z}} J_z$. According to

(2.35), the magnetic vector potential \mathbf{A} is also z -directed and is governed by the following equation in a lossless medium:

$$\nabla^2 A_z + \beta^2 A_z = -\mu J_z. \quad (2.41)$$

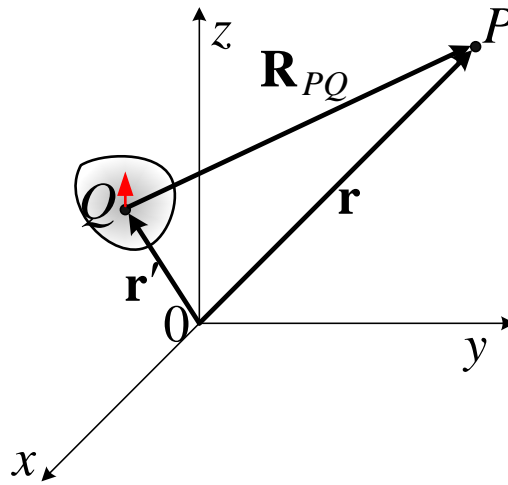
Eq. (2.41) is a Helmholtz equation and its solution in open space is determined by the integral

$$A_z(P) = \iiint_{V_Q} G(P, Q) \cdot [-\mu J_z(Q)] dv_Q \quad (2.42)$$

where $G(P, Q)$ is the open-space Green's function of the Helmholtz equation (see the Appendix), P is the observation point, and Q is the source point. Substituting (2.80) from the Appendix into (2.42) gives

$$A_z(P) = \iiint_{V_Q} \mu J_z(Q) \cdot \left(\frac{e^{-j\beta R_{PQ}}}{4\pi R_{PQ}} \right) dv_Q \quad (2.43)$$

where R_{PQ} is the distance between P and Q .



To further generalize the above formula, one assumes the existence of source currents of arbitrary directions, which would produce partial magnetic vector potentials in any direction. Note that a current element in the $\hat{\xi}$ direction results in a vector potential $\mathbf{A} = A_\xi \hat{\xi}$ in the same direction (unless the medium is inhomogeneous and/or anisotropic). Thus,

$$\mathbf{A}(P) = \iiint_{V_Q} \mu \mathbf{J}(Q) \left(\frac{e^{-j\beta R_{PQ}}}{4\pi R_{PQ}} \right) dv_Q. \quad (2.44)$$

The solution for the electric vector potential due to magnetic current sources $\mathbf{M}(Q)$ is analogous:

$$\mathbf{F}(P) = \iiint_{V_Q} \varepsilon \mathbf{M}(Q) \left(\frac{e^{-j\beta R_{PQ}}}{4\pi R_{PQ}} \right) dv_Q. \quad (2.45)$$

Finally, we recall that not only *volume* sources are used to model current distributions. A useful approximation, especially for currents on a conductor surface, is the *surface* current density (or simply surface current):

$$\mathbf{J}_s(x, y) = \lim_{\delta \rightarrow 0} \int_{-\delta/2}^{\delta/2} \mathbf{J}(x, y, z) dz, \text{ A/m}. \quad (2.46)$$

The magnetic vector potential \mathbf{A} produced by distributed surface currents is then expressed as

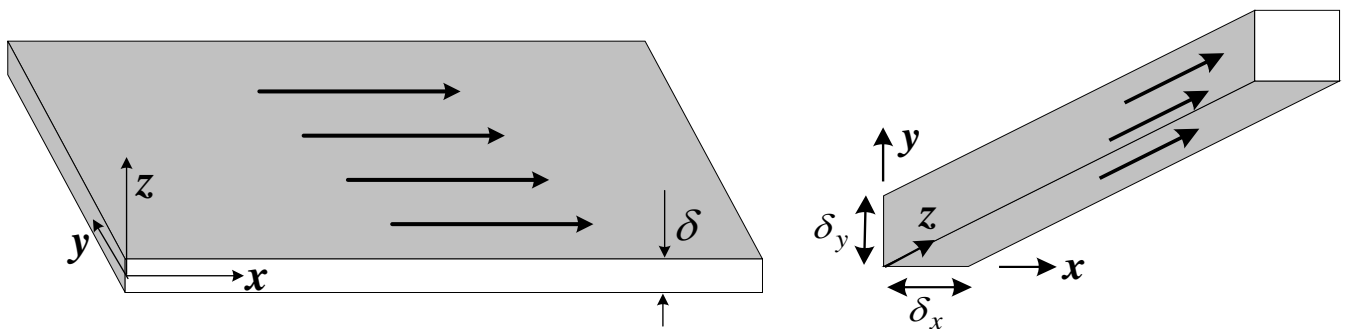
$$\mathbf{A}(P) = \iint_{S_Q} \mu \mathbf{J}_s(Q) \left(\frac{e^{-j\beta R_{PQ}}}{4\pi R_{PQ}} \right) ds_Q. \quad (2.47)$$

Currents on a very thin wire are usually approximated by a linear source, which is the current I flowing through the wire:

$$\mathbf{I}(z) = I(z) \mathbf{a}_l(z) = \lim_{\substack{\delta_x \rightarrow 0 \\ \delta_y \rightarrow 0}} \iint_{\delta_x \delta_y} \mathbf{J}(x, y, z) dx dy, \text{ A}. \quad (2.48)$$

The vector potential of line currents is

$$\mathbf{A}(P) = \int_{L_Q} \mu I(Q) \left(\frac{e^{-j\beta R_{PQ}}}{4\pi R_{PQ}} \right) d\mathbf{l}_Q. \quad (2.49)$$



(a) surface current on a sheet

(b) linear current on a thin wire

7. Far Fields and Vector Potentials

7.1. Potentials

Antennas are sources of finite physical dimensions. The further away from the antenna the observation point is, the more the wave looks like a spherical wave and the more the antenna looks like a point source regardless of its actual shape. For such observation distances, we talk about *far field* and *far zone*. The exact meaning of these terms will be discussed later. For now, we will simply accept that the vector potentials behave like spherical waves, when the observation point is far from the source:

$$\mathbf{A} \approx \underbrace{\left[\hat{\mathbf{r}}A_r(\theta, \varphi) + \hat{\boldsymbol{\theta}}A_\theta(\theta, \varphi) + \hat{\boldsymbol{\phi}}A_\phi(\theta, \varphi) \right]}_{\text{dependence on observation angles only}} \cdot \underbrace{\frac{e^{-jkr}}{r}}_{\text{dependence on distance only}}, \quad r \rightarrow \infty. \quad (2.50)$$

Here, $(\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}})$ are the unit vectors of the spherical coordinate system (SCS) centered on the antenna and $k = \omega\sqrt{\mu\epsilon}$ is the wave number (or the phase constant). The term e^{-jkr} shows propagation along $\hat{\mathbf{r}}$ away from the antenna at the speed of light. The term $1/r$ shows the spherical spread of the potential in space, which results in a decrease of its magnitude with distance.

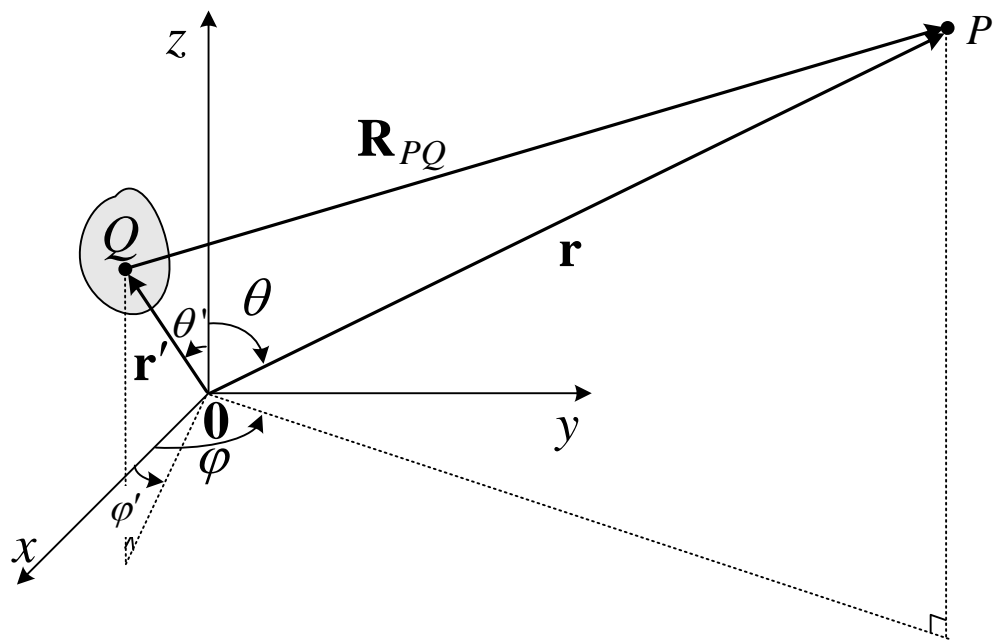
Notice an important feature of the far-field potential: the dependence on the distance r is separable from the dependence on the observation angle (θ, φ) , and it is the same for any antenna: e^{-jkr} / r .

Formula (2.50) is a *far-field approximation* of the vector potential at distant points. We arrive at it starting from the integral in (2.44). When the observation point P is very far from the source, the distance R_{PQ} varies only slightly as Q sweeps the volume of the source. It is almost the same as the distance r from the origin (the antenna center) to P . The following first-order approximation (attributed to Kirchhoff) is made for the integrand:

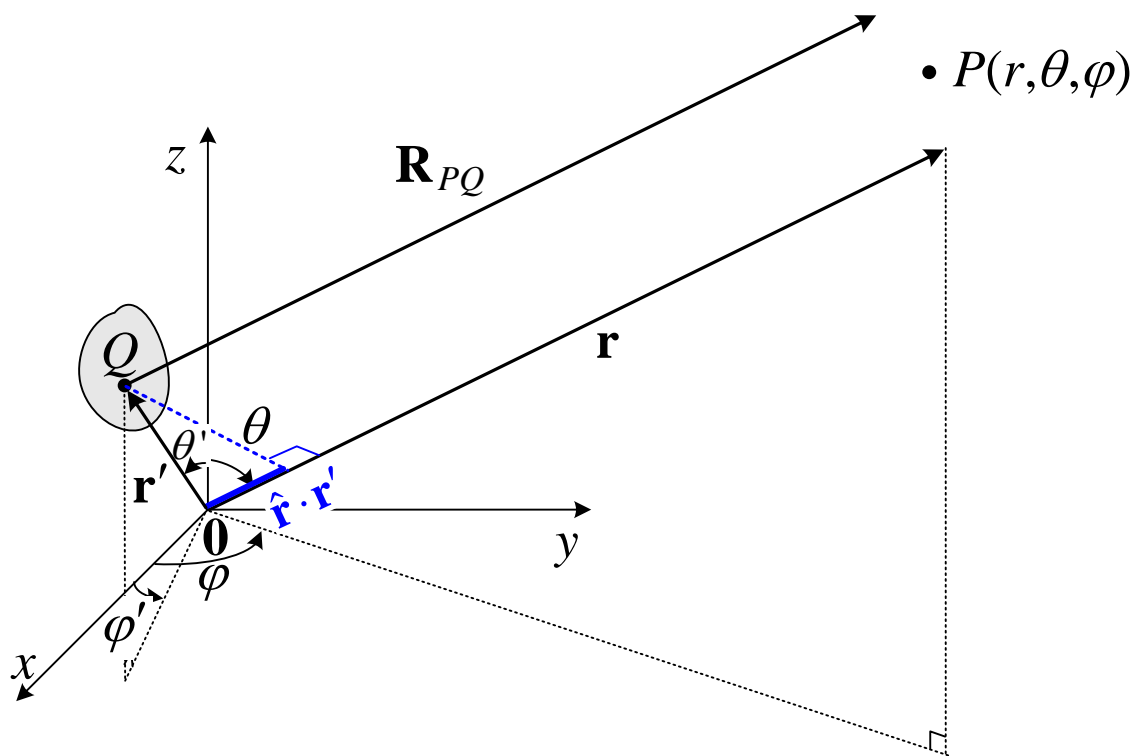
$$\frac{e^{-jkR_{PQ}}}{R_{PQ}} \approx \frac{e^{-jk(r - \hat{\mathbf{r}} \cdot \mathbf{r}')}}{r}. \quad (2.51)$$

Here, \mathbf{r} is the position vector of the observation point P and $r = |\mathbf{r}|$ is its length. Its direction is given by the unit vector $\hat{\mathbf{r}}$, so that $\mathbf{r} = \hat{\mathbf{r}} \cdot r$. The position vector of the integration point Q is \mathbf{r}' . Equation (2.51) is called the *far-field approximation*.

The approximation in the phase term (in the exponent) is illustrated in the figures below. The first figure shows the real problem. The second one shows the approximated problem, where, in effect, the vectors \mathbf{R}_{PQ} and \mathbf{r} are parallel.



(a) original problem



(b) far-field approximation of the original problem

We now apply the far-field approximation to the vector potential in (2.44):

$$\mathbf{A}(P) = \underbrace{\frac{e^{-jkr}}{4\pi r}}_{\text{dependence on distance from origin}} \cdot \underbrace{\iiint_{V_Q} \mu \mathbf{J}(Q) e^{jk\hat{\mathbf{r}} \cdot \mathbf{r}'} dv_Q}_{\text{dependence on source distribution and angular orientation}}. \quad (2.52)$$

The integrand in (2.52) no longer depends on the distance r between the origin and the observation point. It depends only on the current distribution of the source and the angle between the position vector of the integration point \mathbf{r}' and the unit position vector of the observation point $\hat{\mathbf{r}}$. This finally explains the general equation for the *far-field vector potential* in (2.50) and in particular the origin of the term in the square brackets, which is represented by the volume integral in (2.52).

7.2. Far-zone field

The far-field approximation of the vector potential leads to much simpler equations for the far-field vectors. Assume that there are only electrical currents. Then the field is fully described only by the magnetic vector potential \mathbf{A} . We have to substitute (2.50) into the equations of Table 2.1, where $\mathbf{F} = 0$:

$$\mathbf{E} = -j\omega\mathbf{A} - \frac{j}{\omega\mu\epsilon} \nabla \nabla \cdot \mathbf{A}, \quad (2.53)$$

$$\mathbf{H} = \frac{1}{\mu} \nabla \times \mathbf{A}. \quad (2.54)$$

The differential operators $\nabla \times$ and $\nabla \nabla \cdot$ have to be expressed in spherical coordinates. All terms decreasing with the distance as $1/r^2$ and faster are neglected. What remains is

$$\mathbf{E} = \frac{1}{r} \left\{ -j\omega e^{-jkr} \left[\hat{\boldsymbol{\theta}} A_\theta(\theta, \varphi) + \hat{\boldsymbol{\phi}} A_\varphi(\theta, \varphi) \right] \right\} + \frac{1}{r^2} \{ \} + \dots, \quad r \rightarrow \infty, \quad (2.55)$$

$$\mathbf{H} = \frac{1}{r} \left\{ j \frac{\omega}{\eta} e^{-jkr} \left[\hat{\boldsymbol{\theta}} A_\varphi(\theta, \varphi) - \hat{\boldsymbol{\phi}} A_\theta(\theta, \varphi) \right] \right\} + \frac{1}{r^2} \{ \} + \dots, \quad r \rightarrow \infty. \quad (2.56)$$

Here, $\eta = \sqrt{\mu/\epsilon}$ denotes the intrinsic impedance of the medium. We write the far-field equations (2.55) and (2.56) in a more compact way as

$$\left. \begin{array}{l} E_r \approx 0 \\ E_\theta \approx -j\omega A_\theta \\ E_\varphi \approx -j\omega A_\varphi \end{array} \right\} \Rightarrow \mathbf{E}^A \approx -j\omega\mathbf{A}, \quad \text{where } E_r^A \approx 0, \quad (2.57)$$

$$\left. \begin{aligned} H_r &\approx 0 \\ H_\theta &\approx +j\frac{\omega}{\eta}A_\phi = -\frac{E_\phi}{\eta} \\ H_\phi &\approx -j\frac{\omega}{\eta}A_\theta = +\frac{E_\theta}{\eta} \end{aligned} \right\} \Rightarrow \mathbf{H}^A \approx -j\frac{\omega}{\eta}\hat{\mathbf{r}} \times \mathbf{A} = \frac{1}{\eta}\hat{\mathbf{r}} \times \mathbf{E}^A. \quad (2.58)$$

In an analogous manner, we obtain the relations between the field vectors and the electric vector potential \mathbf{F} , when only magnetic sources are present:

$$\left. \begin{aligned} H_r &\approx 0 \\ H_\theta &\approx -j\omega F_\theta \\ H_\phi &\approx -j\omega F_\phi \end{aligned} \right\} \Rightarrow \mathbf{H}^F \approx -j\omega\mathbf{F}, \quad H_r^F \approx 0, \quad (2.59)$$

$$\left. \begin{aligned} E_r &\approx 0 \\ E_\theta &\approx -j\omega\eta F_\phi = \eta H_\phi \\ E_\phi &\approx +j\omega\eta F_\theta = -\eta H_\theta \end{aligned} \right\} \Rightarrow \mathbf{E}^F \approx j\omega\eta\hat{\mathbf{r}} \times \mathbf{F} = -\eta\hat{\mathbf{r}} \times \mathbf{H}^F. \quad (2.60)$$

In summary, the far field of any antenna has the following important features, which follow from equations (2.57) through (2.60):

- The far field has negligible radial components, $E_r \approx 0$ and $H_r \approx 0$. Since the radial direction is also the direction of propagation, the far field is a typical TEM (Transverse Electro-Magnetic) wave.
- The far-field \mathbf{E} vector and \mathbf{H} vector are mutually orthogonal, both of them being also orthogonal to the direction of propagation.
- The magnitudes of the electric field and the magnetic field are related always as $|\mathbf{E}| = \eta |\mathbf{H}|$.

APPENDIX

Green's Function for the Helmholtz Equation

Suppose the following PDE must be solved:

$$L\Phi(\mathbf{x}) = f(\mathbf{x}) \quad (2.61)$$

where \mathbf{x} denotes the set of variables, e.g., $\mathbf{x} = (x, y, z)$. Suppose also that a Green's function exists such that it allows for the integral solution

$$\Phi(\mathbf{x}) = \iiint_{V'} G(\mathbf{x}, \mathbf{x}') \cdot f(\mathbf{x}') dv' \quad (2.62)$$

Applying the operator L to both sides of (2.62), leads to

$$L\Phi(\mathbf{x}) = \iiint_{V'} [LG(\mathbf{x}, \mathbf{x}')] \cdot f(\mathbf{x}') dv' = f(\mathbf{x}). \quad (2.63)$$

Note that L operates on the variable \mathbf{x} while the integral in (2.63) is over \mathbf{x}' . This allows for the insertion of L inside the integral. From (2.63), we conclude that the Green's function must satisfy the same PDE as Φ with a point source described by Dirac's delta function:

$$LG(\mathbf{x}, \mathbf{x}') = \delta(\mathbf{x} - \mathbf{x}'). \quad (2.64)$$

Here, $\delta(\mathbf{x} - \mathbf{x}')$ is Dirac's delta function in 3-D space, e.g., $\delta(\mathbf{x} - \mathbf{x}') = \delta(x - x')\delta(y - y')\delta(z - z')$. If the Green's function of a problem is known and the source function $f(\mathbf{x})$ is known, the construction of an integral solution is possible via (2.62).

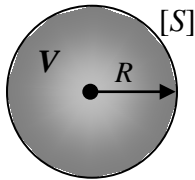
Consider the Green's function for the Helmholtz equation in open space. It must satisfy

$$\nabla^2 G + \beta^2 G = \delta(x)\delta(y)\delta(z) \quad (2.65)$$

together with the scalar radiation condition

$$\lim_{r \rightarrow \infty} r \cdot \left(\frac{\partial G}{\partial r} + j\beta G \right) = 0 \quad (2.66)$$

if the source is centered at the origin of the coordinate system, i.e., $x' = y' = z' = 0$. Integrate (2.65) within a sphere with its center at (0,0,0) and a radius R :



$$\iiint_V \nabla^2 G dv + \iiint_V \beta^2 G dv = 1 \quad (2.67)$$

The function G is due to a point source and thus has a spherical symmetry, i.e., it depends on r only. The Laplacian ∇^2 in spherical coordinates is reduced to derivatives with respect to r only:

$$\frac{d^2 G}{dr^2} + \frac{2}{r} \frac{dG}{dr} + \beta^2 G = \delta(x)\delta(y)\delta(z). \quad (2.68)$$

Everywhere except at the point (x, y, z) , G must satisfy the homogeneous equation

$$\frac{d^2 G}{dr^2} + \frac{2}{r} \frac{dG}{dr} + \beta^2 G = 0 \quad (2.69)$$

whose solution for outgoing waves is well known:

$$G(r) = C \frac{e^{-jkr}}{r}. \quad (2.70)$$

Here, C is a constant to be determined. Consider first the integral from (2.67):

$$I_1 = \iiint_V \beta^2 G dv. \quad (2.71)$$

$$\Rightarrow I_1 = \iiint_V \beta^2 C \frac{e^{-j\beta r}}{r} dv = \int_0^R \int_0^{2\pi} \int_0^\pi \beta^2 C \frac{e^{-j\beta r}}{r} r^2 \sin \theta d\theta d\varphi dr \quad (2.72)$$

$$\Rightarrow I_1(R) = j4\pi\beta C \left(R \cdot e^{-j\beta R} + \frac{e^{-j\beta R}}{j\beta} - \frac{1}{j\beta} \right). \quad (2.73)$$

To evaluate the integral in the point of singularity (0,0,0), we let $R \rightarrow 0$, i.e., we let the sphere collapse into a point. We see that

$$\lim_{R \rightarrow 0} I_1(R) = 0. \quad (2.74)$$

Secondly, consider the other integral in (2.67),

$$I_2 = \iiint_V \nabla^2 G dv = \iiint_V \nabla \cdot (\nabla G) dv = \oiint_S \nabla G \cdot ds. \quad (2.75)$$

Here, $ds = R^2 \sin \theta dr d\theta d\varphi \cdot \hat{\mathbf{r}}$ is a surface element on S , and

$$\nabla G = \frac{\partial G}{\partial r} \hat{\mathbf{r}} = -C \left(jk \frac{e^{-jkr}}{r} + \frac{e^{-jkr}}{r^2} \right) \hat{\mathbf{r}}. \quad (2.76)$$

Substitute (2.76) in (2.75) and carry out the integration over the spherical surface:

$$I_2(R) = -C \left(jkR \cdot e^{-jkR} + e^{-jkR} \right) \int_0^\pi \int_0^{2\pi} \sin \theta d\varphi d\theta \quad (2.77)$$

$$\lim_{R \rightarrow 0} I_2(R) = -4\pi C. \quad (2.78)$$

Substituting (2.78) and (2.74) into (2.67) and taking $\lim_{R \rightarrow 0}$, yields

$$C = -\frac{1}{4\pi}. \quad (2.79)$$

Finally,

$$G(r) = -\frac{e^{-jkr}}{4\pi r}. \quad (2.80)$$

It is not difficult to show that in the general case when the source is at a point $Q(x', y', z')$,

$$\nabla^2 G + \beta^2 G = \delta(x - x') \delta(y - y') \delta(z - z') \quad (2.81)$$

the Green function is

$$G(P, Q) = -\frac{e^{-jkR_{PQ}}}{4\pi R_{PQ}}, \quad (2.82)$$

where R_{PQ} is the distance between the observation point P and the source point Q ,

$$R_{PQ} = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}. \quad (2.83)$$

LECTURE 3: Radiation from Infinitesimal (Elementary) Sources

(Radiation from an infinitesimal dipole. Duality of Maxwell's equations. Radiation from an infinitesimal loop. Radiation zones.)

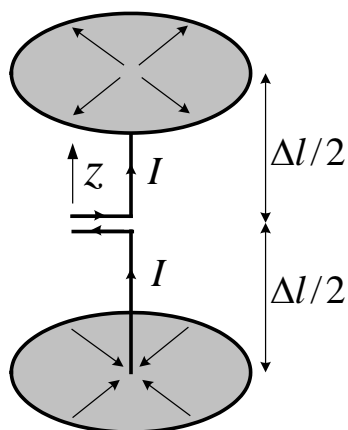
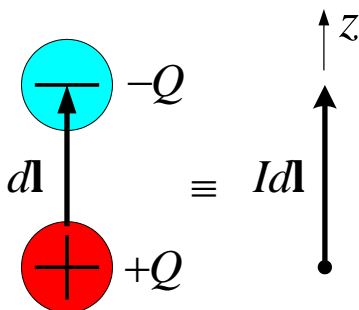
1. Radiation from Infinitesimal Dipole (Electric-current Element)

Definition: The infinitesimal dipole is a straight line segment of length Δl , which is much smaller than the wavelength λ of the excited wave, i.e. $\Delta l \ll \lambda$ ($\Delta l < \lambda/50$), and which supports constant current distribution I along its length. The assumed positive direction of the current I determines the orientation of the line segment: $\Delta \mathbf{l} = \Delta l \hat{\mathbf{i}}$.

The infinitesimal dipole is mathematically described by a current element:

$$I d\mathbf{l} = -\frac{dQ}{dt} d\mathbf{l}.$$

A current element is best illustrated by a very short (compared to λ) piece of infinitesimally thin wire with constant current I . The ideal current element is difficult to realize in practice, but a good approximation of it is the short top-hat antenna. To realize a uniform current distribution along the wire, capacitive plates are used to provide enough charge storage at the end of the wire, so that the current is not zero there.



1.1. Magnetic vector potential due to a current element

The magnetic vector potential (VP) \mathbf{A} due to a linear source is (see Lecture 2):

$$\mathbf{A}(P) = \int_L \mu I(Q) \frac{e^{-j\beta R_{PQ}}}{4\pi R_{PQ}} d\mathbf{l}_Q \quad (3.1)$$

$$\Rightarrow \mathbf{A}(P) = \hat{\mathbf{z}} \frac{\mu_0 I}{4\pi} \int_{\Delta l} \frac{e^{-j\beta R_{PQ}}}{R_{PQ}} dl. \quad (3.2)$$

If we assume that the dipole's length Δl is much smaller than the distance from its center to the observation point P , then $R_{PQ} \approx r$ holds both in the exponential term and in the denominator. Therefore,

$$\mathbf{A} = \mu_0 I \Delta l \frac{e^{-j\beta r}}{4\pi r} \hat{\mathbf{z}}. \quad (3.3)$$

Equation (3.3) gives the vector potential due to an electric current element (infinitesimal dipole). This is an important result because *the field radiated by any complex antenna in a linear medium is a superposition of the fields due to the current elements on the antenna surface*.

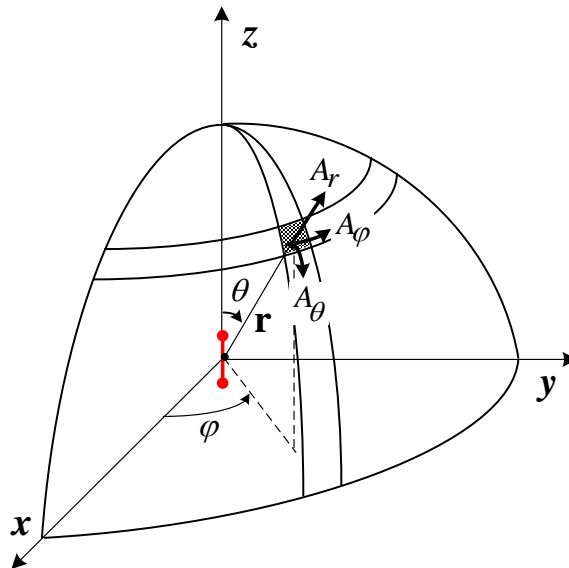
We represent \mathbf{A} with its spherical components. In antenna theory, the preferred coordinate system is the spherical one. This is because the *far field* radiation is of interest where the field dependence on the distance r from the source is decoupled from its angular dependence. This angular dependence is described conveniently in terms of the two angles in the spherical coordinate system (SCS) φ and θ . Also, this field propagates radially (along $\hat{\mathbf{r}}$) when the source is located at the origin of the coordinate system.

The transformation from rectangular to spherical coordinates is given by

$$\begin{bmatrix} A_r \\ A_\theta \\ A_\varphi \end{bmatrix} = \begin{bmatrix} \sin \theta \cos \varphi & \sin \theta \sin \varphi & \cos \theta \\ \cos \theta \cos \varphi & \cos \theta \sin \varphi & -\sin \theta \\ -\sin \varphi & \cos \varphi & 0 \end{bmatrix} \begin{bmatrix} A_x \\ A_y \\ A_z \end{bmatrix}. \quad (3.4)$$

Applying (3.4) to \mathbf{A} in (3.3) produces

$$\begin{aligned} A_r &= A_z \cos \theta = \mu_0 I \Delta l \frac{e^{-j\beta r}}{4\pi r} \cos \theta \\ A_\theta &= -A_z \sin \theta = -\mu_0 I \Delta l \frac{e^{-j\beta r}}{4\pi r} \sin \theta \\ A_\varphi &= 0 \end{aligned} \quad (3.5)$$



Note that:

- 1) \mathbf{A} does not depend on φ (due to the cylindrical symmetry of the dipole);
- 2) the dependence on r , $e^{-j\beta r} / r$, is separable from the dependence on θ .

1.2. Field vectors due to current element

Next we find the field vectors \mathbf{H} and \mathbf{E} from \mathbf{A} .

$$\text{a) } \mathbf{H} = \frac{1}{\mu} \nabla \times \mathbf{A} \quad (3.6)$$

The curl operator is expressed in spherical coordinates to obtain

$$\mathbf{H} = \frac{1}{\mu r} \left[\frac{\partial}{\partial r} (r \cdot A_\theta) - \frac{\partial A_r}{\partial \theta} \right] \hat{\boldsymbol{\phi}}. \quad (3.7)$$

Thus, the magnetic field \mathbf{H} has only a φ -component, i.e.,

$$\begin{cases} H_\varphi = j\beta \cdot (I\Delta l) \cdot \sin \theta \cdot \left(1 + \frac{1}{j\beta r} \right) \frac{e^{-j\beta r}}{4\pi r}, \\ H_\theta = H_r = 0. \end{cases} \quad (3.8)$$

$$\text{b) } \mathbf{E} = \frac{1}{j\omega\epsilon} \nabla \times \mathbf{H} = -j\omega\mathbf{A} - \frac{j}{\omega\mu\epsilon} \nabla \nabla \cdot \mathbf{A} \quad (3.9)$$

In spherical coordinates, the \mathbf{E} field components are:

$$\begin{cases} E_r = j2\eta\beta (I\Delta l) \cdot \cos \theta \cdot \left[\frac{1}{j\beta r} - \frac{1}{(\beta r)^2} \right] \frac{e^{-j\beta r}}{4\pi r} \\ E_\theta = j\eta\beta (I\Delta l) \cdot \sin \theta \cdot \left[1 + \frac{1}{j\beta r} - \frac{1}{(\beta r)^2} \right] \frac{e^{-j\beta r}}{4\pi r} \\ E_\varphi = 0. \end{cases} \quad (3.10)$$

Notes: 1) Equations (3.8) and (3.10) show that the EM field generated by the current element is quite complicated unlike the VP \mathbf{A} . The use of the VP instead of the field vectors is often advantageous in antenna studies.

2) The field vectors contain terms, which depend on the distance from the source as $1/r$, $1/r^2$ and $1/r^3$; the higher-order terms can be neglected at large distances from the dipole.

3) The longitudinal $\hat{\mathbf{r}}$ -component of the \mathbf{E} field vector decreases fast as the field propagates away from the source (as $1/r^2$ and $1/r^3$): it is neglected in the far zone. The longitudinal \mathbf{H} field component of the infinitesimal electric dipole is zero everywhere.

4) The nonzero transverse field components, E_θ and H_ϕ , are orthogonal to each other, and they have terms that depend on the distance as $1/r$. These terms relate through the intrinsic impedance η and they describe a TEM wave. They represent the so-called *far field* which satisfies the Sommerfeld vector radiation boundary conditions. The concept of far field will be re-visited later, when the radiation zones are defined.

1.3. Power density and overall radiated power of the infinitesimal dipole

The complex vector of Poynting \mathbf{P} describes the complex power density flux. In the case of infinitesimal dipole, it is

$$\mathbf{P} = \frac{1}{2}(\mathbf{E} \times \mathbf{H}^*) = \frac{1}{2}(E_r \hat{\mathbf{r}} + E_\theta \hat{\boldsymbol{\theta}}) \times (H_\phi^* \hat{\boldsymbol{\phi}}) = \frac{1}{2}(E_\theta H_\phi^* \hat{\mathbf{r}} - E_r H_\phi^* \hat{\boldsymbol{\theta}}). \quad (3.11)$$

Substituting (3.8) and (3.10) into (3.11) yields

$$\begin{cases} P_r = \frac{\eta}{8} \left| \frac{I\Delta l}{\lambda} \right|^2 \frac{\sin^2 \theta}{r^2} \left[1 - j \frac{1}{(\beta r)^3} \right], \\ P_\theta = j\eta\beta \frac{|I\Delta l|^2 \cos \theta \sin \theta}{16\pi^2 r^3} \left[1 + \frac{1}{(\beta r)^2} \right]. \end{cases} \quad (3.12)$$

The overall power Π is calculated over a sphere, and, therefore, only the radial component P_r contributes:

$$\Pi = \oiint_S \mathbf{P} \cdot d\mathbf{s} = \oiint_S (P_r \hat{\mathbf{r}} + P_\theta \hat{\boldsymbol{\theta}}) \cdot \hat{\mathbf{r}} r^2 \sin \theta d\theta d\phi, \quad (3.13)$$

$$\Pi = \frac{\pi}{3} \eta \left| \frac{I\Delta l}{\lambda} \right|^2 \left[1 - \frac{j}{(\beta r)^3} \right], \text{ W.} \quad (3.14)$$

The radiated power is equal to the real part of the complex power (the time-average of the total power flow, see Lecture 2). Therefore, the radiated power of an infinitesimal electric dipole is

$$\Pi_{rad} = \frac{\pi}{3} \eta \left(\frac{I \Delta l}{\lambda} \right)^2, \text{ W.} \quad (3.15)$$

Here, we introduce the concept of radiation resistance R_r , which describes the power loss due to radiation in the equivalent circuit of the antenna:

$$\Pi = \frac{1}{2} R_r I^2 \Rightarrow R_r = \frac{2\Pi}{I^2} \quad (3.16)$$

$$\Rightarrow R_r^{id} = \frac{2\pi}{3} \eta \left(\frac{\Delta l}{\lambda} \right)^2, \Omega. \quad (3.17)$$

Note that (3.17) holds only for an infinitesimal dipole, i.e., when the current is assumed constant over the length Δl of the dipole.

2. Duality in Maxwell's Equations

Duality in electromagnetics means that the EM field is described by two sets of quantities, which correspond to each other in such a manner that substituting the quantities from one set with the respective quantities from the other set in any given equation produces a valid equation (the dual of the given one).

We deduce these dual sets by comparing the equations describing two dual fields: the field of electric sources and the field of magnetic sources. Note that duality exists even if there are no sources present in the region of interest. Tables 2.1 and 2.2 summarize the duality of the EM equations and quantities.

TABLE 2.1. DUALITY IN ELECTROMAGNETIC EQUATIONS

Electric sources ($\mathbf{J} \neq 0, \mathbf{M} = 0$)	Magnetic sources ($\mathbf{J} = 0, \mathbf{M} \neq 0$)
$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H}$	$\nabla \times \mathbf{H} = j\omega\varepsilon\mathbf{E}$
$\nabla \times \mathbf{H} = j\omega\varepsilon\mathbf{E} + \mathbf{J}$	$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H} - \mathbf{M}$
$\nabla \cdot \mathbf{D} = \rho$	$\nabla \cdot \mathbf{B} = \rho_m$
$\nabla \cdot \mathbf{B} = 0$	$\nabla \cdot \mathbf{D} = 0$
$\nabla \cdot \mathbf{J} = -j\omega\rho$	$\nabla \cdot \mathbf{M} = -j\omega\rho_m$
$\nabla^2 \mathbf{A} + \beta^2 \mathbf{A} = -\mu\mathbf{J}$	$\nabla^2 \mathbf{F} + \beta^2 \mathbf{F} = -\varepsilon\mathbf{M}$
$\mathbf{A} = \iiint_v \mu\mathbf{J} \frac{e^{-j\beta R}}{4\pi R} dv$	$\mathbf{F} = \iiint_v \varepsilon\mathbf{M} \frac{e^{-j\beta R}}{4\pi R} dv$
$\mathbf{H} = \mu^{-1} \nabla \times \mathbf{A}$	$\mathbf{E} = -\varepsilon^{-1} \nabla \times \mathbf{F}$
$\mathbf{E} = -j\omega\mathbf{A} + (j\omega\mu\varepsilon)^{-1} \nabla \nabla \cdot \mathbf{A}$	$\mathbf{H} = -j\omega\mathbf{F} + (j\omega\mu\varepsilon)^{-1} \nabla \nabla \cdot \mathbf{F}$

TABLE 2.2. DUAL QUANTITIES IN ELECTROMAGNETICS

given	E	H	J	M	A	F	ε	μ	η	$1/\eta$	β
dual	H	-E	M	-J	F	-A	μ	ε	$1/\eta$	η	β

3. Radiation from Infinitesimal Magnetic Dipole (Electric-current Loop)

3.1. The vector potential and the field vectors of a magnetic dipole (magnetic current element) $I_m \Delta l$

Using the duality theorem, the field of a magnetic dipole $I_m \Delta l$ is readily found by a simple substitution of the dual quantities in equations (3.5), (3.8) and (3.10) according to Table 2.2. We denote the magnetic current, which is the dual of the electric current I , by I_m (measured in *volts*).

(a) the electric vector potential

$$\left\{ \begin{array}{l} F_r = F_z \cos \theta = \varepsilon_0 (I_m \Delta l) \frac{e^{-j\beta r}}{4\pi r} \cos \theta \\ F_\theta = -F_z \sin \theta = -\varepsilon_0 (I_m \Delta l) \frac{e^{-j\beta r}}{4\pi r} \sin \theta \\ F_\varphi = 0 \end{array} \right. \quad (3.18)$$

(b) the electric field of the magnetic dipole

$$\left\{ \begin{array}{l} E_\varphi = -j\beta \cdot (I_m \Delta l) \cdot \sin \theta \cdot \left(1 + \frac{1}{j\beta r} \right) \frac{e^{-j\beta r}}{4\pi r} \\ E_\theta = E_r = 0 \end{array} \right. \quad (3.19)$$

(c) the magnetic field of the magnetic dipole

$$\left\{ \begin{array}{l} H_r = \frac{2(I_m \Delta l) \cos \theta}{\eta} \left(\frac{1}{r} + \frac{1}{j\beta r^2} \right) \frac{e^{-j\beta r}}{4\pi r} \\ H_\theta = \frac{j\beta (I_m \Delta l) \sin \theta}{\eta} \left(1 + \frac{1}{j\beta r} - \frac{1}{\beta^2 r^2} \right) \frac{e^{-j\beta r}}{4\pi r} \\ H_\varphi = 0 \end{array} \right. \quad (3.20)$$

3.2. Equivalence between a magnetic dipole (magnetic current element) and an electric current loop

First, we prove the equivalence of the fields excited by particular configurations of electric and magnetic current densities. We write Maxwell's equations for the two cases:

(a) electric current density

$$\begin{cases} -\nabla \times \mathbf{E}_1 = j\omega\mu\mathbf{H}_1 \\ \nabla \times \mathbf{H}_1 = j\omega\varepsilon\mathbf{E}_1 + \mathbf{J} \end{cases} \quad (3.21)$$

$$\Rightarrow \nabla \times \nabla \times \mathbf{E}_1 - \omega^2\mu\varepsilon\mathbf{E}_1 = -j\omega\mu\mathbf{J} \quad (3.22)$$

(b) magnetic current density

$$\begin{cases} -\nabla \times \mathbf{E}_2 = j\omega\mu\mathbf{H}_2 + \mathbf{M} \\ \nabla \times \mathbf{H}_2 = j\omega\varepsilon\mathbf{E}_2 \end{cases} \quad (3.23)$$

$$\Rightarrow \nabla \times \nabla \times \mathbf{E}_2 - \omega^2\mu\varepsilon\mathbf{E}_2 = -\nabla \times \mathbf{M} \quad (3.24)$$

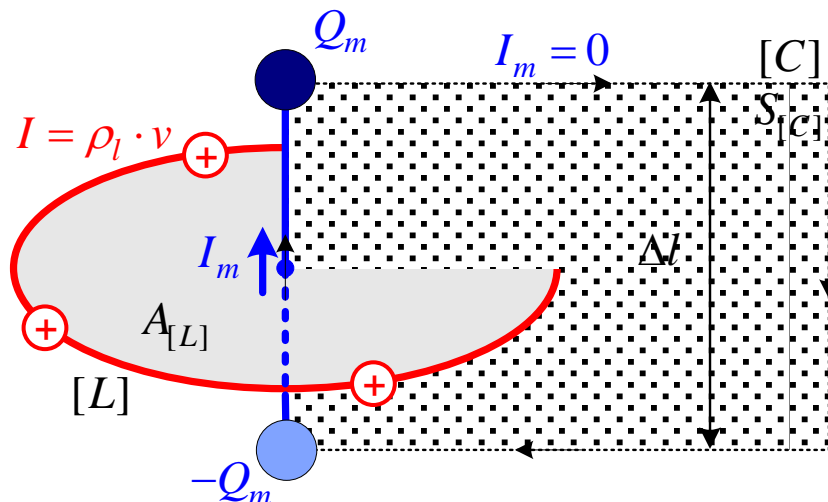
If the boundary conditions (BCs) for \mathbf{E}_1 in (3.22) are the same as the BCs for \mathbf{E}_2 in (3.24), and the excitations of both fields fulfill

$$j\omega\mu\mathbf{J} = \nabla \times \mathbf{M}, \quad (3.25)$$

then both fields are identical, i.e., $\mathbf{E}_1 \equiv \mathbf{E}_2$ and $\mathbf{H}_1 \equiv \mathbf{H}_2$.

Consider a loop $[L]$ of electric current I . Equation (3.25) can be written in its integral form as

$$j\omega\mu \iint_{S_{[C]}} \mathbf{J} \cdot d\mathbf{s} = \oint_C \mathbf{M} \cdot d\mathbf{l}. \quad (3.26)$$



The integral on the left side is the electric current I . \mathbf{M} is assumed non-zero and constant only at the section Δl , which is normal to the loop's plane and passes through the loop's centre. Then,

$$j\omega\mu I = M\Delta l. \quad (3.27)$$

The magnetic current I_m corresponding to the loop $[L]$ is obtained by multiplying the magnetic current density \mathbf{M} by the area of the loop $A_{[L]}$, which yields

$$j\omega\mu IA_{[L]} = I_m\Delta l. \quad (3.28)$$

Thus, we show that *a small loop of electric current I and of area $A_{[L]}$ creates EM field equivalent to that of a small magnetic dipole (magnetic current element) $I_m\Delta l$, such that (3.28) holds.*

Here, it was assumed that the electric current is constant along the loop, which is true only for very small loops ($a < 0.1\lambda$, where a is the loop's radius and the loop has only 1 turn). If the loop is larger, the field expressions below are inaccurate and other solutions should be used. We will discuss the loop antennas in more detail in a dedicated lecture.

3.3. Field vectors of an infinitesimal loop antenna

The expressions below are derived by substituting (3.28) into (3.19)-(3.20):

$$E_\varphi = \eta\beta^2(IA) \cdot \sin\theta \left(1 + \frac{1}{j\beta r}\right) \frac{e^{-j\beta r}}{4\pi r}, \quad (3.29)$$

$$H_r = j2\beta(IA) \cdot \cos\theta \left(\frac{1}{r} + \frac{1}{j\beta r^2}\right) \frac{e^{-j\beta r}}{4\pi r}, \quad (3.30)$$

$$H_\theta = -\beta^2(IA) \cdot \sin\theta \left(1 + \frac{1}{j\beta r} - \frac{1}{\beta^2 r^2}\right) \frac{e^{-j\beta r}}{4\pi r}, \quad (3.31)$$

$$E_r = E_\theta = H_\varphi = 0. \quad (3.32)$$

The far-field terms ($1/r$ dependence on the distance from the source) show the same behaviour as in the case of an infinitesimal dipole antenna: (1) the electric field E_φ is orthogonal to the magnetic field H_θ ; (2) E_φ and H_θ relate through η ; (3) the longitudinal $\hat{\mathbf{r}}$ components have no far-field terms.

The dependence of the Poynting vector and the complex power on the distance r is the same as in the case of an infinitesimal electric dipole. The radiated power can be found to be

$$\Pi_{rad} = \eta\beta^4 (IA)^2 / 12\pi. \quad (3.33)$$

4. Radiation Zones – Introduction

The space surrounding the antenna is divided into three regions according to the dominant field behaviour. The boundaries between the regions are not distinct and the field behaviour changes gradually as these boundaries are crossed. In this course, we are mostly concerned with the far-field characteristics of the antennas.

Next, we illustrate the three radiation zones through the field of the small electric dipole.

4.1. Reactive near-field region

This is the region immediately surrounding the antenna, where the reactive field dominates and the angular field distribution is different at different distances from the antenna. For most antennas, it is assumed that this region is a sphere with the antenna at its centre, and with a radius

$$r_{\text{RNF}} \approx 0.62\sqrt{D^3 / \lambda}, \quad (3.34)$$

where D is the largest dimension of the antenna, and λ is the wavelength of the radiation. The above expression will be derived in Section 5. It must be noted that this limit is most appropriate for wire and waveguide aperture antennas while it is not valid for electrically large reflector antennas.

At this point, we discuss the general field behaviour making use of our knowledge of the infinitesimal electric-dipole field. When (3.34) is true, r is sufficiently small so that $\beta r \ll 1$ (note that $D \ll \lambda$ for the infinitesimal dipole). Then, the most significant terms in the field expressions (3.8) and (3.10) are

$$\left. \begin{aligned} H_\phi &\approx \frac{(I\Delta l)e^{-j\beta r}}{4\pi r^2} \sin\theta \\ E_\theta &\approx -j\eta \frac{(I\Delta l)e^{-j\beta r}}{4\pi\beta r^3} \sin\theta \\ E_r &\approx -j\eta \frac{(I\Delta l)e^{-j\beta r}}{2\pi\beta r^3} \cos\theta \\ H_r &= H_\theta = E_\phi = 0 \end{aligned} \right\}, \beta r \ll 1. \quad (3.35)$$

This approximated field is purely reactive (\mathbf{H} and \mathbf{E} are in phase quadrature). Since $e^{-j\beta r} \approx 1$ we see that: (1) H_ϕ has the distribution of the magnetostatic

field of a current filament $I\Delta l$ (remember Bio-Savart's law); (2) E_θ and E_r have the distribution of the electrostatic field of a dipole.

That the field is almost purely reactive in the near zone is obvious from the power equation (3.14). Its imaginary part is

$$\text{Im}\{\Pi\} = -\frac{\pi}{3}\eta\left(\frac{I\Delta l}{\lambda}\right)^2 \frac{1}{(\beta r)^3}. \quad (3.36)$$

$\text{Im}\{\Pi\}$ dominates over the radiated power,

$$\Pi_{rad} = \text{Re}\{\Pi\} = \frac{\pi}{3}\eta\left(\frac{I\Delta l}{\lambda}\right)^2, \quad (3.37)$$

when $r \rightarrow 0$ because $\beta r \ll 1$ and Π_{rad} does not depend on r .

The radial component of the near-field Poynting vector P_r has negative imaginary value and decreases as $1/r^5$:

$$P_r^{near} = -j\frac{\eta}{8}\left(\frac{I\Delta l}{\lambda}\right)^2 \frac{\sin^2 \theta}{\beta^3 r^5}. \quad (3.38)$$

The near-field P_θ component is also imaginary and has the same order of dependence on r but it is positive:

$$P_\theta^{near} = j\eta\beta \frac{(I\Delta l)^2 \cos \theta \sin \theta}{16\pi^2 r^3} \cdot \frac{1}{(\beta r)^2} \quad (3.39)$$

or

$$P_\theta^{near} = j\frac{\eta}{8}\left(\frac{I\Delta l}{\lambda}\right)^2 \frac{\sin(2\theta)}{\beta^3 r^5}. \quad (3.40)$$

4.2. Radiating near-field (Fresnel) region

This is an intermediate region between the reactive near-field region and the far-field region, where the radiation field is more significant but the angular field distribution is still dependent on the distance from the antenna. In this region, $\beta r \geq 1$. For most antennas, it is assumed that the Fresnel region is enclosed between two spherical surfaces:

$$0.62\sqrt{\frac{D^3}{\lambda}} \leq r \leq \frac{2D^2}{\lambda}. \quad (3.41)$$

Here, D is the largest dimension of the antenna. This region is called the *Fresnel region* because its field expressions reduce to Fresnel integrals.

The fields of an infinitesimal dipole in the Fresnel region are obtained by neglecting the higher-order $(1/\beta r)^n$ -terms, $n \geq 2$, in (3.8) and (3.10):

$$\left\{ \begin{array}{l} H_{\varphi} \approx \frac{j\beta(I\Delta l) \cdot e^{-j\beta r}}{4\pi r} \cdot \sin \theta \\ E_r \approx \eta \frac{\beta(I\Delta l) \cdot e^{-j\beta r}}{2\pi\beta r^2} \cdot \cos \theta \\ E_{\theta} \approx j\eta \frac{\beta(I\Delta l) \cdot e^{-j\beta r}}{4\pi r} \cdot \sin \theta \\ H_{\theta} = H_r = E_{\varphi} = 0 \end{array} \right. , \beta r \geq 1. \quad (3.42)$$

The radial component E_r is not negligible yet but the transverse components E_{θ} and H_{φ} are dominant.

4.3. Far-field (Fraunhofer) region

Only the terms $\sim 1/r$ are considered when $\beta r \gg 1$. The angular field distribution does not depend on the distance from the source any more, i.e., the *far-field pattern* is already well established. The field is a transverse EM wave. For most antennas, the far-field region is defined as

$$r \geq 2D^2 / \lambda. \quad (3.43)$$

The far-field of the infinitesimal dipole is obtained as

$$\left\{ \begin{array}{l} H_{\varphi} \approx \frac{j\beta \cdot (I\Delta l) \cdot e^{-j\beta r}}{4\pi r} \cdot \sin \theta \\ E_{\theta} \approx j\eta \frac{\beta \cdot (I\Delta l) \cdot e^{-j\beta r}}{4\pi r} \cdot \sin \theta, \beta r \gg 1. \\ E_r \approx 0 \\ H_{\theta} = H_r = E_{\varphi} = 0 \end{array} \right. \quad (3.44)$$

The features of the far field are summarized below:

- 1) no radial components;
- 2) the angular field distribution is independent of r ;
- 3) $\mathbf{E} \perp \mathbf{H}$;
- 4) $\eta = Z_0 = E_{\theta} / H_{\varphi}$;
- 5) $\mathbf{P} = (\mathbf{E} \times \mathbf{H}^*) / 2 = \hat{\mathbf{r}} 0.5 |E_{\theta}|^2 / \eta = \hat{\mathbf{r}} 0.5 \eta |H_{\varphi}|^2.$ (3.45)

5. Region Separation and Accuracy of the Approximations

In most practical cases, a closed form solution of the radiation integral (the VP integral) does not exist. For the evaluation of the far fields or the fields in the Fraunhofer region, standard approximations are applied, from which the boundaries of these regions are derived.

Consider the VP integral for a linear current source:

$$\mathbf{A} = \frac{\mu}{4\pi} \int_{L'} I(l') \frac{e^{-j\beta R}}{R} d\mathbf{l}', \quad (3.46)$$

where $R = \sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2}$. The observation point is at $P(x, y, z)$, and the source point is at $Q(x', y', z')$, which belongs to the integration contour L' .

So far, we have analyzed the infinitesimal dipole whose current is constant along L' . In practical antennas, the current distribution is not constant and the solution of (3.46) can be very complicated depending on the vector function $I(l')d\mathbf{l}'$. Besides, because of the infinitesimal size of this source, the distance R between the integration point and the observation point was considered constant and equal to the distance from the centre of the dipole, $R \approx r = (x^2 + y^2 + z^2)^{1/2}$. However, if D_{\max} (the maximum dimension of the antenna) is larger and commensurate with the wavelength λ , the error, especially in the phase term βR , due to the above assumption for R would be unacceptable.

Let us divide the integral kernel $e^{-j\beta R} / R$ into two factors: (1) the amplitude factor ($1/R$), and (2) the phase factor $e^{-j\beta R}$. The amplitude factor is not very sensitive to errors in R . In both, the Fresnel and the Fraunhofer regions, the approximation

$$1/R \approx 1/r \quad (3.47)$$

is acceptable, provided $r \gg D_{\max}$.

The above approximation, however, is unacceptable in the phase term. **To keep the phase term error low enough, the maximum error in (βR) must be kept below $\pi/8 = 22.5^\circ$.**

Neglect the antenna dimensions along the x - and y -axes (infinitesimally thin wire). Then,

$$x' = y' = 0 \Rightarrow R = \sqrt{x^2 + y^2 + (z - z')^2}, \quad (3.48)$$

$$\Rightarrow R = \sqrt{x^2 + y^2 + z^2 + (z'^2 - 2zz')} = \sqrt{r^2 + (z'^2 - 2rz' \cdot \cos \theta)}. \quad (3.49)$$

Using the binomial expansion, R is expanded as

$$R = (r^2)^{1/2} + \frac{1}{2}(r^2)^{-1/2} (z'^2 - 2rz' \cos \theta) + \frac{0.5(-0.5)}{2} (r^2)^{-3/2} (z'^2 - 2rz' \cos \theta)^2 + \dots,$$

$$\Rightarrow R = r - z' \cos \theta + \frac{z'^2}{2r} - \frac{z'^2 \cos^2 \theta}{2r} + \frac{1}{2r^2} z'^3 \cos \theta \sin^2 \theta + O^3. \quad (3.50)$$

O^3 denotes terms of the order $(1/r^3)$ and higher. Neglecting these terms and simplifying further leads to the approximation

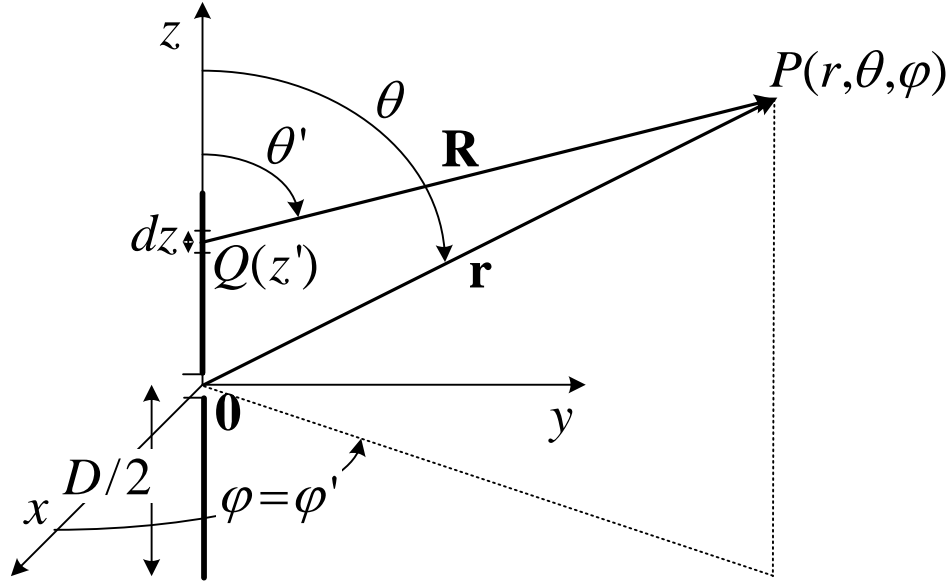
$$R \approx r - z' \cos \theta + \frac{1}{2r} z'^2 \sin^2 \theta + \frac{1}{2r^2} z'^3 \cos \theta \sin^2 \theta. \quad (3.51)$$

This expansion is used below to mathematically define the reactive near-field region, the radiating near-field region and the far-field region.

(a) **Far-field approximation**

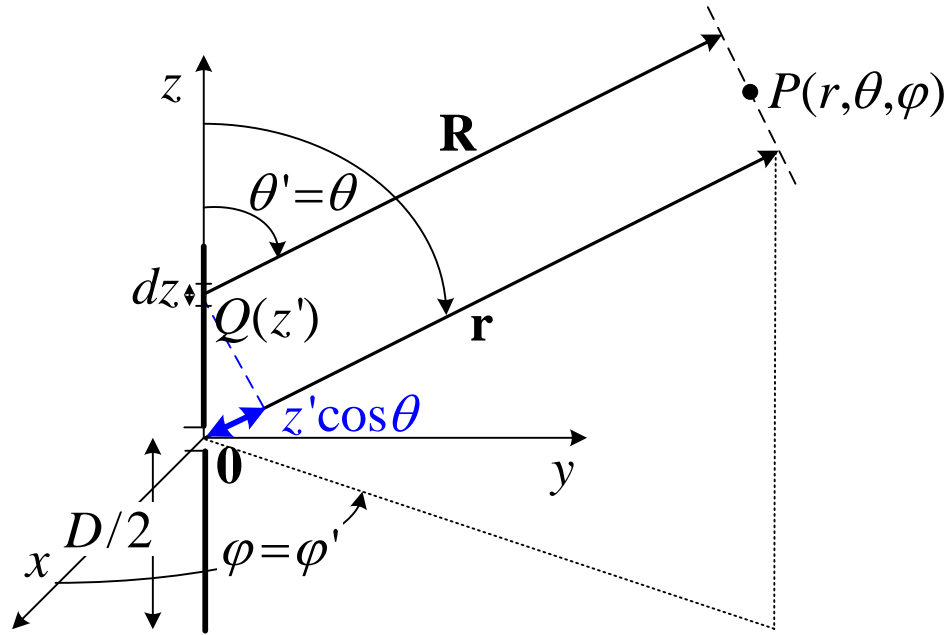
Only the first two terms in the expansion (3.51) are taken into account:

$$R \approx r - z' \cos \theta. \quad (3.52)$$



(a) z -oriented dipole of length D

* $(a + b)^n = a^n + na^{n-1}b + \frac{n(n-1)}{2!} a^{n-2}b^2 + \frac{n(n-1)(n-2)}{3!} a^{n-3}b^3 + \dots$



(b) z -oriented dipole: far-field approximation

The most significant error term in R that was neglected in (3.52) is

$$e(r) = \frac{1}{2} \frac{(z')^2}{r} \sin^2 \theta,$$

which has its maximum at $\theta = \pi/2$ and $z' = z'_{\max} = D/2$:

$$e_{\max}(r) = \frac{(z'_{\max})^2}{2r}. \quad (3.53)$$

The minimum r , at which the phase error (βR) is below $\pi/8$, is derived from:

$$\beta \cdot \frac{(z'_{\max})^2}{2r} \leq \frac{\pi}{8}.$$

Thus, the smallest distance from the antenna centre r , at which the phase error is acceptable is

$$r_{\min}^{\text{far}} = 2D^2 / \lambda. \quad (3.54)$$

This is the far-zone limit defined in (3.43).

As a word of caution, sometimes equation (3.54) produces too small values, which are in conflict with the assumptions made before. For example, in order the amplitude-term approximation $1/R \approx 1/r$ to hold, the ratio of the maximum antenna dimension D and the distance R must fulfill $D/R \ll 1$. Otherwise, the first-order approximation based on the binomial expansion is too inaccurate.

Besides, in order to neglect all field components except the far-field ones, the condition $r \gg \lambda$ must hold, too. Therefore, in addition to (3.54), the

calculated inner boundary of the far-field region should comply with two more conditions:

$$r \gg D \text{ and } r \gg \lambda. \quad (3.55)$$

Finally, we can generalize the far-zone limit as

$$r_{\min}^{\text{far}} = \max(2D^2 / \lambda, \sim 5\lambda, \sim 5D). \quad (3.56)$$

(b) **Radiating near-field (Fresnel region) approximation**

This region is adjacent to the Fraunhofer region, so its upper boundary is specified by

$$r \leq r_{\min}^{\text{far}} = 2D^2 / \lambda. \quad (3.57)$$

When the observation point belongs to this region, we must take one more term in the expansion of R as given by (3.51) to reduce sufficiently the phase error. The approximation this time is

$$R \approx r - z' \cos \theta + \frac{1}{2r} z'^2 \sin^2 \theta. \quad (3.58)$$

The most significant error term is

$$e = \frac{1}{2} \frac{z'^3}{r^2} \cos \theta \sin^2 \theta. \quad (3.59)$$

The angles θ_o must be found, at which e has its extrema:

$$\frac{\partial e}{\partial \theta} = \frac{z'^3}{2r^2} \sin \theta (-\sin^2 \theta + 2\cos^2 \theta) = 0. \quad (3.60)$$

The roots of (3.60) are

$$\begin{aligned} \theta_o^{(1)} &= 0 \rightarrow \text{min}, \\ \theta_o^{(2),(3)} &= \arctan(\pm\sqrt{2}) \approx \pm 54.7^\circ \rightarrow \text{max}. \end{aligned} \quad (3.61)$$

Following a procedure similar to case (a), we obtain:

$$\begin{aligned} \beta e_{\max}(r) &= \frac{2\pi}{\lambda} \cdot \frac{1}{2} \frac{z'^3}{r^2} \cos \theta_o^{(2)} \sin^2 \theta_o^{(2)} = \frac{\pi}{\lambda} \frac{z'^3}{r^2} \frac{1}{\sqrt{3}} \frac{2}{3}, \\ \Rightarrow \beta e_{\max}(r) &= \frac{\pi}{12\sqrt{3}} \frac{D^3}{\lambda r^2} \leq \frac{\pi}{8}, \text{ note: } z' = D/2 \\ \Rightarrow r &\geq \sqrt{\frac{2}{3\sqrt{3}}} \frac{D^3}{\lambda} = 0.62 \sqrt{\frac{D^3}{\lambda}}. \end{aligned} \quad (3.62)$$

Equation (3.62) states the lower boundary of the Fresnel region (for wire antennas) and is identical to the left-hand side of (3.41).

LECTURE 4: Fundamental Antenna Parameters

(Radiation pattern. Pattern beamwidths. Radiation intensity. Directivity. Gain. Antenna efficiency and radiation efficiency. Frequency bandwidth. Input impedance and radiation resistance. Antenna effective area. Relationship between directivity and antenna effective area. Other antenna equivalent areas.)

The antenna parameters describe the antenna performance with respect to space distribution of the radiated energy, power efficiency, matching to the feed circuitry, etc. Many of these parameters are interrelated. There are several parameters not described here, in particular, *antenna temperature and noise characteristics*. They are discussed later in conjunction with radio-wave propagation and system performance.

1. Radiation Pattern

The *radiation pattern* (RP) (or *antenna pattern*) is the representation of the radiation properties of the antenna as a function of the angular coordinates.

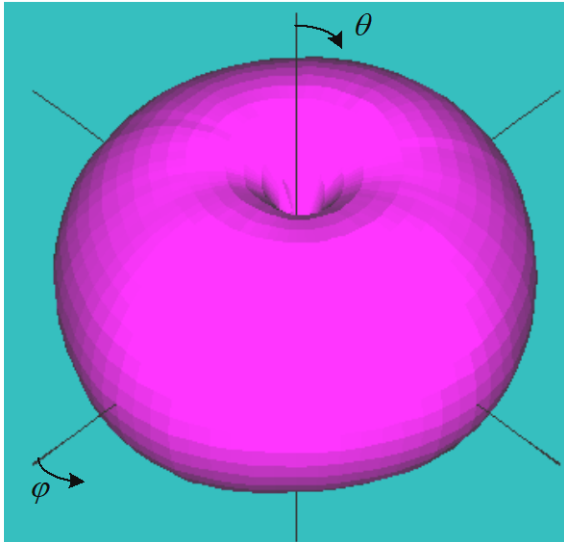
The RP is measured in the far-field region, where the angular distribution of the radiated power does not depend on the distance. We measure and plot either the field intensity, $\sim |\mathbf{E}(\theta, \varphi)|$, or the power $\sim |\mathbf{E}(\theta, \varphi)|^2 / \eta = \eta |\mathbf{H}(\theta, \varphi)|^2$. Usually, the pattern describes the *normalized* field (power) values with respect to the maximum value.

The trace of the angular variation of the magnitude of the electric (or magnetic) field at a constant radius from the antenna is called the *amplitude field pattern*.

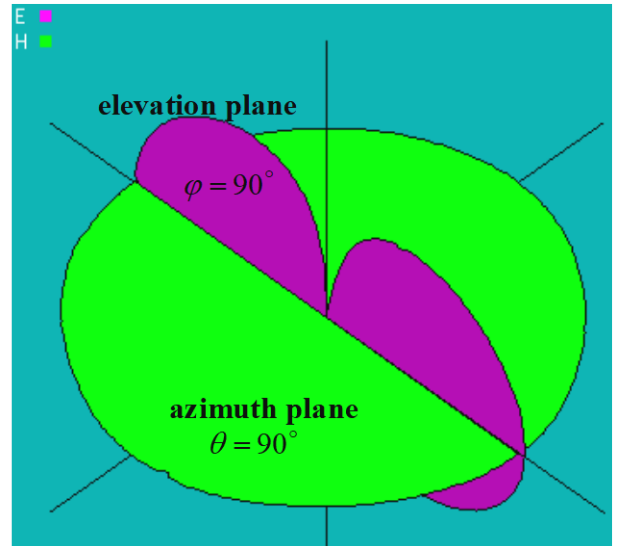
The trace of the angular variation of the received/radiated power at a constant radius from the antenna is called the *power pattern*.

Note: The power pattern and the amplitude field pattern are the same when computed and plotted in dB.

The pattern can be a 3-D plot (both θ and φ vary), or a 2-D plot. A 2-D plot is obtained as an intersection of the 3-D RP with a given plane, usually a $\theta = \text{const.}$ plane or a $\varphi = \text{const.}$ plane that must contain the pattern's maximum.



3-D pattern of a dipole



2-D elevation & azimuth patterns of a dipole

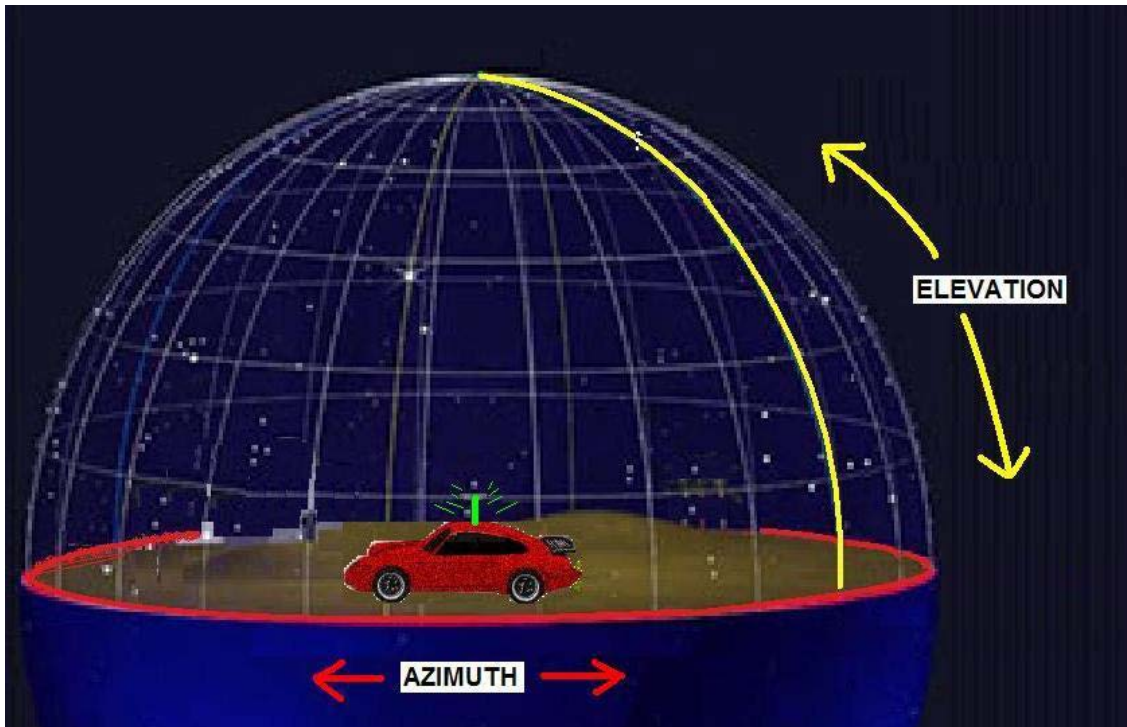
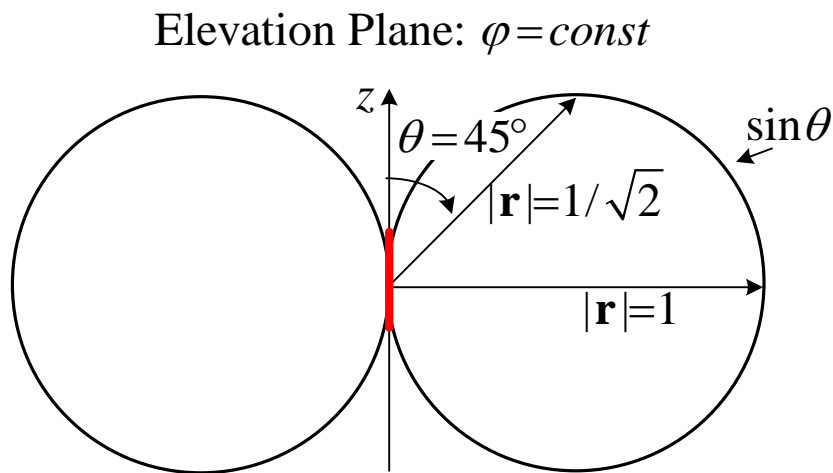


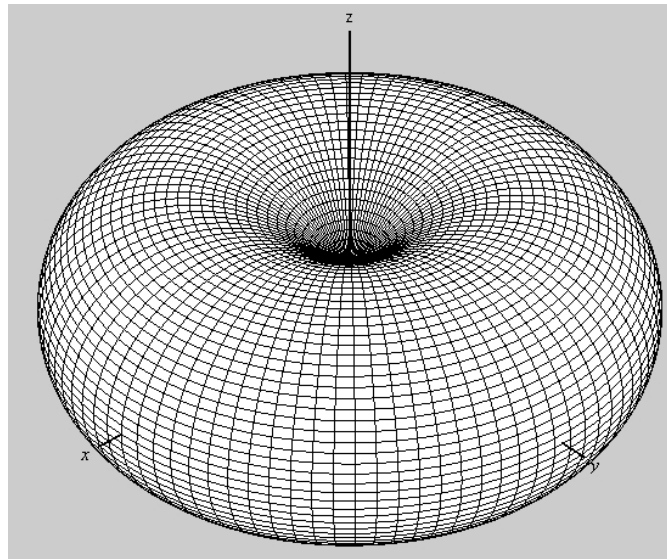
Illustration of azimuth and elevation

Plotting the pattern: the trace of the pattern is obtained by setting the distance from the origin in the direction (θ, φ) to be proportional to the strength of the field $|\mathbf{E}(\theta, \varphi)|$ (in the case of an amplitude field pattern) or proportional to the power density $|\mathbf{E}(\theta, \varphi)|^2$ (in the case of a power pattern).



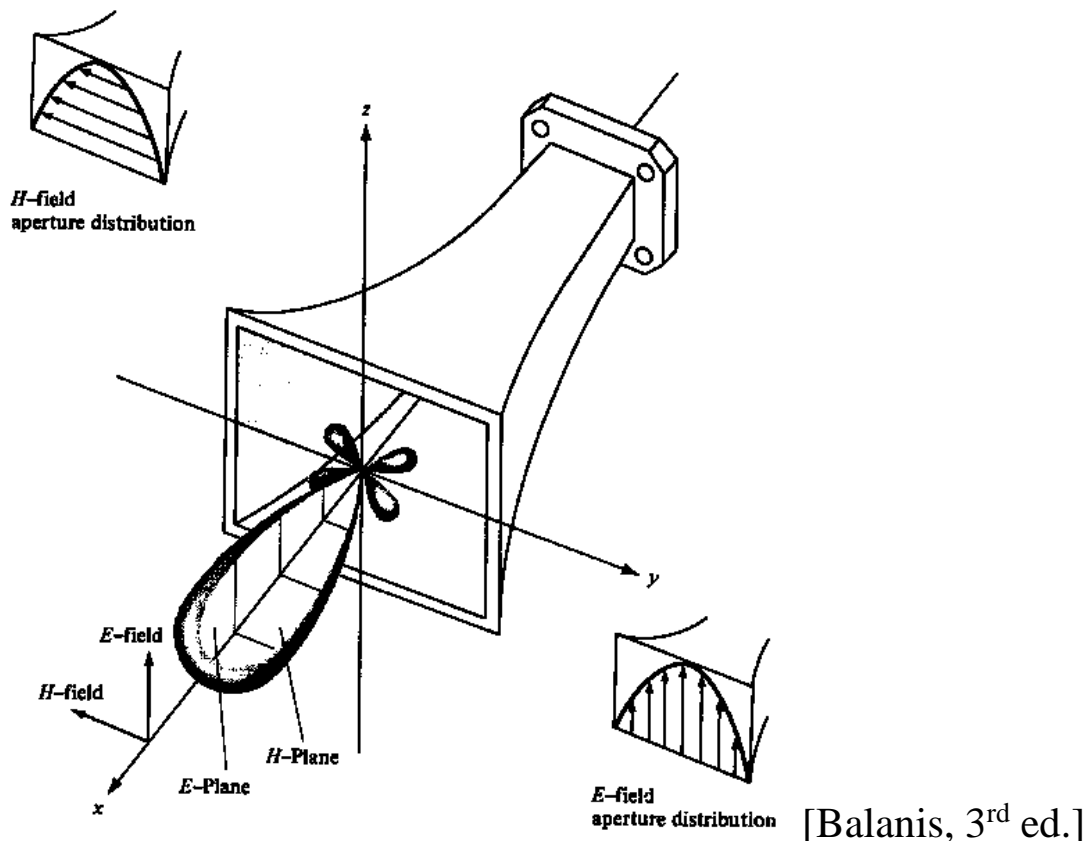
Some concepts related to the pattern terminology

- a) **Isotropic pattern** is the pattern of an antenna having equal radiation in all directions. This is an ideal concept, which, strictly speaking, is achievable only approximately in a narrow frequency band. However, it is used to define other antenna parameters. It is represented simply by a sphere whose center coincides with the location of the isotropic radiator.
- b) **Directional antenna** is an antenna, which radiates (receives) much more efficiently in some directions than in others. Usually, this term is applied to antennas whose directivity is much higher than that of a half-wavelength dipole.
- c) **Omnidirectional antenna** is an antenna, which has a non-directional pattern in a given plane, and a directional pattern in any orthogonal plane (e.g. single-wire antenna). The pattern in the figure below is that of a dipole – it is omnidirectional.

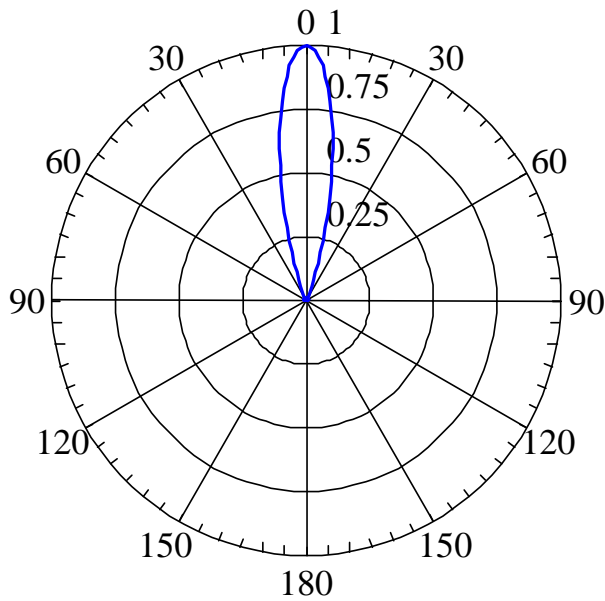


Omnidirectional 3-D pattern

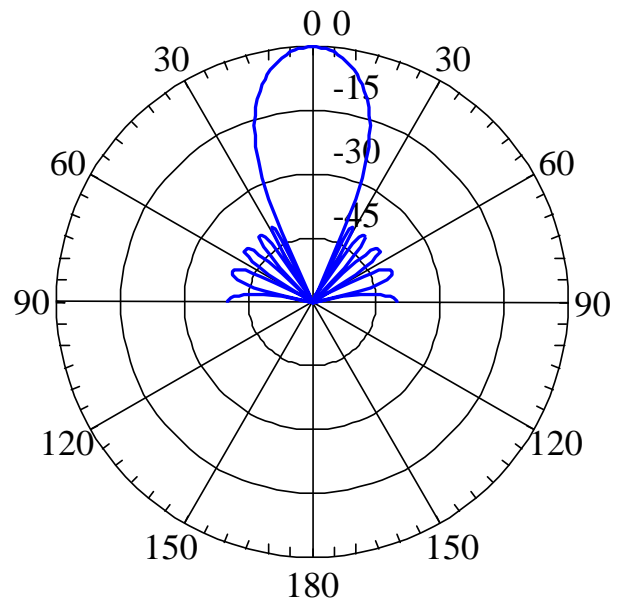
d) **Principal patterns** are the 2-D patterns of linearly polarized antennas, measured in the ***E-plane*** (a plane parallel to the ***E*** vector and containing the direction of maximum radiation) and in the ***H-plane*** (a plane parallel to the ***H*** vector, orthogonal to the ***E-plane***, and containing the direction of maximum radiation).



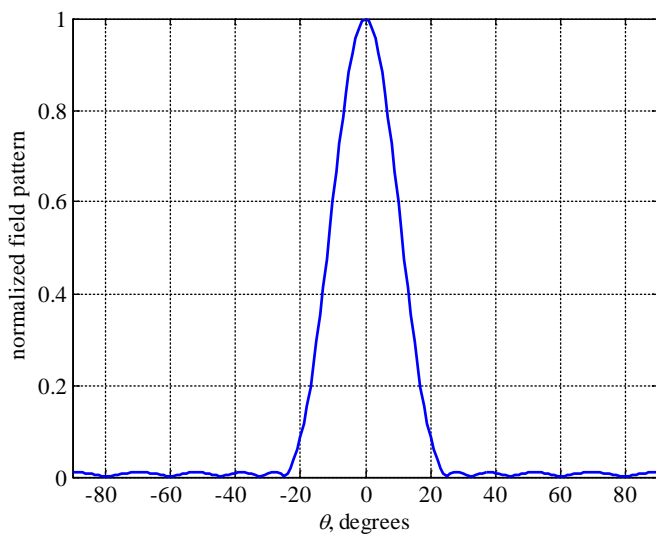
2-D patterns can be *polar* or *rectangular*, depending the way the angle is depicted, and *linear* or *logarithmic* (in dB), depending on the chosen pattern scale. The plots below show the same 2-D pattern in 4 different formats.



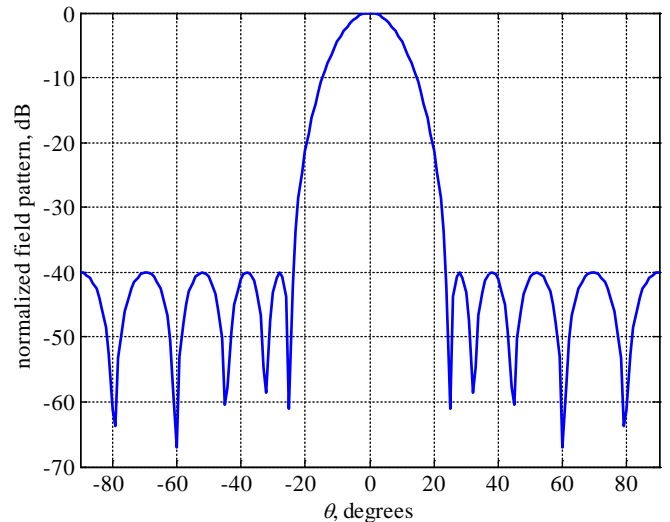
Polar Pattern (linear scale)



Polar Pattern (dB scale, min @ -60 dB)

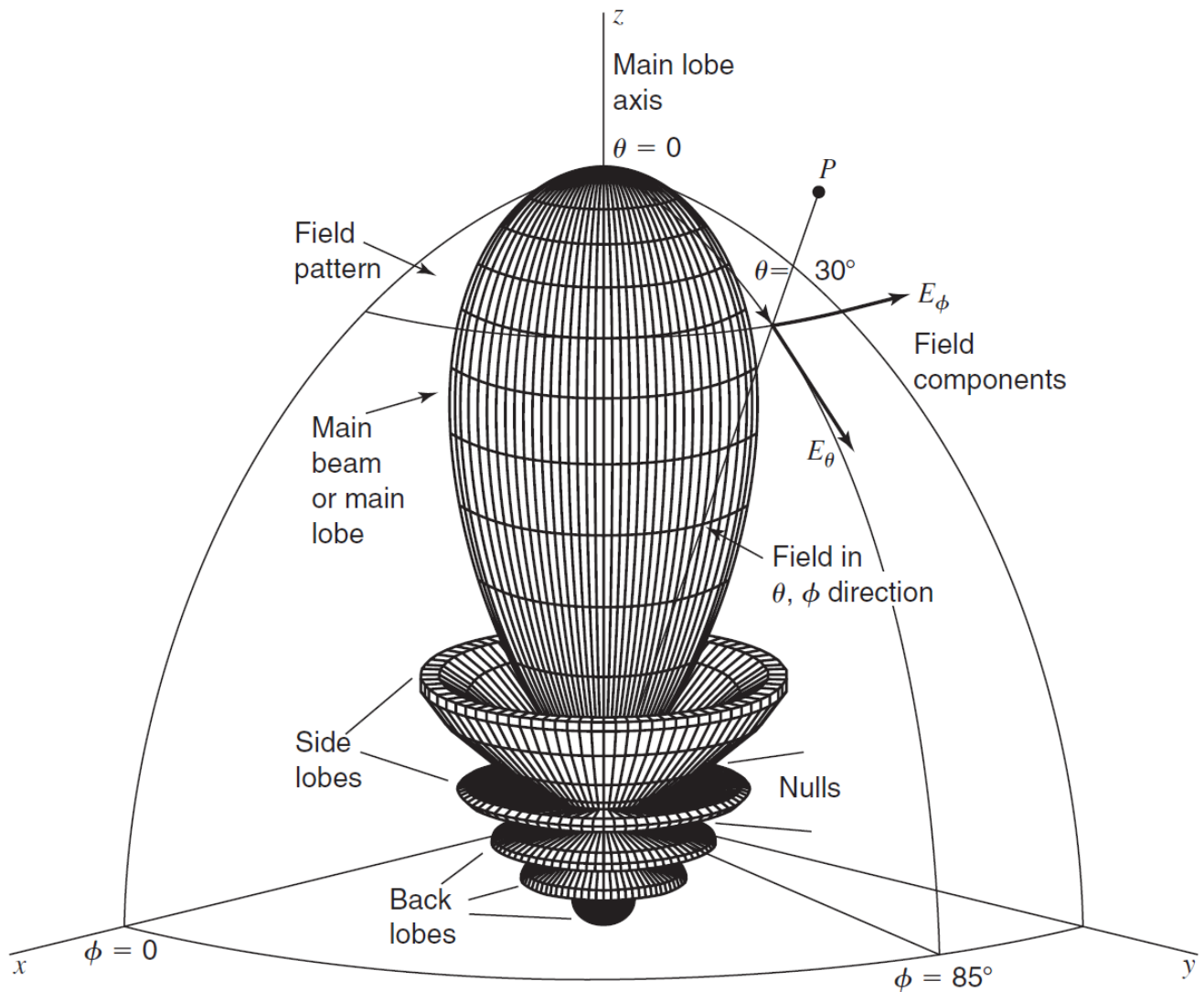


Rectangular Pattern (linear scale)



Rectangular Pattern (dB)

e) **Pattern lobe** is a portion of the RP whose local radiation intensity maximum is relatively weak. Lobes are classified as: major, minor, side lobes, back lobes.

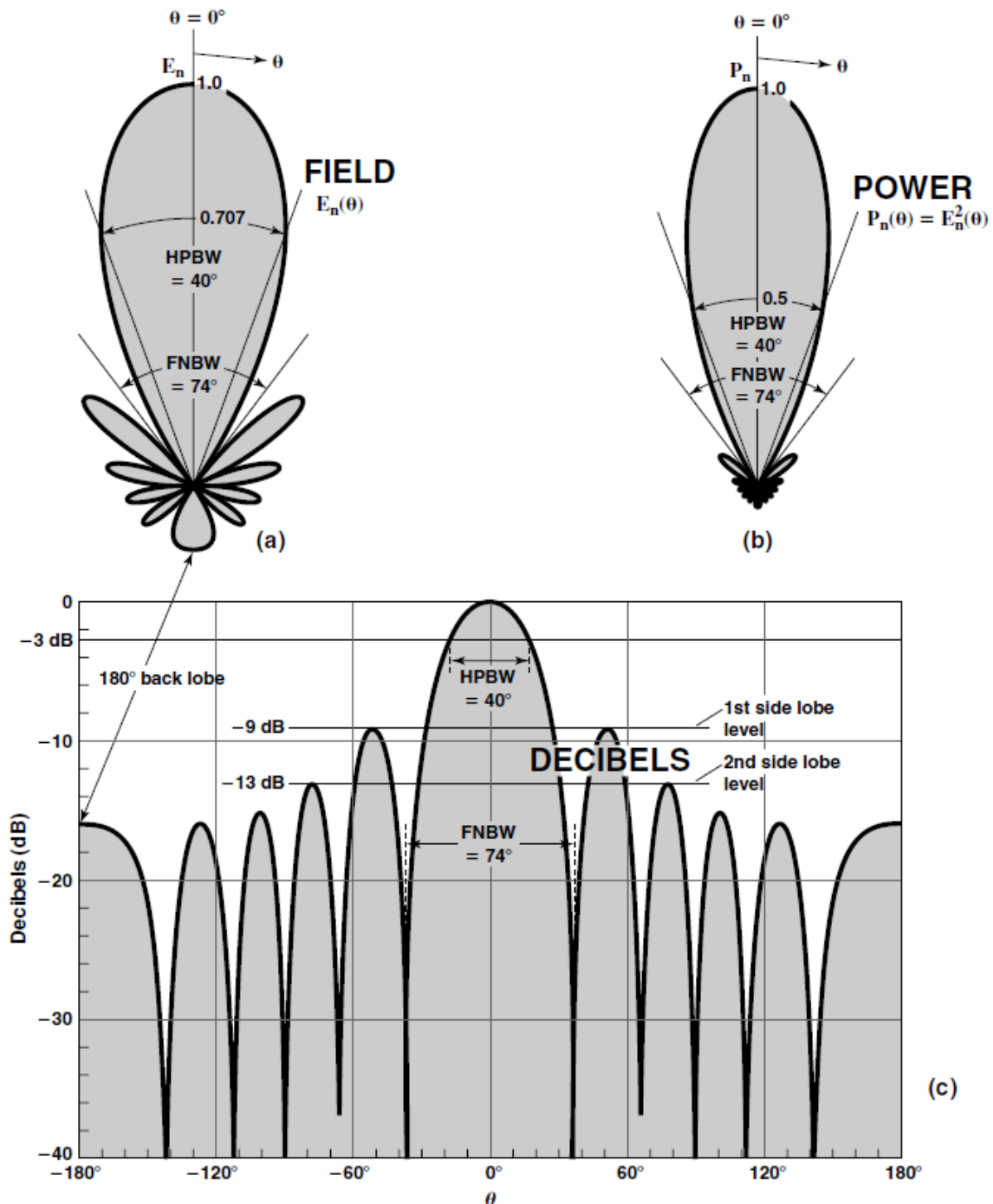


2. Pattern Beamwidth

Half-power beamwidth (HPBW) is the angle between two vectors, originating at the pattern's origin and passing through these points of the major lobe where the radiation intensity is half its maximum.

First-null beamwidth (FNBW) is the angle between two vectors, originating at the pattern's origin and tangent to the main beam at its base. Often, the approximation $FNBW \approx 2 \cdot HPBW$ is used.

The HPBW is the best parameter to describe the antenna resolution properties. In radar technology as well as in radio-astronomy, the antenna resolution capability is of primary importance.



3. Radiation Intensity

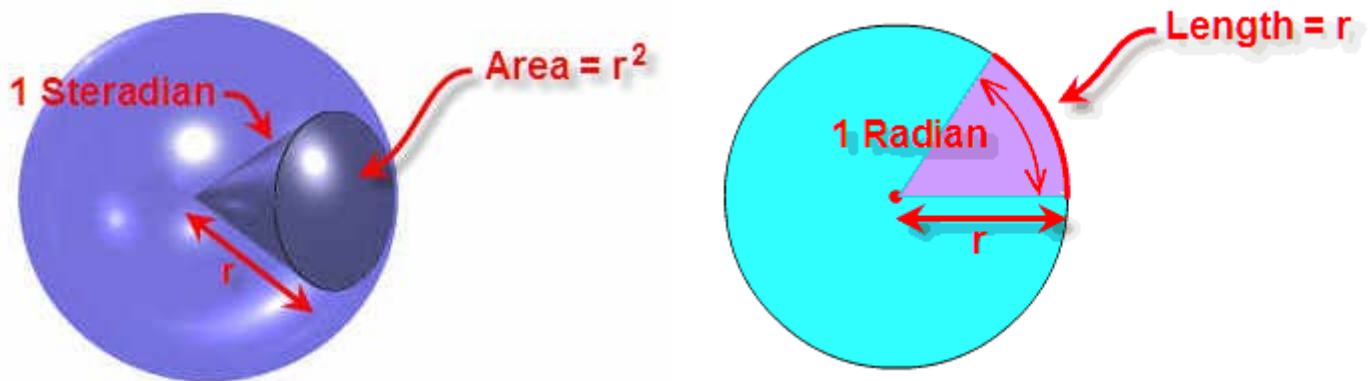
Radiation intensity in a given direction is the power per unit solid angle radiated in this direction by the antenna.

a) Solid angle

One *steradian* (*sr*) is the solid angle with its vertex at the center of a sphere of radius r , which is subtended by a spherical surface of area r^2 . In a closed sphere, there are 4π steradians. A solid angle is defined as

$$\Omega = \frac{S_{\Omega}}{r^2}, \text{ sr} \quad (4.1)$$

Note: The above definition is analogous to the definition of a 2-D angle in radians, $\omega = l_{\omega} / \rho$, where l_{ω} is the length of the arc segment supported by the angle ω in a circle of radius ρ .



The infinitesimal area ds on a surface of a sphere of radius r in spherical coordinates is

$$ds = r^2 \sin \theta d\theta d\varphi, \text{ m}^2. \quad (4.2)$$

Therefore,

$$d\Omega = \sin \theta d\theta d\varphi, \text{ sr}, \quad (4.3)$$

and

$$ds = r^2 d\Omega. \quad (4.4)$$

b) Radiation intensity U

The **radiation intensity** is the power radiated within unit solid angle:

$$U = \lim_{\Delta\Omega \rightarrow 0} \frac{\Delta\Pi_{rad}}{\Delta\Omega} = \frac{d\Pi_{rad}}{d\Omega}, \text{ W/sr.} \quad (4.5)$$

The expression inverse to that in (4.5) is

$$\Pi_{rad} = \oint\oint_{4\pi} U d\Omega, \text{ W.} \quad (4.6)$$

From now on, we will denote the radiated power simply by Π . There is a direct relation between the radiation intensity U and the radiation power density P (that is the Poynting vector magnitude of the far field). Since

$$P = \frac{d\Pi}{ds} = \frac{d\Pi}{r^2 d\Omega} = \frac{1}{r^2} U, \text{ W/m}^2 \quad (4.7)$$

then

$$U = r^2 \cdot P \quad (4.8)$$

It was already shown that the power density of the far field depends on the distance from the source as $1/r^2$, since the far field magnitude depends on r as $1/r$. Thus, the radiation intensity U depends only on the direction (θ, φ) but not on the distance r .

The power pattern is a trace of the function $|U(\theta, \varphi)|$ usually normalized to its maximum value. The normalized pattern will be denoted as $\bar{U}(\theta, \varphi)$.

In the far-field zone, the radial field components vanish, and the remaining \mathbf{E} and \mathbf{H} transverse components are in phase and have magnitudes related by

$$|\mathbf{E}| = \eta |\mathbf{H}|. \quad (4.9)$$

This is why the far-field Poynting vector has only a radial component and it is a real number showing the *radiation power-flow density*:

$$P_{rad} = P = \frac{1}{2} \eta |\mathbf{H}|^2 = \frac{1}{2} \frac{|\mathbf{E}|^2}{\eta}. \quad (4.10)$$

Then, for the *radiation intensity*, we obtain in terms of the electric field

$$U(\theta, \varphi) = \frac{r^2}{2\eta} |\mathbf{E}|^2. \quad (4.11)$$

Equation (4.11) leads to a useful relation between the power pattern and the amplitude field pattern:

$$U(\theta, \varphi) = \frac{r^2}{2\eta} [E_\theta^2(r, \theta, \varphi) + E_\varphi^2(r, \theta, \varphi)] = \frac{1}{2\eta} [E_{\theta_p}^2(\theta, \varphi) + E_{\varphi_p}^2(\theta, \varphi)]. \quad (4.12)$$

Here, $E_{\theta_p}(\theta, \varphi)$ and $E_{\varphi_p}(\theta, \varphi)$ denote the far-zone field patterns for the two orthogonal polarizations.

Examples:

- 1) Radiation intensity and pattern of an isotropic radiator:

$$P(r, \theta, \varphi) = \frac{\Pi}{4\pi r^2}$$

$$U(\theta, \varphi) = r^2 \cdot P = \frac{\Pi}{4\pi} = \text{const.}$$

$$\Rightarrow \underline{\underline{\bar{U}(\theta, \varphi) = 1.}}$$

The normalized pattern of an isotropic radiator is simply a sphere of a unit radius.

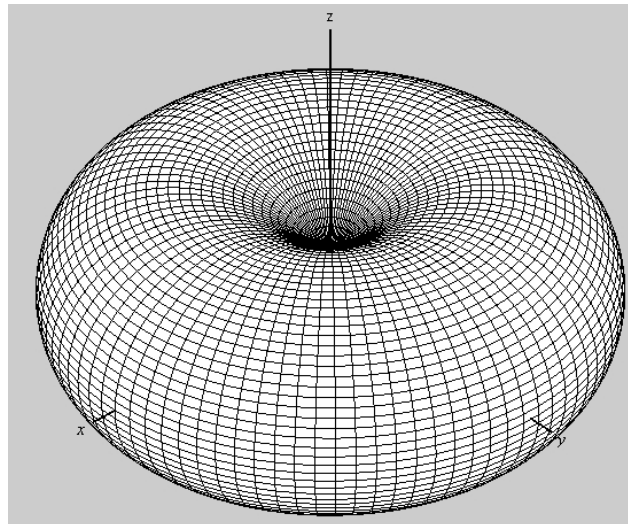
- 2) Radiation intensity and pattern of an infinitesimal dipole:

From Lecture 3, the far-field term of the electric field is:

$$E_\theta = j\eta \frac{\beta \cdot (I\Delta l) \cdot e^{-j\beta r}}{4\pi r} \cdot \sin \theta \Rightarrow \bar{E}(\theta, \varphi) = \sin \theta,$$

$$U = \frac{r^2}{2\eta} \cdot |\mathbf{E}|^2 = \eta \frac{\beta^2 \cdot (I\Delta l)^2}{32\pi^2} \cdot \sin^2 \theta,$$

$$\Rightarrow \underline{\underline{\bar{U}(\theta, \varphi) = \sin^2 \theta.}}$$



4. Directivity

4.1. Definitions and examples

Directivity of an antenna (in a given direction) is the ratio of the radiation intensity in this direction and the radiation intensity averaged over all directions. The radiation intensity averaged over all directions is equal to the total power radiated by the antenna divided by 4π . If a direction is not specified, then the direction of maximum radiation is implied.

It can be also defined as the ratio of the radiation intensity (RI) of the antenna in a given direction and the RI of an isotropic radiator fed by the same amount of power:

$$D(\theta, \varphi) = \frac{U(\theta, \varphi)}{U_{av}} = 4\pi \frac{U(\theta, \varphi)}{\Pi}, \quad (4.13)$$

and

$$D_{\max} = D_0 = 4\pi \frac{U_{\max}}{\Pi}.$$

The directivity is a dimensionless quantity. The maximum directivity is always ≥ 1 .

Examples:

1) Directivity of an isotropic source:

$$\begin{aligned} U(\theta, \varphi) &= U_0 = \text{const.} \\ \Rightarrow \Pi &= 4\pi U_0 \\ \Rightarrow D(\theta, \varphi) &= 4\pi \frac{U(\theta, \varphi)}{\Pi} = 1 \\ \Rightarrow \underline{\underline{D_0}} &= 1. \end{aligned}$$

2) Directivity of an infinitesimal dipole:

$$U(\theta, \varphi) = \eta \frac{\beta^2 \cdot (I\Delta l)^2}{32\pi^2} \cdot \sin^2 \theta$$

$$\Rightarrow \bar{U}(\theta, \varphi) = \sin^2 \theta; \quad U(\theta, \varphi) = M \cdot \bar{U}(\theta, \varphi) = M \sin^2 \theta$$

As shown in (4.6),

$$\Pi = \oint_{4\pi} U d\Omega = M \cdot \int_0^\pi \int_0^{2\pi} \sin^2 \theta \cdot \sin \theta d\varphi d\theta = M \cdot \frac{8\pi}{3}$$

$$D(\theta, \varphi) = 4\pi \frac{U(\theta, \varphi)}{\Pi} = 4\pi \frac{M \sin^2 \theta}{M \cdot 8\pi} \cdot 3 = \frac{3}{2} \sin^2 \theta$$

$$\Rightarrow \underline{\underline{D_0 = 1.5.}}$$

Exercise: Calculate the maximum directivity of an antenna with a radiation intensity $U = M \sin \theta$. (Answer: $D_0 = 4 / \pi \approx 1.27$)

The *partial directivity* of an antenna is specified for a given polarization of the field. It is defined as that part of the radiation intensity, which corresponds to a given polarization, divided by the total radiation intensity averaged over all directions.

The total directivity is the sum of the partial directivities for any two orthogonal polarizations:

$$D = D_\theta + D_\varphi, \quad (4.14)$$

where:

$$D_\theta = 4\pi \frac{U_\theta}{\Pi_\theta + \Pi_\varphi},$$

$$D_\varphi = 4\pi \frac{U_\varphi}{\Pi_\theta + \Pi_\varphi}.$$

4.2. Directivity in terms of normalized radiation intensity $\bar{U}(\theta, \varphi)$

$$U(\theta, \varphi) = M \cdot \bar{U}(\theta, \varphi) \quad (4.15)$$

$$\Pi = \oint\limits_{4\pi} U d\Omega = M \cdot \int_0^\pi \int_0^{2\pi} \bar{U}(\theta, \varphi) \sin\theta d\varphi d\theta \quad (4.16)$$

$$D(\theta, \varphi) = 4\pi \frac{\bar{U}(\theta, \varphi)}{\int_0^\pi \int_0^{2\pi} \bar{U}(\theta', \varphi') \sin\theta' d\varphi' d\theta'} \quad (4.17)$$

$$D_0 = 4\pi \frac{1}{\int_0^\pi \int_0^{2\pi} \bar{U}(\theta, \varphi) \sin\theta d\varphi d\theta} \quad (4.18)$$

4.3. Beam solid angle Ω_A

The **beam solid angle** Ω_A of an antenna is the solid angle through which all the power of the antenna would flow if its radiation intensity were constant and equal to the maximum radiation intensity U_0 for all angles within Ω_A .

$$\Omega_A = \int_0^\pi \int_0^{2\pi} \bar{U}(\theta, \varphi) \sin\theta d\varphi d\theta \quad (4.19)$$

The relation between the maximum directivity and the beam solid angle is obvious from (4.18) and (4.19):

$$D_0 = 4\pi / \Omega_A. \quad (4.20)$$

In order to understand how (4.19) is obtained, follow the derivations below (they reflect the mathematical meaning of the definition above):

$$\begin{aligned} \Pi &= \oint\limits_{4\pi} U d\Omega = \iint\limits_{\Omega_A} U_0 d\Omega = U_0 \Omega_A \\ \Rightarrow \Omega_A &= \frac{\oint\limits_{4\pi} U d\Omega}{U_0} = \oint\limits_{4\pi} \bar{U} d\Omega = \int_0^\pi \int_0^{2\pi} \bar{U}(\theta, \varphi) \sin\theta d\varphi d\theta. \end{aligned}$$

4.4. Approximate expressions for directivity

The complexity of the calculation of the antenna directivity D_0 depends on the power pattern $\bar{U}(\theta, \varphi)$, which has to be integrated over a spherical surface. In most practical cases, this function is not available in closed analytical form (e.g., it might be a data set). Even if it is available in closed analytical form, the integral in (4.18) may not have a closed analytical solution. In practice, simpler although not exact expressions are often used for approximate and fast calculations. These formulas are based on the two orthogonal-plane half-power beamwidths (HPBW) of the pattern. The approximations for the directivity are usually valid for highly directive (pencil-beam) antennas such as large reflectors and horns.

a) *Kraus' formula*

For antennas with narrow major lobe and with negligible minor lobes, the beam solid angle Ω_A is approximately equal to the product of the HPBWs in two orthogonal planes:

$$\Omega_A = \Theta_1 \Theta_2, \quad (4.21)$$

where the HPBW angles are in radians. Another variation of (4.21) is

$$D_0 \approx \frac{41000}{\Theta_1^\circ \Theta_2^\circ}, \quad (4.22)$$

where Θ_1° and Θ_2° are in degrees.

b) *Formula of Tai and Pereira*

$$D_0 \approx \frac{32 \ln 2}{\Theta_1^2 + \Theta_2^2} \quad (4.23)$$

The angles in (4.23) are in radians. For details see: C. Tai and C. Pereira, "An approximate formula for calculating the directivity of an antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-24, No. 2, March 1976, pp. 235-236.

5. Antenna Gain

The gain G of an antenna is the ratio of the radiation intensity U in a given direction and the radiation intensity that would be obtained, if the power fed to the antenna were radiated isotropically.

$$G(\theta, \varphi) = 4\pi \frac{U(\theta, \varphi)}{\Pi_{in}} \quad (4.24)$$

The gain is a dimensionless quantity, which is very similar to the directivity D . When the antenna has no losses, i.e. when $\Pi_{in} = \Pi$, then $G(\theta, \varphi) = D(\theta, \varphi)$. Thus, the gain of the antenna takes into account the losses in the antenna system. It is calculated using the *input power* Π_{in} , which can be measured directly. In contrast, the directivity is calculated via the radiated power Π .

There are many factors that can worsen the transfer of energy from the transmitter to the antenna (or from the antenna to the receiver):

- mismatch losses,
- losses in the transmission line,
- losses in the antenna: dielectric losses, conduction losses, polarization losses.

The power radiated by the antenna is always less than the power fed to it, i.e., $\Pi \leq \Pi_{in}$, unless the antenna has integrated active devices. That is why, usually, $G \leq D$.

! According to the IEEE Standards, the gain does not include losses arising from impedance mismatch and from polarization mismatch.

Therefore, the gain takes into account only the dielectric and conduction losses of the antenna itself.

The radiated power Π is related to the input power Π_{in} through a coefficient called the *radiation efficiency* e :

$$\Pi = e \cdot \Pi_{in}, \quad e \leq 1, \quad (4.25)$$

$$\Rightarrow G(\theta, \varphi) = e \cdot D(\theta, \varphi). \quad (4.26)$$

Partial gains with respect to a given field polarization are defined in the same way as it is done with the partial directivities; see equation (4.14).

6. Antenna Efficiency

The total efficiency of the antenna e_t is used to estimate the total loss of energy at the input terminals of the antenna and within the antenna structure. It includes all mismatch losses and the dielectric/conduction losses (described by the *radiation efficiency* e as defined by the IEEE Standards):

$$e_t = e_p e_r \underbrace{e_c e_d}_e = e_p e_r \cdot e. \quad (4.27)$$

Here: e_r is the reflection (impedance mismatch) efficiency,
 e_p is the polarization mismatch efficiency,
 e_c is the conduction efficiency,
 e_d is the dielectric efficiency.

The reflection efficiency can be calculated through the reflection coefficient Γ at the antenna input:

$$e_r = 1 - |\Gamma|^2. \quad (4.28)$$

Γ can be either measured or calculated, provided the antenna impedance is known:

$$\Gamma = \frac{Z_{in} - Z_c}{Z_{in} + Z_c}. \quad (4.29)$$

Z_{in} is the antenna input impedance and Z_c is the characteristic impedance of the feed line. If there are no polarization losses, then the total efficiency is related to the radiation efficiency as

$$e_t = e \cdot (1 - |\Gamma|^2). \quad (4.30)$$

7. Beam Efficiency

The *beam efficiency* is the ratio of the power radiated in a cone of angle $2\Theta_1$ and the total radiated power. The angle $2\Theta_1$ can be generally any angle, but usually this is the first-null beam width.

$$BE = \frac{\int_0^{2\pi} \int_0^{\Theta_1} U(\theta, \varphi) \sin \theta d\theta d\varphi}{\int_0^{2\pi} \int_0^{\pi} U(\theta, \varphi) \sin \theta d\theta d\varphi} \quad (4.31)$$

If the antenna has its major lobe directed along the z -axis ($\theta = 0$), formula (4.31) defines the main beam efficiency. If Θ_1 is the angle where the first null (or minimum) occurs in two orthogonal planes, then the BE will show what part of the total radiated power is channeled through the main beam.

Very high beam-efficiency antennas are needed in radars, radiometry and astronomy.

8. Frequency Bandwidth (FBW)

This is the range of frequencies, within which the antenna characteristics (input impedance, pattern) conform to certain specifications.

Antenna characteristics, which should conform to certain requirements, might be: input impedance, radiation pattern, beamwidth, polarization, side-lobe level, gain, beam direction and width, radiation efficiency. Separate bandwidths may be introduced: impedance bandwidth, pattern bandwidth, etc.

The FBW of broadband antennas is expressed as the ratio of the upper to the lower frequencies, where the antenna performance is acceptable:

$$\text{FBW} = f_{\max} / f_{\min} . \quad (4.32)$$

Broadband antennas with FBW as large as 40:1 have been designed. Such antennas are referred to as *frequency independent antennas*.

For narrowband antennas, the FBW is expressed as a percentage of the frequency difference over the center frequency:

$$\text{FBW} = \frac{f_{\max} - f_{\min}}{f_0} \cdot 100 \% . \quad (4.33)$$

Usually, $f_0 = (f_{\max} + f_{\min}) / 2$ or $f_0 = \sqrt{f_{\max} f_{\min}}$.

9. Input Impedance

$$Z_A = R_A + jX_A \quad (4.34)$$

Here, R_A is the antenna resistance and X_A is the antenna reactance. Generally, the antenna resistance has two terms:

$$R_A = R_r + R_l, \quad (4.35)$$

where R_r is the radiation resistance and R_l is the loss resistance.

The antenna impedance is related to the radiated power $\Pi \equiv \Pi_{rad}$, the dissipated (loss) power Π_l , and the stored reactive energy as:

$$Z_A = \frac{\Pi_{rad} + \Pi_l + 2j\omega(W_m - W_e)}{0.5I_0I_0^*}. \quad (4.36)$$

Here, I_0 is the current phasor at the antenna terminals; W_m is the time-average magnetic energy, and W_e is the time-average electric energy stored in the near-field region. When the stored magnetic and electric energy values are equal, a condition of resonance occurs and the reactive part of Z_A vanishes. For a thin dipole antenna, this occurs when the antenna length is close to a multiple of a half wavelength.

9.1. Radiation resistance

The radiation resistance relates the radiated power to the voltage (or current) at the antenna terminals. For example, in the Thevenin equivalent of the antenna, the following holds:

$$R_r = 2\Pi / |I|^2, \quad \Omega. \quad (4.37)$$

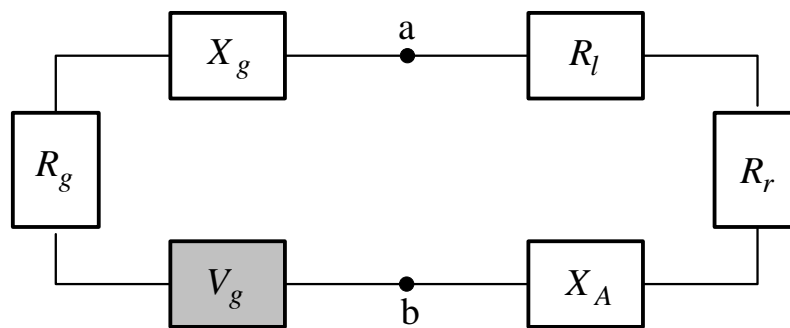
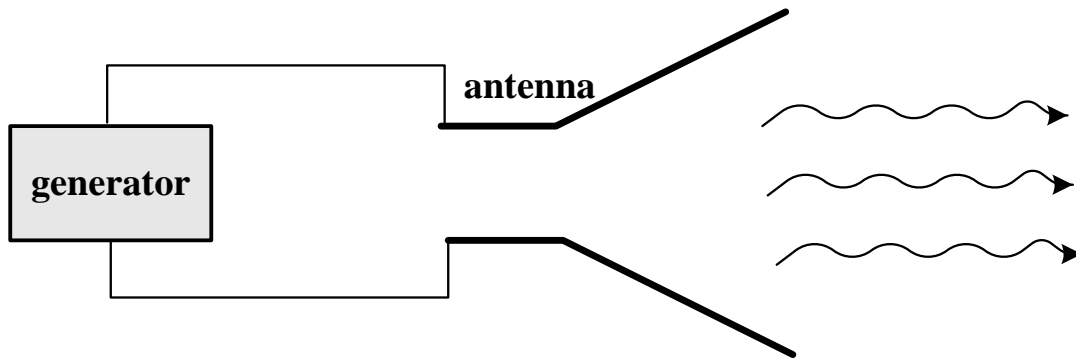
Example: Find the radiation resistance of an infinitesimal dipole in terms of the ratio $(\Delta l / \lambda)$.

We have already derived the radiated power of an infinitesimal dipole in Lecture 3, as:

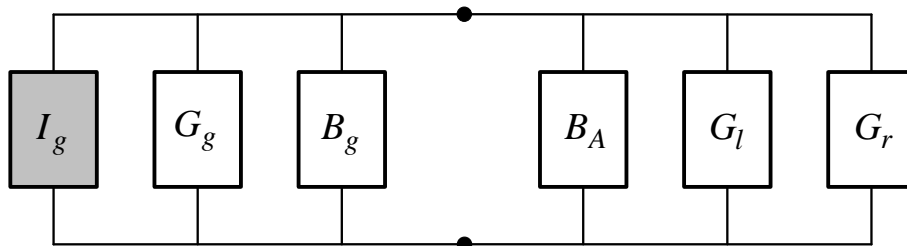
$$\Pi^{id} = \eta \frac{\pi}{3} \left(\frac{I\Delta l}{\lambda} \right)^2 \quad (4.38)$$

$$\Rightarrow R_r^{id} = \eta \frac{2\pi}{3} \left(\frac{\Delta l}{\lambda} \right)^2. \quad (4.39)$$

9.2. Equivalent circuits of the transmitting antenna



(a) Thevenin equivalent



(b) Norton equivalent

In the above model, it is assumed that the generator is connected to the antenna directly. If there is a transmission line between the generator and the antenna, which is usually the case, then $Z_g = R_g + jX_g$ represents the equivalent impedance of the generator transferred to the input terminals of the antenna. Transmission lines themselves often have significant losses.

Reminder: The impedance transformation by a long transmission line is given by

$$Z_{in} = Z_{in} = Z_0 \frac{Z_L + Z_0 \tanh(\gamma L)}{Z_0 + Z_L \tanh(\gamma L)}. \quad (4.40)$$

Here, Z_0 is the characteristic impedance of the line, γ is its propagation constant, Z_L is the load impedance, and Z_{in} is the input impedance. In the case of a loss-free line,

$$Z_{in} = Z_0 \frac{Z_L + jZ_0 \tan(\beta L)}{Z_0 + jZ_L \tan(\beta L)}, \quad (4.41)$$

where $\gamma = j\beta$.

Maximum power is delivered to the antenna when conjugate matching of impedances is achieved:

$$\begin{cases} R_A = R_l + R_r = R_g, \\ X_A = -X_g. \end{cases} \quad (4.42)$$

Using circuit theory, we can derive the following formulas in the case of matched impedances:

a) power delivered to the antenna

$$P_A = \frac{|V_g|^2}{8(R_r + R_l)} \quad (4.43)$$

b) power dissipated as heat in the generator

$$P_g = P_A = \frac{|V_g|^2}{8R_g} = \frac{|V_g|^2}{8(R_r + R_l)} \quad (4.44)$$

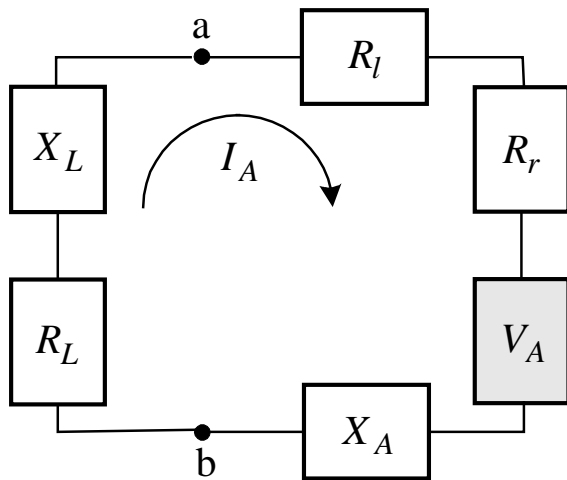
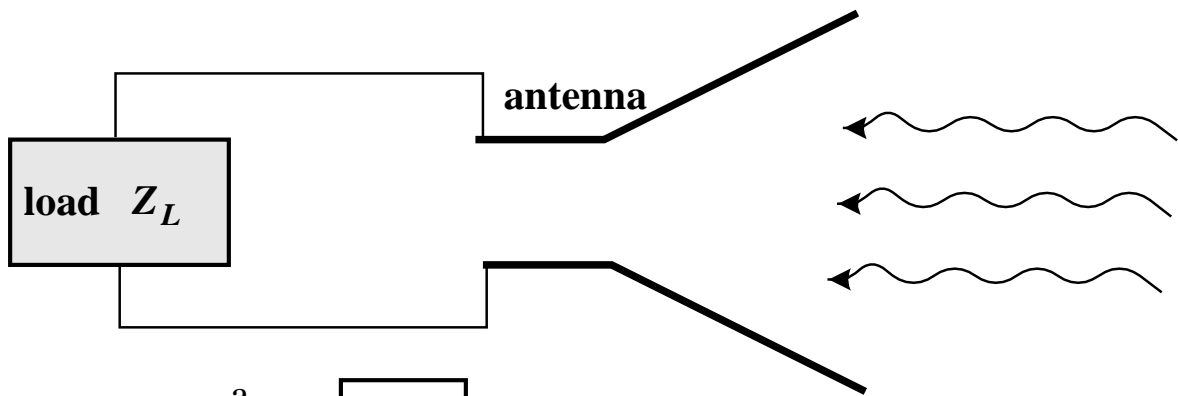
c) radiated power

$$\Pi = P_r = \frac{|V_g|^2}{8} \frac{R_r}{(R_r + R_l)^2} \quad (4.45)$$

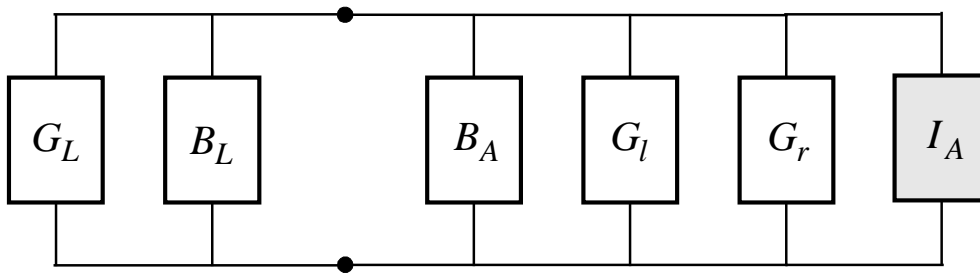
d) power dissipated as heat in the antenna

$$P_l = \frac{|V_g|^2}{8} \frac{R_l}{(R_r + R_l)^2}. \quad (4.46)$$

9.3. Equivalent circuits of the receiving antenna



(a) Thevenin equivalent



(b) Norton equivalent

The incident wave induces voltage V_A at the antenna terminals (measured when the antenna is open circuited). Conjugate impedance matching is required between the antenna and the load (the receiver) to achieve maximum power delivery:

$$\begin{cases} R_L = R_A = R_l + R_r, \\ X_L = -X_A. \end{cases} \quad (4.47)$$

For the case of conjugate matching, the following power expressions hold:

a) power delivered to the load

$$P_L = \frac{|V_A|^2}{8R_L} = \frac{|V_A|^2}{8R_A} \quad (4.48)$$

b) power dissipated as heat in the antenna

$$P_l = \frac{|V_A|^2}{8} \frac{R_l}{R_A^2} \quad (4.49)$$

c) scattered (re-radiated) power

$$P_r = \frac{|V_A|^2}{8} \frac{R_r}{R_A^2} \quad (4.50)$$

d) total captured power

$$P_c = \frac{|V_A|^2}{4(R_r + R_l)} = \frac{|V_A|^2}{4R_A} \quad (4.51)$$

When conjugate matching is achieved, half of the captured power P_c is delivered to the load (the receiver) and half is antenna loss. The antenna losses are heat dissipation P_l and reradiated (scattered) power P_r . When the antenna is non-dissipative, half of the power is delivered to the load and the other half is scattered back into space. Thus a receiving antenna is also a scatterer.

! The antenna input impedance is frequency dependent. Thus, it is matched to its load in a certain frequency band. It can be influenced by the proximity of objects, too.

9.4. Radiation efficiency and antenna losses

The radiation efficiency e takes into account the conductor and dielectric (heat) losses of the antenna. It is the ratio of the power radiated by the antenna and the total power delivered to the antenna terminals (in transmitting mode). In terms of equivalent circuit parameters,

$$e = \frac{R_r}{R_r + R_l}. \quad (4.52)$$

Some useful formulas to calculate conduction losses are given below:

a) dc resistance per unit length

$$R'_{dc} = \frac{1}{\sigma A}, \quad \Omega/\text{m} \quad (4.53)$$

σ - specific conductivity, S/m
 A – conductor's cross-section, m².

b) high-frequency surface resistance

At high frequencies, the current is confined in a thin layer at the conductor's surface (skin effect). This thin layer, called the skin layer, has much smaller cross-section than that of the conductor itself. Its effective thickness, known as the *skin depth* or *penetration depth*, is

$$\delta = \frac{1}{\sqrt{\pi f \sigma \mu}}, \text{ m}, \quad (4.54)$$

where f is the frequency in Hz, and μ is the magnetic permeability, H/m. Remember that (4.54) holds for very good conductors only ($\sigma / \omega \epsilon \gg 1$). The skin depth is inverse proportional to the attenuation constant $\alpha = \text{Re}(\gamma)$ of the conducting medium, $\delta = 1 / \alpha$. Here, $\gamma = j\omega\sqrt{\mu\epsilon}$, $\epsilon = \epsilon' - j(\epsilon'' + \sigma / \omega)$. Due to the exponential decay of the current density in the conductor as $\sim e^{-\alpha x}$, where x denotes the distance from the surface, it can be shown that the total current I flowing along the conductor (along z) is

$$I = \iint_S \mathbf{J} \cdot d\mathbf{s} = \int_C \int_0^\infty J_0 e^{-\alpha x} dx dc = \frac{1}{\alpha} \int_C J_0 dc = \delta \int_C J_0 dc = \int_C J_s dc \quad (4.55)$$

where J_0 is the current density at the conductor surface (in A/m²), $J_s = J_0 \delta$ is the equivalent surface current density (in A/m), and C is the contour of the conductor's cross-section. If the equivalent surface current density is distributed uniformly on the contour of the conductor's cross-section, then $I = J_s p$, where p is the perimeter of the conductor (or the length of its cross-sectional contour).

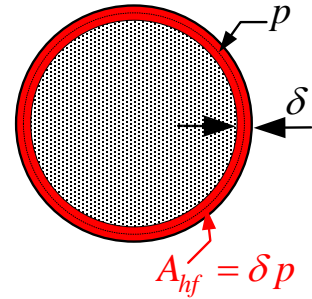
The surface resistance R_s (in Ω) is defined as the real part of the intrinsic impedance of the conductor η_c , which in the case of very good conductors can be found to be

$$R_s = \text{Re} \eta_c \approx \sqrt{\frac{\mu \omega}{2\sigma}} = \frac{1}{\sigma \delta}, \Omega . \quad (4.56)$$

For the case where the current density is uniformly distributed on the conductor's cross-sectional contour, we can find a simple relation between the high-frequency resistance per unit length of a conducting rod, its perimeter p and its surface resistance R_s .

$$R'_{hf} = \frac{1}{\sigma A_{hf}} = \frac{1}{\sigma \delta p} = \frac{R_s}{p}, \Omega/\text{m}. \quad (4.57)$$

Here the area $A_{hf} = \delta p$ is not the actual area of the conducting rod but the effective area through which the high-frequency current flows.



If the surface current distribution is not uniform over the contour of the conductor's cross-section, R'_{hf} appears as a function of R_s and this distribution. The surface density of the loss power in a good conductor is

$$p_\ell = \frac{1}{2} R_s |\mathbf{J}_s|^2 \text{ W/m}^2. \quad (4.58)$$

Then, the power loss per unit length is

$$P'_\ell = \frac{R_s}{2} \int_C |\mathbf{J}_s|^2 dc = \frac{1}{2} R'_{hf} I^2 = \frac{1}{2} R'_{hf} \left(\int_C |\mathbf{J}_s| dc \right)^2 \text{ W/m}. \quad (4.59)$$

It then follows that

$$R'_{hf} = R_s \frac{\int_C |\mathbf{J}_s|^2 dc}{\left(\int_C |\mathbf{J}_s| dc \right)^2} \Omega/\text{m}. \quad (4.60)$$

The above expression reduces to (4.57) if \mathbf{J}_s is constant over C .

Example: A half-wavelength dipole is made of copper ($\sigma = 5.7 \times 10^7 \text{ S/m}$). Determine the radiation efficiency e , if the operating frequency is $f = 100 \text{ MHz}$, the radius of the wire is $b = 3 \times 10^{-4} \cdot \lambda$, and the radiation resistance is $R_r = 73 \Omega$.

$$f = 10^8 \text{ Hz} \Rightarrow \lambda = \frac{c}{f} \approx 3 \text{ m} \Rightarrow l = \frac{\lambda}{2} \approx 1.5 \text{ m}$$

$$p = 2\pi b = 18\pi \times 10^{-4}, \text{ m}$$

If the current along the dipole were uniform, the high-frequency loss power would be uniformly distributed along the dipole. However, the current has a sine distribution along a dipole as we will discuss in Lecture 9:

$$I(z) = I_0 \sin \left[\beta \left(\frac{l}{2} - |z| \right) \right], \quad -\frac{l}{2} \leq z \leq \frac{l}{2}.$$

Equation (4.57) can be now used to express the high-frequency loss resistance per wire differential element of infinitesimal length dz :

$$dR_{hf} = R'_{hf} dz = \frac{dz}{p} \sqrt{\frac{\pi f \mu_0}{\sigma}}.$$

The high-frequency loss power per wire element of infinitesimal length dz is then obtained as

$$dP_{hf}(z) = \frac{1}{2} I^2(z) \cdot \frac{dz}{p} \sqrt{\frac{\pi f \mu_0}{\sigma}}$$

The total loss power is obtained by integrating along the whole dipole. Taking into account the symmetry in the current distribution,

$$\begin{aligned} P_{hf} &= 2 \int_0^{l/2} \frac{I_0^2}{2} \sin^2 \left[\beta \left(\frac{l}{2} - z \right) \right] \cdot \frac{1}{p} \sqrt{\frac{\pi f \mu_0}{\sigma}} dz, \\ \Rightarrow P_{hf} &= \frac{I_0^2}{p} \sqrt{\frac{\pi f \mu_0}{\sigma}} \cdot \int_0^{l/2} \sin^2 \left[\beta \left(\frac{l}{2} - z \right) \right] dz. \end{aligned}$$

Changing variable as

$$x = \beta \left(\frac{l}{2} - z \right)$$

results in

$$\begin{aligned} P_{hf} &= I_0^2 \underbrace{\left(\frac{l}{p} \sqrt{\frac{\pi f \mu_0}{\sigma}} \right)}_{R_{hf}} \cdot \frac{1}{l} \cdot \frac{2}{\beta} \int_0^{\beta l/2} \frac{1 - \cos 2x}{2} dx, \\ \Rightarrow P_{hf} &= I_0^2 R_{hf} \cdot \frac{1}{l} \cdot \frac{l}{4} \left[1 - \frac{\sin(\beta l)}{\beta l} \right] = \frac{I_0^2 R_{hf}}{4} \left[1 - \frac{\sin(\beta l)}{\beta l} \right]. \end{aligned}$$

In the case of $l = \lambda / 2$,

$$\Rightarrow P_{hf} = \frac{I_0^2 R_{hf}}{4}$$

Since the loss resistance R_l is defined through the loss power as

$$P_{hf} = \frac{I_0^2}{2} R_l,$$

we obtain that

$$R_l = 0.5 \cdot R'_{hf} l = 0.5 \frac{l}{p} \sqrt{\frac{\pi f \mu_0}{\sigma}} = 0.349 \ \Omega.$$

The antenna efficiency is:

$$e = \frac{R_r}{R_r + R_l} = \frac{73}{73 + 0.349} = 0.9952 \ (99.52\%)$$

$$e_{[\text{dB}]} = 10 \log_{10} 0.9952 = -0.02.$$

10. Effective Area (Effective Aperture) A_e

The *effective antenna aperture* is the ratio of the available power at the terminals of the antenna to the power flux density of a plane wave incident upon the antenna, which is matched to the antenna in terms of polarization. If no direction is specified, the direction of maximum radiation is implied.

$$A_e = P_A / W_i, \quad (4.61)$$

where

A_e is the effective aperture, m^2 ,

P_A is the power delivered from the antenna to the load, W,

W_i is the power flux density (Poynting vector value) of the incident wave, W/m^2 .

Using the Thevenin equivalent of a receiving antenna, we can show that equation (4.61) relates the antenna impedance and its effective aperture as

$$A_e = \frac{|I_A|^2 R_L / 2}{W_i} = \frac{|V_A|^2}{2W_i} \cdot \frac{R_L}{\left[(R_r + R_l + R_L)^2 + (X_A + X_L)^2 \right]}. \quad (4.62)$$

Under conditions of conjugate matching,

$$A_e = \frac{|V_A|^2}{8W_i \underbrace{(R_r + R_l)}_{R_A=R_L}}. \quad (4.63)$$

For aperture type antennas, the effective area is smaller than the physical aperture area. Aperture antennas with constant amplitude and phase distribution across the aperture have the maximum effective area, which is practically equal to the geometrical area. The effective aperture of wire antennas is much larger than the surface of the wire itself. Sometimes, the *aperture efficiency* of an antenna is estimated as the ratio of the effective antenna aperture and its physical area:

$$\varepsilon_{ap} = \frac{A_e}{A_p}. \quad (4.64)$$

Example: A uniform plane wave is incident upon a very short dipole. Find the effective area A_e assuming that the radiation resistance is $R_{rad} = 80(\pi l / \lambda)^2 \Omega$ and that the field is linearly polarized along the axis of the dipole. Compare A_e with the physical surface of the wire if $l = \lambda / 50$ and $d = \lambda / 300$, where d is the wire's diameter.

Since the dipole is very short, we can neglect the conduction losses. Wire antennas do not have dielectric losses. Therefore, we assume that $R_l = 0$. Under conjugate matching (which is implied unless specified otherwise),

$$A_e = \frac{|V_A|^2}{8W_i R_r}.$$

The dipole is very short and we can assume that the \mathbf{E} -field intensity is the same along the whole wire. Then, the voltage created by the induced electromotive force of the incident wave is

$$V_A = |\mathbf{E}| \cdot l.$$

The Poynting vector has a magnitude of $W_i = |\mathbf{E}|^2 / (2\eta)$. Then, under conditions of conjugate matching, see (4.63),

$$A_e = \frac{|\mathbf{E}|^2 \cdot l^2 \cdot 2\eta}{8 \cdot |\mathbf{E}|^2 \cdot R_r} = \frac{3\lambda^2}{8\pi} = 0.119 \cdot \lambda^2.$$

The physical surface of the dipole is

$$A_p = \pi dl = \frac{\pi}{15} 10^{-3} \lambda^2 = 2.1 \times 10^{-4} \cdot \lambda^2.$$

The aperture efficiency of this dipole is then

$$\varepsilon_{ap} = \frac{A_e}{A_p} = \frac{0.119}{2.1 \times 10^{-4}} = 568.2.$$

It is evident from the above example, that the aperture efficiency is not a suitable parameter for wire antennas, which have very small surface area. However, the effective area is still a useful parameter for wire antennas as it has direct relation with the directivity, as discussed next.

11. Relation Between Directivity D_0 and Effective Aperture A_e

The simplest derivation of this relation goes through two stages.

Stage 1: Prove that the ratio D_0 / A_e is the same for any antenna.

Consider two antennas: A1 and A2. Let A1 be the transmitting antenna, and A2 be the receiving one. Let the distance between the two antennas be R . The power density generated by A1 at A2 is

$$W_1 = \frac{D_1 \Pi_1}{4\pi R^2}.$$

Here, Π_1 is the total power radiated by A1 and D_1 is the directivity of A1. The above follows directly from the definition of directivity:

$$D(\theta, \varphi) = 4\pi \frac{U(\theta, \varphi)}{\Pi} = \frac{4\pi R^2 W(\theta, \varphi)}{\Pi} \Rightarrow W(\theta, \varphi) = \frac{\Pi D(\theta, \varphi)}{4\pi R^2}.$$

The power received by A2 and delivered to its load is

$$P_{1 \rightarrow 2} = A_{e_2} \cdot W_1 = \frac{D_1 \Pi_1 A_{e_2}}{4\pi R^2},$$

where A_{e_2} is the effective area of A2.

$$\Rightarrow D_1 A_{e_2} = 4\pi R^2 \frac{P_{1 \rightarrow 2}}{\Pi_1}.$$

Now, let A1 be the receiving antenna and A2 be the transmitting one. We can derive the following:

$$D_2 A_{e1} = 4\pi R^2 \frac{P_{2 \rightarrow 1}}{\Pi_2}.$$

If $\Pi_1 = \Pi_2$, then, according to the reciprocity principle in electromagnetics*, $P_{1 \rightarrow 2} = P_{2 \rightarrow 1}$. Therefore,

$$D_1 A_{e2} = D_2 A_{e1} \Rightarrow \frac{D_1}{A_{e1}} = \frac{D_2}{A_{e2}} = \gamma.$$

We thus proved that γ is the same for every antenna.

Stage 2: Find the ratio $\gamma = D_0 / A_e$ for an infinitesimal dipole.

The directivity of a very short dipole (infinitesimal dipole) is $D_0^{id} = 1.5$ (see **Examples of Section 4**, this Lecture). The effective aperture of an infinitesimal dipole is $A_e^{id} = 3\lambda^2 / (8\pi)$ (see the **Example of Section 10**, this Lecture). Then,

$$\gamma = \frac{D_0}{A_e} = \frac{1.5}{3\lambda^2} \cdot 8\pi,$$

$$\gamma = \frac{D_0}{A_e} = \frac{4\pi}{\lambda^2}. \quad (4.65)$$

Equation (4.65) is true if there are no dissipation, polarization mismatch, and impedance mismatch in the antenna system. If those factors are present, then

$$A_e = (1 - |\Gamma|^2) |\hat{\mathbf{p}}_w \cdot \hat{\mathbf{p}}_a|^2 \left(\frac{\lambda^2}{4\pi} \right) \underbrace{e D_0}_{G_0}. \quad (4.66)$$

From (4.20) and (4.65), we can obtain a simple relation between the antenna beam solid angle Ω_A and A_e :

$$A_e = \frac{\lambda^2}{4\pi} D_0 = \frac{\lambda^2}{\Omega_A}. \quad (4.67)$$

* Reciprocity in antenna theory states that if antenna #1 is a transmitting antenna and antenna #2 is a receiving antenna, then the ratio of transmitted to received power P_{tra} / P_{rec} will not change if antenna #1 becomes the receiving antenna and antenna #2 becomes the transmitting one.

12. Other Antenna Equivalent Areas

Before, we have defined the antenna effective area (or effective aperture) as the area, which when multiplied by the incident wave power density, produces the power delivered to the load (the terminals of the antenna) P_A . In a similar manner, we define the antenna **scattering area** A_s . It is the area, which when multiplied with the incident wave power density, produces the re-radiated (scattered) power:

$$A_s = \frac{P_s}{W_i} = \frac{|I_A|^2 R_r}{2W_i}, \text{ m}^2. \quad (4.68)$$

In the case of conjugate matching,

$$A_s = \frac{|V_A|^2}{8W_i} \frac{R_r}{(R_r + R_l)^2} = \frac{|V_A|^2}{8W_i} \frac{R_r}{R_A^2}, \text{ m}^2. \quad (4.69)$$

The **loss area** is the area, which when multiplied by the incident wave power density, produces the dissipated (as heat) power of the antenna.

$$A_l = \frac{P_l}{W_i} = \frac{|I_A|^2 R_l}{2W_i}, \text{ m}^2. \quad (4.70)$$

In the case of conjugate matching,

$$A_l = \frac{|V_A|^2}{8W_i} \frac{R_l}{(R_r + R_l)^2} = \frac{|V_A|^2}{8W_i} \frac{R_l}{R_A^2}, \text{ m}^2. \quad (4.71)$$

The **capture area** is the area, which when multiplied with the incident wave power density, produces the total power intercepted by the antenna:

$$A_c = \frac{P_t}{W_i} = \frac{|I_A|^2 (R_r + R_l + R_L)}{2W_i}. \quad (4.72)$$

In the case of conjugate matching,

$$A_c = \frac{|V_A|^2 (R_r + R_l + R_L)}{8W_i (R_r + R_l)^2} = \frac{|V_A|^2 (R_A + R_L)}{8W_i R_A^2} = \frac{|V_A|^2}{4W_i} \frac{1}{R_A}. \quad (4.73)$$

The capture area is the sum of the effective area, the loss area and the scattering area:

$$A_c = A_e + A_l + A_s. \quad (4.74)$$

When conjugate matching is achieved,

$$A_e = A_l + A_s = 0.5A_c. \quad (4.75)$$

If conjugate matching is achieved for a loss-free antenna, then

$$A_e = A_s = 0.5A_c. \quad (4.76)$$

Lecture 5: Polarization and Related Antenna Parameters

(Polarization of EM fields – revision. Polarization vector. Antenna polarization. Polarization loss factor and polarization efficiency.)

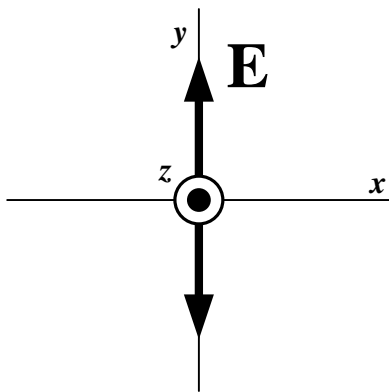
1. Introduction and Definition

The polarization of the EM field describes the orientation of its vectors at a given point and how it varies with time. In other words, it describes the way the direction and magnitude of the field vectors (usually \mathbf{E}) change in time. Polarization is associated with TEM time-harmonic waves where the \mathbf{H} vector relates to the \mathbf{E} vector simply by $\mathbf{H} = \hat{\mathbf{r}} \times \mathbf{E} / \eta$.

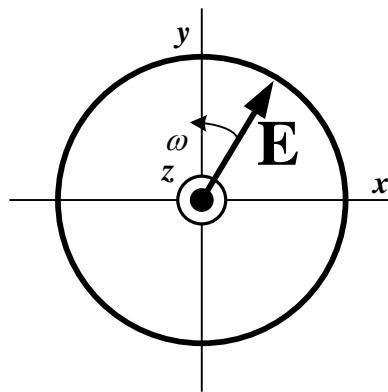
In antenna theory, we are concerned with the polarization of the field in the plane orthogonal to the direction of propagation—this is the plane defined by the far-zone vectors \mathbf{E} and \mathbf{H} . Remember that the far field is a quasi-TEM field.

The **polarization** is the locus traced by the extremity of the time-varying field vector at a fixed observation point.

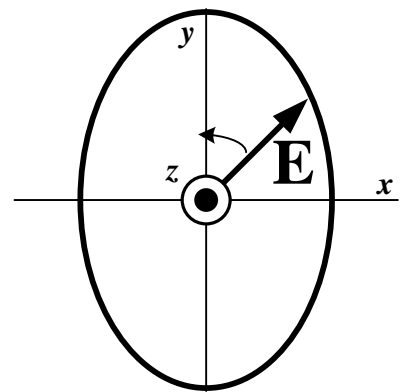
According to the shape of the trace, three types of polarization exist for harmonic fields: **linear**, **circular** and **elliptical**. Any polarization can be represented by two orthogonal linear polarizations, (E_x, E_y) or (E_H, E_V) , the fields of which may, in general, have different magnitudes and may be out of phase by an angle δ_L .



(a) linear polarization

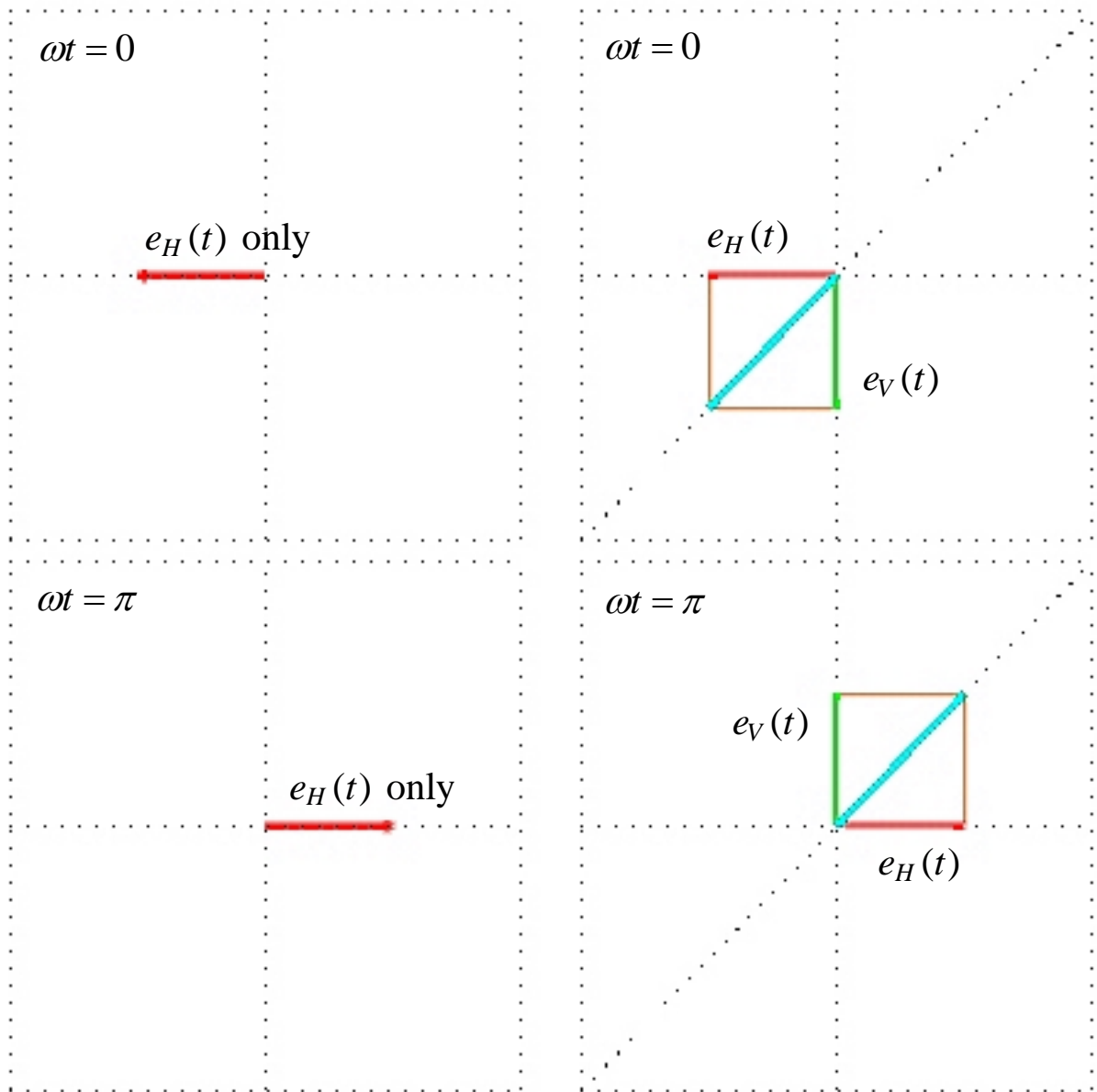


(b) circular polarization



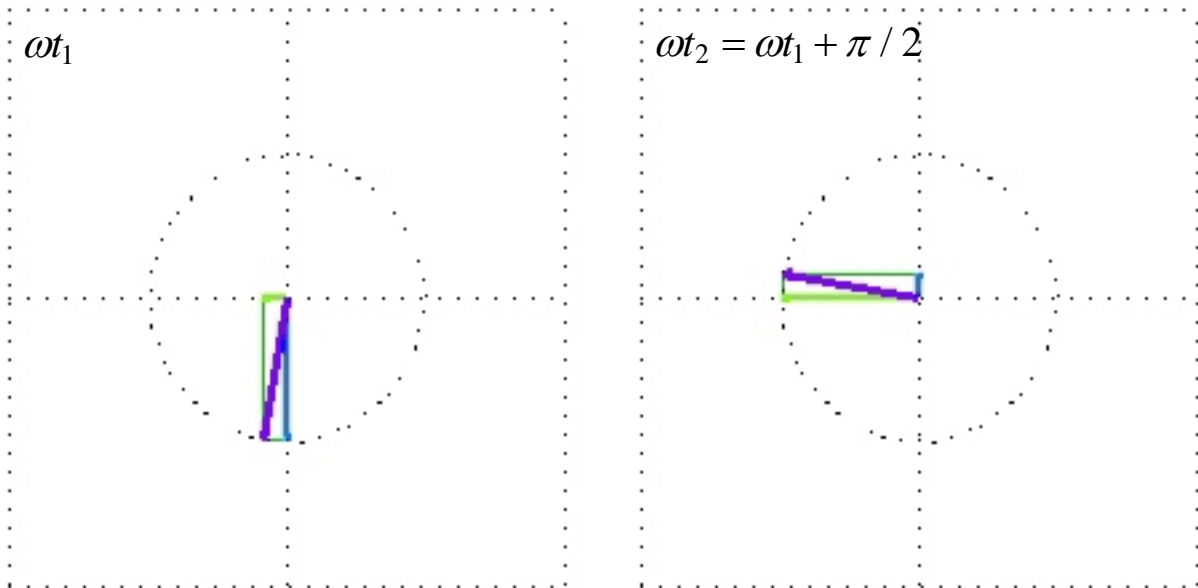
(c) elliptical polarization

- If $\delta_L = 0$ or $n\pi$, then the field is linearly polarized.



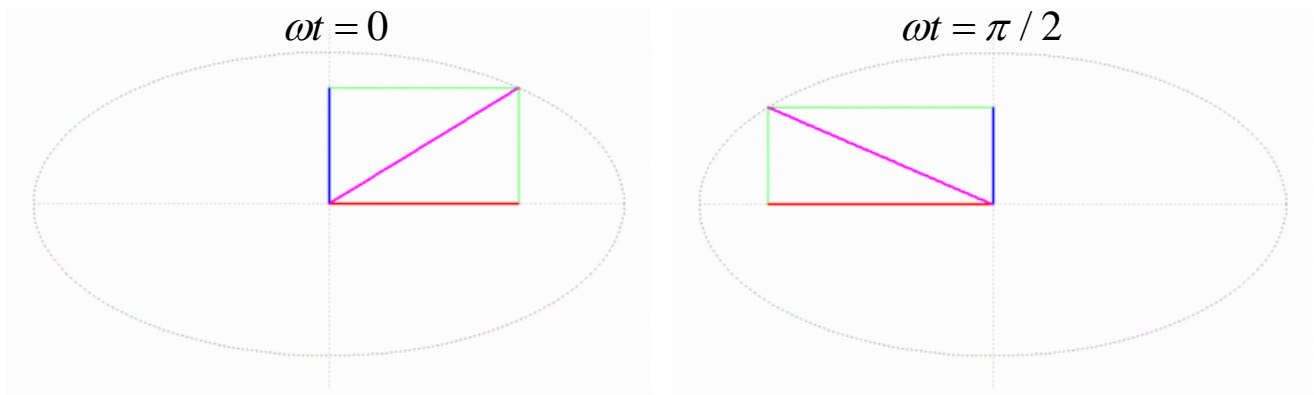
Animation: Linear Polarization, $\delta_L = 0$, $E_x = E_y$

- If $\delta_L = \pi / 2$ (90°) and $|E_x| = |E_y|$, then the field is circularly polarized.



Animation: Clockwise Circular Rotation

- In the most general case, the polarization is elliptical.



Animation: Counter-clockwise Elliptical Rotation

It is also true that any type of polarization can be represented by a right-hand circular and a left-hand circular polarizations (E_L, E_R). **[Animation]**

Next, we review the above statements and definitions, and introduce the new concept of polarization vector.

2. Field Polarization in Terms of Two Orthogonal Linearly Polarized Components

The polarization of any field can be represented by a set of two orthogonal linearly polarized fields. Assume that locally a far-field wave propagates along the z -axis. The far-zone field vectors have only transverse components. Then, the set of two orthogonal linearly polarized fields along the x -axis and along the y -axis, is sufficient to represent any TEM_z field. We use this arrangement to introduce the concept of *polarization vector*.

The field (time-dependent or phasor vector) is decomposed into two orthogonal components:

$$\mathbf{e} = \mathbf{e}_x + \mathbf{e}_y \Rightarrow \mathbf{E} = \mathbf{E}_x + \mathbf{E}_y, \quad (5.1)$$

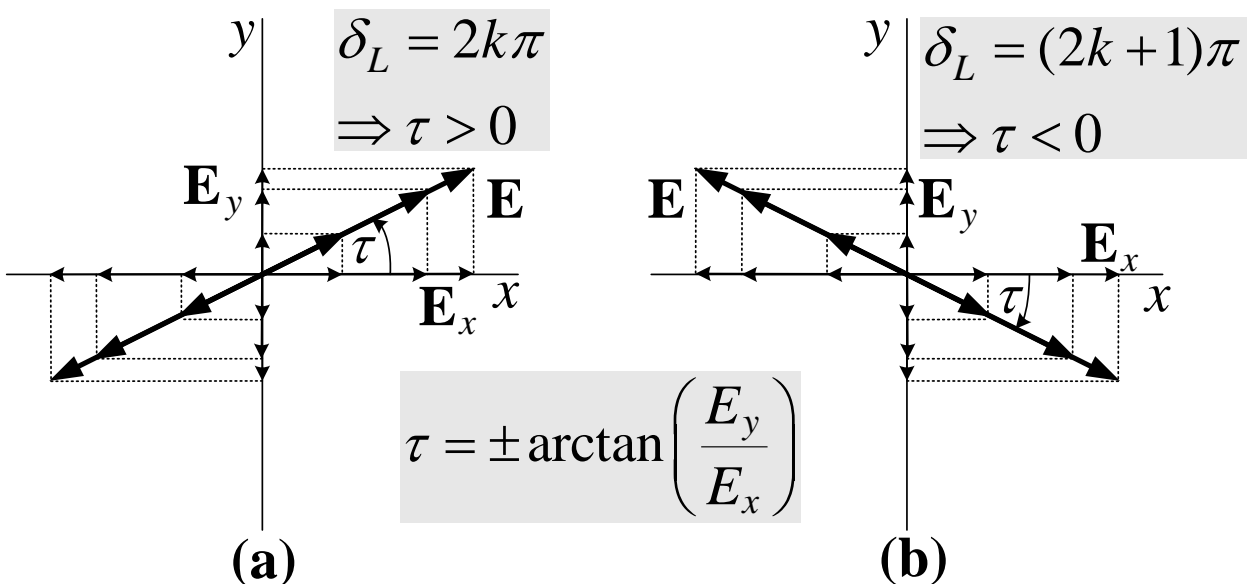
$$\begin{aligned} \mathbf{e}_x &= E_x \cos(\omega t - \beta z) \hat{\mathbf{x}} & \Rightarrow & \mathbf{E}_x = E_x e^{-j\beta z} \hat{\mathbf{x}} \\ \mathbf{e}_y &= E_y \cos(\omega t - \beta z + \delta_L) \hat{\mathbf{y}} & \Rightarrow & \mathbf{E}_y = E_y e^{-j\beta z} e^{j\delta_L} \hat{\mathbf{y}}. \end{aligned} \quad (5.2)$$

At a fixed position (assume $z = 0$), equation (5.1) can be written as

$$\begin{aligned} \mathbf{e}(t) &= \hat{\mathbf{x}} \cdot E_x \cos \omega t + \hat{\mathbf{y}} \cdot E_y \cos(\omega t + \delta_L) \\ &\Rightarrow \boxed{\mathbf{E} = \hat{\mathbf{x}} \cdot E_x + \hat{\mathbf{y}} \cdot E_y e^{j\delta_L}} \end{aligned} \quad (5.3)$$

Case 1: Linear polarization: $\delta_L = n\pi$, $n = 0, 1, 2, \dots$

$$\begin{aligned} \mathbf{e}(t) &= \hat{\mathbf{x}} \cdot E_x \cos(\omega t) + \hat{\mathbf{y}} \cdot E_y \cos(\omega t \pm n\pi) \\ &\Rightarrow \boxed{\mathbf{E} = \hat{\mathbf{x}} \cdot E_x \pm \hat{\mathbf{y}} \cdot E_y} \end{aligned} \quad (5.4)$$



Case 2: Circular polarization:

$$E_x = E_y = E_m \quad \text{and} \quad \delta_L = \pm \left(\frac{\pi}{2} + n\pi \right), \quad n = 0, 1, 2, \dots$$

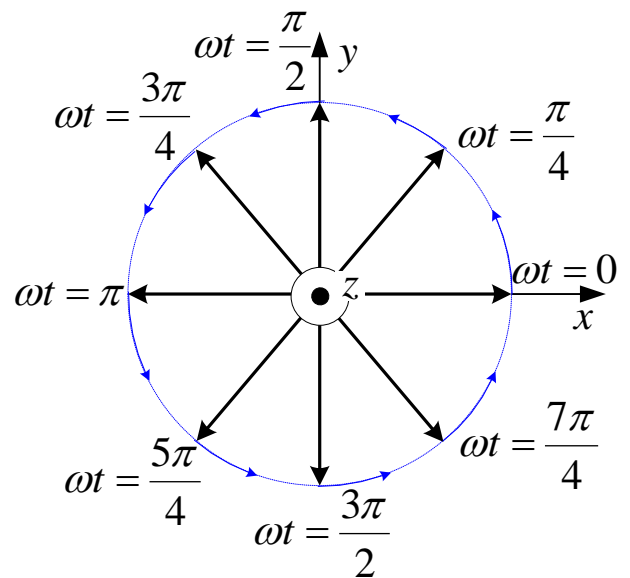
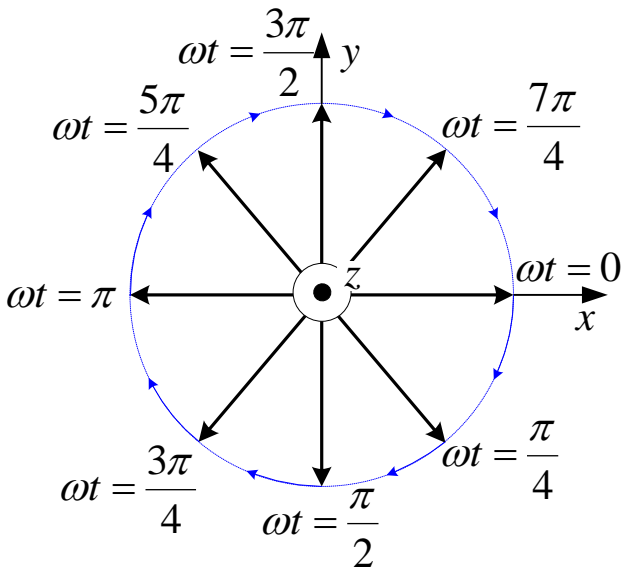
$$\mathbf{e}(t) = \hat{\mathbf{x}}E_x \cos(\omega t) + \hat{\mathbf{y}}E_y \cos[\omega t \pm (\pi / 2 + n\pi)]$$

$$\Rightarrow \boxed{\mathbf{E} = E_m(\hat{\mathbf{x}} \pm j\hat{\mathbf{y}})}$$

(5.5)

$$\boxed{\begin{aligned} \mathbf{E} &= E_m(\hat{\mathbf{x}} + j\hat{\mathbf{y}}) \\ \delta_L &= +\frac{\pi}{2} + 2n\pi \end{aligned}}$$

$$\boxed{\begin{aligned} \mathbf{E} &= E_m(\hat{\mathbf{x}} - j\hat{\mathbf{y}}) \\ \delta_L &= -\frac{\pi}{2} - 2n\pi \end{aligned}}$$

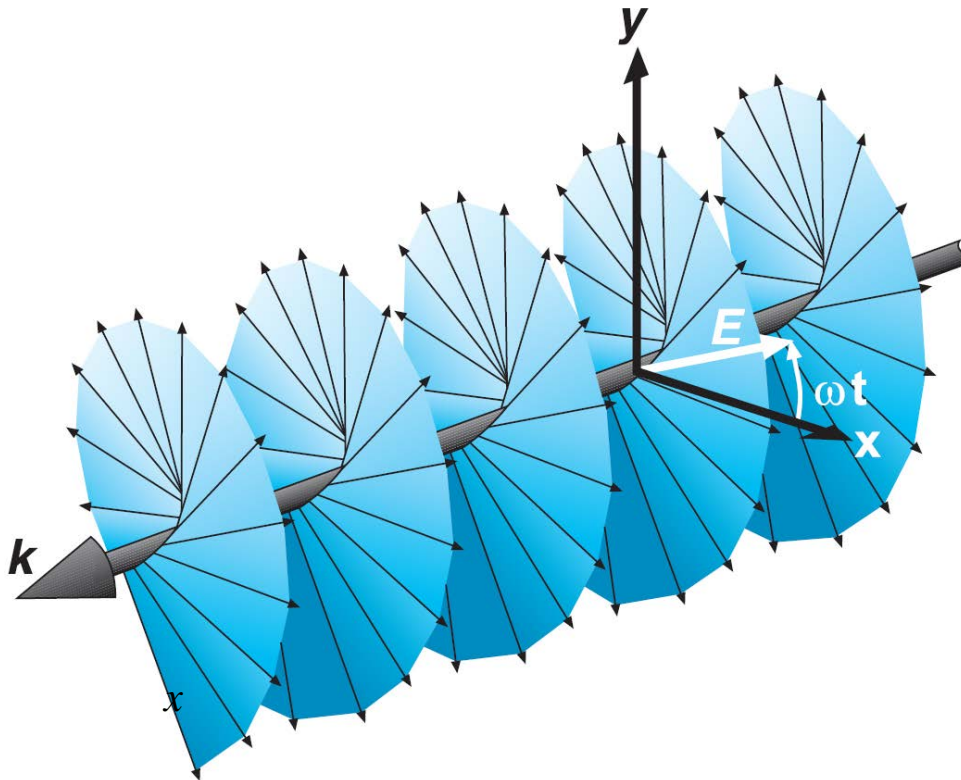


If $+\hat{\mathbf{z}}$ is the direction of propagation:
counterclockwise (CCW) or **left-hand (LH)** polarization

If $+\hat{\mathbf{z}}$ is the direction of propagation: **clockwise (CW)** or **right-hand (RH)** polarization

Note that the sense of rotation changes if the direction of propagation changes. In the example above, if the wave propagates along $-\hat{\mathbf{z}}$, the plot to the left, where $\mathbf{E} = E_m(\hat{\mathbf{x}} + j\hat{\mathbf{y}})$, corresponds to a right-hand (RH) wave, while the plot to the right, where $\mathbf{E} = E_m(\hat{\mathbf{x}} - j\hat{\mathbf{y}})$, corresponds to a left-hand (LH) wave. Vice versa, if the wave propagates along $+\hat{\mathbf{z}}$, then the left plot shows a LH wave, whereas the right plot shows a RH wave.

A snapshot of the field vector along the axis of propagation is given below for a right-hand circularly polarized (RHCP) wave. Pick an observing position along the axis of propagation (see the plane defined by the x and y axes in the plot below) and imagine that the whole helical trajectory of the tip of the field vector moves along the wave vector \mathbf{k} . Are you going to see the vector rotating clockwise or counter-clockwise as you look along \mathbf{k} ? (Ans.: Clockwise, which is equivalent to RH sense of rotation.)



[Hayt, Buck, Engineering Electromagnetics, 8th ed., p. 399]

Case 3: Elliptic polarization

The field vector at a given point traces an ellipse as a function of time. This is the most general type of polarization, obtained for any phase difference δ and any ratio (E_x / E_y) . Mathematically, the linear and the circular polarizations are special cases of the elliptical polarization. In practice, however, the term *elliptical polarization* is used to indicate polarizations *other than linear or circular*.

$$\begin{aligned} \mathbf{e}(t) &= \hat{\mathbf{x}}E_x \cos \omega t + \hat{\mathbf{y}}E_y \cos(\omega t + \delta_L) \\ \Rightarrow \mathbf{E} &= \hat{\mathbf{x}}E_x + \hat{\mathbf{y}}E_y e^{j\delta_L} \end{aligned} \quad (5.6)$$

Show that the trace of the time-dependent vector is an ellipse:

$$\begin{aligned}
 e_y(t) &= E_y(\cos \omega t \cdot \cos \delta_L - \sin \omega t \cdot \sin \delta_L) \\
 \cos \omega t &= \frac{e_x(t)}{E_x} \text{ and } \sin \omega t = \sqrt{1 - \left(\frac{e_x(t)}{E_x}\right)^2} \\
 \sin^2 \delta_L &= \left[\frac{e_x(t)}{E_x}\right]^2 - 2\left[\frac{e_x(t)}{E_x}\right]\left[\frac{e_y(t)}{E_y}\right]\cos \delta_L + \left[\frac{e_y(t)}{E_y}\right]^2
 \end{aligned}$$

or (dividing both sides by $\sin^2 \delta_L$),

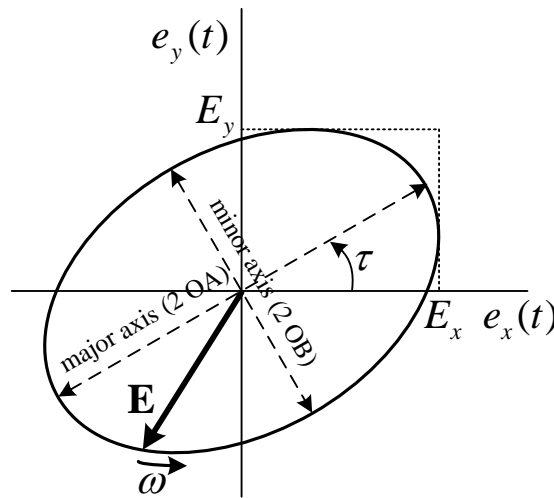
$$1 = x^2(t) - 2x(t)y(t)\cos \delta_L + y^2(t), \quad (5.7)$$

where

$$\begin{aligned}
 x(t) &= \frac{e_x(t)}{E_x \sin \delta_L} = \frac{\cos \omega t}{\sin \delta_L}, \\
 y(t) &= \frac{e_y(t)}{E_y \sin \delta_L} = \frac{\cos(\omega t + \delta_L)}{\sin \delta_L}.
 \end{aligned}$$

Equation (5.7) is the equation of an ellipse centered in the xy plane. It describes the trajectory of a point of coordinates $x(t)$ and $y(t)$, i.e., normalized $e_x(t)$ and $e_y(t)$ values, along an ellipse where the point moves with an angular frequency ω .

As the circular polarization, the elliptical polarization can be *right-handed* or *left-handed*, depending on the relation between the direction of propagation and the sense of rotation.



The parameters of the polarization ellipse are given below. Their derivation is given in Appendix I.

a) major axis ($2 \times \text{OA}$)

$$\text{OA} = \sqrt{\frac{1}{2} \left[E_x^2 + E_y^2 + \sqrt{E_x^4 + E_y^4 + 2E_x^2 E_y^2 \cos(2\delta_L)} \right]} \quad (5.8)$$

b) minor axis ($2 \times \text{OB}$)

$$\text{OB} = \sqrt{\frac{1}{2} \left[E_x^2 + E_y^2 - \sqrt{E_x^4 + E_y^4 + 2E_x^2 E_y^2 \cos(2\delta_L)} \right]} \quad (5.9)$$

c) tilt angle τ

$$\tau = \frac{1}{2} \arctan \left(\underbrace{\frac{2E_x E_y \cos \delta_L}{E_x^2 - E_y^2}}_A \right) \pm \frac{\pi}{2} \quad (5.10)$$

Note: Eq. (5.10) produces an infinite number of angles, $\tau = (\arctan A)/2 \pm n\pi/2$, $n = 1, 2, \dots$. Thus, it gives not only the angle which the major axis of the ellipse forms with the x axis but also the angle of the minor axis with the x axis. In spherical coordinates, τ is usually specified with respect to the $\hat{\theta}$ direction

d) axial ratio

$$\text{AR} = \frac{\text{major axis}}{\text{minor axis}} = \frac{\text{OA}}{\text{OB}} \quad (5.11)$$

Note: The linear and circular polarizations as special cases of the elliptical polarization:

- If $\delta_L = \pm \left(\frac{\pi}{2} + 2n\pi \right)$ and $E_x = E_y$, then $\text{OA} = \text{OB} = E_x = E_y$; the ellipse becomes a circle.
- If $\delta_L = n\pi$, then $\text{OB} = 0$ and $\tau = \pm \arctan(E_y / E_x)$; the ellipse collapses into a line.

3. Field Polarization in Terms of Two Circularly Polarized Components

The representation of a complex vector field in terms of circularly polarized components is somewhat less intuitive but it is actually more useful in the calculation of the polarization ellipse parameters. This time, the total field phasor is represented as the superposition of two circularly polarized waves,

one right-handed and the other left-handed. For the case of a wave propagating along $-z$ [see Case 2 and Eq. (5.5)],

$$\mathbf{E} = E_R(\hat{\mathbf{x}} + j\hat{\mathbf{y}}) + E_L(\hat{\mathbf{x}} - j\hat{\mathbf{y}}). \quad (5.12)$$

Here, E_R and E_L are, in general, complex phasors. Assuming a relative phase difference of $\delta_C = \varphi_R - \varphi_L$, one can write (5.12) as

$$\mathbf{E} = e_R e^{j\delta_C} (\hat{\mathbf{x}} + j\hat{\mathbf{y}}) + e_L (\hat{\mathbf{x}} - j\hat{\mathbf{y}}), \quad (5.13)$$

where e_R and e_L are real numbers.

The relations between the linear-component and the circular-component representations of the field polarization are easily found as

$$\mathbf{E} = \hat{\mathbf{x}} \underbrace{(E_R + E_L)}_{E_x} + \hat{\mathbf{y}} \underbrace{j(E_R - E_L)}_{E_y} \quad (5.14)$$

$$\Rightarrow \begin{cases} E_x = E_R + E_L \\ E_y = j(E_R - E_L) \end{cases} \quad (5.15)$$

$$\Rightarrow \begin{cases} E_R = 0.5(E_x - jE_y) \\ E_L = 0.5(E_x + jE_y). \end{cases} \quad (5.16)$$

4. Polarization Vector and Polarization Ratio of a Plane Wave

The *polarization vector* is the normalized phasor of the electric field vector. It is a complex-valued vector of unit magnitude, i.e., $\hat{\boldsymbol{\rho}}_L \cdot \hat{\boldsymbol{\rho}}_L^* = 1$.

$$\hat{\boldsymbol{\rho}}_L = \frac{\mathbf{E}}{E_m} = \hat{\mathbf{x}} \frac{E_x}{E_m} + \hat{\mathbf{y}} \frac{E_y}{E_m} e^{j\delta_L}, \quad E_m = \sqrt{E_x^2 + E_y^2} \quad (5.17)$$

The expression in (5.17) assumes a wave decomposition into linearly polarized (x and y) components, thereby the subscript L . Polarization vector in terms of RHCP and LHCP components is also used. The polarization vector defined in (5.17) takes the following specific forms in the cases of linearly, circularly and elliptically polarized waves.

Case 1: Linear polarization (the polarization vector is real-valued)

$$\hat{\boldsymbol{\rho}} = \hat{\mathbf{x}} \frac{E_x}{E_m} \pm \hat{\mathbf{y}} \frac{E_y}{E_m}, \quad E_m = \sqrt{E_x^2 + E_y^2} \quad (5.18)$$

where E_x and E_y are real numbers.

Case 2: Circular polarization (the polarization vector is complex-valued)

$$\hat{\mathbf{p}}_L = \frac{1}{\sqrt{2}}(\hat{\mathbf{x}} \pm j\hat{\mathbf{y}}), \quad E_m = \sqrt{2}E_x = \sqrt{2}E_y \quad (5.19)$$

The *polarization ratio* is the ratio of the phasors of the two orthogonal polarization components. In general, it is a complex number:

$$\tilde{r}_L = r_L e^{j\delta_L} = \frac{\tilde{E}_y}{\tilde{E}_x} = \frac{E_y e^{j\delta_L}}{E_x} \quad \text{or} \quad \tilde{r}_L = \frac{\tilde{E}_V}{\tilde{E}_H} \quad (5.20)$$

Point of interest: In the case of circular-component representation, the polarization ratio is defined as

$$\tilde{r}_C = r_C e^{j\delta_C} = \frac{\tilde{E}_R}{\tilde{E}_L} \quad (5.21)$$

The circular polarization ratio \tilde{r}_C is of particular interest since the axial ratio of the polarization ellipse AR can be expressed as

$$AR = \left| \frac{r_C + 1}{r_C - 1} \right| \quad (5.22)$$

Besides, its tilt angle with respect to the y (vertical) axis is simply

$$\tau_V = \delta_C / 2 + n\pi, \quad n = 0, \pm 1, \dots \quad (5.23)$$

Comparing (5.10) and (5.23) readily shows the relation between the phase difference δ_C of the circular-component representation and the linear polarization ratio $\tilde{r}_L = r_L e^{j\delta_L}$:

$$\delta_C = \arctan\left(\frac{2r_L}{1-r_L^2} \cos \delta_L\right) \quad (5.24)$$

We can calculate r_C from the linear polarization ratio \tilde{r}_L making use of (5.11) and (5.22):

$$AR = \frac{r_C + 1}{r_C - 1} = \sqrt{\frac{1 + r_L^2 + \sqrt{1 + r_L^4 + 2r_L^2 \cos(2\delta_L)}}{1 + r_L^2 - \sqrt{1 + r_L^4 + 2r_L^2 \cos(2\delta_L)}}} \quad (5.25)$$

Using (5.24) and (5.25) allows for the switching between the representation of the wave polarization in terms of linear and circular components.

5. Antenna Polarization

The *polarization of a transmitting antenna* is the polarization of its radiated wave in the far zone. *The polarization of a receiving antenna* is the polarization of an incident plane wave, which, for a given power flux density, results in maximum available power at the antenna terminals.

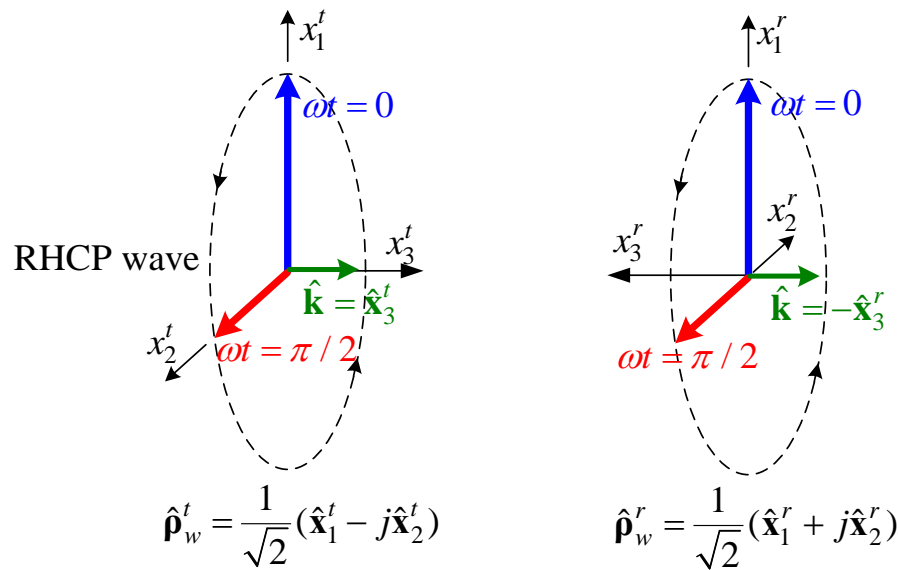
By convention, the antenna polarization is defined by the polarization vector of the wave it transmits. Therefore, the antenna polarization vector is determined according to the definition of antenna polarization in a transmitting mode. Notice that the polarization vector $\hat{\rho}_w^t$ of a wave in the coordinate system of transmission (the wave moves away from the antenna at the origin, i.e., along $\hat{\mathbf{r}}$) is the conjugate of its polarization vector $\hat{\rho}_w^r$ in the coordinate system of reception (the wave moves toward the antenna at the origin, i.e., along $-\hat{\mathbf{r}}$):

$$\hat{\rho}_w^r = (\hat{\rho}_w^t)^*. \quad (5.26)$$

The conjugation is without importance for a linearly polarized wave since its polarization vector is real. It is, however, important in the cases of circularly and elliptically polarized waves.

This is illustrated in the figure below with a RHCP wave. Let the coordinate triplet (x_1^t, x_2^t, x_3^t) represent the coordinate system of the transmitting antenna while (x_1^r, x_2^r, x_3^r) represents that of the receiving antenna. In antenna analysis, the plane of polarization is usually given in spherical coordinates by $(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2) \equiv (\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}})$ and the third axis obeys $\hat{\mathbf{x}}_1 \times \hat{\mathbf{x}}_2 = \hat{\mathbf{x}}_3$, i.e., $\hat{\mathbf{x}}_3 = \hat{\mathbf{r}}$. Since the transmitting and receiving antennas face each other, their coordinate systems are oriented so that $\hat{\mathbf{x}}_3^t = -\hat{\mathbf{x}}_3^r$ (i.e., $\hat{\mathbf{r}}^r = -\hat{\mathbf{r}}^t$). If we align the axes $\hat{\mathbf{x}}_1^t$ and $\hat{\mathbf{x}}_1^r$, then $\hat{\mathbf{x}}_2^t = -\hat{\mathbf{x}}_2^r$ must hold. This changes the sign of the respective (2nd) field vector component. Upon normalization, this results in a change of sign in the imaginary part of the wave polarization vector.

Bearing in mind the definitions of antenna polarization in transmitting and receiving modes, we conclude that *in a common coordinate system the transmitting-mode polarization vector of an antenna is the conjugate of its receiving-mode polarization vector.*



6. Polarization Loss Factor (Polarization Efficiency)

Generally, the polarization of the receiving antenna is not the same as the polarization of the incident wave. This is called *polarization mismatch*. The polarization loss factor (PLF) characterizes the loss of EM power due to the polarization mismatch. The PLF is defined so that it attains a value of 1 (or 100%, or 0 dB) if there is no polarization mismatch, i.e., the antenna receives the maximum possible power for the given incident power density. A PLF equal to 0 ($-\infty$ dB) indicates complete polarization mismatch and inability to capture power from the incident wave. Thus,

$$0 \leq \text{PLF} \leq 1. \quad (5.27)$$

Note that the polarization loss has nothing to do with dissipation. It can be viewed as a “missed opportunity” to capture as much power from the incident wave as possible. The *polarization efficiency* has the same meaning as PLF.

Let us denote the polarization vector of a wave incident upon a receiving antenna as $\hat{\mathbf{p}}_w$. In the coordinate system where the receiving antenna is at the origin, this vector describes a wave propagating along $-\hat{\mathbf{r}}$. Assume also that the polarization vector of the wave that the receiving antenna would produce if it were to operate in transmitting mode is $\hat{\mathbf{p}}_a$. In the coordinate system where the receiving antenna is at the origin, this vector describes a wave propagating along $+\hat{\mathbf{r}}$. Then, the PLF is defined as

$$\text{PLF} = |\hat{\mathbf{p}}_w \cdot \hat{\mathbf{p}}_a|^2. \quad (5.28)$$

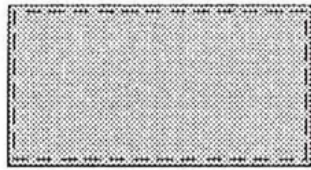
Note that if $\hat{\mathbf{p}}_w^* = \hat{\mathbf{p}}_a$, i.e., the incident field is

$$\mathbf{E}^i = E_m \hat{\mathbf{p}}_a^*$$

PLF = 1, and we obtain maximum possible received power at the antenna terminals. Remember that the transmitting-mode and receiving-mode polarization vectors of a wave are mutually conjugate? This means that $\hat{\mathbf{p}}_a^*$ is nothing but the wave the receiving antenna would generate if it were to transmit in the direction of the incident-wave propagation. Thus, the optimal polarization of the incident wave is the one that matches the polarization of the wave produced by the receiving antenna if it was the one launching the incident wave.

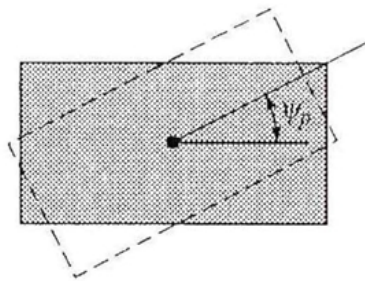
Here are some simple examples:

- 1) if $\hat{\mathbf{p}}_w = \hat{\mathbf{p}}_a^* = \hat{\mathbf{p}}_a = \hat{\mathbf{x}}$, then PLF=1;
- 2) if $\hat{\mathbf{p}}_w = \hat{\mathbf{x}}$ and $\hat{\mathbf{p}}_a = \hat{\mathbf{p}}_a^* = \hat{\mathbf{y}}$, then PLF=0;
- 3) if $\hat{\mathbf{p}}_w = \hat{\mathbf{p}}_a = \hat{\mathbf{x}} + j\hat{\mathbf{y}}$, then PLF=0;
- 4) if $\hat{\mathbf{p}}_w = \hat{\mathbf{x}} + j\hat{\mathbf{y}}$ and $\hat{\mathbf{p}}_a = \hat{\mathbf{x}} - j\hat{\mathbf{y}}$ ($\hat{\mathbf{p}}_w^* = \hat{\mathbf{p}}_a$), then PLF=1.



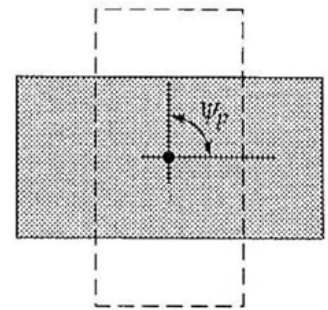
$$\text{PLF} = |\hat{\mathbf{p}}_w \cdot \hat{\mathbf{p}}_a|^2 = 1$$

(aligned)



$$\text{PLF} = |\hat{\mathbf{p}}_w \cdot \hat{\mathbf{p}}_a|^2 = \cos^2 \psi_p$$

(rotated)



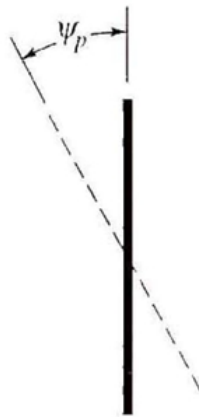
$$\text{PLF} = |\hat{\mathbf{p}}_w \cdot \hat{\mathbf{p}}_a|^2 = 0$$

(orthogonal)



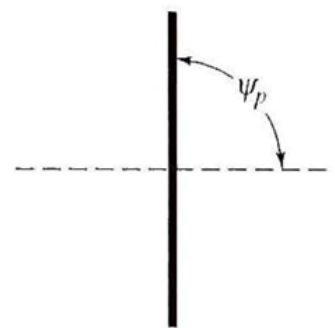
$$\text{PLF} = |\hat{\mathbf{p}}_w \cdot \hat{\mathbf{p}}_a|^2 = 1$$

(aligned)



$$\text{PLF} = |\hat{\mathbf{p}}_w \cdot \hat{\mathbf{p}}_a|^2 = \cos^2 \psi_p$$

(rotated)



$$\text{PLF} = |\hat{\mathbf{p}}_w \cdot \hat{\mathbf{p}}_a|^2 = 0$$

(orthogonal)

[Balanis, 2nd ed.]

In a communication link, the PLF has to be expressed by the polarization vectors of the transmitting and receiving antennas, $\hat{\rho}_{Tx}$ and $\hat{\rho}_{Rx}$, respectively. Both of these are defined in the coordinate system of the respective antenna as the polarization of the transmitted wave. However, these two coordinate systems have their radial unit vectors pointing in opposite directions, i.e., $\hat{\mathbf{r}}_{Rx} = -\hat{\mathbf{r}}_{Tx}$ as illustrated in the figure below. Therefore, either $\hat{\rho}_{Tx}$ or $\hat{\rho}_{Rx}$ has to be conjugated when calculating the PLF (it does not matter which one). For example, if the reference coordinate system is that of the receiving antenna, then

$$\text{PLF} = \left| \hat{\rho}_{Tx}^* \cdot \hat{\rho}_{Rx} \right|^2. \quad (5.29)$$

The expression $\text{PLF} = \left| \hat{\rho}_{Tx} \cdot \hat{\rho}_{Rx}^* \right|^2$ is also correct.



Examples

Example 5.1. The electric field of a linearly polarized EM wave is

$$\mathbf{E}^i = \hat{\mathbf{x}} \cdot E_m(x, y)e^{-j\beta z}.$$

It is incident upon a linearly polarized receiving antenna, which would transmit the field

$$\mathbf{E}_a = (\hat{\mathbf{x}} + \hat{\mathbf{y}}) \cdot e^{j\beta z}$$

if it were to operate in a transmitting instead of receiving mode. Find the PLF.

Notice that \mathbf{E}_a propagates along $-z$ in accordance with the requirement that it represents a transmitted wave.

$$\text{PLF} = \left| \hat{\mathbf{x}} \cdot \frac{1}{\sqrt{2}}(\hat{\mathbf{x}} + \hat{\mathbf{y}}) \right|^2 = \frac{1}{2}$$

$$\text{PLF}_{[\text{dB}]} = 10 \log_{10} 0.5 = -3 \text{ dB}$$

Example 5.2. A transmitting antenna produces a far-zone field, which is RH circularly polarized. This field impinges upon a receiving antenna, whose polarization (in transmitting mode) is also RH circular. Determine the PLF.

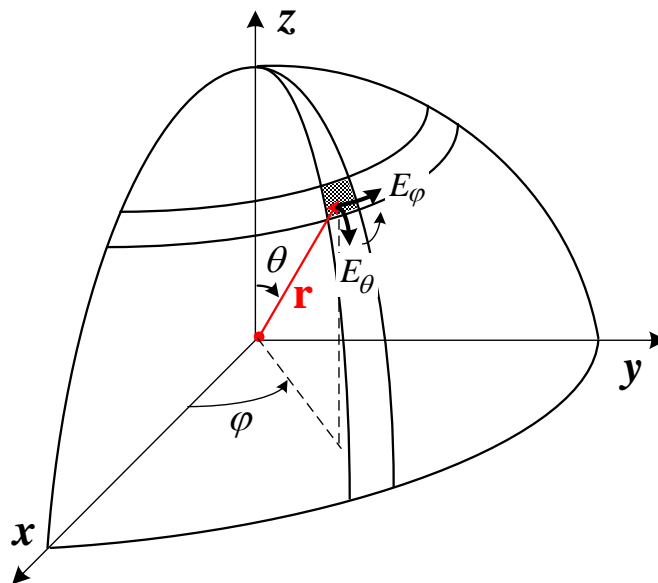
Both antennas (the transmitting one and the receiving one) are RH circularly polarized in transmitting mode. Assume that a transmitting antenna is located at the center of a spherical coordinate system. The far-zone field it would produce is described as

$$\mathbf{E}^{far} = E_m \left[\hat{\boldsymbol{\theta}} \cdot \cos \omega t + \hat{\boldsymbol{\phi}} \cdot \cos(\omega t - \pi / 2) \right].$$

This is a RHCP field with respect to the outward radial direction $\hat{\mathbf{r}}$. Its polarization vector is

$$\hat{\boldsymbol{\rho}}_{Tx} = \frac{\hat{\boldsymbol{\theta}} - j\hat{\boldsymbol{\phi}}}{\sqrt{2}}.$$

This is exactly the polarization vector of the transmitting antenna in its own coordinate system.



Since the receiving antenna is also RHCP, its polarization vector is

$$\hat{\boldsymbol{\rho}}_{Rx} = \frac{\hat{\boldsymbol{\theta}} - j\hat{\boldsymbol{\phi}}}{\sqrt{2}}.$$

The PLF is calculated as per (5.29):

$$\text{PLF} = |\hat{\mathbf{p}}_{\text{Tx}}^* \cdot \hat{\mathbf{p}}_{\text{Rx}}|^2 = \frac{|(\hat{\boldsymbol{\theta}} + j\hat{\boldsymbol{\phi}}) \cdot (\hat{\boldsymbol{\theta}} - j\hat{\boldsymbol{\phi}})|^2}{4} = 1,$$

$$\text{PLF}_{[\text{dB}]} = 10 \log_{10} 1 = 0.$$

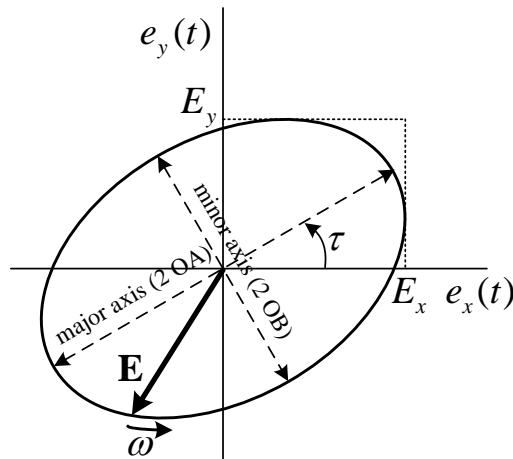
There is no polarization loss.

Exercise: Show that an antenna of RH circular polarization (in transmitting mode) cannot receive LH circularly polarized incident wave (or a wave emitted by a left-circularly polarized antenna).

Appendix I

Find the tilt angle τ , the length of the major axis OA, and the length of the minor axis OB of the ellipse described by the equation:

$$\sin^2 \delta = \left[\frac{e_x(t)}{E_x} \right]^2 - 2 \left[\frac{e_x(t)}{E_x} \right] \left[\frac{e_y(t)}{E_y} \right] \cos \delta + \left[\frac{e_y(t)}{E_y} \right]^2. \quad (\text{A-1})$$



Equation (A-1) can be written as

$$a \cdot x^2 - b \cdot xy + c \cdot y^2 = 1, \quad (\text{A-2})$$

where

$x = e_x(t)$ and $y = e_y(t)$ are the coordinates of a point of the ellipse centered in the xy plane;

$$a = \frac{1}{E_x^2 \sin^2 \delta};$$

$$b = \frac{2 \cos \delta}{E_x E_y \sin^2 \delta};$$

$$c = \frac{1}{E_y^2 \sin^2 \delta}.$$

After dividing both sides of (A-2) by (xy) , one obtains

$$a \frac{x}{y} - b + c \frac{y}{x} = \frac{1}{xy}. \quad (\text{A-3})$$

Introducing $\xi = \frac{y}{x} = \frac{e_y(t)}{e_x(t)}$, one obtains that

$$x^2 = \frac{1}{c\xi^2 - b\xi + a}$$

$$\Rightarrow \rho^2(\xi) = x^2 + y^2 = x^2(1 + \xi^2) = \frac{1 + \xi^2}{c\xi^2 - b\xi + a}. \quad (\text{A-4})$$

Here, ρ is the distance from the center of the coordinate system to the point on the ellipse. We want to know at what values of ξ the maximum and the minimum of ρ occur (ξ_{\min} , ξ_{\max}). This will produce the tilt angle τ . We also want to know the values of ρ_{\max} (major axis) and ρ_{\min} (minor axis). Then, we have to solve

$$\frac{d(\rho^2)}{d\xi} = 0, \text{ or}$$

$$\xi_m^2 - \frac{2(a-c)}{b}\xi_m - 1 = 0, \text{ where } \xi_m \equiv \xi_{\min}, \xi_{\max}. \quad (\text{A-5})$$

(A-5) is solved for the angle τ , which relates to ξ_{\max} as

$$\xi_{\max} = \tan \tau = (y/x)_{\max}. \quad (\text{A-6})$$

Substituting (A-6) in (A-5) yields:

$$\left(\frac{\sin \tau}{\cos \tau} \right)^2 - 2C \left(\frac{\sin \tau}{\cos \tau} \right) - 1 = 0 \quad (\text{A-7})$$

where

$$C = \frac{a - c}{b} = \frac{E_y^2 - E_x^2}{2E_x E_y \cos \delta}.$$

Multiplying both sides of (A-7) by $\cos^2 \tau$ and re-arranging results in

$$\underbrace{\cos^2 \tau - \sin^2 \tau}_{\cos(2\tau)} + \underbrace{2C \sin \tau \cdot \cos \tau}_{C \sin(2\tau)} = 0.$$

Thus, the solution of (A-7) is

$$\tan(2\tau) = -1 / C$$

or

$$\tau_1 = \frac{1}{2} \arctan \left(\frac{2E_x E_y \cos \delta}{E_x^2 - E_y^2} \right); \quad \tau_2 = \tau_1 + \frac{\pi}{2}. \quad (\text{A-8})$$

The angles τ_1 and τ_2 are the angles between the major and minor axes with the x axis. Substituting τ_1 and τ_2 back in ρ (see A-4) yields the expressions for OA and OB.

Lecture 6: Friis Transmission Equation and Radar Range Equation

(Friis equation. EIRP. Maximum range of a wireless link. Radar cross section. Radar equation. Maximum range of a radar.)

1. Friis Transmission Equation

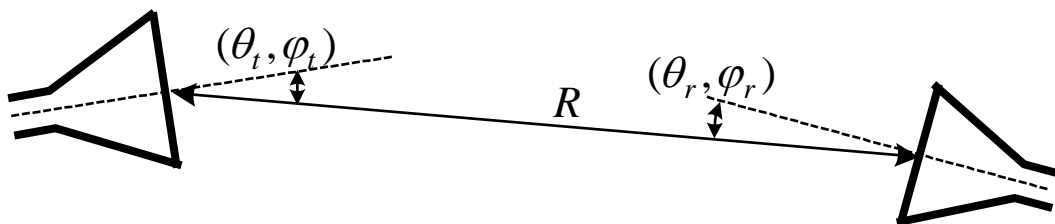
Friis transmission equation is essential in the analysis and design of wireless communication systems. It relates the power fed to the transmitting antenna and the power received by the receiving antenna when the two antennas are separated by a sufficiently large distance ($R \gg 2D_{\max}^2 / \lambda$), i.e., they are in each other's far zones. We derive the Friis equation next.

A transmitting antenna produces power density $W_t(\theta_t, \varphi_t)$ in the direction (θ_t, φ_t) . This power density depends on the transmitting antenna gain in the given direction $G(\theta_t, \varphi_t)$, on the power of the transmitter P_t fed to it, and on the distance R between the antenna and the observation point as

$$W_t = \frac{P_t}{4\pi R^2} G_t(\theta_t, \varphi_t) = \frac{P_t}{4\pi R^2} e_t D_t(\theta_t, \varphi_t). \quad (6.1)$$

Here, e_t denotes the radiation efficiency of the transmitting antenna and D_t is its directivity. The power P_r at the terminals of the receiving antenna can be expressed via its effective area A_r and W_t :

$$P_r = A_r W_t. \quad (6.2)$$



To include polarization and dissipation in the receiving antenna, we add the radiation efficiency of the receiving antenna e_r and the PLF:

$$P_r = e_r \cdot \text{PLF} \cdot A_r W_t = A_r W_t e_r |\hat{\mathbf{p}}_t^* \cdot \hat{\mathbf{p}}_r|^2, \quad (6.3)$$

$$\Rightarrow P_r = \underbrace{D_r(\theta_r, \varphi_r) \cdot \frac{\lambda^2}{4\pi}}_{A_r} \cdot W_t e_r |\hat{\mathbf{p}}_t^* \cdot \hat{\mathbf{p}}_r|^2. \quad (6.4)$$

Here, D_r is the directivity of the receiving antenna. The polarization vectors of the transmitting and receiving antennas, $\hat{\mathbf{p}}_t$ and $\hat{\mathbf{p}}_r$, are evaluated in their respective coordinate systems; this is why, one of them has to be conjugated when calculating the PLF.

The signal is incident upon the receiving antenna from a direction (θ_r, φ_r) , which is defined in the coordinate system of the receiving antenna:

$$\Rightarrow P_r = D_r(\theta_r, \varphi_r) \cdot \frac{\lambda^2}{4\pi} \cdot \underbrace{\frac{P_t}{4\pi R^2} e_t D_t(\theta_t, \varphi_t) \cdot e_r}_{W_t} |\hat{\mathbf{p}}_t^* \cdot \hat{\mathbf{p}}_r|^2. \quad (6.5)$$

The ratio of the received to the transmitted power is obtained as

$$\frac{P_r}{P_t} = e_t e_r |\hat{\mathbf{p}}_t^* \cdot \hat{\mathbf{p}}_r|^2 \left(\frac{\lambda}{4\pi R} \right)^2 D_t(\theta_t, \varphi_t) D_r(\theta_r, \varphi_r). \quad (6.6)$$

If the impedance-mismatch loss factor is included in both the receiving and the transmitting antenna systems, the above ratio becomes

$$\frac{P_r}{P_t} = (1 - |\Gamma_t|^2)(1 - |\Gamma_r|^2) e_t e_r |\hat{\mathbf{p}}_t^* \cdot \hat{\mathbf{p}}_r|^2 \left(\frac{\lambda}{4\pi R} \right)^2 D_t(\theta_t, \varphi_t) D_r(\theta_r, \varphi_r). \quad (6.7)$$

The above equations are variations of Friis' transmission equation, which is widely used in the design of wireless systems as well as the estimation of antenna radiation efficiency (when the antenna directivity is known).

For the case of impedance-matched and polarization-matched transmitting and receiving antennas, Friis equation reduces to

$$\frac{P_r}{P_t} = \left(\frac{\lambda}{4\pi R} \right)^2 G_t(\theta_t, \varphi_t) G_r(\theta_r, \varphi_r). \quad (6.8)$$

The factor $(\lambda / 4\pi R)^2$ is called the **free-space loss factor**. It reflects two effects: (1) the decrease in the power density due to the spherical spread of the wave through the term $1 / (4\pi R^2)$, and (2) the effective aperture dependence on the wavelength as $\lambda^2 / (4\pi)$.

2. Effective Isotropically Radiated Power (EIRP)

From the Friis equation (6.8), it is seen that to estimate the received power P_r we need the product of the transmitter power P_t and the transmitting antenna gain G_t . If the transmission line introduces losses in addition to those of the antenna system, these need to be accounted for as well. This is why often a

transmission system is characterized by its *effective isotropically radiated power (EIRP)*:

$$EIRP = P_t G_t e_{TL}, \text{ W} \quad (6.9)$$

where e_{TL} is the loss efficiency of the transmission line connecting the transmitter to the antenna. Usually, the EIRP is given in dB, in which case (6.9) becomes

$$EIRP_{dB} = P_{tdB} + G_{tdBi} + e_{TL,dB}. \quad (6.10)$$

Bearing in mind that $P_{in,t} = e_{TL} P_t$ and $G_t = 4\pi U_{max,t} / P_{in,t}$, the EIRP can also be written as

$$EIRP = 4\pi U_{max,t}, \text{ W}. \quad (6.11)$$

It is now clear that the *EIRP is a fictitious amount of power that an isotropic radiator would have to emit in order to produce the peak power density observed in the direction of the maximum radiation*. As such, and as evident from (6.9), the EIRP is greater than the actual power an antenna needs in order to achieve a given amount of radiation intensity in its direction of maximum radiation.

3. Maximum Range of a Wireless Link

Friis' transmission equation is frequently used to calculate the *maximum range* at which a wireless link can operate. For that, we need to know the nominal power of the transmitter P_t , all the parameters of the transmitting and receiving antenna systems (such as polarization, gain, losses, impedance mismatch), and the minimum power at which the receiver can operate reliably $P_{r \min}$. Then,

$$R_{max}^2 = (1 - |\Gamma_t|^2)(1 - |\Gamma_r|^2) e_t e_r |\hat{\mathbf{p}}_t^* \cdot \hat{\mathbf{p}}_r|^2 \left(\frac{\lambda}{4\pi} \right)^2 \left(\frac{P_t}{P_{r \min}} \right) D_t(\theta_t, \varphi_t) D_r(\theta_r, \varphi_r). \quad (6.12)$$

The minimum power at which the receiver can operate reliably is dependent on numerous factors, of which very important is the signal-to-noise ratio (SNR). There are different sources of noise but we are mostly concerned with the noise of the antenna itself. This topic is considered in the next lecture.

4. Radar Cross-section (RCS) or Echo Area

The RCS is a far-field characteristic of a radar target, which creates an echo by scattering (reflecting) the radar EM wave.

The RCS of a target σ is the equivalent area capturing that amount of power, which, when scattered isotropically, produces at the receiver an amount of power density, which is equal to that scattered by the target itself:

$$\sigma = \lim_{R \rightarrow \infty} \left[4\pi R^2 \frac{W_s}{W_i} \right] = \lim_{R \rightarrow \infty} \left[4\pi R^2 \frac{|\mathbf{E}_s|^2}{|\mathbf{E}_i|^2} \right], \text{ m}^2. \quad (6.13)$$

Here,

R is the distance from the target, m;

W_i is the incident power density, W/m²;

W_s is the scattered power density at the receiver, W/m².

To understand better the above definition, we can re-write (6.13) in an equivalent form:

$$\lim_{R \rightarrow \infty} \left[\frac{\sigma W_i}{4\pi R^2} \right] = W_s(R). \quad (6.14)$$

The product σW_i represents the equivalent intercepted power, which is assumed to be scattered (re-radiated) isotropically to create a fictitious spherical wave, the power density W_s of which decreases with distance as $1/R^2$ in the far zone. It is then expected that σW_i is a quantity independent of distance. W_s must be equal to the true scattered power density W_s produced by the real scatterer (the radar target).

We note that in general the RCS has little in common with any of the cross-sections of the actual scatterer. However, it is representative of the reflection properties of the target. It depends very much on the angle of incidence, on the angle of observation, on the shape and size of the scatterer, on the EM properties of the materials that it is built of, and on the wavelength. The RCS of targets is similar to the concept of effective aperture of antennas.

Large RCSs result from large metal content in the structure of the object (e.g., trucks and jumbo jet airliners have large RCS, $\sigma > 100 \text{ m}^2$). The RCS increases also due to sharp metallic or dielectric edges and corners. The reduction of the RCS is desired for stealth military aircraft meant to be invisible

to radars. This is achieved by careful shaping and coating (with special materials) of the outer surface of the airplane. The materials are mostly designed to absorb EM waves at the radar frequencies (usually S and X bands). Layered structures can also cancel the backscatter in a particular bandwidth. Shaping aims mostly at directing the backscattered wave at a direction different from the direction of incidence. Thus, in the case of a monostatic radar system, the scattered wave is directed away from the receiver. The stealth aircraft has RCS smaller than 10^{-4} m^2 , which makes it comparable or smaller than the RCS of a penny.

5. Radar Range Equation

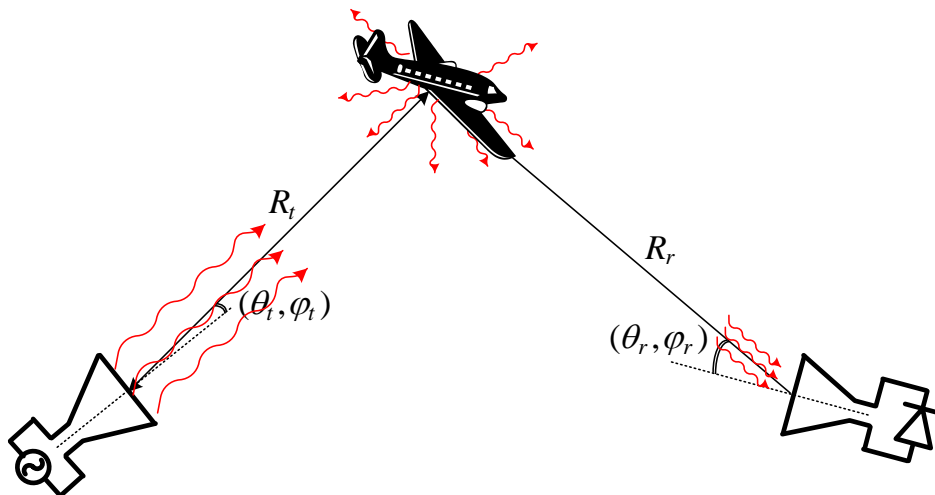
The radar range equation (RRE) gives the ratio of the transmitted power (fed to the transmitting antenna) to the received power, after it has been scattered (re-radiated) by a target of cross-section σ .

In the general radar scattering problem, there is a transmitting and a receiving antenna, and they may be located at different positions as shown in the figure below. This is called *bistatic scattering*.

Often, one antenna is used to transmit an EM pulse and to receive the echo from the target. This case is referred to as *monostatic scattering* or *backscattering*. Bear in mind that the RCS of a target may vary considerably as the location of the transmitting and receiving antennas change.

Assume the power density of the transmitted wave at the target location is

$$W_t = \frac{P_t G_t(\theta_t, \varphi_t)}{4\pi R_t^2} = \frac{P_t e_t D_t(\theta_t, \varphi_t)}{4\pi R_t^2}, \text{ W/m}^2. \quad (6.15)$$



The target is represented by its RCS σ , which is used to calculate the captured power $P_c = \sigma W_t$ (W), which when scattered isotropically gives the power density at the receiving antenna that is due to the target. The density of the re-radiated (scattered) power at the receiving antenna is

$$W_r = \frac{P_c}{4\pi R_r^2} = \frac{\sigma W_t}{4\pi R_r^2} = e_t \sigma \frac{P_t D_t(\theta_t, \varphi_t)}{(4\pi R_t R_r)^2}. \quad (6.16)$$

The power transferred to the receiver is

$$P_r = e_r \cdot A_r \cdot W_r = e_r \cdot \left(\frac{\lambda^2}{4\pi} \right) D_r(\theta_r, \varphi_r) \cdot e_t \sigma \frac{P_t D_t(\theta_t, \varphi_t)}{(4\pi R_t R_r)^2}. \quad (6.17)$$

Re-arranging and including impedance mismatch losses as well as polarization losses, yields the complete radar range equation:

$$\frac{P_r}{P_t} = e_t e_r (1 - |\Gamma_t|^2)(1 - |\Gamma_r|^2) |\hat{\mathbf{p}}_t^* \cdot \hat{\mathbf{p}}_r|^2 \sigma \left(\frac{\lambda}{4\pi R_t R_r} \right)^2 \frac{D_t(\theta_t, \varphi_t) D_r(\theta_r, \varphi_r)}{4\pi}. \quad (6.18)$$

For polarization matched loss-free antennas aligned for maximum directional radiation and reception,

$$\frac{P_r}{P_t} = \sigma \left(\frac{\lambda}{4\pi R_t R_r} \right)^2 \frac{D_{t0} D_{r0}}{4\pi}. \quad (6.19)$$

The radar range equation is used to calculate the *maximum range of a radar system*. As in the case of Friis' transmission equation, we need to know all parameters of both the transmitting and the receiving antennas, as well as the minimum received power at which the receiver operates reliably. Then,

$$(R_t R_r)_{\max}^2 = e_t e_r (1 - |\Gamma_t|^2)(1 - |\Gamma_r|^2) |\hat{\mathbf{p}}_t^* \cdot \hat{\mathbf{p}}_r|^2 \cdot \frac{P_t}{P_{r\min}} \sigma \left(\frac{\lambda}{4\pi} \right)^2 \frac{D_t(\theta_t, \varphi_t) D_r(\theta_r, \varphi_r)}{4\pi}. \quad (6.20)$$

Finally, we note that the above RCS and radar-range calculations are only basic. The subjects of radar system design and EM scattering are huge research areas themselves and are not going to be considered in detail in this course.

Lecture 7: Antenna Noise Temperature and System Signal-to-Noise Ratio

(Noise temperature. Antenna noise temperature. System noise temperature. Minimum detectable temperature. System signal-to-noise ratio.)

1. Noise Temperature of Bright Bodies

The performance of a telecommunication system depends on the signal-to-noise ratio (SNR) at the receiver's input. The electronic circuitry of the RF front end (amplifiers, mixers, etc.) has a significant contribution to the system noise. However, the antenna itself is sometimes a significant source of noise, too. The antenna noise can be divided into two types according to its physical source: noise due to the loss resistance of the antenna and noise, which the antenna picks up from the surrounding environment.

Any object whose temperature is above the absolute zero radiates EM energy. Thus, an antenna is surrounded by noise sources, which create noise power at the antenna terminals. Here, we are not concerned with technological sources of noise, which are the subject of the EM interference (EMI) science. We are also not concerned with intentional sources of EM interference (EM jamming). We are concerned with natural sources of EM noise, which is thermal in nature, such as sky noise and ground noise.

The concept of antenna noise temperature is critical in understanding how the antenna contributes to the system noise in low-noise radio-astronomy systems. It is also important in understanding the relation between an object's temperature and the power it can generate at the receiving antenna terminals. This thermal power is the signal used in passive remote sensing (radiometry) and imaging. A radiometer can create temperature images of objects. Typically, the remote object's temperature is measured by comparison with the noise due to background sources and the receiver itself.

Every object (e.g., a resistor R) with a physical temperature above zero ($0^\circ \text{K} = -273^\circ \text{C}$) possesses heat energy. The **noise power per unit bandwidth** p_h is proportional to the object's temperature and is given by Nyquist's relation:

$$p_h = kT_p, \text{ W/Hz} \quad (7.1)$$

where T_P is the physical temperature of the object in K (Kelvin degrees) and k is Boltzmann's constant ($\approx 1.38 \times 10^{-23}$ J/K).

In the case of a resistor, this is the noise power, which can be measured at the resistor's terminals with a matched load. Thus, a resistor can serve as a noise generator. Often, we assume that heat energy is evenly distributed in the frequency band Δf . Then, the associated heat power in Δf is

$$P_h = kT_P \Delta f, \text{ W.} \quad (7.2)$$

The noise power radiated by the object depends not only on its physical temperature but also on the ability of its surface to let the heat leak out. This radiated heat power is associated with the so-called *equivalent temperature* or *brightness temperature* T_B of the body via the power-temperature relation in (7.2):

$$P_B = kT_B \Delta f, \text{ W.} \quad (7.3)$$

Note, however, that the brightness temperature T_B is not the same as the physical temperature of the body T_P . The two temperatures are proportional:

$$T_B = (1 - |\Gamma_s|^2) \cdot T_P = \varepsilon T_P, \text{ K} \quad (7.4)$$

where

Γ_s is the reflection coefficient of the surface of the body; and

ε is what is called the *emissivity* of the body.

The brightness power P_B relates to the heat power P_h the same way as T_B relates to T_P , i.e., $P_B = \varepsilon P_h$.

2. Antenna Noise Temperature

The power radiated by the body P_B , when intercepted by an antenna, generates power P_A at its terminals. The equivalent temperature associated with the received power P_A at the antenna terminals is called the *antenna temperature* T_A of the object, where again $P_A = kT_A \Delta f$. Here, Δf is a bandwidth, which falls within the antenna bandwidth and is sufficiently narrow to ensure constant noise-power spectral density.

2.1. Antenna noise from large bright bodies

Let us first assume that the entire antenna pattern (beam) “sees” a uniformly “bright” or “warm” object. To simplify matters, we also assume that the antenna is lossless, i.e., it has no loss resistance, and, therefore, it does not generate noise itself. Then, certain noise power can be measured at its terminals, which can be expressed as

$$P_A = kT_B\Delta f, \text{ W.} \quad (7.5)$$

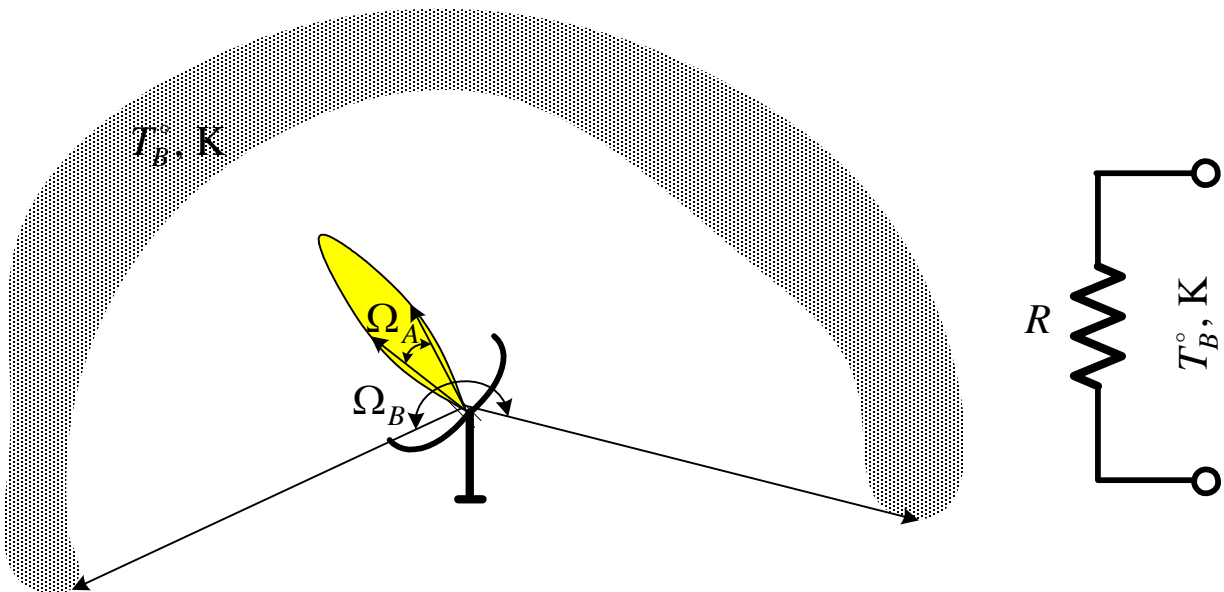
This is the same noise power as that of an equivalent resistor of temperature T_B (K). The temperature T_B is referred to as the ***brightness temperature of the object*** at the antenna terminals.

On the other hand, the antenna temperature is related to the measured noise power as

$$P_A = kT_A\Delta f. \quad (7.6)$$

Thus, in this case (when the solid angle subtended by the noise source Ω_B is much larger than the antenna solid angle Ω_A), the antenna temperature T_A is exactly equal to the object’s temperature T_B (if the antenna is loss-free):

$$T_A = T_B, \text{ if } \Omega_A \ll \Omega_B. \quad (7.7)$$



2.2. Detecting large bright bodies (antenna incremental temperature)

The situation described above is of practical importance. When an antenna is pointed right at the night sky, its noise temperature is very low: $T_A = 3^\circ$ to 5° K at frequencies between 1 and 10 GHz. This is the microwave noise temperature of the night sky. The higher the elevation angle, the lower the sky temperature because of the lower physical temperature of the atmosphere toward the zenith. The sky noise depends on the frequency. It depends on the time of the day, too. Closer to the horizon, it is mostly due to the thermal radiation from the Earth's surface and the atmosphere. Closer to the zenith, it is mostly due to cosmic rays from the sun, the moon and other bright sky objects, as well as the deep-space background temperature commonly referred to as the *cosmic microwave background* ($T_{\text{CMB}} \approx 2.725^\circ$ K).¹ The latter is a left-over thermal effect from the very origin of the universe (the *big bang*).

An antenna may also be pointed toward the ground, e.g., when it is mounted on an airplane or a satellite. The noise temperature of the ground is much higher than that of the night sky because of its much higher physical temperature. The ground noise temperature is about 300° K and it varies during the day. The noise temperature at approximately zero elevation angle (horizon) is about 100° to 150° K.

When a single large bright body is in the antenna beam, (7.7) holds. In practice, however, the antenna temperature may include contributions from several large sources. The source under observation, although large itself, may be superimposed on a background of certain temperature as well as the noise temperature due to the antenna losses, which we initially assumed zero. In order the antenna and its receiver to be able to discern a bright body while “sweeping” the background, this source has to put out more power than the noise power of its background, i.e., it has to be “brighter” than the background noise. Thus, in practice, to obtain the brightness temperature of a large object at the antenna terminals, the antenna temperature is measured with the beam on and off the target. The difference is the *antenna incremental temperature* ΔT_A . If the bright body is large enough to “fill in” the antenna beam completely, the difference between the background-noise antenna temperature and the

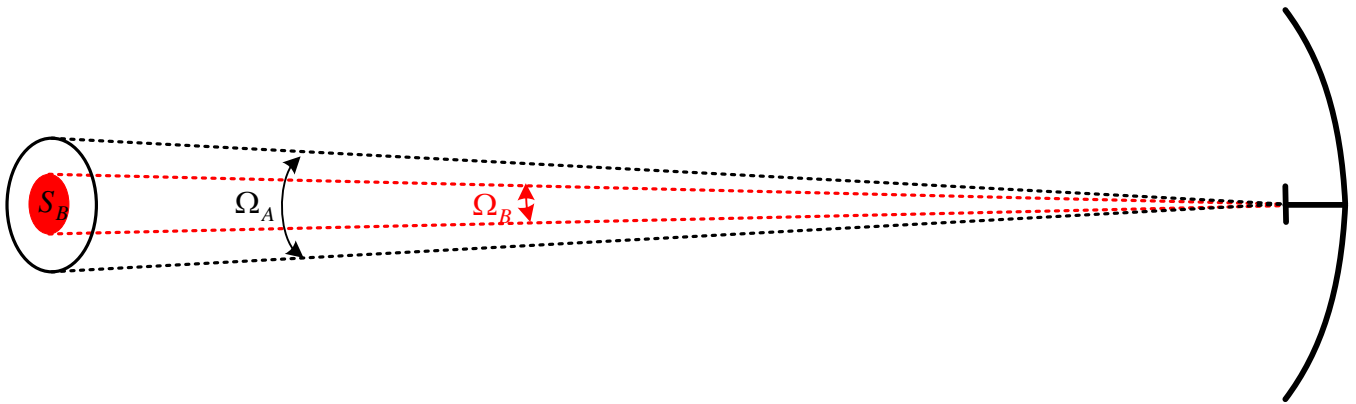
¹ C.T. Stelzried, A.J. Freiley, and M.S. Reid, *Low-noise Receiving Systems*. Artech, 2010.

temperature when the antenna solid angle is on the object is equal to the object's brightness temperature,

$$\Delta T_A = T_B. \quad (7.8)$$

2.3. Antenna noise from small bright bodies

A different case arises in radiometry and radio astronomy. The bright object subtends such a small solid angle that it is well inside the antenna solid angle when the antenna is pointed at it: $\Omega_B \ll \Omega_A$.



To separate the power received from the bright body from the background noise, the difference in the antenna temperature ΔT_A is measured with the beam on and off the object. This time, ΔT_A is *not equal* to the bright body temperature T_B , as it was in the case of a large object. However, both temperatures are proportional. The relation is derived below.

The noise power intercepted by the antenna depends on the antenna effective aperture A_e and on the power density at the antenna's location created by the noise source W_B :

$$P_A = A_e \cdot W_B, \text{ W.} \quad (7.9)$$

Assuming that the bright body radiates isotropically and expressing the effective area by the antenna solid angle, we obtain

$$P_A = \frac{\lambda^2}{\Omega_A} \cdot \frac{P_B}{4\pi R^2}, \text{ W.} \quad (7.10)$$

The distance R between the noise source and the antenna is related to the effective area of the body S_B and the solid angle Ω_B it subtends as

$$R^2 = \frac{S_B}{\Omega_B}, \text{ m}^2 \quad (7.11)$$

$$\Rightarrow P_A \Omega_A = \frac{\lambda^2}{4\pi S_B} P_B \Omega_B. \quad (7.12)$$

Next, we notice that

$$\frac{\lambda^2}{4\pi S_B} = \frac{1}{G_B} = 1. \quad (7.13)$$

Here, G_B is the gain of the bright body (viewed as an antenna), which is unity because we assumed in (7.10) that the body radiates isotropically. In (7.13), we have used the relationship between gain and effective area; see (4.65) in Lecture 4); the effective area of the bright body being simply its cross-section S_B . Finally, substituting (7.13) in (7.12) leads to

$$P_A \Omega_A = P_B \Omega_B, \text{ if } \Omega_B \ll \Omega_A. \quad (7.14)$$

Equation (7.14) leads to the relation between the brightness temperature T_B of the object under observation and the antenna incremental temperature ΔT_A :

$$\Delta T_A = \frac{\Omega_B}{\Omega_A} T_B, \text{ K.} \quad (7.15)$$

For a large bright body, where $\Omega_B = \Omega_A$, (7.15) reduces to (7.8).

2.4. Source flux density from noise sources and noise PLF

The power at the antenna terminals P_A , which corresponds to the antenna incremental temperature ΔT_A , is defined by (7.6). In radio-astronomy and remote sensing, it is often convenient to use the **flux density** S of the noise source at the antenna (the effective area of which is A_e):

$$S = \frac{P_h}{A_e} = \frac{k\Delta T_A}{A_e}, \text{ Wm}^{-2}\text{Hz}^{-1}. \quad (7.16)$$

Notice that S is not the Poynting vector (power flow per unit area) but rather the spectral density of the Poynting vector (power flow per unit area per *hertz*). In

radio-astronomy, the usual unit for flux density is *jansky*, $1 \text{ Jy} = 10^{-26} \text{ Wm}^{-2}\text{Hz}^{-1}$.²

From (7.16), we conclude that the measured incremental antenna temperature ΔT_A relates to the source flux density as

$$\Delta T_A = \frac{1}{k} A_e \cdot S. \quad (7.17)$$

This would be the case indeed if the antenna and the bright-body source were polarization matched. Since the bright-body source is a natural noise source, we cannot expect perfect match. In fact, an astronomical object is typically *unpolarized*, i.e., its polarization is random. Thus, about half of the bright-body flux density cannot be picked up by the receiving antenna, the polarization of which is fixed. For this reason, the relation in (7.17) is modified as

$$\Delta T_A = \frac{1}{2} \cdot \frac{A_e \cdot S}{k}. \quad (7.18)$$

The same correction factor should be inserted in (7.15), where the measured ΔT_A would actually correspond only to one-half of the noise temperature of the bright body:

$$\Delta T_A = \frac{1}{2} \frac{\Omega_B}{\Omega_A} T_B. \quad (7.19)$$

2.5. Antenna noise from a nonuniform noisy background

In the case of a small bright body (see previous subsection), we have tacitly assumed that the gain of the antenna is constant within the solid angle Ω_B subtended by the bright body. This is in accordance with the definition of the antenna solid angle Ω_A , which was used to obtain the ratio between ΔT_A and T_B . The solid-angle representation of the directivity of an antenna is actually quite accurate for high-directivity antennas, e.g., reflector antennas.

In general, however, the antenna gain may be strongly dependent on the observation angle (θ, φ) . In this case, the noise signals arriving from different sectors of space have different contributions to the total antenna temperature. Those arriving from the direction of the maximum directivity contribute the

² Karl G. Jansky was the first one to use radio waves for astronomical observations.

most whereas those arriving from the direction of zero directivity will not contribute at all. The differential contribution from a sector of space of solid angle $d\Omega$ should, therefore, be weighed by the antenna normalized power pattern $\bar{F}(\theta, \varphi)$ in the respective direction:

$$dT_A = \bar{F}(\theta, \varphi) \cdot \frac{T_B(\theta, \varphi) d\Omega}{\Omega_A}. \quad (7.20)$$

The above expression can be understood by considering (7.15) where ΔT_A is replaced by a differential contribution dT_A to the antenna temperature from a bright body subtending a differential solid angle $\Omega_B \rightarrow d\Omega$. The total antenna noise power is finally obtained as

$$T_A = \frac{1}{\Omega_A} \oint\!\!\!\oint_{4\pi} \bar{F}(\theta, \varphi) \cdot T_B(\theta, \varphi) d\Omega. \quad (7.21)$$

The expression in (7.21) is general and the previously discussed special cases are easily derived from it. For example, assume that the brightness temperature surrounding the antenna is the same in all directions, i.e., $T_B(\theta, \varphi) = T_{B0} = \text{const}$. Then,

$$T_A = \frac{T_{B0}}{\Omega_A} \cdot \underbrace{\oint\!\!\!\oint_{4\pi} \bar{F}(\theta, \varphi) d\Omega}_{\Omega_A} = T_{B0}. \quad (7.22)$$

The above situation was already addressed in equation (7.7).

Assume now that $T_B(\theta, \varphi) = \text{const} = T_{B0}$ but only inside a solid angle Ω_B , which is much smaller than the antenna solid angle Ω_A . Outside Ω_B , $T_B(\theta, \varphi) = 0$. Since $\Omega_B \ll \Omega_A$, when the antenna is pointed at the noise source, its normalized power pattern within Ω_B is $\bar{F}(\theta, \varphi) \approx 1$. Then,

$$T_A = \frac{1}{\Omega_A} \oint\!\!\!\oint_{4\pi} \bar{F}(\theta, \varphi) \cdot T_B(\theta, \varphi) d\Omega = \frac{1}{\Omega_A} \int\!\!\int_{\Omega_B} 1 \cdot T_{B0} \cdot d\Omega = T_{B0} \frac{\Omega_B}{\Omega_A}. \quad (7.23)$$

This case was addressed in (7.15).

The antenna pattern strongly influences the antenna temperature. High-gain antennas (such as reflector systems), when pointed at elevation angles close to the zenith at night, have negligible noise level. However, if an antenna has significant side and back lobes, which are pointed toward the ground or the

horizon, its noise power is much higher. The worst case for an antenna is when its main beam points towards the ground or the horizon, as is often the case with satellite or airborne antennas pointed toward the earth.

Example (modified from Kraus, p. 406): A circular reflector antenna of 500 m² effective aperture operating at $\lambda = 20$ cm is directed at the zenith. What is the total antenna temperature assuming the sky temperature close to zenith is equal to 10° K, while at the horizon it is 150° K? Take the ground temperature equal to 300° K and assume that one-half of the minor-lobe beam is in the back direction (toward the ground) and one-half is toward the horizon. The main beam efficiency ($BE = \Omega_M / \Omega_A$) is 0.7.

Such a large reflector antenna is highly directive and, therefore, its main beam “sees” only the sky around the zenith. The main beam efficiency is 70%. Thus, substituting in (7.23), the noise contribution of the main beam is

$$T_A^{MB} = \frac{1}{\Omega_A} (10 \times 0.7 \times \Omega_A) = 7, \text{ K.} \quad (7.24)$$

The contribution from the half back-lobe (which is a half of 30% of the antenna solid angle) directed toward ground is

$$T_A^{GBL} = \frac{1}{\Omega_A} (300 \times 0.15 \times \Omega_A) = 45, \text{ K.} \quad (7.25)$$

The contribution from the half back-lobe directed toward the horizon is

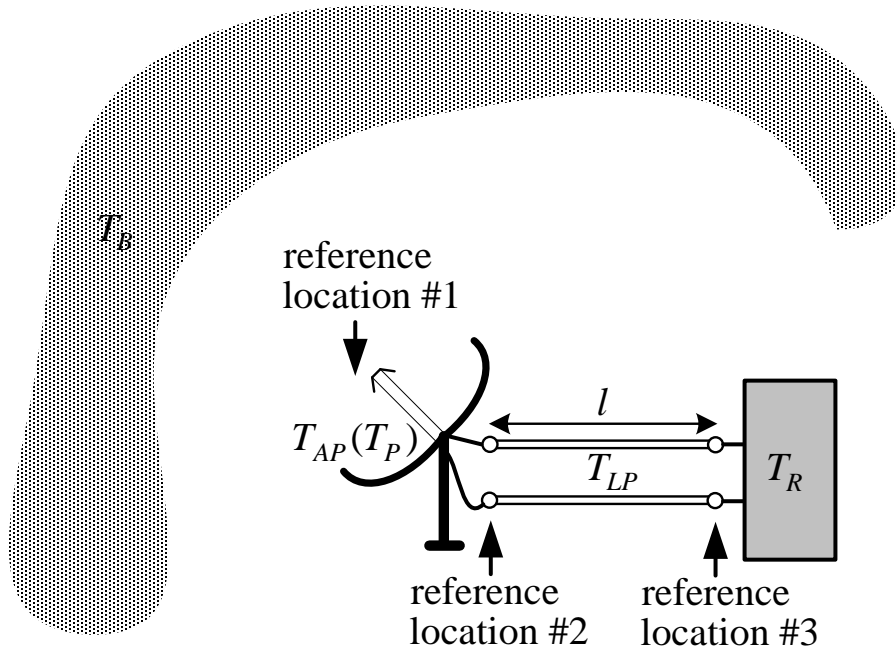
$$T_A^{HBL} = \frac{1}{\Omega_A} (150 \times 0.15 \times \Omega_A) = 22.5, \text{ K.} \quad (7.26)$$

The total noise temperature is

$$T_A = T_A^{MB} + T_A^{GBL} + T_A^{HBL} = 74.5 \text{ K.} \quad (7.27)$$

3. System Noise Temperature

The antenna is a part of a receiving system, which consists of several cascaded components: antenna, transmission line (or waveguide) assembly and receiver (see figure below). All these system components, the antenna included, have their contributions to the system noise. The system noise level is a critical factor in determining its sensitivity and SNR.



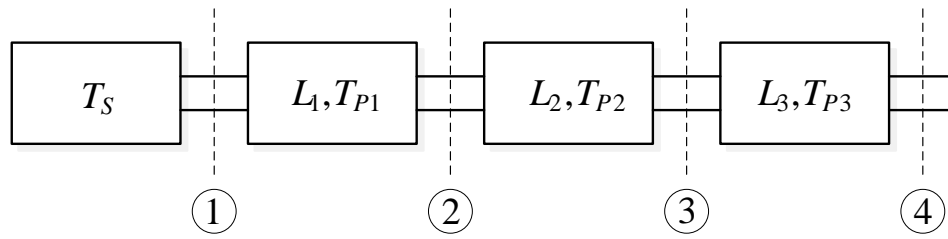
3.1. Noise Analysis of Cascaded Matched Two-port Networks³

To understand the noise analysis of the radio receiver system, we must first review the basics of the noise analysis of cascaded two-port networks. For simplicity, we will assume that all networks are impedance matched, which is close to what is in fact happening in a realistic receiver system.

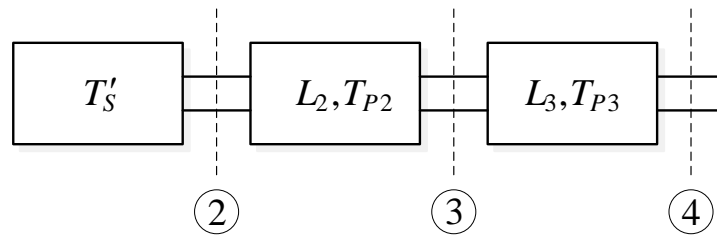
In the figure below (case (a)), a generic cascaded network is shown where the first component on the left is the noise source (e.g., the antenna pointed at the sky) with noise temperature T_S . The remaining two-port components are characterized by their physical temperatures T_{Pi} and by their *loss factors* (or *loss ratios*) L_i , $i = 1, 2, \dots$. In the case of a passive lossy two-port network (such as the waveguide), L is the inverse of the efficiency. In some analyses, the antenna can be viewed as a two-port network as well such that its “input port” is its aperture receiving noise signals from the environment and its output port is at its connection to the transmission line. The efficiency is defined as the output-to-input power ratio $e = P_{ou} / P_{in}$ and it is less or equal to 1. In contrast, $L = P_{in} / P_{ou}$ and it is greater or equal to 1. In the case of an antenna, $L_A = e_A^{-1}$. Also, any two-port network component for which $L \geq 1$, i.e., it exhibits power loss, may be referred to as “attenuator” although this component does not

³ From T.Y. Otoshi, “Calculation of antenna system noise temperatures at different ports—revisited,” *IPN Progress Report*, Aug. 15, 2002.

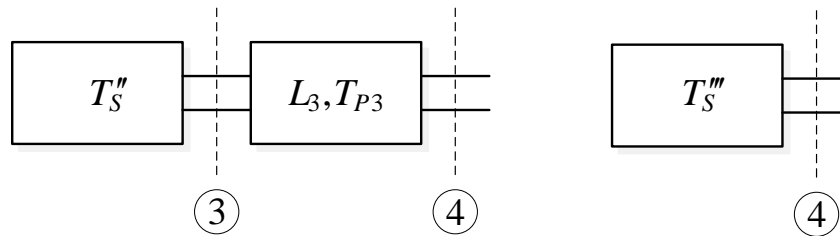
necessarily need to be an attenuator; it could be, for example, the entire antenna-plus-feed assembly. On the other hand, if $L < 1$, we have a component which exhibits gain and it is referred to as an “amplifier”. In this case, the efficiency is replaced by the gain G , which, just like the efficiency is the output-to-input power ratio P_{ou} / P_{in} but it is greater than 1. As with the efficiency, the relationship $L = G^{-1}$ holds.



(a) original network



(b) equivalent source noise temperature at location 2



(c) equivalent source noise temperatures at locations 3 and 4

Case (b) in the figure above, shows a network where an equivalent source of temperature T'_S replaces the original source and its neighboring two-port network (L_1, T_{P1}) . The equivalent source of temperature T'_S at location 2 is

$$T'_S = L_1^{-1}T_S + (1 - L_1^{-1})T_{P1}. \quad (7.28)$$

From (7.28), it is evident that in addition to the usual “attenuated” source noise-power term $L_1^{-1}T_S$ there is a contribution due to the physical temperature of the 1st two-port network. This contribution is referred to as the ***device equivalent noise temperature at its output***,

$$T_{D1}^{\text{ou}} = (1 - L_1^{-1})T_{P1}. \quad (7.29)$$

This contribution is entirely determined by the device physical temperature and its loss ratio, i.e., it does not depend on the source.

To understand where (7.28) comes from, we can re-write it as

$$L_1 = \frac{T_S - T_{P1}}{T'_S - T_{P1}}. \quad (7.30)$$

This is indeed the ratio of input-to-output noise power for the 1st network. T_S represents the power traveling toward the device input while its own noise power, represented by T_{P1} travels away from it. Thus, the total power at the input is represented by $T_S - T_{P1}$.⁴ At the same time, at the output, as per case (b) in the figure, the total noise power incident toward network #2 is given by T'_S . However, the portion that relates to the attenuation L_1 (i.e., the power at the device input) does not include the intrinsic device noise power T_{P1} , which is always present at the device output regardless of whether there is a noise source at the input or not. Thus, T_{P1} has to be subtracted from T'_S .

Using the same methodology, we can find the equivalent source noise temperature T_S'' at location 3 as

$$T_S'' = L_2^{-1}T'_S + (1 - L_2^{-1})T_{P2} \quad (7.31)$$

where

$$T_{D2}^{\text{ou}} = (1 - L_2^{-1})T_{P2} \quad (7.32)$$

is the 2nd device equivalent noise temperature at its output.

We can repeat this step for the network location 4 where we obtain the equivalent source noise temperature T_S''' . In each case, in addition to the “attenuated” source power we will have to add the respective network ***equivalent output device noise temperature***,

$$T_{Di}^{\text{ou}} = (1 - L_i^{-1})T_{Pi}, \quad i = 1, 2, \dots \quad (7.33)$$

As an illustration of the general procedure, we show the results for the equivalent source noise temperature T_S'''' at location 4:

$$T_S'''' = T_S(L_1L_2L_3)^{-1} + (1 - L_1^{-1})T_{P1}(L_2L_3)^{-1} + (1 - L_2^{-1})T_{P2}L_3^{-1} + (1 - L_3^{-1})T_{P3}. \quad (7.34)$$

⁴ Remember the expression $|a|^2 - |b|^2$ for the total power at the input of a microwave network where a and b are the incident and the scattered (outgoing) root-power waves, respectively. $|a|^2$ represents the incoming power whereas $|b|^2$ represents the outgoing power.

3.2. Transferring System Noise Temperature through a Lossy Network

The rule of transferring noise temperature from the output port of a lossy network to its input port (or *vice versa*) is simple:

$$T_{\text{in}} = LT_{\text{ou}} = T_{\text{ou}} / e \quad (7.35)$$

where e is the device efficiency. This rule, while simple, is not immediately obvious. A formal proof can be found in the Appendix of

B.L. Seidel and C.T. Stelzried, “A radiometric method for measuring the insertion loss of radome materials,” *IEEE Trans. Microw. Theory Thech.*, vol. MTT-16, No. 9, Sep. 1968, pp. 625–628.

We can now define the ***equivalent noise temperature of a lossy component at its input*** (also known as ***equivalent input device noise temperature***) by substituting $T_{D_i}^{\text{ou}}$ from (7.33) as T_{ou} in (7.35):

$$T_{D_i}^{\text{in}} = L_i T_{D_i}^{\text{ou}} = (L_i - 1) T_{P_i}. \quad (7.36)$$

It is worth noting that (7.36) suggests that $T_{D_i}^{\text{in}}$ could be much larger than the physical temperature T_{P_i} if the device is very lossy, i.e., if $L_i \gg 1$ ($e_i \ll 1$).

Finally, we discuss the physical meaning of the ***equivalent input device noise temperature*** through an alternative way of deriving the relationship in (7.36). We omit the subscript i hereafter. Consider a noise source of temperature T_S at the device input. Its noise power is then $kT_S \Delta f$. At the output of the device, we add the two input contributions – that of the noise source and that due to the equivalent input device noise temperature, and then multiply the result by the device efficiency:

$$P_{N,\text{ou}} = e(kT_S \Delta f + kT_D^{\text{in}} \Delta f). \quad (7.37)$$

To find the relation between the equivalent input device noise temperature T_D^{in} and its physical temperature T_P , we consider the particular case when the temperature of the source T_S is equal to the physical temperature T_P of the device. In this case, the output noise power must be $P_{N,\text{ou}} = kT_P \Delta f$ because the whole system of the lossy device plus the source is at the physical temperature T_P . Substituting $T_S = T_P$ in (7.37) results in

$$P_{N,\text{ou}} = e(kT_P \Delta f + kT_D^{\text{in}} \Delta f) = kT_P \Delta f \quad (7.38)$$

which, when solved for T_D^{in} , produces (7.36). Note that we have not imposed any restrictions on the actual values of T_S and T_P but have only required that T_D^{in} depends solely on T_P (i.e., it is independent of the noise source at the input) and that (7.37) holds in the special case of $T_S = T_P$.

3.3. The atmosphere as an “attenuator”

An illustration of the above concepts in noise analysis is the impact of the atmosphere on the sky noise, e.g., the cosmic microwave background ($T_{\text{CMB}} \approx 2.725^\circ \text{ K}$). The atmosphere, depending on the time of the day and the weather conditions, exhibits loss, which we describe by the loss factor L_{atm} . L_{atm} can be calculated if we know the averaged attenuation constant in the atmosphere α_{atm} and its thickness H , e.g., $L_{\text{atm}} \approx \exp(2\alpha_{\text{atm}}H)$. This atmospheric “attenuator” lies between the cosmic microwave background noise source and the antenna. Therefore, the actual external noise temperature perceived by the antenna is

$$T_{\text{sky}} = L_{\text{atm}}^{-1}T_{\text{CMB}} + (1 - L_{\text{atm}}^{-1})T_{\text{atm},P} \quad (7.39)$$

where $T_{\text{atm},P}$ is the physical temperature of the atmosphere, as per (7.28). The 1st term in (7.39) is the *space noise* whereas the 2nd one is the *atmospheric noise*. The impact of the atmosphere is often considered negligible. For a pencil-beam antenna pointed at the sky, $T_A = T_{\text{sky}}$.

3.4. Antenna noise due to the antenna physical temperature

If the antenna has losses, the noise temperature at its terminals includes not only the antenna temperature T_A due to the environment surrounding the antenna (the *external* antenna temperature) but also the antenna equivalent noise temperature T_{AP} due to its physical temperature T_P . Here, we note that the antenna acts as an “attenuator” in the cascaded network consisting of the external noise, the antenna, the waveguide and the receiver; see Figure on p. 10.

We first describe the antenna noise contribution at reference location #1, the antenna aperture, or, equivalently, its input. Here, we view the antenna as a lossy two-port component. Its equivalent input noise temperature T_{AP} is

$$T_{AP} = \left(\frac{1}{e_A} - 1 \right) T_P = \frac{R_l}{R_r} T_P, \text{ K} \quad (7.40)$$

where e_A is the radiation efficiency ($0 \leq e_A \leq 1$), R_l is the antenna loss resistance and R_r is its radiation resistance. Eq. (7.40) is essentially an application of Eq. (7.36) to the case of an antenna. It describes the thermal noise contribution of the antenna due to its physical temperature T_P . T_{AP} must be added to T_A in order to obtain the system operating noise temperature at location #1. In fact, additional terms exist due to the noise contributions of the lossy TL (or waveguide) and the receiver electronics.

3.5. Noise due to the physical temperature of the transmission line

We now consider the transmission line (TL) as a source of noise when it has conduction losses. In a manner analogous to the one applied to the antenna, the TL is considered as a two-port “attenuator”. Thus, its noise contribution at the antenna terminals (the input to the TL or reference location #2) is

$$T_{L2} = \left(\frac{1}{e_L} - 1 \right) T_{LP}, \text{ K.} \quad (7.41)$$

Here, $e_L = e^{-2\alpha l}$ is the **line thermal efficiency** ($0 \leq e_L \leq 1$), T_{LP} is the physical temperature of the TL, α (Np/m) is the attenuation constant of the TL, and l is its length.

To transfer the TL noise contribution to the reference location #1, we use (7.35) which leads to

$$T_{L1} = \frac{T_{L2}}{e_A} = \frac{1}{e_A} \left(\frac{1}{e_L} - 1 \right) T_{LP}. \quad (7.42)$$

Together with T_{AP} , T_{L1} must be added to T_A in order to obtain the system operating noise temperature at location #1.

3.6. System noise referred to the antenna aperture (location #1)

The system temperature referred to the antenna aperture includes the contributions of the antenna (external noise temperature plus equivalent input antenna thermal noise temperature), the transmission line and the receiver as

$$T_{\text{sys}}^A = \underbrace{T_A}_{\text{antenna external}} + \underbrace{T_P \left(\frac{1}{e_A} - 1 \right)}_{T_{AP}, \text{ antenna internal}} + \underbrace{\frac{1}{e_A} T_{LP} \left(\frac{1}{e_L} - 1 \right)}_{\text{TL internal}} + \underbrace{\frac{1}{e_A e_L} T_R}_{\text{receiver}}. \quad (7.43)$$

Here, T_A is the external temperature that corresponds to the antenna temperature provided the antenna is loss-free, as discussed in Section 2. T_R is the receiver noise temperature (at its input, reference location #3). It is given by

$$T_R = T_1 + \frac{T_2}{G_1} + \frac{T_3}{G_1 G_2} + \dots, \text{ K.} \quad (7.44)$$

Here,

T_1 is the noise temperature of the first amplifying stage;

G_1 is the gain of the first amplifying stage;

T_2 is the noise temperature of the second amplifying stage;

G_2 is the gain of the second amplifying stage.

Notice that T_R is divided by the TL efficiencies e_L and e_A in order to refer it to the TL input (location #2) and then to the antenna aperture (location #1); see Eq. (7.35).

3.7. System noise referred to the antenna terminals (TL input, location #2)

The reference location is changed by taking into account the efficiency of the antenna. Using (7.35), we arrive at:

$$T_{sys}^{TL} = T_{sys}^A \cdot e_A. \quad (7.45)$$

Therefore,

$$T_{sys}^{TL} = \underbrace{T_A e_A}_{\text{antenna external}} + \underbrace{T_P (1 - e_A)}_{\text{antenna internal}} + \underbrace{T_{LP} \left(\frac{1}{e_L} - 1 \right)}_{\text{TL internal}} + \underbrace{\frac{1}{e_L} T_R}_{\text{receiver}}. \quad (7.46)$$

3.8. System noise referred to the receiver input (location #3)

The reference location is changed once again by taking into account the efficiency of the TL:

$$T_{sys}^R = T_{sys}^{TL} \cdot e_L. \quad (7.47)$$

Therefore,

$$T_{sys}^R = \underbrace{T_A e_A e_L}_{\text{antenna external}} + \underbrace{T_P (1 - e_A) e_L}_{\text{antenna internal}} + \underbrace{T_{LP} (1 - e_L)}_{\text{TL}} + \underbrace{T_R}_{\text{receiver}}, \text{ K.} \quad (7.48)$$

Example (from Kraus, p. 410, modified): A receiver has an antenna with an external noise temperature 50° K , a physical temperature of 300° K , and an efficiency of 99%. Its transmission line has a physical temperature of 300° K and an efficiency of 90%. The first three stages of the receiver all have 80° K noise temperature and 13 dB gain (13 dB is about 20 times the power). Find the system temperature at: (a) the antenna aperture, (b) the antenna terminals, and (c) the receiver input.

The receiver noise temperature is

$$T_R = 80 + \frac{80}{20} + \frac{80}{20^2} = 84.2 \text{ °K.} \quad (7.49)$$

(a) Then, the system temperature at the antenna aperture is

$$T_{sys}^A = T_A + T_P \left(\frac{1}{e_A} - 1 \right) + \frac{1}{e_A} T_{LP} \left(\frac{1}{e_L} - 1 \right) + \frac{1}{e_A e_L} T_R, \quad (7.50)$$

$$T_{sys}^A = 50 + 300 \left(\frac{1}{0.99} - 1 \right) + \frac{300}{0.99} \left(\frac{1}{0.9} - 1 \right) + \frac{84.2}{0.99 \cdot 0.9} \approx 181.2009 \text{ K.}$$

(b) The system temperature at the antenna terminals is

$$T_{sys}^{TL} = T_{sys}^A \cdot e_A \approx 181.2009 \cdot 0.99 \approx 180.3889 \text{ °K.}$$

(c) The system temperature at the receiver input is

$$T_{sys}^R = T_{sys}^{TL} \cdot e_L = 180.3889 \cdot 0.9 \approx 162.35 \text{ °K.}$$

4. Minimum Detectable Temperature (Sensitivity) of the System

The minimum detectable temperature, or sensitivity, of a receiving system ΔT_{\min} is the *RMS* noise temperature of the system ΔT_{rms} , which, when referred to the antenna aperture (reference location #1), is

$$\Delta T_{\min} = \Delta T_{\text{rms}} = \frac{k' T_{sys}^A}{\sqrt{\Delta f \cdot \tau}}, \quad (7.51)$$

where

k' is a system constant (commensurate with unity), dimensionless;

Δf is the pre-detection bandwidth of the receiver, Hz;

τ is the post-detection time constant, s.

The RMS noise temperature ΔT_{rms} is determined experimentally by pointing the antenna at a uniform bright object and recording the signal for a sufficiently long time. Assume the output of the receiver is in terms of real-positive numbers proportional to the received noise power. Modern receivers are digital and their output is actually in the form of integers. Then, the *RMS* deviation D_{rms} of the numbers produced by the receiver represents the *RMS* noise power:

$$D_{\text{rms}} = \sqrt{\frac{1}{N} \sum_{n=1}^N (a_n - a_{av})^2} \approx \Delta T_{\text{rms}}^R \text{ where } a_{av} = \frac{1}{N} \sum_{n=1}^N a_n. \quad (7.52)$$

ΔT_{rms} (at reference location #1) can be obtained from ΔT_{rms}^R by

$$\Delta T_{\text{rms}} = \frac{\Delta T_{\text{rms}}^R}{e_A e_L} = \Delta T_{\text{min}}. \quad (7.53)$$

This is the *sensitivity* of the system in terms of noise temperature.

In order a source to be detected, it has to create an incremental antenna temperature ΔT_A which exceeds ΔT_{min} , $\Delta T_A > \Delta T_{\text{min}}$. The *minimum detectable power* P_{min} is thus

$$P_{\text{min}} = 0.5 A_e p_{\text{min}} = k \Delta T_{\text{min}} \Delta f \quad (7.54)$$

where A_e is the effective antenna area, p_{min} is the power-flux density (magnitude of Poynting vector) due to the source at the location of the antenna, and the factor of 0.5 accounts for the randomness of the wave polarization. It follows that the minimum power-flux density, which can be detected is

$$p_{\text{min}} = \frac{2k \Delta T_{\text{min}} \Delta f}{A_e}. \quad (7.55)$$

The signal-to-noise ratio (SNR) for a signal source of incremental antenna temperature ΔT_A is given by

$$SNR = \frac{\Delta T_A}{\Delta T_{\text{min}}}. \quad (7.56)$$

This SNR is used in radio-astronomy and remote sensing.

5. System Signal-to-Noise Ratio (SNR) in Communication Links

The system noise power at the antenna terminals (location #2) is

$$P_N = k T_{\text{sys}}^{TL} \Delta f_r, \text{ W}. \quad (7.57)$$

Here, Δf_r is the bandwidth of the receiver and $T_{\text{sys}}^{TL} = e_A T_{\text{sys}}^A$. From Friis' transmission equation, we can calculate the received power at the antenna terminals as

$$P_r = (1 - |\Gamma_t|^2)(1 - |\Gamma_r|^2) \text{PLF} \left(\frac{\lambda}{4\pi R} \right)^2 G_t(\theta_t, \varphi_t) G_r(\theta_r, \varphi_r) \cdot P_t. \quad (7.58)$$

Finally, the SNR becomes

$$SNR = \frac{P_r}{P_N} = \frac{(1 - |\Gamma_t|^2)(1 - |\Gamma_r|^2) \text{PLF} \left(\frac{\lambda}{4\pi R} \right)^2 G_t G_r \cdot P_t}{k T_{\text{sys}}^{TL} \Delta f}. \quad (7.59)$$

The above equation is fundamental for the design of telecommunication systems. More specifically, if the SNR necessary for the adequate operation of the receiver is known, Eq. (7.59) allows for determining the maximum range over which the communication link is stable.

LECTURE 8: Basic Methods in Antenna Measurements

(Antenna ranges and anechoic chambers. Measuring far-field patterns, gain, directivity, radiation efficiency, input impedance and polarization.)

1. Introduction*

Many of the basic methods for measuring antenna characteristics were developed before and during World War II. However, new approaches and measurement technologies continue to emerge boosted by the rapid growth of mobile communications and wireless networks. The methods for measuring antenna far-field patterns, polarization, input impedance, gain and directivity have been developed in conjunction with the design of novel radiating structures, which are needed in the telecommunications and radar technologies.

Antenna metrology requires not only sound theoretical background in antenna theory and radiation but also sophisticated equipment capable of providing the necessary accuracy and purity of the measured data. Commercial equipment specifically designed for antenna measurements became available in the 1960s due, in part, to the requirements of the aerospace, space and defence industries.

The antenna measurement equipment includes: antenna ranges, antenna positioners, pattern recorders, scalar and/or vector network analyzers, signal generators, antenna gain standards, etc. Later on, sophisticated computer systems were developed to provide automated control of pattern measurements as well as fast calculations related to antenna directivity, 2-D to 3-D pattern conversion, near-to-far field transformations (in compact antenna ranges), etc.

2. General Requirements for Antenna Measurement Procedures*

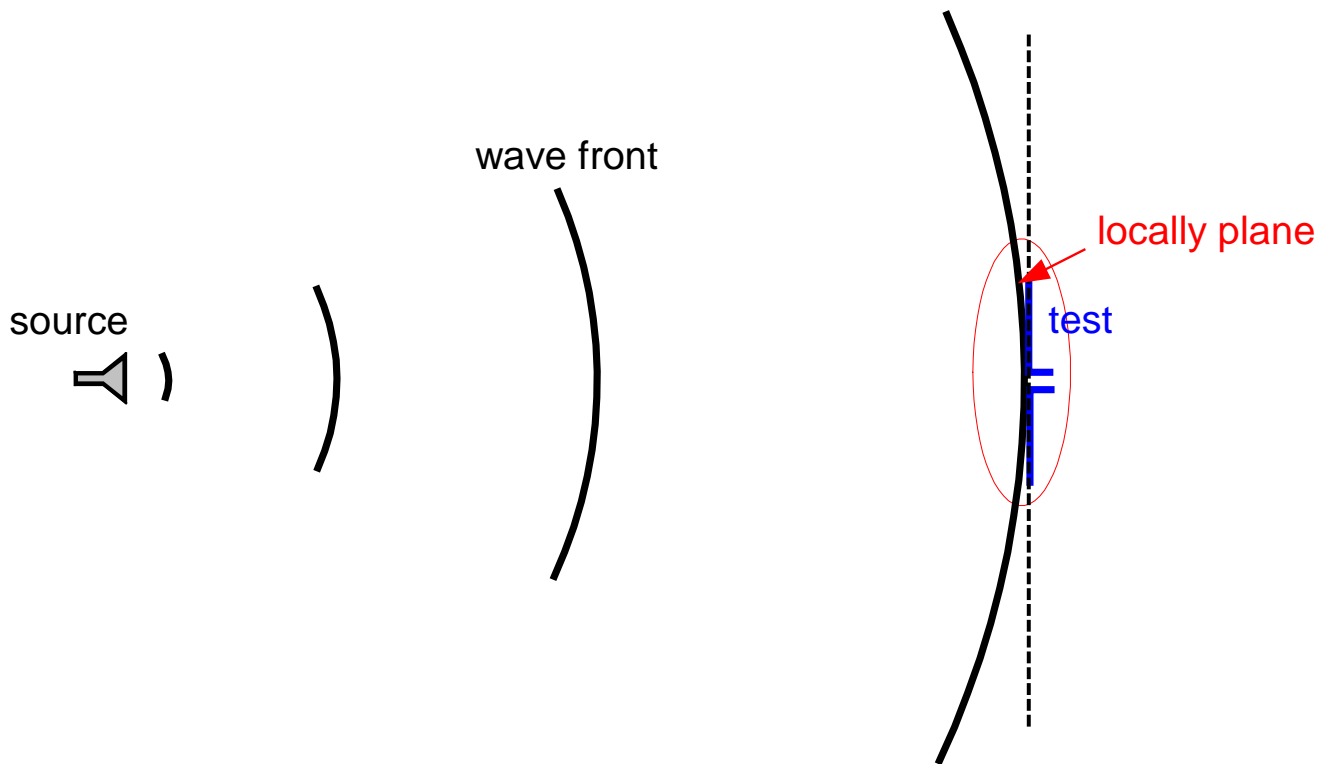
The ideal condition for measuring the far-field pattern and antenna gain is an *illumination by a uniform plane wave*. This is a wave, which has a plane wave front with the field vectors being constant over an area that extends well beyond the aperture of the antenna under test (AUT). For example, the \mathbf{E} field vector of a uniform non-attenuating plane wave propagating in the $+z$ -direction is described by the 1-D wave expression

$$\mathbf{E}(z) = \hat{\mathbf{p}}_w E_m e^{-jkz}. \quad (1)$$

Here, $\hat{\mathbf{p}}_w$ is the wave polarization vector, which must remain constant within

the volume of the AUT. The same holds for the magnitude E_m , which must remain constant across the AUT aperture.

In practice, antennas generate far fields in 3-D space which are closely approximated by spherical wave fronts when the observation point is sufficiently far from the source. Also, at large distances from the source antenna, the curvature of the phase front is small at the aperture of the AUT and it is well approximated by a uniform plane wave.

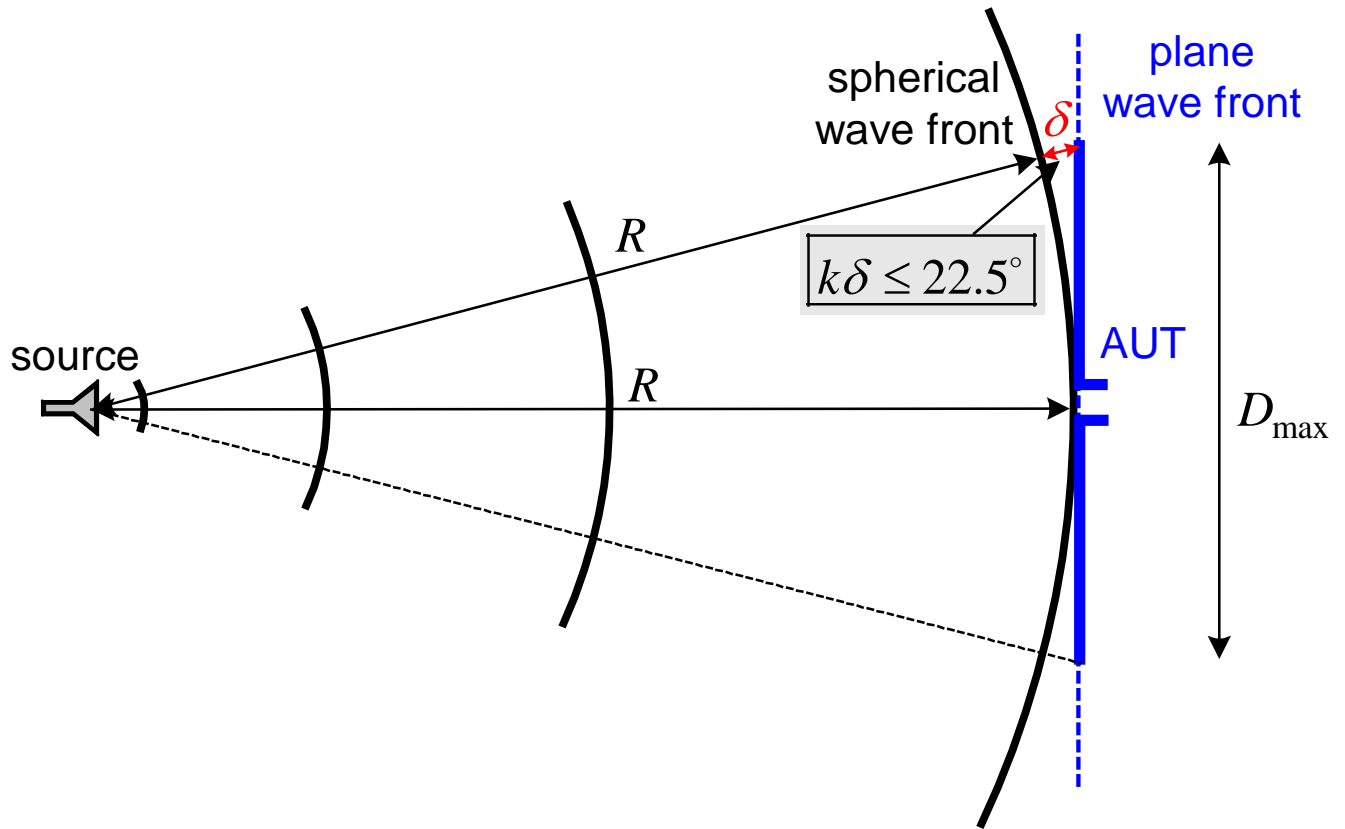


If the distance from the source is equal or greater than the inner boundary of the far-field region $R_{\min} = 2D_{\max, \text{Tx}}^2 / \lambda$, then the maximum phase difference between the actual incident field and its far-zone approximation (remember the 1st order binomial approximation $R \approx r - D_{\max, \text{Tx}} / 2$) does not exceed $e_{\max} \approx 22.5^\circ = \pi / 8$ rad. Here, $D_{\max, \text{Tx}}$ is the maximum dimension of the source (or transmitting) antenna.

Conversely, we can show that if $D_{\max, \text{Rx}} \equiv D_{\max}$ is the maximum dimension of the receiving AUT, a distance

$$R_{\min} = 2D_{\max}^2 / \lambda \quad (2)$$

from the source of a spherical wave ensures that the maximum phase difference between a plane wave and the spherical wave at the aperture of the AUT is $e_{\max} \approx 22.5^\circ = \pi / 8$ rad. Consider a source of a spherical wave and an AUT located a distance R away.



The largest phase difference between the spherical wave and the plane wave appears at the edges of the AUT, which corresponds to the difference in the wave paths δ . This phase difference must fulfil the requirement:

$$k\delta \leq \pi / 8. \quad (3)$$

The difference in the wave paths δ is determined by noticing that

$$(R + \delta)^2 = R^2 + (D_{\max} / 2)^2. \quad (4)$$

The real-positive solution of this quadratic equation for δ is

$$\delta = \sqrt{R^2 + (D_{\max} / 2)^2} - R. \quad (5)$$

Next, the above expression is approximated by the use of the binomial

expansion (the first two terms only) as

$$\delta = R \left[\sqrt{1 + \left(\frac{D_{\max}}{2R} \right)^2} - 1 \right] \approx R \left[1 + \frac{1}{2} \left(\frac{D_{\max}}{2R} \right)^2 - 1 \right] = \frac{D_{\max}^2}{4R}. \quad (6)$$

The minimum distance from the source of the spherical wave is now determined from the requirement in (3),

$$k \frac{D_{\max}^2}{4R} = \frac{2\pi}{\lambda} \frac{D_{\max}^2}{4R} \leq \frac{\pi}{8}. \quad (7)$$

Thus,

$$\boxed{R_{\min} = 2D_{\max}^2 / \lambda}. \quad (8)$$

It is now clear that the antenna far-field characteristics must be measured at a sufficiently large distance between the source antenna and the AUT. ***This distance must be greater than the larger of the two inner limits of the far zones of the transmitting and receiving antennas, i.e., the two antennas must be in each other's far zones.***

The above requirement leads to a major difficulty in antenna measurements – large separation distances are required between the source antenna and the AUT. The larger the AUT, the larger the measurement site. While the size of the site may not be a problem, securing its reflection-free, noise-free, and EM interference-free environment is extremely difficult.

Special attention must be paid to minimizing unwanted reflections from nearby objects (equipment, personnel, buildings), from the ground or the walls of the site. This makes the open sites for antenna measurements (***open ranges***) a rare commodity since they have to provide free-space propagation. Such ideal conditions are found only in unpopulated (desert) areas of predominantly flat terrain. The other alternative is offered by indoor chambers (***anechoic chambers***), which minimize reflections by special wall lining with RF/microwave absorbing material. They are much preferred to open ranges because of their controlled environment. Unfortunately, the anechoic chambers are very expensive and often they cannot accommodate large antennas.

There are cases in which the antenna operates in a very specific environment (mounted on an aircraft, mobile system, etc.). Then, it is better to measure the antenna as it is mounted, i.e., in its own environment. Such measurements are

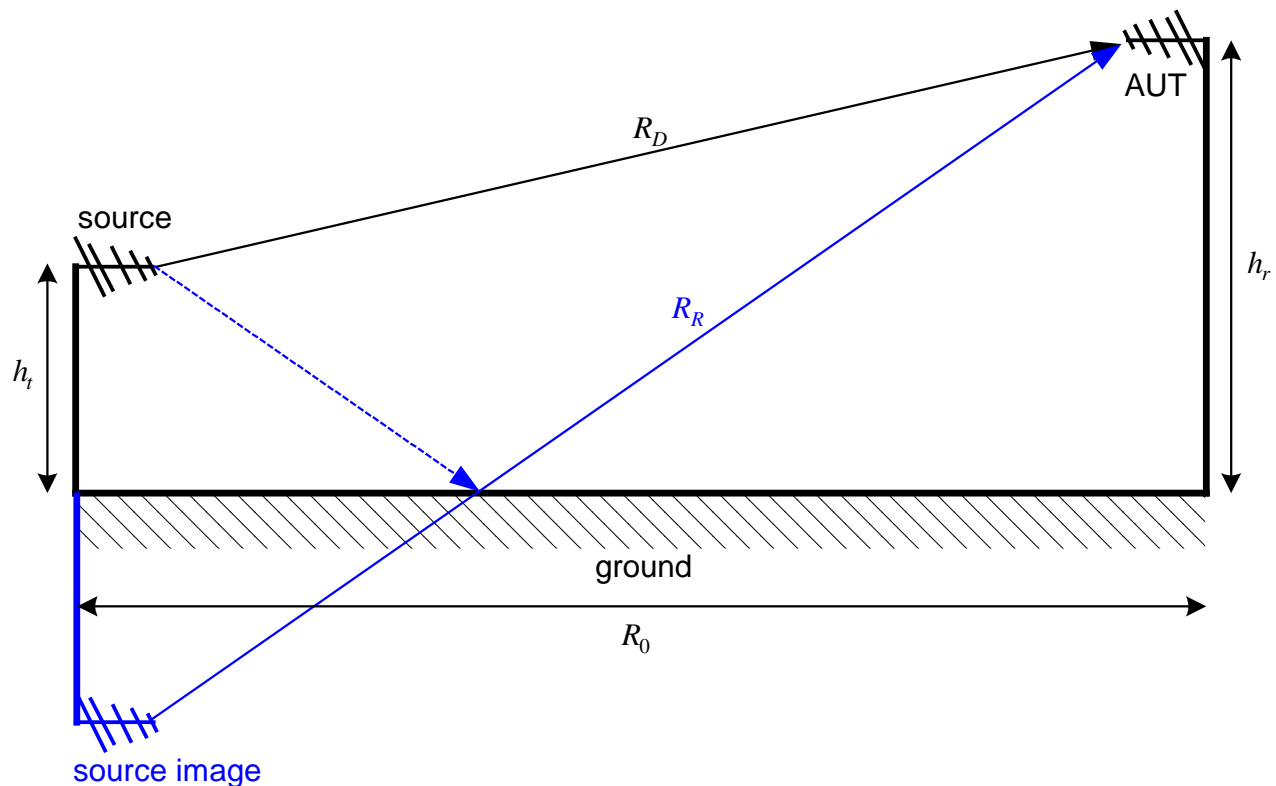
very specific and usually cannot be performed in anechoic chambers.

Below is a summary of the challenges in antenna measurements:

- affected by unwanted reflections;
- often require too large separation distances;
- very complicated when a whole antenna system (e.g., on-craft mounted antenna) is to be measured;
- outdoor sites have uncontrollable EM environment, which, besides all, depends on the weather;
- indoor sites cannot accommodate large antenna systems;
- the instrumentation is expensive.

3. Antenna Ranges (AR)*

The antenna measurement sites are called *antenna ranges* (AR). They can be categorized as outdoor ranges and indoor ranges (anechoic chambers). According to the principle of measurement, they can be also categorized as reflection ranges, free-space ranges, and compact ranges.



The reflection ranges are designed so that the direct and reflected (usually from ground) waves interfere constructively and form a uniform (in both magnitude and phase) wave front in the region of the AUT. Such a region is called the *quite zone*. Reflection ranges are usually of the outdoor type. They are used to measure antennas of moderately broad patterns operating in the UHF frequency bands (500 MHz to 1000 MHz).

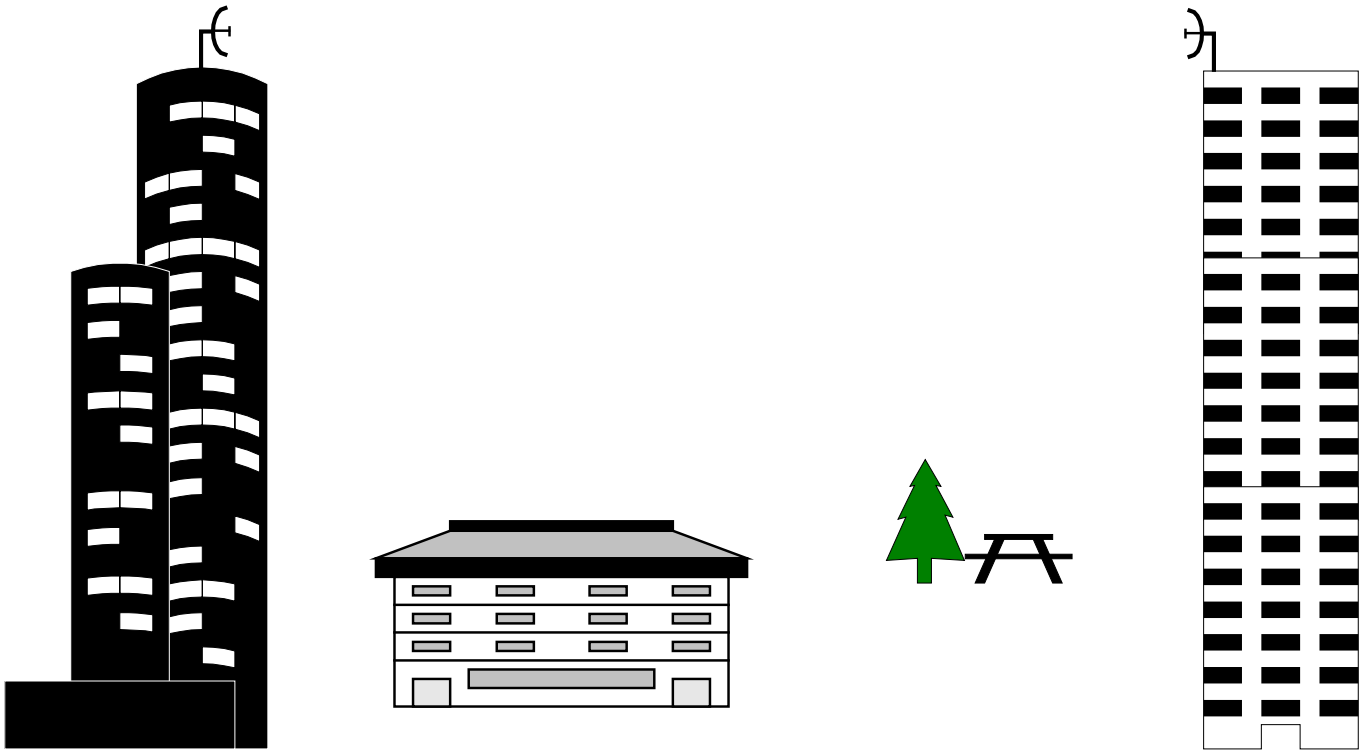
The reflection-range design is complicated and depends on the reflection coefficient of the ground (the range surface), its smoothness, as well as the pattern of the source antenna. The parameter to be determined is the height h_r of the mast, on which the AUT is to be mounted, provided that the height of the transmitting antenna h_t is known. More information can be found in

L.H. Hemming and R.A. Heaton, "Antenna gain calibration on a ground reflection range," *IEEE Trans. on Antennas and Propagation*, vol. AP-21, pp. 532-537, July 1977.

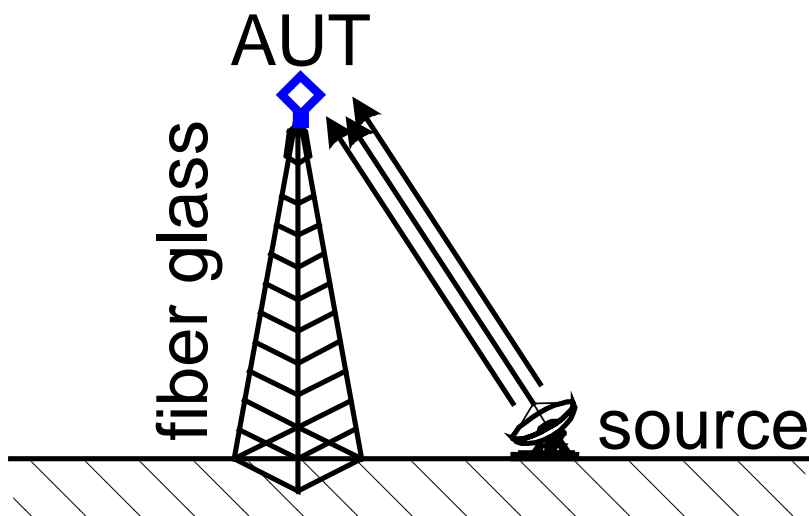
The free-space ranges provide reflection-free propagation. They can be outdoor or indoor. *Outdoor free-space ranges* are carefully built in such a way that reflections from buildings and other objects are minimized. They can be realized as *elevated ranges* and *slant ranges*. *Indoor ranges* (anechoic chambers) suppress reflections (echoes) by lining the walls, the floor and the ceiling with special RF/microwave absorbers.

The elevated ranges are characterized by the following features:

- Both antennas (the transmitting and the receiving) are mounted on high towers or buildings.
- The terrain beneath is smooth.
- The source antenna has very low side lobes so that practically there is no energy directed toward the surface below (the ground) or the buildings behind.
- The line-of-sight is always clear.

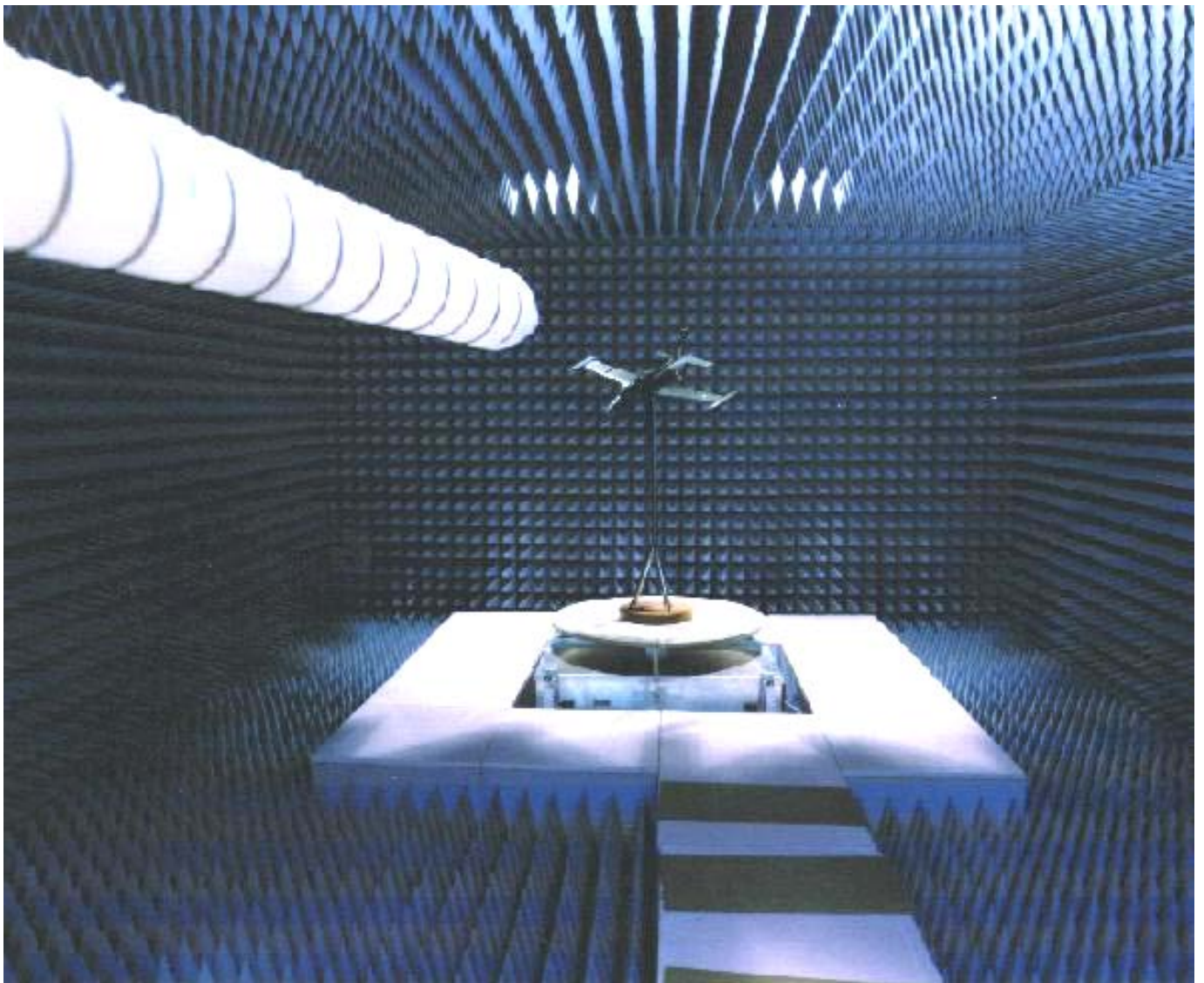


The slant ranges need less space than the elevated ranges. The test antenna is mounted at a fixed height on a non-conducting tower (e.g. made of fiber glass), while the source antenna is mounted near the ground. The source antenna must have its pattern null pointed toward ground. It is desirable that it has very low side lobes and narrow beamwidth. Slant ranges still require wide open space to minimize reflections from surrounding buildings.



The anechoic chambers are the most popular antenna measurement sites especially in the microwave frequency range. They provide convenience and controlled EM environment. However, they are expensive to build and maintain. An anechoic chamber is a large room, the walls, floor and ceiling of which are lined with steel sheets. In effect, an anechoic chamber is a huge Faraday cage, which provides near ideal protection against external EM noise and interference. In addition, all inner surfaces of the chamber are lined with RF/microwave absorbers. An anechoic chamber is shown in the photo below. A comprehensive description of the EM anechoic chambers can be found in

L.H. Hemming, *Electromagnetic Anechoic Chambers: A Fundamental Design and Specifications Guide*, IEEE Press/Wiley, 2002.

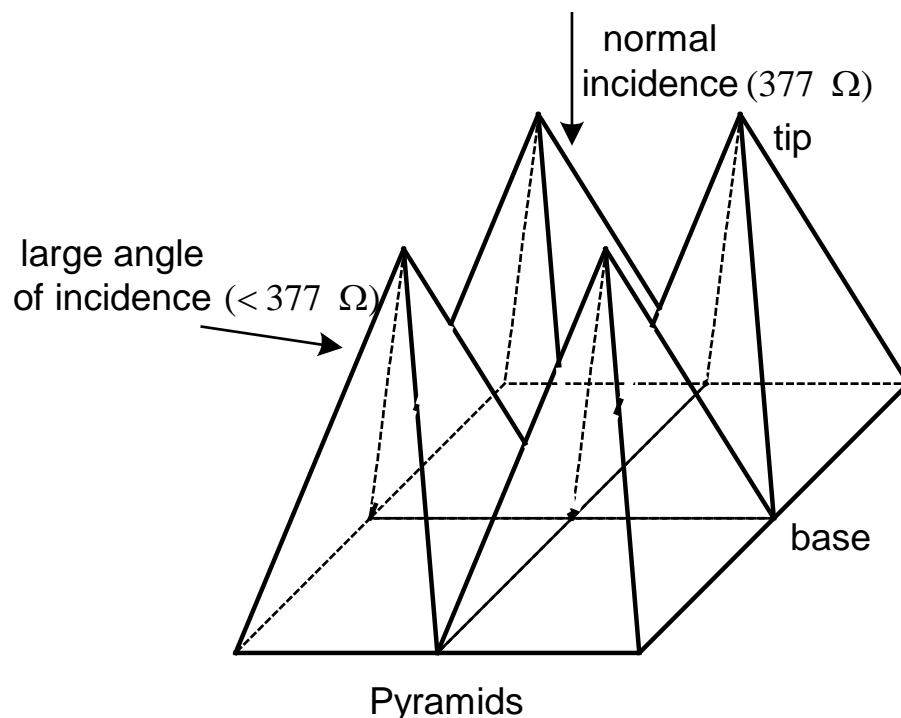


The first EM wave absorbers were developed during World War II in both US and German laboratories. The manufacturing of anechoic chambers became possible after RF/microwave absorbing materials with improved characteristics had become commercially available. The first broadband absorbers were made of a material called *hairflex* consisting of animal fibres sprayed with (or dipped in) conducting carbon in neoprene. A historical summary of the development of EM wave absorbing materials is given by Emerson in his paper:

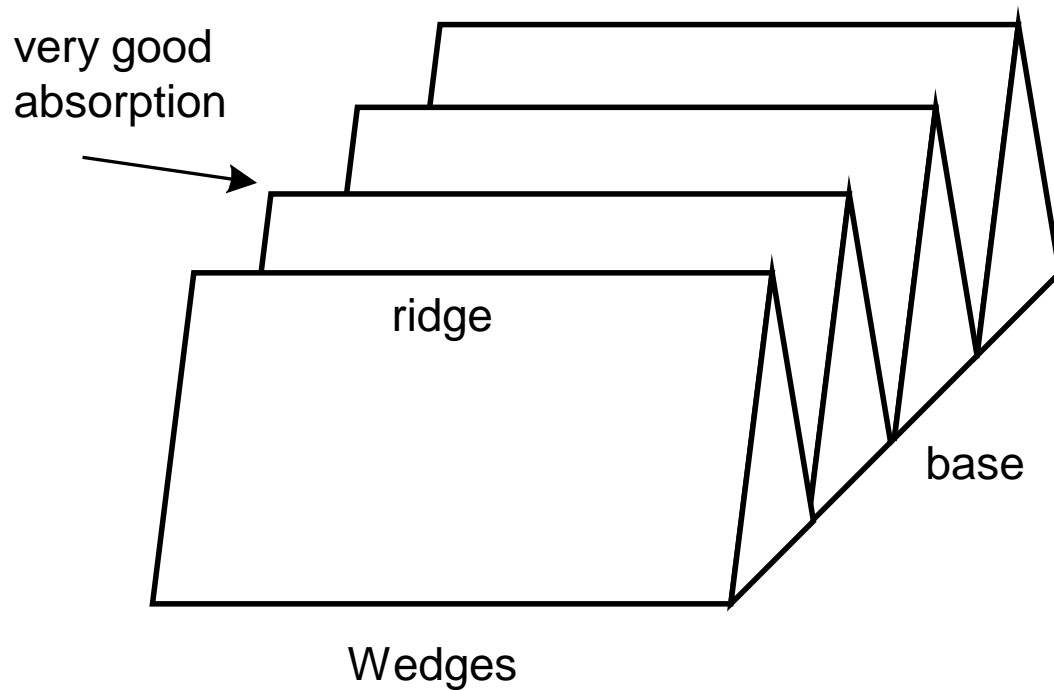
W.H. Emerson, “Electromagnetic wave absorbers and anechoic chambers through the years,” *IEEE Trans. on Antennas and Propagation*, vol. AP-21, pp. 484-489, July 1973.

Nowadays, absorbing elements are with much improved characteristics providing reflection coefficients as low as -50 dB at normal incidence. Reflection increases as the angle of incidence increases. For example, a typical reflection of -25 dB is related to an angle of incidence of about 70 degrees.

A typical absorbing element has the form of a pyramid or a wedge. Pyramids are designed to absorb the waves at normal (nose-on) incidence best. They do not perform well at large angles of incidence. They act, in effect, as a tapered impedance transition for normal incidence of the EM wave from the intrinsic impedance of 377Ω to the short of the chamber’s wall. Their resistance gradually decreases as the pyramid’s cross-section increases.



Wedges, on the other hand, perform much better than pyramids for waves, which travel nearly parallel to their ridges.

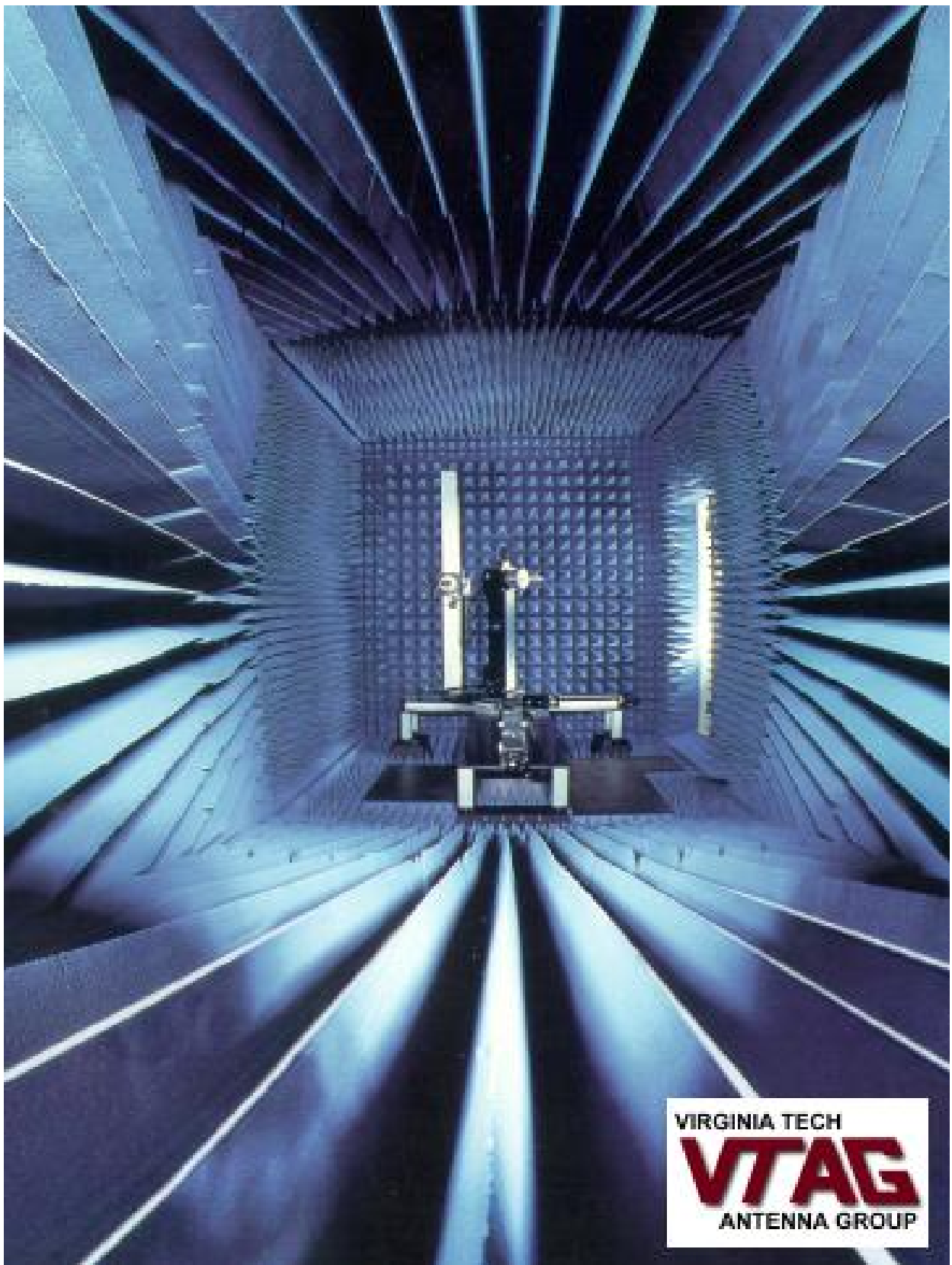


For more detailed information on absorbing materials and shapes see:

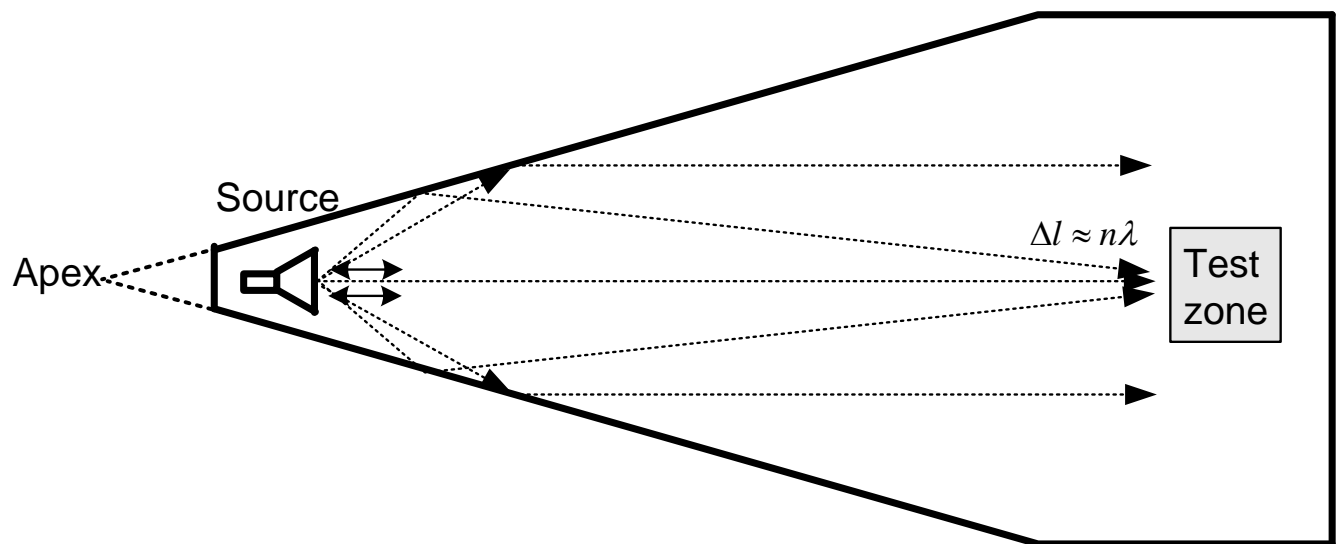
John Kraus, *Antennas*, 2nd edition, McGraw-Hill, Inc.

B.T. DeWitt and W.D. Burnside, "Electromagnetic scattering by pyramidal and wedge absorber," *IEEE Trans. on Antennas and Propagation*, 1988.

An anechoic chamber lined with both types of absorbing shapes is shown below.

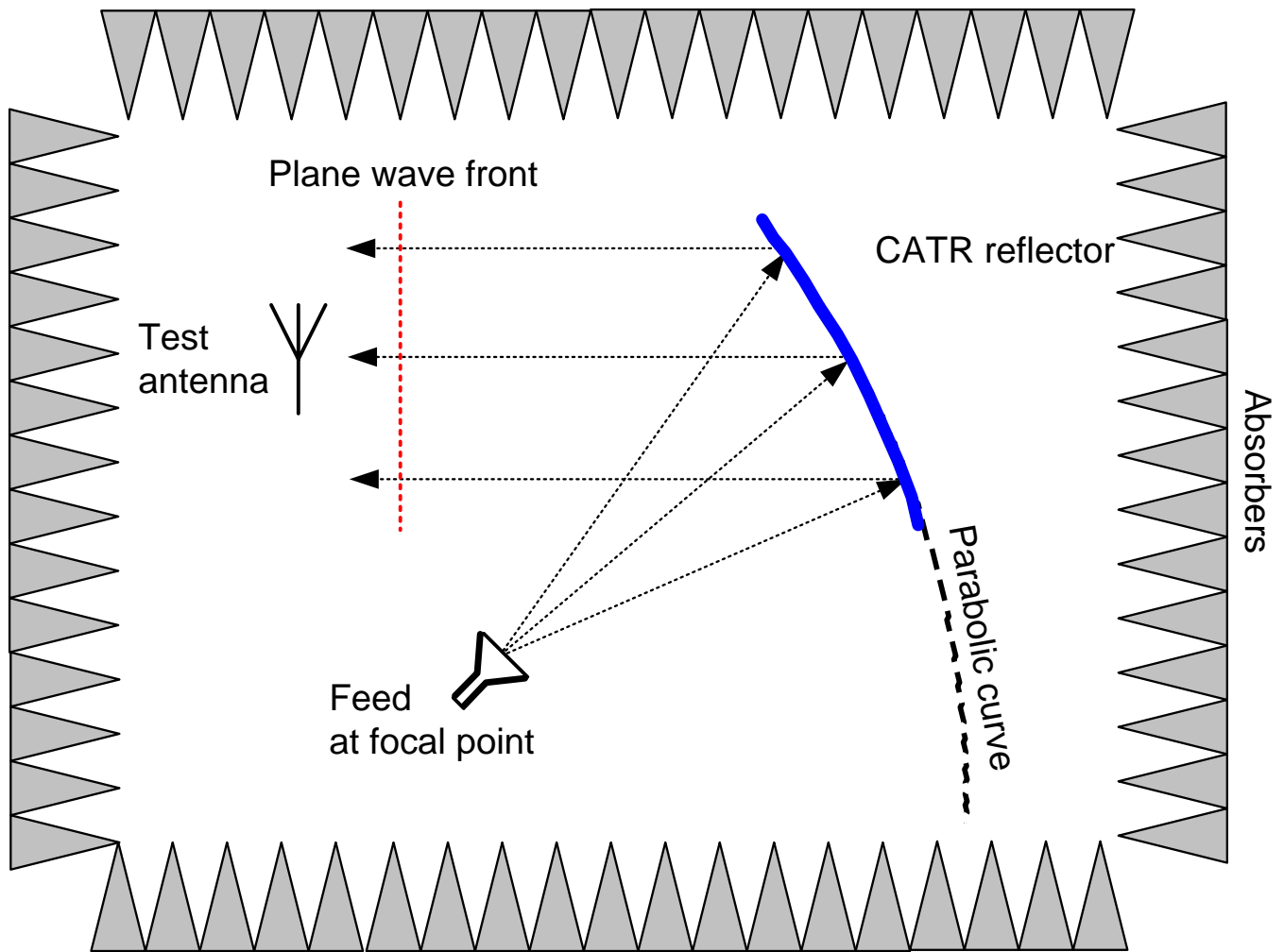


There are two types of anechoic chambers: rectangular and tapered. The design of both chamber types is based on the principles of geometrical optics. The goal is to minimize the amplitude and phase ripples in the test zone (the quiet zone), which are due to the imperfect absorption by the wall lining. The tapered chamber has the advantage of tuning by moving the source antenna closer to (at higher frequencies) or further from (at lower frequencies) the apex of the taper. Thus, the reflected rays are adjusted to produce nearly constructive interference with the direct rays at the test location.



Anechoic chambers are limited by the distance requirements of the far-field measurements of large antennas or scatterers. There are two basic approaches to overcome this limitation. One is presented by the *Compact Antenna Test Ranges (CATRs)*, which produce a nearly uniform plane wave in a very short distance via a system of reflectors (or a single paraboloidal reflector). Another approach is based on near-to-far field transformation, where the measurements are performed in the near-field zone or in the Fresnel zone of the AUT.

The CATR utilizes a precision paraboloidal antenna to collimate the energy of a primary feed antenna in a short distance (about 10 to 20 m). Typical arrangement of a compact range is shown below.



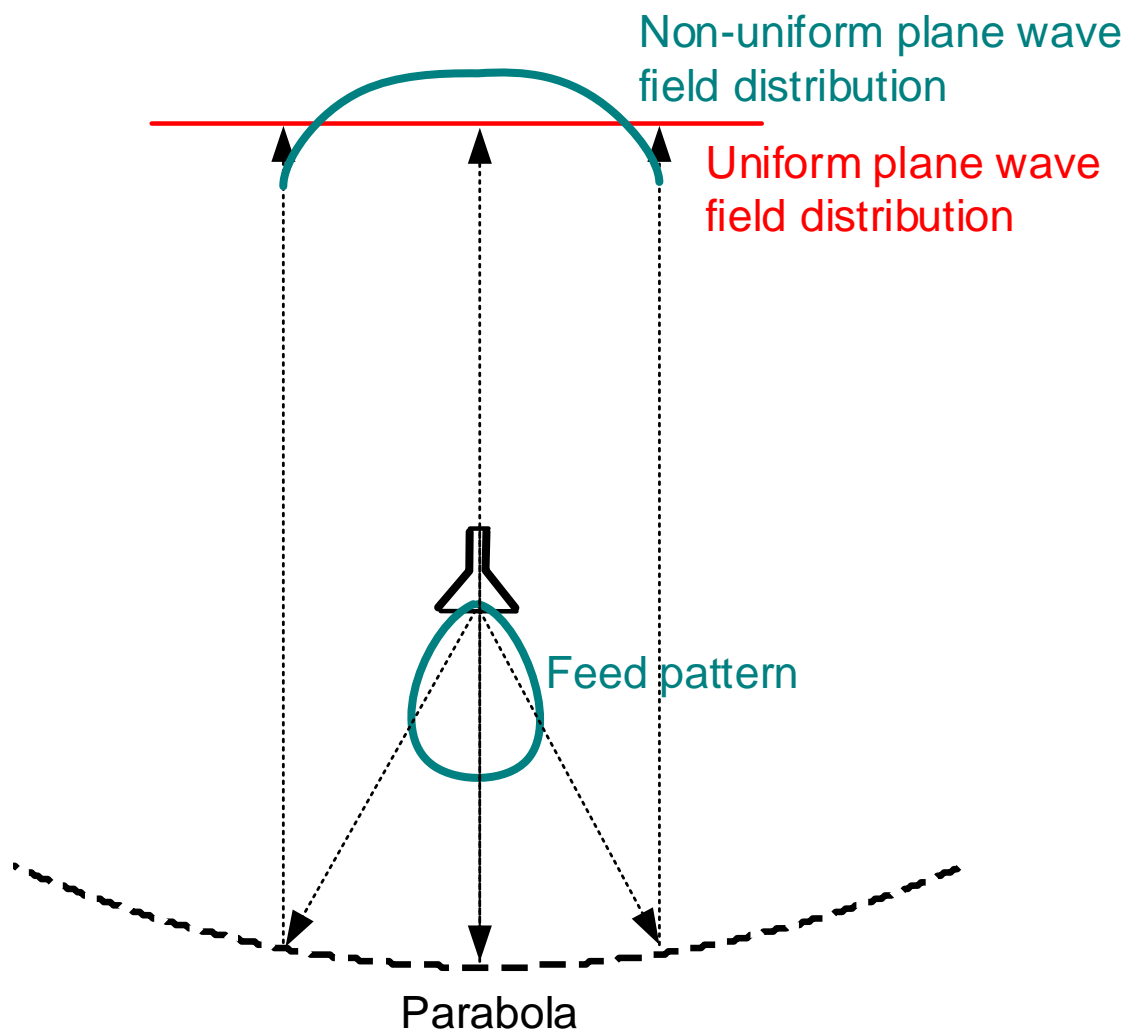
The linear dimensions of the reflector must be at least three to four times those of the AUT so that its illumination is sufficiently close to a uniform plane wave. An offset feed is used for the reflector to prevent aperture blockage and to reduce the diffraction from the primary feed structure. The paraboloidal reflector surface must be fabricated with high precision to obtain fairly uniform amplitude distribution of the incident field at the test antenna.

A perfect plane wave is produced by the CATR if the paraboloidal reflector has a perfect surface, infinite size, and if the feed is a point source with a pattern which compensates for the space attenuation. Of course, such ideal conditions cannot be achieved, and the field distribution in a real CATR deviates from the uniform plane wave. However, it is within acceptable parameters in the quiet zone.

The quiet zone is typically 50% to 60% the aperture of the reflector. The imperfections of the field in the quiet zone are measured in terms of *phase*

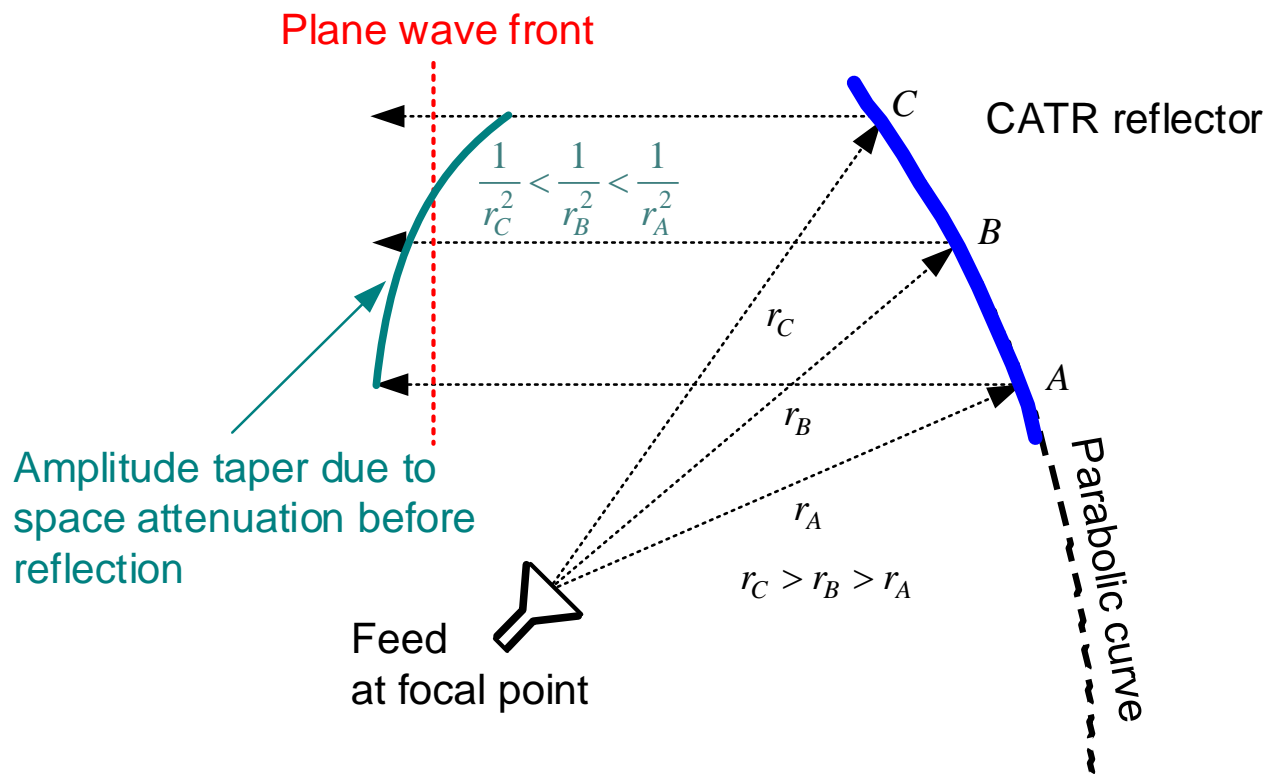
errors, ripple-amplitude deviations, and taper-amplitude deviations. Acceptable deviations for most CATRs are: less than 10% phase error, less than 1 dB ripple and taper amplitude deviations.

Amplitude taper in the quiet zone is due to two reasons: the primary feed pattern and the space attenuation. The primary feed cannot be isotropic; therefore, its pattern has variations with direction. Usually, the pattern gradually decreases as the directional angles point away from the antenna axis. This is called feed-amplitude taper. That portion of the feed pattern, which illuminates the CATR surface, is directly transferred into the quiet zone, thus contributing to the field amplitude-taper deviation from the ideal uniform plane wave.

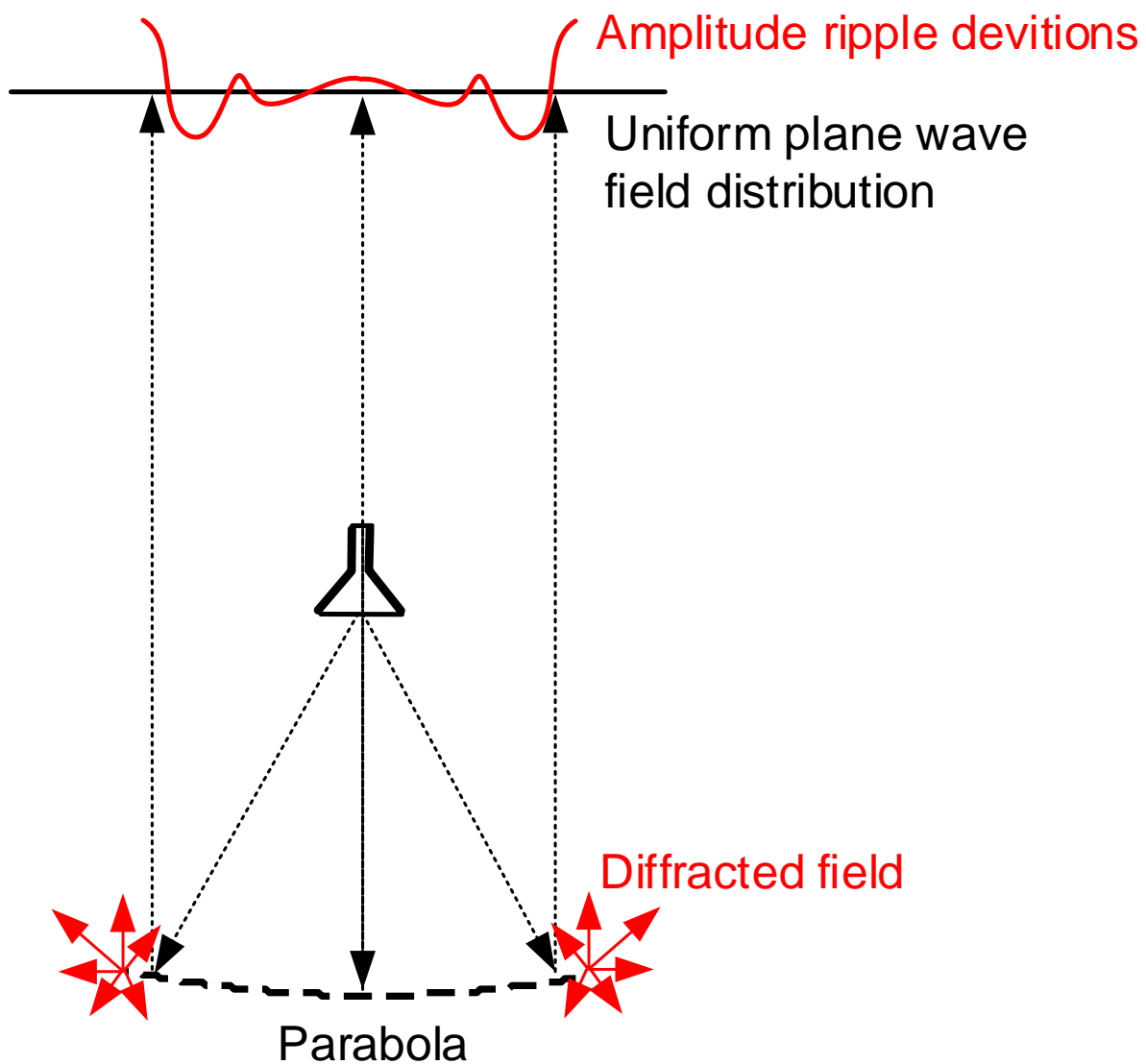


It is obvious that if the feed pattern is nearly isotropic for the angles illuminating the reflector, the feed-amplitude taper will be very small. That is why low-directivity antennas are preferred as feeds. However, the feed cannot be omnidirectional because direct illumination of the AUT by the primary feed is unacceptable. The careful choice of the feed antenna and its location is of paramount importance for the CATR design.

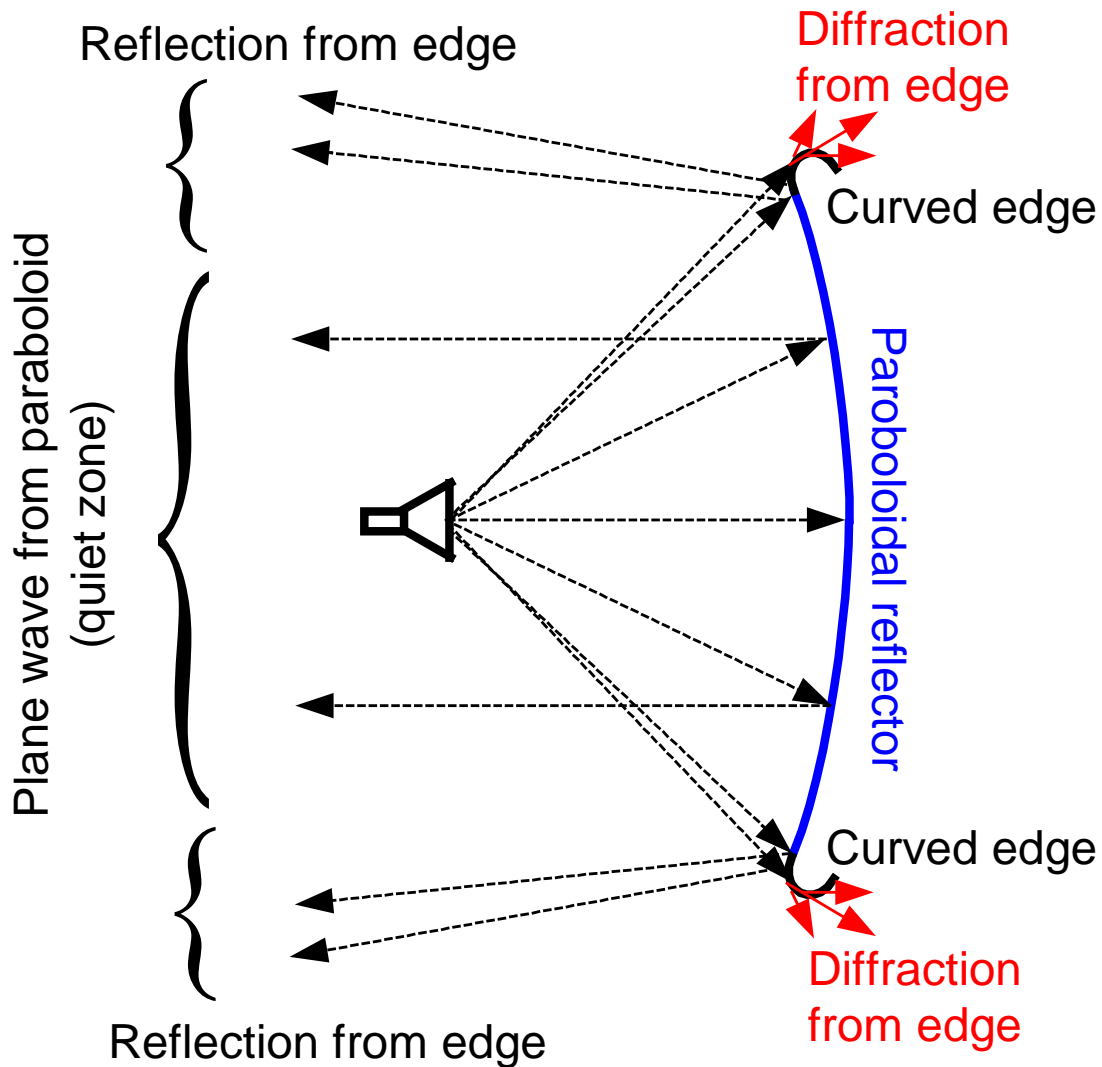
The $1/r^2$ power space attenuation occurs with the spherical spreading of the uncollimated energy radiated by the primary feed toward the reflector. The paths of these primary EM rays from the feed to the reflector are of different lengths, which results in different amplitude across the front of the reflected collimated EM wave. This is yet another reason for amplitude taper deviations in the quiet zone.



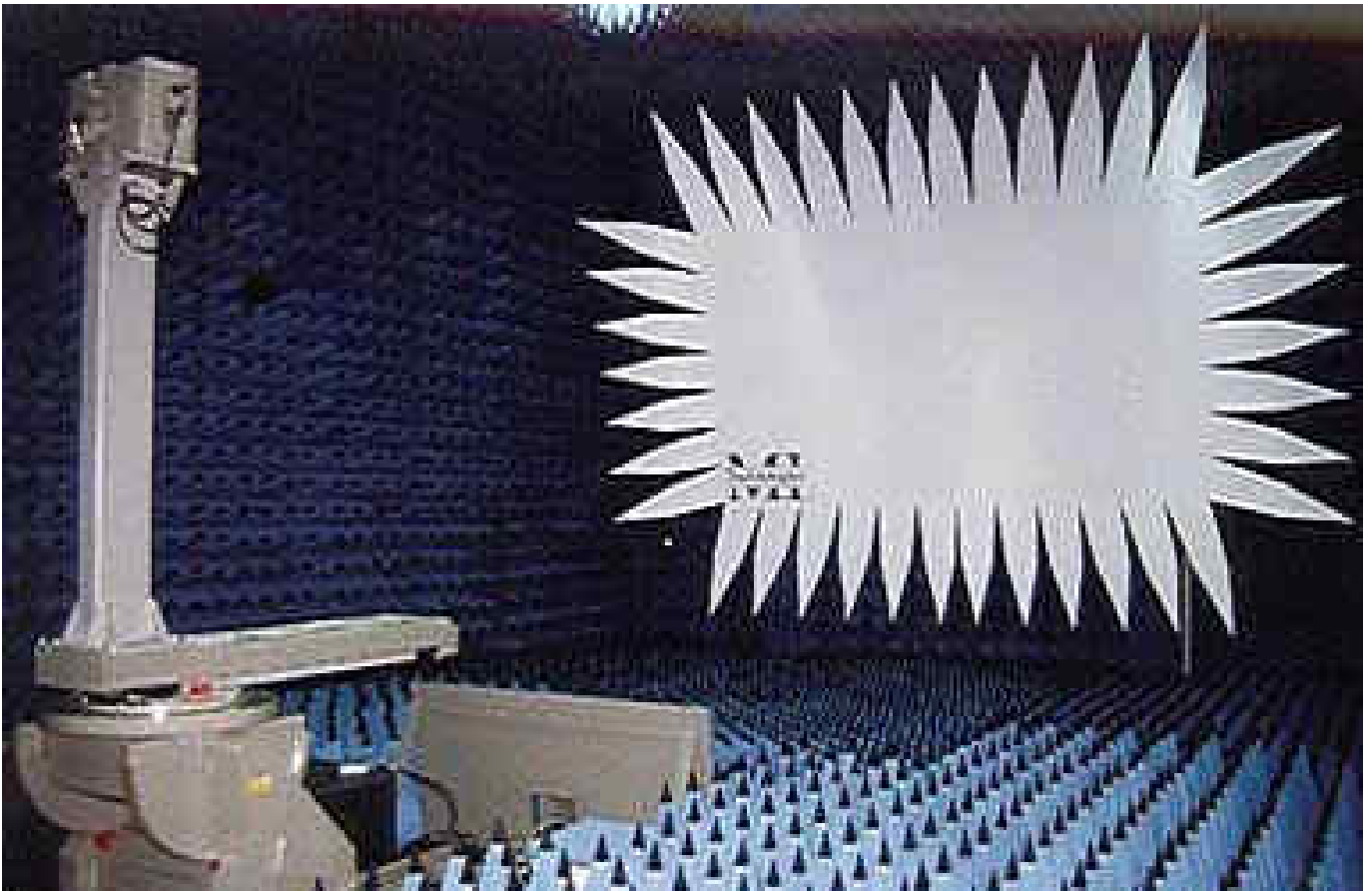
Amplitude and phase ripples in the quiet zone are primarily caused by diffraction from the edges of the reflector. The diffracted field is spread in all directions interfering with the major reflected field in constructive and destructive patterns. The result is the appearance of maxima and minima of the field amplitude across the plane wave front in the quiet zone. Diffraction from edges causes deviation of the phase of the plane wave, too.



There are two common ways to reduce diffraction from reflector edges: serrated-edge reflectors and rolled-edge reflectors. Rolled-edge modifications at the edge of the reflector are introduced to direct the diffracted field mainly to the side and the back of the reflector.



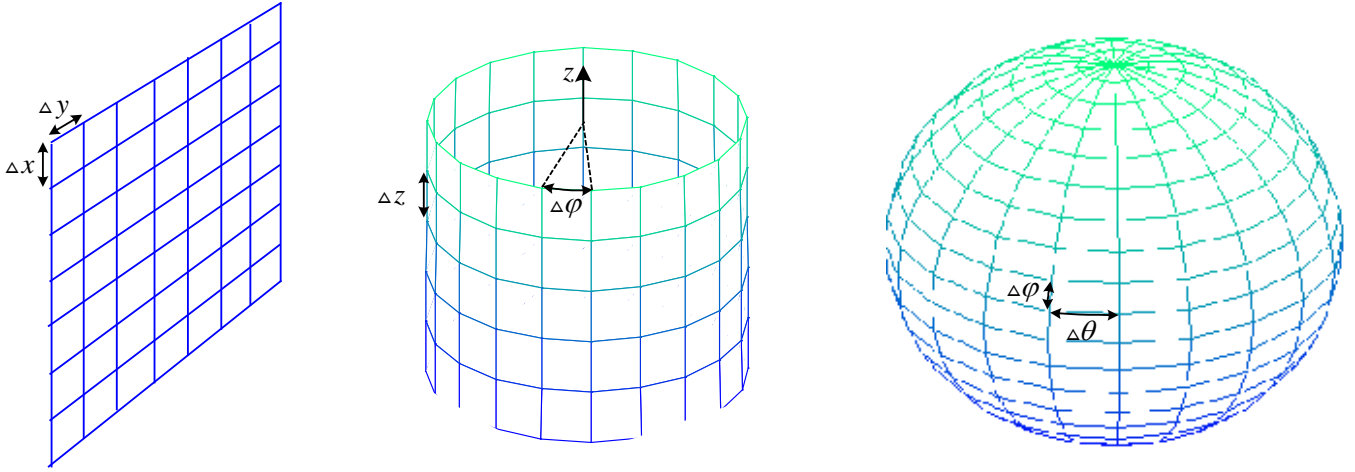
Serrated edges of reflectors produce multiple low-amplitude diffractions, which are randomized in amplitude, phase and polarization. That is why the probability of their cancellation in any point of the quiet zone is high. Serrations are typically of irregular triangular shape. To further reduce the diffraction in the direction of the test zone, the serrated edges may be also rolled backwards. A photograph of a compact range whose reflector has serrated edges is shown below.



4. Near-to-far Field Transformations for Compact Ranges

Another approach for measuring far-field patterns, which allows for the most compact chambers, is the near-field/far-field (NF/FF) method. The field amplitude, phase and polarization are measured in the near field of the AUT, which is in radiating mode. The near-field data is transformed to far-field patterns via analytical techniques implemented in the sophisticated software run by an automated computer system, which controls the measurement procedure.

The magnitude and phase of the tangential \mathbf{E} field are measured at regular intervals over a canonical surface (plane, cylinder, or sphere) located close to the AUT. The sampled \mathbf{E} field is used to calculate the angular spectrum of the plane, the cylindrical or the spherical wave. This spectrum matches closely the radiated field angular distribution. This is called *modal expansion* of the radiated field.



Here, we consider the simplest data acquisition over a planar surface and its modal expansion. We show that *the far-field radiation pattern of any aperture (surface) is the Fourier transform of the aperture field distribution*. We next derive the formulas in the case of a planar acquisition aperture.

Assume that in the near-field measurements, the \mathbf{E} vector is measured over a planar surface, which is our aperture. According to the *equivalence principle*, we can now assume that the field behind the surface (on the side of the antenna) is equal to zero, and its impact on the field on the other side of the surface is due to equivalent surface currents:

$$\begin{aligned}\mathbf{J}_s &= \hat{\mathbf{n}} \times \mathbf{H}_a \\ \mathbf{M}_s &= -\mathbf{n} \times \mathbf{E}_a.\end{aligned}\tag{9}$$

Here, \mathbf{E}_a and \mathbf{H}_a represent the field vectors at the aperture (the surface) due to the antenna behind it. \mathbf{J}_s is the equivalent electric surface current density, \mathbf{M}_s is the equivalent magnetic surface current density, and $\hat{\mathbf{n}}$ is the surface unit normal pointing toward the region of observation (away from the antenna).

Since the field behind the planar surface is now set to zero, we can as well assume that the medium behind the surface is a perfect conductor. In the case of a flat surface of size much larger than a wavelength, the image theory can be applied. Now the equivalent surface sources become

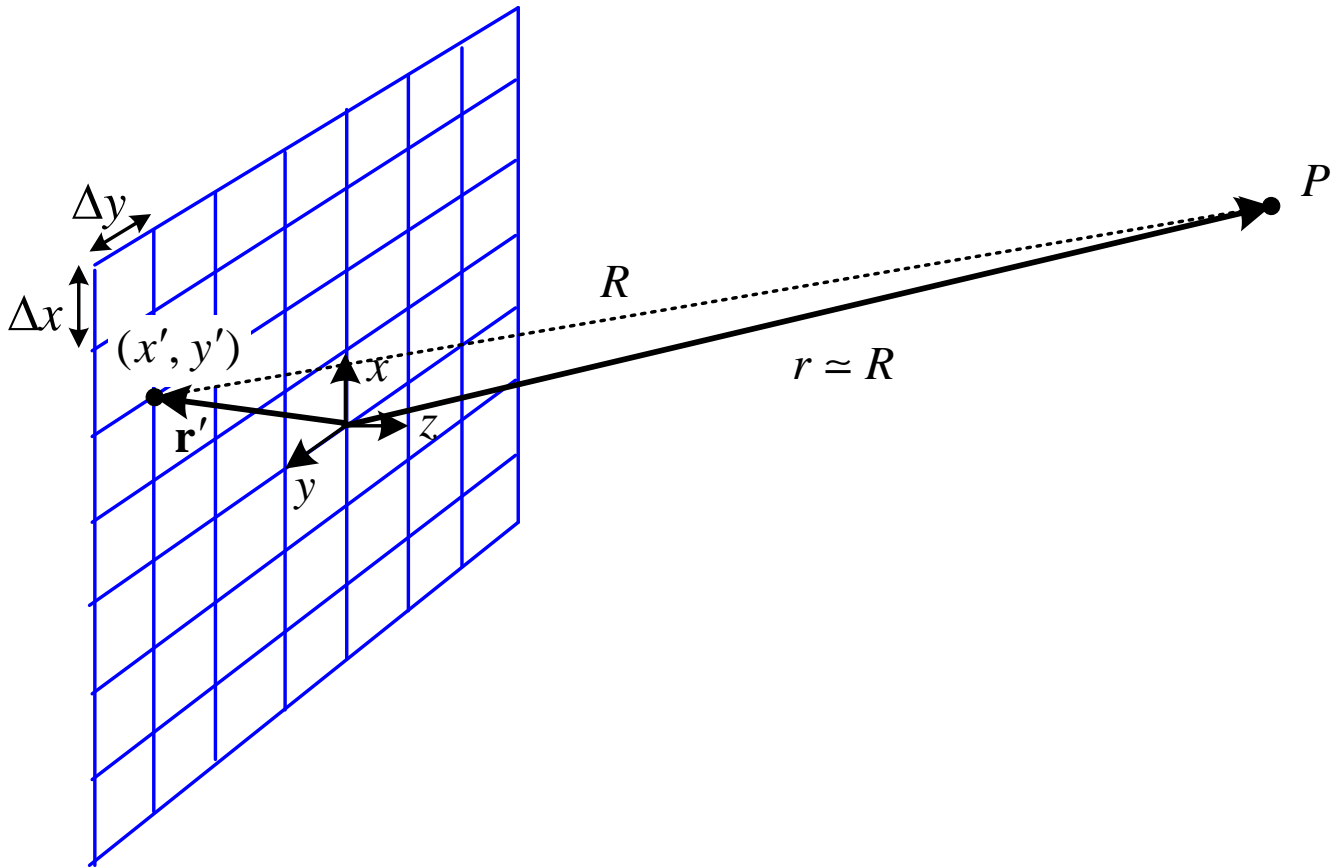
$$\mathbf{J}_s = 0; \quad \mathbf{M}_s = -\hat{\mathbf{n}} \times (2\mathbf{E}_a).\tag{10}$$

The equivalent surface magnetic currents \mathbf{M}_s create an electric vector

potential \mathbf{F} , which, in the far zone, is

$$\mathbf{F}(P) \approx -\varepsilon \frac{e^{-j\beta r}}{4\pi r} \hat{\mathbf{n}} \times \iint_{S_a} 2\mathbf{E}_a(\mathbf{r}') e^{j\mathbf{k} \cdot \mathbf{r}'} ds' = -\varepsilon \frac{e^{-j\beta r}}{2\pi r} \hat{\mathbf{n}} \times \iint_{S_a} \mathbf{E}_a(\mathbf{r}') e^{j\mathbf{k} \cdot \mathbf{r}'} ds' \quad (11)$$

where $\beta = \omega\sqrt{\mu\varepsilon}$ is the wavenumber, $\mathbf{r}' = \hat{\mathbf{x}}x' + \hat{\mathbf{y}}y'$ is the position vector of the integration point, and r is the distance from the observation point P to the origin.



Note that the far-field approximations have been applied to the amplitude and phase terms of the vector-potential integral. The propagation vector $\mathbf{k} = \beta\hat{\mathbf{r}}$ shows the direction of propagation and has a magnitude equal to the wave number β . The scalar product $\mathbf{k} \cdot \mathbf{r}' = \beta r' \hat{\mathbf{r}} \cdot \hat{\mathbf{r}}'$ yields the familiar phase term accounting for the phase delay associated with the source point location.

We now remember that the far-field \mathbf{E} vector is related to the far-field vector potential \mathbf{F} as

$$\mathbf{E}^{\text{far}} = -j\omega\eta\mathbf{F} \times \hat{\mathbf{r}}. \quad (12)$$

Here, $\eta = \sqrt{\mu/\varepsilon}$ is the intrinsic impedance of the medium. Substituting (11) in (12) yields:

$$\mathbf{E}^{\text{far}} \approx -j\beta \frac{e^{-j\beta r}}{2\pi r} \hat{\mathbf{r}} \times \iint_{S_a} (\hat{\mathbf{n}} \times \mathbf{E}_a) e^{j\mathbf{k} \cdot \mathbf{r}'} dx' dy'. \quad (13)$$

In the case of a planar surface, the unit normal is constant, and we can assume that $\hat{\mathbf{n}} = \hat{\mathbf{z}}$. Having in mind that the rectangular components of the radial unit vector are

$$\hat{\mathbf{r}} = \hat{\mathbf{x}} \sin \theta \cos \varphi + \hat{\mathbf{y}} \sin \theta \sin \varphi + \hat{\mathbf{z}} \cos \theta, \quad (14)$$

we can calculate the x and y components of \mathbf{E}^{far} as

$$E_x^{\text{far}} \approx j\beta \frac{e^{-j\beta r}}{2\pi r} \cdot \cos \theta \cdot \iint_{S_a} E_{xa}(x', y') e^{j(k_x x' + k_y y')} dx' dy', \quad (15)$$

$$E_y^{\text{far}} \approx j\beta \frac{e^{-j\beta r}}{2\pi r} \cdot \cos \theta \cdot \iint_{S_a} E_{ya}(x', y') e^{j(k_x x' + k_y y')} dx' dy'. \quad (16)$$

Here, k_x and k_y are the spectral variables, which are the components of the propagation vector \mathbf{k} in the xy plane:

$$\begin{aligned} k_x &= \beta \sin \theta \cos \varphi, \\ k_y &= \beta \sin \theta \sin \varphi. \end{aligned} \quad (17)$$

The z -component of the far \mathbf{E} field is found as

$$E_z^{\text{far}} \approx -j\beta \frac{e^{-j\beta r}}{2\pi r} \cdot \sin \theta \cdot \left[\cos \varphi \iint_{S_a} E_{xa}(x', y') e^{j(k_x x' + k_y y')} dx' dy' + \sin \varphi \iint_{S_a} E_{ya}(x', y') e^{j(k_x x' + k_y y')} dx' dy' \right]. \quad (18)$$

It is obvious from (15), (16) and (18) that if the components E_x^{far} and E_y^{far} are known, the E_z^{far} component can be calculated (if need be) as

$$E_z^{\text{far}} = -\tan \theta \cdot [E_x^{\text{far}} \cos \varphi + E_y^{\text{far}} \sin \varphi]. \quad (19)$$

Let us examine the integrals appearing in (15) and (16):

$$f_x(k_x, k_y) = \iint_{S_a} E_{xa}(x', y') e^{j(k_x x' + k_y y')} dx' dy', \quad (20)$$

$$f_y(k_x, k_y) = \iint_{S_a} E_{ya}(x', y') e^{j(k_x x' + k_y y')} dx' dy'. \quad (21)$$

These integrals are the 2-D Fourier transforms of the tangential field distribution, $E_{xa}(-x', -y')$ and $E_{ya}(-x', -y')$, over the area of the surface S_a . The surface is ideally infinite ($-\infty < x' < +\infty$, $-\infty < y' < +\infty$). In practice, the surface where the field is measured is finite and designed so that the field components outside of it are negligible. The functions f_x and f_y depend on the spectral variables k_x and k_y .

Note that the functions $f_x(k_x, k_y)$ and $f_y(k_x, k_y)$ give the far-field pattern in terms of the field x and y components for small θ when $\cos \theta \approx 1$:

$$\left. \begin{aligned} \bar{E}_x^{\text{far}}(\theta, \varphi) &\approx f_x(k_x, k_y) \\ \bar{E}_y^{\text{far}}(\theta, \varphi) &\approx f_y(k_x, k_y) \end{aligned} \right\} \text{for } \cos \theta \approx 1, \text{ where } \begin{cases} k_x = \beta \sin \theta \cos \varphi, \\ k_y = \beta \sin \theta \sin \varphi. \end{cases} \quad (22)$$

This finally clarifies the statement that *the far-field pattern is the Fourier transform of the aperture field distribution*.

The far-field z -component can be expressed by its spectral counterpart $f_z(k_x, k_y)$ in the same manner as the x and y components:

$$E_z^{\text{far}} = jk \frac{e^{-jkr}}{r} \cdot \cos \theta \cdot f_z(k_x, k_y). \quad (23)$$

Having in mind (18) and (19), it becomes clear that $f_z(k_x, k_y)$ is not an independent function but is related to the other two spectral components as

$$f_z(k_x, k_y) = -\tan \theta \cdot [f_x(k_x, k_y) \cos \varphi + f_y(k_x, k_y) \sin \varphi]. \quad (24)$$

We can now define the vector plane-wave spectral function:

$$\mathbf{f}(k_x, k_y) = \hat{\mathbf{x}}f_x(k_x, k_y) + \hat{\mathbf{y}}f_y(k_x, k_y) + \hat{\mathbf{z}}f_z(k_x, k_y) \quad (25)$$

the spatial components of which are calculated via (20), (21) and (24). The far-field \mathbf{E} vector can be calculated from the spectral function as

$$\mathbf{E}(r, \theta, \varphi) \approx j\beta \frac{e^{-j\beta r}}{2\pi r} \cos \theta \cdot \mathbf{f}(k_x, k_y). \quad (26)$$

We can express the vector equation (26) in terms of the θ and φ components of the far-field \mathbf{E} vector:

$$\begin{aligned} E_{\theta}(r, \theta, \varphi) &\approx j\beta \frac{e^{-j\beta r}}{2\pi r} \cos \theta \cdot f_{\theta}(k_x, k_y), \\ E_{\varphi}(r, \theta, \varphi) &\approx j\beta \frac{e^{-j\beta r}}{2\pi r} \cos \theta \cdot f_{\varphi}(k_x, k_y). \end{aligned} \quad (27)$$

Since the spectral function \mathbf{f} is derived via its rectangular components during the data acquisition over a planar surface, it is desirable to convert f_{θ} and f_{φ} to f_x and f_y . Following the standard transformation from spherical to rectangular components, we obtain

$$\cos \theta \cdot f_{\theta} = \cos \theta (f_x \cos \theta \cos \varphi + f_y \cos \theta \sin \varphi - f_z \sin \theta). \quad (28)$$

After substituting f_z from (24), we arrive at

$$\cos \theta \cdot f_{\theta} = f_x \cos \varphi + f_y \sin \varphi. \quad (29)$$

In analogous manner, it can be shown that

$$\cos \theta \cdot f_{\varphi} = -f_x \sin \varphi + f_y \cos \varphi. \quad (30)$$

The substitution of (29) and (30) into (27) finally gives

$$\begin{aligned} E_{\theta}(r, \theta, \varphi) &\approx jk \frac{e^{-jkr}}{2\pi r} (f_x \cos \varphi + f_y \sin \varphi) \\ E_{\varphi}(r, \theta, \varphi) &\approx jk \frac{e^{-jkr}}{2\pi r} (-f_x \sin \varphi + f_y \cos \varphi). \end{aligned} \quad (31)$$

We can now summarize the procedure of the NF/FF pattern measurement in three basic steps:

- Measure the tangential \mathbf{E} field components $E_{xa}(x', y', z' = 0)$ and $E_{ya}(x', y', z' = 0)$ over the near-field aperture (data acquisition).
- Calculate the plane-wave spectral functions $f_x(k_x, k_y)$ and $f_y(k_x, k_y)$ using (20) and (21).
- Calculate the normalized far-field components using

$$\begin{aligned} \bar{E}_{\theta}(\theta, \varphi) &= f_x \cos \varphi + f_y \sin \varphi, \\ \bar{E}_{\varphi}(\theta, \varphi) &= -f_x \sin \varphi + f_y \cos \varphi, \end{aligned} \quad (32)$$

and the total normalized field pattern using

$$\bar{E}(\theta, \varphi) = \sqrt{\bar{E}_\theta^2(\theta, \varphi) + \bar{E}_\varphi^2(\theta, \varphi)} = \sqrt{f_x^2(k_x, k_y) + f_y^2(k_x, k_y)}. \quad (33)$$

In the actual test procedure, a planar surface is chosen a distance z_0 away from the test antenna, which is in radiating mode. This surface is called the *measurement aperture*. The distance z_0 is at least three wavelengths away from the antenna, so that the measurement is carried out in the radiating near-field region (Fresnel zone) rather than in the reactive near-field region where the amplitude and phase variations of the field are too rapid and the sampling intervals must be very small.

The measurement aperture is rectangular of dimensions $a \times b$. It is divided into $M \times N$ points spaced evenly Δx and Δy apart. The relation between the number of points and the respective spacing is then

$$M = \frac{a}{\Delta x} + 1, \quad N = \frac{b}{\Delta y} + 1. \quad (34)$$

Thus, the sampling points are located at coordinates $(m\Delta x, n\Delta y, 0)$ where $0 \leq m \leq M - 1$ and $0 \leq n \leq N - 1$. The separation distances Δx and Δy must be less than half a wavelength in order to satisfy Nyquist's sampling criterion and such that the equations in (34) yield integer numbers. The measurement aperture must be large enough so that the signal at its edges is at least 45 dB down from the maximum measured signal on the acquisition surface.

The plane-wave spectral function $\mathbf{f}(k_x, k_y)$ can be evaluated at a discrete set of wave numbers as dictated by the discrete Fourier transform:

$$k_x = m \frac{2\pi}{a} = m \frac{2\pi}{\underbrace{(M-1)\Delta x}_{\Delta k_x}} \quad (35)$$

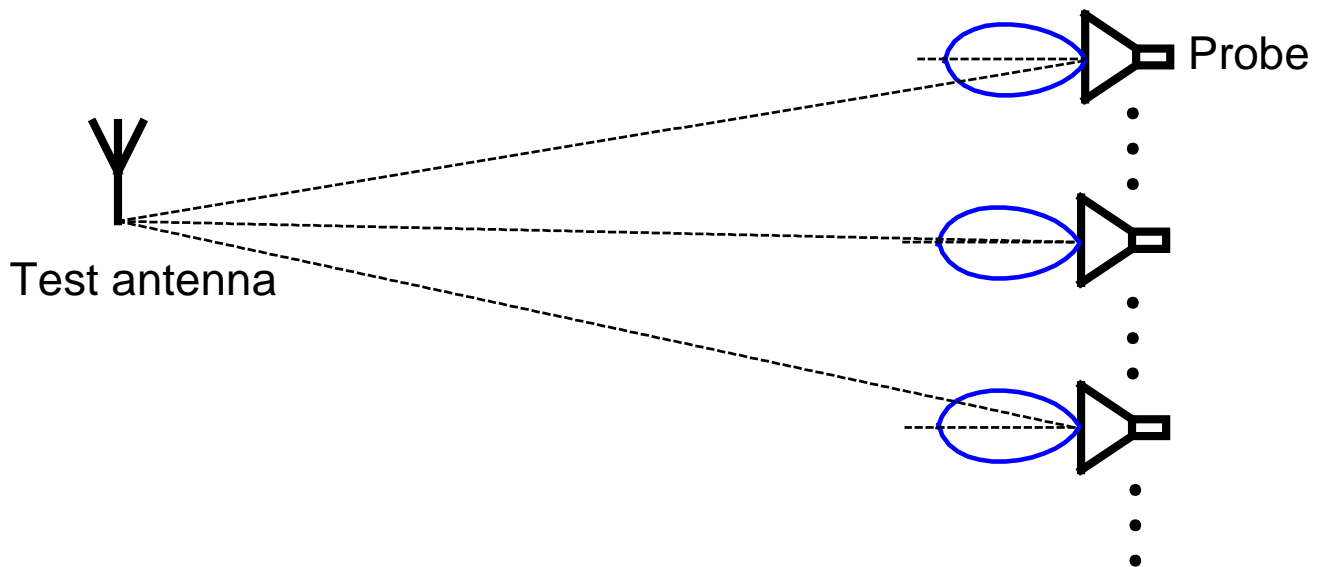
$$k_y = n \frac{2\pi}{b} = n \frac{2\pi}{\underbrace{(N-1)\Delta y}_{\Delta k_y}}.$$

Conventional two-dimensional FFT (Fast Fourier Transform) techniques are used to perform this transformation.

The acquisition of the planar near-field data is done by a computer-controlled probe antenna (typically a waveguide horn or an open waveguide),

which is moved to each grid node over the measurement aperture by a high-precision positioning system (positioner). The probe's axis is held stationary and normal to the measurement aperture. The probe must be linearly polarized so that separate measurements of the two tangential field components E_x and E_y become possible.

As the probe location changes, its pattern orientation with respect to the AUT changes, too, as shown below. The probe's partial directivities in the direction of the test antenna must be taken into account using probe compensation techniques.



The principal advantage of the planar NF/FF transformation over the cylindrical and the spherical one is its mathematical simplicity. Its major disadvantage is that it cannot cover all directional angles. In the ideal case of infinite planar measurement surface, only one hemisphere of the antenna pattern can be measured. Thus, the back lobes and the side lobes of the antenna cannot be measured together with the main beam. Of course, the AUT can be rotated in different positions, so that the overall pattern can be reconstructed.

The reader interested in the subject of NF/FF transforms and measurements is referred to the following introductory sources:

R.C. Johnson, H.A. Ecker, and J.S. Hollis, "Determination of far-field antenna patterns from near-field measurements," *Proc. IEEE*, vol. 61, No. 12, pp. 1668-1694, Dec. 1973.

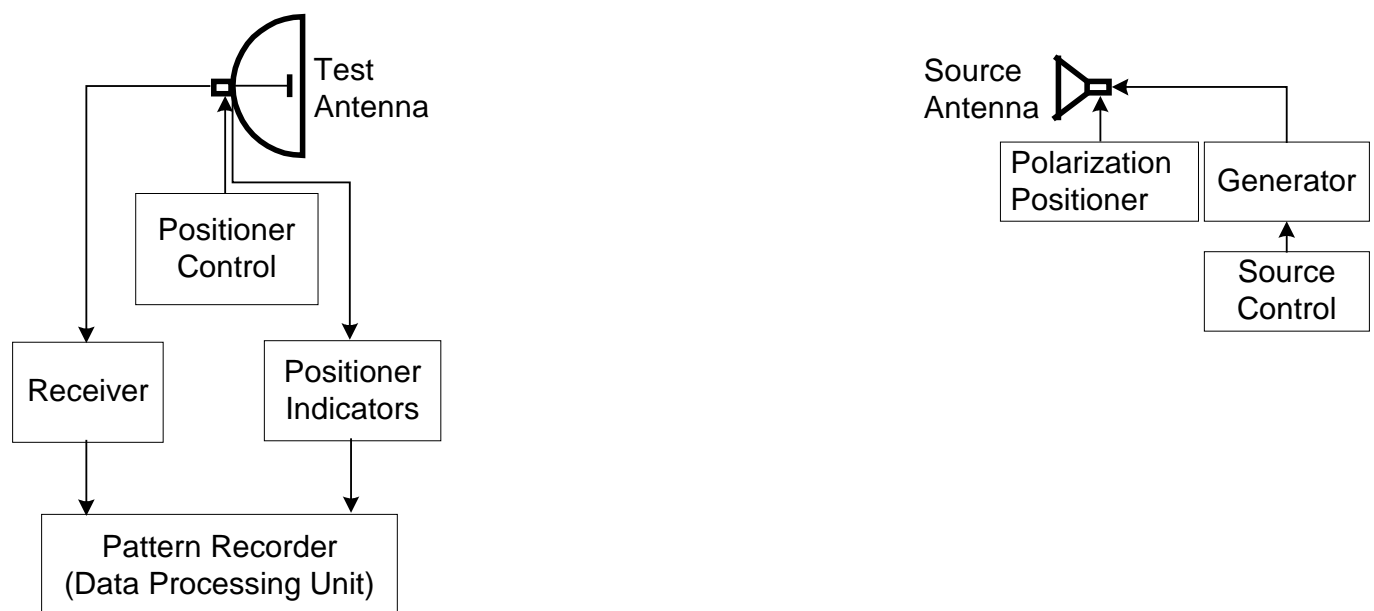
D.T. Paris, W.M. Leach, Jr., and E.B. Joy, “Basic theory of probe compensated near-field measurements,” *IEEE Trans. on Antennas and Propagation*, vol. AP-26, No. 3, pp. 373-379, May 1978.

E.B. Joy, W.M. Leach, Jr., G.P. Rodrigue, and D.T. Paris, “Applications of probe compensated near-field measurements,” *IEEE Trans. on Antennas and Propagation*, vol. AP-26, No. 3, pp. 379-389, May 1978.

A.D. Yaghjian, “An overview of near-field antenna measurements,” *IEEE Trans. on Antennas and Propagation*, vol. AP-34, pp. 30-45, January 1986.

5. Far-field Pattern Measurements*

The far-field patterns are measured on the surface of a sphere of constant radius. Any position on the sphere is identified by the directional angles θ and ϕ of the spherical coordinate system. In general, the pattern of an antenna is 3-D. However, 3-D pattern acquisition is difficult – it involves multiple 2-D pattern measurements. The minimal number of 2-D patterns is two, and these two patterns must be in two orthogonal principal planes. A principal plane must contain the direction of maximum radiation. A simplified block diagram of a pattern measurement system is given below.



The total *amplitude pattern* is described by the vector sum of the two orthogonally polarized radiated field components:

$$|\mathbf{E}| = \sqrt{|E_\theta|^2 + |E_\varphi|^2}. \quad (36)$$

Rarely, the separate patterns for both components are needed. This is the case when the polarization of the test antenna must be characterized in addition to its pattern.

For antennas of low directivity, at least three 2-D pattern cuts are necessary in order to obtain good 3-D pattern approximation: in the two elevation planes at $\varphi = 0^\circ / 180^\circ$ and $\varphi = 90^\circ / 270^\circ$ as well as the azimuth pattern at $\theta = 90^\circ$.

For high-directivity antennas, only two orthogonal 2-D elevation patterns often suffice. Assuming that the antenna boresight is along the z -axis, these are the patterns at $\varphi = 0^\circ / 180^\circ$ and $\varphi = 90^\circ / 270^\circ$. The 3-D pattern approximation from 2-D patterns is discussed below.

High-directivity aperture antennas such as horn and reflector antennas can have their far-field components expressed as

$$E_\theta(\theta, \varphi) = j\beta \frac{e^{-j\beta r}}{4\pi r} \left[\mathcal{J}_x^E \cos \varphi + \mathcal{J}_y^E \sin \varphi + \eta \cos \theta (\mathcal{J}_y^H \cos \varphi - \mathcal{J}_x^H \sin \varphi) \right], \quad (37)$$

$$E_\varphi(\theta, \varphi) = j\beta \frac{e^{-j\beta r}}{4\pi r} \left[-\eta (\mathcal{J}_x^H \cos \varphi + \mathcal{J}_y^H \sin \varphi) + \cos \theta (\mathcal{J}_y^E \cos \varphi - \mathcal{J}_x^E \sin \varphi) \right]. \quad (38)$$

Here, \mathcal{J}_x^E , \mathcal{J}_y^E , \mathcal{J}_x^H and \mathcal{J}_y^H are the plane-wave spectral functions:

$$\mathcal{J}_x^E(\theta, \varphi) = \iint_{S_A} E_{ax}(x', y') e^{j\beta(x' \sin \theta \cos \varphi + y' \sin \theta \sin \varphi)} dx' dy', \quad (39)$$

$$\mathcal{J}_y^E(\theta, \varphi) = \iint_{S_A} E_{ay}(x', y') e^{j\beta(x' \sin \theta \cos \varphi + y' \sin \theta \sin \varphi)} dx' dy', \quad (40)$$

$$\mathcal{J}_x^H(\theta, \varphi) = \iint_{S_A} H_{ax}(x', y') e^{j\beta(x' \sin \theta \cos \varphi + y' \sin \theta \sin \varphi)} dx' dy', \quad (41)$$

$$\mathcal{J}_y^H(\theta, \varphi) = \iint_{S_A} H_{ay}(x', y') e^{j\beta(x' \sin \theta \cos \varphi + y' \sin \theta \sin \varphi)} dx' dy'. \quad (42)$$

From equations (37) and (38) it follows that the field components in the principal planes are

$$E_{\theta}(\theta, 0) = j\beta \frac{e^{-j\beta r}}{4\pi r} \left[\tilde{\mathcal{J}}_x^E(\theta, 0) + \tilde{\mathcal{J}}_y^H(\theta, 0) \cdot \eta \cos \theta \right] \quad (43)$$

$$E_{\theta}(\theta, 90^{\circ}) = j\beta \frac{e^{-j\beta r}}{4\pi r} \left[\tilde{\mathcal{J}}_y^E(\theta, 90^{\circ}) - \tilde{\mathcal{J}}_x^H(\theta, 90^{\circ}) \cdot \eta \cos \theta \right] \quad (44)$$

$$E_{\varphi}(\theta, 0) = j\beta \frac{e^{-j\beta r}}{4\pi r} \left[-\eta \tilde{\mathcal{J}}_x^H(\theta, 0) + \tilde{\mathcal{J}}_y^E(\theta, 0) \cdot \cos \theta \right] \quad (45)$$

$$E_{\varphi}(\theta, 90^{\circ}) = j\beta \frac{e^{-j\beta r}}{4\pi r} \left[-\eta \tilde{\mathcal{J}}_y^H(\theta, 90^{\circ}) - \tilde{\mathcal{J}}_x^E(\theta, 90^{\circ}) \cdot \cos \theta \right]. \quad (46)$$

The 3-D field dependence on the directional angles can be approximated from the 2-D dependences in the equations (43) through (46) as

$$\mathbf{E}(\theta, \varphi) \approx j\beta \frac{e^{-j\beta r}}{4\pi r} \left[\cos \varphi \cdot \mathbf{E}(\theta, 0) + \sin \varphi \cdot \mathbf{E}(\theta, 90^{\circ}) \right], \quad (47)$$

The total 3-D amplitude pattern of the field defined in (47) is obtained as

$$|\mathbf{E}(\theta, \varphi)| \approx \left\{ \cos^2 \varphi \cdot \left[E_{\theta}^2(\theta, 0) + E_{\varphi}^2(\theta, 0) \right] + \sin^2 \varphi \cdot \left[E_{\theta}^2(\theta, 90^{\circ}) + E_{\varphi}^2(\theta, 90^{\circ}) \right] + \sin(2\varphi) \cdot \left[E_{\theta}(\theta, 0) \cdot E_{\theta}(\theta, 90^{\circ}) + E_{\varphi}(\theta, 0) \cdot E_{\varphi}(\theta, 90^{\circ}) \right] \right\}^{1/2}. \quad (48)$$

In the pattern calculation, we drop the factor $j\beta e^{-j\beta r} / (4\pi r)$. Also, it can be shown that the last term in (48) is

$$E_{\theta}(\theta, 0)E_{\theta}(\theta, 90^{\circ}) + E_{\varphi}(\theta, 0)E_{\varphi}(\theta, 90^{\circ}) = (1 - \cos^2 \theta) \cdot \left[\tilde{\mathcal{J}}_x^E(\theta, 0)\tilde{\mathcal{J}}_y^E(\theta, 90^{\circ}) + \eta^2 \tilde{\mathcal{J}}_x^H(\theta, 90^{\circ})\tilde{\mathcal{J}}_y^H(\theta, 0) \right]. \quad (49)$$

For high-directivity antennas, the angles θ , at which the antenna has significant pattern values, are small, and the term given in (49) can be neglected. Thus, the approximation of the 3-D pattern in terms of two orthogonal 2-D patterns reduces to the simple expression

$$|\mathbf{E}(\theta, \varphi)| \approx \sqrt{\cos^2 \varphi \cdot |\mathbf{E}(\theta, 0)|^2 + \sin^2 \varphi \cdot |\mathbf{E}(\theta, 90^{\circ})|^2}. \quad (50)$$

Sometimes, the *phase pattern* of the far field is also measured. This requires phase reference and can be performed using vector network analyzers.

6. Gain Measurements*

The gain measurements require essentially the same environment as the pattern measurements. To measure the gain of antennas operating above 1 GHz, usually, anechoic chambers are used. Between 0.1 GHz and 1 GHz, ground-reflection ranges are used.

Below 0.1 GHz, directive antennas are very large and the ground effects become increasingly pronounced. Usually the gain at these frequencies is measured directly in the environment of operation. Same holds for high-frequency antennas operating in a complicated environment (mounted on vehicles or aircrafts).

We consider three gain-measurement techniques. The first two belong to the so-called *absolute-gain measurements*, and they are: the ***two-antenna method***, and the ***three-antenna method***. The third method is called the ***gain-transfer*** (or ***gain-comparison***) ***method***.

A. The two-antenna method

The ***two-antenna method*** is based on Friis transmission equation and it needs two identical samples of the tested antenna. One is the radiating antenna, and the other one is receiving. Assuming that the antennas are well matched in terms of impedance and polarization, the Friis transmission equation is

$$\frac{P_r}{P_t} = \left(\frac{\lambda}{4\pi R} \right)^2 G_t G_r, \text{ where } G_t = G_r = G, \quad (51)$$

or, in dB,

$$G_{\text{dB}} = \frac{1}{2} \left[20 \log_{10} \left(\frac{4\pi R}{\lambda} \right) + 10 \log_{10} \left(\frac{P_r}{P_t} \right) \right]. \quad (52)$$

One needs to know accurately the distance between the two antennas R , the received power P_r , the transmitted power P_t , and the frequency $f = c / \lambda$.

B. The three-antenna method

The ***three-antenna method*** is used when only one sample of the test antenna is available. Then, any other two antennas can be used to perform three measurements, which allow the calculation of the gains of all three antennas. All three measurements are made at a fixed known distance R between the

radiating and the transmitting antennas.

It does not matter whether an antenna is in a transmitting or in a receiving mode. What matters is that the three measurements involve all three possible pairs of antennas: antenna #1 and antenna #2; antenna #1 and antenna #3; antenna #2 and antenna #3. The calculations are again based on Friis transmission equation, which in the case of two different antennas (antenna # i and antenna # j) measured during experiment # k ($k = 1, 2, 3$) becomes

$$G_{i \text{ dB}} + G_{j \text{ dB}} = 20 \log_{10} \left(\frac{4\pi R}{\lambda} \right) + 10 \log_{10} \left(\frac{P_r}{P_t} \right)^{(k)}. \quad (53)$$

The system of equations describing all three experiments is

$$\begin{aligned} G_{1 \text{ dB}} + G_{2 \text{ dB}} &= 20 \log_{10} \left(\frac{4\pi R}{\lambda} \right) + 10 \log_{10} \left(\frac{P_r}{P_t} \right)^{(1)}, \\ G_{1 \text{ dB}} + G_{3 \text{ dB}} &= 20 \log_{10} \left(\frac{4\pi R}{\lambda} \right) + 10 \log_{10} \left(\frac{P_r}{P_t} \right)^{(2)}, \\ G_{2 \text{ dB}} + G_{3 \text{ dB}} &= 20 \log_{10} \left(\frac{4\pi R}{\lambda} \right) + 10 \log_{10} \left(\frac{P_r}{P_t} \right)^{(3)}. \end{aligned} \quad (54)$$

The right-hand sides of the equations in (54) are known if the distance R and the ratios of the received-to-transmitted power are known. Thus, the following system of three equations with three unknowns is obtained

$$\begin{cases} G_{1 \text{ dB}} + G_{2 \text{ dB}} = A \\ G_{1 \text{ dB}} + G_{3 \text{ dB}} = B \\ G_{2 \text{ dB}} + G_{3 \text{ dB}} = C \end{cases} \quad (55)$$

The solution to the system of equations in (55) is

$$\begin{aligned} G_{1 \text{ dB}} &= \frac{A + B - C}{2}, \\ G_{2 \text{ dB}} &= \frac{A - B + C}{2}, \\ G_{3 \text{ dB}} &= \frac{-A + B + C}{2}. \end{aligned} \quad (56)$$

C. The gain-comparison method

The *gain-comparison method* requires an antenna the gain of which is exactly known (called *gain standard*) and a transmitting antenna the gain of which does not need to be known. Two sets of measurements are performed.

- 1) The test antenna is in a receiving mode, and its received power P_{AUT} is measured.
- 2) The gain standard is in a receiving mode in exactly the same arrangement (the distance R and the transmitted power P_0 are kept the same), and its received power P_{GS} is measured.

In both measurements, the receiving antennas must be matched to their loads (the receiver).

The calculation of the test antenna gain in dB uses Friis' transmission equation. The two measurements lead to the following system of equations:

$$\begin{aligned} G_{\text{AUT dB}} + G_0 \text{ dB} &= 20 \log_{10} \left(\frac{4\pi R}{\lambda} \right) + 10 \log_{10} \left(\frac{P_{\text{AUT}}}{P_0} \right)^{(1)}, \\ G_{\text{GS dB}} + G_0 \text{ dB} &= 20 \log_{10} \left(\frac{4\pi R}{\lambda} \right) + 10 \log_{10} \left(\frac{P_{\text{GS}}}{P_0} \right)^{(2)}. \end{aligned} \quad (57)$$

Here,

$G_{\text{AUT dB}}$ is the gain of the test antenna;

$G_{\text{GS dB}}$ is the gain of the gain standard; and

$G_0 \text{ dB}$ is the gain of the transmitting antenna.

From (57), we derive the expression for the gain of the test antenna:

$$G_{\text{AUT dB}} = G_{\text{GS dB}} + 10 \cdot \log_{10} \left(\frac{P_{\text{AUT}}}{P_{\text{GS}}} \right). \quad (58)$$

If the test antenna is circularly or elliptically polarized, two orthogonal linearly polarized gain standards must be used in order to obtain the partial gains corresponding to each linearly polarized component. The total gain of the test antenna is

$$G_{\text{AUT dB}} = 10 \log_{10} (G_{\text{AUT}_v} + G_{\text{AUT}_h}), \quad (59)$$

where G_{AUT_v} is the dimensionless gain of the test antenna measured with the vertically polarized gain standard and G_{AUT_h} is the dimensionless gain of the test antenna measured with the horizontally polarized gain standard.

7. Directivity Measurements*

The directivity measurements are directly related to the pattern measurements. Once the pattern is found over a sphere, the directivity can be determined using the definition:

$$D_0 = 4\pi \frac{F_{\max}(\theta_0, \varphi_0)}{\int_0^{2\pi} \int_0^{\pi} F(\theta, \varphi) \sin \theta d\theta d\varphi}, \quad (60)$$

where $F(\theta, \varphi)$ is the power pattern of the test antenna and (θ_0, φ_0) is the direction of maximum radiation.

Generally, $F(\theta, \varphi)$ is measured by sampling the field over a sphere of constant radius R . The spacing between the sampling points depends on the directive properties of the antenna and on the desired accuracy. The integral

$$\Pi = \int_0^{2\pi} \int_0^{\pi} F(\theta, \varphi) \sin \theta d\theta d\varphi \quad (61)$$

is computed numerically, e.g.,

$$\Pi \approx \frac{\pi}{N} \frac{2\pi}{M} \sum_{j=1}^M \left[\sum_{i=1}^N F(\theta_i, \varphi_j) \sin \theta_i \right]. \quad (62)$$

If the antenna is circularly or elliptically polarized, two measurements of the above type must be carried out in order to determine the partial directivities, D_θ and D_φ . Then, the total directivity is calculated as

$$D_0 = D_\theta + D_\varphi, \quad (63)$$

where the partial directivities are defined as

$$D_\theta = 4\pi \frac{F_{\theta \max}}{\Pi_\theta + \Pi_\varphi}, \quad (64)$$

$$D_{\varphi} = 4\pi \frac{F_{\varphi \max}}{\Pi_{\theta} + \Pi_{\varphi}}. \quad (65)$$

8. Radiation Efficiency, e_{cd} *

In order to calculate the radiation efficiency, the gain and the directivity must be measured first. Factors like impedance mismatch and polarization mismatch have to be minimized during these measurements. The radiation efficiency is then calculated using its definition:

$$e_{cd} = \frac{\text{Gain}}{\text{Directivity}}. \quad (66)$$

9. Impedance Measurements*

The input impedance of an antenna is calculated via the reflection coefficient at its terminals Γ , which are connected to a transmission line of known characteristic impedance Z_c . If the magnitude and the phase of Γ are known, then, the antenna input impedance is calculated as

$$Z_A = Z_c \left(\frac{1+\Gamma}{1-\Gamma} \right), \Omega. \quad (67)$$

Γ is usually measured using a vector network analyzer (VNA). The VNA measures the complex S -parameters of microwave networks. The antenna is a single-port device, therefore, $\Gamma = S_{11}$.

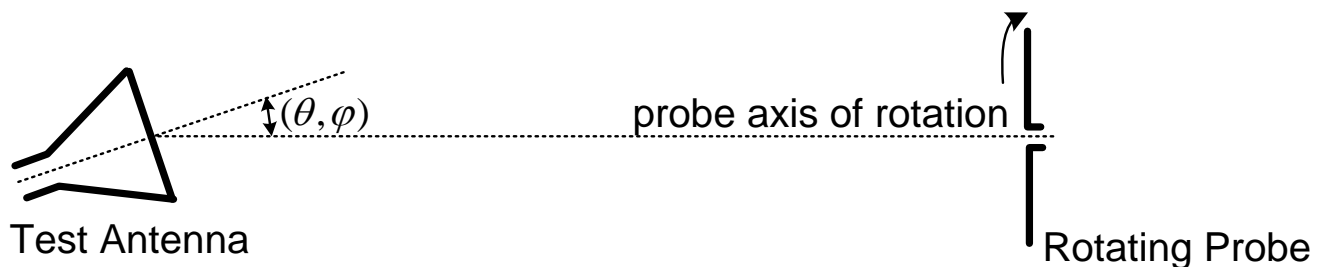
10. Polarization Measurements*

A complete description of the antenna polarization is given by the polarization ellipse (the axial ratio and the tilt angle), as well as the sense of rotation (clockwise, or counter-clockwise). In general, the polarization of an antenna is not the same in every direction, i.e., it depends on the observation angle. That is why, often, many measurements are required according to the desired degree of polarization description.

The polarization measurement methods are classified into three general categories.

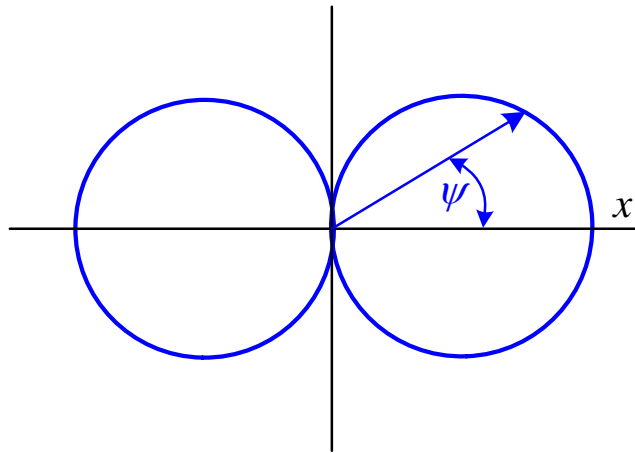
- **Partial methods** give incomplete information about the polarization but are simple and require conventional equipment.
- **Comparison methods** yield complete polarization information; however, they require a polarization standard.
- **Absolute methods** yield complete polarization information; and, they do not require a polarization standard.

The **polarization-pattern method** is a common partial method. It produces the polarization ellipse parameters (the axial ratio and the tilt angle) in a given direction of radiation. It cannot determine however the sense of rotation. The AUT can be either in transmitting or in receiving mode. The other antenna (the probe) must be linearly polarized, e.g., a dipole, and its pattern must be accurately known. A typical arrangement for the polarization-pattern measurement is given below.

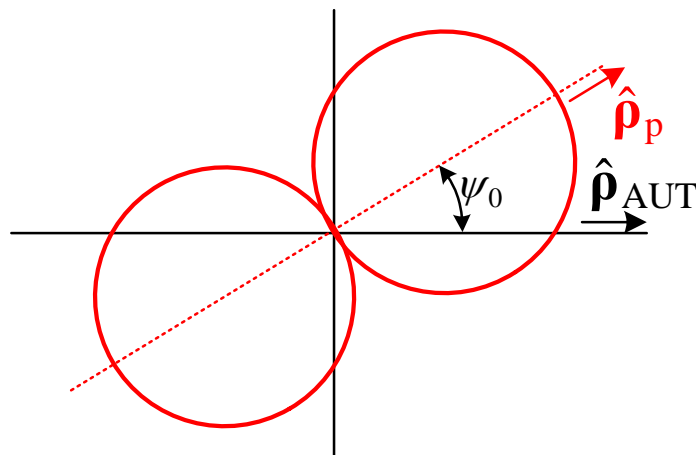


The signal at the output of the probe depends on two factors: the polarization of the test antenna and the angle of the probe's rotation. The signal level is recorded and plotted versus the angle of rotation. Thus, a **polarization pattern** is obtained for the considered direction of radiation.

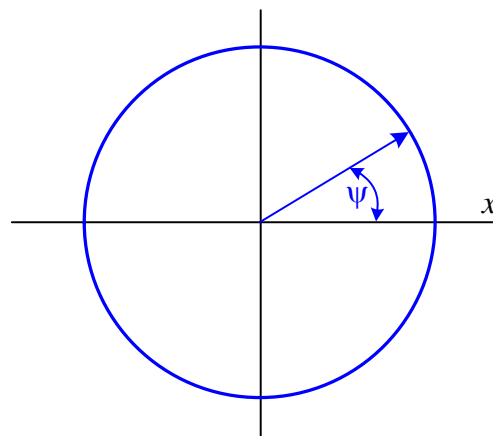
Let us assume that the probe is initially in such an orientation that it is polarized along the x -axis where x is perpendicular to the line connecting the two antennas. The linearly polarized probe rotates and, therefore, its polarization vector $\hat{\mathbf{p}}_p$ also rotates forming a time-dependent angle ψ with the x -axis. If the AUT is a linearly polarized antenna with a polarization vector $\hat{\mathbf{p}}_{\text{AUT}} = \hat{\mathbf{x}}$, then the PLF is $|\hat{\mathbf{p}}_p \cdot \hat{\mathbf{p}}_{\text{AUT}}^*|^2 = \cos^2 \psi$; see the figure below showing the polar plot of this PLF. This PLF pattern is the AUT polarization pattern.



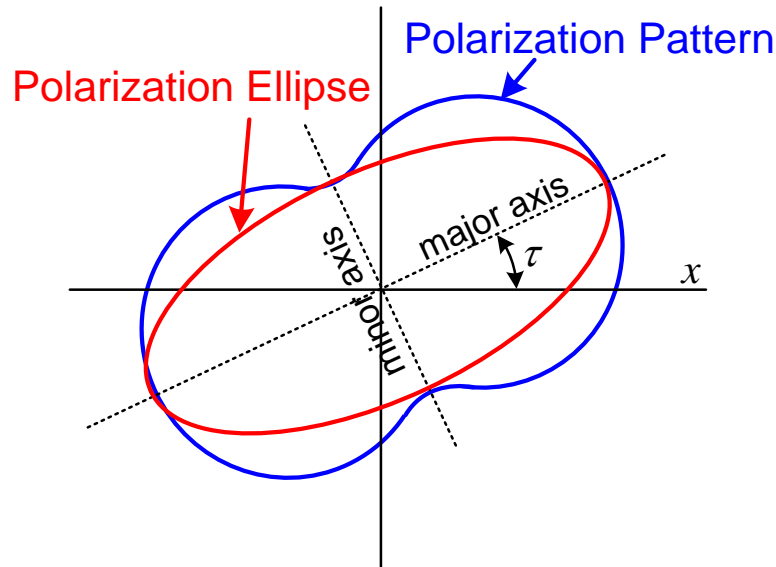
In general, if the AUT is linearly polarized, the polarization pattern will be the same as the cosine pattern shown above but could be tilted depending on the initial angle ψ_0 between the polarization axes of the probe and the AUT; see the illustration below.



If the AUT is circularly polarized, the polarization pattern is a circle regardless of the initial mutual orientation of the probe and the AUT.

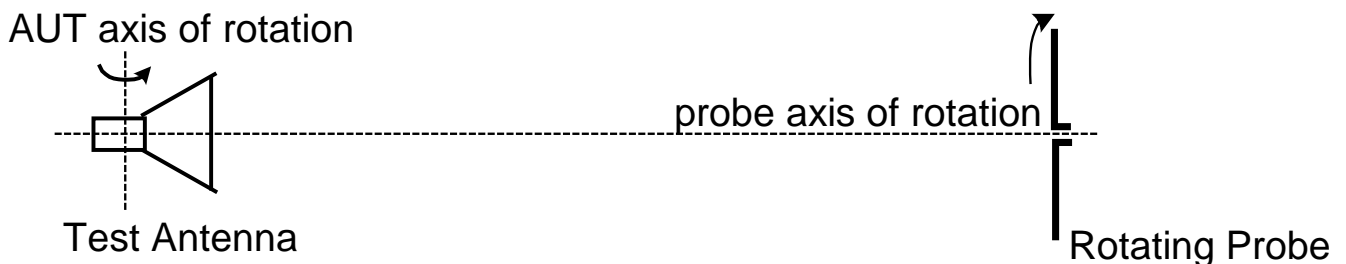


In the general case of elliptically polarized test antenna, the polarization pattern is a dumb-bell contour, which allows for the direct calculation of the axial ratio and the tilt angle τ of the polarization ellipse as is shown in the figure below.



The polarization-pattern method cannot provide information about the sense of rotation. However, this can be easily established by the use of circularly polarized probes (e.g. spiral antennas): one of a clockwise polarization, and the other one of a counter-clockwise polarization. Whichever receives a stronger signal determines the sense of rotation.

Another partial method is the *axial-ratio pattern method*. The arrangement is very similar to that of the polarization-pattern method. The only difference is that now the AUT (which is usually in a receiving mode) is rotated in a desired plane by the antenna positioning mechanism. The probe rotates with much larger angular frequency than the AUT because it should complete one full turn at approximately every degree of rotation of the test antenna.



As a result of the measurement described above, a 2-D pattern is obtained, which allows for the calculation of the axial ratio of the polarization at any direction of the measured 2-D pattern. Such a pattern (in dB) of an antenna, which is nearly circularly polarized along $\theta = 0$, is shown below.

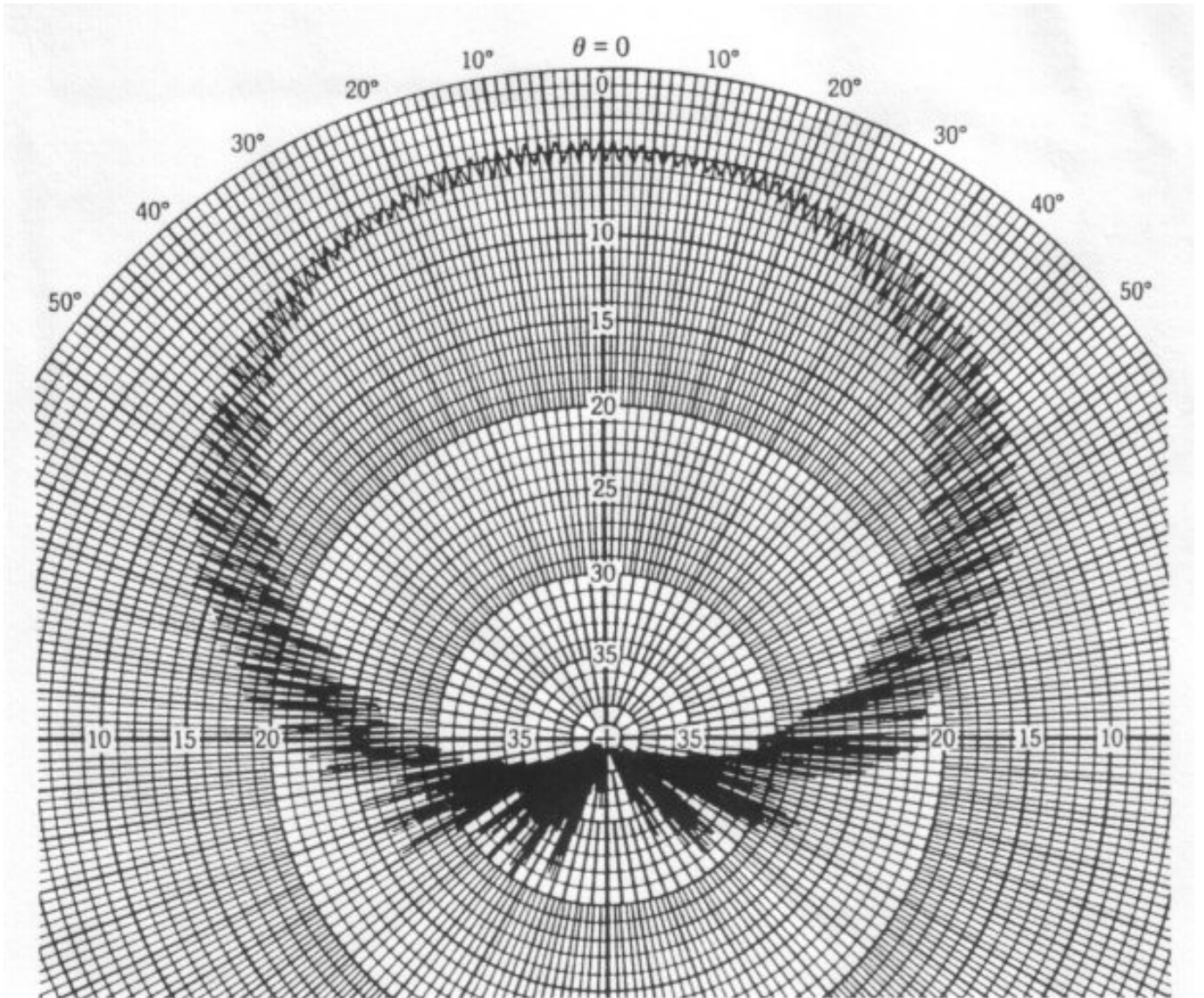


Figure 16.26 Pattern of a circularly polarized test antenna taken with a rotating, linearly polarized, source antenna [E. S. Gillespie, "Measurement of Antenna Radiation Characteristics on Far-Field Ranges," in *Antenna Handbook* (Y. T. Lo & S. W. Lee, eds.), 1988, © Van Nostrand Reinhold Co., Inc.]

[Balanis]

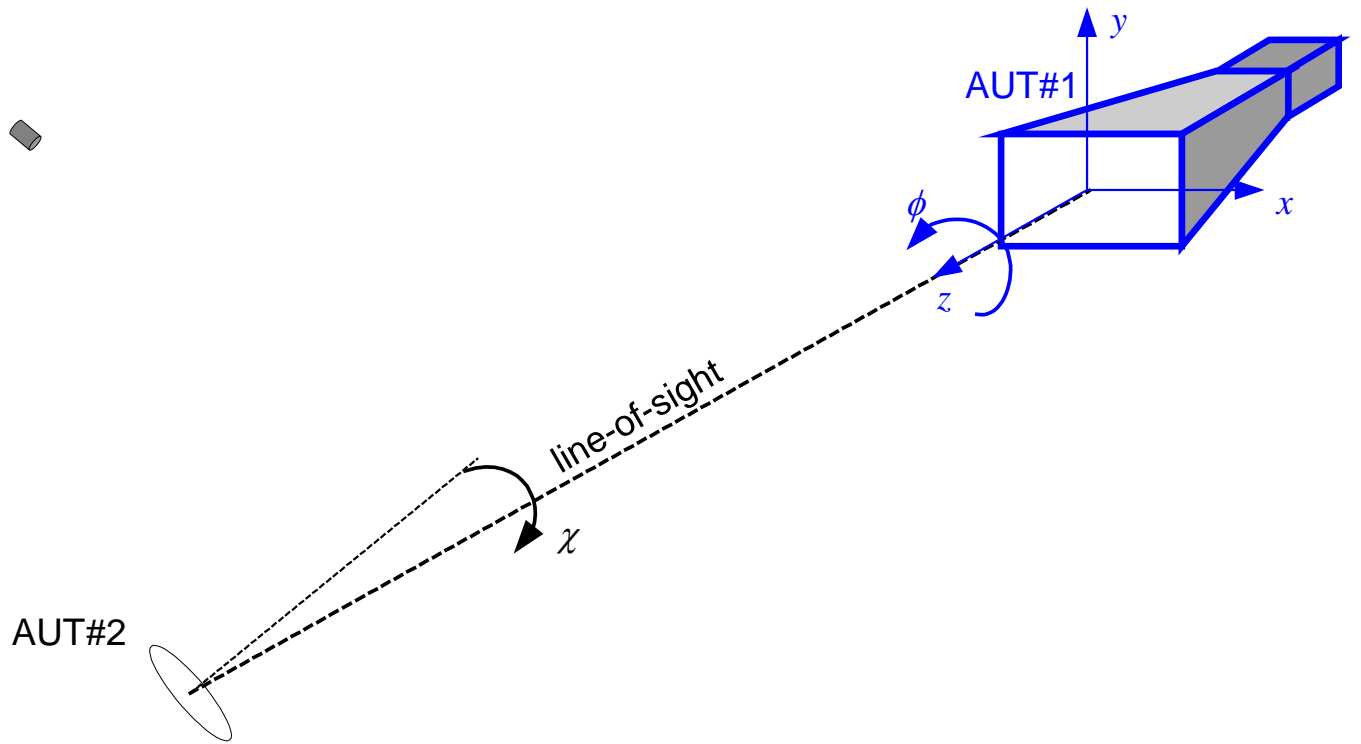
From the plot above, it is obvious that the axial-ratio pattern has an inner envelope and an outer envelope. The ratio of the outer envelope to the inner one for a given angle gives the axial ratio of the field polarization in this direction. For example, the pattern above shows that the test antenna is nearly circularly polarized along boresight ($\theta = 0$), where the axial ratio is close to one. At greater observation angles, however, its polarization becomes elliptical of increasingly larger axial ratio.

The axial-ratio pattern method yields only the axial ratio of the polarization ellipse. It does not give information about the tilt angle and the sense of rotation. However, it is very fast and convenient to implement in any antenna test range. The tilt angle at selected directional angles can be always clarified later with the polarization-pattern method.

A powerful absolute polarization measurement method is the ***three-antenna method***. It yields full polarization information for all three antennas. The only *a-priori* knowledge required is the approximate tilt angle of one of the three antennas.

The method requires the measurement of the amplitude and the phase of the normalized received voltage in three experiments, which involve: 1) antenna #1 and antenna #2; 2) antenna #1 and antenna #3; and 3) antenna #2 and antenna #3. All three experiments must use the same measurement set-up. The three complex voltage phasors are measured as a function of the angles ϕ and χ , which are the angles of rotation of the antennas about the antenna-range axis (usually, this is the line-of-sight between them).

An example set-up is shown in the figure below. First, the AUT#1 is scanned for $\phi \in [0^\circ, 360^\circ]$ usually with a step of $\Delta\phi = 1^\circ$. Then, the angle of AUT#2 is incremented by $\Delta\chi$ (usually, $\Delta\chi \approx 15^\circ$), and AUT#1 is scanned again. This is repeated until the angle χ sweeps the whole range from 0° to 360° .



Three complex quantities $M_{m,n}$ are then calculated from the double Fourier transform of the voltage phasor patterns:

$$M_{m,n} = \frac{\int_0^{2\pi} \int_0^{2\pi} \tilde{V}_{m,n}(\phi, \chi) e^{+j(\phi+\chi)} d\phi d\chi}{\int_0^{2\pi} \int_0^{2\pi} \tilde{V}_{m,n}(\phi, \chi) e^{-j(\phi+\chi)} d\phi d\chi}, \quad (m,n) = (1,2), \text{ or } (1,3), \text{ or } (2,3). \quad (68)$$

It can be shown (see references [6],[7],[8]) that $M_{m,n}$ are equal to the dot products of the circular polarization ratios (see reference [3]; for definition of polarization ratio refer to Lecture 5, eq. 5.21) of the two antennas used in the respective measurement:

$$\begin{aligned} \hat{\mathbf{p}}_{c1} \cdot \hat{\mathbf{p}}_{c2} &= M_{1,2} \\ \hat{\mathbf{p}}_{c1} \cdot \hat{\mathbf{p}}_{c3} &= M_{1,3} \\ \hat{\mathbf{p}}_{c2} \cdot \hat{\mathbf{p}}_{c3} &= M_{2,3} \end{aligned} \quad (69)$$

The system in (69) is used to solve for the three circular polarization ratios:

$$\hat{\rho}_{c1} = \sqrt{\frac{M_{12}M_{13}}{M_{23}}}; \hat{\rho}_{c2} = \sqrt{\frac{M_{12}M_{23}}{M_{13}}}; \hat{\rho}_{c3} = \sqrt{\frac{M_{23}M_{13}}{M_{12}}}. \quad (70)$$

The square root of a complex number implies ambiguity in the phase calculations for the polarization vectors. Here, it becomes necessary to have an approximate knowledge of the tilt angle of one of the antennas. The circular polarization ratios are directly related to the polarization ellipse; see Lecture 5, [2], [3].

References

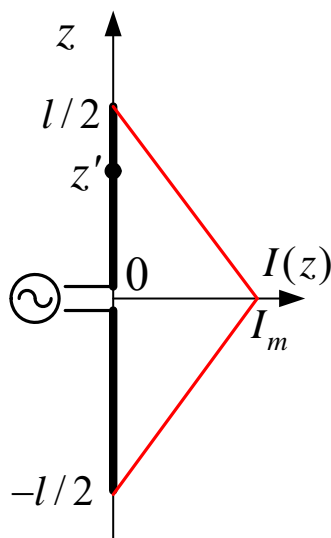
- [1] *IEEE Standard Test Procedures for Antennas*, IEEE Std 149-1979, IEEE Inc., 1979, distributed by Wiley-Interscience.
- [2] J.S. Hollis, T.J. Lyon, and L. Clayton, Jr., *Microwave Antenna Measurements*, Scientific-Atlanta, Inc., Atlanta, Georgia, July 1970.
- [3] W.H. Kummer and E.S. Gillespie, "Antenna measurements-1978," *Proc. IEEE*, vol. 66, No. 4, pp. 483-507, April 1978.
- [4] C.A. Balanis, *Antenna Theory*, 2nd ed., John Wiley & Sons, Inc., New York, 1997.
- [5] J. Kraus, *Antennas*, 2nd ed., McGraw Hill, Inc., New York, 1988.
- [6] J.R. Jones and D.W. Hess, "Automated three-antenna polarization measurements using digital signal processing," white paper available for download at <http://www.mi-technologies.com/techliterature/>
- [7] E.B. Joy, and D.T. Paris, "A practical method for measuring the complex polarization ratio of arbitrary antennas," *IEEE Trans. on Antennas and Propagation*, vol. AP-21, pp. 432-435, 1973.
- [8] A.C. Newell, and D.M. Kerns, "Determination of both polarization and power gain of antennas by a generalised 3-antenna measurement method," *Electronics Letters*, vol. 7, No. 3, pp. 68-70, February 11, 1971.

Lecture 9: Linear Wire Antennas – Dipoles and Monopoles

(Small electric dipole antenna. Finite-length dipoles. Half-wavelength dipole. Method of images – revision. Vertical infinitesimal dipole above a conducting plane. Monopoles. Horizontal infinitesimal dipole above a conducting plane.)

The dipole and the monopole are arguably the two most widely used antennas across the UHF, VHF and lower-microwave bands. Arrays of dipoles are commonly used as base-station antennas in land-mobile systems. The monopole and its variations are common in portable equipment, such as cellular telephones, cordless telephones, automobiles, trains, etc. It has attractive features such as simple construction, sufficiently broadband characteristics for voice communication, small dimensions at high frequencies. Alternatives to the monopole antenna for hand-held units is the inverted F and L antennas, the microstrip patch antenna, loop and spiral antennas, and others. The printed inverted F antenna (PIFA) is arguably the most common antenna design used in modern handheld phones.

1. Small Dipole



$$\frac{\lambda}{50} < l \leq \frac{\lambda}{10} \quad (9.1)$$

If we assume that (9.1) holds in addition to $R \approx r$, the maximum phase error in (βR) that can occur is

$$e_{\max} = \frac{\beta l}{2} = \frac{\pi}{10} \approx 18^\circ,$$

which corresponds to an observation direction at $\theta = 0^\circ$. *Reminder:* A maximum total phase error less than $\pi/8$ is acceptable since it does not affect substantially the integral solution for the vector potential \mathbf{A} . Note that the approximation $R \approx r$ implies that $r \gg l$.

The current is a triangular function of z' :

$$I(z') = \begin{cases} I_m \cdot \left(1 - \frac{z'}{l/2}\right), & 0 \leq z' \leq l/2 \\ I_m \cdot \left(1 + \frac{z'}{l/2}\right), & -l/2 \leq z' \leq 0 \end{cases} \quad (9.2)$$

The VP integral is obtained as

$$\mathbf{A} = \hat{\mathbf{z}} \frac{\mu}{4\pi} \left[\int_{-l/2}^0 I_m \left(1 + \frac{z'}{l/2}\right) \frac{e^{-j\beta R}}{R} dz' + \int_0^{l/2} I_m \left(1 - \frac{z'}{l/2}\right) \frac{e^{-j\beta R}}{R} dz' \right]. \quad (9.3)$$

The solution of (9.3) is simple when we assume that $R \approx r$:

$$\mathbf{A} = \hat{\mathbf{z}} \frac{1}{2} \left[\frac{\mu}{4\pi} I_m l \frac{e^{-j\beta r}}{r} \right]. \quad (9.4)$$

The further away from the antenna the observation point is, the more accurate the expression in (9.4). Note that *the result in (9.4) is exactly one-half of the result obtained for \mathbf{A} of an infinitesimal dipole of the same length*, if I_m were the current uniformly distributed along the dipole. This is expected because we made the same approximation for R , as in the case of the infinitesimal dipole with a constant current distribution, and we integrated a triangular function along l , whose average is $I_0 = I_{av} = 0.5I_m$.

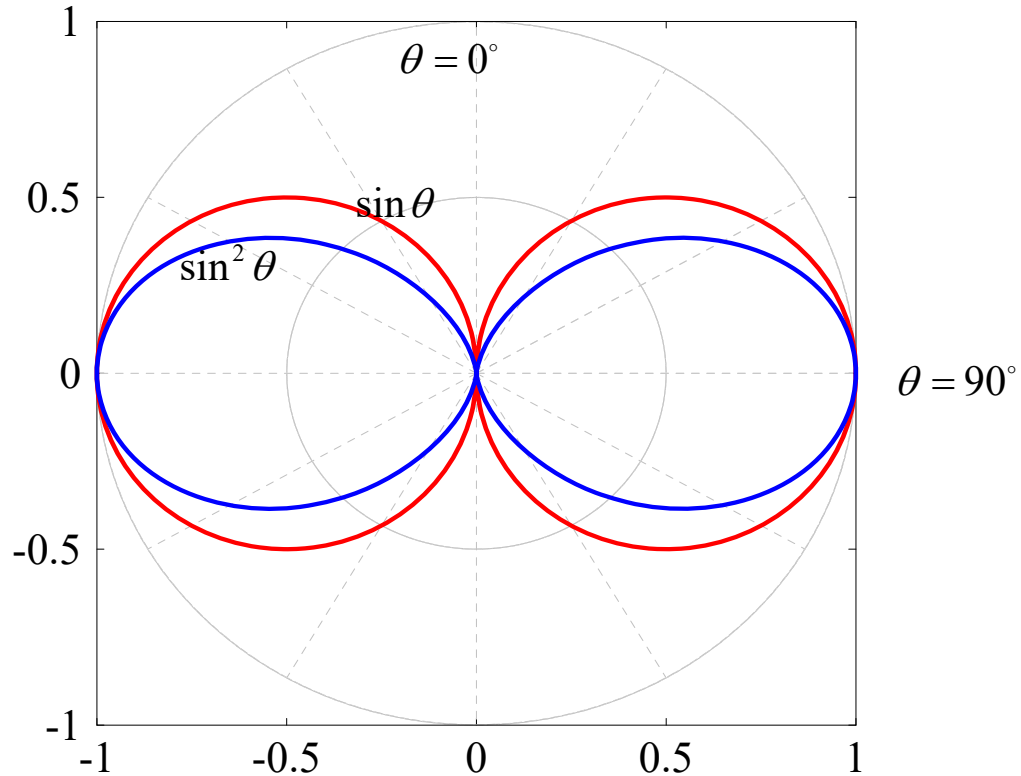
Therefore, we need not repeat all the calculations of the field components, power and antenna parameters; we simply use the infinitesimal-dipole field multiplied by a factor of 0.5:

$$\begin{aligned} E_\theta &\approx j\eta \frac{\beta I_m l}{8\pi} \frac{e^{-j\beta r}}{r} \sin \theta \\ H_\varphi &\approx j \frac{\beta I_m l}{8\pi} \frac{e^{-j\beta r}}{r} \sin \theta, \quad \beta r \gg 1. \\ E_r = E_\varphi = H_r = H_\theta &= 0 \end{aligned} \quad (9.5)$$

The normalized field pattern is the same as that of the infinitesimal dipole:

$$\bar{E}(\theta, \varphi) = \sin \theta. \quad (9.6)$$

The power pattern: $\bar{U}(\theta, \varphi) = \sin^2 \theta$ (9.7)



The beam solid angle:

$$\Omega_A = \int_0^{2\pi} \int_0^{\pi} \sin^2 \theta \cdot \sin \theta d\theta d\varphi,$$

$$\Omega_A = 2\pi \cdot \int_0^{\pi} \sin^3 \theta d\theta = 2\pi \cdot \frac{4}{3} = \frac{8\pi}{3}$$

The directivity:

$$D_0 = \frac{4\pi}{\Omega_A} = \frac{3}{2} = 1.5. \quad (9.8)$$

As expected, the directivity, the beam solid angle as well as the effective aperture are the same as those of the infinitesimal dipole because the normalized patterns of both dipoles are the same.

The radiated power is four times less than that of an infinitesimal dipole of the same length and current $I_0 = I_m$ because the far fields are twice smaller in magnitude:

$$\Pi = \frac{1}{4} \cdot \frac{\pi}{3} \eta \left(\frac{I_m l}{\lambda} \right)^2 = \frac{\pi}{12} \eta \left(\frac{I_m l}{\lambda} \right)^2. \quad (9.9)$$

As a result, the radiation resistance is also four times smaller than that of the infinitesimal dipole:

$$R_r = \frac{\pi}{6} \eta \left(\frac{l}{\lambda} \right)^2 = 20\pi^2 \left(\frac{l}{\lambda} \right)^2. \quad (9.10)$$

2. Finite-length Infinitesimally Thin Dipole

A good approximation of the current distribution along the dipole's length is the sinusoidal one:

$$I(z') = \begin{cases} I_0 \sin \left[\beta \left(\frac{l}{2} - z' \right) \right], & 0 \leq z' \leq l/2 \\ I_0 \sin \left[\beta \left(\frac{l}{2} + z' \right) \right], & -l/2 \leq z' \leq 0. \end{cases} \quad (9.11)$$

It can be shown that the VP integral

$$\mathbf{A} = \hat{\mathbf{z}} \frac{\mu}{4\pi} \int_{-l/2}^{l/2} I(z') \frac{e^{-j\beta R}}{R} dz' \quad (9.12)$$

has an analytical (closed form) solution. Here, however, we follow a standard approach used to calculate the far field for an arbitrary wire antenna. It is based on the solution for the field of the infinitesimal dipole. The finite-length dipole is subdivided into an infinite number of infinitesimal dipoles of length dz' . Each such dipole produces the elementary far field given by

$$\begin{aligned} dE_\theta &\approx j\eta\beta I_e(z') \frac{e^{-j\beta R}}{4\pi R} \sin\theta \cdot dz' \\ dH_\phi &\approx j\beta I_e(z') \frac{e^{-j\beta R}}{4\pi R} \sin\theta \cdot dz' \\ dE_r &\approx dE_\phi \approx dH_r \approx dH_\theta \approx 0 \end{aligned} \quad (9.13)$$

where $R = [x^2 + y^2 + (z - z')^2]^{1/2}$ and $I_e(z')$ denotes the value of the current element at z' . Using the far-zone approximations,

$$\left| \begin{array}{l} \frac{1}{R} \approx \frac{1}{r}, \text{ for the amplitude factor} \\ R \approx r - z' \cos \theta, \text{ for the phase factor} \end{array} \right. \quad (9.14)$$

the following approximation of the elementary far field is obtained:

$$dE_{\theta} \approx j\eta\beta I_e \left(\frac{e^{-j\beta r}}{4\pi r} \right) e^{j\beta z' \cos \theta} \cdot \sin \theta dz'. \quad (9.15)$$

Using the superposition principle, the total far field is obtained as

$$E_{\theta} = \int_{-l/2}^{l/2} dE_{\theta} \approx j\eta\beta \left(\frac{e^{-j\beta r}}{4\pi r} \right) \cdot \sin \theta \cdot \int_{-l/2}^{l/2} I_e(z') e^{j\beta z' \cos \theta} dz'. \quad (9.16)$$

The *first factor*

$$g(\theta) = j\eta\beta \left(\frac{e^{-j\beta r}}{r} \right) \sin \theta \quad (9.17)$$

is called the ***element factor***. The element factor in this case is the far field produced by an infinitesimal dipole of unit current element $Il = 1$ (A × m). The element factor is the same for any current element, provided the angle θ is always associated with the axis of the current flow. The *second factor*

$$f(\theta) = \int_{-l/2}^{l/2} I_e(z') e^{j\beta z' \cos \theta} dz' \quad (9.18)$$

is the ***space factor (or pattern factor, array factor)***. The pattern factor is dependent on the amplitude and phase distribution of the current at the antenna (the source distribution in space).

For the specific current distribution described by (9.11), the pattern factor is

$$f(\theta) = I_0 \left\{ \int_{-l/2}^0 \sin \left[\beta \left(\frac{l}{2} + z' \right) \right] e^{j\beta z' \cos \theta} dz' + \int_0^{l/2} \sin \left[\beta \left(\frac{l}{2} - z' \right) \right] e^{j\beta z' \cos \theta} dz' \right\}. \quad (9.19)$$

The above integrals are solved having in mind that

$$\int \sin(a + b \cdot x) e^{c \cdot x} dx = \frac{e^{cx}}{b^2 + c^2} [c \sin(a + bx) - b \cos(a + bx)]. \quad (9.20)$$

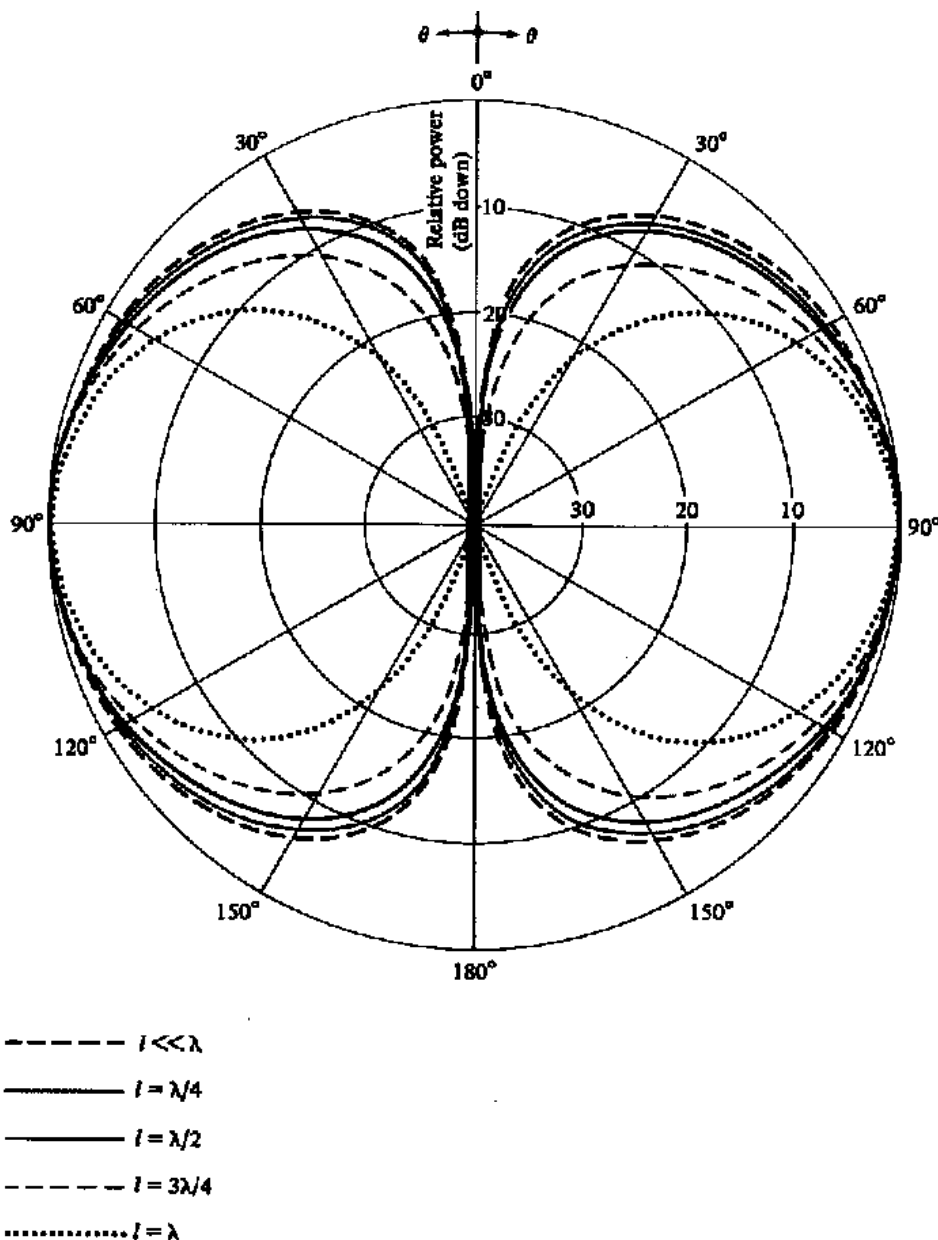
The far field of the finite-length dipole is obtained as

$$E_{\theta} = g(\theta) \cdot f(\theta) = j\eta I_0 \left(\frac{e^{-j\beta r}}{2\pi r} \right) \cdot \frac{\left[\cos\left(\frac{\beta l}{2} \cos\theta\right) - \cos\left(\frac{\beta l}{2}\right) \right]}{\sin\theta}. \quad (9.21)$$

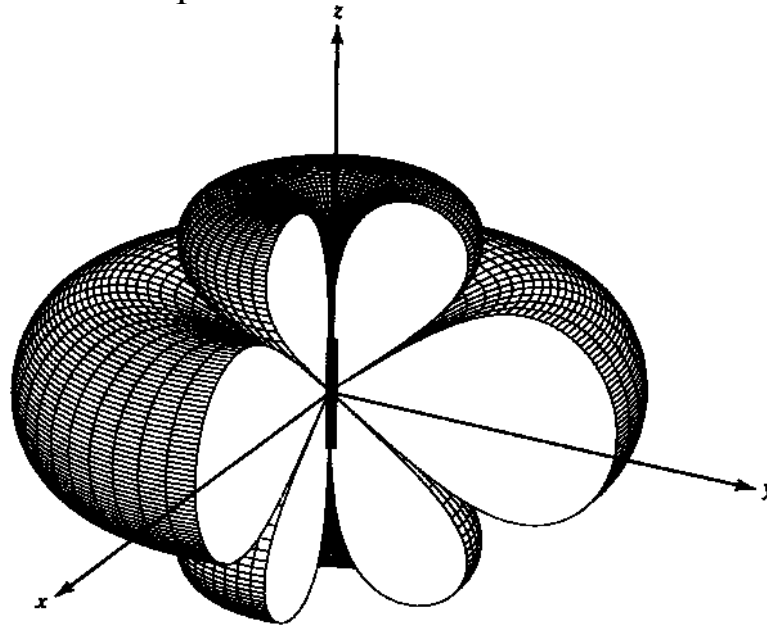
Amplitude pattern:

$$\bar{E}(\theta, \varphi) = \frac{\cos\left(\frac{\beta l}{2} \cos\theta\right) - \cos\left(\frac{\beta l}{2}\right)}{\sin\theta}. \quad (9.22)$$

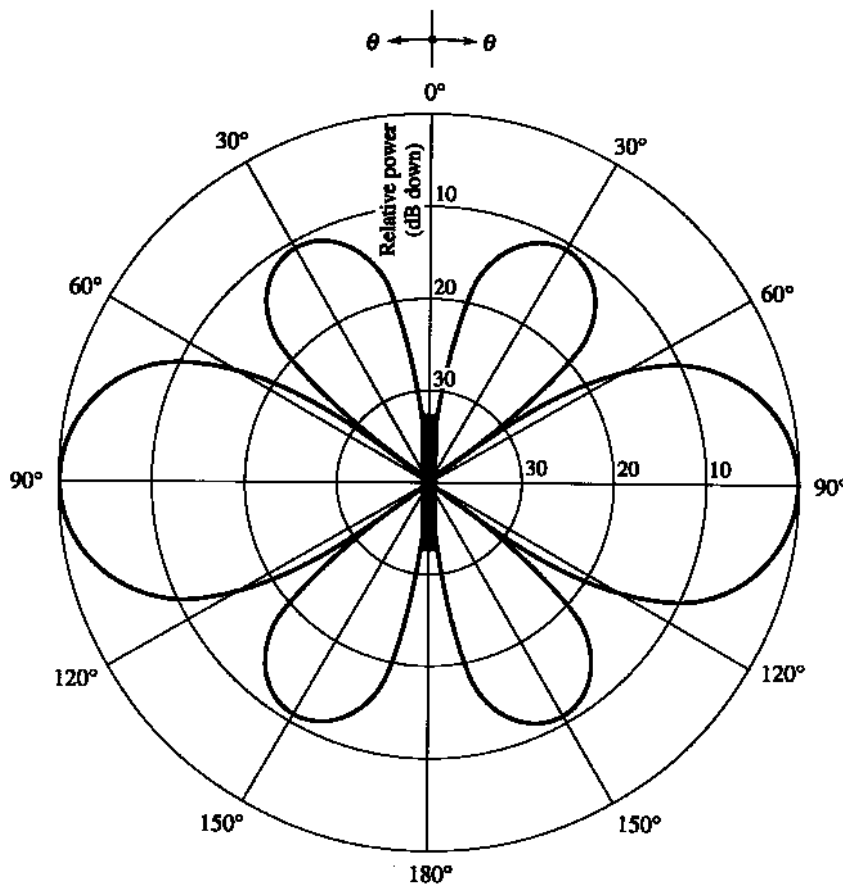
Patterns (in dB) for some dipole lengths $l \leq \lambda$ [from Balanis]:



The 3-D pattern of the dipole $l = 1.25\lambda$:



(a) Three-dimensional



(b) Two-dimensional

[Balanis]

Power pattern:

$$F(\theta, \varphi) = \frac{\left[\cos\left(\frac{\beta l}{2} \cos \theta\right) - \cos\left(\frac{\beta l}{2}\right) \right]^2}{\sin^2 \theta}. \quad (9.23)$$

Note: The maximum of $F(\theta, \varphi)$ is not necessarily unity, but for $l < 2\lambda$ the major maximum is always at $\theta = 90^\circ$.

Radiated power

First, the far-zone power flux density is calculated as

$$\mathbf{P} = \hat{\mathbf{r}} \frac{1}{2\eta} |E_\theta|^2 = \hat{\mathbf{r}} \eta \frac{I_0^2}{8\pi^2 r^2} \left[\frac{\cos(0.5\beta l \cos \theta) - \cos(0.5\beta l)}{\sin \theta} \right]^2. \quad (9.24)$$

The total radiated power is then

$$\Pi = \oiint \mathbf{P} \cdot d\mathbf{s} = \int_0^{2\pi} \int_0^\pi P \cdot r^2 \sin \theta d\theta d\varphi \quad (9.25)$$

$$\Pi = \eta \frac{I_0^2}{4\pi} \underbrace{\int_0^\pi \frac{[\cos(0.5\beta l \cos \theta) - \cos(0.5\beta l)]^2}{\sin \theta} d\theta}_{\mathfrak{I}}. \quad (9.26)$$

\mathfrak{I} is solved in terms of the cosine and sine integrals:

$$\begin{aligned} \mathfrak{I} = & C + \ln(\beta l) - C_i(\beta l) + \frac{1}{2} \sin(\beta l) [S_i(2\beta l) - 2S_i(\beta l)] + \\ & + \frac{1}{2} \cos(\beta l) [C + \ln(\beta l / 2) + C_i(2\beta l) - 2C_i(\beta l)]. \end{aligned} \quad (9.27)$$

Here,

$C \approx 0.5772$ is the Euler's constant,

$C_i(x) = \int_\infty^x \frac{\cos y}{y} dy = -\int_x^\infty \frac{\cos y}{y} dy$ is the cosine integral,

$S_i(x) = \int_0^x \frac{\sin y}{y} dy$ is the sine integral.

Thus, the radiated power can be written as

$$\Pi = \eta \frac{I_0^2}{4\pi} \cdot \mathfrak{S}. \quad (9.28)$$

Radiation resistance

The radiation resistance is defined as

$$R_r = \frac{2\Pi}{I_m^2} = \frac{I_0^2}{I_m^2} \cdot \frac{\eta}{2\pi} \cdot \mathfrak{S} \quad (9.29)$$

where I_m is the maximum current magnitude along the dipole. If the dipole is half-wavelength long or longer ($l \geq \lambda/2$), $I_m = I_0$, see (9.11). However, if $l < \lambda/2$, then $I_m < I_0$ as per (9.11). If $l < \lambda/2$ holds, the maximum current is at the dipole center (the feed point $z' = 0$) and its value is

$$I_m = I_{(z'=0)} = I_0 \sin(\beta l / 2) \quad (9.30)$$

where $\beta l / 2 < \pi / 2$, and, therefore, $\sin(\beta l / 2) < 1$. In summary,

$$\begin{aligned} I_m &= I_0 \sin(\beta l / 2), & \text{if } l \leq \lambda / 2 \\ I_m &= I_0, & \text{if } l > \lambda / 2. \end{aligned} \quad (9.31)$$

Therefore,

$$\begin{aligned} R_r &= \frac{\eta}{2\pi} \cdot \frac{\mathfrak{S}}{\sin^2(\beta l / 2)}, & \text{if } l < \lambda / 2 \\ R_r &= \frac{\eta}{2\pi} \cdot \mathfrak{S}, & \text{if } l \geq \lambda / 2. \end{aligned} \quad (9.32)$$

Directivity

The directivity is obtained as

$$D_0 = 4\pi \frac{U_{\max}}{\Pi} = 4\pi \frac{F_{\max}}{\int_0^{\pi} \int_0^{2\pi} F(\theta, \varphi) \sin \theta d\theta d\varphi} \quad (9.33)$$

where

$$F(\theta, \varphi) = \left[\frac{\cos(0.5\beta l \cos \theta) - \cos(0.5\beta l)}{\sin \theta} \right]^2$$

is the power pattern [see (9.23)]. Finally,

$$D_0 = \frac{2F_{\max}}{\mathfrak{I}}. \quad (9.34)$$

Input resistance of center-fed dipoles

The radiation resistance given in (9.32) is not necessarily equal to the input resistance because the current at the dipole center I_{in} (if its center is the feed point) is not necessarily equal to I_m . In particular, $I_{in} \neq I_m$ if $l > \lambda / 2$ and $l \neq (2n + 1)\lambda / 2$, n is any integer. Note that when $l > \lambda / 2$, $I_m = I_0$.

To obtain a general expression for the current magnitude I_{in} at the center of the dipole (assumed also to be a feed point), we note that if the dipole is lossless, the input power is equal to the radiated power. Therefore, in the case of a dipole longer than half a wavelength,

$$P_{in} = \frac{|I_{in}|^2}{2} R_{in} = \Pi = \frac{|I_0|^2}{2} R_r \text{ for } l > \lambda / 2, \quad (9.35)$$

and the input and radiation resistances relate as

$$R_{in} = \frac{|I_0|^2}{|I_{in}|^2} R_r \text{ for } l > \lambda / 2. \quad (9.36)$$

Since the current at the center of the dipole ($z' = 0$) is [see (9.11)]

$$I_{in} = I_0 \sin(\beta l / 2), \quad (9.37)$$

then,

$$R_{in} = \frac{R_r}{\sin^2(\beta l / 2)}. \quad (9.38)$$

Using the 2nd expression for R_r in (9.32), we obtain

$$R_{in} = \frac{\eta}{2\pi} \cdot \frac{\mathfrak{I}}{\sin^2(\beta l / 2)}, \quad l > \lambda / 2. \quad (9.39)$$

For a short dipole ($l \leq \lambda / 2$), $I_{in} = I_m$. It then follows from

$$P_{in} = \frac{|I_{in}|^2}{2} R_{in} = \frac{|I_m|^2}{2} R_r \text{ and } I_{in} = I_m, l \leq l/2, \quad (9.40)$$

that

$$R_{in} = R_r = \frac{\eta}{2\pi} \cdot \frac{\mathfrak{S}}{\sin^2(\beta l/2)}, l \leq \lambda/2, \quad (9.41)$$

where we have taken into account the first equation in (9.32).

In summary, the dipole's input resistance, regardless of its length, depends on the integral \mathfrak{S} as in (9.39) or (9.41), as long as the feed point is at the center.

Loss can be easily incorporated in the calculation of R_{in} bearing in mind that the power-balance relation (9.35) can be modified as

$$P_{in} = \frac{|I_{in}|^2}{2} R_{in} = \Pi + P_{loss} = \frac{|I_m|^2}{2} R_r + P_{loss}. \quad (9.42)$$

Remember that in Lecture 4, we obtained the expression for the loss of a dipole of length l as:

$$P_{loss} = \frac{I_0^2 R_{hf}}{4} \left[1 - \frac{\sin(\beta l)}{\beta l} \right]. \quad (9.43)$$

3. Half-wavelength Dipole

This is a classical and widely used thin wire antenna: $l \approx \lambda/2$.

$$\begin{aligned} E_\theta &= j\eta \frac{I_0 e^{-j\beta r}}{2\pi r} \cdot \frac{\cos(0.5\pi \cos \theta)}{\sin \theta} \\ H_\phi &= E_\theta / \eta \end{aligned} \quad (9.44)$$

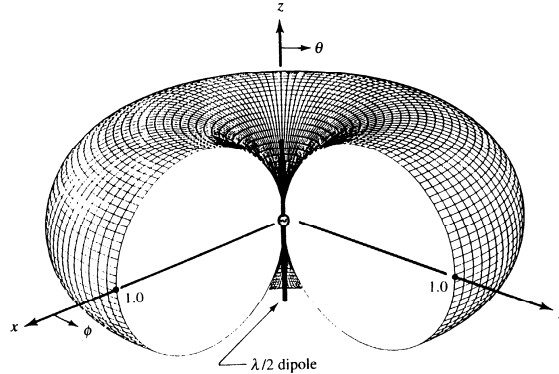
Radiated power flux density:

$$P = \frac{1}{2\eta} |E_\theta|^2 = \eta \frac{|I_0|^2}{8\pi^2 r^2} \underbrace{\left[\frac{\cos(0.5\pi \cos \theta)}{\sin \theta} \right]^2}_{F(\theta) - \text{normalized power pattern}} \approx \eta \frac{|I_0|^2}{8\pi^2 r^2} \sin^3 \theta. \quad (9.45)$$

Radiation intensity:

$$U = r^2 P = \eta \frac{|I_0|^2}{8\pi^2} \underbrace{\left[\frac{\cos(0.5\pi \cos \theta)}{\sin \theta} \right]^2}_{F(\theta) - \text{normalized power pattern}} \approx \eta \frac{|I_0|^2}{8\pi^2} \sin^3 \theta. \quad (9.46)$$

3-D power pattern (not in dB) of the half-wavelength dipole:



Radiated power

The radiated power of the half-wavelength dipole is a special case of the integral in (9.26):

$$\Pi = \eta \frac{|I_0|^2}{4\pi} \int_0^\pi \frac{\cos^2(0.5\pi \cos \theta)}{\sin \theta} d\theta \quad (9.47)$$

$$\Pi = \eta \frac{|I_0|^2}{8\pi} \int_0^{2\pi} \frac{1 - \cos y}{y} dy \quad (9.48)$$

$$\mathcal{J} = 0.5772 + \ln(2\pi) - C_i(2\pi) \approx 2.435 \quad (9.49)$$

$$\Rightarrow \Pi = \underline{\underline{2.435 \frac{\eta}{8\pi} |I_0|^2 = 36.525 |I_0|^2}}. \quad (9.50)$$

Radiation resistance:

$$R_r = \frac{2\Pi}{|I_0|^2} \approx 73 \ \Omega. \quad (9.51)$$

Directivity:

$$D_0 = 4\pi \frac{U_{\max}}{\Pi} = 4\pi \frac{U_{/\theta=90^\circ}}{\Pi} = \frac{4}{\mathfrak{J}} = \frac{4}{2.435} = 1.643. \quad (9.52)$$

Maximum effective area:

$$A_e = \frac{\lambda^2}{4\pi} D_0 \approx 0.13\lambda^2. \quad (9.53)$$

Input impedance

Since $l = \lambda / 2$, the input resistance is

$$R_{in} = R_r \approx 73 \ \Omega. \quad (9.54)$$

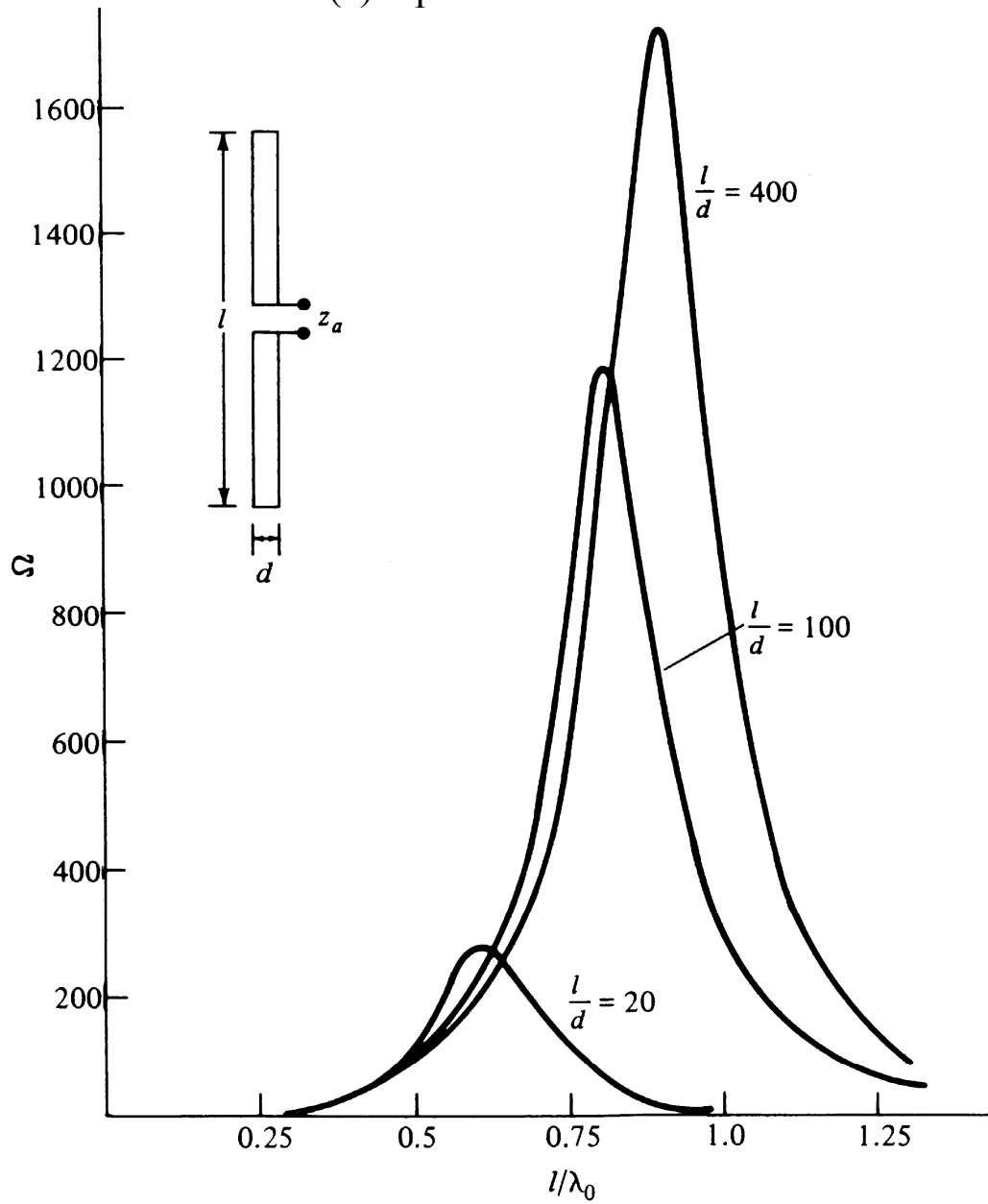
The imaginary part of the input impedance is approximately $+j42.5 \ \Omega$. To acquire maximum power transfer, this reactance has to be removed by matching (e.g., shortening) the dipole:

- thick dipole $l \approx 0.47\lambda$
- thin dipole $l \approx 0.48\lambda$.

The input reactance of the dipole is very frequency sensitive; i.e., it depends strongly on the ratio l / λ . This is to be expected from a resonant narrow-band structure operating at or near resonance such as the half-wavelength dipole. We should also keep in mind that the input impedance is influenced by the capacitance associated with the physical junction to the transmission line. The structure used to support the antenna, if any, can also influence the input impedance. That is why the curves below describing the antenna impedance are only representative.

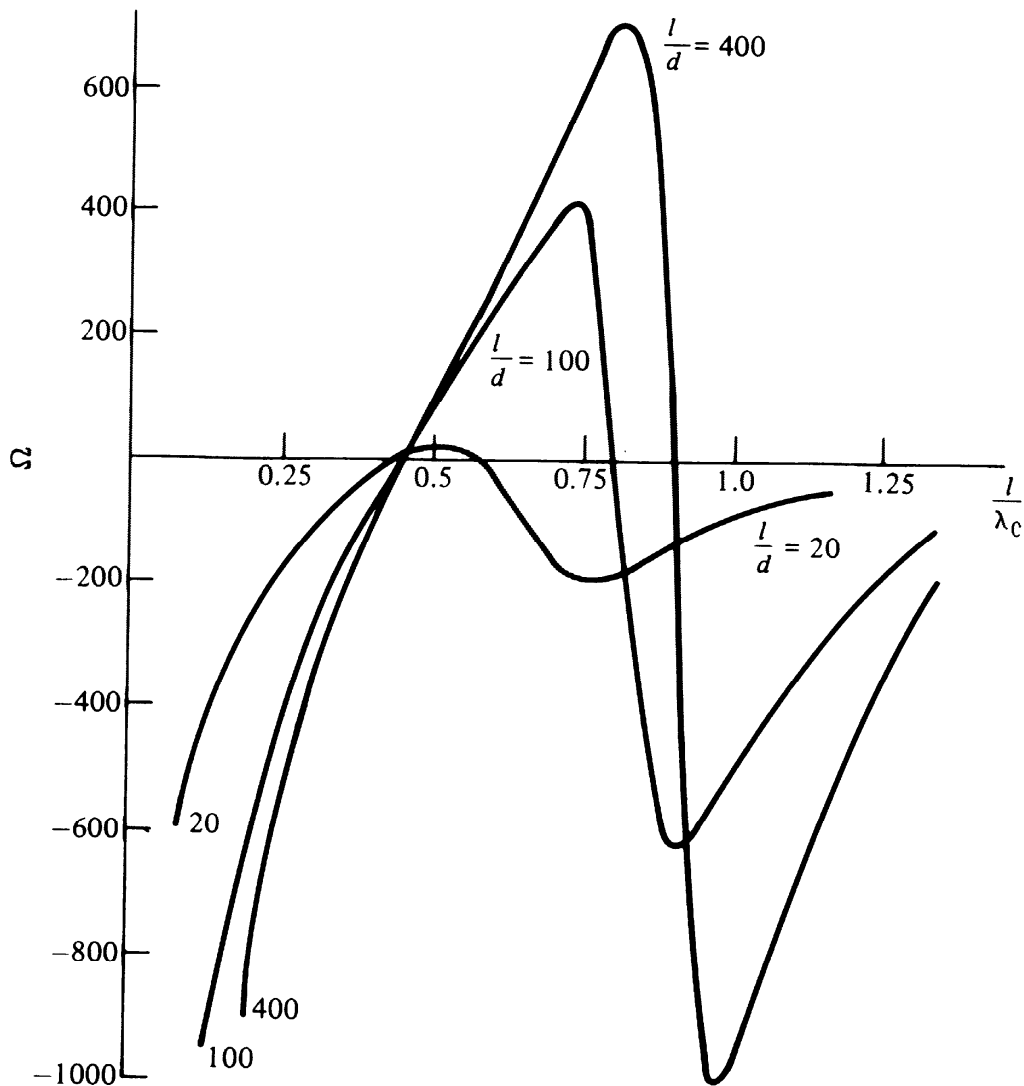
Measurement results for the input impedance of a dipole vs. its electrical length

(a) input resistance



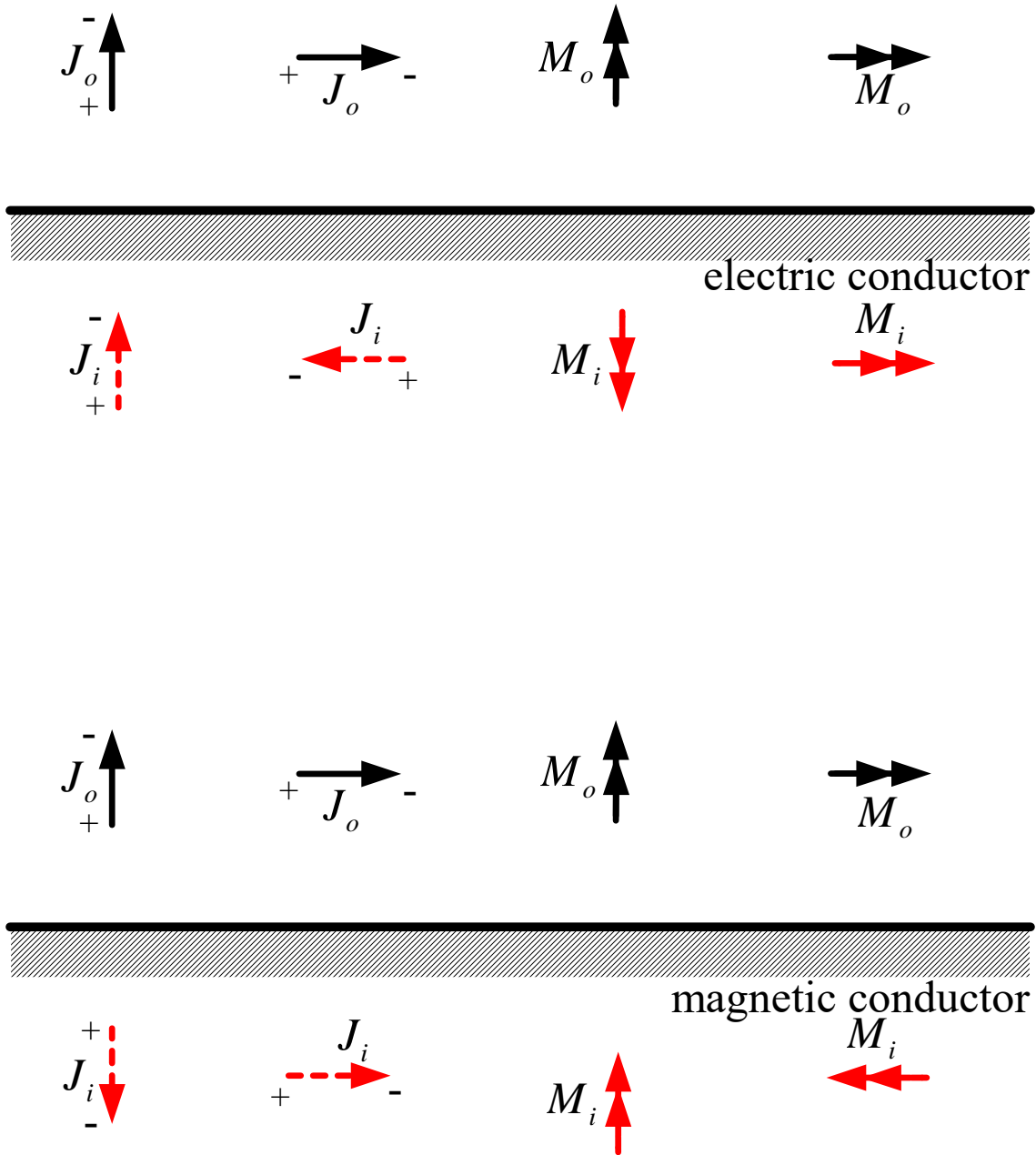
Note the strong influence of the dipole diameter on its resonant properties.

(b) input reactance

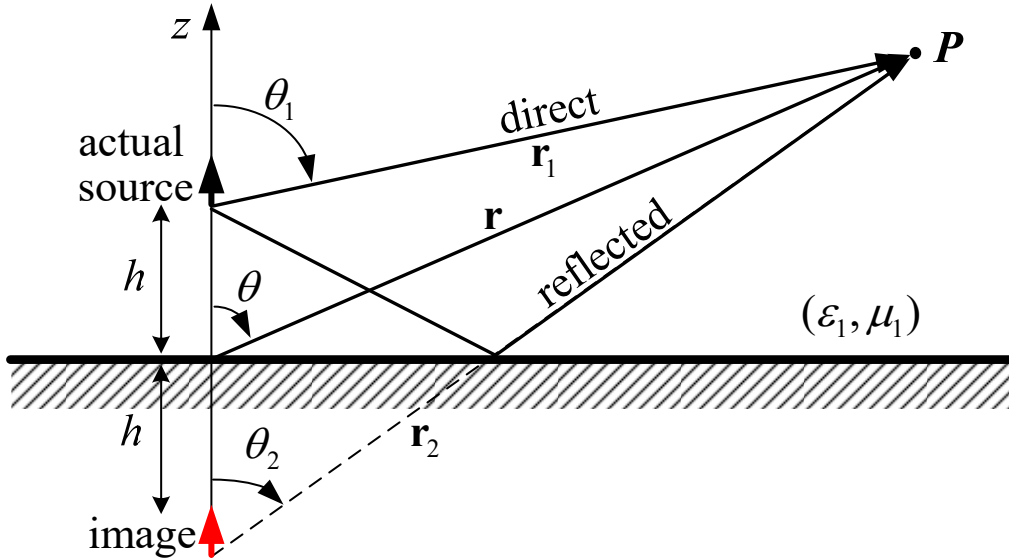


We can calculate the input resistance as a function of l/λ using (9.29) and (9.39). These equations, however, are valid only for infinitesimally thin dipoles. Besides, they do not produce the reactance. In practice, dipoles are most often tubular and they have substantial diameter d . General-purpose numerical methods such as the *method of moments* (MoM) or the *finite-difference time-domain* (FDTD) method can be used to calculate the complex antenna input impedance. When finite-thickness wire antennas are analyzed and no assumption is made for the current distribution along the wire, the MoM is applied to Pocklington's equation or to its variation, the Hallen equation. A classical method producing closed form solutions for the self-impedance and the mutual impedance of straight-wire antennas is the *induced electromotive force (emf) method*, which is discussed later.

4. Method of Images – Revision



5. Vertical Electric Current Element Above Perfect Conductor



The field at the observation point P is a superposition of the fields of the actual source and the image source, both radiating in a homogeneous medium of constitutive parameters (ϵ_1, μ_1) . The actual (or original) source is a current element $I_0 \Delta l$ (infinitesimal dipole). Therefore, the image source is also an infinitesimal dipole. The respective field components are:

$$E_{\theta}^d = j\eta\beta(I_0\Delta l) \frac{e^{-j\beta r_1}}{4\pi r_1} \cdot \sin \theta_1, \quad (9.55)$$

$$E_{\theta}^r = j\eta\beta(I_0\Delta l) \frac{e^{-j\beta r_2}}{4\pi r_2} \cdot \sin \theta_2 .$$

Expressing the distances $r_1 = |\mathbf{r}_1|$ and $r_2 = |\mathbf{r}_2|$ in terms of $r = |\mathbf{r}|$ and h (using the cosine theorem) gives

$$r_1 = \sqrt{r^2 + h^2 - 2rh \cos \theta}, \quad (9.56)$$

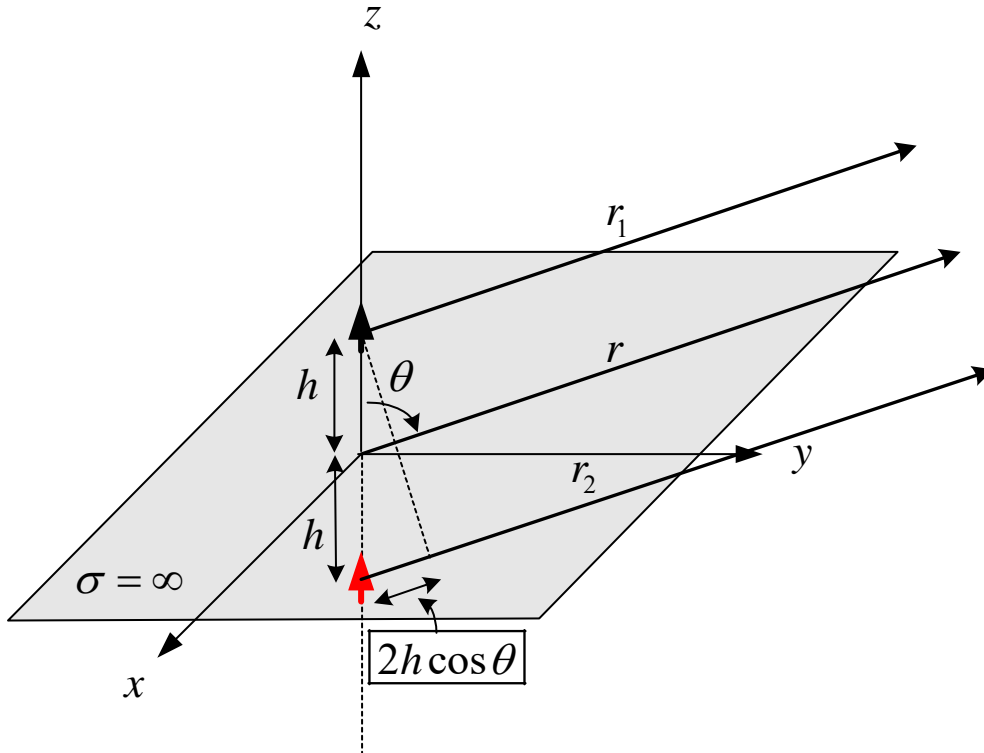
$$r_2 = \sqrt{r^2 + h^2 - 2rh \cos(\pi - \theta)} .$$

We make use of the binomial expansions of r_1 and r_2 to approximate the amplitude and the phase terms, which simplify the evaluation of the total far field and the VP integral. For the amplitude term,

$$\frac{1}{r_1} \approx \frac{1}{r_2} \approx \frac{1}{r}. \quad (9.57)$$

For the phase term, we use the second-order approximation (see also the geometrical interpretation below),

$$\begin{aligned} r_1 &\approx r - h \cos \theta \\ r_2 &\approx r + h \cos \theta. \end{aligned} \quad (9.58)$$



The total far field is

$$E_\theta = E_\theta^d + E_\theta^r \quad (9.59)$$

$$E_\theta = j\eta\beta \frac{(I_0\Delta l)}{4\pi r} \cdot \sin \theta \left[e^{-j\beta(r-h\cos\theta)} + e^{-j\beta(r+h\cos\theta)} \right] \quad (9.60)$$

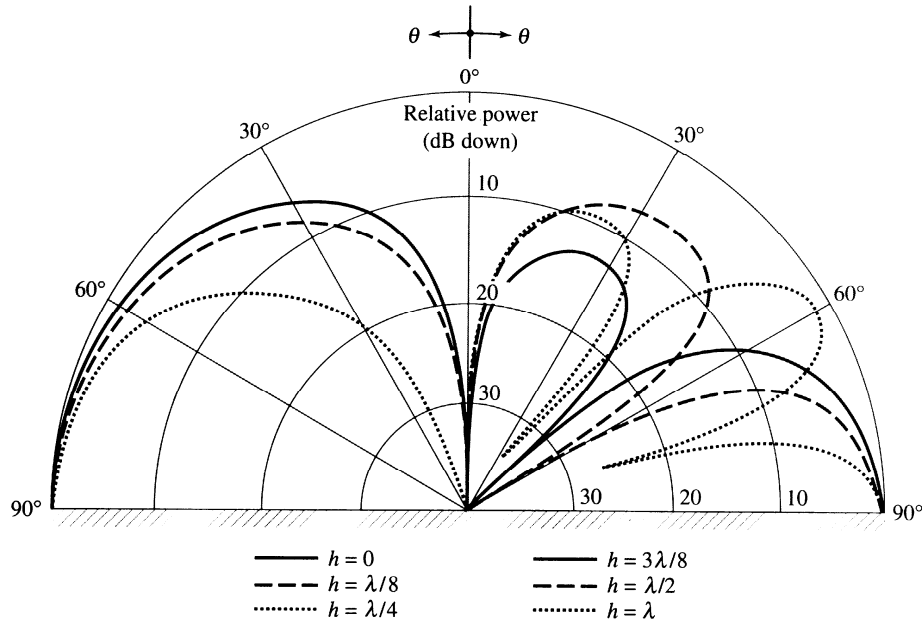
$$\boxed{\begin{aligned} E_\theta &= \underbrace{j\eta\beta (I_0\Delta l) \frac{e^{-j\beta r}}{4\pi r}}_{\text{element factor } g(\theta)} \cdot \underbrace{\sin \theta \cdot [2 \cos(\beta h \cos \theta)]}_{\text{array factor } f(\theta)}, \quad z \geq 0 \\ E_\theta &= 0, \quad z < 0 \end{aligned}} \quad (9.61)$$

Note that the far field can be decomposed into two factors: the field of the elementary source $g(\theta)$ and the pattern factor (also array factor) $f(\theta)$.

The normalized power pattern is

$$F(\theta) = [\sin \theta \cdot \cos(\beta h \cos \theta)]^2. \quad (9.62)$$

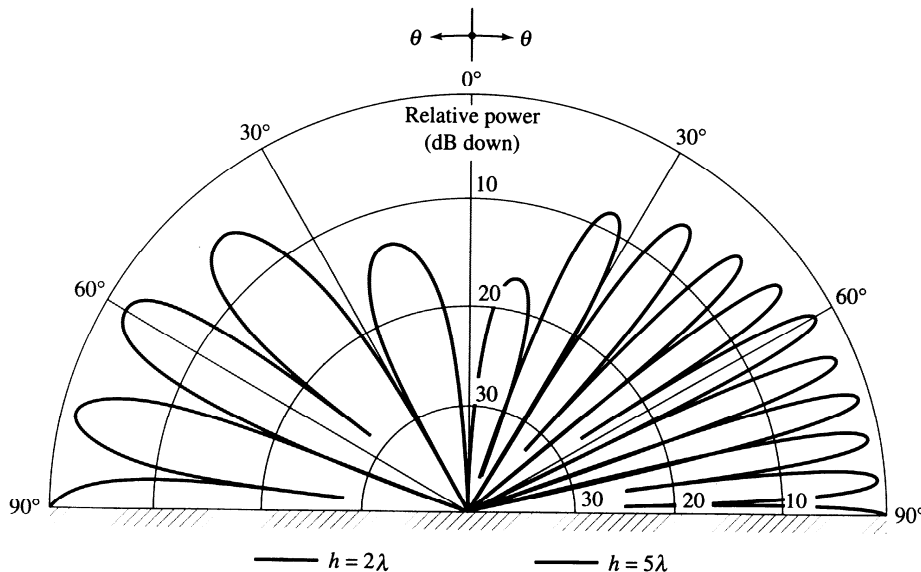
Below, the elevation plane patterns are plotted for vertical infinitesimal electric dipoles of different heights above a perfectly conducting plane:



[Balanis]

As the vertical dipole moves further away from the infinite conducting (ground) plane, more and more lobes are introduced in the power pattern. This effect is called *scalloping* of the pattern. The number of lobes is

$$n = \text{nint}[(2h / \lambda) + 1].$$



Total radiated power

$$\begin{aligned}\Pi &= \oiint \mathbf{P} \cdot d\mathbf{s} = \frac{1}{2\eta} \int_0^{2\pi} \int_0^{\pi/2} |E_\theta|^2 r^2 \sin\theta d\theta d\varphi, \\ \Pi &= \frac{\pi}{\eta} \int_0^{\pi/2} |E_\theta|^2 r^2 \sin\theta d\theta, \end{aligned} \quad (9.63)$$

$$\Pi = \eta\beta^2 (I_0 \Delta l)^2 \int_0^{\pi/2} \sin^2 \theta \cdot \cos^2(\beta h \cos \theta) d\theta,$$

$$\Pi = \pi\eta \left(\frac{I_0 \Delta l}{\lambda} \right)^2 \left[\frac{1}{3} - \frac{\cos(2\beta h)}{(2\beta h)^2} + \frac{\sin(2\beta h)}{(2\beta h)^3} \right]. \quad (9.64)$$

- As $\beta h \rightarrow 0$, the radiated power of the vertical dipole above ground approaches twice the value of the radiated power of a dipole of the same length in free space.
- As $\beta h \rightarrow \infty$, the radiated power of the vertical dipole above ground tends toward that of the vertical dipole in open space.

Note:

$$\lim_{h \rightarrow 0} \left[-\frac{\cos(2\beta h)}{(2\beta h)^2} + \frac{\sin(2\beta h)}{(2\beta h)^3} \right] = \frac{1}{3}, \quad (9.65)$$

$$\lim_{h \rightarrow \infty} \left[-\frac{\cos(2\beta h)}{(2\beta h)^2} + \frac{\sin(2\beta h)}{(2\beta h)^3} \right] = 0. \quad (9.66)$$

Radiation resistance

$$R_r = \frac{2\Pi}{|I_0|^2} = 2\pi\eta \left(\frac{\Delta l}{\lambda} \right)^2 \left[\frac{1}{3} - \frac{\cos(2\beta h)}{(2\beta h)^2} + \frac{\sin(2\beta h)}{(2\beta h)^3} \right]. \quad (9.67)$$

- As $\beta h \rightarrow 0$, the radiation resistance of the vertical dipole above ground approaches twice the value of the radiation resistance of a dipole of the same length in free space:

$$R_{in}^{vdp} = 2R_{in}^{dp}, \quad \beta h = 0. \quad (9.68)$$

- As $\beta h \rightarrow \infty$, the radiation resistance of both dipoles becomes the same.

Radiation intensity

$$U = r^2 P = r^2 \frac{|E_\theta|^2}{2\eta} = \frac{\eta}{2} \left(\frac{I_0 \Delta l}{\lambda} \right)^2 \sin^2 \theta \cos^2(\beta h \cos \theta). \quad (9.69)$$

The maximum of $U(\theta)$ occurs at $\theta = \pi / 2$:

$$U_{\max} = \frac{\eta}{2} \left(\frac{I_0 \Delta l}{\lambda} \right)^2. \quad (9.70)$$

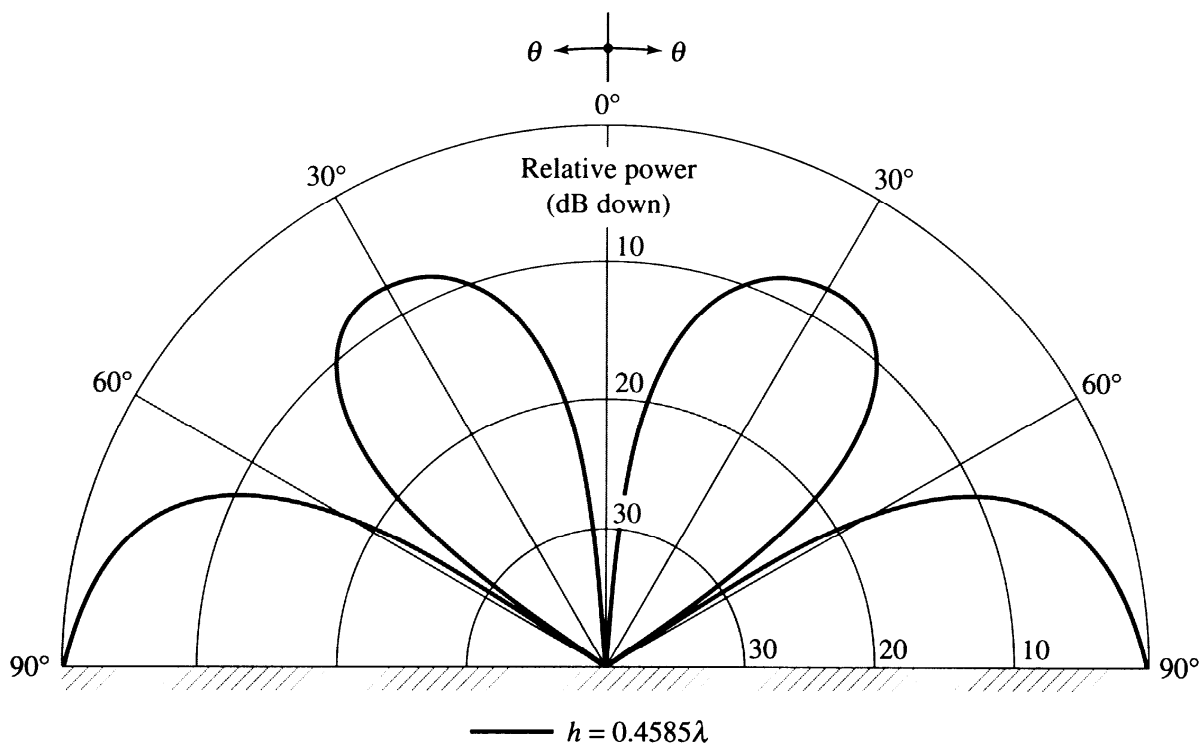
This value is 4 times greater than U_{\max} of a free-space dipole of the same length.

Maximum directivity

$$D_0 = 4\pi \frac{U_{\max}}{\Pi} = \frac{2}{\frac{1}{3} - \frac{\cos(2\beta h)}{(2\beta h)^2} + \frac{\sin(2\beta h)}{(2\beta h)^3}}. \quad (9.71)$$

If $\beta h = 0$, $D_0 = 3$, which is twice the maximum directivity of a free-space current element ($D_0^{id} = 1.5$).

The maximum of D_0 as a function of the height h occurs when $\beta h \approx 2.881$ ($h \approx 0.4585\lambda$). Then, $D_0 \approx 6.566$ for $\beta h = 2.881$.



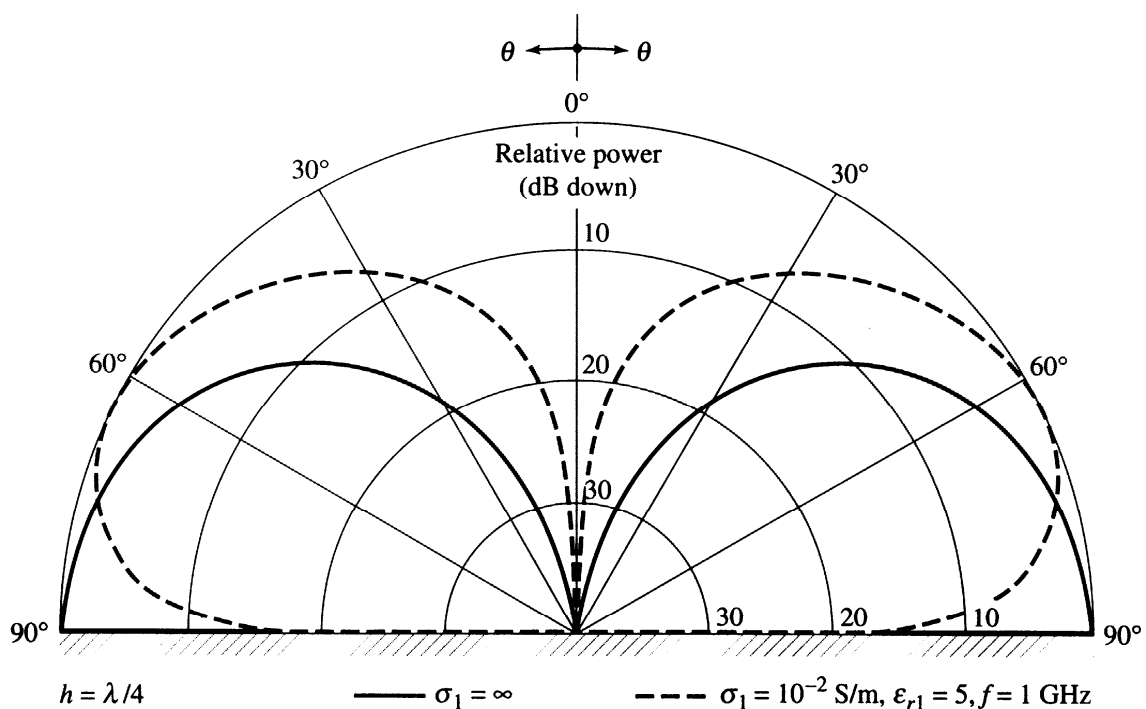
6. Monopoles

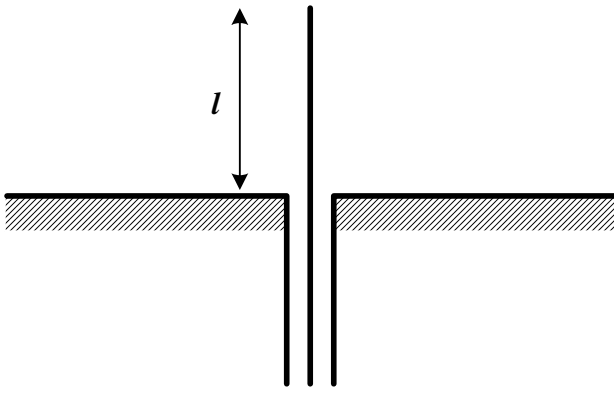
A monopole is a dipole that has been reduced by one-half and is fed against a ground plane. It is normally $\lambda / 4$ long (a *quarter-wavelength monopole*), but it might be shorter if there are space restrictions. Then, the monopole is a *small monopole* the counterpart of which is the *small dipole* (see Section 1). Its current has linear distribution with its maximum at the feed point and its null at the end.

The vertical monopole is a common antenna for AM broadcasting ($f = 500$ to 1500 kHz, $\lambda = 200$ to 600 m), because it is the shortest most efficient antenna at these frequencies as well as because vertically polarized waves suffer less attenuation at close-to-ground propagation. Vertical monopoles are widely used as base-station antennas in mobile communications, too.

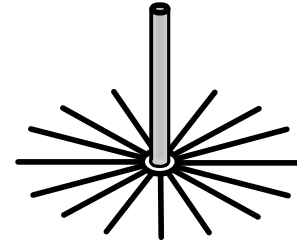
Monopoles at base stations and radio-broadcast stations are supported by towers and guy wires. The guy wires must be separated into short enough ($\leq \lambda / 8$) pieces insulated from each other to suppress parasitic currents.

Special care is taken when grounding the monopole. Usually, multiple radial wires or rods, each $0.25 - 0.35\lambda$ long, are buried at the monopole base in the ground to simulate perfect ground plane, so that the pattern approximates closely the theoretical one, i.e., the pattern of the $\lambda / 2$ -dipole. Losses in the ground plane cause undesirable deformation of the pattern as shown below (infinitesimal dipole above an imperfect ground plane).





Monopole fed against a large solid ground plane



Practical monopole with radial wires to simulate perfect ground

Several important conclusions follow from the image theory and the discussion in Section 5:

- The field distribution in the upper half-space is the same as that of the respective free-space dipole.
- The currents and charges on a monopole are the same as on the upper half of its dipole counterpart but the terminal voltage is only one-half that of the dipole. The input impedance of a monopole is therefore only half that of the respective dipole:

$$Z_{in}^{mp} = \frac{1}{2} Z_{in}^{dp}. \quad (9.72)$$

- The radiation pattern of a monopole is one-half the dipole's pattern since it radiates in half-space and, at the same time, the field normalized distribution in this half-space is the same as that of the dipole. As a result, the beam solid angle of the monopole is half that of the respective dipole and its directivity is twice that of the dipole:

$$D_0^{mp} = \frac{4\pi}{\Omega_A^{mp}} = \frac{4\pi}{0.5\Omega_A^{dp}} = 2D_0^{dp}. \quad (9.73)$$

The quarter-wavelength monopole

This is a straight wire of length $l = \lambda / 4$ mounted over a ground plane. From the discussion above, it follows that the quarter-wavelength monopole is the counterpart of the half-wavelength dipole as far as the radiation in the hemisphere above the ground plane is concerned.

- Its radiation pattern is the same as that of a free-space $\lambda / 2$ -dipole, but it is non-zero only for $0^\circ \leq \theta \leq 90^\circ$ (above ground).
- The field expressions are the same as those of the $\lambda / 2$ -dipole.
- The total radiated power of the $\lambda / 4$ -monopole is half that of the $\lambda / 2$ -dipole.
- The radiation resistance of the $\lambda / 4$ -monopole is half that of the $\lambda / 2$ -dipole: $Z_{in}^{mp} = 0.5Z_{in}^{dp} = 0.5(73 + j42.5) = 36.5 + j21.25, \Omega$.
- The directivity of the $\lambda / 4$ -monopole is

$$D_0^{mp} = 2D_0^{dp} = 2 \cdot 1.643 = 3.286.$$

Some approximate formulas for rapid calculations of the input resistance of a dipole and the respective monopole:

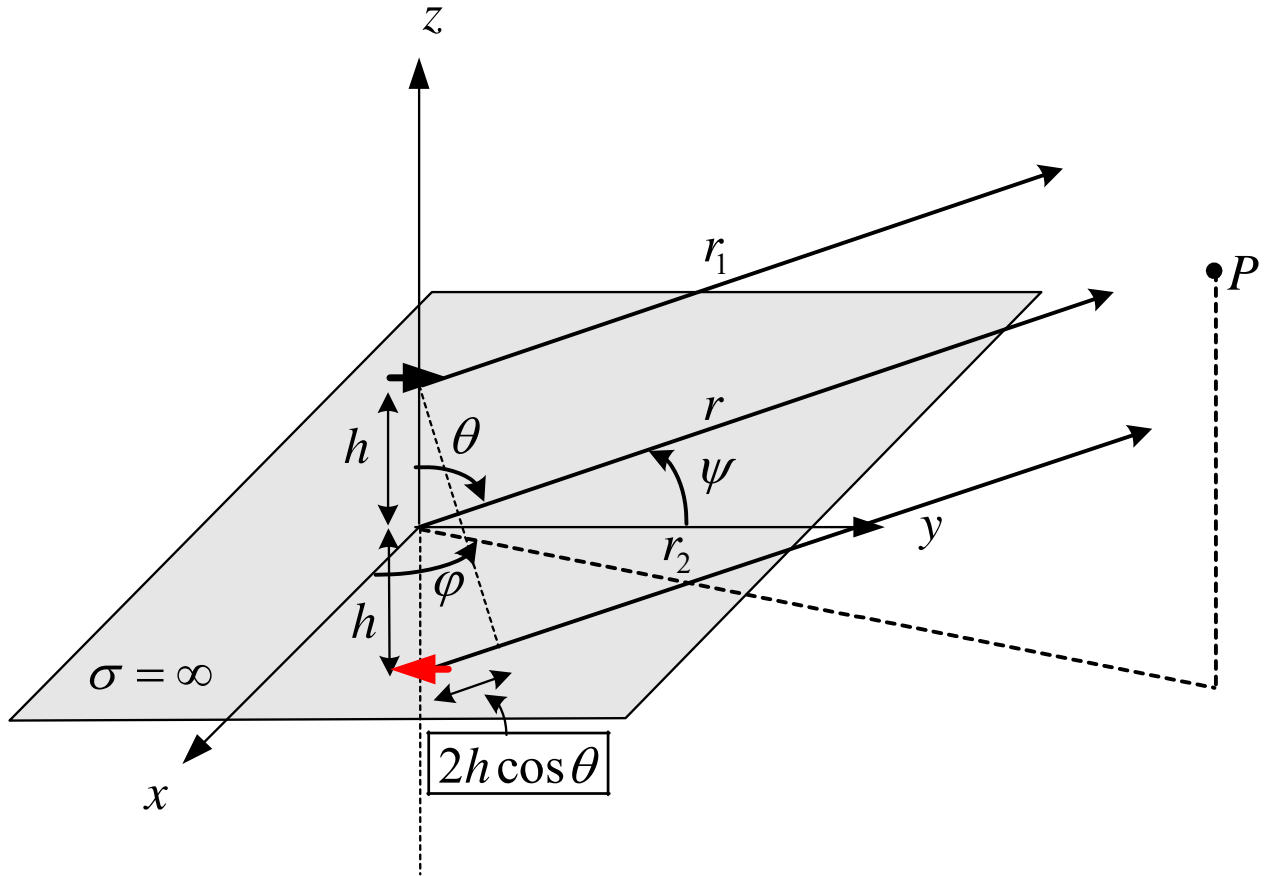
$$\text{Let } \left\{ \begin{array}{l} G = \frac{\beta l}{2} = \pi \frac{l}{\lambda}, \text{ for dipole} \\ G = \beta l = 2\pi \frac{l}{\lambda}, \text{ for monopole.} \end{array} \right.$$

Approximate formulas:

- If $0 < G < \frac{\pi}{4}$, then $\left\{ \begin{array}{l} R_{in} = 20G^2, \text{ dipole} \\ R_{in} = 10G^2, \text{ monopole} \end{array} \right.$
- If $\frac{\pi}{4} < G < \frac{\pi}{2}$, then $\left\{ \begin{array}{l} R_{in} = 24.7G^{2.5}, \text{ dipole} \\ R_{in} = 12.35G^{2.5}, \text{ monopole} \end{array} \right.$
- If $\frac{\pi}{2} < G < 2$, then $\left\{ \begin{array}{l} R_{in} = 11.14G^{4.17}, \text{ dipole} \\ R_{in} = 5.57G^{4.17}, \text{ monopole} \end{array} \right.$

7. Horizontal Current Element Above a Perfectly Conducting Plane

The analysis is analogous to that of a vertical current element above a ground plane. The difference arises in the element factor $g(\theta)$ because of the horizontal orientation of the current element. Let us assume that the current element is oriented along the y -axis, and the angle between \mathbf{r} and the dipole's axis (y -axis) is ψ .



$$\mathbf{E}(P) = \mathbf{E}^d(P) + \mathbf{E}^r(P), \quad (9.74)$$

$$E_{\psi}^d = j\eta\beta(I_0\Delta l) \frac{e^{-j\beta r_1}}{4\pi r_1} \sin\psi, \quad (9.75)$$

$$E_{\psi}^r = -j\eta\beta(I_0\Delta l) \frac{e^{-j\beta r_2}}{4\pi r_2} \sin\psi. \quad (9.76)$$

We can express the angle ψ in terms of (θ, φ) :

$$\cos\psi = \hat{\mathbf{y}} \cdot \hat{\mathbf{r}} = \hat{\mathbf{y}} \cdot (\hat{\mathbf{x}} \sin\theta \cos\varphi + \hat{\mathbf{y}} \sin\theta \sin\varphi + \hat{\mathbf{z}} \cos\theta)$$

$$\begin{aligned} \Rightarrow \cos \psi &= \sin \theta \sin \varphi \\ \Rightarrow \sin \psi &= \sqrt{1 - \sin^2 \theta \sin^2 \varphi}. \end{aligned} \quad (9.77)$$

The far-field approximations are:

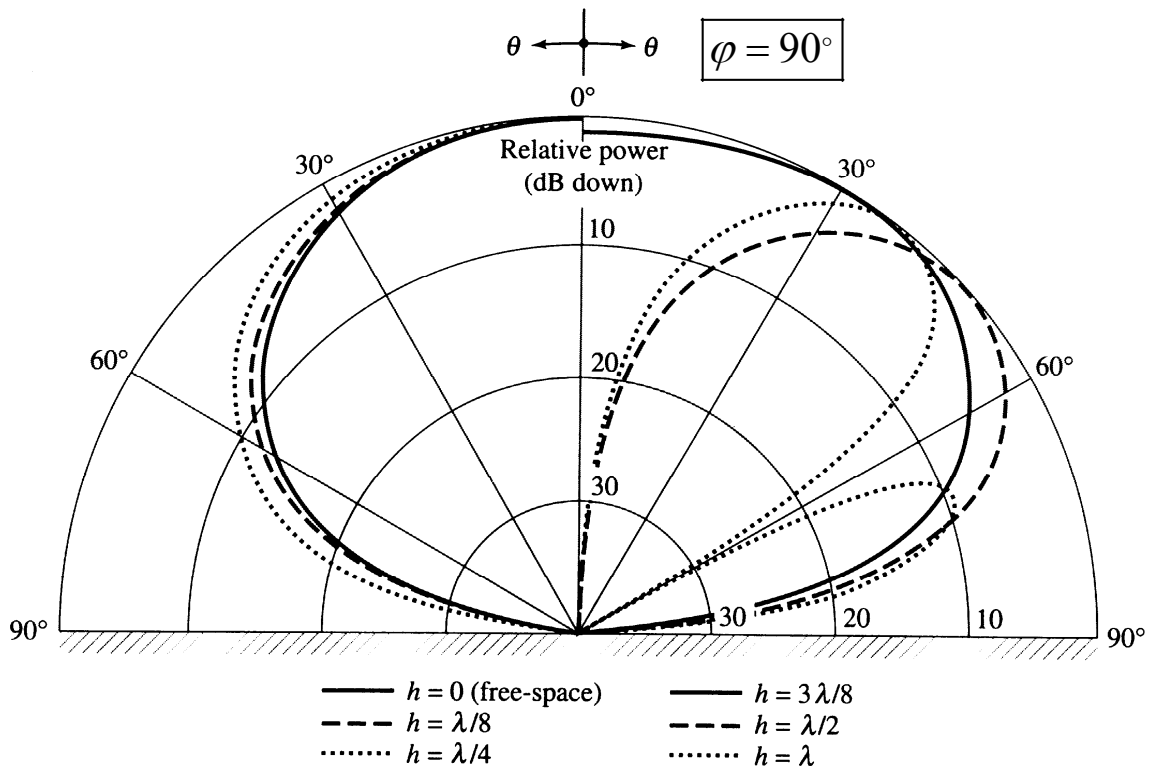
$$\left. \begin{aligned} \frac{1}{r_1} &= \frac{1}{r_2} = \frac{1}{r}, \text{ for the amplitude term} \\ r_1 &\approx r - h \cos \theta \\ r_2 &\approx r + h \cos \theta \end{aligned} \right\} \text{ for the phase term.}$$

The substitution of the far-field approximations and equations (9.75), (9.76), (9.77) into the total field expression (9.74) yields

$$E_{\psi}(\theta, \varphi) = \underbrace{j\eta\beta(I_0\Delta l) \frac{e^{-j\beta r}}{4\pi r}}_{\text{element factor } g(\theta, \varphi)} \sqrt{1 - \sin^2 \theta \sin^2 \varphi} \cdot \underbrace{[2j \sin(\beta h \cos \theta)]}_{\text{array factor } f(\theta, \varphi)}. \quad (9.78)$$

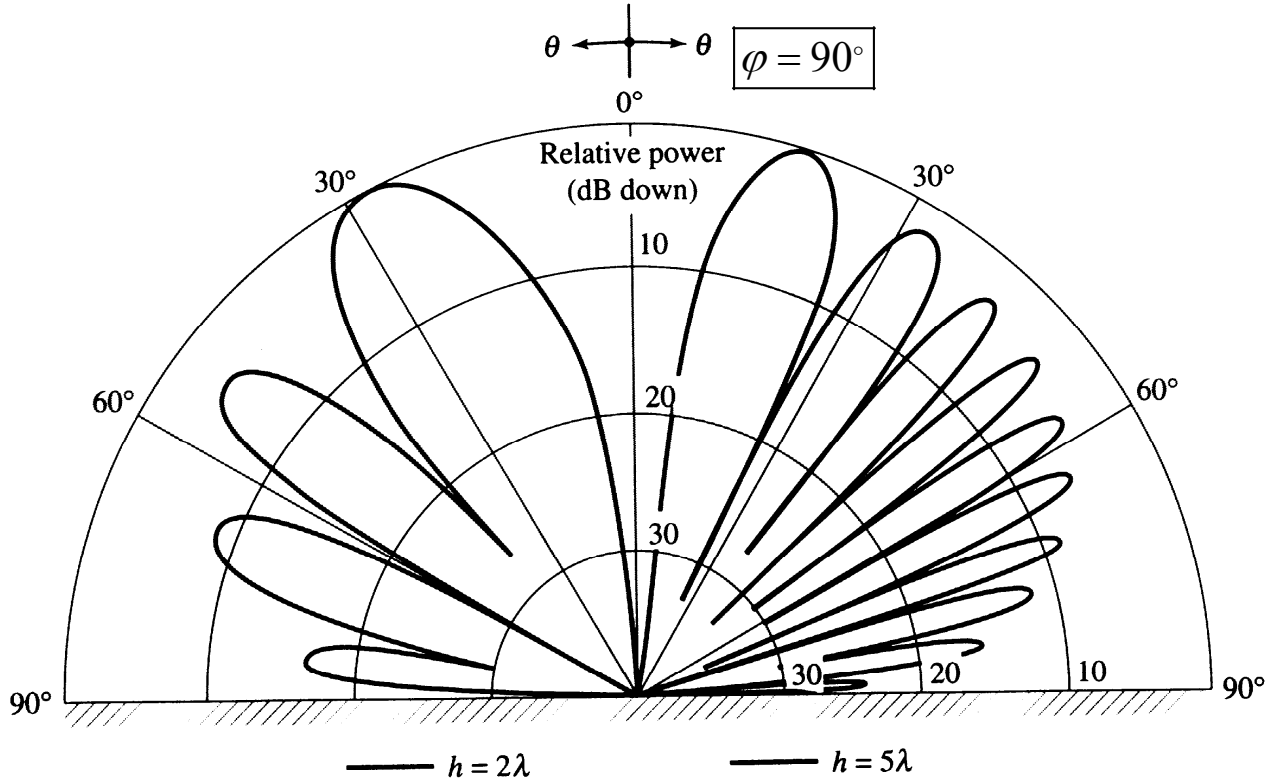
The normalized power pattern

$$F(\theta, \varphi) = (1 - \sin^2 \theta \cdot \sin^2 \varphi) \cdot \sin^2(\beta h \cos \theta) \quad (9.79)$$



As the height increases beyond a wavelength ($h > \lambda$), scalloping appears with the number of lobes being

$$n = \text{nint}\left(2\frac{h}{\lambda}\right). \quad (9.80)$$



Following a procedure similar to that of the vertical dipole, the radiated power and the radiation resistance of the horizontal dipole can be found:

$$\Pi = \frac{\pi}{2} \eta \left(\frac{I_0 \Delta l}{\lambda} \right)^2 \underbrace{\left[\frac{2}{3} - \frac{\sin(2\beta h)}{2\beta h} - \frac{\cos(2\beta h)}{(2\beta h)^2} + \frac{\sin(2\beta h)}{(2\beta h)^3} \right]}_{R(\beta h)} \quad (9.81)$$

$$R_r = \pi \eta \left(\frac{\Delta l}{\lambda} \right)^2 \cdot R(\beta h). \quad (9.82)$$

By expanding the sine and the cosine functions into series, it can be shown that for small values of (βh) the following approximation holds:

$$R_{r/\beta h \rightarrow 0} \approx \frac{32\pi^2}{15} \left(\frac{h}{\lambda} \right)^2. \quad (9.83)$$

It is also obvious that if $h = 0$, then $R_r = 0$ and $\Pi = 0$. This is to be expected because the dipole is short-circuited by the ground plane.

Radiation intensity

$$U = \frac{r^2}{2\eta} |\mathbf{E}_\psi|^2 = \frac{\eta}{2} \left(\frac{I_0 \Delta l}{\lambda} \right)^2 (1 - \sin^2 \theta \cdot \sin^2 \varphi) \cdot \sin^2(\beta h \cos \theta) \quad (9.84)$$

The maximum value of (9.84) depends on whether (βh) is less than $\pi/2$ or greater:

- If $\beta h \leq \frac{\pi}{2}$ $\left(h \leq \frac{\lambda}{4} \right)$

$$U_{\max} = \frac{\eta}{2} \left(\frac{I_0 \Delta l}{\lambda} \right)^2 \sin^2(\beta h)_{/\theta=0^\circ}. \quad (9.85)$$

- If $\beta h > \frac{\pi}{2}$ $\left(h > \frac{\lambda}{4} \right)$

$$U_{\max} = \frac{\eta}{2} \left(\frac{I_0 \Delta l}{\lambda} \right)^2_{/\theta=\arccos\left(\frac{\pi}{2\beta h}\right), \varphi=0^\circ}. \quad (9.86)$$

Maximum directivity

- If $h \leq \frac{\lambda}{4}$, then U_{\max} is obtained from (9.85) and the directivity is

$$D_0 = 4\pi \frac{U_{\max}}{\Pi} = \frac{4 \sin^2(\beta h)}{R(\beta h)}. \quad (9.87)$$

- If $h > \frac{\lambda}{4}$, then U_{\max} is obtained from (9.86) and the directivity is

$$D_0 = 4\pi \frac{U_{\max}}{\Pi} = \frac{4}{R(\beta h)}. \quad (9.88)$$

For very small βh , the approximation $D_0 \approx 7.5 \left(\frac{\sin(\beta h)}{\beta h} \right)^2$ is often used.

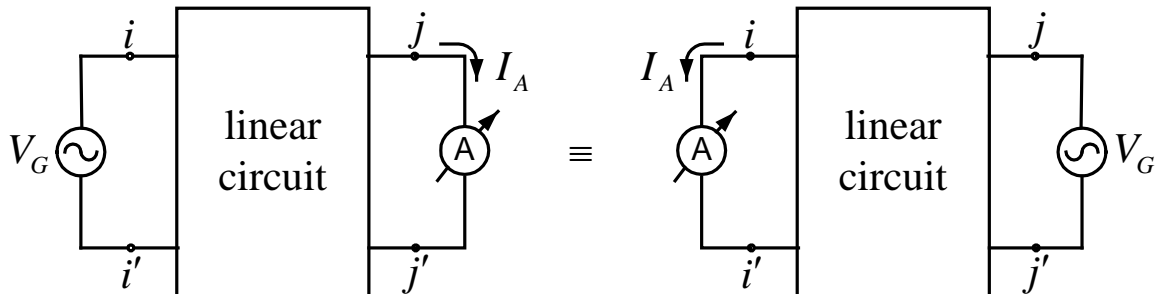
LECTURE 10: Reciprocity. Cylindrical Antennas – Analytical Models

(Reciprocity theorem. Implications of reciprocity in antenna measurements. Self-impedance of a dipole using the induced *emf* method. Pocklington's equation. Hallén's equation.)

1. Reciprocity Theorem for Antennas

1.1. Reciprocity theorem in circuit theory

If a voltage (current) generator is placed between any pair of nodes of a linear circuit, and a current (voltage) response is measured between any other pair of nodes, the interchange of the generator's and the measurement's locations would lead to the same measurement results.



$$\frac{V_i}{I_j} = \frac{V_j}{I_i} \quad \text{or} \quad Z_{ji} = Z_{ij}. \quad (10.1)$$

1.2. Reciprocity theorem in EM field theory (Lorentz' reciprocity theorem)

Consider a volume $V_{[S]}$ bounded by the surface S , where two pairs of sources exist: $(\mathbf{J}_1, \mathbf{M}_1)$ and $(\mathbf{J}_2, \mathbf{M}_2)$. The medium is linear. We denote the field associated with the $(\mathbf{J}_1, \mathbf{M}_1)$ sources as $(\mathbf{E}_1, \mathbf{H}_1)$, and the field generated by $(\mathbf{J}_2, \mathbf{M}_2)$ as $(\mathbf{E}_2, \mathbf{H}_2)$.

$$\begin{cases} \nabla \times \mathbf{E}_1 = -j\omega\mu\mathbf{H}_1 - \mathbf{M}_1 & / \cdot \mathbf{H}_2 \\ \nabla \times \mathbf{H}_1 = j\omega\varepsilon\mathbf{E}_1 + \mathbf{J}_1 & / \cdot \mathbf{E}_2 \end{cases} \quad (10.2)$$

$$\begin{cases} \nabla \times \mathbf{E}_2 = -j\omega\mu\mathbf{H}_2 - \mathbf{M}_2 & / \cdot \mathbf{H}_1 \\ \nabla \times \mathbf{H}_2 = j\omega\varepsilon\mathbf{E}_2 + \mathbf{J}_2 & / \cdot \mathbf{E}_1 \end{cases} \quad (10.3)$$

The vector identity

$\mathbf{H}_2 \cdot \nabla \times \mathbf{E}_1 - \mathbf{E}_1 \cdot \nabla \times \mathbf{H}_2 - \mathbf{H}_1 \cdot \nabla \times \mathbf{E}_2 + \mathbf{E}_2 \cdot \nabla \times \mathbf{H}_1 = \nabla \cdot (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1)$
is used along with (10.2) and (10.3) to obtain

$$\nabla \cdot (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1) = -\mathbf{E}_1 \cdot \mathbf{J}_2 + \mathbf{H}_1 \cdot \mathbf{M}_2 + \mathbf{E}_2 \cdot \mathbf{J}_1 - \mathbf{H}_2 \cdot \mathbf{M}_1. \quad (10.4)$$

Equation (10.4) is written in its integral form as

$$\oiint_S (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1) \cdot d\mathbf{s} = \iiint_{V_{[S]}} (-\mathbf{E}_1 \cdot \mathbf{J}_2 + \mathbf{H}_1 \cdot \mathbf{M}_2 + \mathbf{E}_2 \cdot \mathbf{J}_1 - \mathbf{H}_2 \cdot \mathbf{M}_1) dv. \quad (10.5)$$

Equations (10.4) and (10.5) represent the general ***Lorentz reciprocity theorem in differential and integral forms***, respectively.

One special case of the reciprocity theorem is of fundamental importance to antenna theory, namely, its application to unbounded (open) problems. In this case, the surface S is a sphere of infinite radius. Therefore, the fields integrated over it are far-zone fields. This means that the left-hand side of (10.5) vanishes:

$$\oiint_S \left(\frac{|\mathbf{E}_1| |\mathbf{E}_2|}{\eta} \cos \gamma - \frac{|\mathbf{E}_1| |\mathbf{E}_2|}{\eta} \cos \gamma \right) ds = 0. \quad (10.6)$$

Here, γ is the angle between the polarization vectors of both fields, \mathbf{E}_1 and \mathbf{E}_2 . Note that in the far zone, the field vectors are orthogonal to the direction of propagation and, therefore are orthogonal to $d\mathbf{s}$. Thus, in the case of open problems, the reciprocity theorem reduces to

$$\iiint_{V_{[S]}} (\mathbf{E}_1 \cdot \mathbf{J}_2 - \mathbf{H}_1 \cdot \mathbf{M}_2) dv = \iiint_{V_{[S]}} (\mathbf{E}_2 \cdot \mathbf{J}_1 - \mathbf{H}_2 \cdot \mathbf{M}_1) dv. \quad (10.7)$$

Each of the integrals in (10.7) can be interpreted as ***coupling energy*** between the field produced by some sources and another set of sources generating another field. The quantity

$$\langle 1, 2 \rangle = \iiint_{V_{[S]}} (\mathbf{E}_1 \cdot \mathbf{J}_2 - \mathbf{H}_1 \cdot \mathbf{M}_2) dv$$

is called the ***reaction*** of the field $(\mathbf{E}_1, \mathbf{H}_1)$ to the sources $(\mathbf{J}_2, \mathbf{M}_2)$. Similarly,

$$\langle 2, 1 \rangle = \iiint_{V_{[S]}} (\mathbf{E}_2 \cdot \mathbf{J}_1 - \mathbf{H}_2 \cdot \mathbf{M}_1) dv$$

is the *reaction* of the field $(\mathbf{E}_2, \mathbf{H}_2)$ to the sources $(\mathbf{J}_1, \mathbf{M}_1)$. Thus, in a shorthand notation, the reciprocity equation (10.7) is $\langle 1, 2 \rangle = \langle 2, 1 \rangle$.

The Lorentz reciprocity theorem is the most general form of reciprocity in linear EM systems. Circuit reciprocity is a special case of lumped element sources and responses (local voltage or current measurements).

To illustrate the above statement, consider the following scenario. Assume that the sources in two measurements have identical amplitude-phase distributions in their respective volumes: \mathbf{J}_1 and \mathbf{M}_1 reside in V_1 whereas \mathbf{J}_2 and \mathbf{M}_2 reside in V_2 . Note that the volumes V_1 and V_2 may or may not overlap. We can associate a local coordinate system with each source volume where the position is given by $\mathbf{x}_i = (r_i, \theta_i, \varphi_i)$, $i = 1, 2$. If the sources have identical distributions, then the source volumes are the same in shape and size, $V_1 = V_2 = V_s$, and $\mathbf{J}_1(\mathbf{x}_1) = \mathbf{J}_2(\mathbf{x}_2) = \mathbf{J}$, and $\mathbf{M}_1(\mathbf{x}_1) = \mathbf{M}_2(\mathbf{x}_2) = \mathbf{M}$. Then, according to (10.7),

$$\iiint_{V_s} (\mathbf{E}_1 \cdot \mathbf{J} - \mathbf{H}_1 \cdot \mathbf{M}) dv_2 = \iiint_{V_s} (\mathbf{E}_2 \cdot \mathbf{J} - \mathbf{H}_2 \cdot \mathbf{M}) dv_1. \quad (10.8)$$

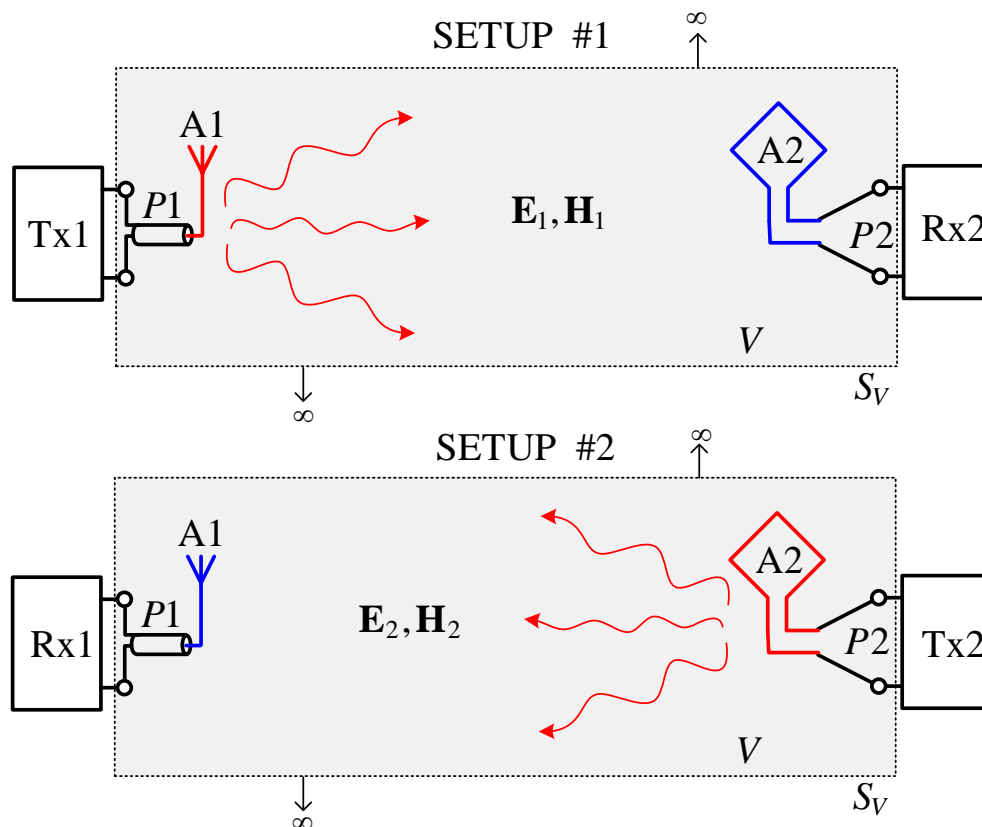
It follows that $\mathbf{E}_1 = \mathbf{E}_2$ and $\mathbf{H}_1 = \mathbf{H}_2$. Here, \mathbf{E}_1 and \mathbf{H}_1 describe the observed field in V_2 (the volume where the sources \mathbf{J}_2 and \mathbf{M}_2 reside but are inactive), this field being due to the sources \mathbf{J}_1 and \mathbf{M}_1 (in V_1) which are active. Conversely, \mathbf{E}_2 and \mathbf{H}_2 describe the observed field in V_1 (the volume where \mathbf{J}_1 and \mathbf{M}_1 reside but are inactive), that field being due to the sources \mathbf{J}_2 and \mathbf{M}_2 (in V_2) which are active. These are two measurement scenarios which differ only in the interchanged locations of the source and the observation: in the former scenario, the observation is in V_2 whereas the source is in V_1 ; in the latter scenario, the observation is in V_1 whereas the source is in V_2 . The field equality, $\mathbf{E}_1 = \mathbf{E}_2$ and $\mathbf{H}_1 = \mathbf{H}_2$, tells us that *interchanging the locations of excitation and observation leaves the observed field unchanged*. This result is general in the sense that it holds in a heterogeneous medium. This is essentially the same principle that is postulated as reciprocity in circuit theory (see Section 1.1). Only that Lorentz' EM reciprocity considers volumes instead of nodes and branches, and field vectors instead of voltages and currents.

The reciprocity theorem can be postulated also as: *any network constructed of linear isotropic matter has a symmetrical impedance matrix*. This “network” can be two antennas and the space between them.

1.3. Implications of reciprocity for the received-to-transmitted power ratio

Using the reciprocity theorem, we next prove that *the ratio of received to transmitted power P_r / P_t does not depend on whether antenna #1 transmits and antenna #2 receives or vice versa*. We should reiterate that the reciprocity theorem holds only if the whole system (antennas + propagation environment) is isotropic and linear.

In this case, we view the two antenna system as a two-port microwave network; see the figure below. Port 1 (P1) connects to antenna 1 (A1) while port 2 (P2) is at the terminals of antenna 2 (A2). Depending on whether an antenna transmits or receives, its terminals are connected to a transmitter (Tx) or a receiver (Rx), respectively. We consider two measurement setups. In Setup #1, A1 transmits and A2 receives whereas in Setup #2 A1 receives and A2 transmits.



The volume V in both setups excludes the power sources in the respective transmitters and, therefore, it does not have impressed currents sources, i.e., $\mathbf{J}_1 = \mathbf{J}_2 = 0$ and $\mathbf{M}_1 = \mathbf{M}_2 = 0$. The reciprocity integral (10.5) becomes

$$\oiint_{S_V} (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1) \cdot d\mathbf{s} = 0 \quad (10.9)$$

where the subscripts refer to the measurement setups. Part of the surface S_V extends to infinity away from the antennas (top and bottom lines in the plots above) but it also crosses through P1 and P2. At infinity, the surface integration in (10.9) produces zero; however, at the cross-sections S_1 and S_2 of ports 1 and 2, respectively, the contributions are not zero. Then,

$$\iint_{S_1} (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1) \cdot d\mathbf{s} + \iint_{S_2} (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1) \cdot d\mathbf{s} = 0. \quad (10.10)$$

Let us now assume that the transmit power in both setups is set to 1 W. This is not going to affect the generality of the final result. Let us denote the field vectors in the transmission lines of ports 1 and 2 corresponding to 1-W transferred power as $(\mathbf{e}_{P1}, \mathbf{h}_{P1})$ and $(\mathbf{e}_{P2}, \mathbf{h}_{P2})$, respectively.¹ We assume that these vectors correspond to power transfer *from* the antenna (out of V and toward the Tx or the Rx circuit). When the power is transferred toward the antenna (from the Tx or Rx circuit), due to the opposite direction of propagation, we have to change the sign of either the \mathbf{e} or the \mathbf{h} vector (but not both!) in the respective pair.

At P1, in Setup #1, the incident field is the 1-W field generated by Tx1, which is $(\mathbf{e}_{P1}, -\mathbf{h}_{P1})$. There could be a reflected field due to impedance mismatch at the A1 terminals, which can be expressed as $\Gamma_1(\mathbf{e}_{P1}, \mathbf{h}_{P1})$ where Γ_1 is the reflection coefficient at P1. At P1, in Setup #2, there is the field $(\mathbf{E}_2, \mathbf{H}_2)$ due to the radiation from A2.

Analogous field components can be identified at P2 in both setups: (i) $(\mathbf{e}_{P2}, -\mathbf{h}_{P2})$ is the field when in Setup #2 the Tx at A2 provides 1 W of power, (ii) $(\mathbf{E}_1, \mathbf{H}_1)$ is the field at P2 in Setup #1. Equation (10.10) now becomes

$$\begin{aligned} & \iint_{S_1} [(\mathbf{e}_{P1} \times \mathbf{H}_2 + \mathbf{E}_2 \times \mathbf{h}_{P1}) + \Gamma_1(\mathbf{e}_{P1} \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{h}_{P1})] \cdot d\mathbf{s} + \\ & \iint_{S_2} [(-\mathbf{E}_1 \times \mathbf{h}_{P2} - \mathbf{e}_{P2} \times \mathbf{H}_1) + \Gamma_2(\mathbf{E}_1 \times \mathbf{h}_{P2} - \mathbf{e}_{P2} \times \mathbf{H}_1)] \cdot d\mathbf{s} = 0. \end{aligned} \quad (10.11)$$

¹ It can be shown that a propagating mode in a transmission line or a waveguide can be represented by real-valued phasor vectors \mathbf{e} and \mathbf{h} .

Next, the received fields in both scenarios, $(\mathbf{E}_1, \mathbf{H}_1)$ at P2 and $(\mathbf{E}_2, \mathbf{H}_2)$ at P1, can be expressed in terms of $(\mathbf{e}_{P2}, \mathbf{h}_{P2})$ and $(\mathbf{e}_{P1}, \mathbf{h}_{P1})$, which represent 1-W received powers at the respective ports:

$$\begin{aligned} (\mathbf{E}_2, \mathbf{H}_2)|_{P1} &= R_{1,2}(\mathbf{e}_{P1}, \mathbf{h}_{P1}) \\ (\mathbf{E}_1, \mathbf{H}_1)|_{P2} &= R_{2,1}(\mathbf{e}_{P2}, \mathbf{h}_{P2}). \end{aligned} \quad (10.12)$$

Note that (10.12) implies that the respective received-to-transmitted power ratios in Setup #1 and Setup #2 are

$$P_{r1} / P_{t1} = R_{2,1}^2 \quad (10.13)$$

$$P_{r2} / P_{t2} = R_{1,2}^2. \quad (10.14)$$

Substituting (10.12) into (10.11) leads to

$$\begin{aligned} R_{1,2} \iint_{S_1} [(\underbrace{\mathbf{e}_{P1} \times \mathbf{h}_{P1} + \mathbf{e}_{P1} \times \mathbf{h}_{P1}}_{=4}) + \Gamma_1(\underbrace{\mathbf{e}_{P1} \times \mathbf{h}_{P1} - \mathbf{e}_{P1} \times \mathbf{h}_{P1}}_{=0})] \cdot d\mathbf{s} + \\ R_{2,1} \iint_{S_2} [(\underbrace{-\mathbf{e}_{P2} \times \mathbf{h}_{P2} - \mathbf{e}_{P2} \times \mathbf{h}_{P2}}_{=-4}) + \Gamma_2(\underbrace{\mathbf{e}_{P2} \times \mathbf{h}_{P2} - \mathbf{e}_{P2} \times \mathbf{h}_{P2}}_{=0})] \cdot d\mathbf{s} = 0. \end{aligned} \quad (10.15)$$

Since the fields $(\mathbf{e}_{Pn}, \mathbf{h}_{Pn})$, $n = 1, 2$, correspond to 1 W of transferred power, their respective integrals over the port cross-sections (integration over the Poynting vector) have the same value:

$$\frac{1}{2} \iint_{S_n} (\mathbf{e}_{Pn} \times \mathbf{h}_{Pn}) \cdot d\mathbf{s} = 1, \quad n = 1, 2. \quad (10.16)$$

Note that here we have assumed that the fields $(\mathbf{e}_{Pn}, \mathbf{h}_{Pn})$, $n = 1, 2$, are “magnitude” (not RMS) phasors. It follows from (10.15) and (10.16) that

$$R_{1,2} = R_{2,1}. \quad (10.17)$$

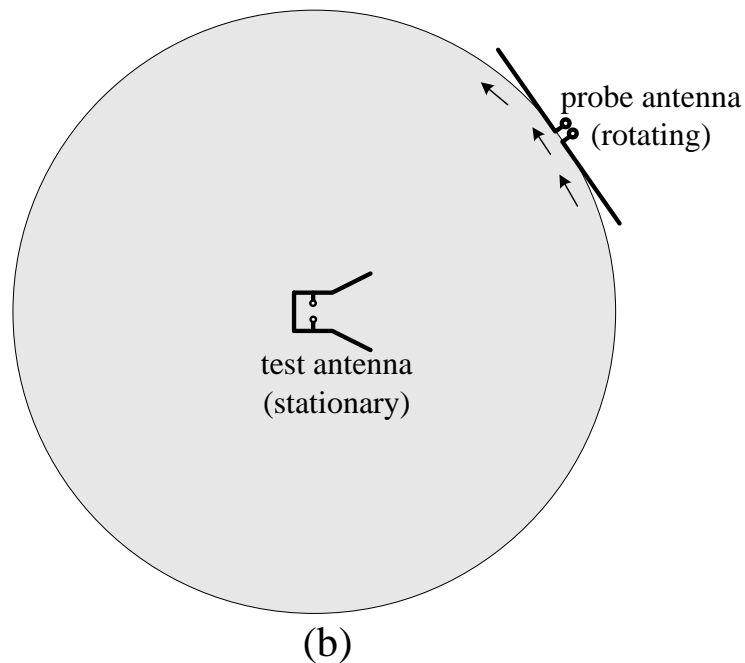
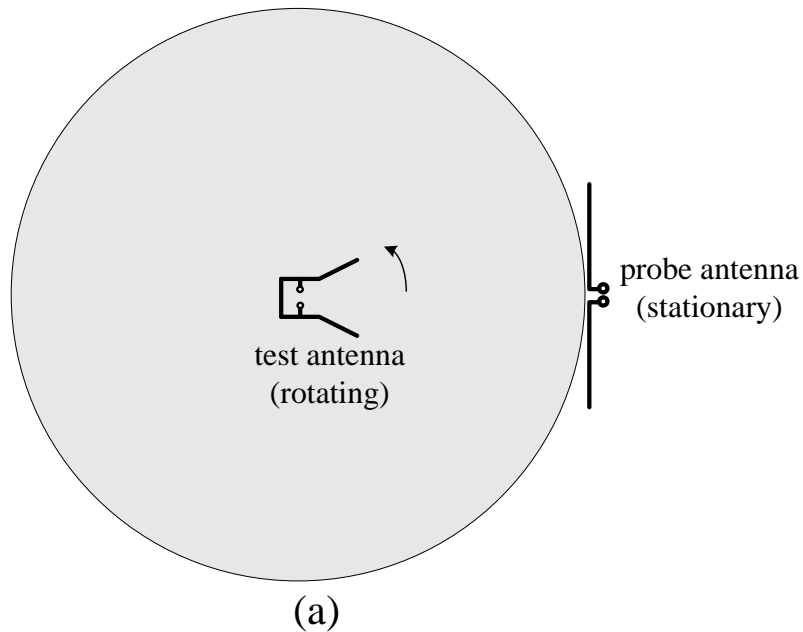
This result together with (10.13) and (10.14) leads to the conclusion that the received-to-transmitted power ratio in a two-antenna system does not depend on which antenna transmits and which receives.

1.4. Reciprocity of the radiation pattern

The measured radiation pattern of an antenna is the same in receiving and in transmitting mode if the system is linear. Nonlinear devices such as diodes and transistors make the system nonlinear, therefore, nonreciprocal.

In a two-antenna pattern measurement system, the pattern would not depend on whether the antenna under test (AUT) receives and the other antenna transmits, or *vice versa*. The pattern depends only on the mutual angular

orientation of the two antennas (the distance between the two antennas must remain the same regardless of the angular orientation of the antennas). It also does not matter whether the AUT rotates and the other antenna is stationary, or *vice versa*.

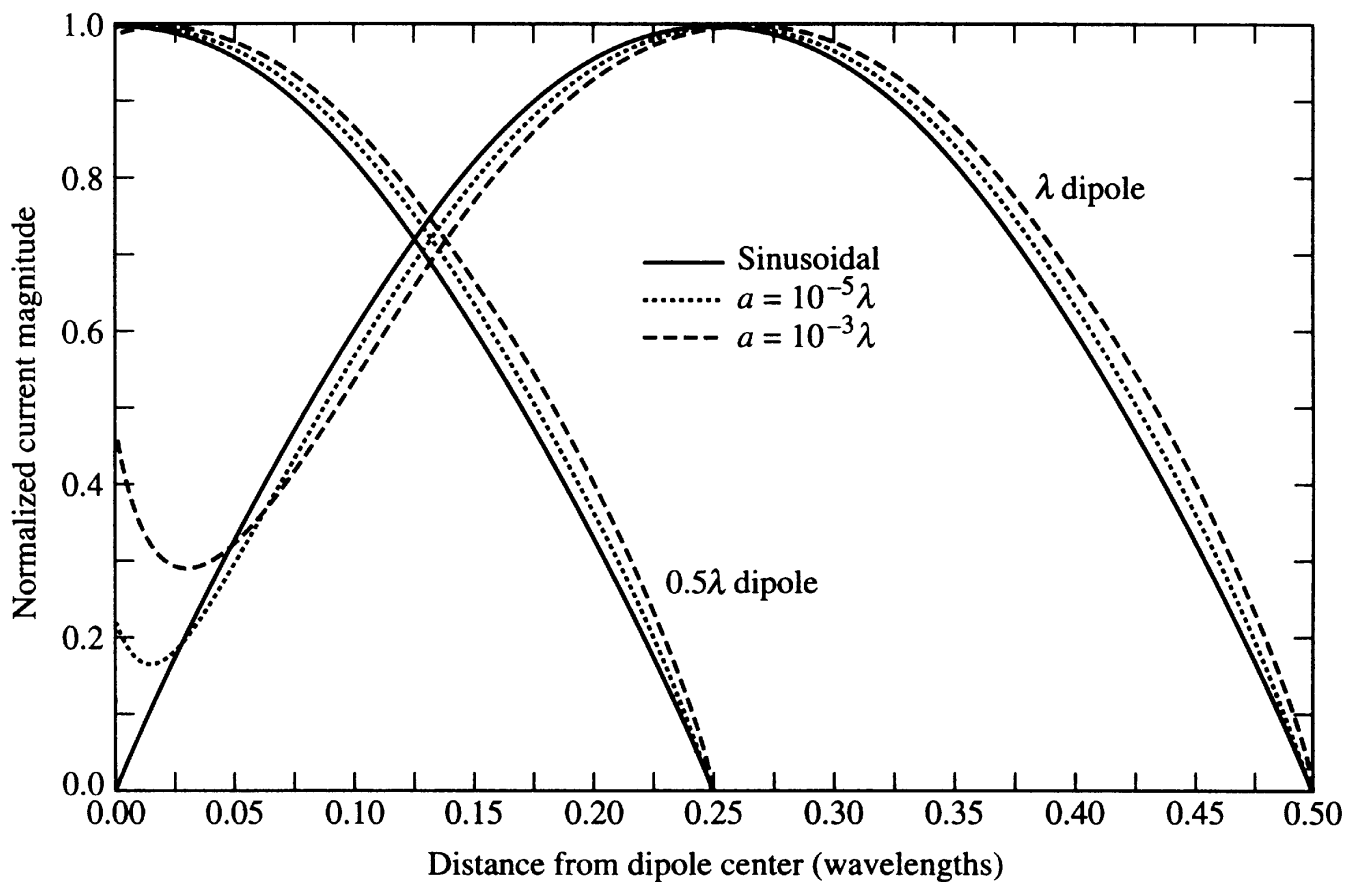


Scenario (a) is obviously more practical especially because the distance between the antennas must be sufficiently large to ensure a measurement in the far zone.

2. Self-impedance of a Dipole Using the Induced *EMF* Method

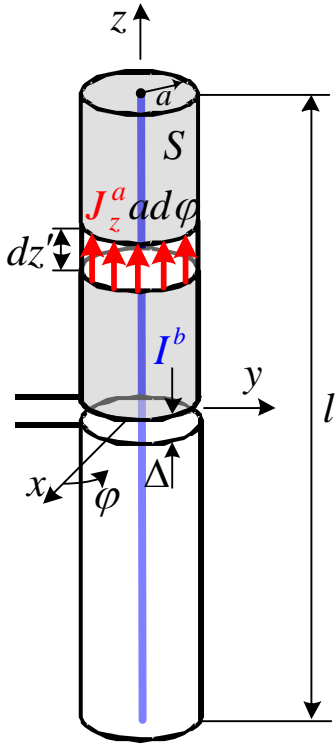
The induced *emf* (electro-motive force) method was developed by Carter² in 1932, when computers were not available and analytical (closed-form) solutions were much needed to calculate the self-impedance of wire antennas. The method was later extended to calculate mutual impedances of multiple wires (see, e.g., Elliot, *Antenna Theory and Design*). The *emf* method is restricted to straight parallel wires.

Measurements and full-wave simulations indicate that the current distribution on thin dipoles is nearly sinusoidal (except at the current minima). The induced *emf* method assumes this type of idealized distribution. It results in satisfactory accuracy for dipoles with length-diameter ratios as small as 100.



[Balanis]

² P.S. Carter, "Circuit relations in radiating systems and applications to antenna problems," *Proc. IRE*, **20**, pp.1004-1041, June 1932.



Consider a tubular dipole the arms of which are made of perfect electric conductor (PEC). When excited by a voltage-gap source at its base, the dipole supports surface current along z , which radiates. This surface current density is $J_{sz}(z') = H_\phi(z')$ as per the boundary conditions at the PEC surface where $E_z(z') = 0$.

Using the equivalence principle, we consider an equivalent problem where $J_{sz}^a(z') = J_{sz}(z')$ is a cylindrical current sheet that exists over a closed cylindrical surface S tightly enveloping the dipole. It radiates in open space generating the field $(\mathbf{E}^a, \mathbf{H}^a)$ such that

$$E_z^a(\rho \leq a, z') = \begin{cases} V_{in} / \Delta, & -\Delta/2 \leq z' \leq \Delta/2, \\ 0, & \text{elsewhere.} \end{cases} \quad (10.18)$$

Here, Δ is the feed gap length.

Consider also a linear current source $I^b(z')$ along the axis of the cylinder (the z axis) where I^b is nonzero only for $-l/2 < z' < l/2$. It also radiates in open space and its field is denoted as $(\mathbf{E}^b, \mathbf{H}^b)$.

In the volume bound by S , we apply the reciprocity formula (10.7):

$$\iiint_{V_{|S|}} (\mathbf{E}^a \cdot \mathbf{J}^b - \mathbf{E}^b \cdot \mathbf{J}^a) dv = 0. \quad (10.19)$$

Bearing in mind the surface nature of source a , the linear nature of source b , and equation (10.18), we write (10.19) as

$$\int_0^{2\pi} \int_{-l/2}^{l/2} E_z^b J_{sz}^a a dz' d\phi = \int_{-\Delta/2}^{\Delta/2} E_z^a I^b dz'. \quad (10.20)$$

In (10.20), we have neglected the edge effects from the disc-like end surfaces of S since a (the dipole radius) is at least 100 times smaller than its length l . We also assumed that the electric fields of the two sources have only z components.

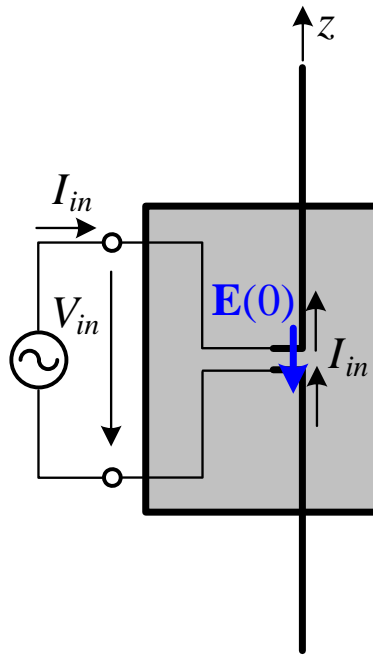
Assuming constant current distribution in the feed gap, we obtain

$$\int_0^{2\pi} \int_{-l/2}^{l/2} E_z^b J_{sz}^a a dz' d\phi = -I_{in} V_{in} \quad (10.21)$$

where

$$V_{in} = - \int_{-\Delta/2}^{\Delta/2} E_z^a dz' \quad (10.22)$$

is the voltage at the terminals of the generator driving the current $J_{sz}^a(z')$. The minus sign in (10.22) reflects the fact that a positive V_{in} , which implies a “positive” current I_{in} , i.e., current flowing in the positive z direction, relates to a “negative” electric field at the dipole’s base, i.e., E_z points in the negative z direction. This is illustrated below where the dipole antenna is viewed as a 1-port network.



Further, due to the cylindrical symmetry, all quantities in the integral in (10.21) are independent of φ . Thus,

$$\int_{-l/2}^{l/2} E_z^b (2\pi a J_{sz}^a) dz' = -I_{in} V_{in}. \quad (10.23)$$

The quantity in the brackets in (10.23) is the total current I^a at position z' :

$$\int_{-l/2}^{l/2} E_z^b I^a dz' = -I_{in} V_{in}. \quad (10.24)$$

We now require that I^a represents the actual current distribution along the dipole's arms and we drop the superscripts:

$$\int_{-l/2}^{l/2} E_z I dz' = -I_{in} V_{in}. \quad (10.25)$$

Note that E_z is the field at the cylindrical surface enveloping the dipole due to I^b , the distribution of which along z' is also representing the actual current distribution (i.e., it is identical to I^a). The above result leads to the following self-impedance expression:

$$Z_{in}|_{z'=0} = \frac{V_{in}}{I_{in}} = \frac{V_{in} \cdot I_{in}}{I_{in}^2} = -\frac{1}{I_{in}^2} \int_{-l/2}^{l/2} E_z(z') \cdot I(z') dz'. \quad (10.26)$$

In the classical *emf* method, we assume that the current has a sinusoidal distribution:

$$I(z') = \begin{cases} I_0 \sin \left[\beta \left(\frac{l}{2} - z' \right) \right], & 0 \leq z' \leq l/2 \\ I_0 \sin \left[\beta \left(\frac{l}{2} + z' \right) \right], & -l/2 \leq z' \leq 0. \end{cases} \quad (10.27)$$

So far, we have obtained only the far-field components of the field generated by the current in (10.27) (see Lecture 9). However, when the input resistance and reactance are needed, the near field must be known. In our case, we are interested in E_z , which is the field produced by $I(z')$ as if there is no conductor surface present. If we know it, we can calculate the integral in (10.26) since we already know $I(z')$ from (10.27).

We use cylindrical coordinates to describe the locations of the integration point (primed coordinates) and the observation point. The electric field can be expressed in terms of the VP \mathbf{A} and the scalar potential ϕ (see Lecture 2):

$$\mathbf{E} = -\nabla \phi - j\omega \mathbf{A}, \quad (10.28)$$

$$\Rightarrow E_z = -\frac{\partial \phi}{\partial z} - j\omega A_z. \quad (10.29)$$

The VP \mathbf{A} is z -polarized,

$$A_z = \frac{\mu}{4\pi} \int_{-l/2}^{l/2} I_z(z') \frac{e^{-j\beta R}}{R} dz'. \quad (10.30)$$

The scalar potential is

$$\phi = \frac{1}{4\pi\epsilon} \int_{-l/2}^{l/2} q_l(z') \frac{e^{-j\beta R}}{R} dz'. \quad (10.31)$$

Here, q_l stands for linear charge density (C/m). Knowing that the current depends only on z' , the continuity relation is written as

$$j\omega q_l = -\frac{\partial I_z}{\partial z'}. \quad (10.32)$$

$$\Rightarrow q_l(z') = \begin{cases} -j\frac{I_0}{c} \cos\left[\beta\left(\frac{l}{2} - z'\right)\right], & 0 \leq z' \leq l/2 \\ +j\frac{I_0}{c} \cos\left[\beta\left(\frac{l}{2} + z'\right)\right], & -l/2 \leq z' \leq 0 \end{cases} \quad (10.33)$$

where $c = \omega / \beta$ is the speed of light. Now we express \mathbf{A} and ϕ as

$$A_z = \frac{\mu}{4\pi} I_0 \left\{ \int_{-l/2}^0 \sin\left[\beta\left(\frac{l}{2} + z'\right)\right] \frac{e^{-j\beta R}}{R} dz' + \int_0^{l/2} \sin\left[\beta\left(\frac{l}{2} - z'\right)\right] \frac{e^{-j\beta R}}{R} dz' \right\} \quad (10.34)$$

$$\phi = j\frac{\eta I_0}{4\pi} \left\{ -\int_{-l/2}^0 \cos\left[\beta\left(\frac{l}{2} + z'\right)\right] \frac{e^{-j\beta R}}{R} dz' + \int_0^{l/2} \cos\left[\beta\left(\frac{l}{2} - z'\right)\right] \frac{e^{-j\beta R}}{R} dz' \right\}. \quad (10.35)$$

Here, $\eta = \sqrt{\mu / \epsilon}$ is the intrinsic impedance of the medium.

The distance between integration and observation point is

$$R = \sqrt{\rho^2 + (z - z')^2}. \quad (10.36)$$

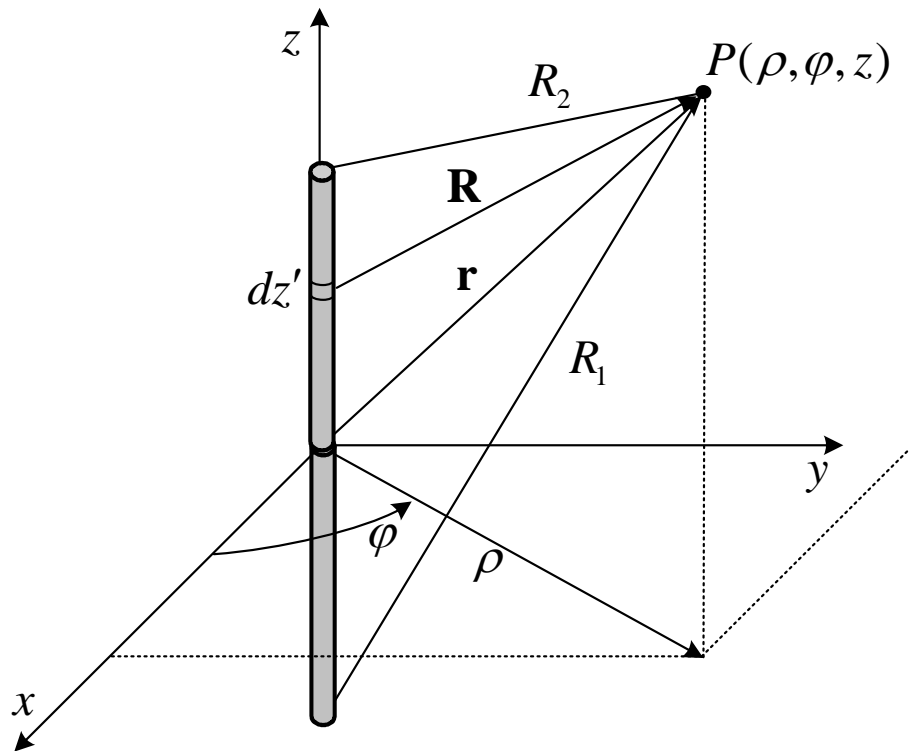
Equation (10.36) is substituted in (10.34) and (10.35). In addition, the resulting equations for A_z and ϕ are modified making use of Moivre's formulas:

$$\begin{aligned} \cos x &= \frac{1}{2}(e^{jx} + e^{-jx}) \\ \sin x &= \frac{1}{2j}(e^{jx} - e^{-jx}) \end{aligned}, \text{ where } x = \beta\left(\frac{l}{2} \pm z'\right). \quad (10.37)$$

Then, the equations for A_z and ϕ are substituted in (10.29) to derive the expression for E_z valid at any observation point. This is a rather lengthy derivation and we give the final result only:

$$E_z = -j \frac{\eta I_0}{4\pi} \left[\frac{e^{-j\beta R_1}}{R_1} + \frac{e^{-j\beta R_2}}{R_2} - 2 \cos\left(\frac{\beta l}{2}\right) \frac{e^{-j\beta r}}{r} \right]. \quad (10.38)$$

Here, r is the distance from the observation point to the dipole's center, while R_1 and R_2 are the distances to the lower and upper edges of the dipole, respectively (see figure below).



We need $E_z(z')$ at the dipole's surface where

$$r \approx z', \quad R_1 = z' + l/2, \quad \text{and} \quad R_2 = (l/2) - z'. \quad (10.39)$$

Notice the thin-wire approximation! The final goal of this development is to find the self-impedance (10.26) of the dipole. We substitute (10.39) in (10.38). The result for $E_z(z')$ is then substituted in (10.26), and the integration is performed. We give the final results for the real and imaginary parts of Z_{in} :

$$R_{in} = k \frac{\eta}{2\pi} \cdot \left\{ C + \ln(\beta l) - C_i(\beta l) + \frac{1}{2} \sin(\beta l) [S_i(2\beta l) - 2S_i(\beta l)] + \right. \\ \left. + \frac{1}{2} \cos(\beta l) \left[C + \ln\left(\frac{\beta l}{2}\right) + C_i(2\beta l) - 2C_i(\beta l) \right] \right\} = k \frac{\eta}{2\pi} \cdot \mathfrak{R}, \quad (10.40)$$

$$X_{in} = k \frac{\eta}{4\pi} \left\{ 2S_i(\beta l) - \cos(\beta l) [S_i(2\beta l) - 2S_i(\beta l)] + \right. \\ \left. + \sin(\beta l) [C_i(2\beta l) - 2C_i(\beta l) + C_i(2\beta a^2 / l)] \right\}, \quad (10.41)$$

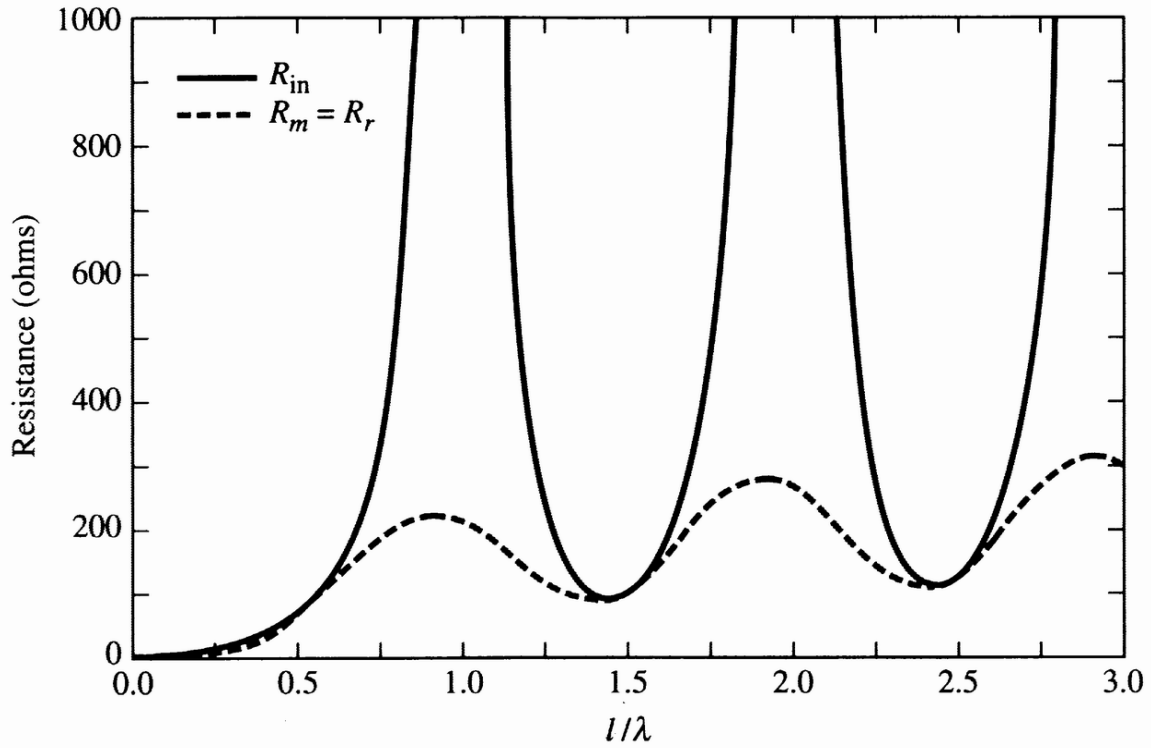
where $k = 1/\sin^2(\beta l/2)$ is the coefficient accounting for the difference between the maximum current magnitude along the dipole and the magnitude of the input current at the dipole's center [see Lecture 9, section 2]. Also, C is the Euler's constant, S_i is the sine integral and C_i is the cosine integral.

Equation (10.40) is identical with the expression found for the input resistance of an infinitesimally thin wire [see Lecture 9, Eqs. (9.37) and (9.38)]. Expression (10.41) for the dipole's reactance however is new. For a short dipole, the input reactance can be approximated by

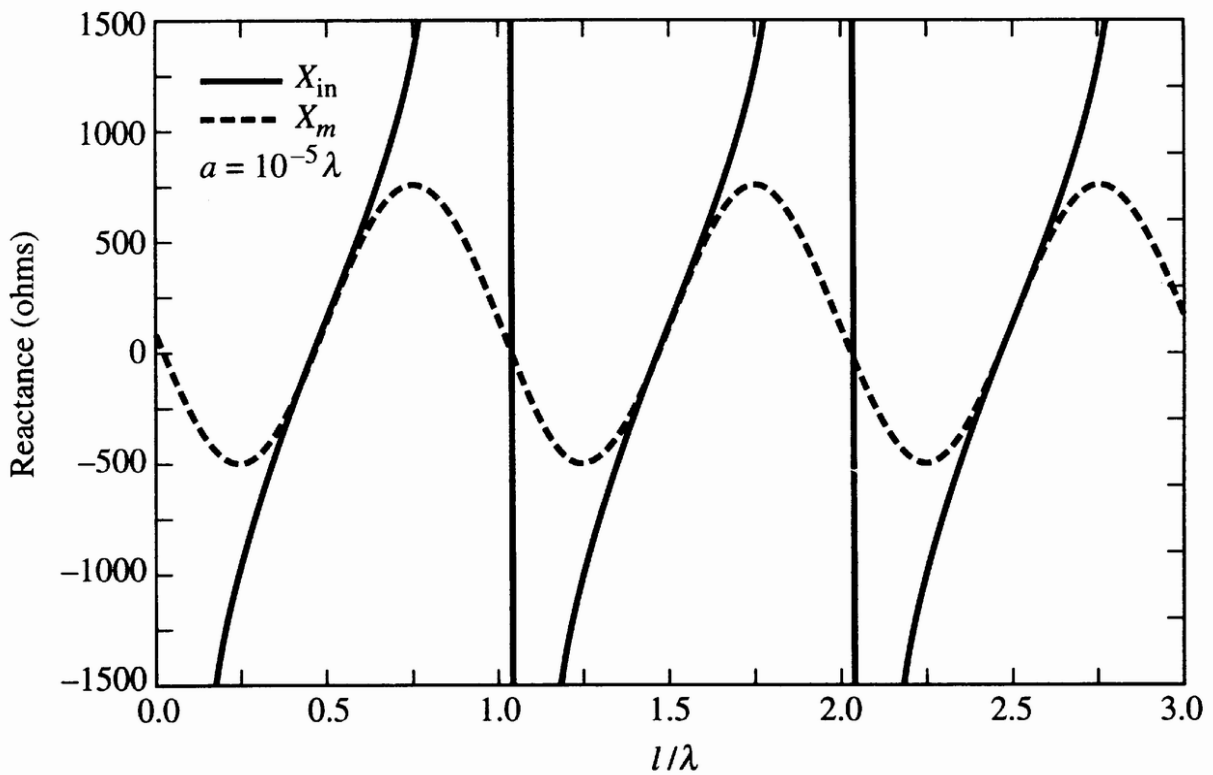
$$X_{in} \approx -120 \frac{[\ln(l/a) - 1]}{\tan(\beta l)}. \quad (10.42)$$

The results produced by (10.40) and (10.41) for different ratios l/λ are given in the plots below.

INPUT IMPEDANCE OF A THIN DIPOLE (*EMF* METHOD) OF RADIUS $a = 10^{-5} \lambda$



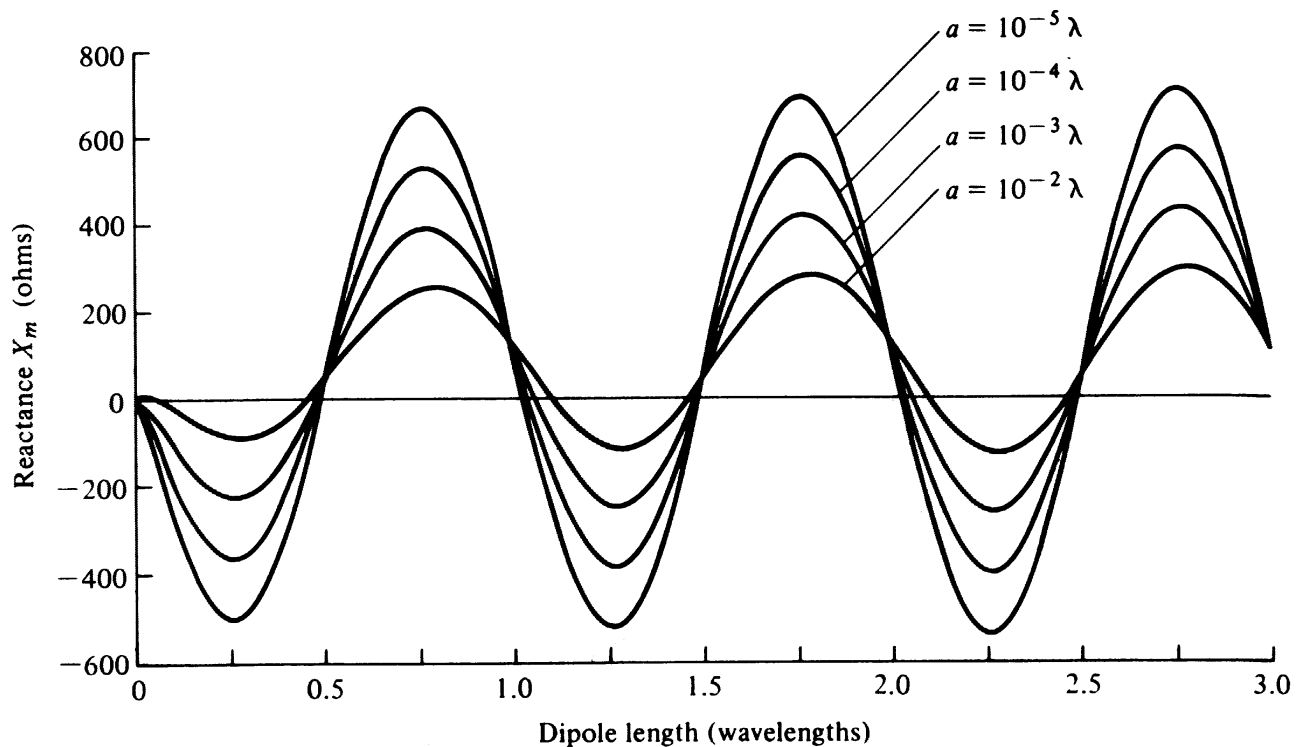
(a) Resistance



(b) Reactance

[Balanis]

INPUT REACTANCE OF A THIN DIPOLE (*EMF* METHOD) FOR DIFFERENT RADII a



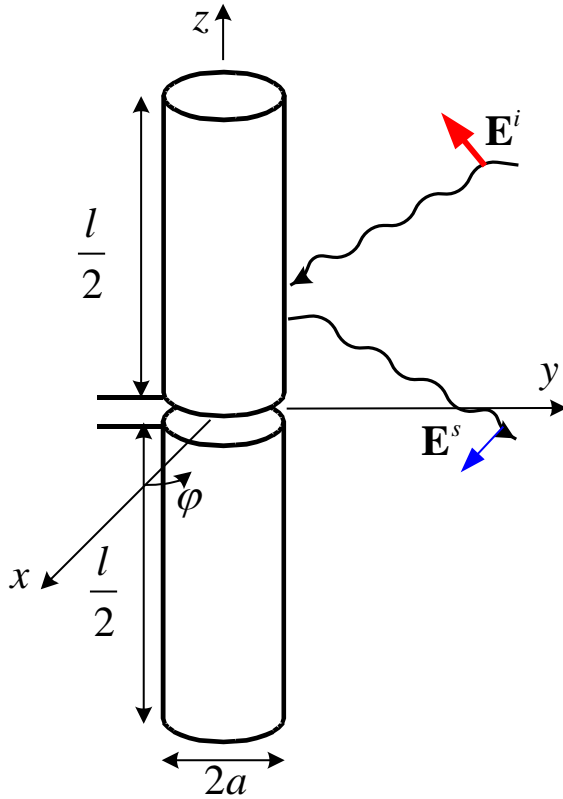
Note that:

- the reactance does not depend on the radius a , when the dipole length is a multiple of a half-wavelength ($l = n\lambda / 2$), as follows from (10.41);
- the resistance does not depend on a according to the assumptions made in the *emf* method (see equation (10.40)).

3. Pocklington's Equation

The assumption of a sinusoidal current distribution along the dipole is considered accurate enough for wire diameters $d < 0.05\lambda$. The current distribution is not quite sinusoidal in the case of thicker wires. The currents must be computed using some general numerical approach. Below, we introduce two integral equations, which can produce the current distribution on any straight wire antenna of finite diameter. These equations are classical in wire antenna theory. We do not discuss their numerical solutions here, only their derivations and applications.

To derive Pocklington's equation, the concepts of incident and scattered field are introduced first.



The *incident* wave is a wave produced by some external sources. This wave would have existed in the location of the scatterer, if the scatterer were not present. The scatterer however is present, and, since it is a conducting body, it requires vanishing electric field components tangential to its surface,

$$\mathbf{E}_{\tau}^i = 0. \quad (10.43)$$

The vector \mathbf{E}^i denotes the so-called total electric field. This means that as the non-zero incident field impinges upon the conducting scatterer, it induces on its surface currents \mathbf{J}_s , which in their turn produce a field, the *scattered* field \mathbf{E}^s . The scattered and the incident fields superimpose to produce the total field

$$\mathbf{E}^t = \mathbf{E}^i + \mathbf{E}^s. \quad (10.44)$$

The scattered field is such that (10.43) is fulfilled, i.e.,

$$\mathbf{E}_{\tau}^s = -\mathbf{E}_{\tau}^i. \quad (10.45)$$

Any object presenting a discontinuity in the wave's path is a scatterer, and so is any receiving antenna.

The above concepts hold for transmitting antennas, too. In the case of a wire dipole, the incident field exists only at the base of the dipole (in its feed gap). In the case of a cylindrical dipole with excitation of cylindrical symmetry, the \mathbf{E} field has no φ -component and is independent of the φ coordinate. The only tangential component is E_z . The boundary condition at the dipole's surface is

$$E_z^s = -E_z^i / \rho = a, -l/2 \leq z \leq l/2. \quad (10.46)$$

The scattered field can be expressed in terms of \mathbf{A} and ϕ as it was already done in (10.29):

$$E_z^s = -\frac{\partial \phi}{\partial z} - j\omega A_z = -j \frac{1}{\omega \mu \epsilon} \frac{\partial^2 A_z}{\partial z^2} - j\omega A_z, \quad (10.47)$$

or

$$E_z^s = -j \frac{1}{\omega \mu \varepsilon} \left(\beta^2 A_z + \frac{\partial^2 A_z}{\partial z^2} \right). \quad (10.48)$$

We assume only z -components of the surface currents and no edge effects:

$$A_z(\rho, \varphi, z) = \frac{\mu}{4\pi} \int_{-l/2}^{l/2} \int_0^{2\pi} J_{sz} \frac{e^{-j\beta R}}{R} \underbrace{ad\varphi' dz'}_{ds}. \quad (10.49)$$

If the cylindrical symmetry of the dipole and the excitation are preserved, the current J_z does not depend on the azimuthal angle φ . It can be shown that the field due to a cylindrical sheet of surface currents J_{sz} is equivalent to the field due to a current filament of current I_z ,

$$2\pi a J_{sz} = I_z \Rightarrow J_{sz}(z') = \frac{1}{2\pi a} I_z(z'). \quad (10.50)$$

Then, (10.49) reduces to

$$A_z(\rho, \varphi, z) = \frac{\mu}{4\pi} \int_{-l/2}^{l/2} \frac{1}{2\pi a} \int_0^{2\pi} I_z(z') \frac{e^{-j\beta R}}{R} ad\varphi' dz'. \quad (10.51)$$

Here, the distance between observation and integration points is

$$\begin{aligned} R &= \sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2} = \\ &= \sqrt{\rho^2 + a^2 - 2\rho a \cos(\varphi - \varphi') + (z-z')^2}, \end{aligned} \quad (10.52)$$

where $x = \rho \cos \varphi$, $y = \rho \sin \varphi$, $x' = a \cos \varphi'$, $y' = a \sin \varphi'$.

The cylindrical geometry of the problem implies the cylindrical symmetry of the observed fields, i.e., \mathbf{A} does not depend on φ . We assume that $\varphi = 0$. Besides, we are interested in the scattered field produced by this equivalent current at the dipole's surface, i.e., the observation point is at $\rho = a$. Then,

$$A_z(a, 0, z) = \mu \int_{-l/2}^{l/2} I_z(z') \left(\frac{1}{2\pi} \int_0^{2\pi} \frac{e^{-j\beta R}}{4\pi R} d\varphi' \right) dz' = \mu \int_{-l/2}^{l/2} I_z(z') G(z, z') dz', \quad (10.53)$$

where

$$G(z, z') = \frac{1}{2\pi} \int_0^{2\pi} \frac{e^{-j\beta R}}{4\pi R} d\varphi', \quad (10.54)$$

$$R_{(\rho=a, \varphi=0)} = \sqrt{4a^2 \sin^2\left(\frac{\varphi'}{2}\right) + (z - z')^2}. \quad (10.55)$$

Substituting (10.53) in (10.48) yields

$$E_z^s(\rho = a) = -j \frac{1}{\omega \varepsilon} \left(\beta^2 + \frac{d^2}{dz^2} \right) \int_{-l/2}^{l/2} I_z(z') G(z, z') dz'. \quad (10.56)$$

Imposing the boundary condition (10.46) on the field in (10.56) leads to

$$\left(\beta^2 + \frac{d^2}{dz^2} \right) \int_{-l/2}^{l/2} I_z(z') G(z, z') dz' = -j \omega \varepsilon E_z^i(\rho = a). \quad (10.57)$$

The source I_z does not depend on z and (10.57) can be rewritten as

$$\int_{-l/2}^{l/2} I_z(z') \left(\beta^2 G(z, z') + \frac{d^2 G(z, z')}{dz^2} \right) dz' = -j \omega \varepsilon E_z^i(\rho = a). \quad (10.58)$$

Equation (10.58) is called Pocklington's³ integro-differential equation. It is used to compute the equivalent filamentary current distribution $I_z(z')$ by knowing the incident field on the dipole's surface.

When the gap of length b is the only place where E_z^i exists, equation (10.58) is written as

$$\int_{-l/2}^{l/2} I_z(z') \left(\beta^2 G(z, z') + \frac{d^2 G(z, z')}{dz^2} \right) dz' = \begin{cases} -j \omega \varepsilon E_z^i, & -\frac{b}{2} \leq z \leq \frac{b}{2} \\ 0, & \frac{b}{2} < |z| < \frac{l}{2} \end{cases} \quad (10.59)$$

If we assume that the wire is very thin, then Green's function $G(z, z')$ of (10.54) simplifies to

$$G(z, z') = \frac{1}{2\pi} \int_0^{2\pi} \frac{e^{-j\beta R}}{4\pi R} d\varphi' = \frac{e^{-j\beta R}}{4\pi R}, \quad (10.60)$$

where R reduces to $R_{(\rho=a \rightarrow 0)} = \sqrt{4a^2 \sin^2(\varphi' / 2) + (z - z')^2} \approx z - z'$. Assuming

³ H.C. Pocklington, "Electrical oscillation in wires", *Camb. Phil. Soc. Proc.*, **9**, 1897, pp.324-332.

(10.60), Richmond⁴ has differentiated and re-arranged (10.58) in a form more convenient for programming:

$$\int_{-l/2}^{+l/2} I_z(z') \frac{e^{-j\beta R}}{4\pi R^5} \left[(1 + j\beta R)(2R^2 - 3a^2) + (\beta a R)^2 \right] dz' = -j\omega\epsilon E_z^i. \quad (10.61)$$

Equation (10.61) can be solved numerically by the Method of Moments (MoM), after the structure is discretized into small linear segments.

4. Hallén's equation

Hallén's equation⁵ can be derived as a modification of Pocklington's equation. It is easier to solve numerically, but it makes some additional assumptions. Consider again equation (10.59). It can be written in terms of A_z explicitly as (see also (10.48)):

$$\frac{d^2 A_z}{dz^2} + \beta^2 A_z = \begin{cases} -j\omega\epsilon\mu E_z^i, & -\frac{b}{2} \leq z \leq \frac{b}{2} \\ 0, & \frac{b}{2} < |z| < \frac{l}{2} \end{cases} \quad (10.62)$$

When $b \rightarrow 0$, we can express the incident field in the gap via the voltage applied to the gap:

$$V_g = \lim_{b \rightarrow 0} b E_z^i. \quad (10.63)$$

The $E_z^i(z)$ function is an impulse function of z , such that

$$E_z^i = V_g \delta(z). \quad (10.64)$$

The excitation term in (10.62) collapses into a δ -function:

$$\frac{d^2 A_z}{dz^2} + \beta^2 A_z = -j\omega\epsilon\mu V_g \delta(z). \quad (10.65)$$

If $z \neq 0$,

$$\frac{d^2 A_z}{dz^2} + \beta^2 A_z = 0. \quad (10.66)$$

⁴ J.H. Richmond, "Digital computer solutions of the rigorous equations for scattering problems," *Proc. IEEE*, **53**, pp.796-804, August 1965.

⁵ E. Hallén, "Theoretical investigation into the transmitting and receiving qualities of antennae," *Nova Acta Regiae Soc. Sci. Upsaliensis*, Ser. IV, No. 4, 1938, pp. 1-44.

Because the current density on the cylinder is symmetrical with respect to z' , i.e., $J_z(z') = J_z(-z')$, the potential A_z must also be symmetrical. Then, the general solution of the ODE in (10.66) along z ($x = y = 0$) has the form:

$$A_z(z) = B \cos(\beta z) + C \sin(\beta |z|). \quad (10.67)$$

From (10.65) it follows that

$$\left. \frac{dA_z}{dz} \right|_{0_-}^{0_+} = -j\omega\mu\varepsilon V_g. \quad (10.68)$$

From (10.67) and (10.68), we calculate the constant C :

$$\begin{aligned} \left. \frac{dA_z}{dz} \right|_{0_-}^{0_+} &= C\beta \cos(0_+) - C(-\beta) \cos(0_-) = -j\omega\mu\varepsilon V_g, \\ &\Rightarrow 2C\beta = -j\omega\mu\varepsilon V_g, \\ &\Rightarrow C = -j\sqrt{\mu\varepsilon} \frac{V_g}{2} = -j \frac{\mu}{\eta} \frac{V_g}{2}. \end{aligned} \quad (10.69)$$

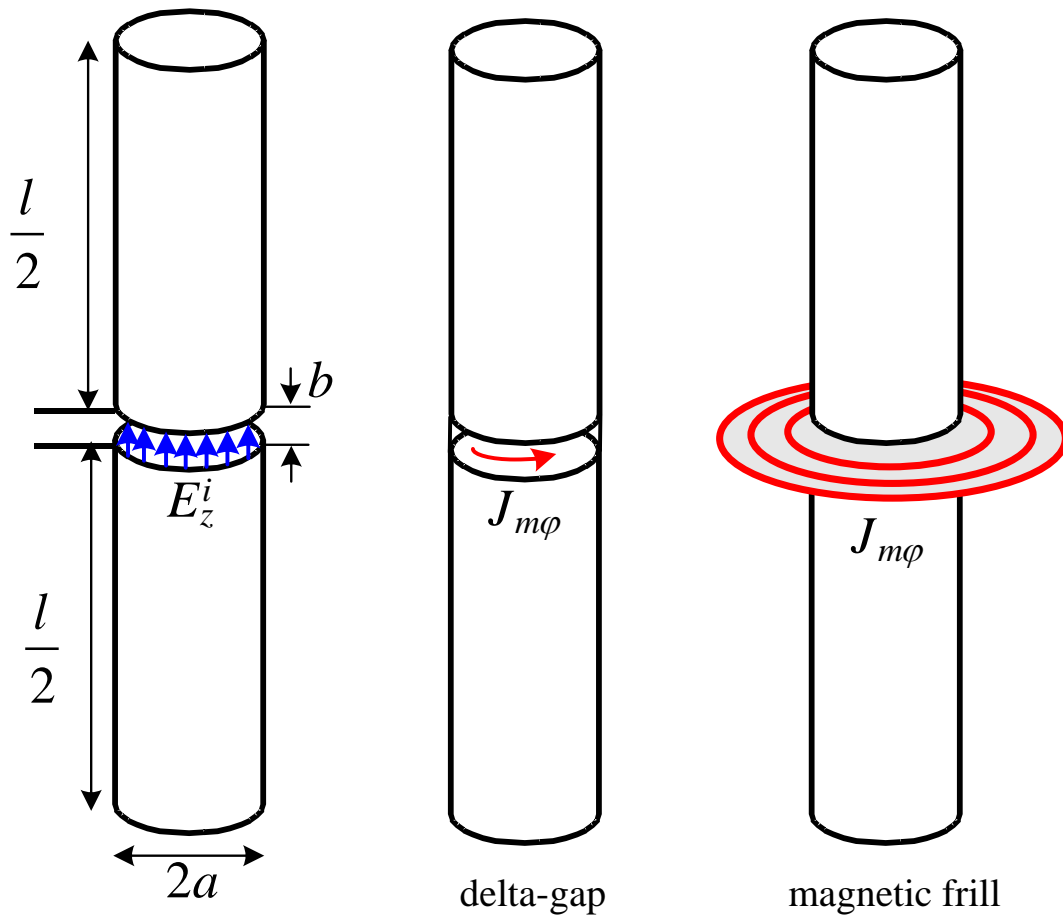
Equation (10.69) is substituted in (10.67), and A_z is expressed with its integral over the currents, to obtain the final form of Hallén's integral equation:

$$\int_{-l/2}^{+l/2} I_z(z') \frac{e^{-j\beta R}}{4\pi R} dz' = -j \frac{V_g}{2\eta} \sin(\beta |z|) + B \cos(\beta z). \quad (10.70)$$

Here, $R = \sqrt{a^2 + (z - z')^2}$. We must reiterate that *Hallén's equation assumes that the incident field exists only in the infinitesimal dipole gap*, while in Pocklington's equation there are no restrictions on the distribution of the incident field at the dipole.

5. Modeling the excitation field

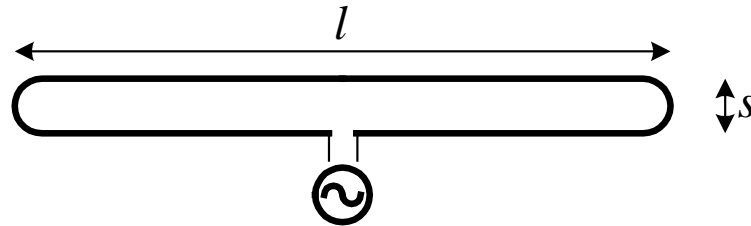
- Delta-gap source (Pocklington and Hallén)
- Magnetic frill source (Pocklington)



LECTURE 11: Practical Dipole/Monopole Geometries. Matching Techniques for Dipole/Monopole Feeds

(The folded dipole antenna. Conical skirt monopoles. Sleeve antennas. Turnstile antenna. Impedance matching techniques. Dipoles with traps.)

1. Folded Dipoles

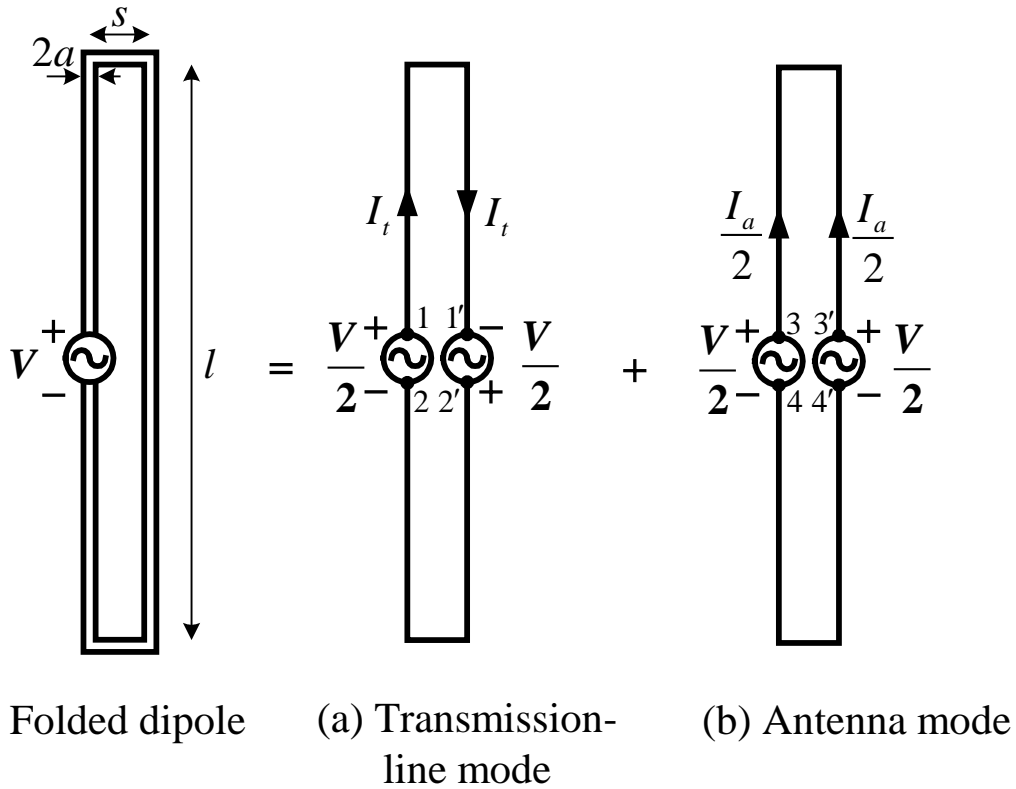


The folded dipole is a popular antenna for reception of TV broadcast signals. It has essentially the same radiation pattern as the dipole of the same length l but it provides four times greater input impedance when $l \approx \lambda / 2$. The input resistance of the conventional half-wavelength dipole is $R_{in} \approx 73 \Omega$ while that of the half-wavelength folded dipole is about 292Ω . Wire antennas do not fit well with coaxial feed lines because of the different field mode; thus, balanced-to-unbalanced transition is required. However, they are ideally suited for twin-lead (two-wire) feed lines. These lines (two parallel thin wires separated by a distance of about 8 to 10 mm) have characteristic impedance $Z_0 \approx 300 \Omega$. Therefore, an input antenna impedance of $(4 \times 73) \Omega$ matches well the 2-wire feed lines. The separation distance between the two wires of the folded dipole s should not exceed 0.05λ .

The folded dipole can be analyzed by decomposing its current into two modes: the transmission-line (TL) mode and the antenna mode. This analysis, although approximate¹, illustrates the four-fold impedance transformation.

In the TL mode, the source terminals $1-2'$ and $2-1'$ are at the same potential and can be connected by a short without changing the mode of operation. The equivalent source (of voltage V) is now feeding in parallel two 2-wire transmission lines, each terminated by a short.

¹ G.A. Thiele, E.P. Ekelman, Jr., L.W. Henderson, "On the accuracy of the transmission line model for the folded dipole," *IEEE Trans. on Antennas and Propagation*, vol. AP-28, No. 5, pp. 700-703, Sep. 1980.



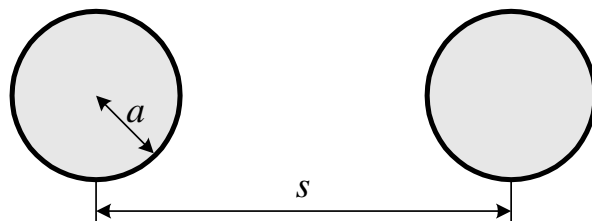
The input impedance of each shorted transmission line of length $l/2$ is

$$Z_t = \left[Z_0 \left(\frac{Z_L + jZ_0 \tan(\beta l / 2)}{Z_0 + jZ_L \tan(\beta l / 2)} \right) \right]_{Z_L=0}, \quad (11.1)$$

$$\Rightarrow Z_t = jZ_0 \tan\left(\frac{\beta l}{2}\right). \quad (11.2)$$

Here, Z_0 is the characteristic impedance of the 2-wire transmission line formed by the two segments of the folded wire. It can be calculated as

$$Z_0 = \frac{\eta}{\pi} \operatorname{arccosh}\left(\frac{s}{2a}\right) = \frac{\eta}{\pi} \ln \left[\frac{s/2 + \sqrt{(s/2)^2 - a^2}}{a} \right]. \quad (11.3)$$



Usually, the folded dipole has a length of $l \approx \lambda / 2$. Then,

$$Z_{t(l=\lambda/2)} = jZ_0 \tan(\pi / 2) \rightarrow \infty. \quad (11.4)$$

If $l \neq \lambda / 2$, the more general expression (11.2) should be used. The current in the transmission-line mode is

$$I_t = \frac{V}{2Z_t} \quad (11.5)$$

because the voltage applied to the equivalent transmission lines of the upper and lower dipole arms is $V/2$ (see figure (a) on previous page).

We now consider the antenna mode. The generators' terminals 3–3' (and 4–4') are with identical potentials. Therefore, they can be connected electrically without changing the conditions of operation. The following assumption is made: an equivalent dipole of effective radius

$$a_e = \sqrt{as} \quad (11.6)$$

is radiating excited by $V / 2$ voltage. Since usually $a \ll \lambda$ and $s \ll \lambda$, the input impedance of the equivalent dipole Z_a is assumed equal to the input impedance of an infinitesimally thin dipole of the respective length l . If $l = \lambda / 2$, then $Z_a = 73 \ \Omega$. The current in the antenna mode is

$$I_a = \frac{V}{2Z_a}. \quad (11.7)$$

The current on each arm of the equivalent dipole is

$$\frac{I_a}{2} = \frac{V}{4Z_a}. \quad (11.8)$$

The total current of a folded dipole is obtained by superimposing both modes. At the input

$$I_{in} = I_t + \frac{I_a}{2} = V \left(\frac{1}{2Z_t} + \frac{1}{4Z_a} \right), \quad (11.9)$$

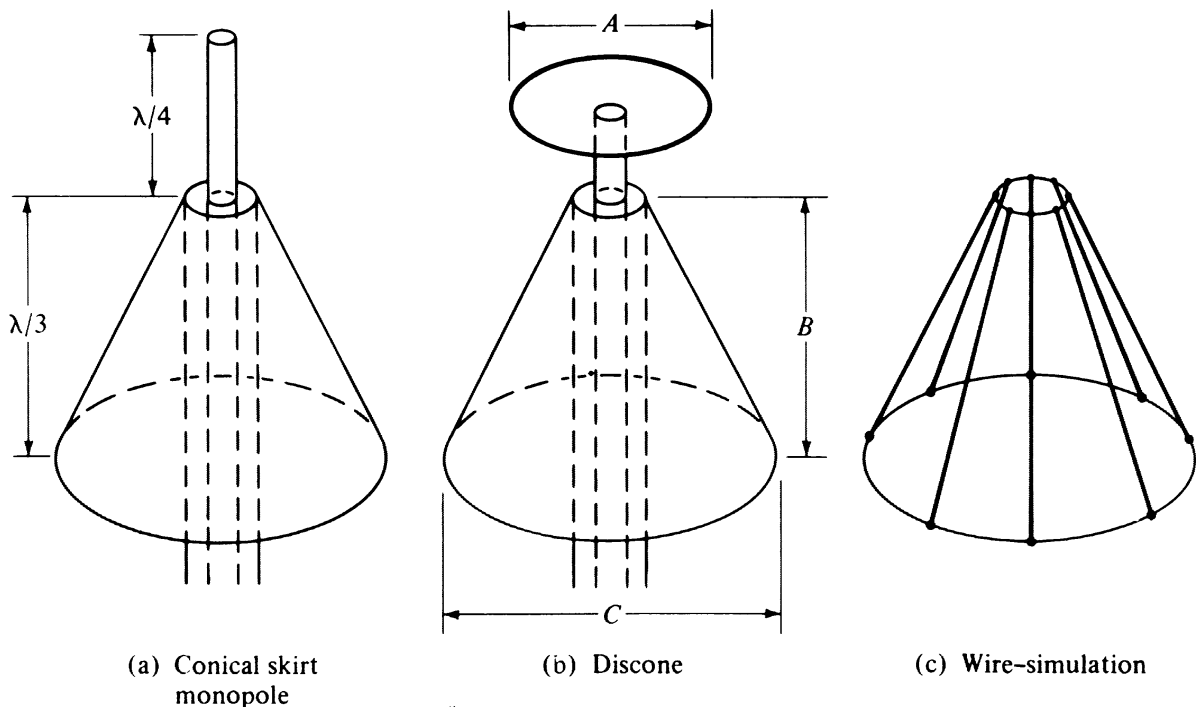
$$\Rightarrow Z_{in} = \frac{4Z_t Z_a}{2Z_a + Z_t}. \quad (11.10)$$

When $l = \lambda / 2$ (half-wavelength folded dipole), then $Z_t \rightarrow \infty$, and

$$\Rightarrow Z_{in} = 4Z_a |_{l=\lambda/2} \approx 292 \ \Omega. \quad (11.11)$$

Thus, the half-wavelength folded dipole is well suited for direct connection to a twin-lead line of $Z_0 \approx 300 \ \Omega$. It is often made in a simple way: a suitable portion (the end part of the twin-lead cable of length $l = \lambda / 2$) is separated into two single wire leads, which are bent to form the folded dipole.

2. Conical (Skirt) Monopoles and Discones

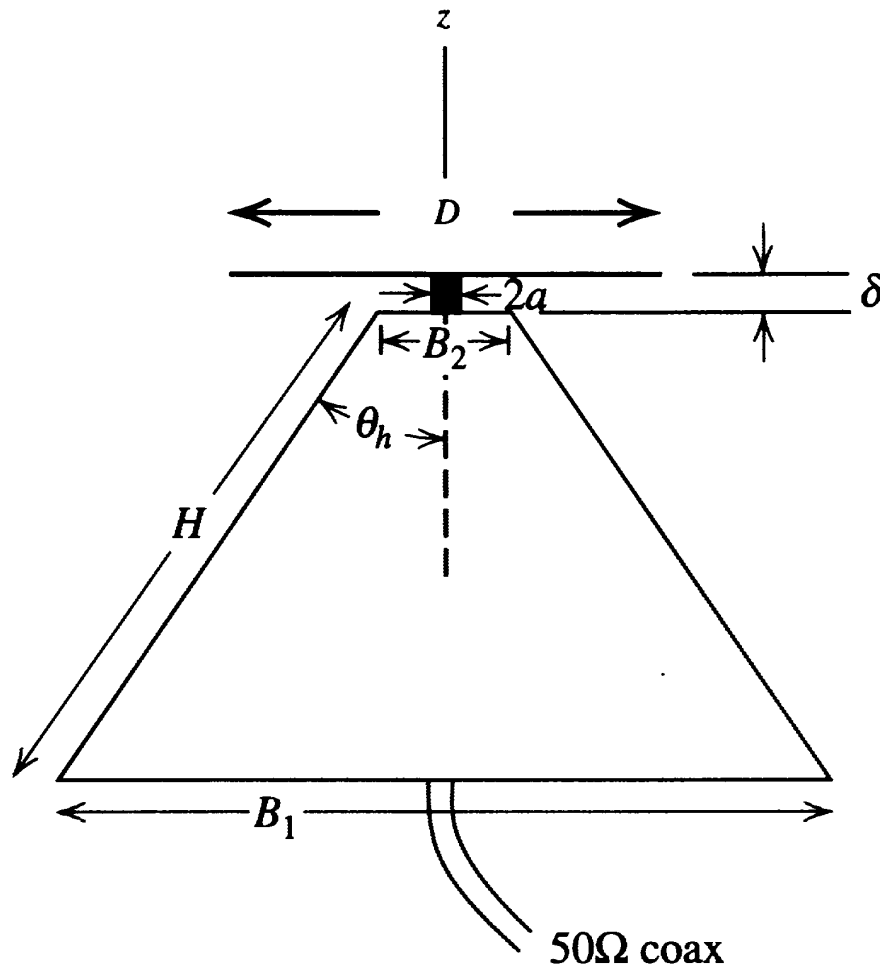


These monopoles have much broader impedance frequency band (a couple of octaves) than the ordinary quarter-wavelength monopoles. They are a combination of the two basic antennas: the monopole/dipole antenna and the biconical antenna. The discone and conical skirt monopoles find wide application in the VHF (30 to 300 MHz) and the UHF (300 MHz to 3 GHz) spectrum for FM broadcast, television and mobile communications.

There are numerous variations of the dipole/monopole/cone geometries, which aim at broader bandwidth rather than shaping the radiation pattern. All these antennas provide omnidirectional radiation.

The discone (disk-cone) is the most broadband among these types of

antennas. This antenna was first designed by Kandoian² in 1945. The performance of the discone in frequency is similar to that of a high-pass filter. Below certain effective cutoff frequency, it has a considerable reactance and produces severe standing waves in the feed line. This happens approximately at wavelength such that the slant height of the cone is $\approx \lambda / 4$.

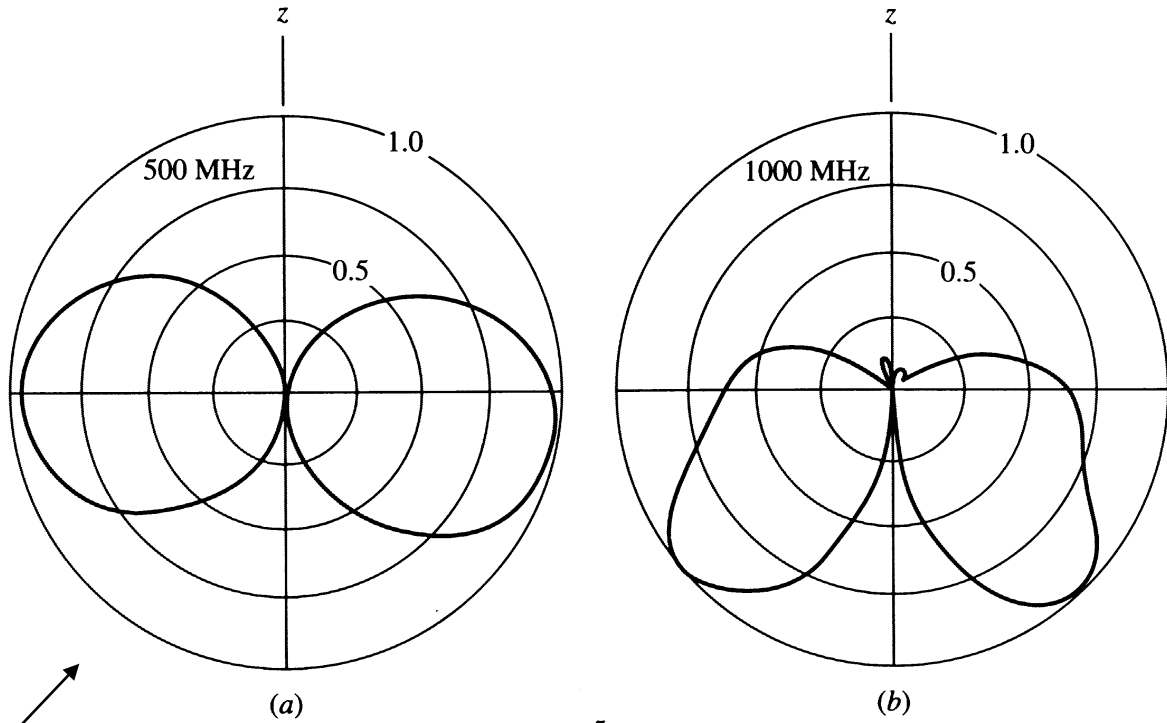


Typical dimensions of a discone antenna at the central frequency are: $D \approx 0.4\lambda$, $B_1 \approx 0.6\lambda$, $H = 0.7\lambda$, $45^\circ \leq 2\theta_h \leq 75^\circ$ and $\delta \ll \lambda$. The typical input impedance is designed to be 50Ω . Optimum design formulas are given by Nail³: $B_2 \approx \lambda_u / 75$ at the highest operating frequency, $\delta \approx (0.3 \div 0.5)B_2$, and $D \approx 0.7B_1$.

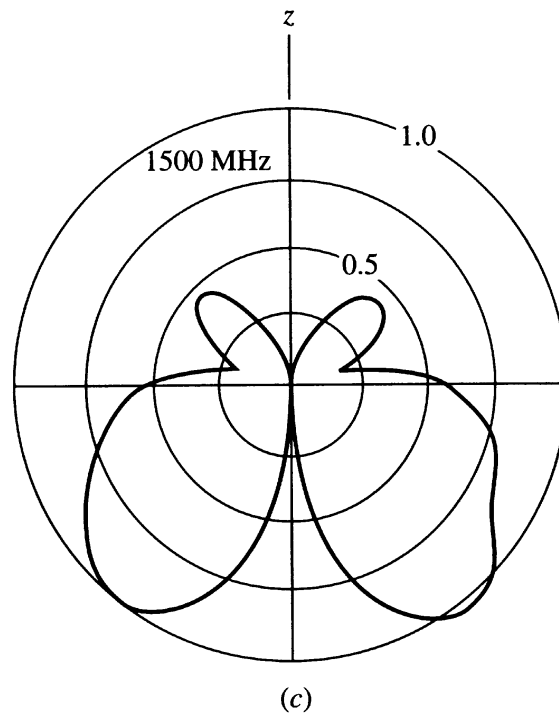
² A.G. Kandoian, "Three new antenna types and their application," *Proc. IRE*, vol. 34, pp. 70W-75W, Feb. 1946.

³ J.J. Nail, "Designing discone antennas," *Electronics*, vol. 26, pp. 167-169, Aug. 1953.

Measured patterns for a discone: $H = 21.3$ cm, $B = 19.3$ cm, $\theta_h = 25^\circ$:



Similar to a short dipole

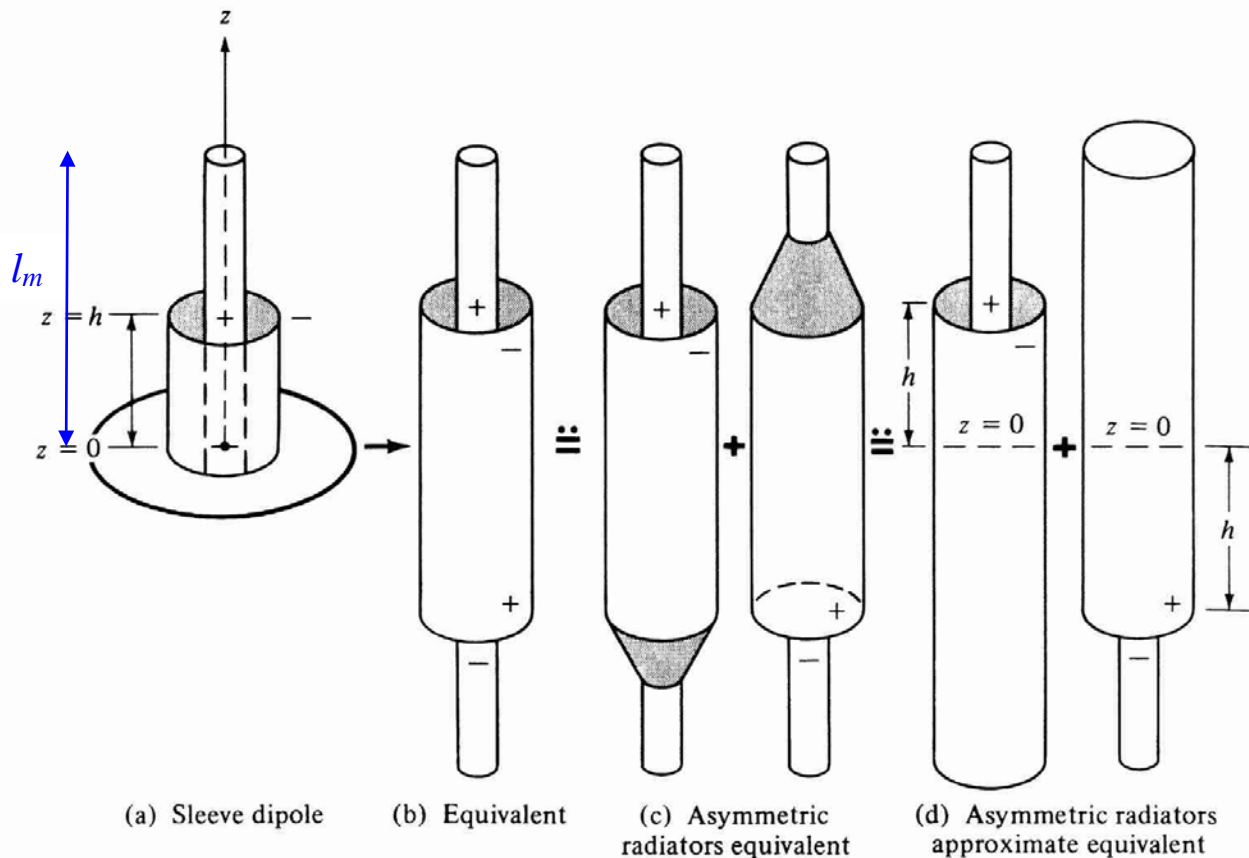


Similar to an infinite conical monopole

[Balanis]

3. Sleeve (Coaxial) Dipoles and Monopoles

The impedance of dipole/monopole antennas is very frequency sensitive. The addition of a sleeve to a dipole or a monopole can increase the bandwidth up to more than an octave and fine-tune the input impedance.



Sleeve dipole and its equivalents. (SOURCE: W. L. Weeks, *Antenna Engineering*, McGraw-Hill, New York, 1968)

This type of antenna closely resembles two asymmetrically fed dipoles and can be analyzed using the approximation in (d). The outer shield of the coaxial line is connected to the ground plane, but it also extends above it a distance h [see (a)] in order to provide mechanical strength, impedance tuning and impedance broadband characteristics. The equivalent in (d) consists of two dipoles, which are asymmetrically driven at $z' = +h$ or $z' = -h$. When analyzing the field of the two asymmetrically driven dipoles, we can ignore the change in diameter occurring at the feed point.

The input impedance Z_{as} of an asymmetric dipole can be related to the impedance Z_s of a center-fed (symmetric) dipole of the same length l using the assumption for sinusoidal current distribution [see Lecture 9],

$$I(z') = \begin{cases} I_0 \sin[\beta(0.5l - z')] & , 0 \leq z' \leq l/2 \\ I_0 \sin[\beta(0.5l + z')] & , -l/2 \leq z' \leq 0. \end{cases} \quad (11.12)$$

The impedances Z_{as} and Z_s relate to the radiated power Π through their respective current magnitudes, $I(z' = h)$ and $I(z' = 0)$. Imposing the condition that for a given current distribution the power delivered by the transmitter must be equal to the radiated (active) and stored (reactive) power of the antenna, leads to

$$Z_s I_s^2 = Z_{as} I_{as}^2. \quad (11.13)$$

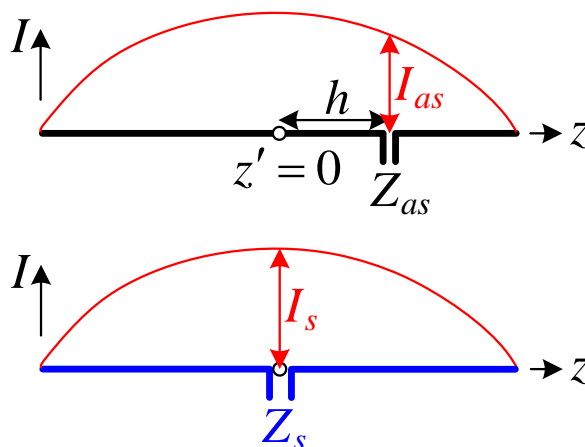
Thus,

$$Z_{as}(h) = Z_s \left[\frac{I(z' = 0)}{I(z' = h)} \right]^2 = Z_s \frac{\sin^2(\beta l / 2)}{\sin^2 \left[\beta \left(\frac{l}{2} - h \right) \right]}. \quad (11.14)$$

For a half-wavelength dipole, $\beta l / 2 = \pi / 2$ and $\sin(0.5\pi - \beta h) = \cos(\beta h)$. Thus, the relation (11.14) reduces to

$$Z_{as}(h) = \frac{Z_s}{\cos^2(\beta h)}, \text{ for } l = \frac{\lambda}{2}. \quad (11.15)$$

The relation between the input impedances of the symmetric and asymmetric dipole feeds is illustrated in the figure below.



The relation between Z_{as} and the impedance Z_m associated with the current maximum I_m (remember that the radiation resistance of the dipole is $R_r = \text{Re } Z_m$) can be found through Z_s bearing in mind that

$$Z_s = Z_m (I_m / I_s)^2. \quad (11.16)$$

The input current I_s for the center-fed dipole relates to the maximum in the dipole's current distribution I_m as

$$\begin{aligned} I_s &= I_m, & \text{if } l \leq \lambda / 2 \\ I_s &= I_m \sin(\beta l / 2), & \text{if } l \geq \lambda / 2. \end{aligned} \quad (11.17)$$

Remember from Lecture 9 that

$$\begin{aligned} I_m &= I_0 \sin(\beta l / 2), & \text{if } l \leq \lambda / 2 \\ I_m &= I_0, & \text{if } l > \lambda / 2. \end{aligned} \quad (11.18)$$

Thus, in terms of I_0 , we always have $I_s = I_0 \sin(\beta l / 2)$.

From (11.16) and (11.17), it follows that

$$\begin{aligned} Z_s &= Z_m, & \text{if } l \leq \lambda / 2 \\ Z_s &= \frac{Z_m}{\sin^2(\beta l / 2)}, & \text{if } l \geq \lambda / 2. \end{aligned} \quad (11.19)$$

Now (11.14) can be written in terms of Z_m as

$$Z_{as}(h) = \begin{cases} Z_m \frac{\sin^2(\beta l / 2)}{\sin^2 \left[\beta \left(\frac{l}{2} - h \right) \right]}, & \text{if } l \leq \frac{\lambda}{2} \\ Z_m \frac{1}{\sin^2 \left[\beta \left(\frac{l}{2} - h \right) \right]}, & \text{if } l \geq \frac{\lambda}{2} \end{cases} \quad (11.20)$$

It is now obvious that we can tune the input impedance of a dipole by moving the feed point off-center. In the case of a sleeve monopole, this is achieved by changing h , i.e., shortening or extending the sleeve along the stub.

Let us examine the equivalent antenna structure in Figure (d). It consists of two asymmetrically driven dipoles in parallel. The total input current is

$$I_{in} = I_{as}(z' = +h) + I_{as}(z' = -h). \quad (11.21)$$

The input admittance is

$$Y_{in} = \frac{I_{in}}{V_{in}} = \frac{I_{as}(z' = +h) + I_{as}(z' = -h)}{V_{in}} = \frac{I_{as}(z' = h)}{V_{in}} \left[1 + \frac{I_{as}(z' = -h)}{I_{as}(z' = +h)} \right] \quad (11.22)$$

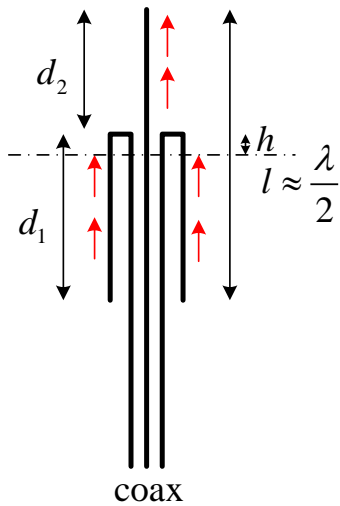
$$\Rightarrow Y_{in} = Y_{as} \left[1 + \frac{I_{as}(z' = -h)}{I_{as}(z' = +h)} \right]. \quad (11.23)$$

Since the two dipoles in (d) are geometrically identical and their currents are equal according to image theory, $I_{as}(z' = -h) = I_{as}(z' = +h)$, it follows from (11.22) that

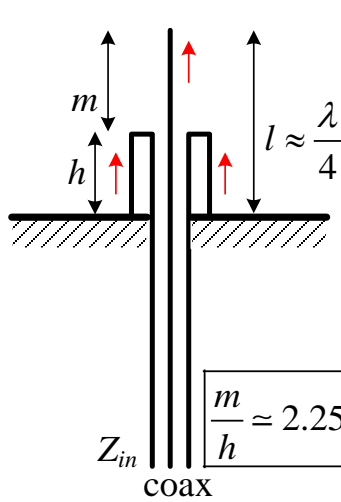
$$Y_{in} \approx 2Y_{as}. \quad (11.24)$$

Thus, the impedance of the sleeve antenna in (a) is twice smaller than the impedance of the respective asymmetrically driven dipole [one of the dipoles in (d)]. This conclusion is in agreement with the general relation between the impedance of a monopole above a ground plane and its respective dipole (of doubled length) radiating in open space (see Lecture 10).

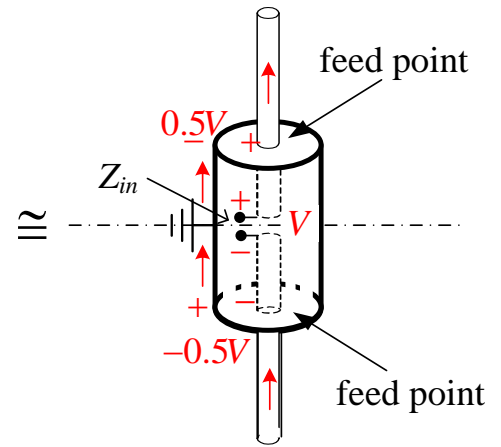
The first sleeve-dipole resonance occurs at a length $l_m \approx \lambda / 4$. The other important design variable is the monopole-to-sleeve ratio $\eta = (l_m - h) / h$. It has been experimentally established that $\eta = 2.25$ yields optimum (nearly constant with frequency) radiation patterns over a 4:1 band. The value of η has little effect on the radiation pattern if $l_m \leq \lambda / 2$, since the current on the outside of the sleeve has approximately the same phase as that on the top portion of the monopole. However, for longer lengths, the ratio η has notable effect on the pattern since the current on the outside of the sleeve is not necessarily in phase with that on the top portion of the monopole. Some practical sleeve dipoles and monopoles are shown below.



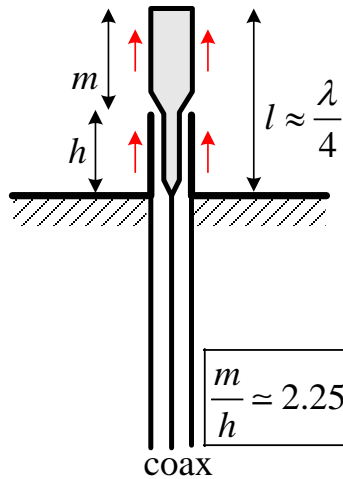
(a) asymmetrically-fed sleeve dipole



(b) sleeve monopole



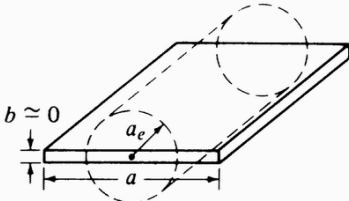
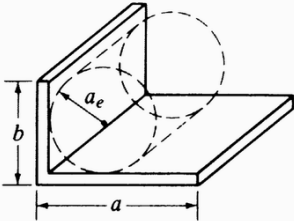
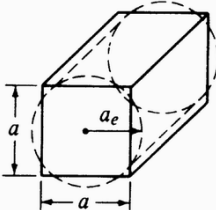
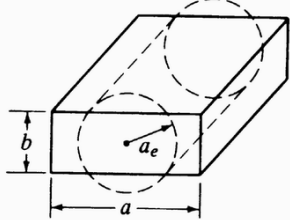
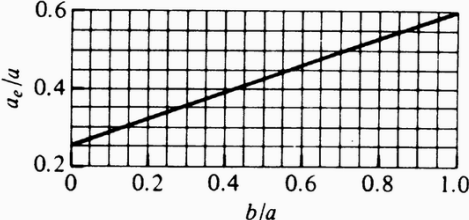
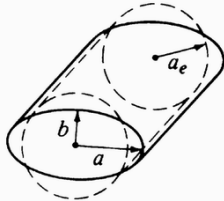
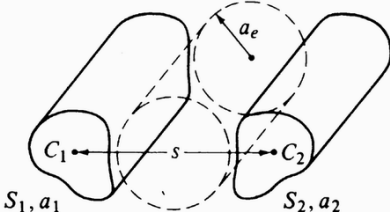
(c) sleeve dipole equivalent to monopole in (b)



(d) another sleeve monopole

So far, we have assumed that the cross-section of the wire is circular of radius a . An electrically equivalent radius can be obtained for some uniform wires of non-circular cross-sections. This is helpful when calculating the impedance of dipoles made of non-circular wires. The equivalent radii for certain wires are given below.

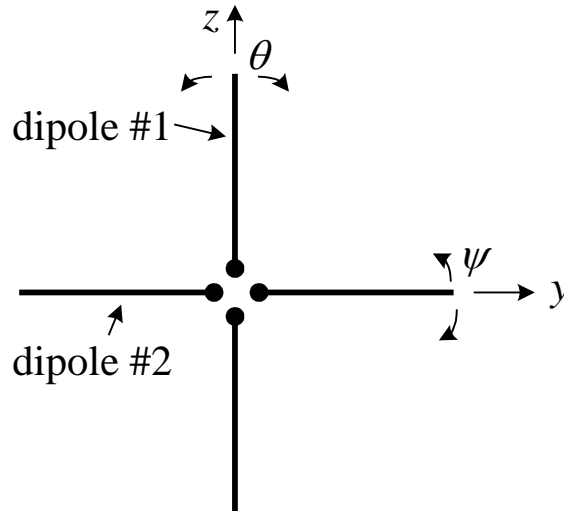
Table 9.3 CONDUCTOR GEOMETRICAL SHAPES AND THEIR EQUIVALENT CIRCULAR CYLINDER RADII

Geometrical Shape	Electrical Equivalent Radius
	$a_e = 0.25a$
	$a_e \approx 0.2(a + b)$
	$a_e = 0.59a$
	
	$a_e = \frac{1}{2}(a + b)$
	$\ln a_e \approx \frac{1}{(S_1 + S_2)^2} \times [S_1^2 \ln a_1 + S_2^2 \ln a_2 + 2S_1 S_2 \ln s]$ <p>S_1, S_2 = peripheries of conductors C_1, C_2 a_1, a_2 = equivalent radii of conductors C_1, C_2</p>

[Balanis]

4. Turnstile Antenna

The *turnstile* antenna is a combination of two orthogonal dipoles fed in phase-quadrature. This antenna is capable of producing circularly polarized field in the direction, which is normal to the dipoles' plane. It produces an isotropic pattern in the dipoles' plane (the θ -plane) of linearly (along $\hat{\theta}$) polarized wave. In all other directions, the wave is elliptically polarized.



In the $\varphi = 90^\circ, 270^\circ$ plane (the yz plane in which the dipoles lie), the field is a superposition of the separate dipole fields the patterns of which as a function of time are

$$\bar{E}_\theta^{(1)}(t) = \sin \theta \cos \omega t, \quad (11.25)$$

$$\bar{E}_\psi^{(2)}(t) = \sin \psi \cdot \cos(\omega t \pm \pi / 2) = \pm \sin(\pi / 2 - \theta) \cdot \sin \omega t = \pm \cos \theta \cdot \sin \omega t. \quad (11.26)$$

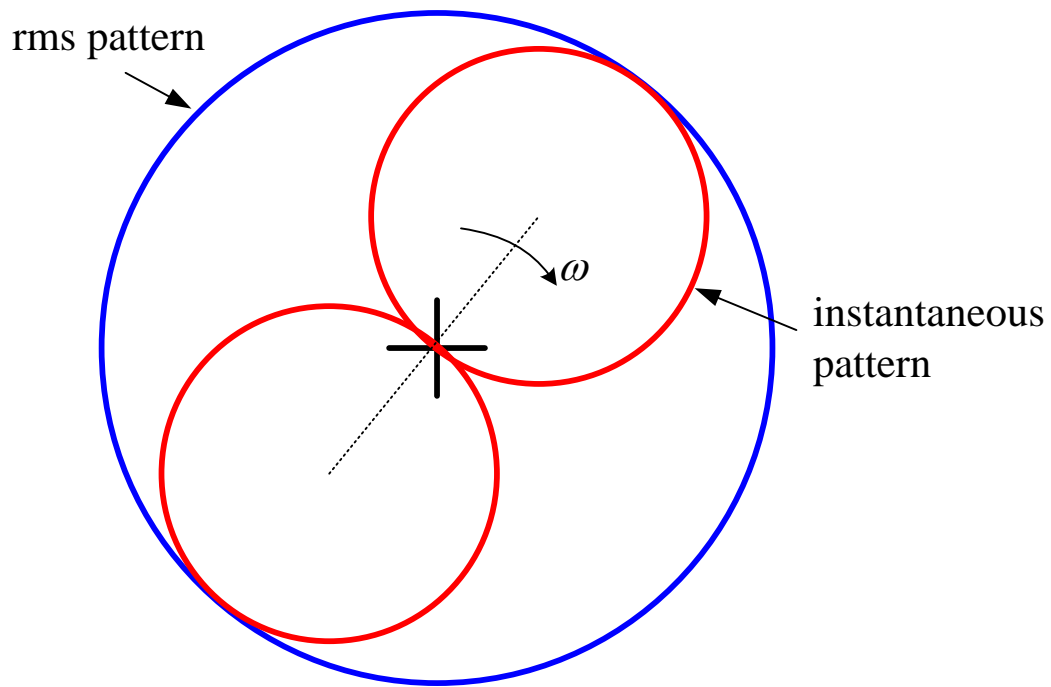
In the yz plane, the ψ -component of a vector is actually a θ -component. Equations (11.25) and (11.26) define the total field as

$$\bar{E}_\theta(t, \theta) = \sin \theta \cos \omega t \pm \cos \theta \sin \omega t, \quad (11.27)$$

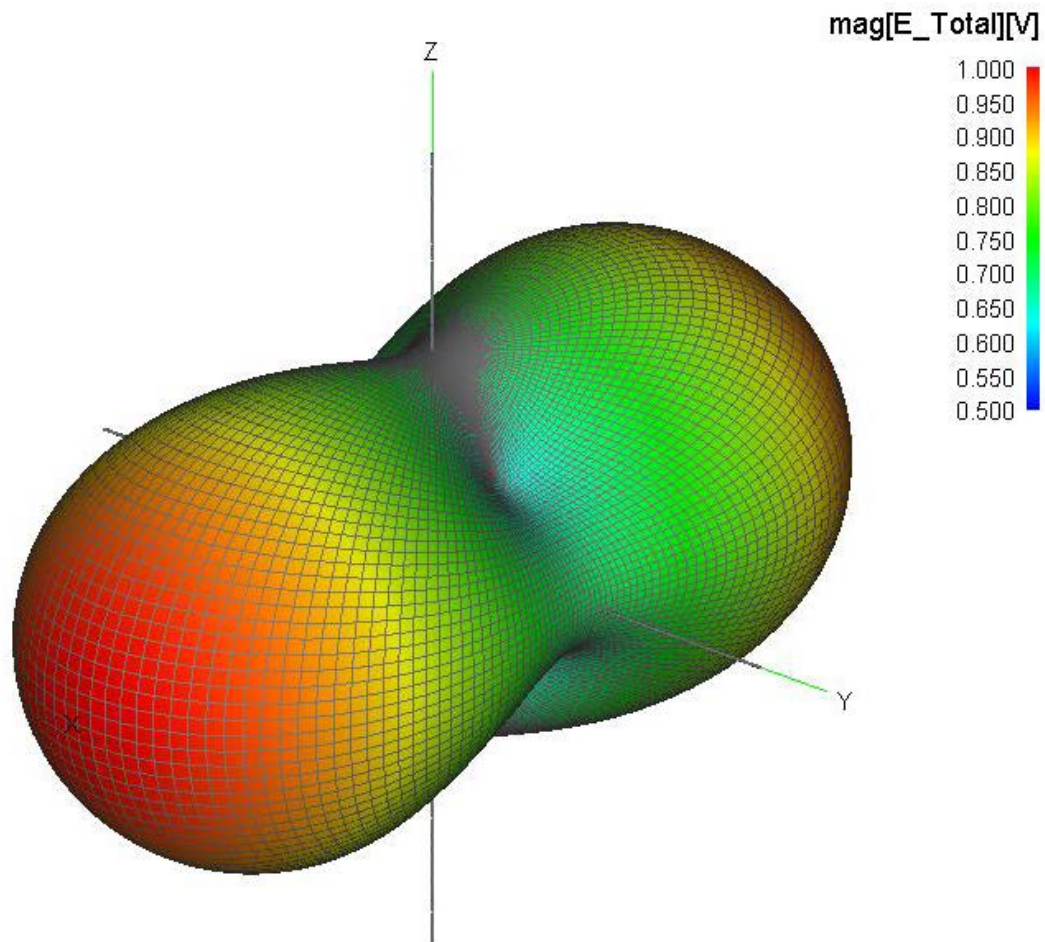
which reduces to

$$\bar{E}_\theta(t, \theta) = \sin(\theta \pm \omega t). \quad (11.28)$$

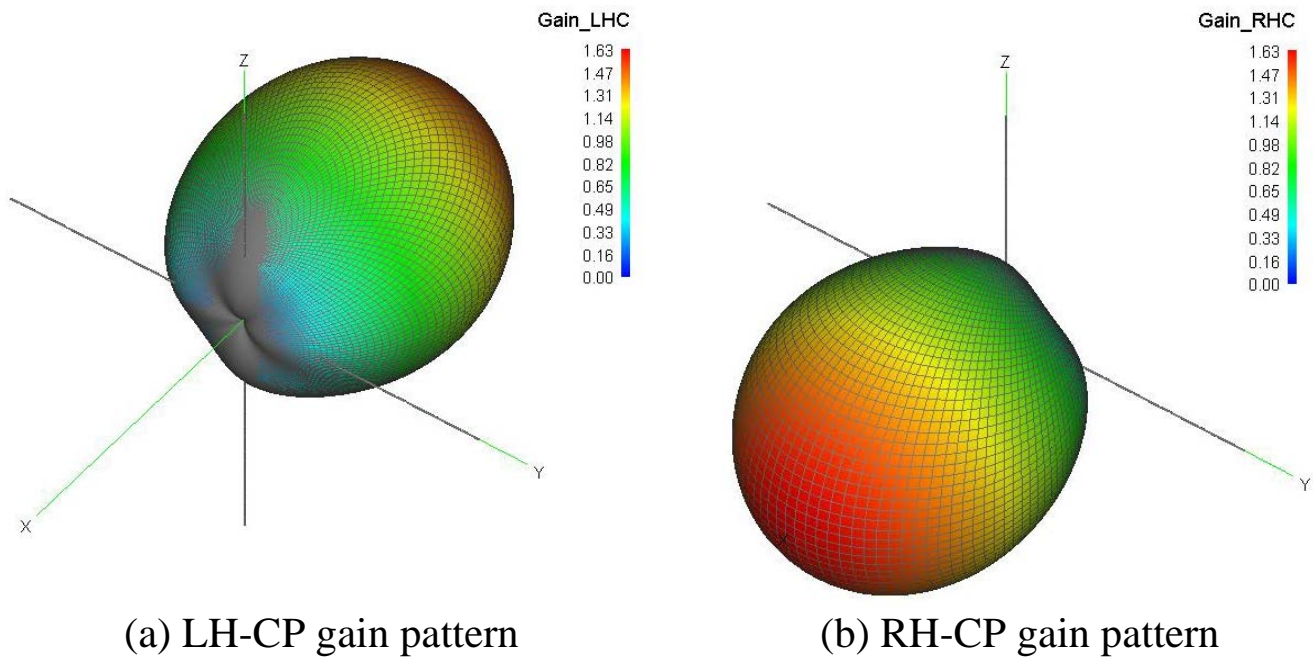
The *rms* pattern is circular, although the instantaneous pattern rotates.



3-D Turnstile Total Field Magnitude Pattern:

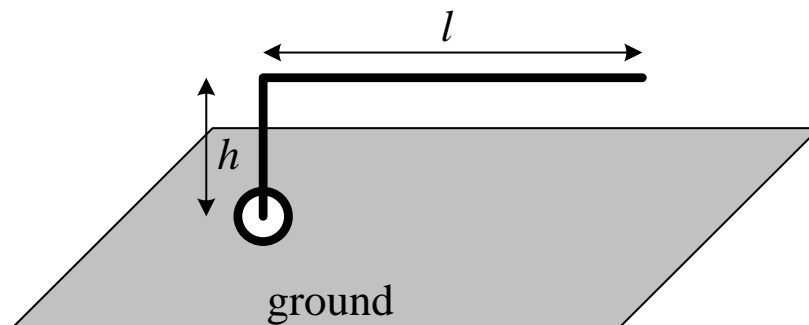


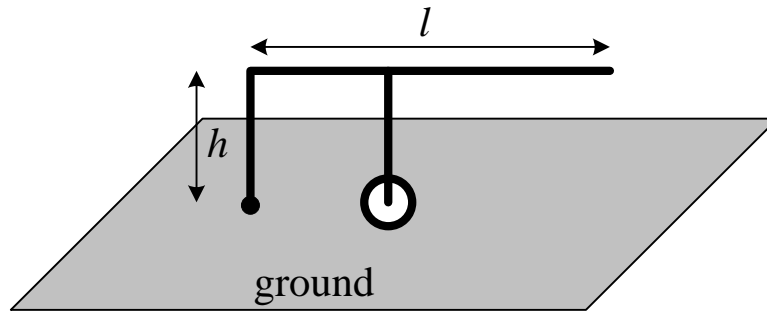
3D Gain Patterns of Trunstile ($V_z = 1, V_y = 1e^{j\pi/2}$)



5. Inverted-L and Inverted-F Antennas

The L-antenna (or inverted-L antenna) can be viewed as a bent monopole. The bending of the monopole results in a reduced size and low profile. In fact the first designs were made for missile applications (R. King *et al.*, “Transmission-line missile antennas,” *IRE Trans. Antennas Propagat.*, Jan. 1960, pp. 88-90). The reduction of the monopole’s height results in reduced radiation resistance and bandwidth. Besides, the main angle of radiation is depressed as there is substantial radiation not only from the monopole h but also from the arm l . A popular amateur-radio antenna is the $\lambda/4$ inverted-L where $h = l = \lambda/8$.





The inverted-F antenna uses a shifted feed point (a “tap”) along the bent arm l , to obtain better impedance match (offset feed).

Both the inverted-L and the inverted-F antennas can be analyzed using equivalent transmission line models. Their patterns in both principal planes are not much different from those of a monopole.

A popular antenna for mobile handsets is the planar inverted-F (PIFA) or its variations:

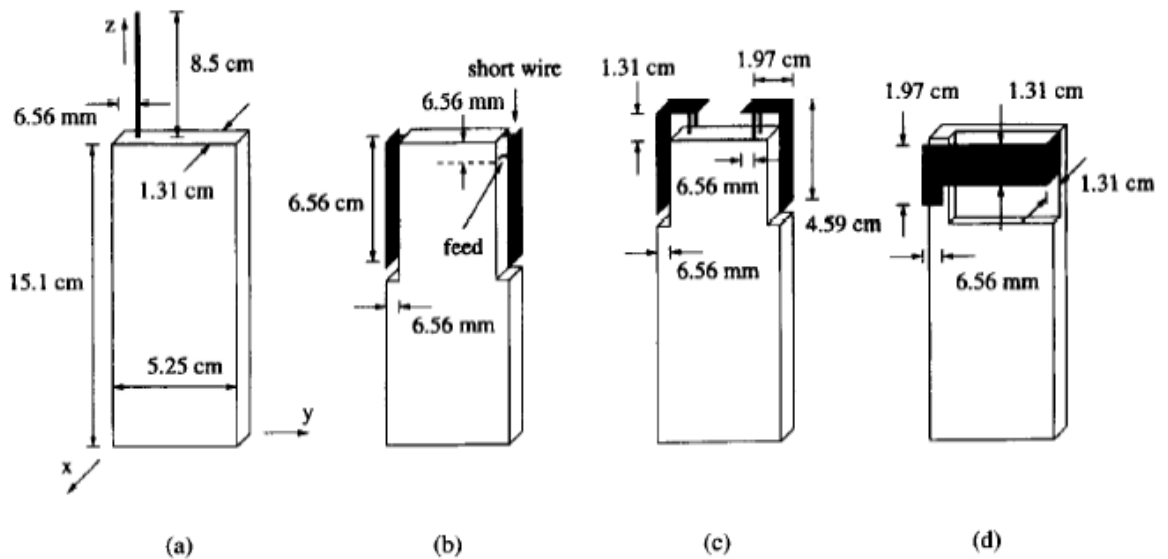


Fig. 1. Antenna geometries and dimensions for hand-held transceiver units: (a) monopole; (b) side-mounted dual PIFA; (c) top-mounted dual BIFA; (d) back-mounted PIFA. The chassis dimensions (105 cm^3) shown in (a) apply to all four configurations.

Figure from M.A. Jensen *et al.*, “EM interaction of handset antennas and a human in personal communications,” *Proceedings of the IEEE*, vol. 83 , No. 1, Jan. 1995, pp. 7-17. (Note: BIFA stands for *bent inverted F antenna*)

6. Matching Techniques for Wire Antennas

There are two major issues when constructing the feed circuit: impedance matching and balanced-unbalanced matching.

6.1. Impedance matching

$$\text{VSWR} = \frac{1 + |\Gamma|}{1 - |\Gamma|}. \quad (11.29)$$

The reflected-to-incident power ratio is given by $|\Gamma|^2$ (where Γ is the reflection coefficient). In terms of the VSWR, this is

$$|\Gamma|^2 = \left(\frac{\text{VSWR} - 1}{\text{VSWR} + 1} \right)^2. \quad (11.30)$$

The transmitted-to-incident power ratio is given by

$$|T|^2 = 1 - |\Gamma|^2. \quad (11.31)$$

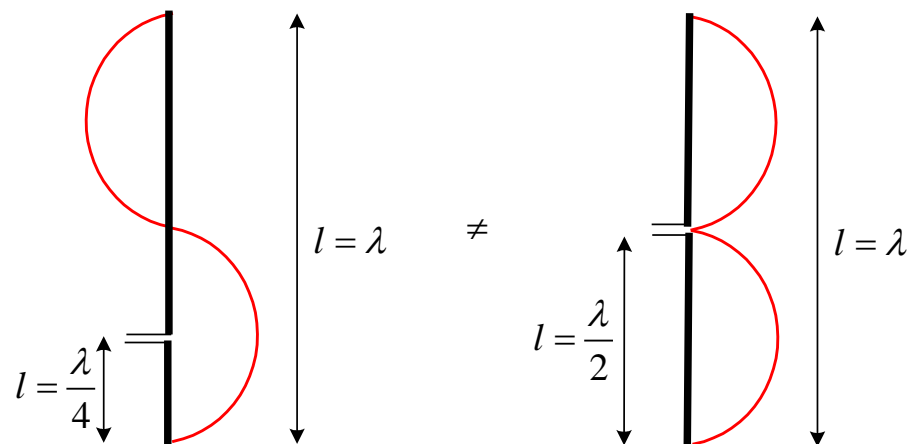
Impedance mismatch is undesirable not only because of the inefficient power transfer. In high-power transmitting systems, high VSWR leads to maxima of the standing wave which can cause arcing. Sometimes, the frequency of the transmitter can be affected by severe impedance mismatch (“frequency pulling”). Excessive reflections can damage the amplifying stages in the transmitter.

TABLE: VSWR AND TRANSMITTED POWER

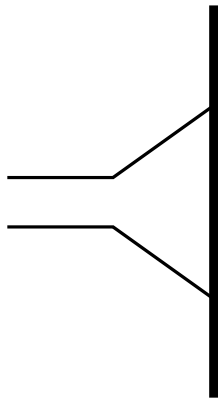
VSWR	$ \Gamma ^2 \times 100\%$	$ T ^2 \times 100\%$
1.0	0.0	100.0
1.1	0.2	99.8
1.2	0.8	99.2
1.5	4.0	96.0
2.0	11.1	88.9
3.0	25.0	75.0
4.0	36.0	64.0
5.0	44.4	55.6
5.83	50.0	50.0
10.0	66.9	33.1

A common way to find the proper feed location along a dipole or monopole is to feed off-center, which provides increase of the input impedance with respect to the center-feed impedance according to (11.14) and (11.15). For example, the input resistance of a center-fed half-wavelength dipole is approximately 73Ω , which is well suited for a $75\text{-}\Omega$ coaxial line if proper care is taken of the balanced-to-unbalanced transition. However, it is not well matched to a $300\text{-}\Omega$ antenna cable where we do not have to worry about a balanced-to-unbalanced transition since both the antenna and the cable are balanced. Greater values of the dipole input impedance (close to 300Ω) are easily achieved by moving the feed off center. Similarly, the quarter-wavelength monopole has an input resistance of approximately 37Ω and usually the sleeve-type of feed is used to achieve greater values of the antenna input impedance such as 50Ω to achieve better impedance match to a $50\text{-}\Omega$ coaxial cable.

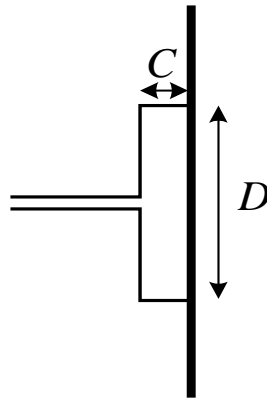
As a word of caution, the off-center feed is not symmetrical and can lead to undesirable phase reversal in the antenna if $l > \lambda/2$. This may profoundly change the radiation pattern.



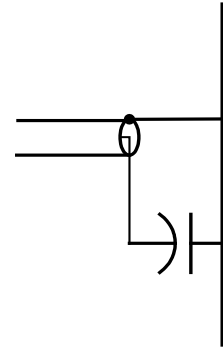
To avoid current phase reversal, symmetrical feeds for increased impedance are used. A few forms of shunt matching (or shunt feed) are shown below:



(a) Delta match



(b) Tee match



(c) Gamma match

We explain the principles of operation of the T-match only, which is the simplest of all to design and which gives the basic idea for all shunt feeds. Similarly to the folded-dipole analysis, the T-match interconnection together with the antenna can be viewed as two shorted transmission lines (in TL mode of operation) and a dipole (in an antenna mode of operation), which is longer than the two shorted TLs. The shorted TLs are less than quarter-wavelength long, and, therefore, they have an inductive reactance. This reactance is usually greater than the capacitive reactance of the dipole and an additional tuning lumped capacitor might be necessary to achieve better match. As the distance D increases, the input impedance increases (current magnitude drops). It has a maximum at about $D = l / 2$ (half the dipole's length). Then, it starts decreasing again, and when $D = l$, it equals the folded-dipole input impedance. In practice, sliding contacts are made between the shunt arms and the dipole for impedance adjustments and tuning. Note that shunt matches radiate and may alter the patterns and the directivity.

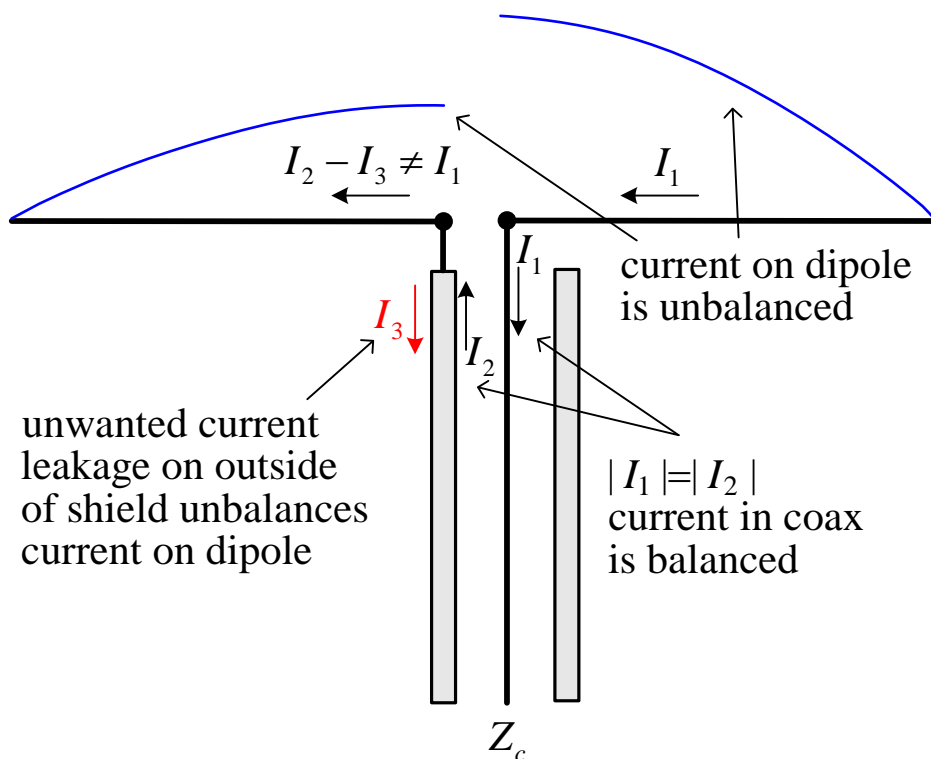
The Gamma-match is essentially the same as the T-match, only that it is designed for unbalanced-balanced connection.

Additional matching devices are sometimes used such as quarter-wavelength impedance transformers, reactive stubs for compensating antenna reactance, etc. These devices are well studied and described in Microwave Engineering courses.

6.2. Balanced-to-unbalanced feed

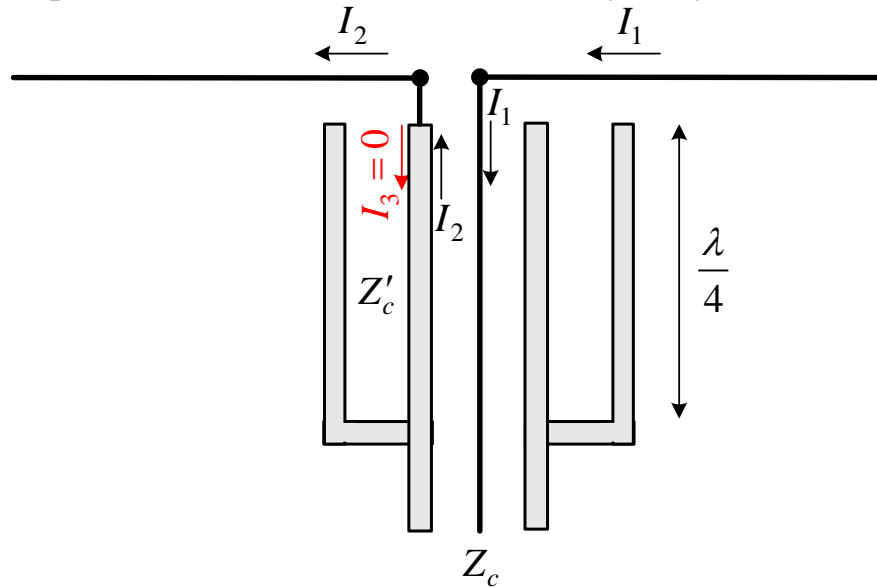
Sometimes, when high-frequency devices are connected, their impedances may be well matched and still we may observe significant reflections. This is sometimes referred to as “field mismatch.”

A typical example in antennas is the interconnect between a coaxial line of $Z_c = 75 \ \Omega$ and a half-wavelength dipole of $Z_{in} = 73 \ \Omega$. The reflections are much more severe than one would predict using equation (11.30). This is because the field and the current distributions in the coaxial line and at the input of the wire dipole are very different [see figure below]. The unequal currents on the dipole’s arms unbalance the antenna and the coaxial feed and induce currents on the outside of the coax shield which are the reason for parasitic radiation. To balance the currents, various devices are used, called **baluns** (balanced-to-unbalanced transformer).



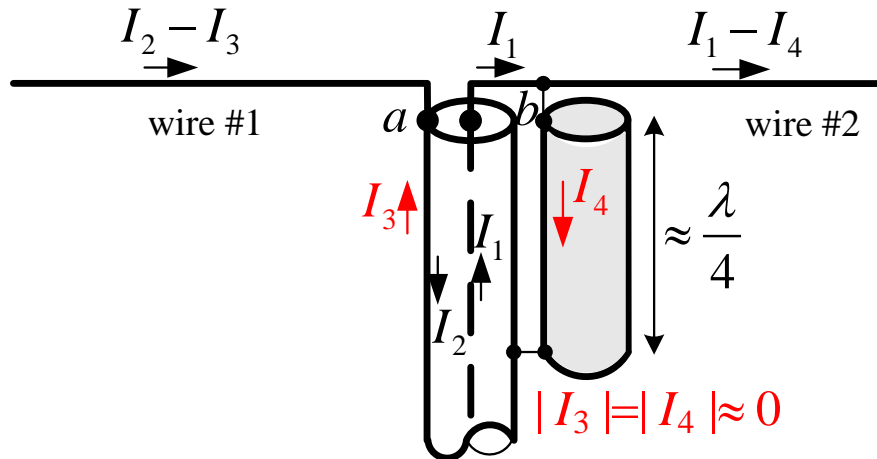
A. Sleeve (bazooka) balun 1:1

The sleeve and the outer conductor of the coaxial feed form another coaxial line, which has a characteristic impedance of Z'_c . This line is shorted quarter-wavelength away from the antenna input terminals. Thus, its input impedance is very large and results in: (i) suppression of the currents on the outer shield (I_3), and (ii) no interference with the antenna input impedance, which is in parallel with respect to the coaxial feed. This is a narrowband balun, which does not transform the impedance (1:1 balun). It is not very easy to construct.



B. Folded balun 1:1 (split-coax balun, $\lambda / 4$ -coax balun)

This 1:1 balun is easier to make. It is also narrowband.

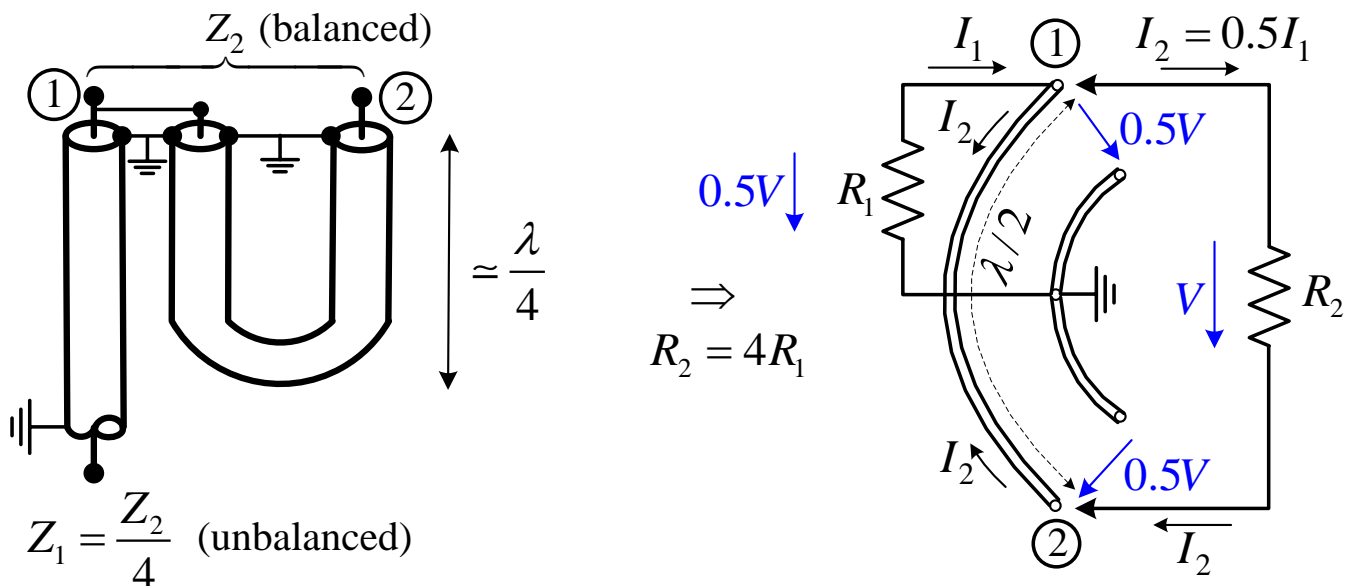


The outer shields of the feeding coaxial line and the additional coax-line section form a twin-lead transmission line, shorted a distance $\approx \lambda/4$ away from the antenna input. This line is in parallel with the antenna but does not affect the overall impedance because it has infinite impedance at the antenna terminals. The additional piece of coaxial line re-directs a portion of the I_1 current, which induces the twin-lead current I_4 . The currents I_3 and I_4 are well balanced ($I_3 = I_4$) because the current of wire #1 ($I_2 - I_3$) would induce as much current at the outer coaxial shield I_3 , as the current of wire #2 ($I_1 - I_4$) would induce in the outer shield of the auxiliary coaxial piece I_4 . This is due to the structural similarity of the two interconnects; see nodes (a) and (b) in the Figure. Thus,

$$\frac{I_3}{I_2 - I_3} = \frac{I_4}{I_1 - I_4}.$$

Since $I_1 = I_2$ in the feeding coaxial line, it is also true that $I_3 = I_4$. Thus, the current at the outer coaxial shield is effectively canceled from a certain point on ($\approx \lambda/4$).

C. Half-wavelength coaxial balun 1:4



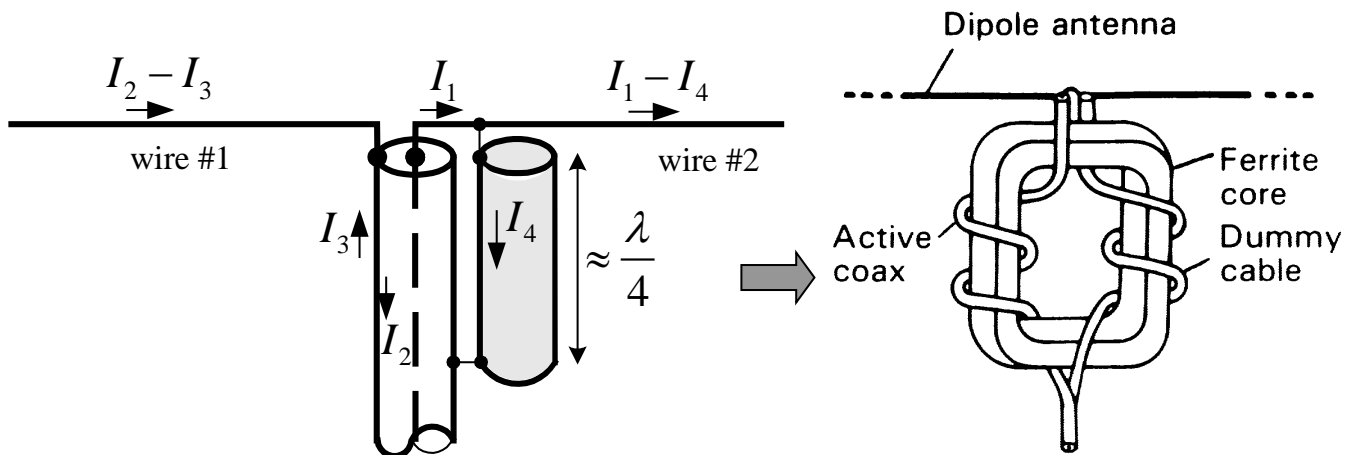
Typically, a coaxial feed of $Z_c = 75 \Omega$ would be connected with such a balun to a folded dipole of $Z_A \approx 292 \Omega$ [see equation (11.11)]. The principle of operation is explained by the equivalent circuit on the right. The auxiliary

piece of coaxial line ($\lambda/2$ long) transforms the input voltage at terminal 1 to a voltage of the same magnitude but opposite polarity at terminal 2. It also splits the input current into two equal parts (why?). Thus, the load “sees” twice larger voltage and twice smaller current compared to the case without balun. Hence the 4-fold increase in impedance.

D. Broadband baluns

All baluns described above are narrowband because of the dependence on the wavelength of the auxiliary transmission-line sections. Broadband baluns for high-frequency applications can be constructed by tapering a balanced transmission line to an unbalanced one gradually over a distance of several wavelengths (microstrip-to-twin-lead, coax-to-twin-lead).

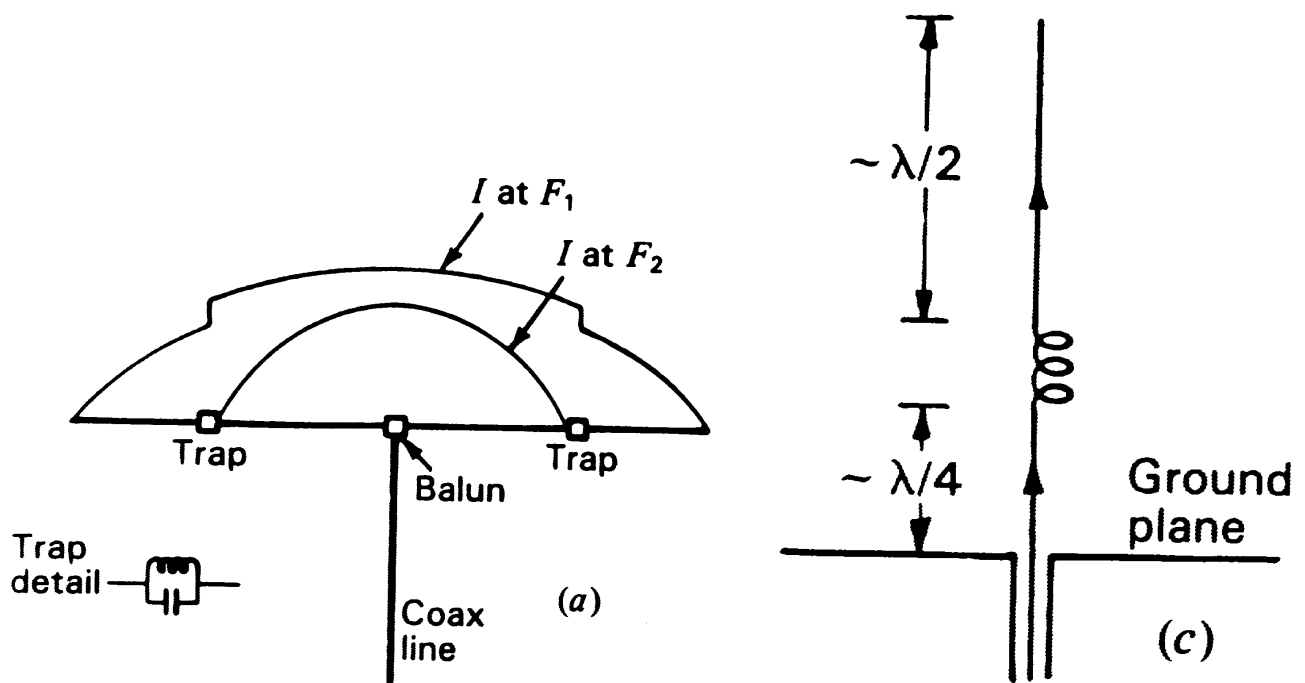
At lower frequencies (below UHF), tapered baluns are impractical, and transformers are used for impedance adjustment and balancing the feed. Often, ferrite-core bifilar wound-wire baluns are preferred for their small dimensions and broadband characteristics (bandwidths of 10:1 are achievable). A ferrite-core transformer 1:1, which is equivalent to the folded balun 1:1, but is much more broadband, is shown below.

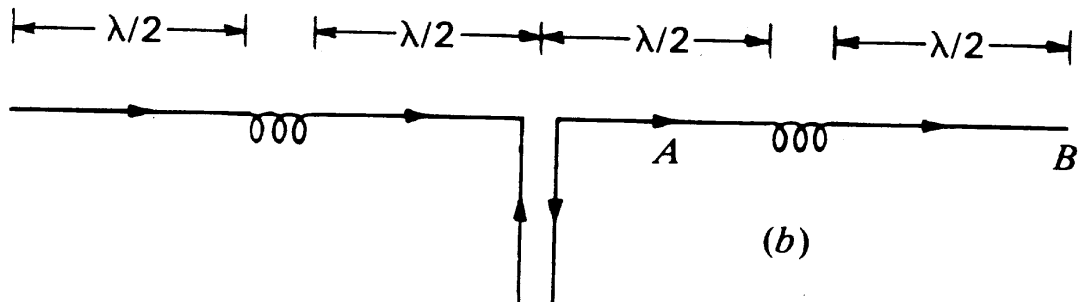


The transmission line formed by the outer shields of the two coaxial cables is now a very high-impedance line because of the high relative permeability of the ferrite core. Thus, its length does not depend critically on λ , in order not to disturb the antenna input impedance.

7. Dipoles with Traps

In many wideband applications, it is not necessary to have frequency-independent antennas (which are often expensive) but rather antennas that can operate at two (or more) different bands. Typical example is the type of multi-band antennas in cellular communication systems. A dual-band antenna can be constructed from a single center-fed dipole (or its respective monopole) by means of tuned traps [see figure (a) below]. Each trap is a tuned parallel LC circuit. At frequency f_1 , for which the whole dipole is $\approx \lambda / 2$ long, the trap is typically an inductor. This reduces slightly the resonant length of the dipole, and has to be taken into account. At another frequency $f_2 > f_1$, the traps become resonant and effectively cut off the outer portions of the dipole, making the dipole much shorter and resonant at this new frequency. If the traps, for example, are in the middle of the dipole arms, then $f_2 = 2f_1$ and the antenna can operate equally well at two frequencies separated by an octave. It should be noted that the isolation of the outer portions of the dipole depends not only on the high impedance of the traps but also on the impedance of this outer portion. When the outer portions are about $\lambda / 4$ long, they have very low impedance compared to the trap's impedance and are effectively mismatched, i.e., their currents are negligible. This is not the case if the outer portions are $\lambda / 2$ long.





[Figure from Kraus, *Antennas*, 2nd ed., p. 744]

When the outer portions of the dipole are about $\lambda/2$ each, they represent very high impedance themselves in series with the trap. They are no longer isolated and support strong current. The trap, however, introduces phase reversal which is essential in keeping the current in phase along the whole dipole.

A coil alone can form a trap at certain (very high) frequency because of its own distributed capacitance. It can also act as a 180° phase shifter (the coil may be viewed as a coiled-up $\lambda/2$ element). The use of this property is illustrated in Figure (b) which shows how one can construct an array of 4 in-phase $\lambda/2$ -elements with a single feed and achieve a gain of 6.4 dBi. Figure (c) shows the $3\lambda/4$ monopole, which is obtained from the dipole in (b) by cutting the dipole at point A, and mounting it above a ground plane. This is a common antenna for wireless phones and handsets. Its gain can be as high as 8.3 dBi and it has an input resistance of $\approx 150 \Omega$.

LECTURE 12: Loop Antennas

(Radiation parameters of a small loop. Circular loop of constant current. Equivalent circuit of the loop antenna. The small loop as a receiving antenna. Ferrite loops.)

1. Introduction

Loop antennas feature simplicity, low cost and versatility. They may have various shapes: circular, triangular, square, elliptical, etc. They are widely used in communication links up to the microwave bands (up to ≈ 3 GHz). They are also used as electromagnetic (EM) field probes in the microwave bands. Electrically small loops are widely used as compact transmitting and receiving antennas in the low MHz range (3 MHz to 30 MHz, or wavelengths of about 10 m to 100 m).

Loop antennas are usually classified as electrically small ($C < \lambda / 3$) and electrically large ($C \sim \lambda$). Here, C denotes the loop's circumference.

The small loops of a single turn have small radiation resistance ($< 1 \Omega$) usually comparable to their loss resistance. Their radiation resistance, however, can be improved by adding more turns. Also, the small loops are narrowband. Typical bandwidths are less than 1%. However, clever impedance matching can provide low-reflection transition from a coaxial cable to a loop antenna with a tuning frequency range as high as 1:10.¹ Moreover, in the HF and VHF bands where the loop diameters are on the order of a half a meter to several meters, the loop can be made of large-diameter tubing or coaxial cable, or wide copper tape, which can drastically reduce the loss.

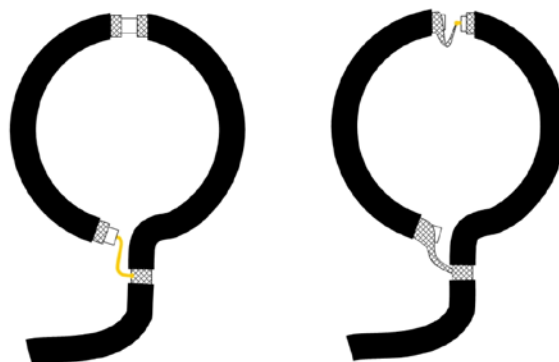


Fig. 1: Shielded Faraday loops used to inductively feed electrically small loop antennas. [©2012, Frank Dörenberg, used with permission; see https://www.nonstopsystems.com/radio/frank_radio_antenna_magloop.htm. Additional resource: L. Turner VK5KLT, “An overview of the underestimated magnetic loop HF antenna,”]

¹ John H. Dunlavy Jr., US Patent 13,588,905: “Wide range tunable transmitting loop antenna”, 1967.

The small loops, regardless of their shape, have a far-field pattern very similar to that of a small electric dipole normal to the plane of the loop. This is expected because they are equivalent to a magnetic dipole. Note, however, that the field polarization is orthogonal to that of the electric dipole.

As the circumference of the loop increases, the pattern maximum shifts towards the loop's normal, and when $C \approx \lambda$, the maximum of the pattern is along the loop's normal.

2. Radiation Characteristics of a Small Loop

A small loop is by definition a loop of constant current. Its radius satisfies

$$a < \frac{\lambda}{6\pi}, \quad (12.1)$$

or, equivalently, $C < \lambda / 3$. The limit (12.1) is mathematically derived later in this Lecture from the first-order approximation of the Bessel function of the first order $J_1(x)$ in the general solution for a loop of constant current. Actually, to make sure that the current has near-constant distribution along the loop, a tighter limit must be imposed:

$$a < 0.03\lambda, \quad (12.2)$$

or, $C < \lambda / 5$. A good approximate model of a small loop is provided by the infinitesimal loop (or the infinitesimal magnetic dipole).

The expressions for the field components of an infinitesimal loop of electric current of area A were already derived in Lecture 3. Here, we give only the far-field components of the loop, the axis of which is along z :

$$E_\varphi = \eta\beta^2 \cdot (IA) \cdot \frac{e^{-j\beta r}}{4\pi r} \cdot \sin \theta, \quad (12.3)$$

$$H_\theta = -\beta^2 \cdot (IA) \cdot \frac{e^{-j\beta r}}{4\pi r} \cdot \sin \theta. \quad (12.4)$$

It is obvious that the far-field pattern,

$$\bar{E}_\varphi(\theta) = \sin \theta, \quad (12.5)$$

is identical to that of a z -directed infinitesimal electric dipole although the polarization is orthogonal. The power pattern is identical to that of the infinitesimal electric dipole:

$$F(\theta) = \sin^2 \theta. \quad (12.6)$$

Radiated power:

$$\begin{aligned} \Pi &= \oint \frac{1}{2\eta} |E_\phi|^2 \cdot \underbrace{r^2 \sin \theta d\theta d\phi}_{ds}, \\ \Pi &= \frac{1}{12\pi} \eta \beta^4 (IA)^2. \end{aligned} \quad (12.7)$$

Radiation resistance:

$$R_r = \eta \frac{8}{3} \pi^3 \left(\frac{A}{\lambda^2} \right)^2. \quad (12.8)$$

In free space, $\eta = 120\pi \Omega$, and

$$R_r \approx 31171(A / \lambda^2)^2. \quad (12.9)$$

Equation (12.9) gives the radiation resistance of a single loop. If the loop antenna has N turns, then the radiation resistance increases with a factor of N^2 (because the radiated power increases as I^2):

$$R_r = \eta \frac{8}{3} \pi^3 \left(N \frac{A}{\lambda^2} \right)^2. \quad (12.10)$$

The relation in (12.10) provides a handy mechanism to increase R_r and the radiated power Π . Unfortunately, the losses of the loop antenna also increase (although only as $\sim N$) and this may result in low efficiency.

The directivity is the same as that of an infinitesimal dipole:

$$D_0 = 4\pi \frac{U_{\max}}{\Pi_{rad}} = 1.5. \quad (12.11)$$

3. Circular Loop of Constant Current – General Solution

So far, we have assumed that the loop is of infinitesimal radius a , which allows the use of the expressions for the infinitesimal magnetic dipole. Now, we derive the far field of a circular loop, which might not be necessarily very small, but still has constant current distribution. This derivation illustrates the general

loop-antenna analysis as the approach is used in the solutions to circular loop problems of nonuniform distributions, too.

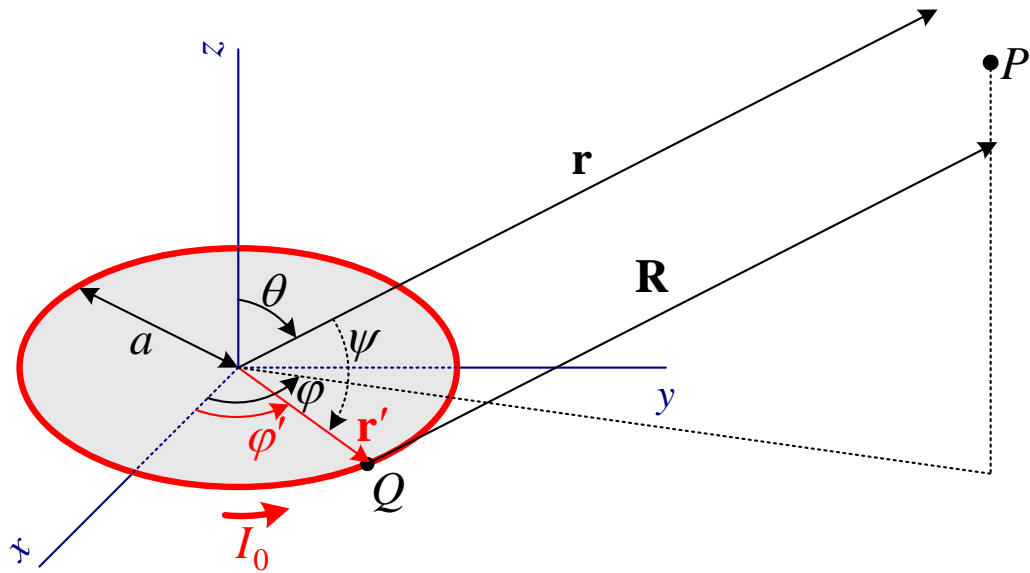
The circular loop can be divided into an infinite number of infinitesimal current elements. With reference to the figure below, the position of a current element in the xy plane is characterized by $0^\circ \leq \varphi' < 360^\circ$ and $\theta' = 90^\circ$. The position of the observation point P is defined by (θ, φ) .

The far-field approximations are

$$\left| \begin{array}{l} R \approx r - a \cos \psi, \text{ for the phase term,} \\ \frac{1}{R} \approx \frac{1}{r}, \text{ for the amplitude term.} \end{array} \right. \quad (12.12)$$

In general, the solution for \mathbf{A} does not depend on φ because of the cylindrical symmetry of the problem. Here, we set $\varphi = 0$. The angle between the position vector of the source point Q and that of the observation point P is determined as

$$\cos \psi = \hat{\mathbf{r}} \cdot \hat{\mathbf{r}}' = (\hat{\mathbf{x}} \sin \theta \cos \varphi + \hat{\mathbf{y}} \sin \theta \sin \varphi + \hat{\mathbf{z}} \cos \theta) \cdot (\hat{\mathbf{x}} \cos \varphi' + \hat{\mathbf{y}} \sin \varphi'),$$



$$\Rightarrow \cos \psi = \sin \theta \cos \varphi'. \quad (12.13)$$

Now the vector potential integral can be solved for the far zone:

$$\mathbf{A}(r, \theta, \varphi) = \frac{\mu}{4\pi} \oint_C I_0 \frac{e^{-j\beta(r - a \sin \theta \cos \varphi')}}{r} d\mathbf{l} \quad (12.14)$$

where $d\mathbf{l} = \hat{\boldsymbol{\phi}}' a d\varphi'$ is the linear element of the loop contour. The current element changes its direction along the loop and its contribution depends on the angle between its direction and the respective \mathbf{A} component. Since all current elements are directed along $\hat{\boldsymbol{\phi}}$, we conclude that the vector potential has only A_φ component, i.e., $\mathbf{A} = A_\varphi \hat{\boldsymbol{\phi}}$, where

$$A_\varphi(r, \theta, \varphi) = \hat{\boldsymbol{\phi}} \cdot \mathbf{A}(r, \theta, \varphi) = \frac{\mu}{4\pi} (I_0 a) \frac{e^{-j\beta r}}{r} \int_0^{2\pi} (\hat{\boldsymbol{\phi}} \cdot \hat{\boldsymbol{\phi}}') e^{j\beta a \sin \theta \cos \varphi'} d\varphi'. \quad (12.15)$$

Since

$$\begin{aligned} \hat{\boldsymbol{\phi}} \cdot \hat{\boldsymbol{\phi}}' &= (\hat{\mathbf{x}} \cos \varphi + \hat{\mathbf{y}} \sin \varphi) \cdot (\hat{\mathbf{x}} \cos \varphi' + \hat{\mathbf{y}} \sin \varphi') = \\ &= \cos \varphi \cos \varphi' + \sin \varphi \sin \varphi' = \\ &= \cos(\varphi - \varphi') \Big|_{\varphi=0} = \cos \varphi', \end{aligned} \quad (12.16)$$

the vector potential is

$$A_\varphi(\theta, 0) = \frac{\mu}{4\pi} (I_0 a) \frac{e^{-j\beta r}}{r} \int_0^{2\pi} \cos \varphi' \cdot e^{j\beta a \sin \theta \cos \varphi'} d\varphi', \quad (12.17)$$

$$A_\varphi(\theta) = \frac{\mu}{4\pi} (I_0 a) \frac{e^{-j\beta r}}{r} \left[\int_0^\pi \cos \varphi' \cdot e^{j\beta a \sin \theta \cos \varphi'} d\varphi' + \int_\pi^{2\pi} \cos \varphi' \cdot e^{j\beta a \sin \theta \cos \varphi'} d\varphi' \right].$$

We apply the following substitution in the second integral: $\varphi' = \varphi'' + \pi$. Then,

$$A_\varphi(\theta) = \frac{\mu I_0 a}{4\pi} \frac{e^{-j\beta r}}{r} \left[\int_0^\pi \cos \varphi' \cdot e^{j\beta a \sin \theta \cos \varphi'} d\varphi' - \int_0^\pi \cos \varphi'' \cdot e^{-j\beta a \sin \theta \cos \varphi''} d\varphi'' \right]. \quad (12.18)$$

The integrals in (12.18) can be expressed in terms of Bessel functions, which are defined as

$$\int_0^\pi \cos(n\varphi) e^{jz \cos \varphi} d\varphi = \pi j^n J_n(z). \quad (12.19)$$

Here, $J_n(z)$ is the Bessel function of the first kind of order n . From (12.18) and (12.19), it follows that

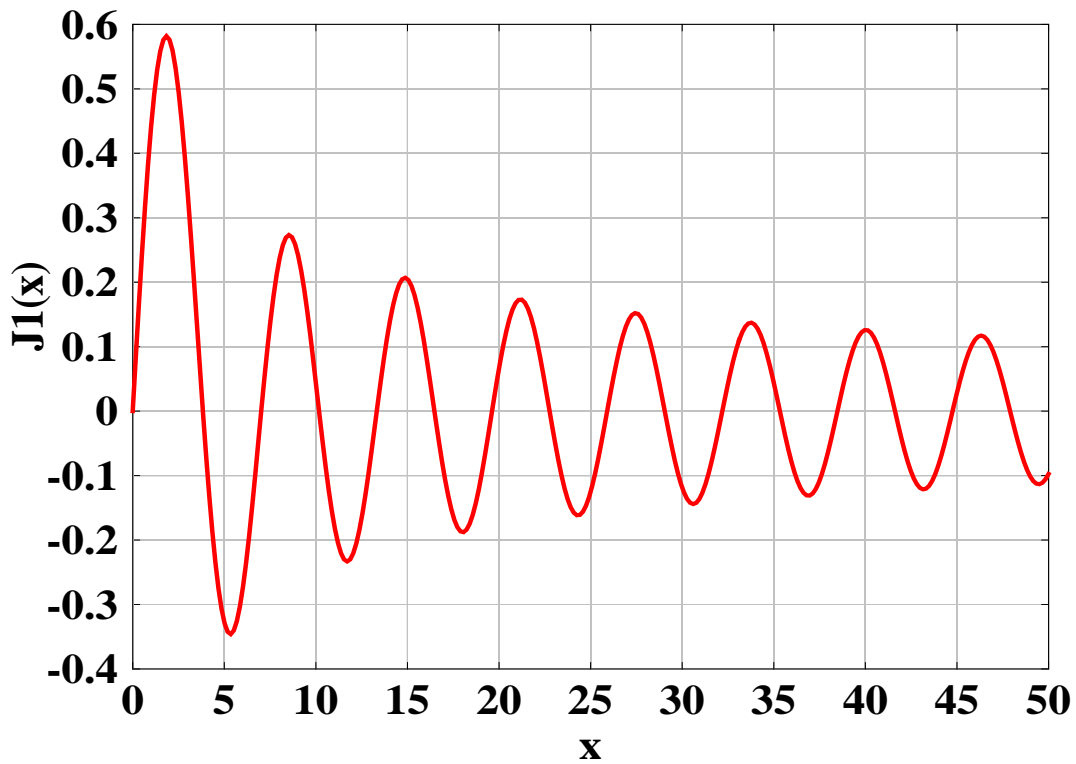
$$A_\varphi(\theta) = \frac{\mu}{4\pi} (I_0 a) \frac{e^{-j\beta r}}{r} \pi j \left[J_1(\beta a \sin \theta) - J_1(-\beta a \sin \theta) \right]. \quad (12.20)$$

Since

$$J_n(-z) = (-1)^n J_n(z), \quad (12.21)$$

equation (12.20) reduces to

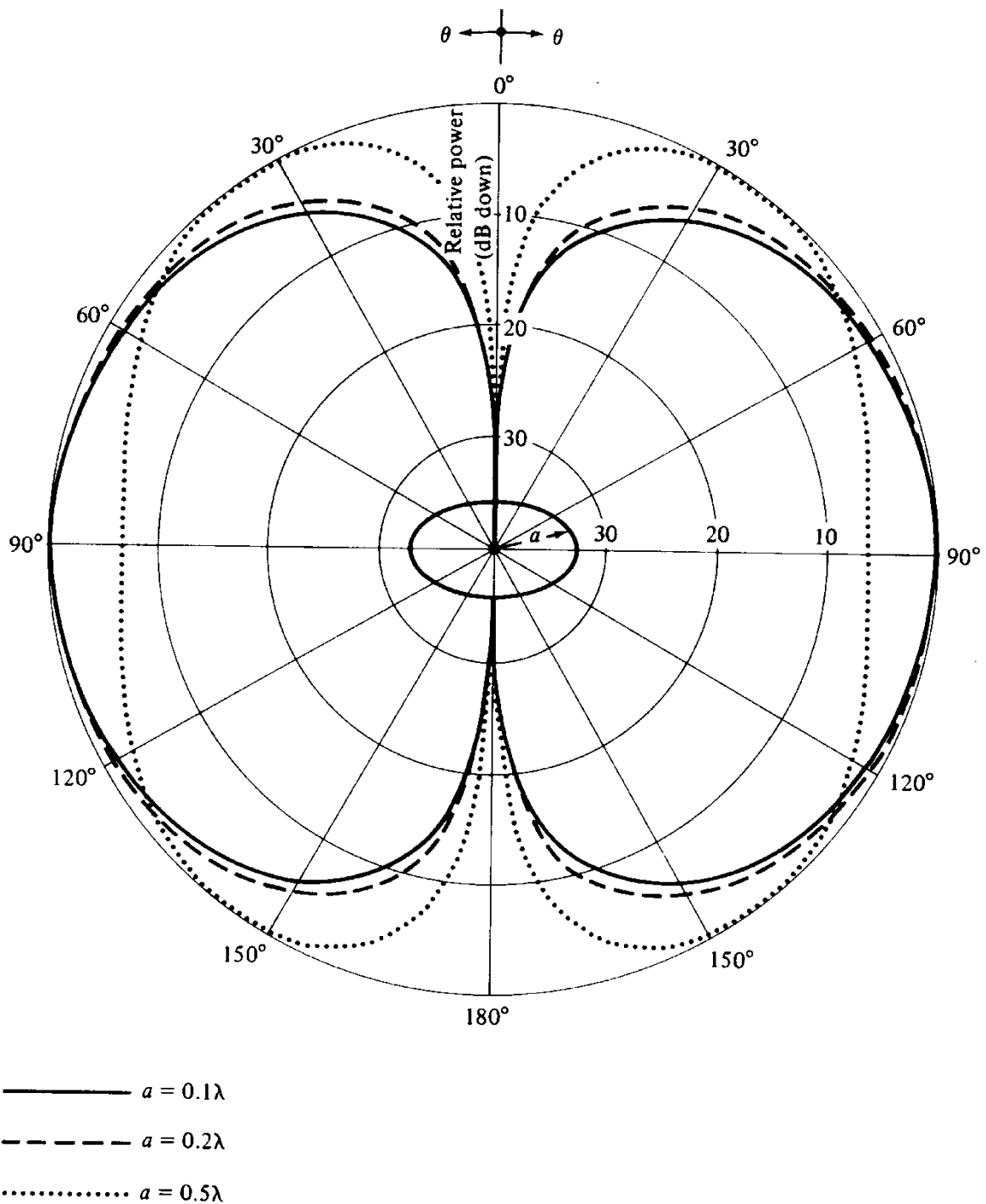
$$A_\varphi(\theta) = j \frac{\mu}{2} (I_0 a) \frac{e^{-j\beta r}}{r} J_1(\beta a \sin \theta). \quad (12.22)$$



The far-zone fields are derived as

$$\begin{cases} E_\varphi(\theta) = \beta \eta (I_0 a) \frac{e^{-j\beta r}}{2r} J_1(\beta a \sin \theta), \\ H_\theta(\theta) = -\frac{E_\varphi}{\eta} = -\beta (I_0 a) \frac{e^{-j\beta r}}{2r} J_1(\beta a \sin \theta). \end{cases} \quad (12.23)$$

The patterns of constant-current loops obtained from (12.23) are shown below:



[Balanis]

The small-loop field solution in (12.3)-(12.4) is actually a first-order approximation of the solution in (12.23). This becomes obvious when the Bessel function is expanded in series as

$$J_1(\beta a \sin \theta) = \frac{1}{2}(\beta a \sin \theta) - \frac{1}{16}(\beta a \sin \theta)^3 + \dots \quad (12.24)$$

For small values of the argument ($\beta a < 1/3$), the first-order approximation is acceptable, i.e.,

$$J_1(\beta a \sin \theta) \approx \frac{1}{2}(\beta a \sin \theta). \quad (12.25)$$

The substitution of (12.25) in (12.23) yields (12.3)-(12.4).

It can be shown that the maximum of the pattern given by (12.23) is in the direction $\theta = 90^\circ$ for all loops, which have circumference $C < 1.84\lambda$.

Radiated power and radiation resistance

We substitute the E_φ expression (12.23) in

$$\Pi = \oint\oint \frac{1}{2\eta} |E_\varphi|^2 \cdot \underbrace{r^2 \sin \theta d\theta d\varphi}_{ds},$$

which yields

$$\Pi = \frac{(\omega\mu)^2}{4\eta} (I_0^2 A) \cdot \int_0^\pi J_1^2(\beta a \sin \theta) \sin \theta d\theta. \quad (12.26)$$

Here, $A = \pi a^2$ is the loop's area. The integral in (12.26) does not have a closed form solution. Often, the following transformation is applied:

$$\int_0^\pi J_1^2(\beta a \sin \theta) \sin \theta d\theta = \frac{1}{\beta a} \int_0^{2\beta a} J_2(x) dx. \quad (12.27)$$

The second integral in (12.27) does not have a closed form solution either but it can be approximated with a highly convergent series:

$$\int_0^{2\beta a} J_2(x) dx = 2 \sum_{m=0}^{\infty} J_{2m+3}(2\beta a). \quad (12.28)$$

The radiation resistance is obtained as

$$R_r = \frac{2\Pi}{I_0^2} = \frac{(\omega\mu)^2}{2\eta} A \cdot \int_0^\pi J_1^2(\beta a \sin \theta) \sin \theta d\theta. \quad (12.29)$$

The radiation resistance of small loops is very small. For example, for $\lambda/100 < a < \lambda/30$ the radiation resistance varies from $\approx 3 \times 10^{-3} \Omega$ to $\approx 0.5 \Omega$.

This is often less than the loss resistance of the loop. That is why small loop antennas are constructed with multiple turns and on ferromagnetic cores. Such loop antennas have large inductive reactance, which is compensated by a capacitor. This is convenient in narrowband receivers, where the antenna itself is a very efficient filter (together with the tuning capacitor), which can be tuned for different frequency bands. Low-loss capacitors must be used to prevent further increase in the loss.

4. Circular Loop of Nonuniform Current

When the loop radius becomes larger than 0.2λ , the constant-current assumption does not hold. A common assumption is the cosine distribution.^{2,3} Lindsay, Jr.,⁴ considers the circular loop to be a deformation of a shorted parallel-wire line. If I_s is the current magnitude at the “shorted” end, i.e., the point opposite to the feed point where $\varphi' = \pi$, then

$$I(\alpha) = I_s \cosh(\gamma a \alpha) \quad (12.30)$$

where $\alpha = \pi - \varphi'$ is the angle with respect to the shorted end, γ is the line propagation constant and a is the loop radius. If we assume loss-free transmission-line model, then $\gamma = j\beta$ and $\cosh(\gamma a \alpha) = \cos(\beta a \alpha)$. For a loop in open space, β is assumed to be the free-space wave number ($\beta = \omega\sqrt{\mu_0\epsilon_0}$).

The cosine distribution is not very accurate, especially close to the terminals, and this has a negative impact on the accuracy of the computed input impedance. That is why the current is often represented by a Fourier series:^{5,6}

$$I(\varphi') = I_0 + 2 \sum_{n=1}^N I_n \cos(n\varphi'). \quad (12.31)$$

Here, φ' is measured from the feed point. This way, the derivative of the current distribution with respect to φ' at $\varphi' = \pi$ (the point diametrically opposite to the feed point) is always zero. This imposes the requirement for a symmetrical current distribution on both sides of the diameter from $\varphi' = 0$ to $\varphi' = \pi$. The

² E.A. Wolff, *Antenna Analysis*, Wiley, New York, 1966.

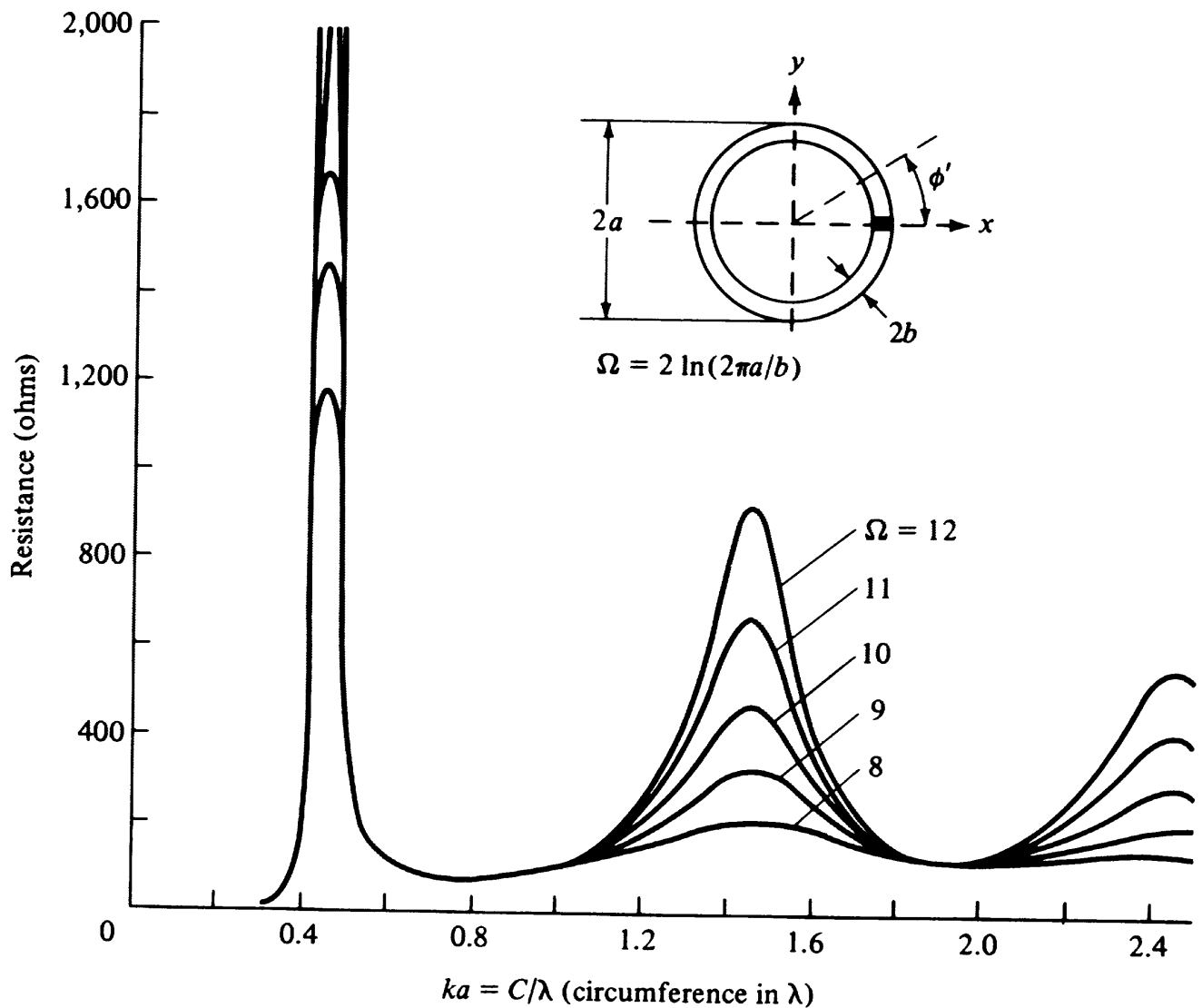
³ A. Richtscheid, “Calculation of the radiation resistance of loop antennas with sinusoidal current distribution,” *IEEE Trans. Antennas Propagat.*, Nov. 1976, pp. 889-891.

⁴ J. E. Lindsay, Jr., “A circular loop antenna with non-uniform current distribution,” *IRE Trans. Antennas Propagat.*, vol. AP-8, No. 4, July 1960, pp. 439-441.

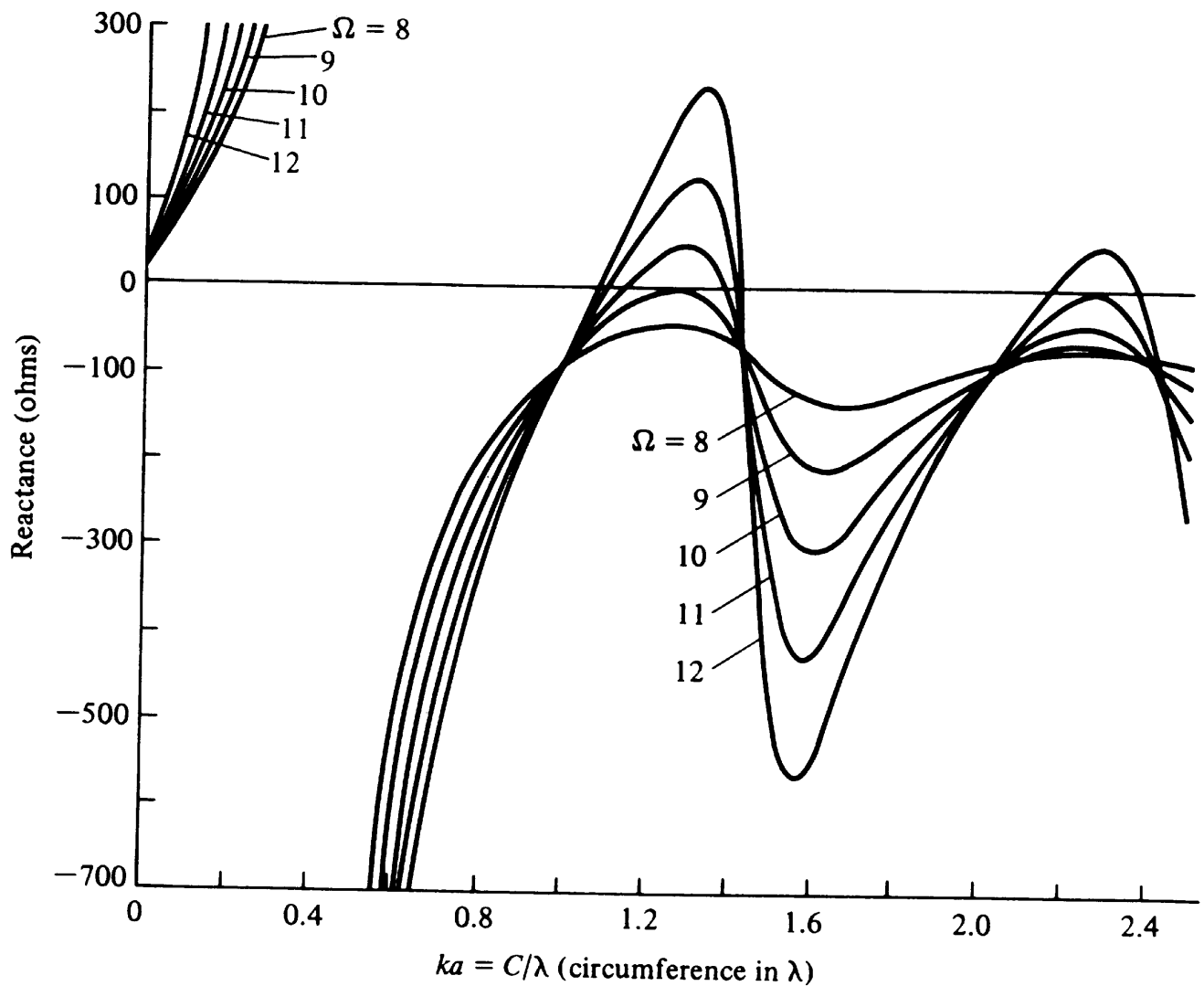
⁵ H. C. Pocklington, “Electrical oscillations in wire,” in *Cambridge Phil. Soc. Proc.*, vol. 9, 1897, pp. 324–332.

⁶ J. E. Storer, “Input impedance of circular loop antennas,” *Am. Inst. Electr. Eng. Trans.*, vol. 75, Nov. 1956.

complete analysis of this general case will be left out, and only some important results will be given. When the circumference of the loop approaches λ , the maximum of the radiation pattern shifts exactly along the loop's normal. Then, the input resistance of the antenna is also good (about 50 to 70 Ω). The maximum directivity occurs when $C \approx 1.4\lambda$ but then the input impedance is too large. The input resistance and reactance of the large circular loop are given below.



(a) Resistance



(b) Reactance

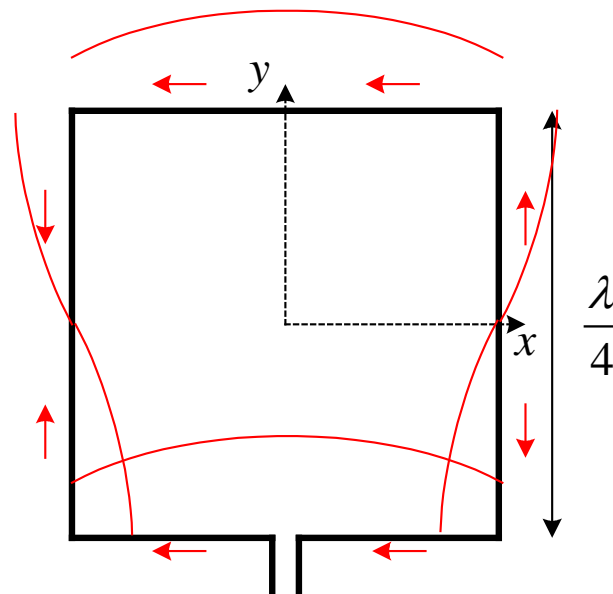
Figure 5.11 Input impedance of circular loop antennas. (SOURCE: J. E. Storer, "Impedance of Thin-Wire Loop Antennas," *AIEE Trans.*, Vol. 75, November 1956. © 1956 IEEE).

(Note: typo in author's name, read as J. E. Storer)

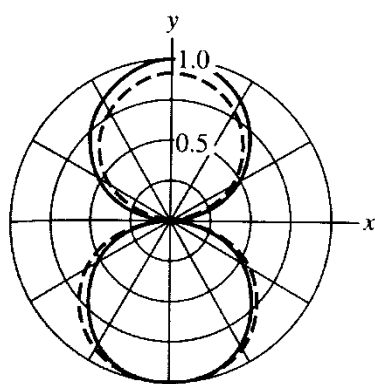
The large circular loop is very similar in its performance to the large square loop. An approximate solution of very good accuracy for the square-loop antenna can be found in

W.L. Stutzman and G.A. Thiele, *Antenna Theory and Design*, 2nd Ed., John Wiley & Sons, New York, 1998.

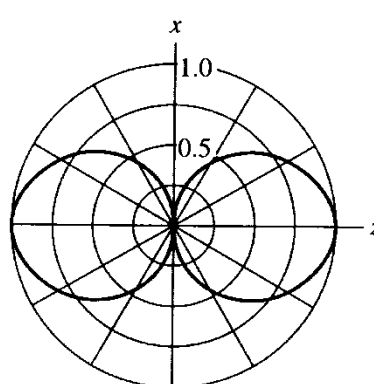
There, it is assumed that the total antenna loop is exactly one wavelength and has a cosine current distribution along the loop's wire.



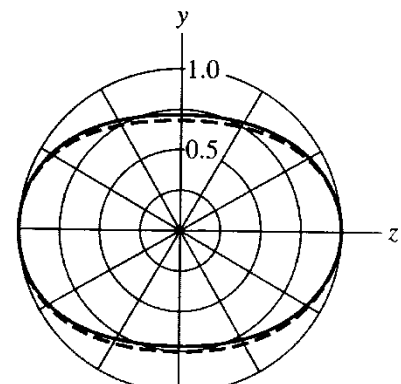
The principal plane patterns obtained through the cosine-current assumption (solid line) and using numerical methods (dash line) are shown below:



(a) The xy -plane (the plane of the loop and an E -plane) normalized pattern plot of E_ϕ . In this plane, $HP = 94^\circ$.

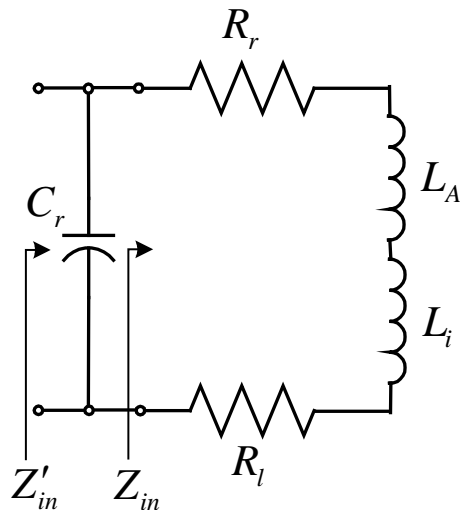


(b) The xz -plane (an E -plane) normalized pattern plot of E_θ . In this plane, $HP = 85^\circ$. The patterns from the two methods coincide in this plane.



(c) The yz -plane (the H -plane) pattern plot of E_ϕ .

5. Equivalent Circuit of a Loop Antenna



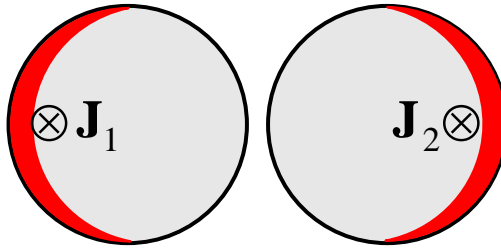
- C_r - resonance capacitor
- R_l - loss resistance of the loop antenna
- R_r - radiation resistance
- L_A - inductance of the loop
- L_i - inductance of the loop conductor (wire)

(a) Loss resistance

Usually, it is assumed that the loss resistance of loosely wound loop equals the high-frequency loss resistance of a straight wire of the same length as the loop and of the same current distribution. In the case of a uniform current distribution, the high-frequency resistance is calculated as

$$R_{hf} = \frac{l}{p} R_s, \quad R_s = \sqrt{\frac{\pi f \mu}{\sigma}}, \quad \Omega \quad (12.32)$$

where l is the length of the wire, and p is the perimeter of the wire's cross-section. We are not concerned with the current distribution now because it can be always taken into account in the same way as it is done for the dipole/monopole antennas. However, another important phenomenon has to be taken into account, namely the *proximity effect*.



When the spacing between the turns of the wound wire is very small, the loss resistance due to the proximity effect is larger than that due to the skin effect. The following formula is used to calculate exactly the loss resistance of a loop with N turns, wire radius b , and turn separation $2c$:

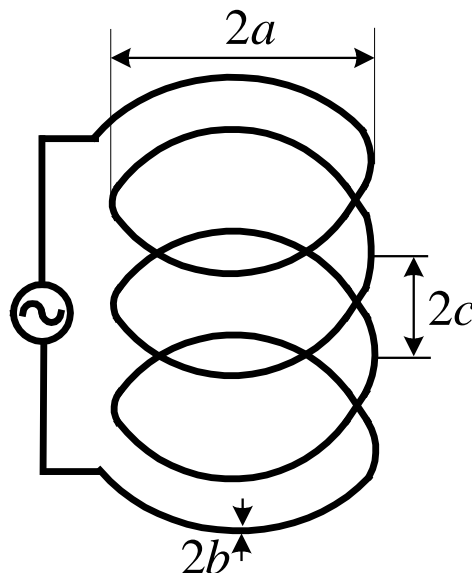
$$R_l = \frac{Na}{b} R_s \left(\frac{R_p}{R_0} + 1 \right) \quad (12.33)$$

where

R_s , Ω , is the surface resistance (see (12.32)),

R_p , Ω / m , is the ohmic resistance per unit length due to the proximity effect,

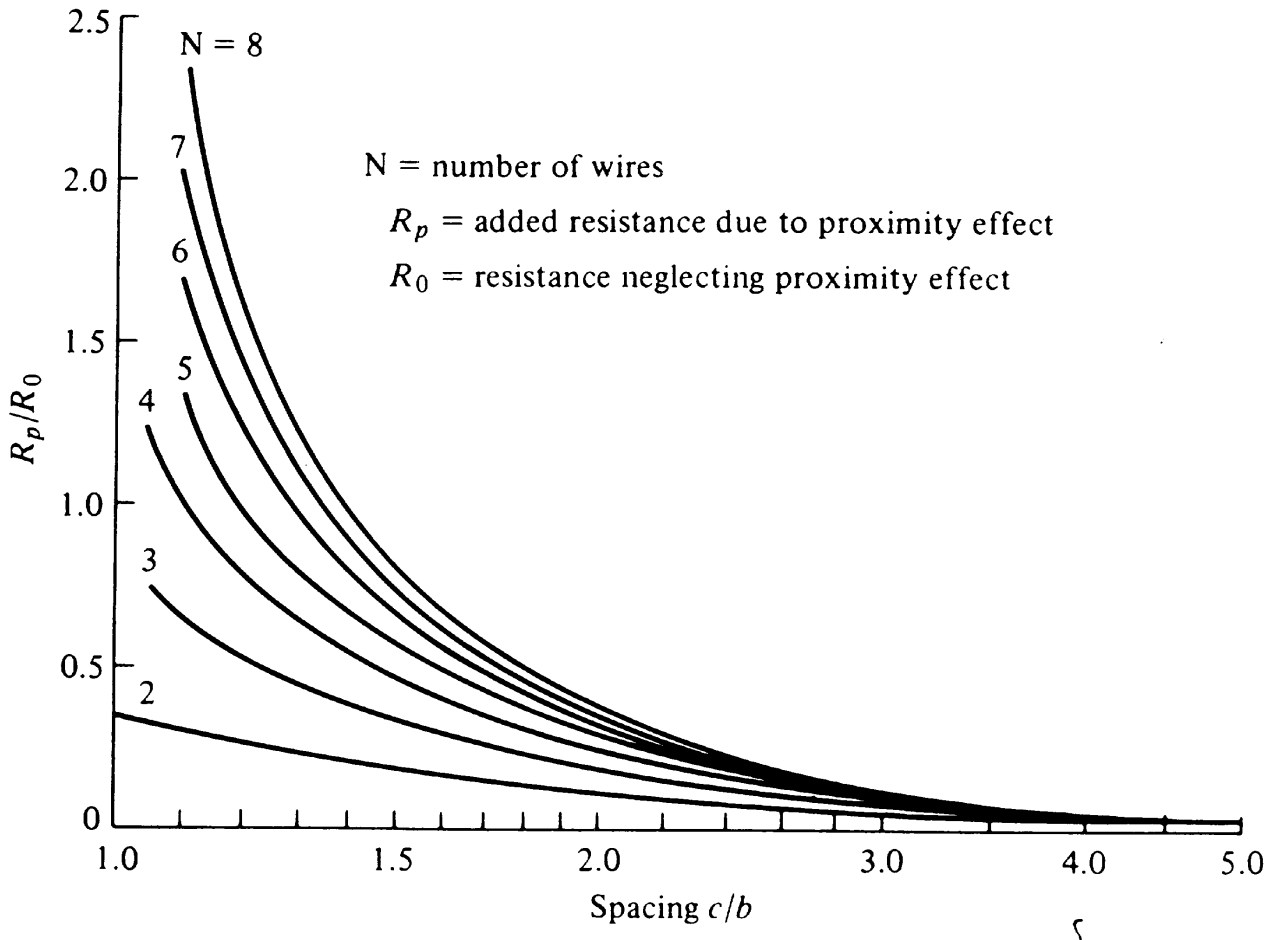
$R_0 = \frac{NR_s}{2\pi b}$, Ω / m , is the ohmic resistance per unit length due to the skin effect.



The ratio R_p / R_0 has been calculated for different relative spacings c / b , for loops with $2 \leq N \leq 8$ in:

G.N. Smith, "The proximity effect in systems of parallel conductors," *J. Appl. Phys.*, vol. 43, No. 5, May 1972, pp. 2196-2203.

The results are shown below:



(b) Ohmic resistance due to proximity (after G. N. Smith)

(b) Loop inductance

The inductance of a single circular loop of radius a made of wire of radius b is

$$L_{A1}^{\text{circ}} = \mu a \left[\ln \left(\frac{8a}{b} \right) - 2 \right] \text{ H.} \quad (12.34)$$

The inductance of a square loop with sides a and wire radius b is calculated as

$$L_{A1}^{\text{sq}} = 2\mu \frac{a}{\pi} \left[\ln\left(\frac{a}{b}\right) - 0.774 \right] \text{ H.} \quad (12.35)$$

The inductance of a multi-turn coil is obtained from the inductance of a single-turn loop multiplied by N^2 , where N is the number of turns.

The inductance of the wire itself (internal inductance) is very small and is often neglected. It can be shown that the HF self-inductance per unit length of a straight wire of cylindrical cross-section is

$$L'_{\text{int}} = \frac{\mu_0}{8\pi} \left[\frac{a^4 - 4a^2c^2 + 3c^4 + 4c^4 \ln(a/c)}{(a^2 - c^2)^2} \right] \text{ H/m,} \quad (12.36)$$

where $c = a - \delta$ and δ is the skin depth. To obtain the total internal inductance of the wire, simply multiply L'_{int} by the overall length of the wire used to construct the multi-turn loop antenna.

(c) Tuning capacitor

The susceptance of the capacitor B_r must be chosen to eliminate the susceptance of the loop. Assume that the equivalent admittance of the loop is

$$Y_{in} = \frac{1}{Z_{in}} = \frac{1}{R_{in} + jX_{in}} \quad (12.37)$$

where

$$R_{in} = R_r + R_l,$$

$$X_{in} = j\omega(L_A + L_{\text{int}}).$$

The following transformation holds:

$$Y_{in} = G_{in} + jB_{in} \quad (12.38)$$

where

$$\begin{cases} G_{in} = \frac{R_{in}}{R_{in}^2 + X_{in}^2}, \\ B_{in} = \frac{-X_{in}}{R_{in}^2 + X_{in}^2}. \end{cases} \quad (12.39)$$

The susceptance of the capacitor is

$$B_r = \omega C_r. \quad (12.40)$$

For resonance to occur at $f_0 = \omega_0 / (2\pi)$ when the capacitor is in parallel with the loop, the condition

$$B_r = -B_{in} \quad (12.41)$$

must be fulfilled. Therefore,

$$2\pi f_0 C_r = \frac{X_{in}}{R_{in}^2 + X_{in}^2}, \quad (12.42)$$

$$\Rightarrow C_r = \frac{1}{2\pi f} \frac{X_{in}}{(R_{in}^2 + X_{in}^2)}. \quad (12.43)$$

Under resonance, the input impedance Z'_{in} becomes

$$Z'_{in} = R'_{in} = \frac{1}{G'_{in}} = \frac{1}{G_{in}} = \frac{R_{in}^2 + X_{in}^2}{R_{in}}, \quad (12.44)$$

$$\Rightarrow Z'_{in} = R_{in} + \frac{X_{in}^2}{R_{in}}, \quad \Omega. \quad (12.45)$$

5. The Small Loop as a Receiving Antenna

The small loop antennas have the following features:

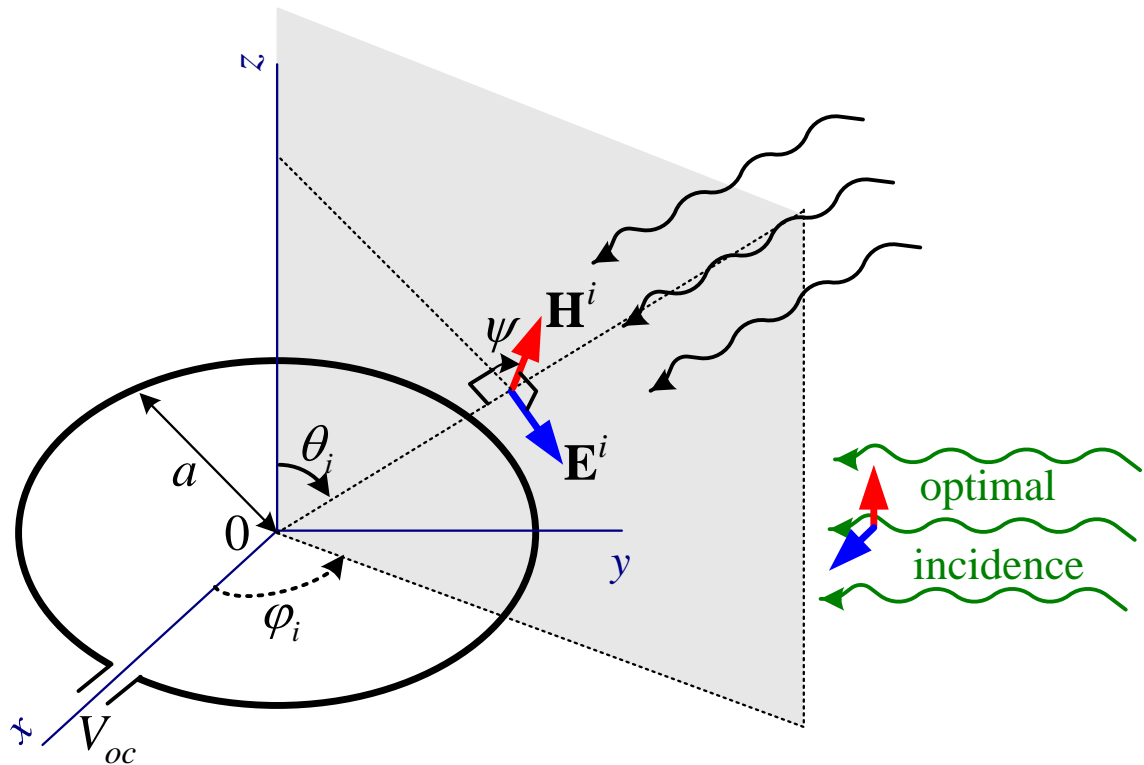
- 1) high radiation resistance provided multi-turn ferrite-core constructions are used;
- 2) high losses, therefore, low radiation efficiency;
- 3) simple construction, small size and weight.

Small loops are usually not used as transmitting antennas due to their low efficiency e_{cd} . However, they are much preferred as receiving antennas in AM

radio-receivers because of their high signal-to-noise ratio (they can be easily tuned to form a very high- Q resonant circuit), their small size and low cost.

Loops are constructed as magnetic field probes to measure magnetic flux densities. At higher frequencies (UHF and microwave), loops are used to measure the EM field intensity. In this case, ferrite rods are not used.

Since the loop is a typical linearly polarized antenna, it has to be oriented properly to optimize reception. The optimal case is a linearly polarized wave with the \mathbf{H} -field aligned with the loop's axis.



The open-circuit voltage at the loop terminals is induced by the time-varying magnetic flux through the loop:

$$V_{oc} = j\omega\Psi_m = j\omega\mathbf{B} \cdot \mathbf{s} = j\omega\mu H_z \cdot \pi a^2, \quad (12.46)$$

$$H_z = H^i \cos\psi \sin\theta_i. \quad (12.47)$$

Here,

Ψ_m is the magnetic flux, Wb;

(θ_i, φ_i) are the angles specifying the direction of incidence;
 ψ is the angle between the \mathbf{H}^i vector and the plane of incidence.

Finally, the open-circuit voltage can be expressed as

$$V_{oc} = j\omega\mu SH^i \cos\psi \sin\theta_i = j\beta SE^i \cos\psi \sin\theta_i. \quad (12.48)$$

Here, $S = \pi a^2$ denotes the area of the loop, and $\beta = \omega\sqrt{\mu\varepsilon}$ is the phase constant. V_{oc} is maximum for $\theta_i = 90^\circ$ and $\psi = 0^\circ$.

6. Ferrite Loops

The radiation resistance and radiation efficiency can be raised by inserting a ferrite core, which has high magnetic permeability in the operating frequency band. Large magnetic permeability $\mu = \mu_0\mu_r$ means large magnetic flux Ψ_m , and therefore large induced voltage V_{oc} . The radiation resistance of a small loop was already derived in (12.10) to include the number of turns, and it was shown that it increases as $\sim N^2$. Now the magnetic properties of the loop will be included in the expression for R_r .

The magnetic properties of a ferrite core depend not only on the relative magnetic permeability μ_r of the material it is made of but also on its geometry. The increase in the magnetic flux is then more realistically represented by the *effective relative permeability (effective magnetic constant)* $\mu_{r_{eff}}$. We show next that the radiation resistance of a ferrite-core loop is $(\mu_{r_{eff}})^2$ times larger than the radiation resistance of the air-core loop of the same geometry. When we calculated the far fields of a small loop, we used the equivalence between an electric current loop and a magnetic current element:

$$j\omega\mu(IA) = I_m l. \quad (12.49)$$

From (12.49) it is obvious that the equivalent magnetic current is proportional to μ . The field magnitudes are proportional to I_m , and therefore they are proportional to μ as well. This means that the radiated power Π_{rad} is proportional to μ^2 , and therefore the radiation resistance increases as $\sim (\mu_{r_{eff}})^2$.

Finally, we can express the radiation resistance as

$$R_r = \eta_0 \frac{8}{3} \pi^3 \left(N \mu_{r_{eff}} \frac{A}{\lambda^2} \right)^2. \quad (12.50)$$

Here, $A = \pi a^2$ is the loop area, and $\eta_0 = \sqrt{\mu_0 / \epsilon_0}$ is the intrinsic impedance of vacuum. An equivalent form of (12.50) is

$$R_r \approx 20\pi^2 (N \mu_{r_{\text{eff}}})^2 \left(\frac{C}{\lambda} \right)^4 \quad (12.51)$$

where we have used the approximate expression $\eta_0 \approx 120\pi$ and C is the circumference of the loop, $C = 2\pi a$.

Some notes are made below with regard to the properties of ferrite cores:

- The effective magnetic constant of a ferrite core is always less than the magnetic constant of the ferromagnetic material it is made of, i.e., $\mu_{r_{\text{eff}}} < \mu_r$. Toroidal cores have the highest $\mu_{r_{\text{eff}}}$, and ferrite-stick cores have the lowest $\mu_{r_{\text{eff}}}$.
- The effective magnetic constant is frequency dependent. One has to be careful when picking the right core for the application at hand.
- The magnetic losses of ferromagnetic materials increase with frequency. At very high (microwave) frequencies, the magnetic losses are very high. They have to be calculated and represented in the equivalent circuit of the antenna as a shunt conductance G_m .

LECTURE 13: LINEAR ARRAY THEORY - PART I

(Linear arrays: the two-element array. N-element array with uniform amplitude and spacing. Broad-side array. End-fire array. Phased array.)

1. Introduction

Usually the radiation patterns of single-element antennas are relatively wide, i.e., they have relatively low directivity (gain). In long distance communications, antennas with high directivity are often required. Such antennas are possible to construct by enlarging the dimensions of the radiating aperture (maximum size much larger than λ). This approach however may lead to the appearance of multiple side lobes. Besides, the antenna is usually large and difficult to fabricate.

Another way to increase the electrical size of an antenna is to construct it as an assembly of radiating elements in a proper electrical and geometrical configuration – **antenna array**. Usually, the array elements are identical. This is not necessary but it is practical and simpler for design and fabrication. The individual elements may be of any type (wire dipoles, loops, apertures, etc.)

The total field of an array is a vector superposition of the fields radiated by the individual elements. To provide very directive pattern, it is necessary that the partial fields (generated by the individual elements) interfere constructively in the desired direction and interfere destructively in the remaining space.

There are five basic methods to control the overall antenna pattern:

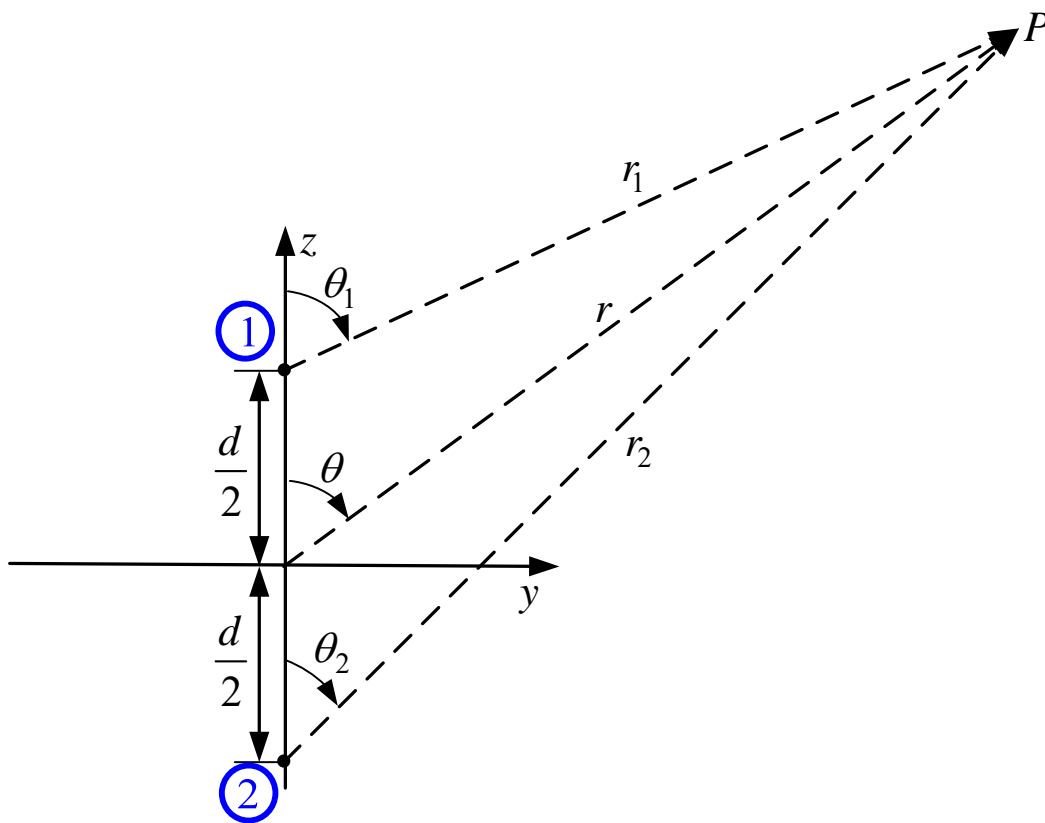
- a) the geometrical configuration of the overall array (linear, circular, spherical, rectangular, etc.),
- b) the relative placement of the elements,
- c) the excitation amplitude of the individual elements,
- d) the excitation phase of each element,
- e) the individual pattern of each element.

2. Two-element Array

Let us represent the electric fields in the far zone of the array elements in the form

$$\mathbf{E}_1 = M_1 E_{n1}(\theta_1, \phi_1) \frac{e^{-j\left(kr_1 - \frac{\beta}{2}\right)}}{r_1} \hat{\mathbf{p}}_1, \quad (13.1)$$

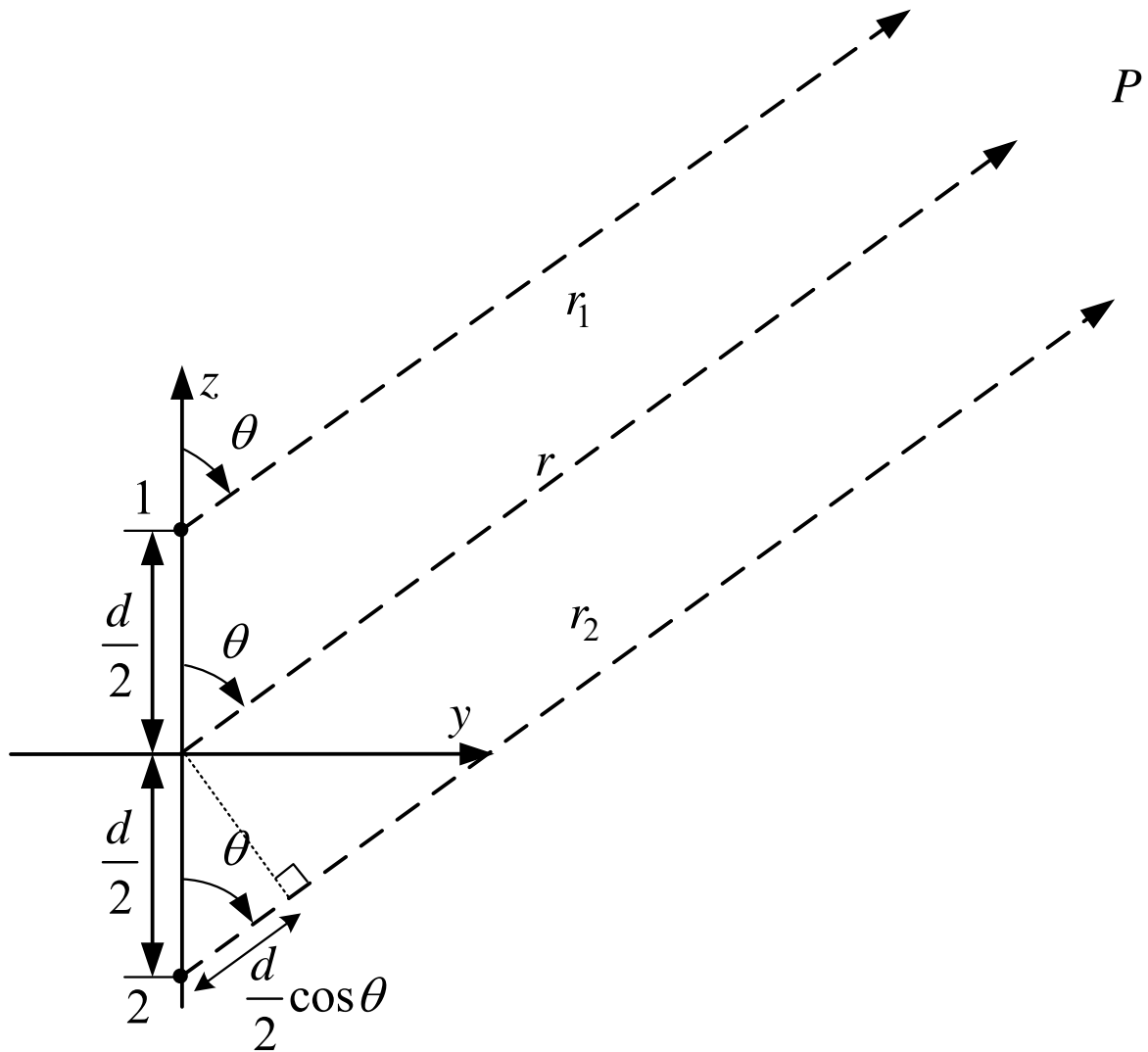
$$\mathbf{E}_2 = M_2 E_{n2}(\theta_2, \phi_2) \frac{e^{-j\left(kr_2 + \frac{\beta}{2}\right)}}{r_2} \hat{\mathbf{p}}_2. \quad (13.2)$$



Here,

- M_1, M_2 field magnitudes (do not include the $1/r$ factor);
- E_{n1}, E_{n2} normalized field patterns;
- r_1, r_2 distances to the observation point P ;
- β phase difference between the feed of the two array elements;
- $\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2$ polarization vectors of the far-zone \mathbf{E} fields.

The far-field approximation of the two-element array problem:



Let us assume that:

- 1) the array elements are identical, i.e.,

$$E_{n1}(\theta, \phi) = E_{n2}(\theta, \phi) = E_n(\theta, \phi), \quad (13.3)$$

- 2) they are oriented in the same way in space (they have identical polarizations), i.e.,

$$\hat{\rho}_1 = \hat{\rho}_2 = \hat{\rho}, \quad (13.4)$$

- 3) their excitation is of the same amplitude, i.e.,

$$M_1 = M_2 = M. \quad (13.5)$$

Then, the total field can be derived as

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2, \quad (13.6)$$

$$\mathbf{E} = \hat{\mathbf{p}} M E_n(\theta, \phi) \frac{1}{r} \left[e^{-jk\left(r - \frac{d}{2} \cos \theta\right) + j\frac{\beta}{2}} + e^{-jk\left(r + \frac{d}{2} \cos \theta\right) - j\frac{\beta}{2}} \right],$$

$$\mathbf{E} = \hat{\mathbf{p}} \frac{M}{r} e^{-jkr} E_n(\theta, \phi) \left[e^{j\left(\frac{kd}{2} \cos \theta + \frac{\beta}{2}\right)} + e^{-j\left(\frac{kd}{2} \cos \theta + \frac{\beta}{2}\right)} \right],$$

$$\boxed{\mathbf{E} = \hat{\mathbf{p}} M \underbrace{\frac{e^{-jkr}}{r} E_n(\theta, \phi)}_{EF} \times 2 \underbrace{\cos\left(\frac{kd \cos \theta + \beta}{2}\right)}_{AF}}. \quad (13.7)$$

The total field of the array is equal to the product of the field created by a single element located at the origin and the **array factor** (AF):

$$AF = 2 \cos\left(\frac{kd \cos \theta + \beta}{2}\right). \quad (13.8)$$

Using the normalized field pattern of a single element, $E_n(\theta, \phi)$, and the normalized AF,

$$AF_n = \cos\left(\frac{kd \cos \theta + \beta}{2}\right), \quad (13.9)$$

the normalized field pattern of the array is expressed as their product:

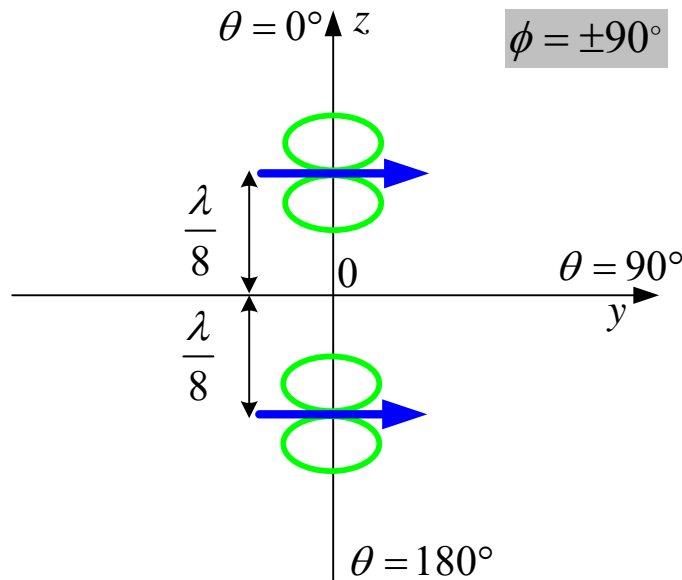
$$f_n(\theta, \phi) = E_n(\theta, \phi) \times AF_n(\theta, \phi). \quad (13.10)$$

The concept expressed by (13.10) is the so-called **pattern multiplication rule** valid for arrays of identical elements. This rule holds for any array consisting of decoupled identical elements, where the excitation magnitudes, the phase shift between the elements and the displacement between them are not necessarily the same. The total pattern, therefore, can be controlled via the single–element pattern $E_n(\theta, \phi)$ or via the AF. The AF, in general, depends on the:

- number of elements,
- mutual placement,
- relative excitation magnitudes and phases.

Example 1: An array consists of two horizontal infinitesimal dipoles located at a distance $d = \lambda / 4$ from each other. Find the nulls of the total field in the elevation plane $\phi = \pm 90^\circ$, if the excitation magnitudes are the same and the phase difference is:

- a) $\beta = 0$
- b) $\beta = \pi / 2$
- c) $\beta = -\pi / 2$



The element factor $E_n(\theta, \phi) = \sqrt{1 - \sin^2 \theta \sin^2 \phi}$ does not depend on β , and it produces in all three cases the same null. For $\phi = \pm 90^\circ$, $E_n(\theta, \phi) = |\cos \theta|$ and the null is at

$$\theta_1 = \pi / 2. \quad (13.11)$$

The AF depends on β and produces different results in the 3 cases:

- a) $\beta = 0$

$$AF_n = \cos\left(\frac{kd \cos \theta_n}{2}\right) = 0 \Rightarrow \cos\left(\frac{\pi}{4} \cos \theta_n\right) = 0,$$

$$\Rightarrow \frac{\pi}{4} \cos \theta_n = (2n + 1) \frac{\pi}{2} \Rightarrow \cos \theta_n = (2n + 1) \cdot 2, \quad n = 0, \pm 1, \pm 2, \dots$$

A solution with a real-valued angle does not exist. In this case, the total field pattern has only 1 null at $\theta = 90^\circ$.

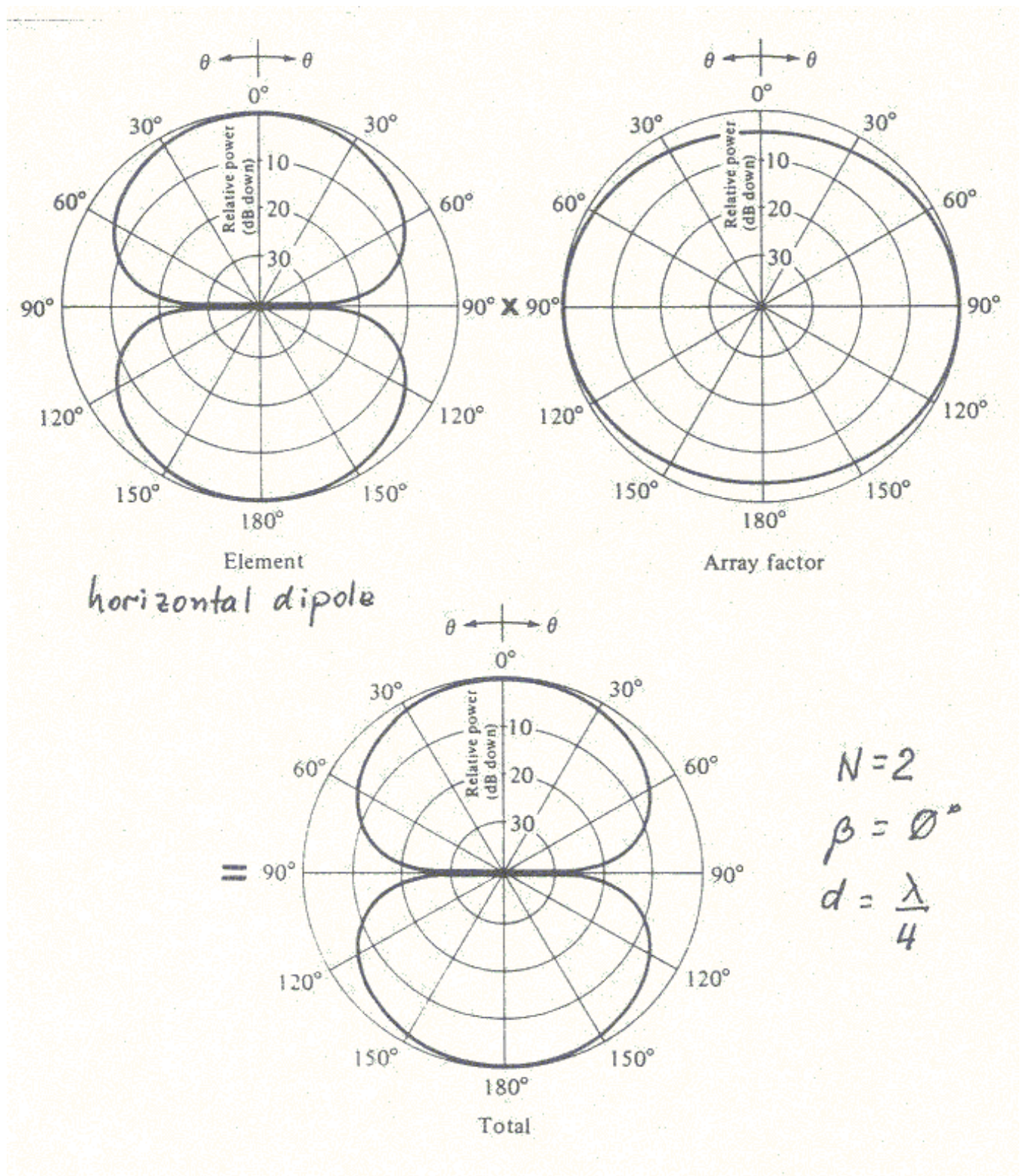


Fig. 6.3, p. 255, Balanis

b) $\beta = \pi / 2$

$$AF_n = \cos\left(\frac{\pi}{4}\cos\theta_n + \frac{\pi}{4}\right) = 0 \Rightarrow \frac{\pi}{4}(\cos\theta_n + 1) = (2n + 1)\frac{\pi}{2},$$

$$\Rightarrow \cos\theta_n + 1 = (2n + 1) \cdot 2 \Rightarrow \cos\theta_{(n=0)} = 1 \Rightarrow \boxed{\theta_2 = 0^\circ}.$$

The solution for $n = 0$ is the only real-valued solution. Thus, the total field pattern has 2 nulls: at $\theta_1 = 90^\circ$ and at $\theta_2 = 0^\circ$:

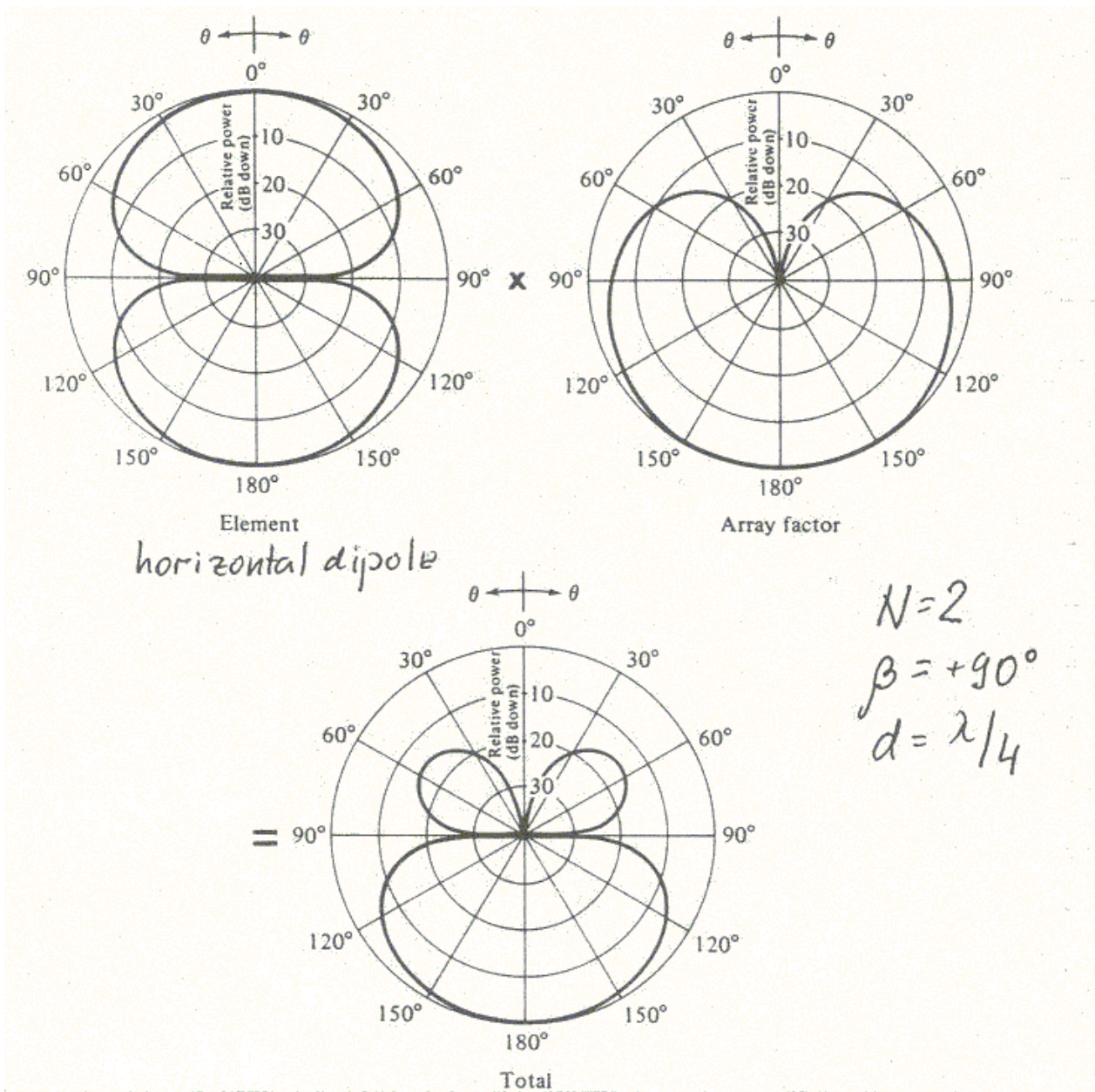


Fig. 6.4, p. 256, Balanis

c) $\beta = -\pi / 2$

$$AF_n = \cos\left(\frac{\pi}{4}\cos\theta_n - \frac{\pi}{4}\right) = 0 \Rightarrow \frac{\pi}{4}(\cos\theta_n - 1) = (2n + 1)\frac{\pi}{2},$$

$$\Rightarrow \cos\theta_n - 1 = (2n + 1) \cdot 2 \Rightarrow \cos\theta_{(n=-1)} = -1 \Rightarrow \boxed{\theta_2 = \pi}.$$

The total field pattern has 2 nulls: at $\theta_1 = 90^\circ$ and at $\theta_2 = 180^\circ$.

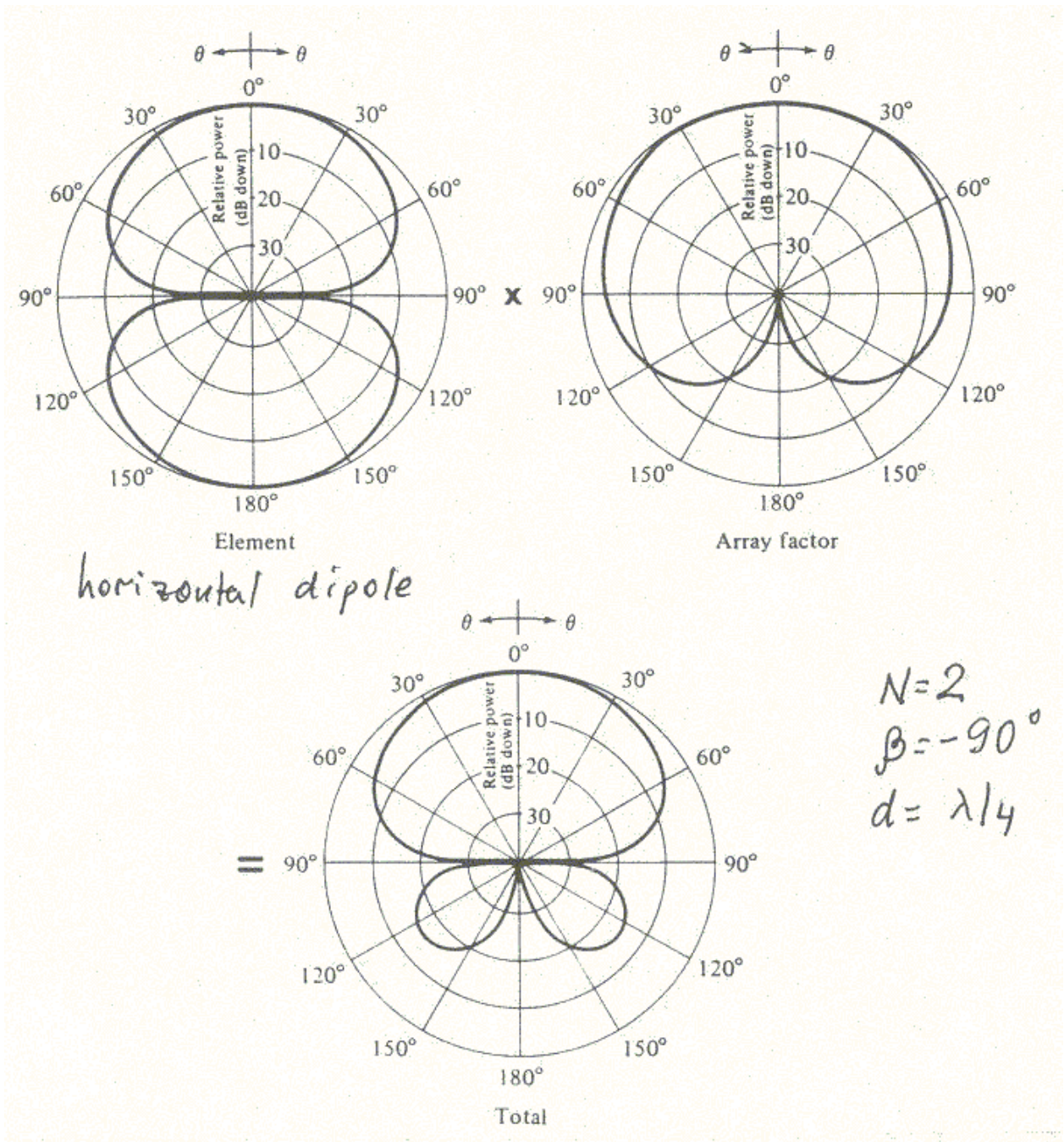


Fig. 6.4b, p. 257, Balanis

Example 2: Consider a 2-element array of identical (infinitesimal) dipoles oriented along the y -axis. Find the angles of observation where the nulls of the pattern occur in the plane $\phi = \pm 90^\circ$ as a function of the distance d between the dipoles and the phase difference β .

The normalized total field pattern is

$$f_n = |\cos \theta| \times \cos\left(\frac{kd \cos \theta + \beta}{2}\right). \quad (13.12)$$

In order to find the nulls, the equation

$$f_n = |\cos \theta| \cdot \cos\left(\frac{kd \cos \theta + \beta}{2}\right) = 0 \quad (13.13)$$

is solved.

The element factor $|\cos \theta|$ produces one null at

$$\theta_1 = \pi / 2. \quad (13.14)$$

The AF leads to the following solution:

$$\cos\left(\frac{kd \cos \theta + \beta}{2}\right) = 0 \Rightarrow \frac{kd \cos \theta + \beta}{2} = (2n + 1)\frac{\pi}{2}, \quad n = 0, \pm 1, \pm 2, \dots$$

$$\underline{\underline{\theta_n = \arccos\left\{\frac{\lambda}{2\pi d}[-\beta \pm (2n + 1)\pi]\right\}}}. \quad (13.15)$$

When there is no phase difference between the two element feeds ($\beta = 0$), the separation d must satisfy

$$d \geq \frac{\lambda}{2}$$

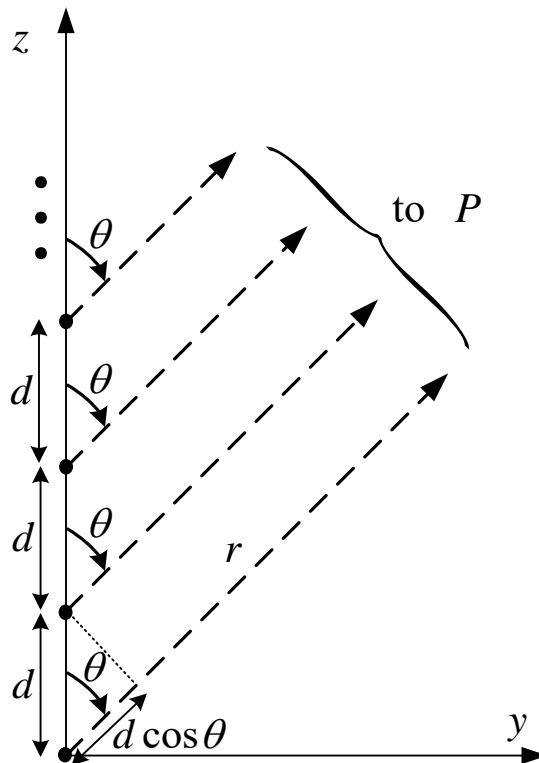
in order at least one null to occur due to (13.15).

3. N -element Linear Array with Uniform Amplitude and Spacing

We assume that each succeeding element has a β progressive phase lead in the excitation relative to the preceding one. An array of identical elements with identical magnitudes and with a progressive phase is called a **uniform array**. The AF of the uniform array can be obtained by considering the individual elements as point (isotropic) sources. Then, the total field pattern can be obtained by simply multiplying the AF by the normalized field pattern of the individual element (provided the elements are not coupled).

The AF of an N -element linear array of isotropic sources is

$$AF = 1 + e^{j(kd \cos \theta + \beta)} + e^{j2(kd \cos \theta + \beta)} + \dots + e^{j(N-1)(kd \cos \theta + \beta)}. \quad (13.16)$$



Phase terms of the partial fields:

$$\begin{aligned}
 1^{\text{st}} &\rightarrow e^{-jkr} \\
 2^{\text{nd}} &\rightarrow e^{-jk(r-d \cos \theta)} \\
 3^{\text{rd}} &\rightarrow e^{-jk(r-2d \cos \theta)} \\
 &\dots \\
 N^{\text{th}} &\rightarrow e^{-jk(r-(N-1)d \cos \theta)}
 \end{aligned}$$

Equation (13.16) can be re-written as

$$AF = \sum_{n=1}^N e^{j(n-1)(kd \cos \theta + \beta)}, \quad (13.17)$$

$$AF = \sum_{n=1}^N e^{j(n-1)\psi}, \quad (13.18)$$

where $\psi = kd \cos \theta + \beta$.

From (13.18), it is obvious that the AF s of uniform linear arrays can be controlled by the relative phase β between the elements. The AF in (13.18) can be expressed in a closed form, which is more convenient for pattern analysis:

$$AF \cdot e^{j\psi} = \sum_{n=1}^N e^{jn\psi}, \quad (13.19)$$

$$AF \cdot e^{j\psi} - AF = e^{jN\psi} - 1,$$

$$AF = \frac{e^{jN\psi} - 1}{e^{j\psi} - 1} = \frac{e^{j\frac{N}{2}\psi} \left(e^{j\frac{N}{2}\psi} - e^{-j\frac{N}{2}\psi} \right)}{e^{j\frac{\psi}{2}} \left(e^{j\frac{\psi}{2}} - e^{-j\frac{\psi}{2}} \right)},$$

$$AF = e^{j\left(\frac{N-1}{2}\right)\psi} \cdot \frac{\sin(N\psi / 2)}{\sin(\psi / 2)}. \quad (13.20)$$

Here, the phase factor $\exp[j(N-1)\psi / 2]$ reflects a phase advancement associated with the last (N th) array element relative to the center of the linear array. It also represents the phase shift of the array's centre relative to the

origin, and it would be equal to one if the origin were to coincide with the array centre. This factor is not important unless the array output signal is further combined with the output signal of another antenna. As we aim at obtaining the normalized AF, we neglect this phase factor, leading to

$$AF = \frac{\sin(N\psi / 2)}{\sin(\psi / 2)}. \quad (13.21)$$

For small values of ψ , (13.21) reduces to

$$AF = \frac{\sin(N\psi / 2)}{\psi / 2}. \quad (13.22)$$

To normalize (13.22) or (13.21), we need the maximum of the AF . We re-write (13.21) as

$$AF = N \cdot \frac{\sin(N\psi / 2)}{N \sin(\psi / 2)}. \quad (13.23)$$

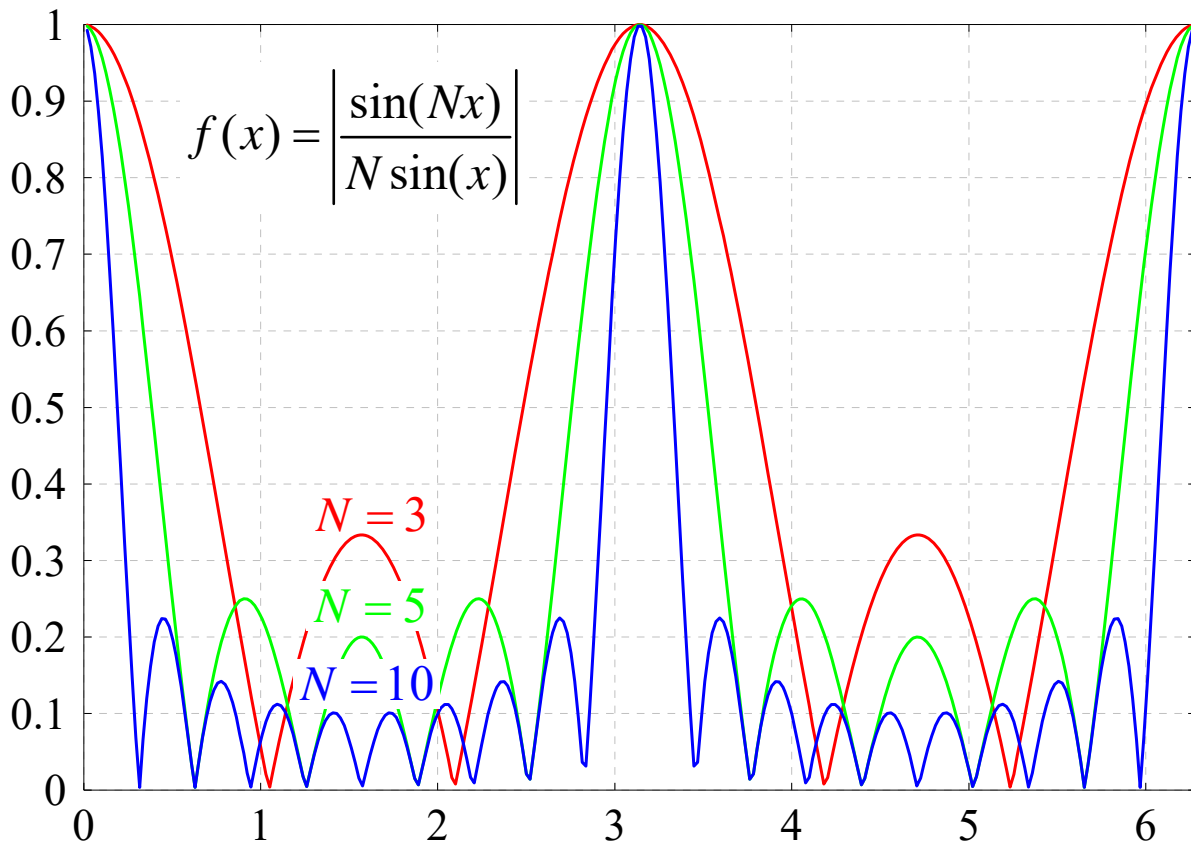
The function

$$f(x) = \frac{\sin(Nx)}{N \sin(x)}$$

has its maximum at $x = 0, \pi, \dots$, and the value of this maximum is $f_{\max} = 1$. Therefore, $AF_{\max} = N$. The normalized AF is thus obtained as

$$AF_n = \frac{\sin(N\psi / 2)}{N \sin(\psi / 2)}. \quad (13.24)$$

The function $|f(x)|$ is plotted below.



For small ψ ,

$$AF_n = \frac{1}{N} \left[\frac{\sin(N\psi/2)}{\psi/2} \right]. \quad (13.25)$$

Nulls of the AF

To find the nulls of the AF, equation (13.24) is set equal to zero:

$$\sin\left(\frac{N}{2}\psi\right) = 0 \Rightarrow \frac{N}{2}\psi = \pm n\pi \Rightarrow \frac{N}{2}(kd \cos\theta_n + \beta) = \pm n\pi, \quad (13.26)$$

$$\theta_n = \arccos\left[\frac{\lambda}{2\pi d} \left(-\beta \pm \frac{2n}{N}\pi\right)\right], \quad n = 1, 2, 3, \dots (n \neq 0, N, 2N, 3N \dots). \quad (13.27)$$

When $n = 0, N, 2N, 3N \dots$, the AF attains its maximum values not nulls (see the case below). The values of n determine the order of the nulls. For a null to exist, *the argument of the arccosine must be between -1 and $+1$.*

Major maxima of the AF

They are studied in order to determine the maximum directivity, the *HPBW*s, the direction of maximum radiation. The maxima of (13.24) occur when (see the plot in page 13, note that $x = \psi / 2$)

$$\frac{\psi}{2} = \frac{1}{2}(kd \cos \theta_m + \beta) = \pm m\pi, \quad (13.28)$$

$$\theta_m = \arccos \left[\frac{\lambda}{2\pi d} (-\beta \pm 2m\pi) \right], \quad m = 0, 1, 2, \dots \quad (13.29)$$

When (13.28) is true, $AF_n = 1$, i.e., these are not maxima of minor lobes. The index m shows the maximum's order. It is usually desirable to have a single major lobe, i.e. $m = 0$ only. This can be achieved by choosing d / λ sufficiently small. Then the argument of the *arccosine* function in (13.29) becomes greater than unity for $m = 1, 2, \dots$ and equation (13.29) has a single real-valued solution:

$$\theta_0 = \arccos \left(-\frac{\beta\lambda}{2\pi d} \right). \quad (13.30)$$

The *HPBW* of a major lobe

The *HPBW* of a major lobe is calculated by setting the value of AF_n equal to $1 / \sqrt{2}$. For the approximate AF_n in (13.25),

$$\frac{N}{2}\psi = \frac{N}{2}(kd \cos \theta_h + \beta) \approx \pm 1.391.$$

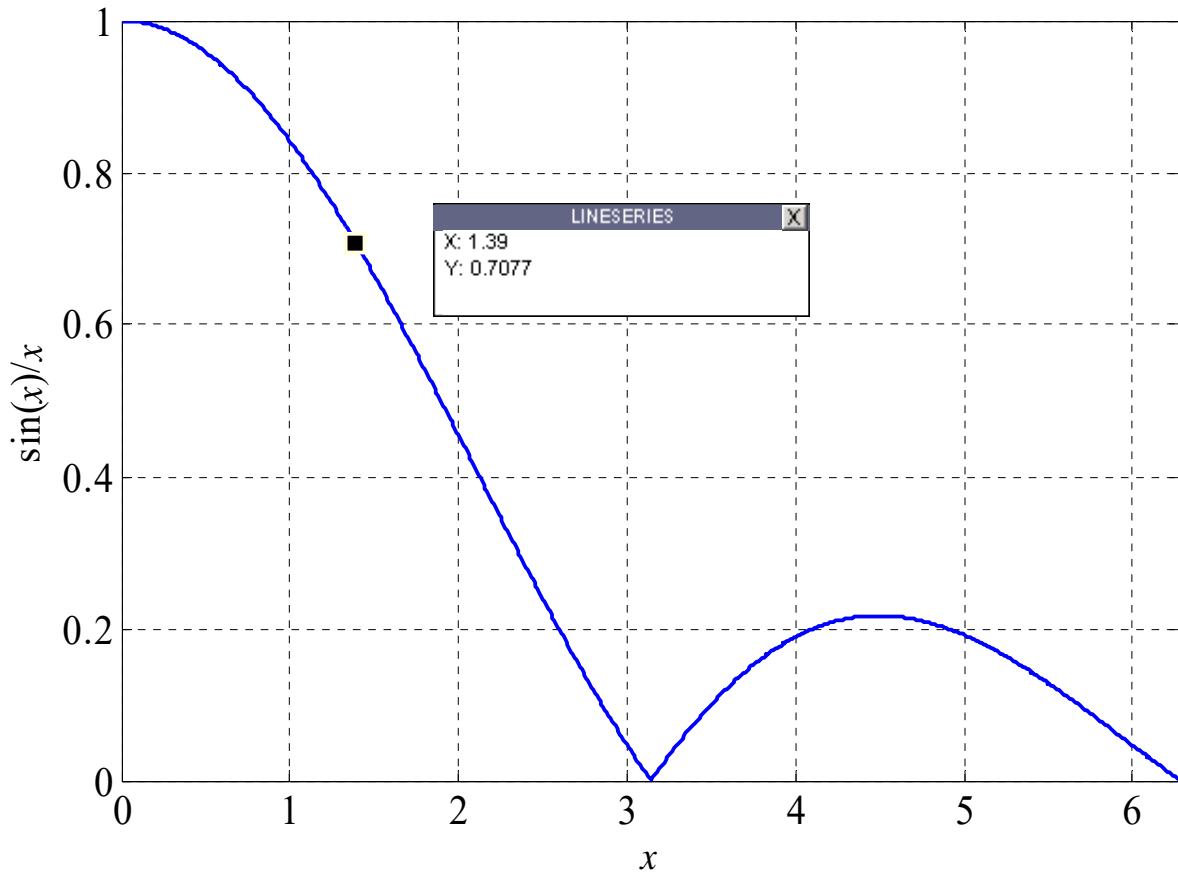
See the plot of $(\sin x) / x$ below.

$$\Rightarrow \theta_h = \arccos \left[\frac{\lambda}{2\pi d} \left(-\beta \pm \frac{2.782}{N} \right) \right]. \quad (13.31)$$

For a symmetrical pattern around θ_m (the angle at which maximum radiation occurs), the *HPBW* is calculated as

$$HPBW = 2 | \theta_m - \theta_h |. \quad (13.32)$$

For a broadside array, for example, $\theta_m = \theta_0 = \pi / 2$.



Maxima of minor lobes (secondary maxima)

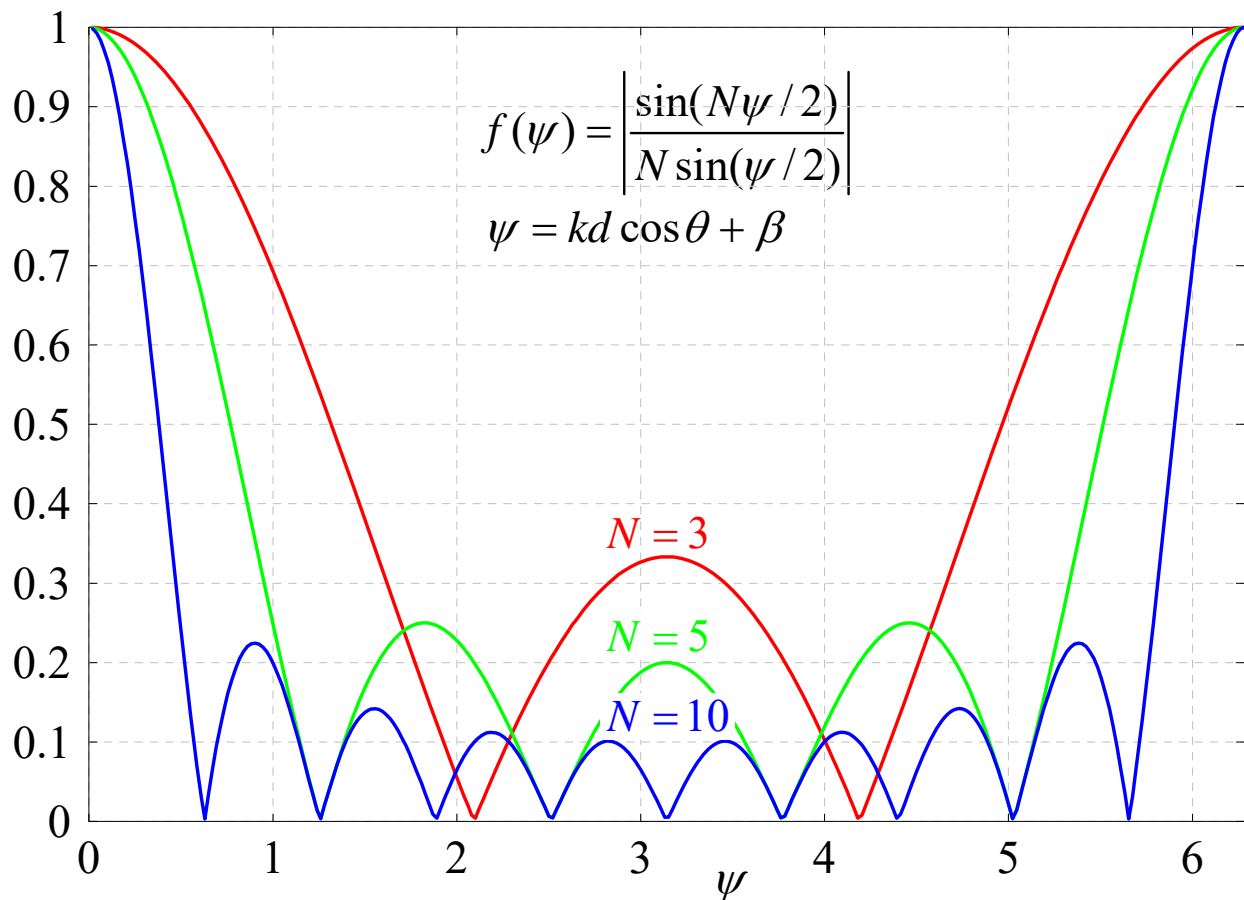
They are the maxima of AF_n , where $AF_n < 1$. These are illustrated in the plot below, which shows the array factors as a function of $\psi = kd \cos \theta + \beta$ for a uniform equally spaced linear array with $N = 3, 5, 10$.

The secondary maxima occur where the numerator attains a maximum and the AF is beyond its 1st null:

$$\sin\left(\frac{N}{2}\psi\right) = \pm 1 \Rightarrow \frac{N}{2}(kd \cos \theta + \beta) = \pm(2s+1)\frac{\pi}{2}, \quad (13.33)$$

$$\Rightarrow \theta_s = \arccos\left\{\frac{\lambda}{2\pi d}\left(-\beta \pm \left(\frac{2s+1}{N}\right)\pi\right)\right\} \text{ or} \quad (13.34)$$

$$\theta_s = \frac{\pi}{2} - \arccos\left\{\frac{\lambda}{2\pi d}\left(-\beta \pm \left(\frac{2s+1}{N}\right)\pi\right)\right\}. \quad (13.35)$$



4. Broadside Array

A broadside array is an array, which has maximum radiation at $\theta = 90^\circ$ (normal to the axis of the array). For optimal solution, both the element factor and the AF , should have their maxima at $\theta = 90^\circ$.

From (13.28), it follows that the maximum of the AF occurs when

$$\psi = kd \cos \theta_m + \beta = 0. \quad (13.36)$$

Equation (13.36) is valid for the 0th order maximum, $m = 0$. If $\theta_m = \pi / 2$, then

$$\boxed{\beta = 0}. \quad (13.37)$$

The uniform linear array has its maximum radiation at $\theta = 90^\circ$, if all array elements have their excitation with the same phase.

To ensure that there are no maxima in other directions (grating lobes), the separation between the elements d must be smaller than the wavelength:

$$\boxed{d < \lambda}. \quad (13.38)$$

To illustrate the appearance of additional maxima, $AF_n = 1$, let us consider the case of $d = \xi\lambda$, where $\xi \geq 1$. Then,

$$\psi_{\beta=0} = kd \cos \theta = \frac{2\pi}{\lambda} \xi\lambda \cos \theta = 2\pi\xi \cos \theta. \quad (13.39)$$

The condition for $AF_n = 1$ from (13.28),

$$\psi_m = 2\pi \cdot m, \quad m = 0, \pm 1, \pm 2 \dots \quad (13.40)$$

is then fulfilled not only for $\theta_0 = \pi / 2$ but also for

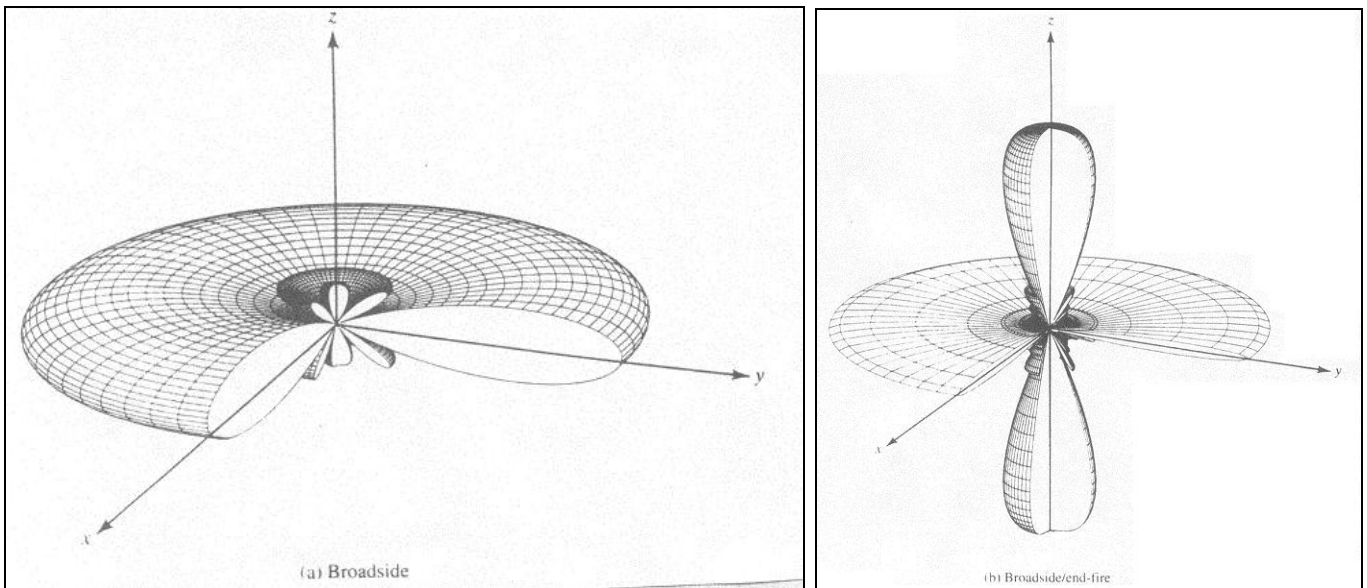
$$\theta_g = \arccos(m / \xi), \quad m = \pm 1, \pm 2 \dots \quad (13.41)$$

As long as $m \leq \xi$, real-valued solutions for θ_g exist and grating lobes appear.

If, for example, $d = \lambda$ ($\xi = 1$), equation (13.41) results in two additional major lobes at

$$\theta_g = \arccos(\pm 1) \Rightarrow \theta_{g1,2} = 0^\circ, 180^\circ.$$

The resulting AF is illustrated in figure (b) below.



If $d = 2\lambda$ ($\xi = 2$), equation (13.41) results in four additional major lobes at

$$\theta_g = \arccos(\pm 0.5, \pm 1) \Rightarrow \theta_{g1,2,3,4} = 0^\circ, 60^\circ, 120^\circ, 180^\circ.$$

If $d = 1.25\lambda$ ($\xi = 1.25$),

$$\theta_g = \arccos(\pm 0.8) \Rightarrow \theta_{g1,2} \approx 37^\circ, 143^\circ.$$

5. Ordinary End-fire Array

An end-fire array is an array, which has its maximum radiation along the axis of the array ($\theta = 0^\circ, 180^\circ$). It may be required that the array radiates only in one direction – either $\theta = 0^\circ$ or $\theta = 180^\circ$. For an *AF* maximum at $\theta = 0^\circ$,

$$\psi = kd \cos \theta + \beta \Big|_{\theta=0^\circ} = kd + \beta = 0, \quad (13.42)$$

$$\Rightarrow \boxed{\beta = -kd, \text{ for } \theta_{\max} = 0^\circ}. \quad (13.43)$$

For an *AF* maximum at $\theta = 180^\circ$,

$$\psi = kd \cos \theta + \beta \Big|_{\theta=180^\circ} = -kd + \beta = 0,$$

$$\Rightarrow \boxed{\beta = kd, \text{ for } \theta_{\max} = 180^\circ}. \quad (13.44)$$

If the element separation is multiple of a wavelength, $d = n\lambda$, then in addition to the end-fire maxima there also exist maxima in the broadside direction ($\theta = 90^\circ$). As with the broadside array, in order to avoid grating lobes, the maximum spacing between the element should be less than λ :

$$d < \lambda.$$

(Show that an end-fire array with $d = \lambda / 2$ has 2 maxima for $\beta = -kd$: at $\theta = 0^\circ$ and at $\theta = 180^\circ$.)

AF pattern of an *EFA*: $N = 10$, $d = \lambda / 4$

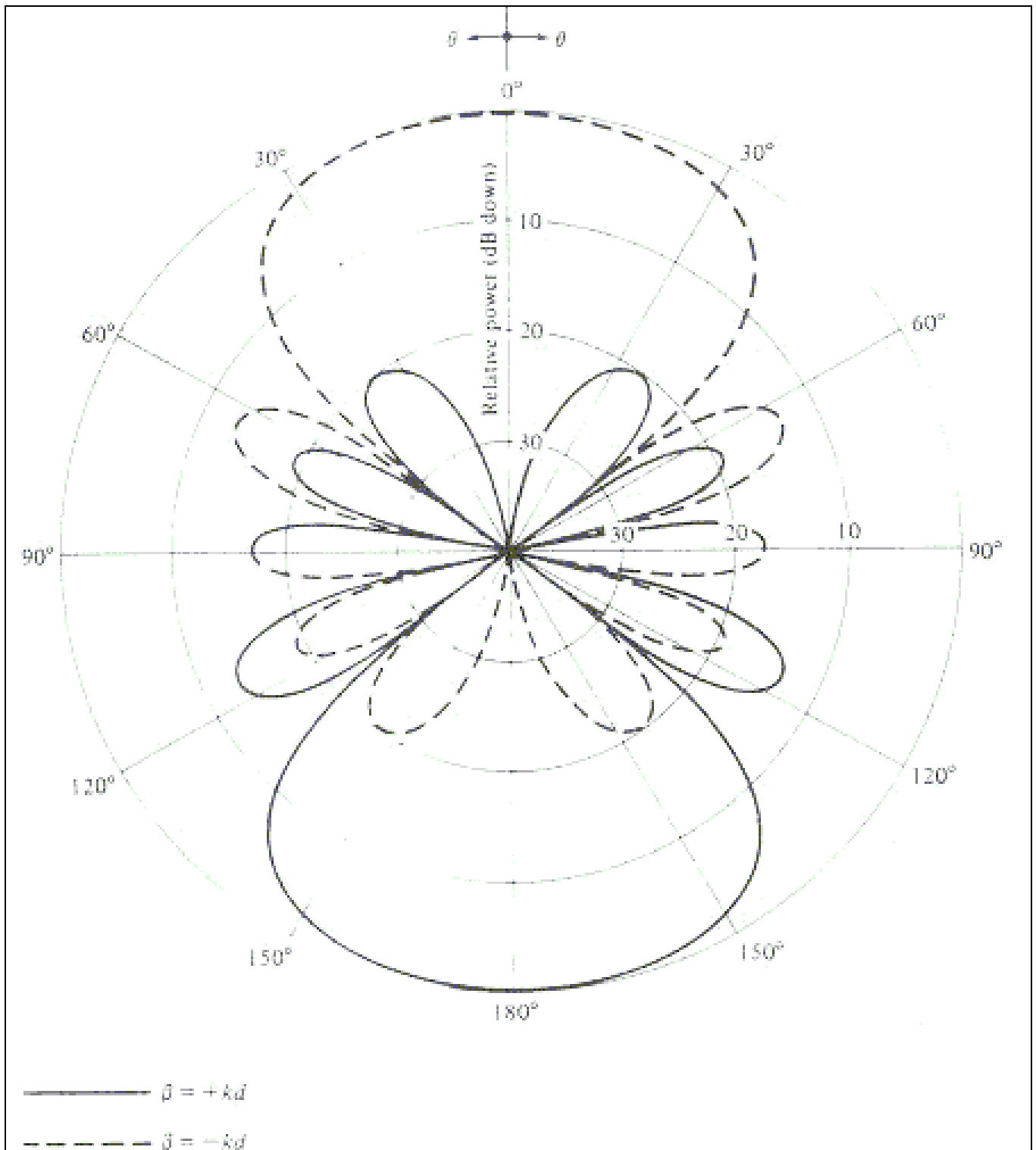


Fig. 6-11, p. 270, Balanis

6. Phased (Scanning) Arrays

It was already shown that the 0th order maximum ($m=0$) of AF_n occurs when

$$\psi = kd \cos \theta_0 + \beta = 0. \quad (13.45)$$

This gives the relation between the direction of the main beam θ_0 and the phase difference β . The direction of the main beam can be controlled by the phase shift β . This is the basic principle of **electronic scanning** for phased arrays.

When the scanning is required to be continuous, the feeding system must be capable of continuously varying the progressive phase β between the elements. This is accomplished by ferrite or diode shifters (varactors).

Example: Derive the values of the progressive phase shift β as dependent on the direction of the main beam θ_0 for a uniform linear array with $d = \lambda / 4$.

From equation (13.45):

$$\beta = -kd \cos \theta_0 = -\frac{2\pi}{\lambda} \frac{\lambda}{4} \cos \theta_0 = -\frac{\pi}{2} \cos \theta_0$$

θ_0	β
0°	-90°
60°	-45°
120°	+45°
180°	+90°

The approximate *HPBW* of a scanning array is obtained using (13.31) with $\beta = -kd \cos \theta_0$:

$$\theta_{h_{1,2}} = \arccos \left[\frac{\lambda}{2\pi d} \left(-\beta \pm \frac{2.782}{N} \right) \right]. \quad (13.46)$$

The total beamwidth is

$$HPBW = \theta_{h1} - \theta_{h2}, \quad (13.47)$$

$$HPBW = \arccos \left[\frac{\lambda}{2\pi d} \left(kd \cos \theta_0 - \frac{2.782}{N} \right) \right] - \arccos \left[\frac{\lambda}{2\pi d} \left(kd \cos \theta_0 + \frac{2.782}{N} \right) \right] \quad (13.48)$$

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

$$HPBW = \arccos \left[\cos \theta_0 - \frac{2.782}{Nkd} \right] - \arccos \left[\cos \theta_0 + \frac{2.782}{Nkd} \right]. \quad (13.49)$$

We can use the substitution $N = (L + d) / d$ to obtain

$$HPBW = \arccos \left[\cos \theta_0 - 0.443 \left(\frac{\lambda}{L + d} \right) \right] - \arccos \left[\cos \theta_0 + 0.443 \left(\frac{\lambda}{L + d} \right) \right]. \quad (13.50)$$

Here, L is the length of the array.

Equations (13.49) and (13.50) can be used to calculate the *HPBW* of a broadside array ($\theta_0 = 90^\circ = \text{const}$). However, they are *not valid for end-fire arrays*, where

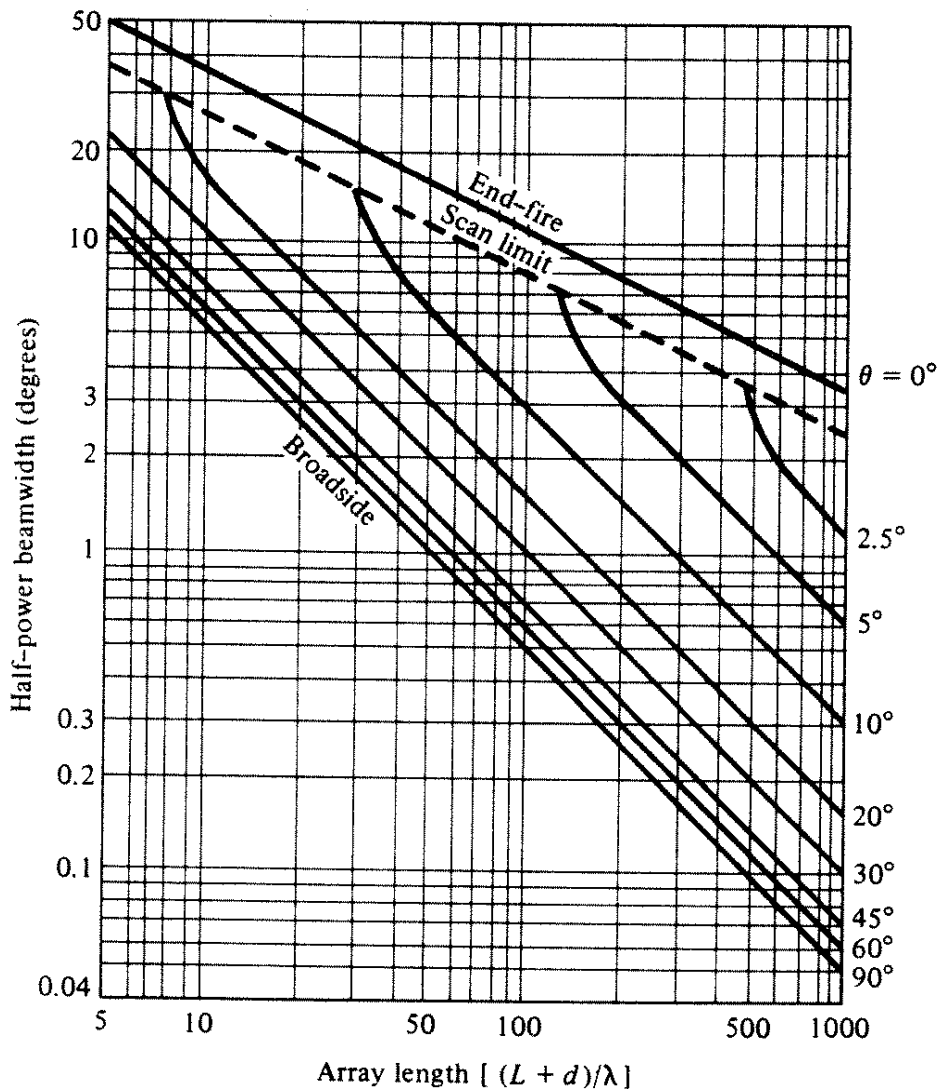
$$HPBW = 2 \arccos \left(1 - \frac{2.782}{Nkd} \right). \quad (13.51)$$

LECTURE 14: LINEAR ARRAY THEORY - PART II

(Linear arrays: Hansen-Woodyard end-fire array, directivity of a linear array, linear array pattern characteristics – recapitulation; 3-D characteristics of an N -element linear array.)

1. Hansen-Woodyard End-fire Array (HWEFA)

The end-fire arrays (EFA) have relatively large HPBW as compared to broadside arrays.



Half-power beamwidth for broadside, ordinary end-fire, and scanning uniform linear arrays. (SOURCE: R. S. Elliott, "Beamwidth and Directivity of Large Scanning Arrays," First of Two Parts, *The Microwave Journal*, December 1963)

[Fig. 6-11, p. 270, Balanis]

To enhance the directivity of an end-fire array, Hansen and Woodyard proposed that the phase shift of an ordinary EFA

$$\beta = \pm kd \quad (14.1)$$

be increased as

$$\beta = -\left(kd + \frac{2.94}{N}\right) \text{ for a maximum at } \theta = 0^\circ, \quad (14.2)$$

$$\beta = +\left(kd + \frac{2.94}{N}\right) \text{ for a maximum at } \theta = 180^\circ. \quad (14.3)$$

Conditions (14.2)–(14.3) are known as the Hansen–Woodyard conditions for end-fire radiation. They follow from a procedure for maximizing the directivity, which we outline below.

The normalized pattern AF_n of a uniform linear array is

$$AF_n \approx \frac{\sin\left[\frac{N}{2}(kd \cos \theta + \beta)\right]}{\frac{N}{2}(kd \cos \theta + \beta)} \quad (14.4)$$

if the argument $\psi = kd \cos \theta + \beta$ is sufficiently small (see previous lecture). We are looking for an optimal β , which results in maximum directivity. Let

$$\beta = -pd, \quad (14.5)$$

where d is the array spacing and p is the optimization parameter. Then,

$$AF_n = \frac{\sin\left[\frac{Nd}{2}(k \cos \theta - p)\right]}{\frac{Nd}{2}(k \cos \theta - p)}.$$

For brevity, use the notation $Nd / 2 = q$. Then,

$$AF_n = \frac{\sin[q(k \cos \theta - p)]}{q(k \cos \theta - p)}, \quad (14.6)$$

$$\text{or } AF_n = \frac{\sin Z}{Z}, \text{ where } Z = q(k \cos \theta - p).$$

The radiation intensity becomes

$$U(\theta) = |AF_n|^2 = \frac{\sin^2 Z}{Z^2}, \quad (14.7)$$

$$U(\theta = 0) = \left\{ \frac{\sin[q(k-p)]}{q(k-p)} \right\}^2, \quad (14.8)$$

$$U_n(\theta) = \frac{U(\theta)}{U(\theta = 0)} = \left(\frac{z}{\sin z} \cdot \frac{\sin Z}{Z} \right)^2, \quad (14.9)$$

where

$$z = q(k-p),$$

$$Z = q(k \cos \theta - p), \text{ and}$$

$U_n(\theta)$ is normalized power pattern with respect to $\theta = 0^\circ$.

The directivity at $\theta = 0^\circ$ is

$$D_0 = \frac{4\pi U(\theta = 0)}{P_{rad}} \quad (14.10)$$

where $P_{rad} = \oiint_{\Omega} U_n(\theta) d\Omega$. To maximize the directivity, $U_0 = P_{rad} / 4\pi$ is minimized.

$$U_0 = \frac{1}{4\pi} \int_0^{2\pi} \int_0^\pi \left(\frac{z}{\sin z} \frac{\sin Z}{Z} \right)^2 \sin \theta d\theta d\theta, \quad (14.11)$$

$$U_0 = \frac{1}{2} \left(\frac{z}{\sin z} \right)^2 \int_0^\pi \left\{ \frac{\sin[q(k \cos \theta - p)]}{q(k \cos \theta - p)} \right\}^2 \sin \theta d\theta, \quad (14.12)$$

$$U_0 = \frac{1}{2kq} \left(\frac{z}{\sin z} \right)^2 \left[\frac{\pi}{2} + \frac{\cos 2z - 1}{2z} + \text{Si}(2z) \right] = \frac{1}{2kq} g(z). \quad (14.13)$$

Here, $\text{Si}(z) = \int_0^z (\sin t / t) dt$. The minimum of $g(z)$ occurs when

$$z = q(k-p) \approx -1.47, \quad (14.14)$$

$$\begin{aligned}
&\Rightarrow \frac{Nd}{2}(k-p) \approx -1.47. \\
&\Rightarrow \frac{Ndk}{2} - \frac{Ndp}{2} \approx -1.47, \text{ where } dp = -\beta \\
&\Rightarrow \frac{N}{2}(dk + \beta) \approx -1.47 \\
&\beta \approx -\frac{2.94}{N} - kd = -\left(kd + \frac{2.94}{N}\right). \tag{14.15}
\end{aligned}$$

Equation (14.15) gives the Hansen-Woodyard condition for improved directivity along $\theta = 0^\circ$. Similarly, for $\theta = 180^\circ$,

$$\beta = +\left(\frac{2.94}{N} + kd\right). \tag{14.16}$$

Usually, conditions (14.15) and (14.16) are approximated by

$$\beta \approx \pm\left(kd + \frac{\pi}{N}\right), \tag{14.17}$$

which is easier to remember and gives almost identical results since the curve $g(z)$ at its minimum is fairly flat.

Conditions (14.15)-(14.16), or (14.17), ensure minimum beamwidth (maximum directivity) in the end-fire direction. There is, however, a trade-off in the side-lobe level, which is higher than that of the ordinary EFA. Besides, conditions (14.15)-(14.16) have to be complemented by additional requirements, which would ensure low level of the radiation in the direction opposite to the main lobe.

(a) Maximum at $\theta = 0^\circ$ [reminder: $\psi = kd \cos \theta + \beta$]

$$\beta = -\left(kd + \frac{2.94}{N}\right) \Big|_{\theta=0^\circ} \Rightarrow \begin{cases} \psi_{\theta=0^\circ} = -\frac{2.94}{N} \\ \psi_{\theta=180^\circ} = -2kd - \frac{2.94}{N}. \end{cases} \tag{14.18}$$

Since we want to have a minimum of the pattern in the $\theta = 180^\circ$ direction, we must ensure that

$$|\psi|_{\theta=180^\circ} \approx \pi. \quad (14.19)$$

It is easier to remember the Hansen-Woodyard conditions for maximum directivity in the $\theta = 0^\circ$ direction as

$$|\psi|_{\theta=0^\circ} = \frac{2.94}{N} \approx \frac{\pi}{N}, \quad |\psi|_{\theta=180^\circ} \approx \pi. \quad (14.20)$$

(b) Maximum at $\theta = 180^\circ$

$$\beta = kd + \frac{2.94}{N} \Big|_{\theta=180^\circ} \Rightarrow \begin{cases} \psi_{\theta=180^\circ} = \frac{2.94}{N} \\ \psi_{\theta=0^\circ} = 2kd + \frac{2.94}{N}. \end{cases} \quad (14.21)$$

In order to have a minimum of the pattern in the $\theta = 0^\circ$ direction, we must ensure that

$$|\psi|_{\theta=0^\circ} \approx \pi. \quad (14.22)$$

We can now summarize the Hansen-Woodyard conditions for maximum directivity in the $\theta = 180^\circ$ direction as

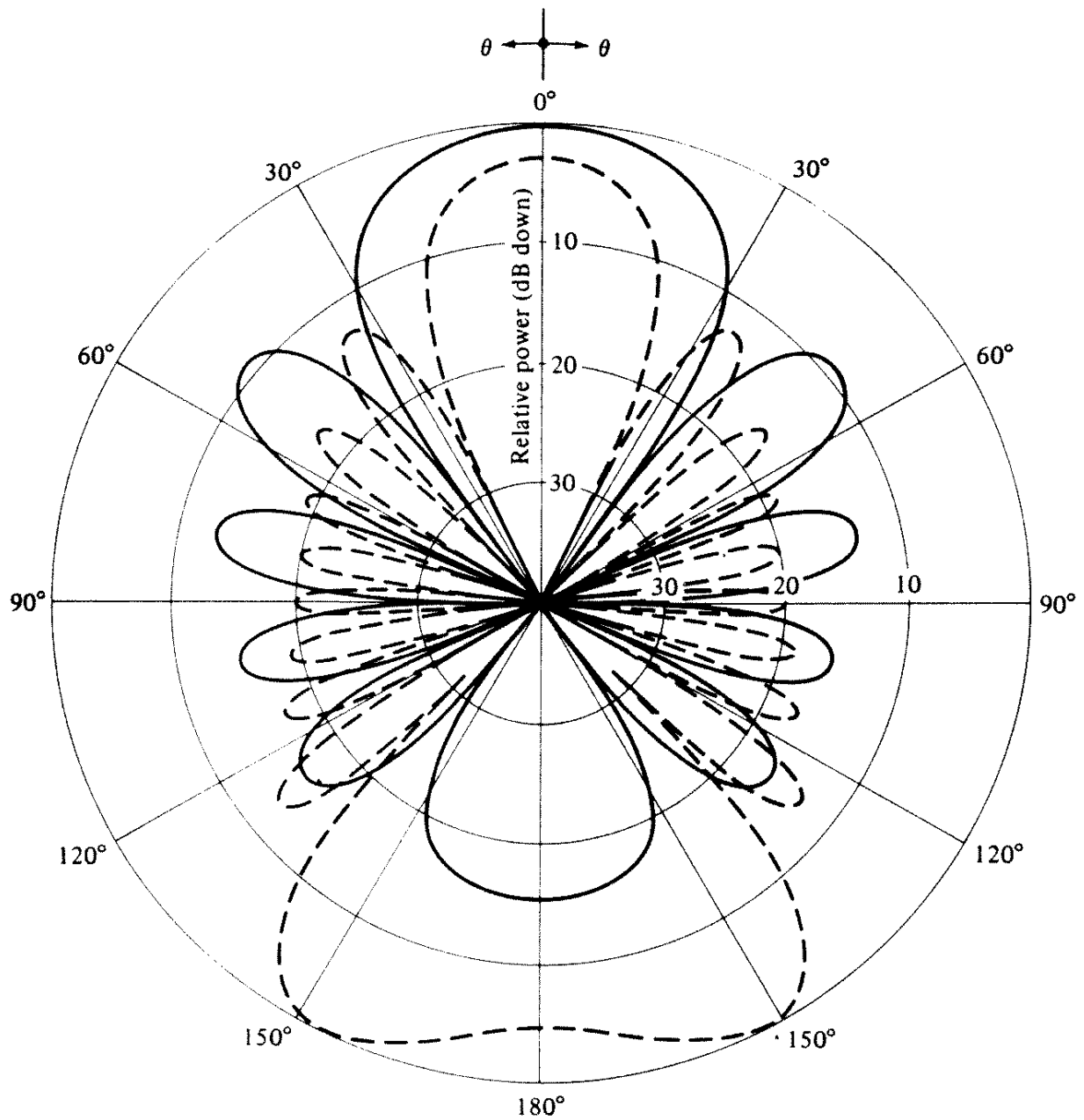
$$|\psi|_{\theta=180^\circ} = \frac{2.94}{N} \approx \frac{\pi}{N}, \quad |\psi|_{\theta=0^\circ} \approx \pi. \quad (14.23)$$

If (14.19) and (14.22) are not observed, the radiation in the opposite of the desired direction might even exceed the main beam level. It is easy to show that the complementary requirement $|\psi| = \pi$ at the opposite direction can be met if the following relation is observed:

$$d \approx \left(\frac{N-1}{N} \right) \frac{\lambda}{4}. \quad (14.24)$$

If N is large, $d \approx \lambda/4$. Thus, for a large uniform array, Hansen-Woodyard condition can yield improved directivity only if the spacing between the array elements is approximately $\lambda/4$.

ARRAY FACTORS OF A 10-ELEMENT UNIFORM-AMPLITUDE HW EFA



Solid line: $d = \lambda / 4$

Dash line: $d = \lambda / 2$

$N = 10$

$$\beta = -\left(kd + \frac{\pi}{N}\right)$$

Fig. 6.12, p. 273, Balanis

2. Directivity of a Linear Array

2.1. Directivity of a BSA

$$U(\theta) = |AF_n|^2 = \left[\frac{\sin\left(\frac{N}{2}kd \cos \theta\right)}{\frac{N}{2}kd \cos \theta} \right]^2 = \left(\frac{\sin Z}{Z}\right)^2 \quad (14.25)$$

$$D_0 = 4\pi \frac{U_0}{P_{rad}} = \frac{U_0}{U_{av}}, \quad (14.26)$$

where $U_{av} = P_{rad} / (4\pi)$. The radiation intensity in the direction of maximum radiation $\theta = \pi / 2$ in terms of AF_n is unity:

$$\begin{aligned} U_0 &= U_{\max} = U(\theta = \pi / 2) = 1, \\ \Rightarrow D_0 &= U_{av}^{-1}. \end{aligned} \quad (14.27)$$

The radiation intensity averaged over all directions is calculated as

$$U_{av} = \frac{1}{4\pi} \int_0^{2\pi} \int_0^\pi \frac{\sin^2 Z}{Z^2} \sin \theta d\theta d\phi = \frac{1}{2} \int_0^\pi \left| \frac{\sin\left(\frac{N}{2}kd \cos \theta\right)}{\frac{N}{2}kd \cos \theta} \right|^2 \sin \theta d\theta.$$

Change variable:

$$Z = \frac{N}{2}kd \cos \theta \Rightarrow dZ = -\frac{N}{2}kd \sin \theta d\theta. \quad (14.28)$$

Then,

$$U_{av} = -\frac{1}{2N kd} \int_{\frac{Nkd}{2}}^{\frac{Nkd}{2}} \left(\frac{\sin Z}{Z}\right)^2 dZ, \quad (14.29)$$

$$U_{av} = \frac{1}{Nkd} \int_{-\frac{Nkd}{2}}^{\frac{Nkd}{2}} \left(\frac{\sin Z}{Z}\right)^2 dZ. \quad (14.30)$$

The function $(Z^{-1} \sin Z)^2$ is a relatively fast decaying function as Z increases. That is why, for large arrays, where $Nkd / 2$ is big enough (≥ 20), the integral (14.30) can be approximated by

$$U_{av} \approx \frac{1}{Nkd} \int_{-\infty}^{\infty} \left(\frac{\sin Z}{Z} \right)^2 dZ = \frac{\pi}{Nkd}, \quad (14.31)$$

$$D_0 = \frac{1}{U_{av}} \approx \frac{Nkd}{\pi} = 2N \left(\frac{d}{\lambda} \right). \quad (14.32)$$

Substituting the length of the array $L = (N - 1)d$ in (14.32) yields

$$D_0 \approx 2 \underbrace{\left(1 + \frac{L}{d} \right)}_N \left(\frac{d}{\lambda} \right). \quad (14.33)$$

For a large array ($L \gg d$),

$$D_0 \approx 2L / \lambda. \quad (14.34)$$

2.2. Directivity of ordinary EFA

Consider an EFA with maximum radiation at $\theta = 0^\circ$, i.e., $\beta = -kd$.

$$U(\theta) = |AF_n|^2 = \left\{ \frac{\sin \left[\frac{N}{2} kd (\cos \theta - 1) \right]}{\left[\frac{N}{2} kd (\cos \theta - 1) \right]} \right\}^2 = \left(\frac{\sin Z}{Z} \right)^2, \quad (14.35)$$

where $Z = \frac{N}{2} kd (\cos \theta - 1)$. The averaged radiation intensity is

$$U_{av} = \frac{P_{rad}}{4\pi} = \frac{1}{4\pi} \int_0^{2\pi} \int_0^\pi \left(\frac{\sin Z}{Z} \right)^2 \sin \theta d\theta d\phi = \frac{1}{2} \int_0^\pi \left(\frac{\sin Z}{Z} \right)^2 \sin \theta d\theta.$$

Since

$$Z = \frac{N}{2} kd (\cos \theta - 1) \text{ and } dZ = -\frac{N}{2} kd \sin \theta d\theta, \quad (14.36)$$

it follows that

$$U_{av} = -\frac{1}{2} \frac{2}{Nkd} \int_0^{-Nkd/2} \left(\frac{\sin Z}{Z} \right)^2 dZ,$$

$$U_{av} = \frac{1}{Nkd} \int_0^{Nkd/2} \left(\frac{\sin Z}{Z} \right)^2 dZ. \quad (14.37)$$

If (Nkd) is sufficiently large, the above integral can be approximated as

$$U_{av} \approx \frac{1}{Nkd} \int_0^{\infty} \left(\frac{\sin Z}{Z} \right)^2 dZ = \frac{1}{Nkd} \cdot \frac{\pi}{2}. \quad (14.38)$$

The directivity then becomes

$$\Rightarrow D_0 \approx \frac{1}{U_{av}} = \frac{2Nkd}{\pi} = 4N \left(\frac{d}{\lambda} \right). \quad (14.39)$$

The comparison of (14.39) and (14.32) shows that the directivity of an EFA is approximately twice as large as the directivity of the BSA.

Another (equivalent) expression can be derived for D_0 of the EFA in terms of the array length $L = (N-1)d$:

$$D_0 = 4 \left(1 + \frac{L}{d} \right) \left(\frac{d}{\lambda} \right). \quad (14.40)$$

For large arrays, the following approximation holds:

$$D_0 = 4L / \lambda \quad \text{if } L \gg d. \quad (14.41)$$

2.3. Directivity of HW EFA

If the radiation has its maximum at $\theta = 0^\circ$, then the minimum of U_{av} is obtained as in (14.13):

$$U_{av}^{\min} = \frac{1}{2k} \frac{2}{Nd} \left[\frac{Z_{\min}}{\sin Z_{\min}} \right]^2 \left[\frac{\pi}{2} + \frac{\cos(2Z_{\min}) - 1}{2Z_{\min}} + \text{Si}(2Z_{\min}) \right], \quad (14.42)$$

where $Z_{\min} = -1.47 \approx -\pi / 2$.

$$\Rightarrow U_{av}^{\min} = \frac{1}{Nkd} \left(\frac{\pi}{2} \right)^2 \left[\frac{\pi}{2} + \frac{2}{\pi} - 1.8515 \right] = \frac{0.878}{Nkd}. \quad (14.43)$$

The directivity is then

$$D_0 = \frac{1}{U_{av}^{\min}} = \frac{Nkd}{0.878} = 1.789 \left[4N \left(\frac{d}{\lambda} \right) \right]. \quad (14.44)$$

From (14.44), we can see that using the HW conditions leads to improvement of the directivity of the EFA with a factor of 1.789. Equation (14.44) can be expressed via the length L of the array as

$$D_0 = 1.789 \left[4 \left(1 + \frac{L}{d} \right) \left(\frac{d}{\lambda} \right) \right] = 1.789 \left[4 \left(\frac{L}{\lambda} \right) \right]. \quad (14.45)$$

Example: Given a linear uniform array of N isotropic elements ($N = 10$), find the directivity D_0 if:

- $\beta = 0$ (BSA)
- $\beta = -kd$ (Ordinary EFA)
- $\beta = -kd - \pi / N$ (Hansen-Woodyard EFA)

In all cases, $d = \lambda / 4$.

a) BSA

$$D_0 \approx 2N \left(\frac{d}{\lambda} \right) = 5 \quad (6.999 \text{ dB})$$

b) Ordinary EFA

$$D_0 \approx 4N \left(\frac{d}{\lambda} \right) = 10 \quad (10 \text{ dB})$$

c) HW EFA

$$D_0 \approx 1.789 \left[4N \left(\frac{d}{\lambda} \right) \right] = 17.89 \quad (12.53 \text{ dB})$$

3. Pattern Characteristics of Linear Uniform Arrays – Recapitulation

A. Broad-side array

NULLS ($AF_n = 0$):

$$\theta_n = \arccos\left(\pm \frac{n \lambda}{N d}\right), \text{ where } n = 1, 2, 3, 4, \dots \text{ and } n \neq N, 2N, 3N, \dots$$

MAXIMA ($AF_n = 1$):

$$\theta_n = \arccos\left(\pm \frac{m \lambda}{d}\right), \text{ where } m = 0, 1, 2, 3, \dots$$

HALF-POWER POINTS:

$$\theta_h \approx \arccos\left(\pm \frac{1.391 \lambda}{\pi N d}\right), \text{ where } \frac{\pi d}{\lambda} \ll 1$$

HALF-POWER BEAMWIDTH:

$$\Delta \theta_h = 2 \left[\frac{\pi}{2} - \arccos\left(\frac{1.391 \lambda}{\pi N d}\right) \right], \frac{\pi d}{\lambda} \ll 1$$

MINOR LOBE MAXIMA:

$$\theta_s \approx \arccos\left[\pm \frac{\lambda}{2d} \left(\frac{2s+1}{N}\right)\right], \text{ where } s = 1, 2, 3, \dots \text{ and } \frac{\pi d}{\lambda} \ll 1$$

FIRST-NULL BEAMWIDTH (FNBW):

$$\Delta \theta_n = 2 \left[\frac{\pi}{2} - \arccos\left(\frac{\lambda}{N d}\right) \right]$$

FIRST SIDE LOBE BEAMWIDTH (FSLBW):

$$\Delta \theta_s = 2 \left[\frac{\pi}{2} - \arccos\left(\frac{3 \lambda}{2 N d}\right) \right], \frac{\pi d}{\lambda} \ll 1$$

B. Ordinary end-fire array

NULLS ($AF_n = 0$):

$$\theta_n = \arccos\left(1 - \frac{n\lambda}{Nd}\right), \text{ where } n = 1, 2, 3, \dots \text{ and } n \neq N, 2N, 3N, \dots$$

MAXIMA ($AF_n = 1$):

$$\theta_n = \arccos\left(1 - \frac{m\lambda}{d}\right), \text{ where } m = 0, 1, 2, 3, \dots$$

HALF-POWER POINTS:

$$\theta_h = \arccos\left(1 - \frac{1.391\lambda}{\pi Nd}\right), \text{ where } \frac{\pi d}{\lambda} \ll 1$$

HALF-POWER BEAMWIDTH:

$$\Delta\theta_h = 2\arccos\left(1 - \frac{1.391\lambda}{\pi Nd}\right), \frac{\pi d}{\lambda} \ll 1$$

MINOR LOBE MAXIMA:

$$\theta_s = \arccos\left[1 - \frac{(2s+1)\lambda}{2Nd}\right], \text{ where } s = 1, 2, 3, \dots \text{ and } \frac{\pi d}{\lambda} \ll 1$$

FIRST-NULL BEAMWIDTH:

$$\Delta\theta_n = 2\arccos\left(1 - \frac{\lambda}{Nd}\right)$$

FIRST SIDE LOBE BEAMWIDTH:

$$\Delta\theta_s = 2\arccos\left(1 - \frac{3\lambda}{2Nd}\right), \frac{\pi d}{\lambda} \ll 1$$

C. Hansen-Woodyard end-fire array

NULLS:

$$\theta_n = \arccos \left[1 + (1 - 2n) \frac{\lambda}{2Nd} \right], \text{ where } n = 1, 2, \dots \text{ and } n \neq N, 2N, \dots$$

MINOR LOBE MAXIMA:

$$\theta_s = \arccos \left(1 - \frac{s\lambda}{Nd} \right), \text{ where } s = 1, 2, 3, \dots \text{ and } \frac{\pi d}{\lambda} \ll 1$$

SECONDARY MAXIMA:

$$\theta_m = \arccos \left\{ 1 + [1 - (2m + 1)] \frac{\lambda}{2Nd} \right\}, \text{ where } m = 1, 2, \dots \text{ and } \frac{\pi d}{\lambda} \ll 1$$

HALF-POWER POINTS:

$$\theta_h = \arccos \left(1 - 0.1398 \frac{\lambda}{Nd} \right), \text{ where } \frac{\pi d}{\lambda} \ll 1, \text{ } N\text{-large}$$

HALF-POWER BEAMWIDTH:

$$\Delta\theta_h = 2 \arccos \left(1 - 0.1398 \frac{\lambda}{Nd} \right), \text{ where } \frac{\pi d}{\lambda} \ll 1, \text{ } N\text{-Large}$$

FIRST-NULL BEAMWIDTH:

$$\Delta\theta_n = 2 \arccos \left(1 - \frac{\lambda}{2Nd} \right)$$

4. 3-D Characteristics of a Linear Array

In the previous considerations, it was always assumed that the linear-array elements are located along the z -axis, thus, creating a problem, which is symmetrical around the z -axis. If the array axis has an arbitrary orientation, the array factor can be expressed as

$$AF = \sum_{n=1}^N a_n e^{j(n-1)(kd \cos \gamma + \beta)} = \sum_{n=1}^N a_n e^{j(n-1)\psi}, \quad (14.46)$$

where a_n is the excitation amplitude and $\psi = kd \cos \gamma + \beta$.

The angle γ is subtended between the array axis and the position vector to the observation point. Thus, if the array axis is along the unit vector $\hat{\mathbf{a}}$,

$$\hat{\mathbf{a}} = \sin \theta_a \cos \phi_a \hat{\mathbf{x}} + \sin \theta_a \sin \phi_a \hat{\mathbf{y}} + \cos \theta_a \hat{\mathbf{z}} \quad (14.47)$$

and the position vector to the observation point is

$$\hat{\mathbf{r}} = \sin \theta \cos \phi \hat{\mathbf{x}} + \sin \theta \sin \phi \hat{\mathbf{y}} + \cos \theta \hat{\mathbf{z}} \quad (14.48)$$

the angle γ can be found as

$$\begin{aligned} \cos \gamma &= \hat{\mathbf{a}} \cdot \hat{\mathbf{r}} = \sin \theta \cos \phi \sin \theta_a \cos \phi_a \hat{\mathbf{x}} + \sin \theta \sin \phi \sin \theta_a \sin \phi_a \hat{\mathbf{y}} + \cos \theta \cos \theta_a \hat{\mathbf{z}}, \\ &\Rightarrow \cos \gamma = \sin \theta \sin \theta_a \cos(\phi - \phi_a) + \cos \theta \cos \theta_a. \end{aligned} \quad (14.49)$$

If $\hat{\mathbf{a}} = \hat{\mathbf{z}}$ ($\theta_a = 0^\circ$), then $\cos \gamma = \cos \theta$, $\gamma = \theta$.

LECTURE 15: LINEAR ARRAYS – PART III

(*N*-element linear arrays with uniform spacing and non-uniform amplitude: Binomial array; Dolph–Tschebyscheff array. Directivity and design.)

1. Advantages of Linear Arrays with Nonuniform Amplitude Distribution

The most often met BSAs, classified according to the type of their excitation amplitudes, are:

- the uniform BSA – relatively high directivity, but the side-lobe levels are high;
- Dolph–Tschebyscheff (or Chebyshev)¹ BSA – for a given number of elements, maximum directivity is next after that of the uniform BSA; side-lobe levels are the lowest in comparison with the other two types of arrays for a given directivity;
- binomial BSA – does not have good directivity but has very low side-lobe levels (when $d = \lambda / 2$, there are no side lobes at all).

2. Array Factor of Linear Arrays with Nonuniform Amplitude Distribution

Let us consider a linear array with an even number ($2M$) of elements, located symmetrically along the z -axis, with excitation, which is also symmetrical with respect to $z = 0$. For a *broadside array* ($\beta = 0$),

$$AF^e = a_1 e^{j\frac{1}{2}kd \cos \theta} + a_2 e^{j\frac{3}{2}kd \cos \theta} + \dots + a_M e^{j\frac{2M-1}{2}kd \cos \theta} + a_1 e^{-j\frac{1}{2}kd \cos \theta} + a_2 e^{-j\frac{3}{2}kd \cos \theta} + \dots + a_M e^{-j\frac{2M-1}{2}kd \cos \theta}, \quad (15.1)$$

$$\Rightarrow AF^e = 2 \sum_{n=1}^M a_n \cos \left[\left(\frac{2n-1}{2} \right) kd \cos \theta \right]. \quad (15.2)$$

If the linear array consists of an odd number ($2M+1$) of elements, located symmetrically along the z -axis, the array factor is

$$AF^o = 2a_1 + a_2 e^{jkd \cos \theta} + a_3 e^{j2kd \cos \theta} + \dots + a_{M+1} e^{jMkd \cos \theta} + a_2 e^{-jkd \cos \theta} + a_3 e^{-j2kd \cos \theta} + \dots + a_{M+1} e^{-jMkd \cos \theta}, \quad (15.3)$$

¹ Russian spelling is Чебышёв.

$$\Rightarrow AF^o = 2 \sum_{n=1}^{M+1} a_n \cos[(n-1)kd \cos \theta]. \quad (15.4)$$

EVEN- AND ODD-NUMBER ARRAYS

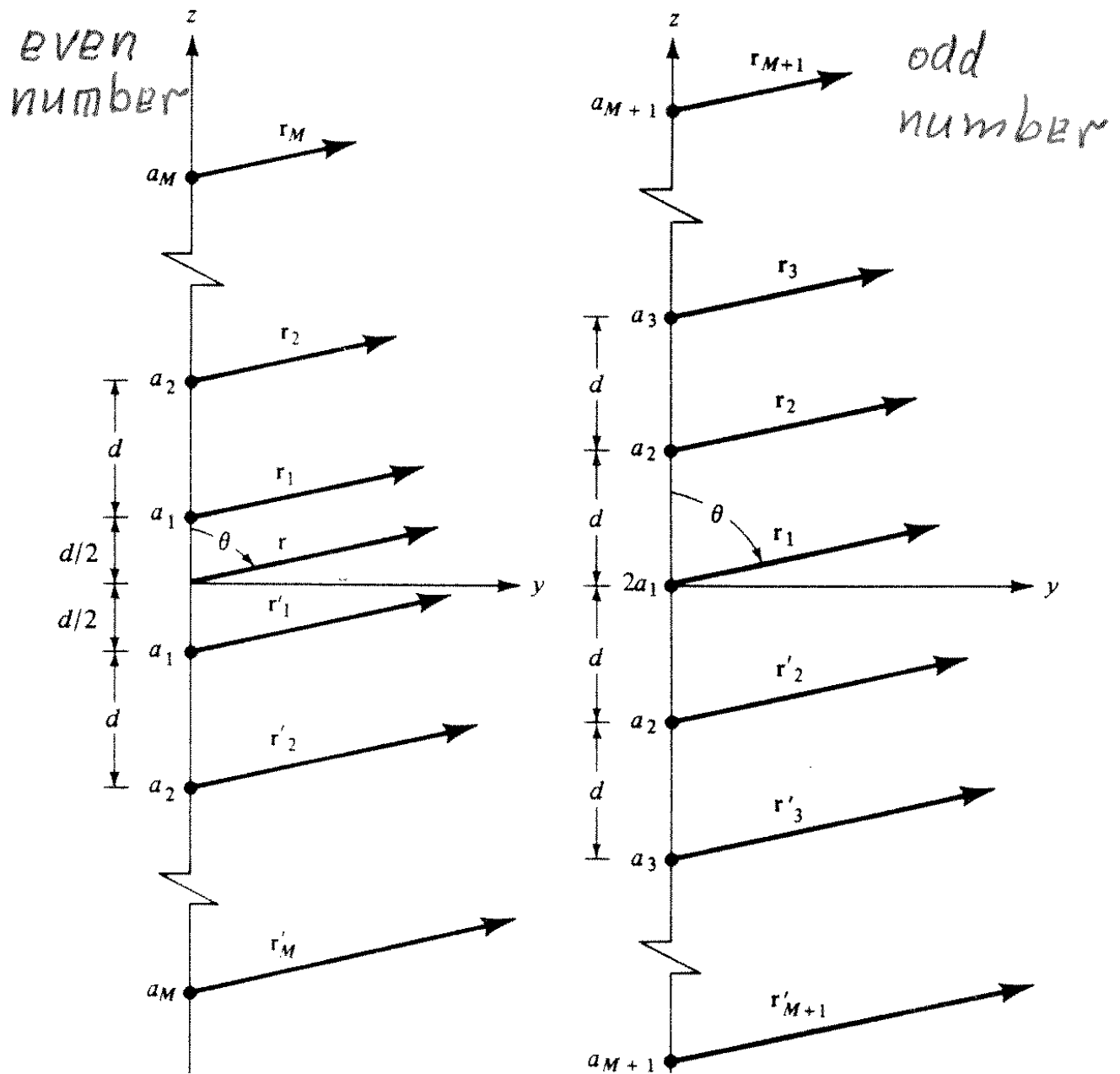


Fig. 6.17, p. 291, Balanis

The normalized AF derived from (15.2) and (15.4) can be written in the form

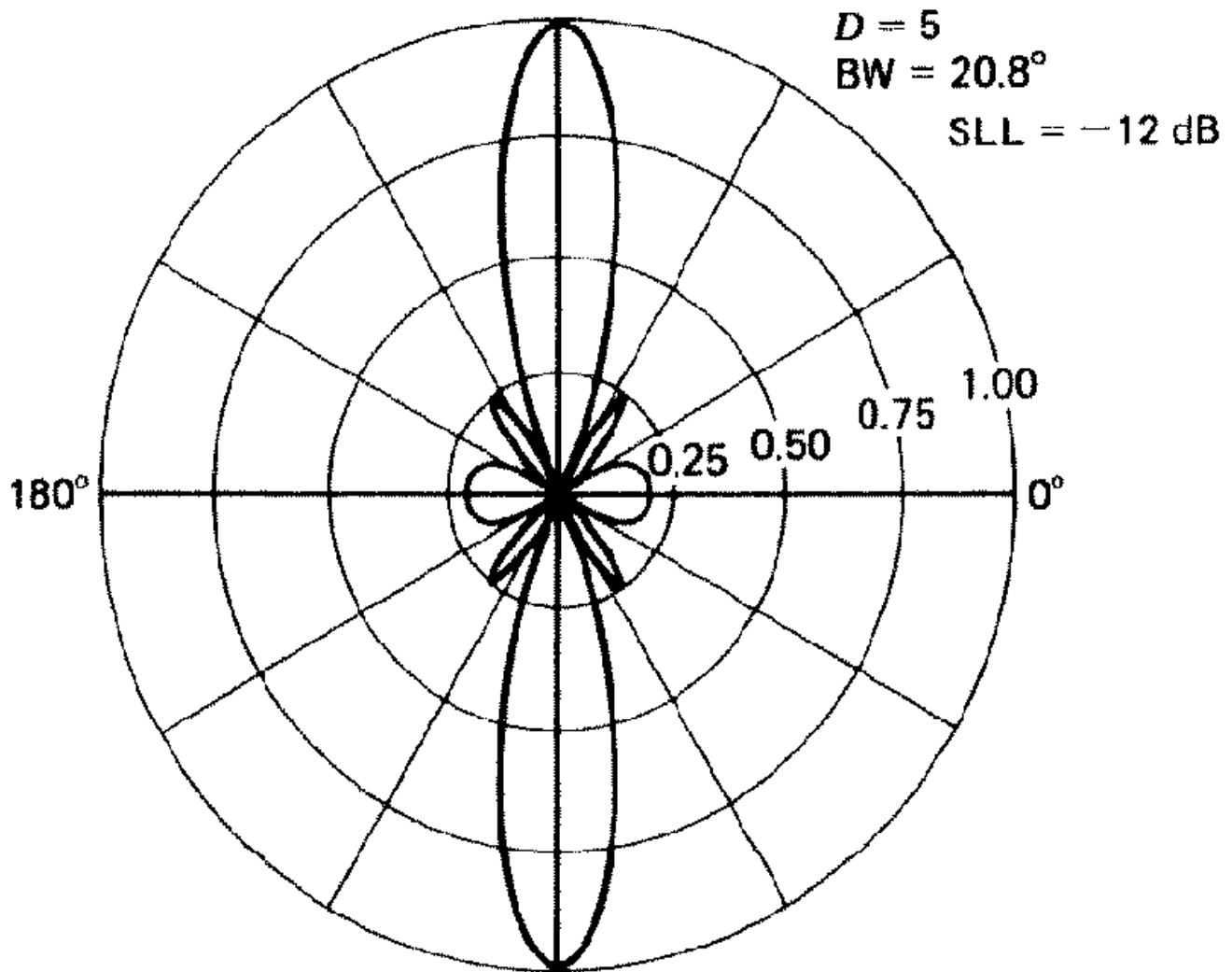
$$AF^e = \sum_{n=1}^M a_n \cos[(2n-1)u], \text{ for } N = 2M, \quad (15.5)$$

$$AF^o = \sum_{n=1}^{M+1} a_n \cos[2(n-1)u], \text{ for } N = 2M + 1, \quad (15.6)$$

where $u = \frac{1}{2}kd \cos \theta = \frac{\pi d}{\lambda} \cos \theta$.

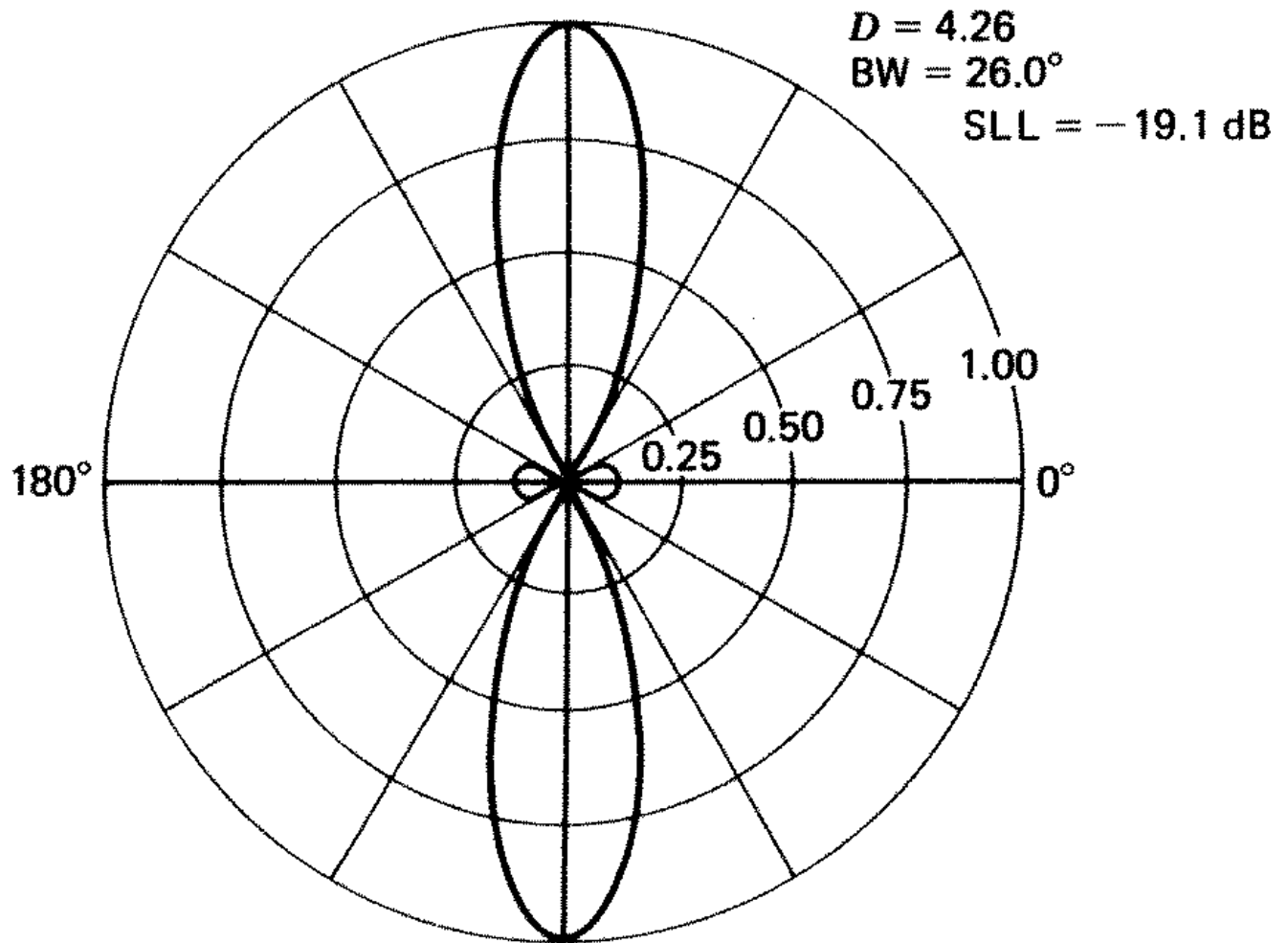
Examples of AFs of arrays of nonuniform amplitude distribution

a) **uniform** amplitude distribution ($N = 5$, $d = \lambda / 2$, max. at $\theta_0 = 90^\circ$)



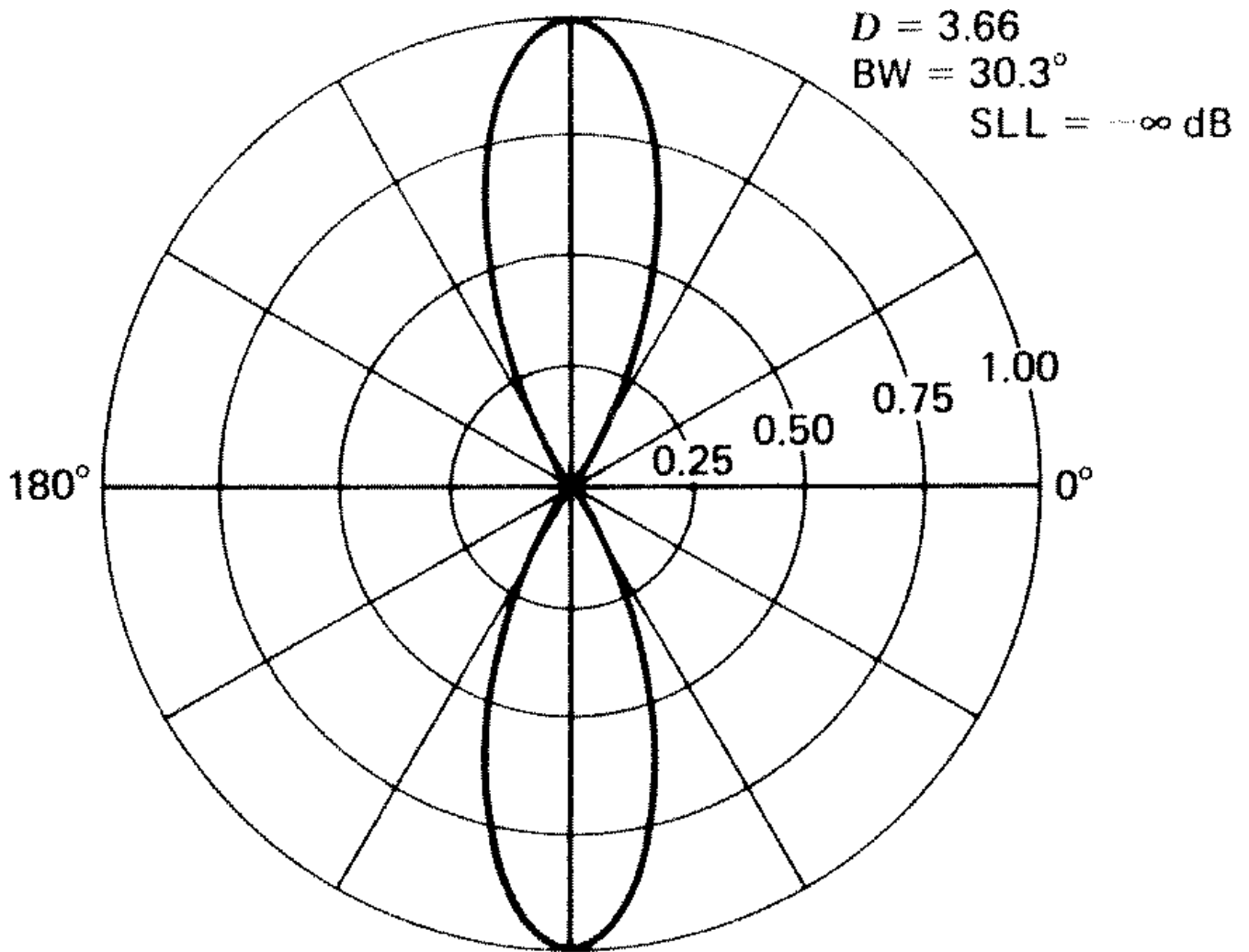
pp. 148-149, Stutzman

b) **triangular** (1:2:3:2:1) amplitude distribution ($N = 5$, $d = \lambda / 2$, max. at $\theta_0 = 90^\circ$)



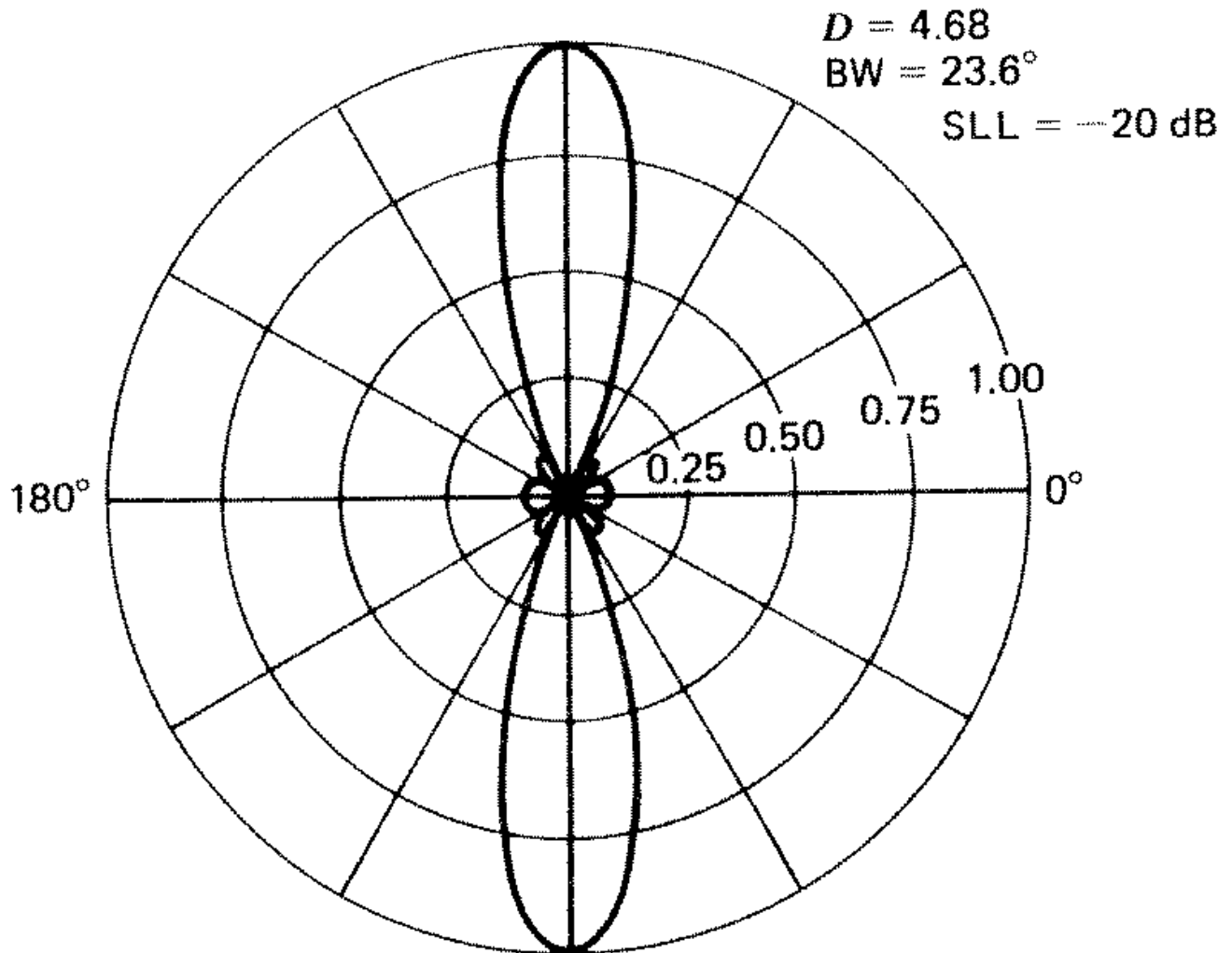
pp. 148-149, Stutzman

c) **binomial** (1:4:6:4:1) amplitude distribution ($N = 5$, $d = \lambda/2$, max. at $\theta_0 = 90^\circ$)



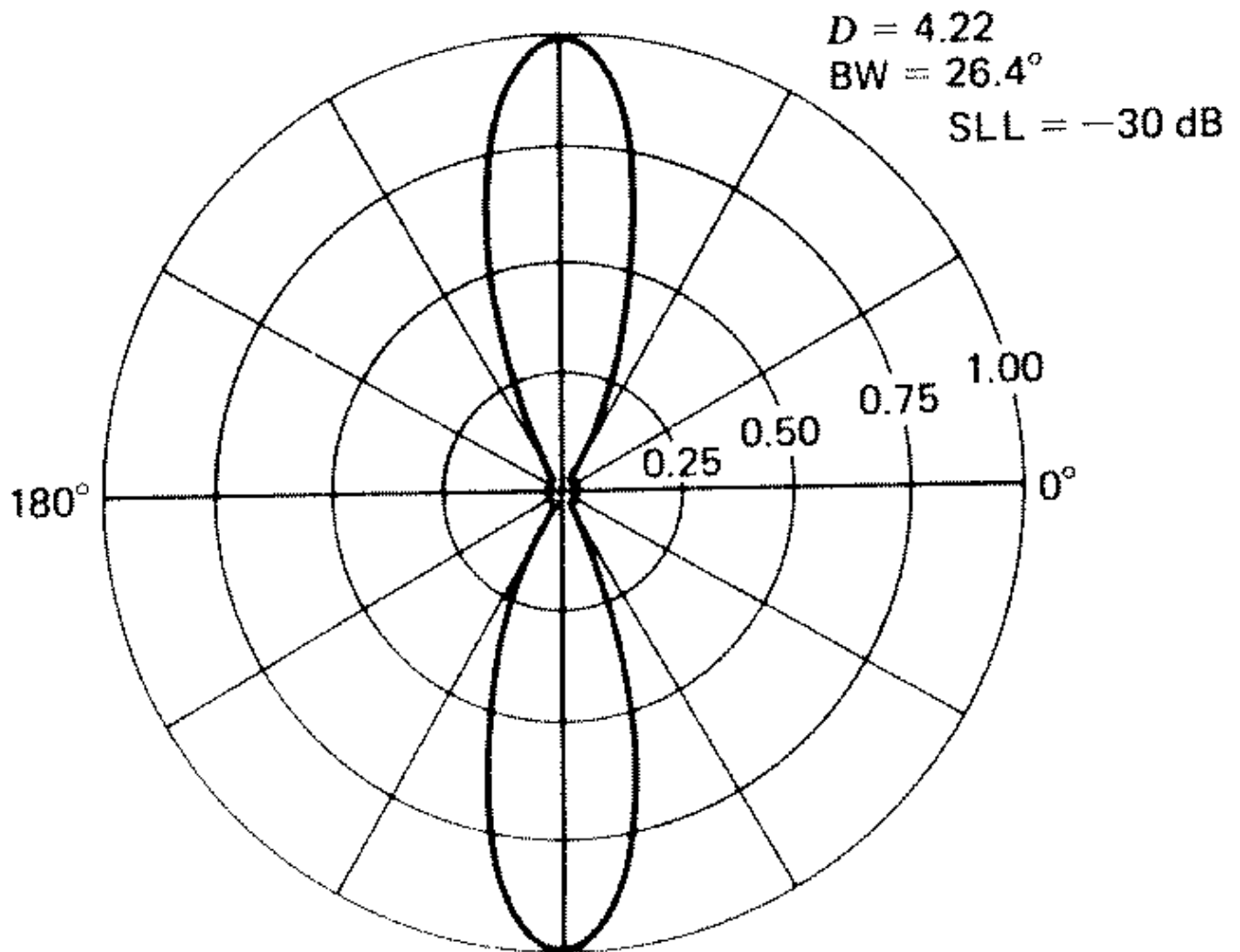
pp. 148-149, Stutzman

d) **Dolph-Tschebyschev** (1:1.61:1.94:1.61:1) amplitude distribution ($N = 5$,
 $d = \lambda / 2$, max. at $\theta_0 = 90^\circ$)



pp. 148-149, Stutzman

e) **Dolph-Tschebyschev** (1:2.41:3.14:2.41:1) amplitude distribution ($N = 5$, $d = \lambda / 2$, max. at $\theta_0 = 90^\circ$)

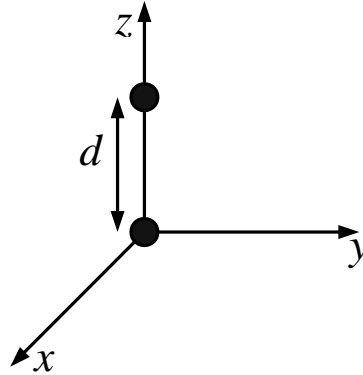


pp. 148-149, Stutzman

Notice that as the current amplitude is tapered more gradually toward the edges of the array, the side lobes tend to decrease and the beamwidth tends to increase.

3. Binomial Broadside Array

The binomial BSA was investigated and proposed by J. S. Stone² to synthesize patterns without side lobes. First, consider a 2–element array (along the z -axis).



The elements of the array are identical and their excitations are the same. The array factor is of the form

$$AF = 1 + Z, \text{ where } Z = e^{j\psi} = e^{j(kd \cos \theta + \beta)}. \quad (15.7)$$

If the spacing is $d \leq \lambda / 2$ and $\beta = 0$ (broad-side maximum), the array pattern $|AF|$ has no side lobes at all. This is proven as follows.

$$|AF|^2 = (1 + \cos \psi)^2 + \sin^2 \psi = 2(1 + \cos \psi) = 4 \cos^2(\psi / 2) \quad (15.8)$$

where $\psi = kd \cos \theta$. The first null of the array factor is obtained from (15.8) as

$$\frac{1}{2} \cdot \frac{2\pi}{\lambda} \cdot d \cos \theta_{n1,2} = \pm \frac{\pi}{2} \Rightarrow \theta_{n1,2} = \pm \arccos\left(\frac{\lambda}{2d}\right). \quad (15.9)$$

As long as $d < \lambda / 2$, the first null does not exist. If $d = \lambda / 2$, then $\theta_{n1,2} = 0, 180^\circ$. Thus, in the “visible” range of θ , all secondary lobes are eliminated.

Second, consider a 2–element array whose elements are identical and the same as the array given above. The distance between the two arrays is again d .

² US Patents #1,643,323, #1,715,433.

An approximate closed-form expression for the HPBW of a BA with $d = \lambda / 2$ is

$$HPBW \approx \frac{1.06}{\sqrt{N-1}} = \frac{1.06}{\sqrt{2L/\lambda}} = \frac{1.75}{\sqrt{L/\lambda}}, \quad (15.12)$$

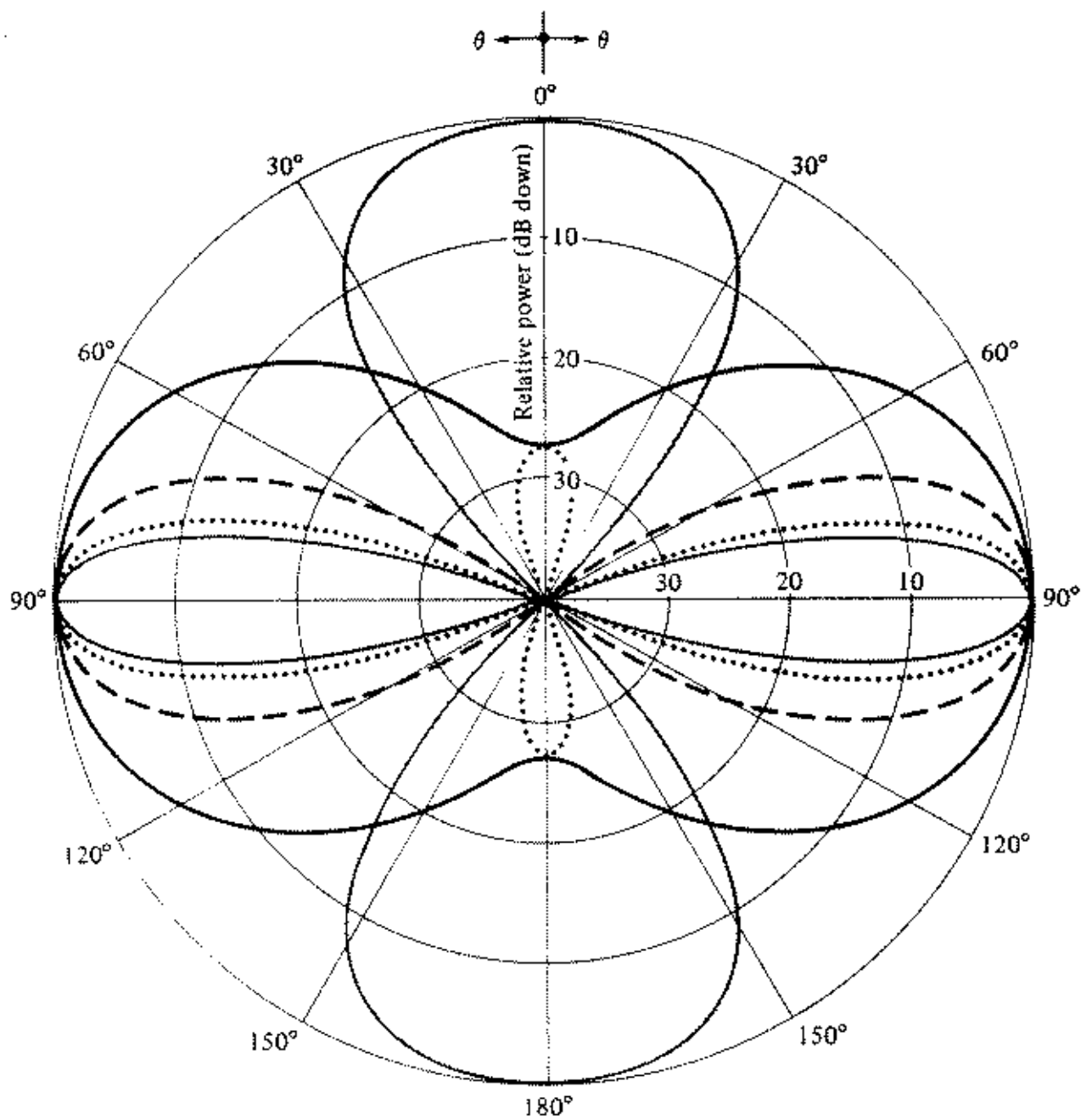
where $L = (N-1)d$ is the array's length. The AFs of 10-element broadside binomial arrays ($N = 10$) are given below.

The directivity of a broadside BA with spacing $d = \lambda / 2$ can be calculated as

$$D_0 = \frac{2}{\int_0^\pi \left[\cos\left(\frac{\pi}{2} \cos\theta\right) \right]^{2(N-1)} d\theta}, \quad (15.13)$$

$$D_0 = \frac{(2N-2) \cdot (2N-4) \cdot \dots \cdot 2}{(2N-3) \cdot (2N-5) \cdot \dots \cdot 1}, \quad (15.14)$$

$$D_0 \approx 1.77\sqrt{N} = 1.77\sqrt{1+2L/\lambda}. \quad (15.15)$$



- $d = \lambda/4$
- - - - - $d = \lambda/2$
- $d = 3\lambda/4$
- . - . - $d = \lambda$

10-element broadside binomial array ($N=10$)

Fig. 6.18, p.293, Balanis

4. Dolph–Chebyshev Array (DCA)

Dolph proposed (in 1946) a method to design arrays with any desired side-lobe levels and any HPBW. This method is based on the approximation of the pattern of the array by a Chebyshev polynomial of order m , high enough to meet the requirement for the side-lobe levels. A DCA with no side lobes (side-lobe level of $-\infty$ dB) reduces to the binomial design.

4.1. Chebyshev polynomials

The Chebyshev polynomial of order m is defined by

$$T_m(z) = \begin{cases} (-1)^m \cosh(m \cdot \operatorname{arccosh} |z|), & z \leq -1, \\ \cos(m \cdot \arccos(z)), & -1 \leq z \leq 1, \\ \cosh(m \cdot \operatorname{arccosh}(z)), & z \geq 1. \end{cases} \quad (15.16)$$

A Chebyshev polynomial $T_m(z)$ of any order m can be derived via a recursion formula, provided $T_{m-1}(z)$ and $T_{m-2}(z)$ are known:

$$T_m(z) = 2zT_{m-1}(z) - T_{m-2}(z). \quad (15.17)$$

Explicitly, from (15.16) we see that

$$m = 0, \quad T_0(z) = 1$$

$$m = 1, \quad T_1(z) = z.$$

Then, (15.17) produces:

$$m = 2, \quad T_2(z) = 2z^2 - 1$$

$$m = 3, \quad T_3(z) = 4z^3 - 3z$$

$$m = 4, \quad T_4(z) = 8z^4 - 8z^2 + 1$$

$$m = 5, \quad T_5(z) = 16z^5 - 20z^3 + 5z, \text{ etc.}$$

If $|z| \leq 1$, then the Chebyshev polynomials are related to the cosine functions, see (15.16). We can always expand the function $\cos(mx)$ as a polynomial of $\cos(x)$ of order m , e.g., for $m = 2$,

$$\cos 2x = 2 \cos^2 x - 1. \quad (15.18)$$

The expansion of $\cos(mx)$ can be done by observing that $(e^{jx})^m = e^{jmx}$ and by making use of Euler's formula as

$$(\cos x + j \sin x)^m = \cos(mx) + j \sin(mx). \quad (15.19)$$

The left side of the equation is then expanded and its real and imaginary parts are equated to those on the right. Similar relations hold for the hyperbolic cosine function cosh.

Comparing the trigonometric relation in (15.18) with the expression for $T_2(z)$ above (see the expanded Chebyshev polynomials after (15.17)), we see that the Chebyshev argument z is related to the cosine argument x by

$$z = \cos x \quad \text{or} \quad x = \arccos z. \quad (15.20)$$

For example, (15.18) can be written as:

$$\begin{aligned} \cos(2 \arccos z) &= 2[\cos(\arccos z)]^2 - 1, \\ \Rightarrow \cos(2 \arccos z) &= 2z^2 - 1 = T_2(z). \end{aligned} \quad (15.21)$$

Properties of the Chebyshev polynomials:

- 1) All polynomials of any order m pass through the point (1,1).
- 2) Within the range $-1 \leq z \leq 1$, the polynomials have values within $[-1,1]$.
- 3) All nulls occur within $-1 \leq z \leq 1$.
- 4) The maxima and minima in the $z \in [-1,1]$ range have values +1 and -1, respectively.
- 5) The higher the order of the polynomial, the steeper the slope for $|z| > 1$.

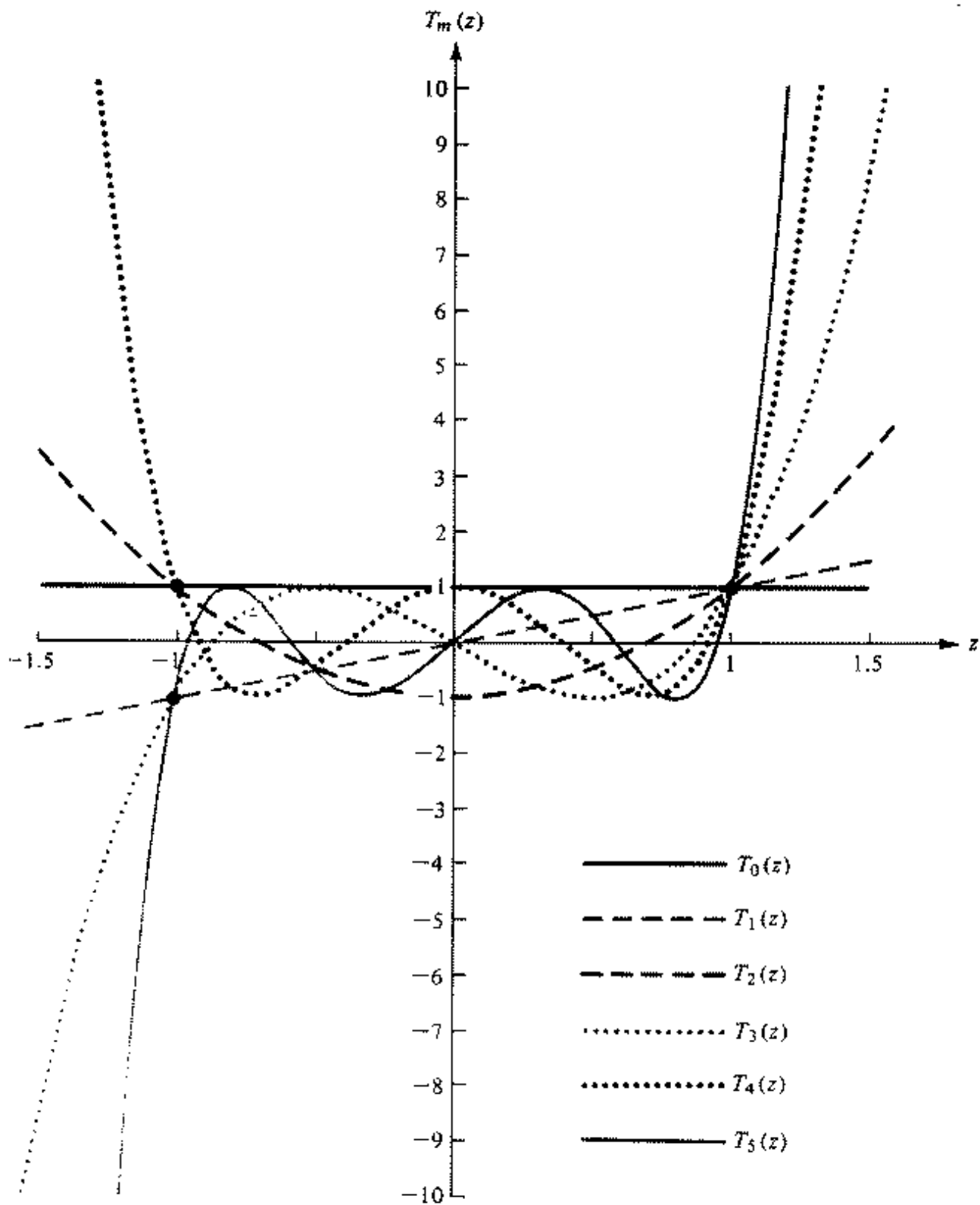


Fig. 6.19, pp. 296, Balanis

4.2. Chebyshev array design

The main goal is to approximate the desired AF with a Chebyshev polynomial such that

- the side-lobe level meets the requirements, and
- the main beam width is as small as possible.

An array of N elements has an AF approximated with a Chebyshev polynomial of order m , which is

$$m = N - 1. \quad (15.22)$$

In general, for a given side-lobe level, the higher the order m of the polynomial, the narrower the beamwidth. However, for $m > 10$, the difference is not substantial – see the slopes of $T_m(z)$ in the previous figure. The AF of an N -element array (15.5) or (15.6) is shaped by a Chebyshev polynomial by requiring that

$$T_{N-1}(z) = \begin{cases} \sum_{n=1}^M a_n \cos[(2n-1)u], & M = N/2, \quad \text{even} \\ \sum_{n=1}^{M+1} a_n \cos[2(n-1)u], & M = (N-1)/2, \quad \text{odd} \end{cases} \quad (15.23)$$

Here, $u = (\pi d / \lambda) \cos \theta$. Let the side-lobe level be

$$R_0 = \frac{E_{\max}}{E_{sl}} = \frac{1}{AF_{sl}} \quad (\text{voltage ratio}). \quad (15.24)$$

Then, we require that the maximum of T_{N-1} is fixed at an argument z_0 ($|z_0| > 1$), where

$$T_{N-1}^{\max}(z_0) = R_0. \quad (15.25)$$

Equation (15.25) corresponds to $AF(u) = AF^{\max}(u_0)$. Obviously, z_0 must satisfy the condition:

$$|z_0| > 1, \quad (15.26)$$

where $T_{N-1} > 1$. The maxima of $|T_{N-1}(z)|$ for $|z| \leq 1$ are equal to unity and they correspond to the side lobes of the AF. Thus, $AF(u)$ has side-lobe levels equal to R_0 . The AF is a polynomial of $\cos u$, and the $T_{N-1}(z)$ is a polynomial of z where z is limited to the range

$$-1 \leq z \leq z_0. \quad (15.27)$$

Since

$$-1 \leq \cos u \leq 1, \quad (15.28)$$

the relation between z and $\cos u$ must be normalized as

$$\cos u = z / z_0. \quad (15.29)$$

Design of a DCA of N elements – general procedure:

- 1) Expand the AF as given by (15.5) or (15.6) by replacing each $\cos(mu)$ term ($m = 1, 2, \dots, M$) with the power series of $\cos u$.
- 2) Determine z_0 such that $T_{N-1}(z_0) = R_0$ (voltage ratio).
- 3) Substitute $\cos u = z / z_0$ in the AF found in step 1.
- 4) Equate the AF found in Step 3 to $T_{N-1}(z)$ and determine the coefficients for each power of z .

Example: Design a DCA (broadside) of $N=10$ elements with a major-to-minor lobe ratio of $R_0 = 26$ dB. Find the excitation coefficients and form the AF.

Solution:

The order of the Chebyshev polynomial is $m = N - 1 = 9$. The AF for an even-number array is:

$$AF_{2M} = \sum_{n=1}^5 a_n \cos[(2n-1)u], \quad u = \frac{\pi d}{\lambda} \cos \theta, \quad M = 5.$$

Step 1: Write AF_{10} explicitly:

$$AF_{10} = a_1 \cos u + a_2 \cos 3u + a_3 \cos 5u + a_4 \cos 7u + a_5 \cos 9u.$$

Expand the $\cos(mu)$ terms as powers of $\cos u$:

$$\cos 3u = 4 \cos^3 u - 3 \cos u,$$

$$\cos 5u = 16 \cos^5 u - 20 \cos^3 u + 5 \cos u,$$

$$\cos 7u = 64 \cos^7 u - 112 \cos^5 u + 56 \cos^3 u - 7 \cos u,$$

$$\cos 9u = 256 \cos^9 u - 576 \cos^7 u + 432 \cos^5 u - 120 \cos^3 u + 9 \cos u.$$

Note that the above expansions can be readily obtained from the recursive Chebyshev relation (15.17), $-1 \leq z \leq z_0$, and the substitution $z = \cos u$. For example,

$$m = 3, \quad T_3(z) = 4z^3 - 3z$$

translates into

$$\cos(3u) = 4\cos^3 u - 3\cos u.$$

Step 2: Determine z_0 :

$$\begin{aligned} R_0 = 26 \text{ dB} &\Rightarrow R_0 = 10^{26/20} \approx 20 \Rightarrow T_9(z_0) = 20, \\ \cosh[9\operatorname{arccosh}(z_0)] &= 20, \\ 9\operatorname{arccosh}(z_0) &= \operatorname{arccosh}20 = 3.69, \\ \operatorname{arccosh}(z_0) &= 0.41, \\ z_0 = \cosh 0.41 &\Rightarrow z_0 = 1.08515. \end{aligned}$$

Step 3: Express the AF from Step 1 in terms of $\cos u = z / z_0$ and make equal to the Chebyshev polynomial:

$$\begin{aligned} AF_{10} &= \frac{z}{z_0} (a_1 - 3a_2 + 5a_3 - 7a_4 + 9a_5) \\ &+ \frac{z^3}{z_0^3} (4a_2 - 20a_3 + 56a_4 - 120a_5) \\ &+ \frac{z^5}{z_0^5} (16a_3 - 112a_4 + 432a_5) \\ &+ \frac{z^7}{z_0^7} (64a_4 - 576a_5) \\ &+ \frac{z^9}{z_0^9} (256a_5) = \\ &= \underbrace{9z - 120z^3 + 432z^5 - 576z^7 + 256z^9}_{T_9(z)} \end{aligned}$$

Step 4: Find the coefficients by matching the power terms:

$$256a_5 = 256z_0^9 \Rightarrow a_5 = 2.0860$$

$$64a_4 - 576a_5 = -576z_0^7 \Rightarrow a_4 = 2.8308$$

$$16a_3 - 112a_4 + 432a_5 = 432z_0^5 \Rightarrow a_3 = 4.1184$$

$$4a_2 - 20a_3 + 56a_4 - 120a_5 = -120z_0^3 \Rightarrow a_2 = 5.2073$$

$$a_1 - 3a_2 + 5a_3 - 7a_4 + 9a_5 = 9z_0^1 \Rightarrow a_1 = 5.8377$$

Normalize coefficients with respect to edge element ($N=5$):

$$a_5 = 1; a_4 = 1.357; a_3 = 1.974; a_2 = 2.496; a_1 = 2.789$$

$$\Rightarrow AF_{10} = 2.789 \cos(u) + 2.496 \cos(3u) + 1.974 \cos(5u) + 1.357 \cos(7u) + \cos(9u)$$

where $u = \frac{\pi d}{\lambda} \cos \theta$.

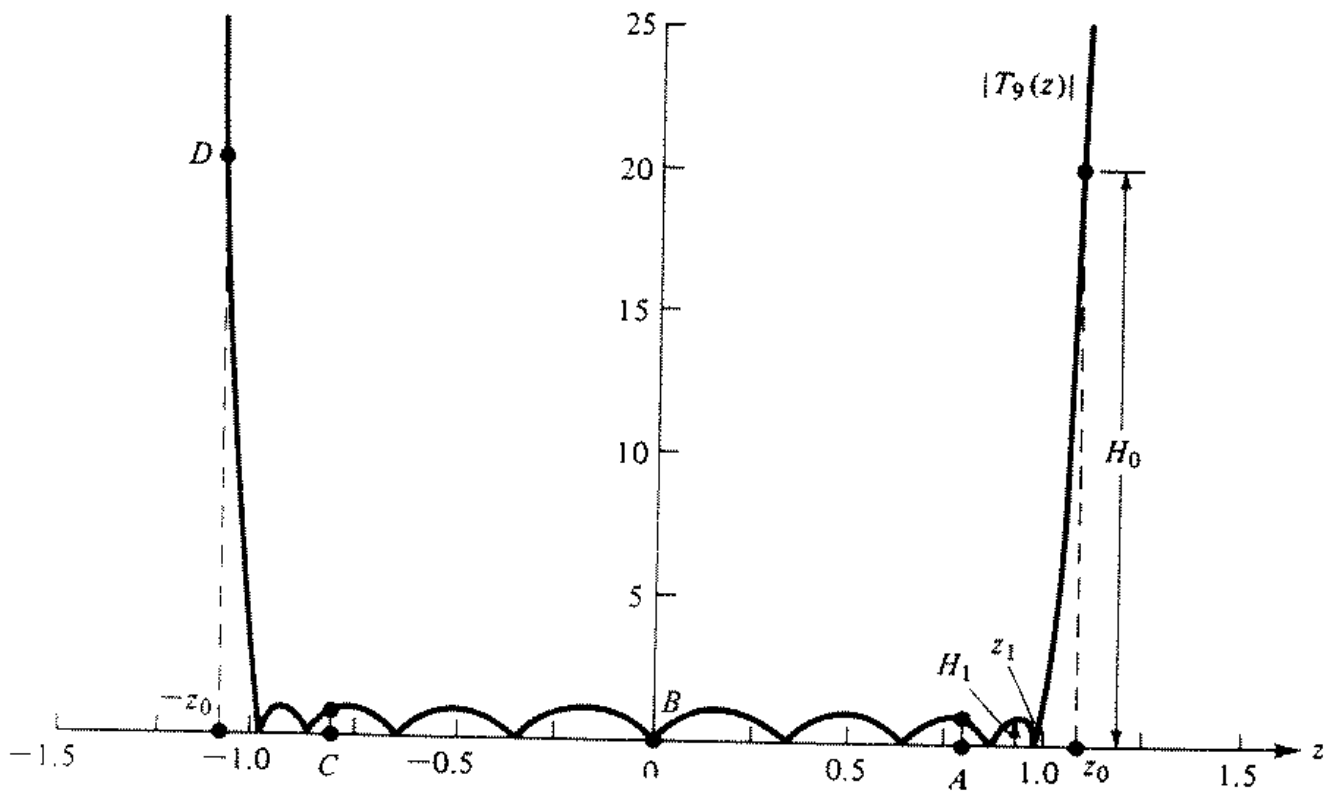


Fig. 6.20b, p. 298, Balanis

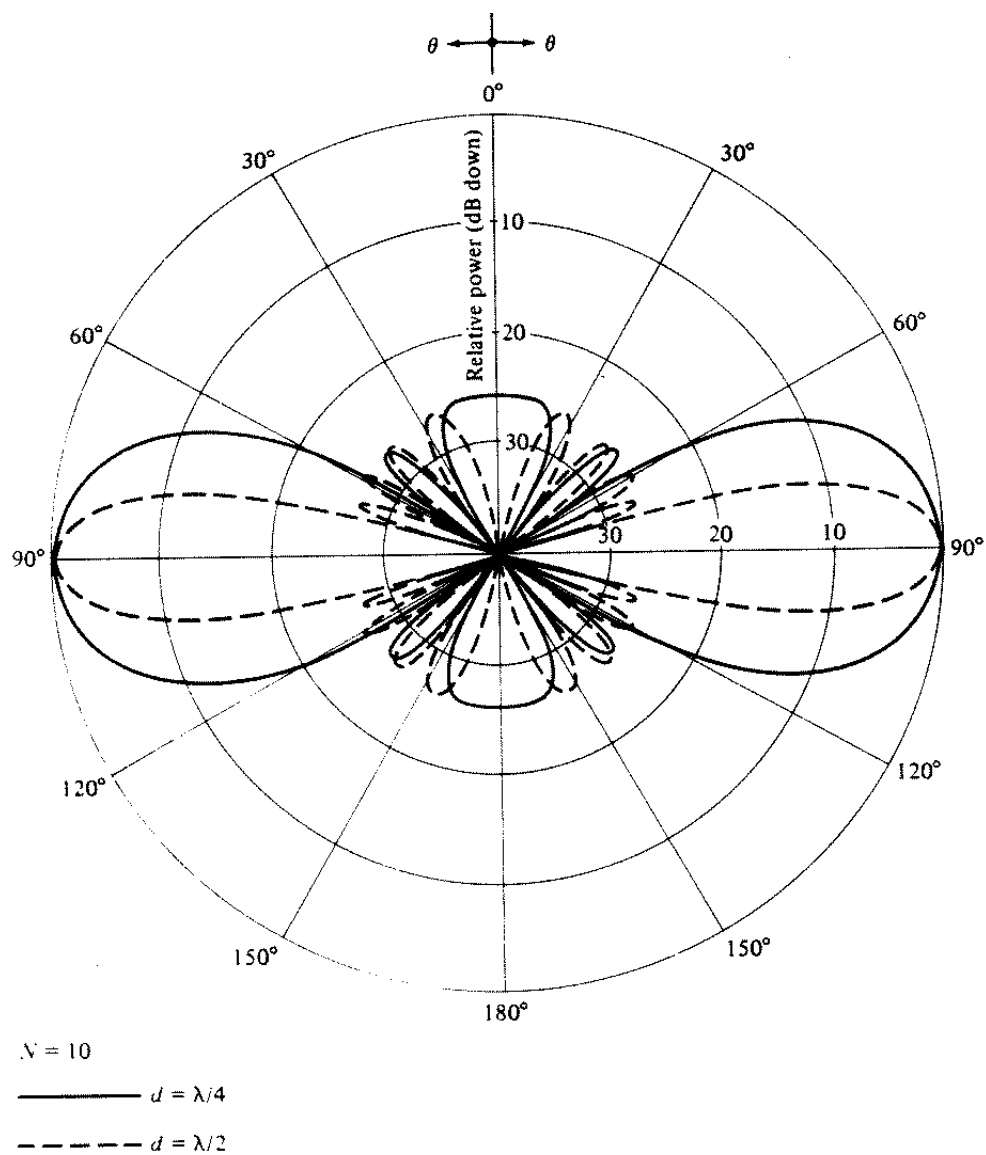


Fig. 6.21, p. 300, Balanis

4.3. Maximum affordable d for Dolph-Chebyshev arrays

This restriction arises from the requirement for a single major lobe – see also equation (15.27), $-1 \leq z \leq z_0$:

$$z \geq -1, \cos u = z / z_0, u = \frac{\pi d}{\lambda} \cos \theta,$$

$$\Rightarrow z = z_0 \cos\left(\frac{\pi d}{\lambda} \cos \theta\right) \geq -1. \quad (15.30)$$

For a given array, when θ varies from 0° to 180° , the argument z assumes values

$$\text{from } z_{(\theta=0^\circ)} = z_0 \cos\left(\frac{\pi d}{\lambda}\right) \quad (15.31)$$

$$\text{through } z_{(\theta=90^\circ)} = z_0 \quad (15.32)$$

$$\text{to } z_{(\theta=180^\circ)} = z_0 \cos\left(-\frac{\pi d}{\lambda}\right) = z_{(\theta=0^\circ)}. \quad (15.33)$$

The extreme value of z to the left on the abscissa corresponds to the end-fire directions of the AF. This value must not go beyond $z = -1$. Otherwise, end-fire lobes of levels higher than 1 (higher than R_0) will appear. Therefore, the inequality (15.30) must hold for $\theta = 0^\circ$ or 180° :

$$z_0 \cos\left(\frac{\pi d}{\lambda}\right) \geq -1 \Rightarrow \cos\left(\frac{\pi d}{\lambda}\right) \geq -\frac{1}{z_0}. \quad (15.34)$$

Let

$$\gamma = \arccos\left(z_0^{-1}\right). \quad (15.35)$$

Remember that $z_0 > 1$; thus γ is a real-valued angle. Then,

$$\frac{\pi d}{\lambda} < \pi - \gamma = \pi - \arccos\left(\frac{1}{z_0}\right) \quad (15.36)$$

or

$$\frac{\pi d_{\max}}{\lambda} = \pi - \underbrace{\arccos\left(z_0^{-1}\right)}_{\gamma} \Rightarrow \boxed{\frac{d_{\max}}{\lambda} = 1 - \frac{1}{\pi} \arccos\left(\frac{1}{z_0}\right)} \quad (15.37)$$

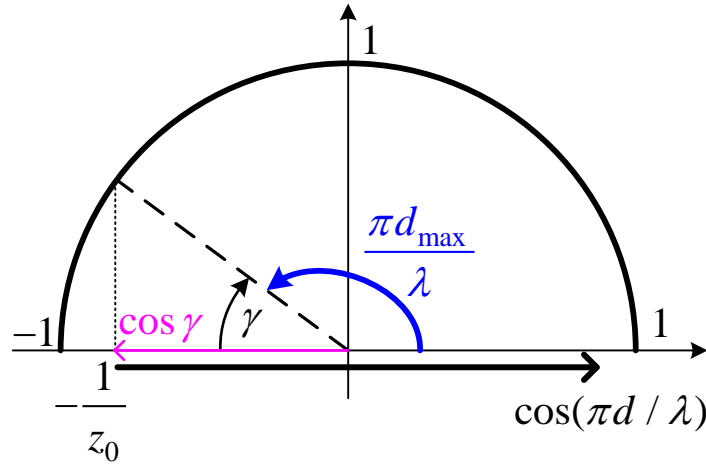


ILLUSTRATION OF EQUATION (15.34) AND THE REQUIREMENT IN (15.36)

For the case of the previous example,

$$\frac{d}{\lambda} < 1 - \frac{1}{\pi} \arccos\left(\frac{1}{1.08515}\right) = 1 - \frac{0.39879}{\pi} = 0.873,$$

$$d_{\max} = 0.873\lambda.$$

5. Directivity of Non-uniform Arrays

It is difficult to derive closed form expressions for the directivity of non-uniform arrays. Here, we derive expressions in the form of series in the most general case of a linear array when the excitation coefficients are known.

The non-normalized array factor is

$$AF = \sum_{n=0}^{N-1} a_n e^{j\beta_n} e^{jkz_n \cos\theta}, \quad (15.38)$$

where

- a_n is the amplitude of the excitation of the n -th element;
- β_n is the phase angle of the excitation of the n -th element;
- z_n is the z -coordinate of the n -th element.

The maximum AF is

$$AF_{\max} = \sum_{n=0}^{N-1} a_n. \quad (15.39)$$

The normalized AF is

$$AF_n = \frac{AF}{AF_{\max}} = \frac{\sum_{n=0}^{N-1} a_n e^{j\beta_n} e^{jkz_n \cos \theta}}{\sum_{n=0}^{N-1} a_n}. \quad (15.40)$$

The beam solid angle of a linear array is

$$\Omega_A = 2\pi \int_0^{\pi} |AF_n(\theta)|^2 \sin \theta d\theta, \quad (15.41)$$

$$\Omega_A = \frac{2\pi}{\left(\sum_{n=0}^{N-1} a_n\right)^2} \sum_{m=0}^{N-1} \sum_{p=0}^{N-1} a_m a_p e^{j(\beta_m - \beta_p)} \int_0^{\pi} e^{jk(z_m - z_p) \cos \theta} \sin \theta d\theta,$$

where

$$\int_0^{\pi} e^{jk(z_m - z_p) \cos \theta} \sin \theta d\theta = \frac{2 \sin[k(z_m - z_p)]}{k(z_m - z_p)}.$$

$$\Rightarrow \Omega_A = \frac{4\pi}{\left(\sum_{n=0}^{N-1} a_n\right)^2} \sum_{m=0}^{N-1} \sum_{p=0}^{N-1} a_m a_p e^{j(\beta_m - \beta_p)} \cdot \frac{\sin[k(z_m - z_p)]}{k(z_m - z_p)}. \quad (15.42)$$

From

$$D_0 = 4\pi / \Omega_A,$$

we obtain

$$\Rightarrow D_0 = \frac{\left(\sum_{n=0}^{N-1} a_n\right)^2}{\sum_{m=0}^{N-1} \sum_{p=0}^{N-1} a_m a_p e^{j(\beta_m - \beta_p)} \cdot \frac{\sin[k(z_m - z_p)]}{k(z_m - z_p)}}. \quad (15.43)$$

For equispaced linear ($z_n = nd$) arrays, (15.43) reduces to

$$D_0 = \frac{\left(\sum_{n=0}^{N-1} a_n \right)^2}{\sum_{m=0}^{N-1} \sum_{p=0}^{N-1} a_m a_p e^{j(\beta_m - \beta_p)} \cdot \frac{\sin[(m-p)kd]}{(m-p)kd}}. \quad (15.44)$$

For equispaced broadside arrays, where $\beta_m = \beta_p$ for any (m,p) , (15.44) reduces to

$$D_0 = \frac{\left(\sum_{n=0}^{N-1} a_n \right)^2}{\sum_{m=0}^{N-1} \sum_{p=0}^{N-1} a_m a_p \cdot \frac{\sin[(m-p)kd]}{(m-p)kd}}. \quad (15.45)$$

For equispaced broadside uniform arrays,

$$D_0 = \frac{N^2}{\sum_{m=0}^{N-1} \sum_{p=0}^{N-1} \frac{\sin[(m-p)kd]}{(m-p)kd}}. \quad (15.46)$$

When the spacing d is a multiple of $\lambda / 2$, equation (15.45) reduces to

$$D_0 = \frac{\left(\sum_{n=0}^{N-1} a_n \right)^2}{\sum_{n=0}^{N-1} (a_n)^2}, \quad d = \frac{\lambda}{2}, \lambda, \dots \quad (15.47)$$

Example: Calculate the directivity of the Dolph–Chebyshev array designed in the previous example if $d = \lambda / 2$.

The 10-element DCA has the following amplitude distribution:

$$a_5 = 1; \quad a_4 = 1.357; \quad a_3 = 1.974; \quad a_2 = 2.496; \quad a_1 = 2.798.$$

We make use of (15.47):

$$D_0 = \frac{4 \left(\sum_{n=1}^5 a_n \right)^2}{2 \sum_{n=1}^5 (a_n)^2} = 2 \cdot \frac{(9.625)^2}{20.797} = 8.9090 \quad (9.5 \text{ dB}).$$

Output from ARRAYS.m: $D_0 = 8.9276$.

6. Half-power Beamwidth of a BS DCA

For large DCAs with side lobes in the range from -20 dB to -60 dB, the HPBW $HPBW_{DCA}$ can be found from the HPBW of a uniform array $HPBW_{UA}$ by introducing a beam-broadening factor f given by

$$f = 1 + 0.636 \left\{ \frac{2}{R_0} \cosh \left[\sqrt{(\operatorname{arccosh} R_0)^2 - \pi^2} \right] \right\}^2, \quad (15.48)$$

so that

$$HPBW_{DCA} = f \times HPBW_{UA}. \quad (15.49)$$

In (15.48), R_0 denotes the side-lobe level (the voltage ratio).

LECTURE 16: PLANAR ARRAYS AND CIRCULAR ARRAYS

1. Planar Arrays

Planar arrays provide directional beams, symmetrical patterns with low side lobes, much higher directivity (narrow main beam) than that of their individual element. In principle, they can point the main beam toward any direction.

Applications – tracking radars, remote sensing, communications, etc.

A. The array factor of a rectangular planar array

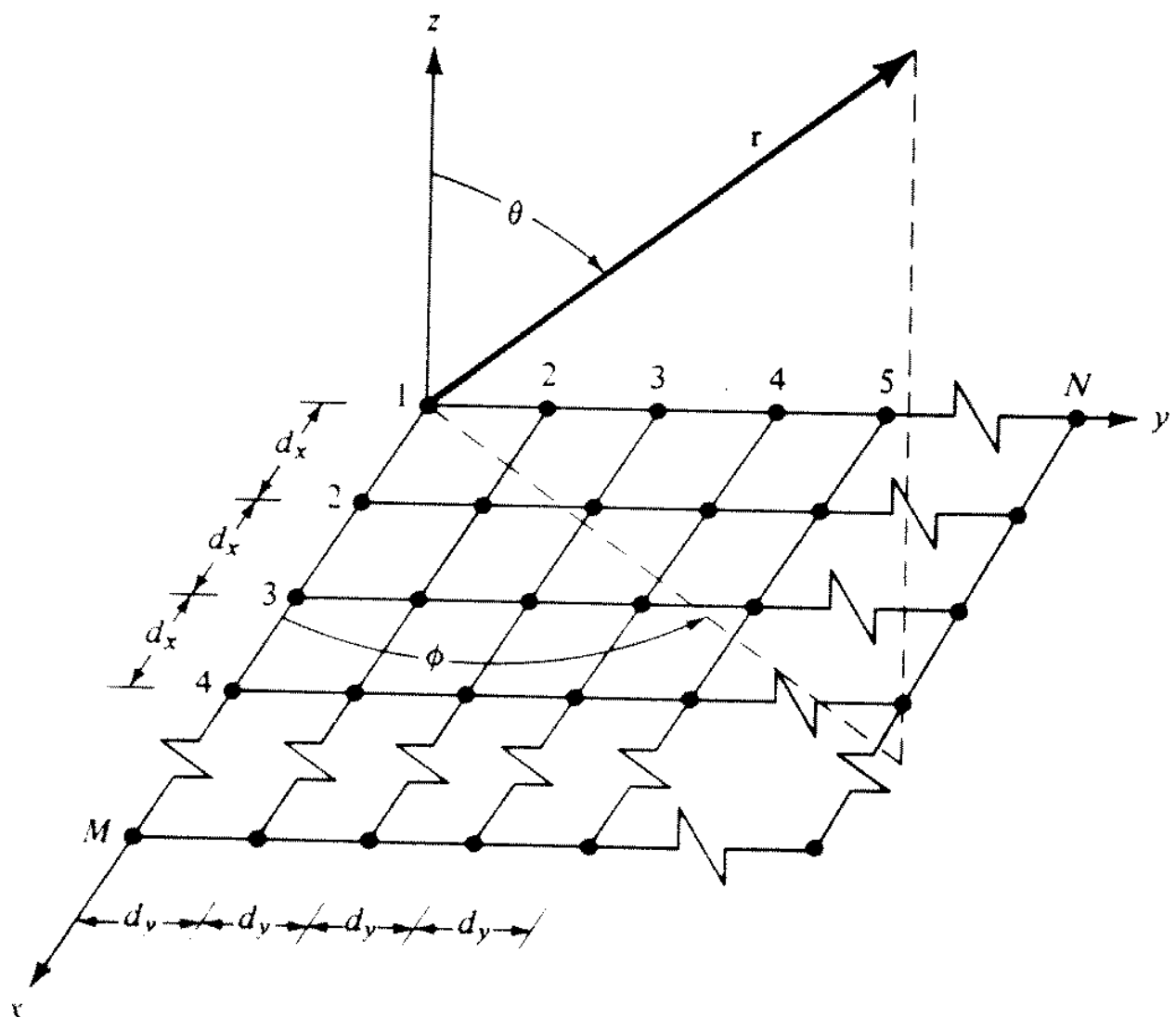


Fig. 6.23b, p. 310, Balanis

The AF of a linear array of M elements along the x -axis is

$$AF_{x1} = \sum_{m=1}^M I_{m1} e^{j(m-1)(kd_x \sin \theta \cos \phi + \beta_x)} \quad (16.1)$$

where $\sin \theta \cos \phi = \cos \gamma_x$ is the directional cosine with respect to the x -axis (γ_x is the angle between \mathbf{r} and the x axis). It is assumed that all elements are equispaced with an interval of d_x and a progressive shift β_x . I_{m1} denotes the excitation amplitude of the element at the point with coordinates $x = (m-1)d_x$, $y = 0$. In the figure above, this is the element of the m -th row and the 1st column of the array matrix. Note that the 1st row corresponds to $x = 0$.

If N such arrays are placed at even intervals along the y direction, a rectangular array is formed. We assume again that they are equispaced at a distance d_y and there is a progressive phase shift β_y along each row. We also assume that the normalized current distribution along each of the x -directed arrays is the same but the absolute values correspond to a factor of I_{1n} ($n = 1, \dots, N$). Then, the AF of the entire $M \times N$ array is

$$AF = \sum_{n=1}^N I_{1n} \left[\sum_{m=1}^M I_{m1} e^{j(m-1)(kd_x \sin \theta \cos \phi + \beta_x)} \right] \cdot e^{j(n-1)(kd_y \sin \theta \sin \phi + \beta_y)}, \quad (16.2)$$

or

$$AF = S_{xM} \cdot S_{yN}, \quad (16.3)$$

where

$$S_{xM} = AF_{x1} = \sum_{m=1}^M I_{m1} e^{j(m-1)(kd_x \sin \theta \cos \phi + \beta_x)}, \text{ and}$$

$$S_{yN} = AF_{1y} = \sum_{n=1}^N I_{1n} e^{j(n-1)(kd_y \sin \theta \sin \phi + \beta_y)}.$$

In the array factors above,

$$\begin{aligned} \sin \theta \cos \phi &= \hat{\mathbf{x}} \cdot \hat{\mathbf{r}} = \cos \gamma_x, \\ \sin \theta \sin \phi &= \hat{\mathbf{y}} \cdot \hat{\mathbf{r}} = \cos \gamma_y. \end{aligned} \quad (16.4)$$

Thus, the pattern of a rectangular array is the product of the array factors of the linear arrays in the x and y directions.

In the case of a uniform planar rectangular array, $I_{m1} = I_{1n} = I_0$ for all m and n , i.e., all elements have the same excitation amplitudes. Thus,

$$AF = I_0 \sum_{m=1}^M e^{j(m-1)(kd_x \sin \theta \cos \phi + \beta_x)} \times \sum_{n=1}^N e^{j(n-1)(kd_y \sin \theta \sin \phi + \beta_y)}. \quad (16.5)$$

The normalized array factor is obtained as

$$AF_n(\theta, \phi) = \left[\frac{\sin\left(M \frac{\psi_x}{2}\right)}{M \sin\left(\frac{\psi_x}{2}\right)} \right] \cdot \left[\frac{\sin\left(N \frac{\psi_y}{2}\right)}{N \sin\left(\frac{\psi_y}{2}\right)} \right], \quad (16.6)$$

where

$$\psi_x = kd_x \sin \theta \cos \phi + \beta_x,$$

$$\psi_y = kd_y \sin \theta \sin \phi + \beta_y.$$

The major lobe (principal maximum) and grating lobes of the terms

$$S_{xM} = \frac{\sin\left(M \frac{\psi_x}{2}\right)}{M \sin\left(\frac{\psi_x}{2}\right)} \quad (16.7)$$

and

$$S_{yN} = \frac{\sin\left(N \frac{\psi_y}{2}\right)}{N \sin\left(\frac{\psi_y}{2}\right)} \quad (16.8)$$

are located at angles such that

$$kd_x \sin \theta_m \cos \phi_m + \beta_x = \pm 2m\pi, \quad m = 0, 1, \dots, \quad (16.9)$$

$$kd_y \sin \theta_n \sin \phi_n + \beta_y = \pm 2n\pi, \quad n = 0, 1, \dots \quad (16.10)$$

The principal maximum corresponds to $m = 0, n = 0$.

In general, β_x and β_y are independent from each other. But, if it is required that the main beams of S_{x_M} and S_{y_N} intersect (which is usually the case), then the common main beam is in the direction:

$$\theta = \theta_0 \text{ and } \phi = \phi_0, m = n = 0. \quad (16.11)$$

If the principal maximum is specified by (θ_0, ϕ_0) , then the progressive phase shifts β_x and β_y must satisfy

$$\beta_x = -kd_x \sin \theta_0 \cos \phi_0, \quad (16.12)$$

$$\beta_y = -kd_y \sin \theta_0 \sin \phi_0. \quad (16.13)$$

If β_x and β_y are specified, then the direction of the main beam can be found by simultaneously solving (16.12) and (16.13):

$$\tan \phi_0 = \frac{\beta_y d_x}{\beta_x d_y}, \quad (16.14)$$

$$\sin \theta_0 = \pm \sqrt{\left(\frac{\beta_x}{kd_x}\right)^2 + \left(\frac{\beta_y}{kd_y}\right)^2}. \quad (16.15)$$

The grating lobes can be located by substituting (16.12) and (16.13) in (16.9) and (16.10):

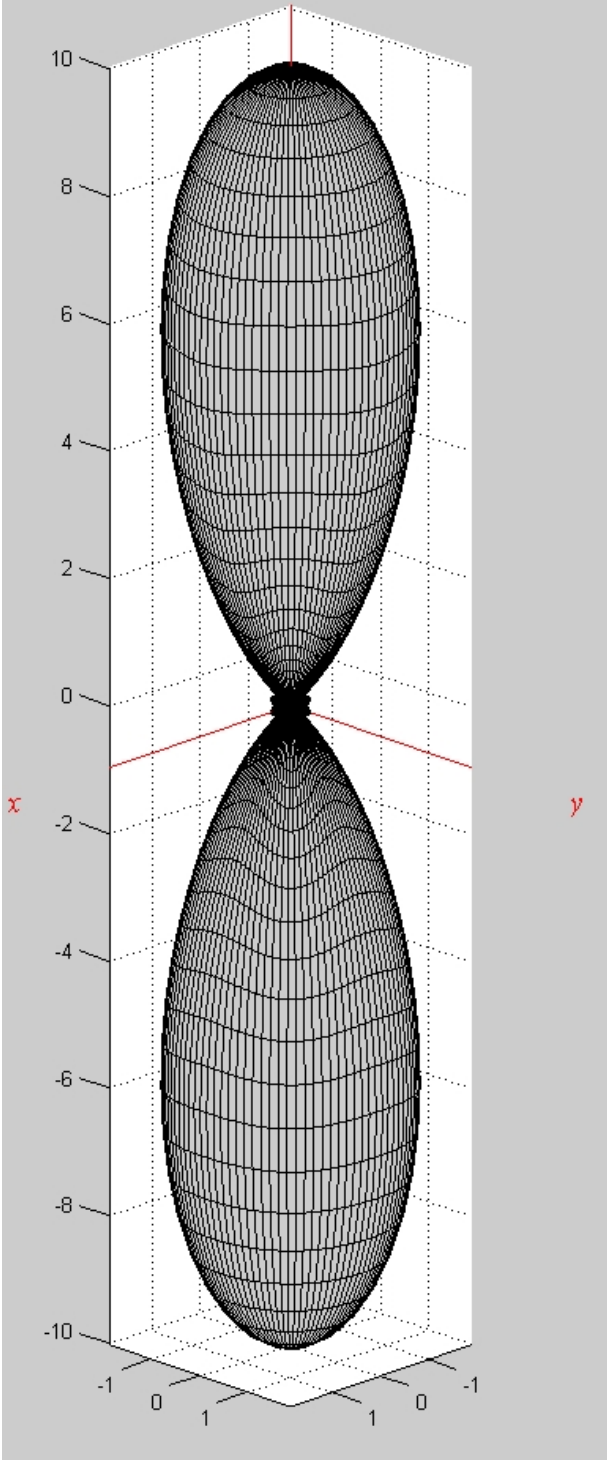
$$\tan \phi_{mn} = \frac{\sin \theta_0 \sin \phi_0 \pm n\lambda/d_y}{\sin \theta_0 \cos \phi_0 \pm m\lambda/d_x}, \quad (16.16)$$

$$\sin \theta_{mn} = \frac{\sin \theta_0 \cos \phi_0 \pm m\lambda/d_x}{\cos \phi_{mn}} = \frac{\sin \theta_0 \sin \phi_0 \pm n\lambda/d_y}{\sin \phi_{mn}}. \quad (16.17)$$

To avoid grating lobes, the spacing between the elements must be less than λ , i.e., $d_x < \lambda$ and $d_y < \lambda$. In order a true grating lobe to occur, both equations (16.16) and (16.17) must have a real solution (θ_{mn}, ϕ_{mn}) .

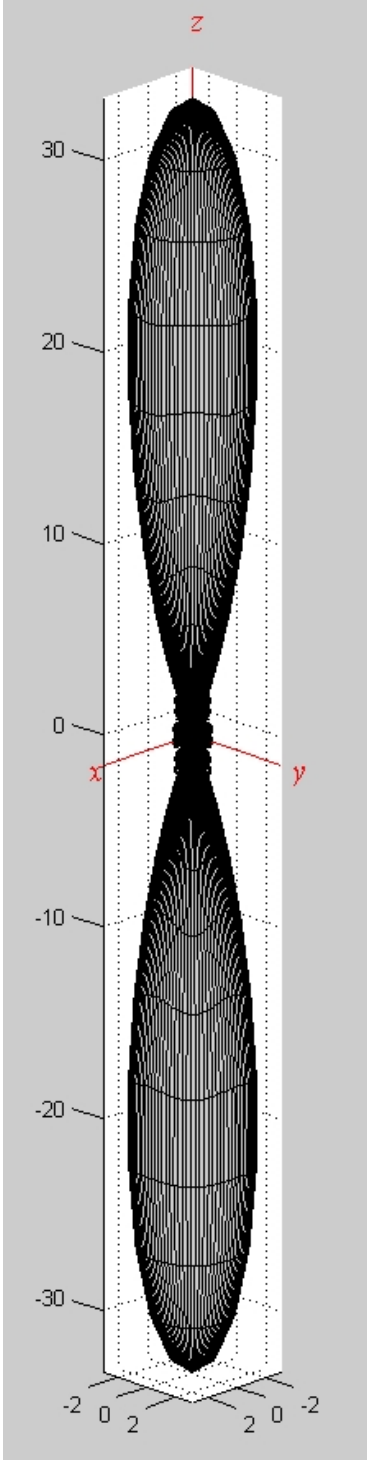
The array factors of a 5 by 5 uniform array are shown below for two spacing values: $d = \lambda/4$ and $d = \lambda/2$. Notice the considerable decrease in the beamwidth as the spacing is increased from $\lambda/4$ to $\lambda/2$.

DIRECTIVITY PATTERNS OF A 5-ELEMENT SQUARE PLANAR UNIFORM ARRAY
 WITHOUT GRATING LOBES $\beta_x = \beta_y = 0$: (a) $d = \lambda / 4$, (b) $d = \lambda / 2$



$D_0 = 10.0287$ (10.0125 dB)

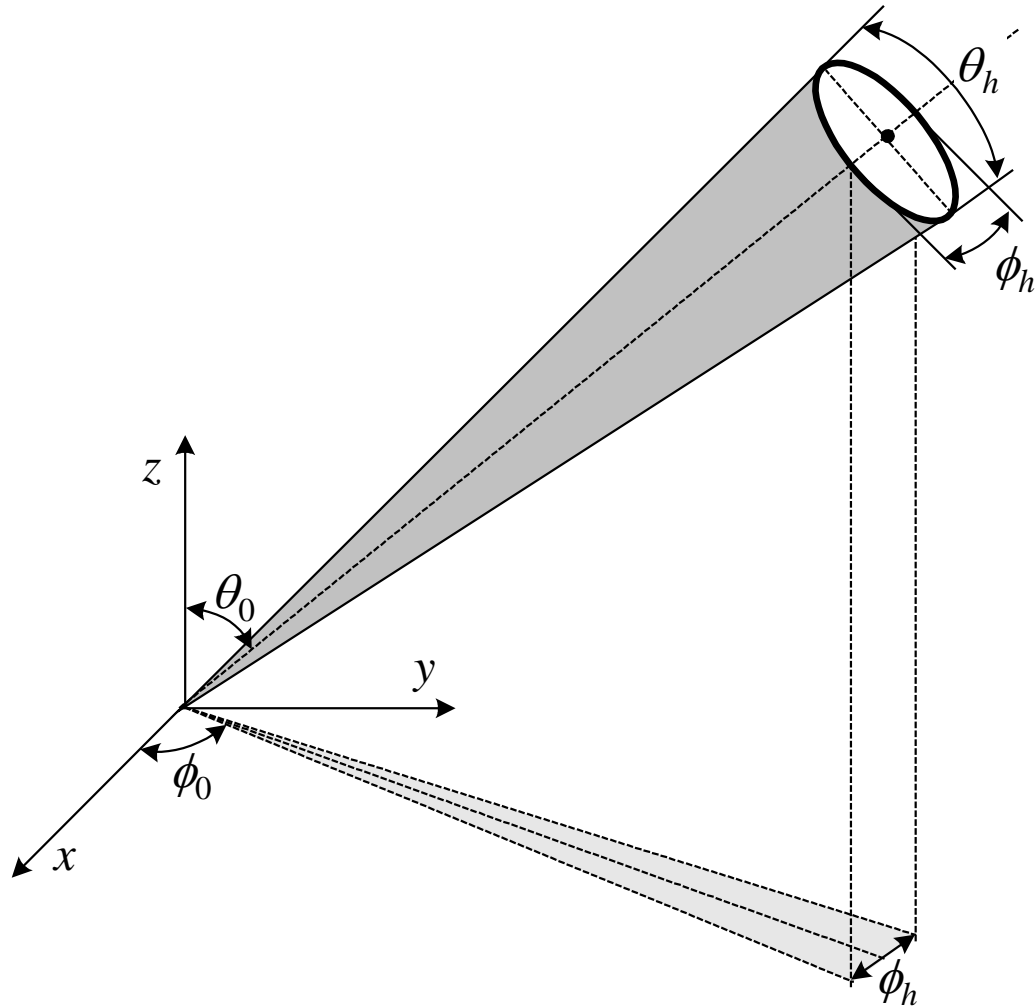
(a)



$D_0 = 33.2458$ (15.2174 dB)

(b)

B. The beamwidth of a planar array



A simple procedure, proposed by R.S. Elliot¹ is outlined below. It is based on the use of the beamwidths of the linear arrays building the planar array.

For a large array, the maximum of which is near the broad side, the *elevation plane HPBW* is approximately

$$\theta_h = \frac{1}{\cos \theta_0 \sqrt{\Delta \theta_x^{-2} \cos^2 \phi_0 + \Delta \theta_y^{-2} \sin^2 \phi_0}} \quad (16.18)$$

where

¹ "Beamwidth and directivity of large scanning arrays", *The Microwave Journal*, Jan. 1964, pp.74-82.

(θ_0, ϕ_0) specifies the main-beam direction;

$\Delta\theta_x$ is the HPBW of a linear BSA of M elements and an amplitude distribution which is the same as that of the x -axis linear arrays building the planar array;

$\Delta\theta_y$ is the HPBW of a linear BSA of N elements and amplitude distribution is the same as those of the y -axis linear arrays building the planar array.

The azimuth HPBW is the HPBW, which is in the plane orthogonal to the elevation plane and contains the maximum. It is

$$\phi_h = \sqrt{\frac{1}{\Delta\theta_x^{-2} \sin^2 \phi_0 + \Delta\theta_y^{-2} \cos^2 \phi_0}}. \quad (16.19)$$

For a square array ($M = N$) with the same amplitude distributions along the x and y axes, equations (16.18) and (16.19) reduce to

$$\theta_h = \frac{\Delta\theta_x}{\cos \theta_0} = \frac{\Delta\theta_y}{\cos \theta_0}, \quad (16.20)$$

$$\phi_h = \Delta\theta_x = \Delta\theta_y. \quad (16.21)$$

From (16.20), it is obvious that the HPBW in the elevation plane very much depends on the elevation angle θ_0 of the main beam. The HPBW in the azimuthal plane ϕ_h does not depend on the elevation angle θ_0 .

The beam solid angle of the planar array can be approximated by

$$\Omega_A = \theta_h \cdot \phi_h, \quad (16.22)$$

or

$$\Omega_A = \frac{\Delta\theta_x \Delta\theta_y}{\cos \theta_0 \sqrt{\left[\sin^2 \phi_0 + \frac{\Delta\theta_y^2}{\Delta\theta_x^2} \cos^2 \phi_0 \right] \left[\sin^2 \phi_0 + \frac{\Delta\theta_x^2}{\Delta\theta_y^2} \cos^2 \phi_0 \right]}}. \quad (16.23)$$

C. Directivity of planar rectangular array

The general expression for the calculation of the directivity of an array is

$$D_0 = 4\pi \frac{|AF(\theta_0, \phi_0)|^2}{\int_0^{2\pi} \int_0^\pi |AF(\theta, \phi)|^2 \sin \theta d\theta d\phi}. \quad (16.24)$$

For large planar arrays, which are nearly broadside, (16.24) reduces to

$$D_0 = \pi D_x D_y \cos \theta_0 \quad (16.25)$$

where

D_x is the directivity of the respective linear BSA, x -axis;

D_y is the directivity of the respective linear BSA, y -axis.

We can also use the array solid beam angle Ω_A in (16.23) to calculate the approximate directivity of a nearly broadside planar array:

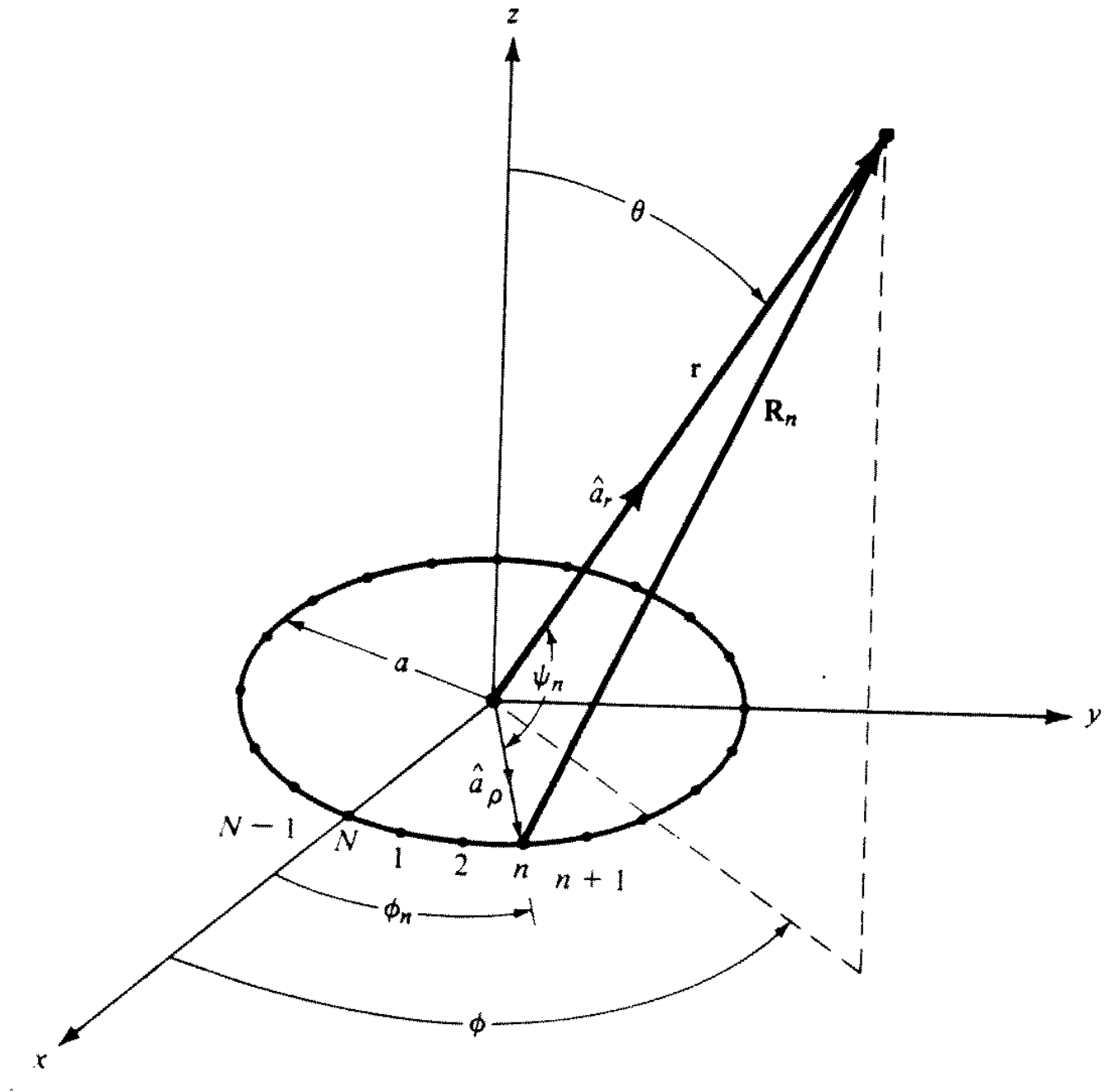
$$D_0 \approx \frac{\pi^2}{\Omega_{A[\text{Sr}]}} \approx \frac{32400}{\Omega_{A[\text{deg}^2]}}.^2 \quad (16.26)$$

Remember:

- 1) The main beam direction is controlled through the phase shifts, β_x and β_y .
- 2) The beamwidth and side-lobe levels are controlled through the amplitude distribution.

² A *steradian* relates to square degrees as $1 \text{ sr} = (180/\pi)^2 \approx 3282.80635 \text{ deg}$. Note that this formula is only approximate and the relationship between the exact values of D_0 and Ω_A is $D_0 = 4\pi/\Omega_A$.

2. Circular Array



A. Array factor of circular array

The normalized field can be written as

$$E(r, \theta, \phi) = \sum_{n=1}^N a_n \frac{e^{-jkR_n}}{R_n}, \quad (16.27)$$

where

$$R_n = \sqrt{r^2 + a^2 - 2ar \cos \psi_n}. \quad (16.28)$$

For $r \gg a$, (16.28) reduces to

$$R_n \approx r - a \cos \psi_n = r - a(\hat{\mathbf{a}}_{\rho_n} \cdot \hat{\mathbf{r}}). \quad (16.29)$$

In a rectangular coordinate system,

$$\begin{cases} \hat{\mathbf{a}}_{\rho_n} = \hat{\mathbf{x}} \cos \phi_n + \hat{\mathbf{y}} \sin \phi_n \\ \hat{\mathbf{r}} = \hat{\mathbf{x}} \sin \theta \cos \phi + \hat{\mathbf{y}} \sin \theta \sin \phi + \hat{\mathbf{z}} \cos \theta. \end{cases}$$

Therefore,

$$R_n \approx r - a \sin \theta (\cos \phi_n \cos \phi + \sin \phi_n \sin \phi), \quad (16.30)$$

or

$$R_n \approx r - a \sin \theta \cos(\phi - \phi_n). \quad (16.31)$$

For the amplitude term, the approximation

$$\frac{1}{R_n} \approx \frac{1}{r}, \text{ all } n \quad (16.32)$$

is made.

Assuming the approximations (16.31) and (16.32) are valid, the far-zone array field is reduced to

$$E(r, \theta, \phi) = \frac{e^{-jkr}}{r} \sum_{n=1}^N a_n e^{jka \sin \theta \cos(\phi - \phi_n)}, \quad (16.33)$$

where

a_n is the complex excitation coefficient (amplitude and phase);

$\phi_n = 2\pi n / N$ is the angular position of the n -th element.

In general, the excitation coefficient can be represented as

$$a_n = I_n e^{j\alpha_n}, \quad (16.34)$$

where I_n is the amplitude term, and α_n is the phase of the excitation of the n -th element relative to a chosen array element of zero phase,

$$\Rightarrow E(r, \theta, \phi) = \frac{e^{-jkr}}{r} \sum_{n=1}^N I_n e^{j[ka \sin \theta \cos(\phi - \phi_n) + \alpha_n]}. \quad (16.35)$$

The AF is then

$$\boxed{AF(\theta, \phi) = \sum_{n=1}^N I_n e^{j[ka \sin \theta \cos(\phi - \phi_n) + \alpha_n]}. \quad (16.36)}$$

Expression (16.36) represents the AF of a circular array of N equispaced elements. The maximum of the AF occurs when all the phase terms in (16.36) equal unity, or,

$$ka \sin \theta \cos(\phi - \phi_n) + \alpha_n = 2m\pi, \quad m = 0, \pm 1, \pm 2, \text{ all } n. \quad (16.37)$$

The principal maximum ($m = 0$) is defined by the direction (θ_0, ϕ_0) , for which

$$\alpha_n = -ka \sin \theta_0 \cos(\phi_0 - \phi_n), \quad n = 1, 2, \dots, N. \quad (16.38)$$

If a circular array is required to have maximum radiation along (θ_0, ϕ_0) , then the phases of its excitations have to fulfil (16.38). The AF of such an array is

$$AF(\theta, \phi) = \sum_{n=1}^N I_n e^{jka[\sin \theta \cos(\phi - \phi_n) - \sin \theta_0 \cos(\phi_0 - \phi_n)]}, \quad (16.39)$$

$$\Rightarrow AF(\theta, \phi) = \sum_{n=1}^N I_n e^{jka(\cos \psi_n - \cos \psi_{0n})}. \quad (16.40)$$

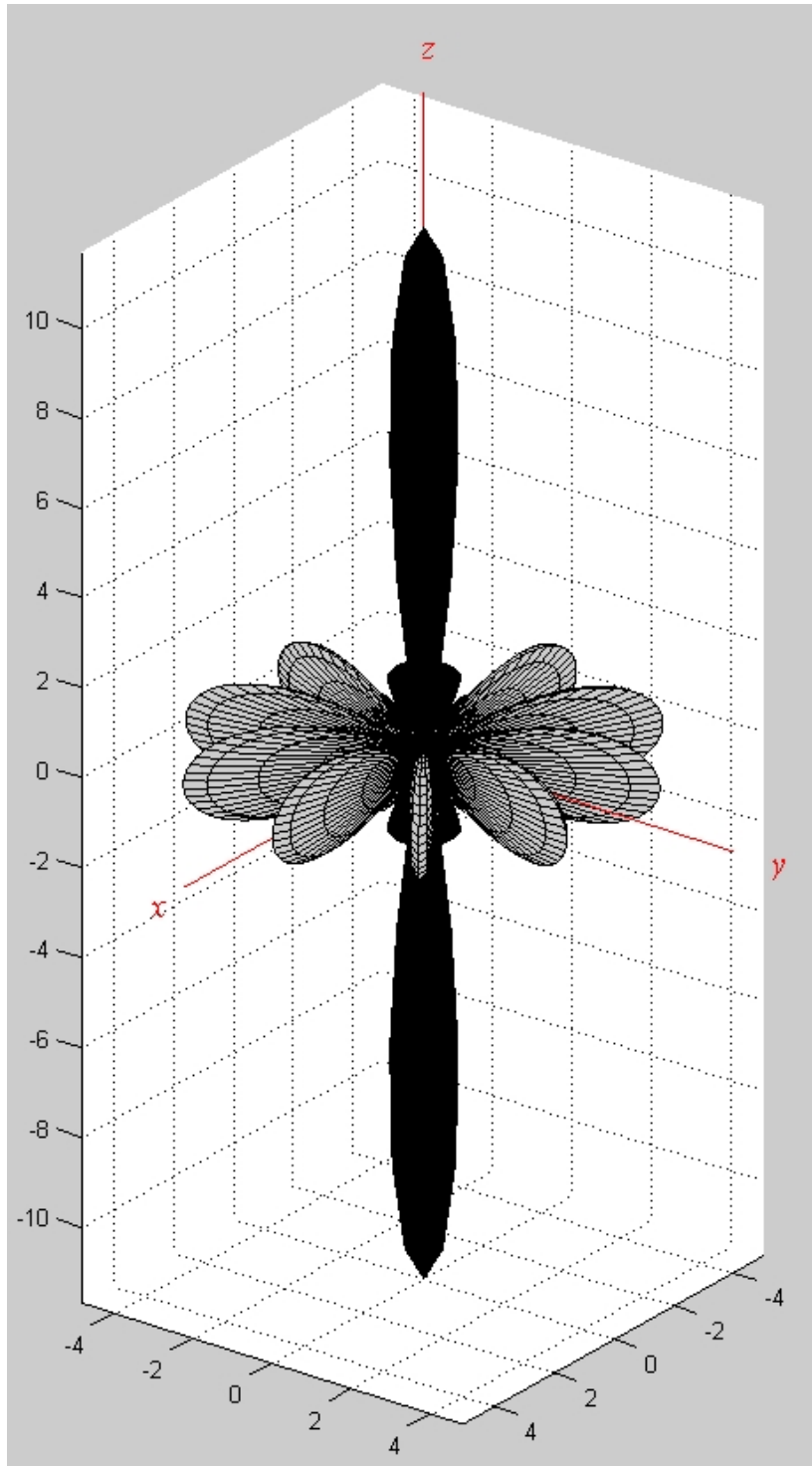
Here,

$\psi_n = \arccos[\sin \theta \cos(\phi - \phi_n)]$ is the angle between $\hat{\mathbf{r}}$ and $\hat{\mathbf{a}}_{\rho_n}$;

$\psi_{0n} = \arccos[\sin \theta_0 \cos(\phi_0 - \phi_n)]$ is the angle between $\hat{\mathbf{a}}_{\rho_n}$ and $\hat{\mathbf{r}}_{\max}$ pointing in the direction of maximum radiation.

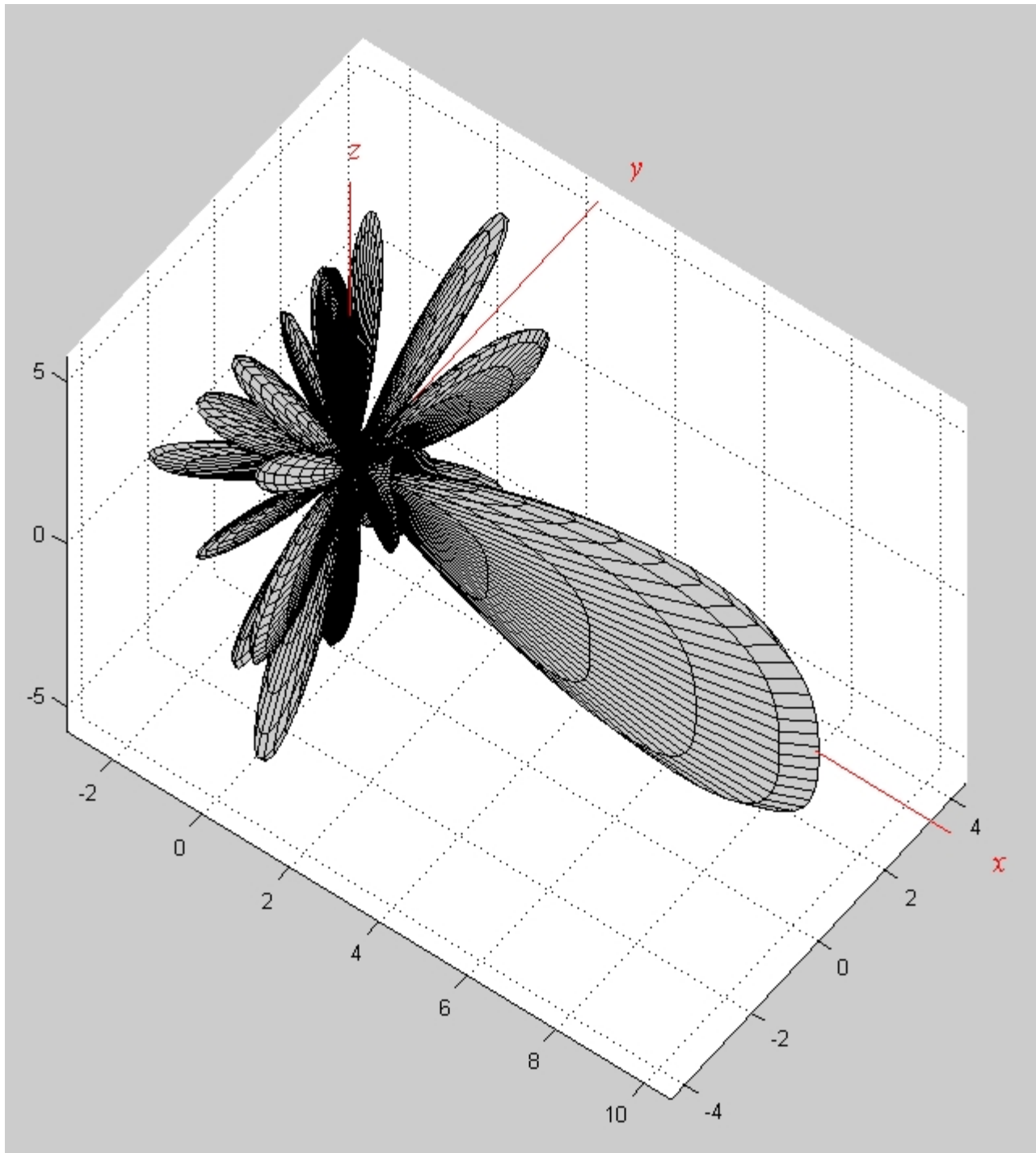
As the radius of the array a becomes large compared to λ , the directivity of the uniform circular array ($I_n = I_0$, all n) approaches the value of N .

UNIFORM CIRCULAR ARRAY 3-D PATTERN ($N = 10$, $ka = 2\pi a / \lambda = 10$):
MAXIMUM AT $\theta = 0^\circ, 180^\circ$



$$D_0 = 11.6881 \text{ (10.6775 dB)}$$

UNIFORM CIRCULAR ARRAY 3-D PATTERN ($N = 10, ka = 2\pi a / \lambda = 10$):
MAXIMUM AT $\theta = 90^\circ, \phi = 0^\circ$



$$D_0 = 10.589 \text{ (10.2485 dB)}$$

LECTURE 17: Radiation from Apertures

(The uniqueness theorem. The equivalence principle. The application of the equivalence principle to aperture problem. The uniform rectangular aperture and the radiating slit. The tapered rectangular aperture.)

1. Introduction

Aperture antennas constitute a large class of antennas, which emit EM waves through an opening (or aperture). These antennas have close analogs in acoustics, namely, the megaphone and the parabolic microphone. The pupil of the human eye is a typical aperture receiver for optical radiation. At radio and microwave frequencies, horns, waveguide apertures, reflectors and microstrip patches are examples of aperture antennas. Aperture antennas are commonly used at UHF and above where their size is reasonable. Their gain increases as $\sim f^2$. For an aperture antenna to be efficient and to have high directivity, it has to have an area $\geq \lambda^2$. Thus, these antennas are impractical at low frequencies.

To facilitate the analysis of these antennas, the equivalence principle is applied. This allows for carrying out the far-field analysis in the outer (unbounded) region only, which is external to the antenna. This requires the knowledge of the tangential field components at the aperture.

2. Uniqueness Theorem

A solution is said to be unique if it is the only one possible among a given class of solutions. The EM field in a given region $V_{[S]}$ is unique if

- all sources are given;
- either the tangential \mathbf{E}_{tan} components or the tangential \mathbf{H}_{tan} components are specified at the boundary S .¹

The uniqueness theorem follows from Poynting's theorem in its integral form:

$$\oiint_S (\mathbf{E} \times \mathbf{H}^*) \cdot d\mathbf{s} + j\omega \iiint_{V_S} (\mu |\mathbf{H}|^2 - \varepsilon |\mathbf{E}|^2) dv + \iiint_{V_S} \sigma |\mathbf{E}|^2 dv = - \iiint_{V_S} (\mathbf{E} \cdot \mathbf{J}^* + \mathbf{H}^* \cdot \mathbf{M}^i) dv. \quad (17.1)$$

¹ A more general statement of the theorem asserts that any one of the following boundary conditions at S ensure the solution's uniqueness: (1) $\mathbf{E}_{\text{tan}}|_S$, or (2) $\mathbf{H}_{\text{tan}}|_S$, or (3) $E_{\text{tan1}}|_S$ and $H_{\text{tan1}}|_S$, or (4) $E_{\text{tan2}}|_S$ and $H_{\text{tan2}}|_S$. Here, $\mathbf{E}_{\text{tan}} = \mathbf{E} - \mathbf{E} \cdot \hat{\mathbf{n}}$ is the tangential component of \mathbf{E} at the surface S while E_{tan1} and E_{tan2} are its components. The same notations hold for \mathbf{H} .

[N.K. Nikolova, "Electromagnetic boundary conditions and uniqueness revisited," *IEEE Antennas & Propagation Magazine*, vol. 46, no. 5, pp. 141–149, Oct. 2004.]

We start with the supposition that a given EM problem has two solutions (due to the same sources and the same boundary conditions): $(\mathbf{E}^a, \mathbf{H}^a)$ and $(\mathbf{E}^b, \mathbf{H}^b)$. The difference field is then formed:

$$\begin{cases} \delta\mathbf{E} = \mathbf{E}^a - \mathbf{E}^b, \\ \delta\mathbf{H} = \mathbf{H}^a - \mathbf{H}^b. \end{cases} \quad (17.2)$$

The difference field has no sources; thus, it satisfies the source-free form of (17.1):

$$\oint_S (\delta\mathbf{E} \times \delta\mathbf{H}^*) \cdot d\mathbf{s} + j\omega \iiint_{V_S} (\mu |\delta\mathbf{H}|^2 - \varepsilon |\delta\mathbf{E}|^2) dv + \iiint_{V_S} \sigma |\delta\mathbf{E}|^2 dv = 0. \quad (17.3)$$

Since both fields satisfy the same boundary conditions, then $\delta\mathbf{E}_{\text{tan}} = 0$ or $\delta\mathbf{H}_{\text{tan}} = 0$ over S , which makes the surface integral in (17.3) zero. This results in

$$\underbrace{j\omega \iiint_{V_S} (\mu |\delta\mathbf{H}|^2 - \varepsilon |\delta\mathbf{E}|^2) dv}_{\text{imaginary}} + \underbrace{\iiint_{V_S} \sigma |\delta\mathbf{E}|^2 dv}_{\text{real}} = 0, \quad (17.4)$$

which is true only if

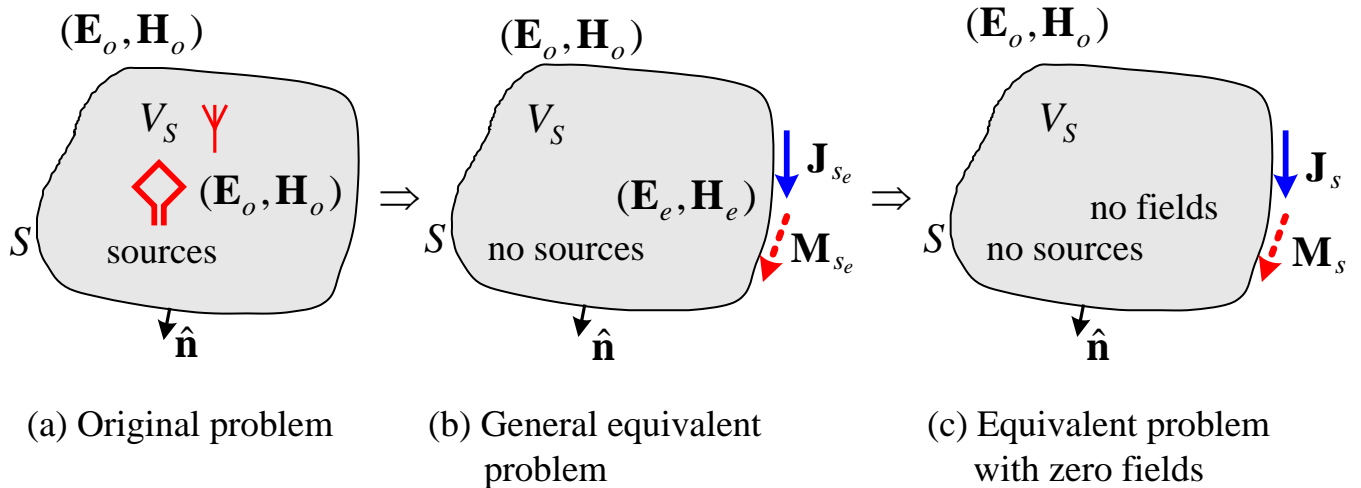
$$\begin{cases} \omega \iiint_{V_S} (\mu |\delta\mathbf{H}|^2 - \varepsilon |\delta\mathbf{E}|^2) dv = 0, \\ \iiint_{V_S} \sigma |\delta\mathbf{E}|^2 dv = 0. \end{cases} \quad (17.5)$$

If we assume some dissipation ($\sigma > 0$), however slight, equations (17.5) are satisfied only if $\delta\mathbf{E} = \delta\mathbf{H} = 0$ everywhere in the volume V_S . This implies the uniqueness of the solution. If $\sigma = 0$ (a common approximation), multiple solutions $(\delta\mathbf{E}, \delta\mathbf{H})$ may exist in the form of resonant modes. However, these resonant modes can be derived using eigenvalue analysis and they are not considered as the particular solution for the given sources. The particular unique solution for the loss-free case can be obtained from a problem where σ is assumed nonzero and then the limit is found as $\sigma \rightarrow 0$.

3. Equivalence Principles

The equivalence principle follows from the uniqueness theorem. It allows for the simplification of certain EM problems. As long as a problem is re-formulated

so that it preserves the boundary conditions at S , it is going to produce the only one possible solution for the region V_S bounded by S . Such a re-formulated problem is referred to as an *equivalent problem*.



For the equivalent problem in (b),

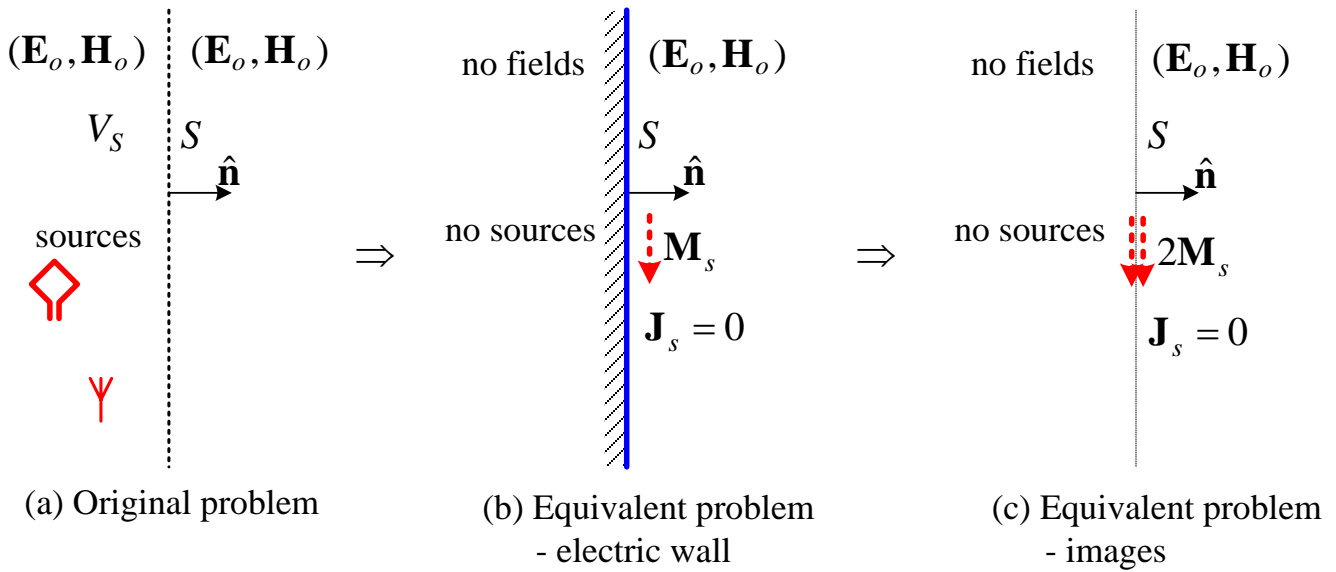
$$\begin{cases} \mathbf{J}_{s_e} = \hat{\mathbf{n}} \times (\mathbf{H}_o - \mathbf{H}_e), \\ \mathbf{M}_{s_e} = (\mathbf{E}_o - \mathbf{E}_e) \times \hat{\mathbf{n}}. \end{cases} \quad (17.6)$$

For the equivalent problem in (c),

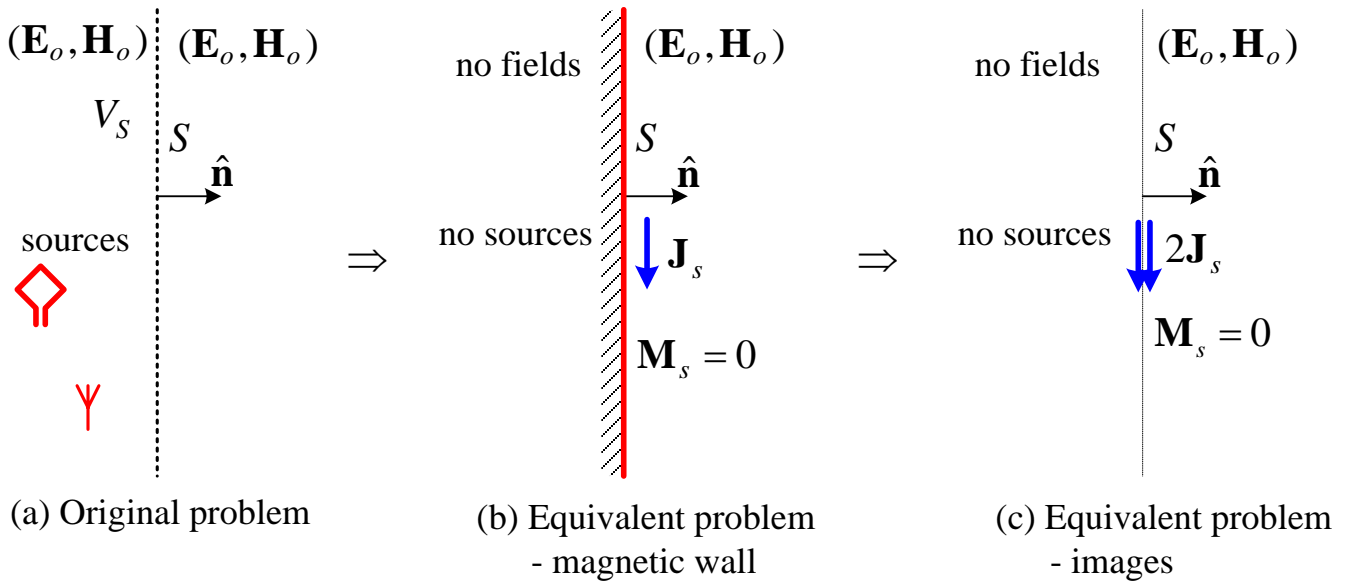
$$\begin{cases} \mathbf{J}_s = \hat{\mathbf{n}} \times \mathbf{H}_o, \\ \mathbf{M}_s = \mathbf{E}_o \times \hat{\mathbf{n}}. \end{cases} \quad (17.7)$$

The zero-field formulation is often referred to as *Love's equivalence principle*. We can apply Love's equivalence principle in three different ways.

- (a) We can assume that the boundary S is a perfect conductor. As per image theory, in an equivalent open problem, this eliminates the surface electric currents, i.e., $\mathbf{J}_s = 0$, and leaves just surface magnetic currents of double strength $2\mathbf{M}_s$. Such an equivalent problem is illustrated below.



(b) We can assume that the boundary S is a perfect magnetic conductor. As per image theory, in an equivalent open problem, this eliminates the surface magnetic currents, i.e., $\mathbf{M}_s = 0$, and leaves just surface electric currents of double strength $2\mathbf{J}_s$. This approach is illustrated below.



(c) Make no assumptions about the materials inside S , and define both \mathbf{J}_s and \mathbf{M}_s currents, which radiate in free space (no fictitious conductors behind them). It can be shown that these equivalent currents create zero fields inside V_S . [Ewald-Oseen extinction theorem: A. Ishimaru, *Electromagnetic Wave Propagation, Radiation, and Scattering*, Prentice Hall, 1991, p. 173]

The first two approaches are not very accurate in the general case of a curved boundary surface S because the image theory can be applied to curved surfaces only if the curvature radius is large compared to the wavelength. However, in the case of flat infinite planes (walls), the image theory holds exactly and all three approaches should produce the same external field according to the uniqueness theorem.

The above approaches are used to compute fields in half-space as excited by apertures. The field behind S is assumed known and are used to define the equivalent surface currents. The open-region far-zone solutions for the vector potentials \mathbf{A} (resulting from \mathbf{J}_s) and \mathbf{F} (resulting from \mathbf{M}_s) are

$$\mathbf{A}(P) = \mu \frac{e^{-j\beta r}}{4\pi r} \iint_S \mathbf{J}_s(\mathbf{r}') e^{j\beta \hat{\mathbf{r}} \cdot \mathbf{r}'} ds', \quad (17.8)$$

$$\mathbf{F}(P) = \varepsilon \frac{e^{-j\beta r}}{4\pi r} \iint_S \mathbf{M}_s(\mathbf{r}') e^{j\beta \hat{\mathbf{r}} \cdot \mathbf{r}'} ds'. \quad (17.9)$$

Here, $\hat{\mathbf{r}}$ denotes the unit vector pointing from the origin of the coordinate system to the point of observation P . The integration point Q is specified through the position vector \mathbf{r}' . In the far zone, it is assumed that the field propagates radially away from the antenna. It is convenient to introduce the *propagation vector* or *wave vector*,

$$\boldsymbol{\beta} = \beta \hat{\mathbf{r}}, \quad (17.10)$$

which characterizes both the phase constant and the direction of propagation of the wave. The vector potentials can then be written as

$$\mathbf{A}(P) = \mu \frac{e^{-j\beta r}}{4\pi r} \iint_S \mathbf{J}_s(\mathbf{r}') e^{j\boldsymbol{\beta} \cdot \mathbf{r}'} ds', \quad (17.11)$$

$$\mathbf{F}(P) = \varepsilon \frac{e^{-j\beta r}}{4\pi r} \iint_S \mathbf{M}_s(\mathbf{r}') e^{j\boldsymbol{\beta} \cdot \mathbf{r}'} ds'. \quad (17.12)$$

The relations between the far-zone field vectors and the vector potentials are

$$\mathbf{E}_A^{far} = -j\omega(A_\theta \hat{\boldsymbol{\theta}} + A_\varphi \hat{\boldsymbol{\phi}}), \quad (17.13)$$

$$\mathbf{H}_F^{far} = -j\omega(F_\theta \hat{\boldsymbol{\theta}} + F_\varphi \hat{\boldsymbol{\phi}}). \quad (17.14)$$

Since

$$\mathbf{E}_F^{far} = \eta \mathbf{H}_F^{far} \times \hat{\mathbf{r}}, \quad (17.15)$$

the total far-zone electric field (due to both \mathbf{A} and \mathbf{F}) is found as

$$\mathbf{E}^{far} = \mathbf{E}_A^{far} + \mathbf{E}_F^{far} = -j\omega \left[(A_\theta + \eta F_\phi) \hat{\boldsymbol{\theta}} + (A_\phi - \eta F_\theta) \hat{\boldsymbol{\phi}} \right]. \quad (17.16)$$

Equation (17.16) involves both vector potentials as arising from both types of surface currents. Computations are reduced in half if image theory is used in conjunction with an electric or magnetic wall assumption.

4. Application of the Equivalence Principle to Aperture Problems

The equivalence principle is widely used in the analysis of aperture antennas. To calculate exactly the far field, the exact field distribution at the (infinite) aperture is needed. In the case of exact knowledge of the aperture field distribution, all three approaches given above produce the same results. However, the aperture field distribution is usually not known exactly and approximations are used. Then, the three equivalence-principle approaches produce slightly different results, the consistency being dependent on how accurate our knowledge about the aperture field is. Usually, it is assumed that the field is to be determined in half-space, leaving the feed and the antenna behind an infinite wall S . The aperture of the antenna S_A is this portion of S where we have an approximate knowledge of the field distribution based on the type of the feed line or the incident wave illuminating the aperture. This is the so-called *physical optics* approximation, which is more accurate than the *geometrical optics* approach of ray tracing. The larger the aperture (as compared to the wavelength), the more accurate the approximation based on the incident wave.

Let us assume that the field at the aperture S_A is known: $\mathbf{E}_a, \mathbf{H}_a$, and it is zero everywhere on S except at S_A . The equivalent current densities are:

$$\begin{cases} \mathbf{J}_s = \hat{\mathbf{n}} \times \mathbf{H}_a, \\ \mathbf{M}_s = \mathbf{E}_a \times \hat{\mathbf{n}}. \end{cases} \quad (17.17)$$

The substitution of (17.17) into (17.11) and (17.12) produces

$$\mathbf{A}(P) = \mu \frac{e^{-j\beta r}}{4\pi r} \iint_{S_A} \hat{\mathbf{n}} \times \mathbf{H}_a e^{j\mathbf{B} \cdot \mathbf{r}'} ds', \quad (17.18)$$

$$\mathbf{F}(P) = -\varepsilon \frac{e^{-j\beta r}}{4\pi r} \iint_{S_A} \hat{\mathbf{n}} \times \mathbf{E}_a e^{j\beta \cdot \mathbf{r}'} ds'. \quad (17.19)$$

We can work with the general vector expression for the far field \mathbf{E} [see (17.16)] written as

$$\mathbf{E}^{far} = -j\omega\mathbf{A} - j\omega\eta\mathbf{F} \times \hat{\mathbf{r}}, \quad (17.20)$$

where *the longitudinal A_r component is to be neglected*. Substituting (17.18) and (17.19) into (17.20) yields

$$\mathbf{E}^{far} = -j\beta \frac{e^{-j\beta r}}{4\pi r} \hat{\mathbf{r}} \times \iint_{S_A} [\hat{\mathbf{n}} \times \mathbf{E}_a - \eta \hat{\mathbf{r}} \times (\hat{\mathbf{n}} \times \mathbf{H}_a)] e^{j\beta \cdot \mathbf{r}'} ds'. \quad (17.21)$$

This is the full vector form of the radiated field resulting from the aperture field, and it is referred to as the *vector diffraction integral* (or *vector Kirchhoff integral*).

We now consider a practical case of a flat aperture lying in the xy plane with $\hat{\mathbf{n}} \equiv \hat{\mathbf{z}}$. For brevity, the radiation integrals in (17.18) and (17.19) are denoted as

$$\mathbf{I}^H = I_x^H \hat{\mathbf{x}} + I_y^H \hat{\mathbf{y}} = \iint_S \mathbf{H}_a e^{j\beta \cdot \mathbf{r}'} ds', \quad (17.22)$$

$$\mathbf{I}^E = I_x^E \hat{\mathbf{x}} + I_y^E \hat{\mathbf{y}} = \iint_S \mathbf{E}_a e^{j\beta \cdot \mathbf{r}'} ds'. \quad (17.23)$$

Then,

$$\mathbf{A} = \mu \frac{e^{-j\beta r}}{4\pi r} (-I_y^H \hat{\mathbf{x}} + I_x^H \hat{\mathbf{y}}), \quad (17.24)$$

$$\mathbf{F} = -\varepsilon \frac{e^{-j\beta r}}{4\pi r} (-I_y^E \hat{\mathbf{x}} + I_x^E \hat{\mathbf{y}}). \quad (17.25)$$

The integrals in the above expressions can be explicitly written for the case $\hat{\mathbf{n}} \equiv \hat{\mathbf{z}}$ bearing in mind that the source-point position is $\mathbf{r}' = x'\hat{\mathbf{x}} + y'\hat{\mathbf{y}}$:

$$I_x^E(\theta, \varphi) = \iint_{S_A} E_{a_x}(x', y') e^{j\beta(x' \sin \theta \cos \varphi + y' \sin \theta \sin \varphi)} dx' dy', \quad (17.26)$$

$$I_y^E(\theta, \varphi) = \iint_{S_A} E_{a_y}(x', y') e^{j\beta(x' \sin \theta \cos \varphi + y' \sin \theta \sin \varphi)} dx' dy', \quad (17.27)$$

$$I_x^H(\theta, \varphi) = \iint_{S_A} H_{a_x}(x', y') e^{j\beta(x' \sin \theta \cos \varphi + y' \sin \theta \sin \varphi)} dx' dy', \quad (17.28)$$

$$I_y^H(\theta, \varphi) = \iint_{S_A} H_{a_y}(x', y') e^{j\beta(x' \sin \theta \cos \varphi + y' \sin \theta \sin \varphi)} dx' dy'. \quad (17.29)$$

Note that the above integrals can be viewed as 2-D Fourier transforms of the aperture field components where x transforms into $\beta_x = -\beta \sin \theta \cos \varphi$ and y transforms into $\beta_y = -\beta \sin \theta \sin \varphi$.

The transverse components of the magnetic vector potential in spherical terms are obtained from (17.24) as

$$A_\theta = \mu \frac{e^{-j\beta r}}{4\pi r} (-I_y^H \cdot \cos \theta \cos \varphi + I_x^H \cdot \cos \theta \sin \varphi), \quad (17.30)$$

$$A_\varphi = \mu \frac{e^{-j\beta r}}{4\pi r} (I_y^H \cdot \sin \varphi + I_x^H \cdot \cos \varphi), \quad (17.31)$$

which can also be written in the vector form:

$$\mathbf{A}_\perp = \mu \frac{e^{-j\beta r}}{4\pi r} \left[\hat{\boldsymbol{\theta}} \cos \theta (I_x^H \sin \varphi - I_y^H \cos \varphi) + \hat{\boldsymbol{\phi}} (I_x^H \cos \varphi + I_y^H \sin \varphi) \right]. \quad (17.32)$$

Analogously,

$$\mathbf{F}_\perp = -\varepsilon \frac{e^{-j\beta r}}{4\pi r} \left[\hat{\boldsymbol{\theta}} \cos \theta (I_x^E \sin \varphi - I_y^E \cos \varphi) + \hat{\boldsymbol{\phi}} (I_x^E \cos \varphi + I_y^E \sin \varphi) \right]. \quad (17.33)$$

By substituting the above expressions in (17.16), we obtain the far-zone \mathbf{E} field:

$$E_\theta = j\beta \frac{e^{-j\beta r}}{4\pi r} \left[I_x^E \cos \varphi + I_y^E \sin \varphi + \eta \cos \theta \cdot (I_y^H \cos \varphi - I_x^H \sin \varphi) \right], \quad (17.34)$$

$$E_\varphi = j\beta \frac{e^{-j\beta r}}{4\pi r} \left[-\eta (I_x^H \cos \varphi + I_y^H \sin \varphi) + \cos \theta \cdot (I_y^E \cos \varphi - I_x^E \sin \varphi) \right]. \quad (17.35)$$

For apertures mounted on a conducting plane (e.g., slot antennas), the preferred equivalent model is the one using an electric wall with doubled magnetic current density

$$\mathbf{M}_s = 2 \cdot (\mathbf{E}_a \times \hat{\mathbf{n}}), \quad (17.36)$$

radiating in open space. The solution, of course, is valid only for $z \geq 0$. In this case, $\mathbf{I}^H = 0$ and the field is given by

$$E_{\theta}(\theta, \varphi) = j\beta \frac{e^{-j\beta r}}{4\pi r} (I_x^E \cos \varphi + I_y^E \sin \varphi), \quad (17.37)$$

$$E_{\varphi}(\theta, \varphi) = j\beta \frac{e^{-j\beta r}}{4\pi r} \cos \theta (I_y^E \cos \varphi - I_x^E \sin \varphi). \quad (17.38)$$

For apertures illuminated from open space (e.g., reflector antennas), the dual current formulation is used. Then, the usual assumption is that the aperture field resembles that of a locally-plane wave, i.e.,

$$\mathbf{H}_a = \hat{\mathbf{z}} \times \mathbf{E}_a / \eta. \quad (17.39)$$

This implies that

$$\mathbf{I}^H = \frac{1}{\eta} \hat{\mathbf{z}} \times \mathbf{I}^E \text{ or } I_x^H = -\frac{I_y^E}{\eta}, I_y^H = \frac{I_x^E}{\eta}. \quad (17.40)$$

This assumption is valid for apertures that are at least a couple of wavelengths in extent. Then, (17.34)-(17.35) reduce to

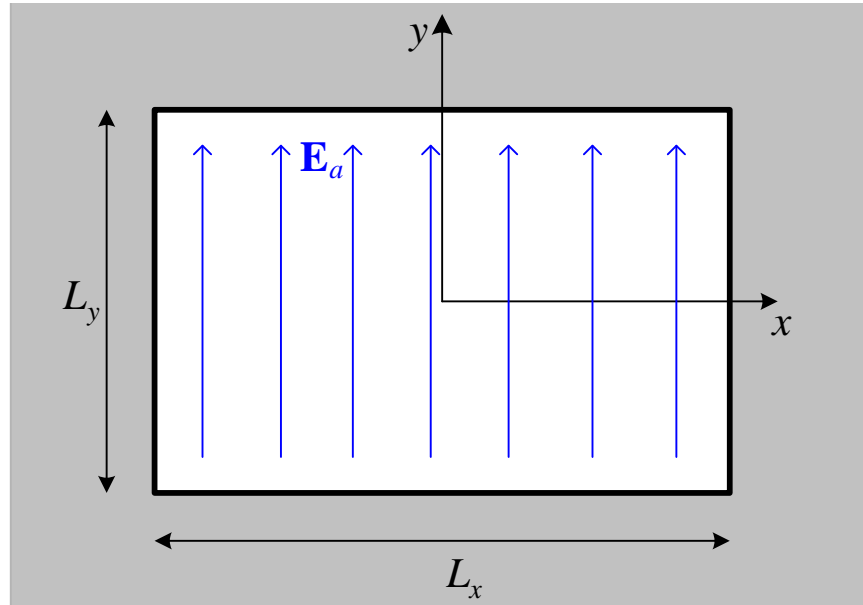
$$E_{\theta}(\theta, \varphi) = j\beta \frac{e^{-j\beta r}}{4\pi r} \frac{(1 + \cos \theta)}{2} (I_x^E \cos \varphi + I_y^E \sin \varphi), \quad (17.41)$$

$$E_{\varphi}(\theta, \varphi) = j\beta \frac{e^{-j\beta r}}{4\pi r} \frac{(1 + \cos \theta)}{2} (I_y^E \cos \varphi - I_x^E \sin \varphi). \quad (17.42)$$

Compare (17.41)-(17.42) to (17.37)-(17.38). The terms in the brackets are identical. If the aperture has high gain, the factors containing $\cos \theta$ are not going to affect the pattern significantly and the two sets of formulas are going to be nearly equivalent.

5. The Uniform Rectangular Aperture on an Infinite Ground Plane

A rectangular aperture is defined in the xy plane as shown below.



If the field is uniform in amplitude and phase across the aperture, it is referred to as a *uniform rectangular aperture*. Let us assume that the aperture field is y -polarized:

$$\mathbf{E}_a = E_0 \hat{\mathbf{y}}, \text{ for } |x| \leq \frac{L_x}{2} \text{ and } |y| \leq \frac{L_y}{2}, \quad (17.43)$$

$$\mathbf{E}_a = 0, \text{ elsewhere .}$$

According to the equivalence principle, we assume an electric wall at $z = 0$, where the equivalent magnetic current density is given by $\mathbf{M}_{s,e} = \mathbf{E}_0 \times \hat{\mathbf{n}}$. Applying image theory, we double the equivalent source radiating in open space:

$$\mathbf{M}_s = 2\mathbf{M}_{s,e} = 2E_0 \hat{\mathbf{y}} \times \hat{\mathbf{z}} = 2E_0 \hat{\mathbf{x}}. \quad (17.44)$$

The only non-zero radiation integral is [see (17.27)]

$$I_y^E(\theta, \varphi) = 2E_0 \int_{-L_x/2}^{L_x/2} e^{j\beta x' \sin \theta \cos \varphi} dx' \cdot \int_{-L_y/2}^{L_y/2} e^{j\beta y' \sin \theta \sin \varphi} dy', \quad (17.45)$$

the solution of which yields

$$I_y^E(\theta, \varphi) = 2E_0L_xL_y \frac{\sin\left(\frac{\beta L_x}{2} \sin\theta \cos\varphi\right) \sin\left(\frac{\beta L_y}{2} \sin\theta \sin\varphi\right)}{\left(\frac{\beta L_x}{2} \sin\theta \cos\varphi\right) \left(\frac{\beta L_y}{2} \sin\theta \sin\varphi\right)}. \quad (17.46)$$

To shorten the notations, let us introduce the pattern variables:

$$\begin{cases} u(\theta, \varphi) = 0.5\beta L_x \sin\theta \cos\varphi, \\ v(\theta, \varphi) = 0.5\beta L_y \sin\theta \sin\varphi. \end{cases} \quad (17.47)$$

The complete radiation field is found by substituting (17.46) in (17.37)-(17.38):

$$\begin{cases} E_\theta(\theta, \varphi) = j\beta \frac{e^{-j\beta r}}{2\pi r} E_0L_xL_y \sin\varphi \left(\frac{\sin u}{u}\right) \left(\frac{\sin v}{v}\right), \\ E_\varphi(\theta, \varphi) = j\beta \frac{e^{-j\beta r}}{2\pi r} E_0L_xL_y \cos\theta \cos\varphi \left(\frac{\sin u}{u}\right) \left(\frac{\sin v}{v}\right). \end{cases} \quad (17.48)$$

The total-field amplitude pattern is, therefore,

$$\begin{aligned} |\bar{E}(\theta, \varphi)| = F(\theta, \varphi) &= \sqrt{\sin^2\varphi + \cos^2\theta \cos^2\varphi} \cdot \left(\frac{\sin u}{u}\right) \left(\frac{\sin v}{v}\right) = \\ &= \sqrt{1 - \sin^2\theta \cos^2\varphi} \cdot \left(\frac{\sin u}{u}\right) \left(\frac{\sin v}{v}\right). \end{aligned} \quad (17.49)$$

The principal plane patterns are:

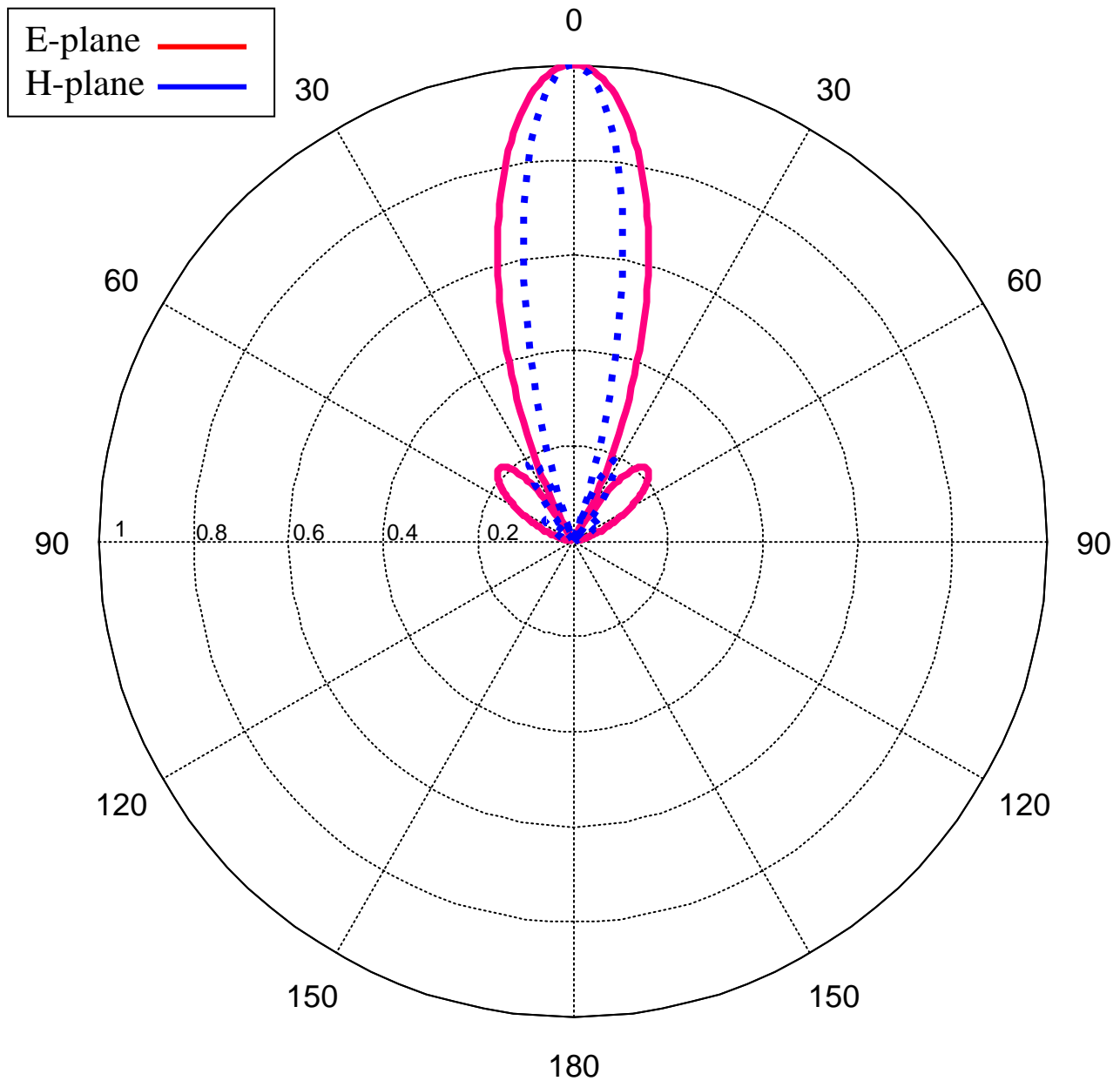
E-plane pattern ($\varphi = \pi/2$)

$$\bar{E}_\theta(\theta) = \frac{\sin(0.5\beta L_y \sin\theta)}{(0.5\beta L_y \sin\theta)} \quad (17.50)$$

H-plane pattern ($\varphi = 0$)

$$\bar{E}_\varphi(\theta) = \cos\theta \cdot \frac{\sin(0.5\beta L_x \sin\theta)}{(0.5\beta L_x \sin\theta)} \quad (17.51)$$

PRINCIPLE PATTERNS FOR APERTURE OF SIZE: $L_x = 3\lambda$, $L_y = 2\lambda$

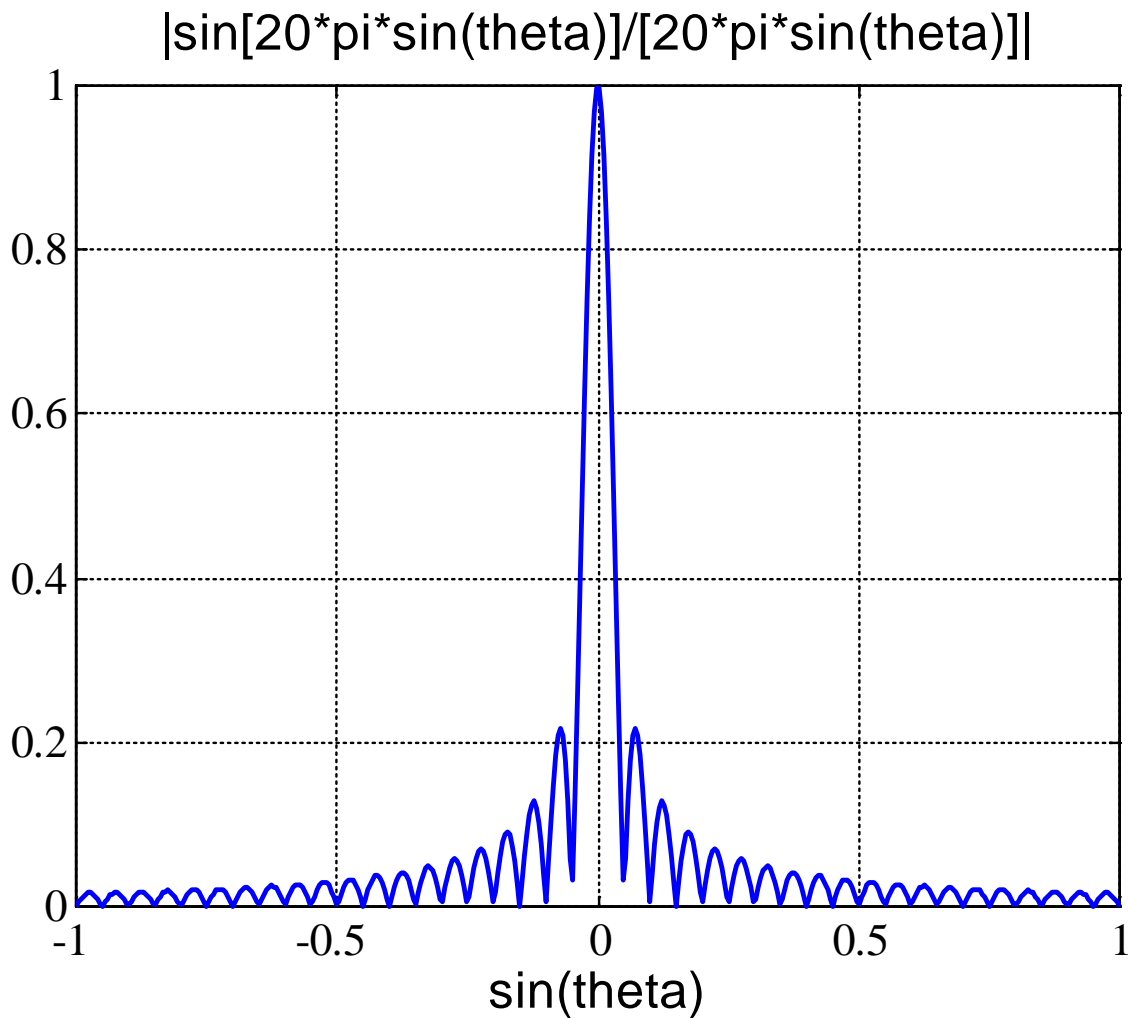


For electrically large apertures, the main beam is narrow and the $(1 - \sin^2 \theta \cos^2 \varphi)^{1/2}$ in (17.49) is negligible, i.e., it is roughly equal to 1 for all observation angles within the main beam. That is why, in the theory of large apertures and arrays, it is assumed that the amplitude pattern is

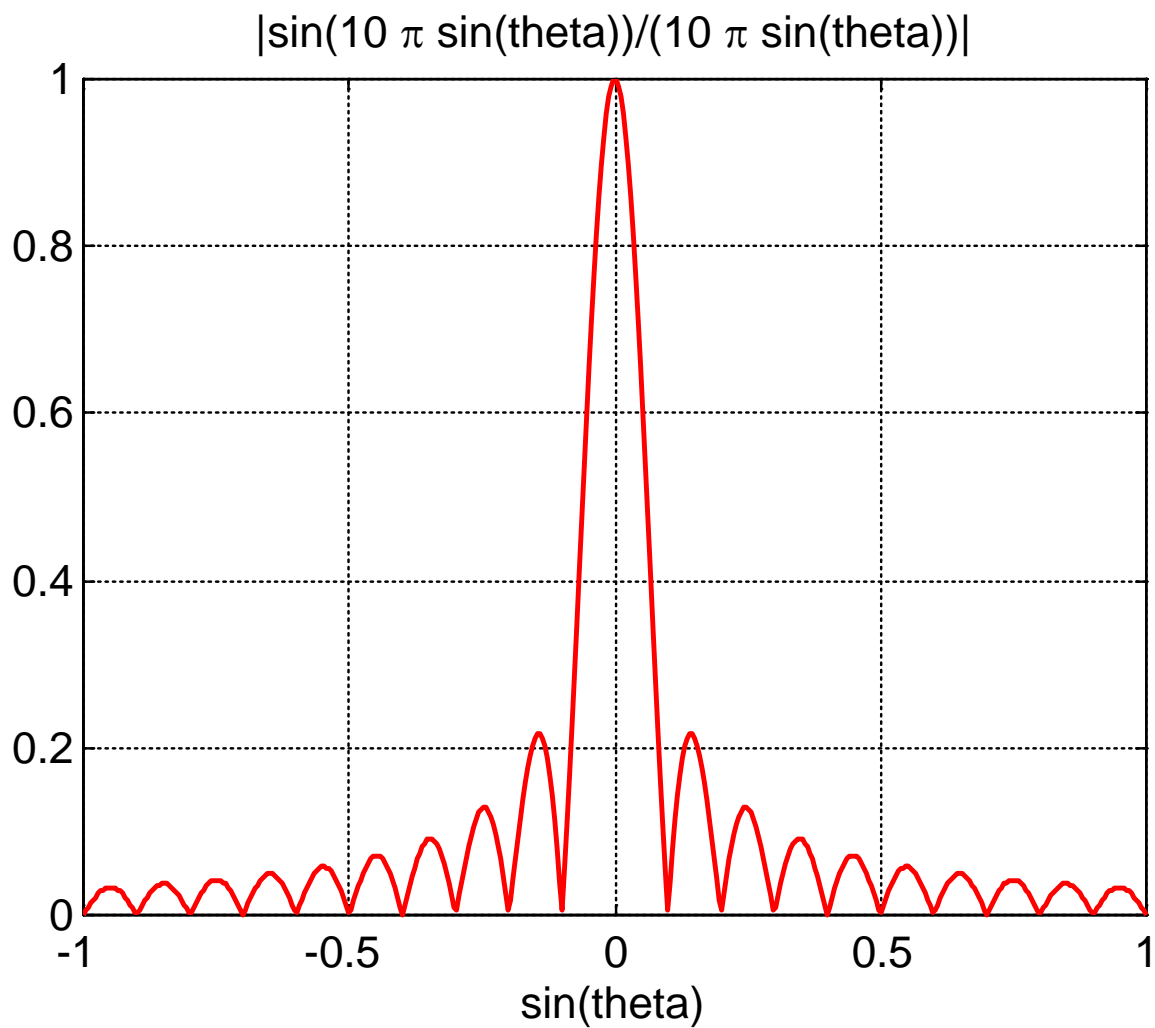
$$f(u, v) \approx \left| \frac{\sin u}{u} \cdot \frac{\sin v}{v} \right|, \quad (17.52)$$

where $u = 0.5\beta L_x \sin \theta \cos \varphi$ and $v = 0.5\beta L_y \sin \theta \sin \varphi$.

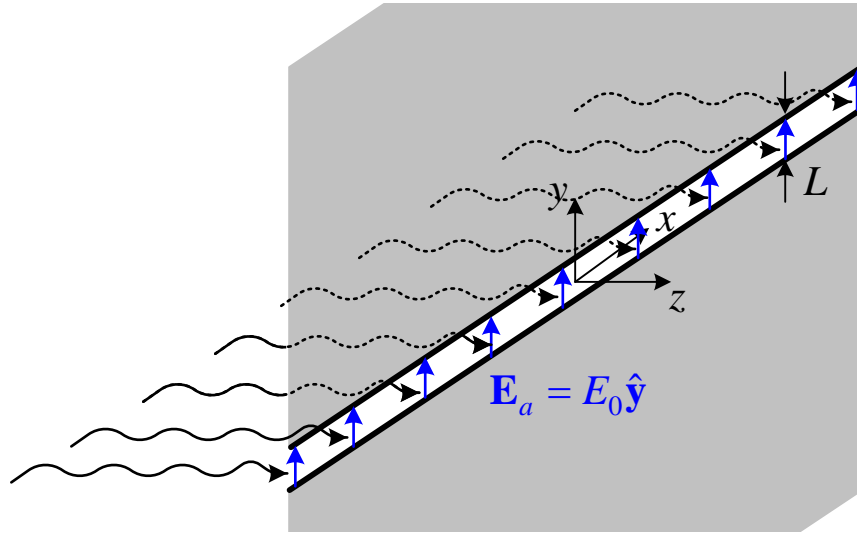
Below is a view of the $|\sin u / u|$ function for $L_x = 20\lambda$ and $\varphi = 0^\circ$ (H -plane pattern):



Here is a view of the $|\sin v/v|$ function for $L_y = 10\lambda$ and $\varphi = 90^\circ$ (E -plane pattern):



Point for Discussion: The field of a narrow slot (a slit) ($L_y \ll \lambda$).



The radiation integral for the case of a slit is a particular case of (17.46):

$$(I_y^E)_{\text{slit}} = 2E_0 L_x L_y \frac{\sin(0.5\beta L_x \sin \theta \cos \varphi)}{(0.5\beta L_x \sin \theta \cos \varphi)} \cdot \lim_{L_y \rightarrow 0} \left[\frac{\sin(0.5\beta L_y \sin \theta \sin \varphi)}{(0.5\beta L_y \sin \theta \sin \varphi)} \right], \quad (17.53)$$

which leads to

$$(I_y^E)_{\text{slit}} = 2E_0 L_x L_y \frac{\sin(0.5\beta L_x \sin \theta \cos \varphi)}{(0.5\beta L_x \sin \theta \cos \varphi)}. \quad (17.54)$$

The total field pattern of the slit is then

$$F(\theta, \varphi) = \sqrt{1 - \sin^2 \theta \cos^2 \varphi} \cdot \left[\frac{\sin(0.5\beta L_x \sin \theta \cos \varphi)}{0.5\beta L_x \sin \theta \cos \varphi} \right] \quad (17.55)$$

The principal plane patterns are:

E-plane pattern ($\varphi = \pi / 2$) - omnidirectional!

$$\bar{E}_\theta = 1 \quad (17.56)$$

H-plane pattern ($\varphi = 0$)

$$\bar{E}_\varphi = \cos \theta \cdot \frac{\sin(0.5\beta L_x \sin \theta)}{(0.5\beta L_x \sin \theta)} \quad (17.57)$$

Beamwidths

(a) First-null beamwidth (FNBW)

We need the locations of the first nulls in the pattern in order to calculate the FNBW. The nulls of the E -plane pattern are determined from (17.50) as

$$\frac{\beta L_y}{2} \sin \theta_{|\theta=\theta_n} = n\pi, \quad n = 1, 2, \dots, \quad (17.58)$$

$$\Rightarrow \theta_n = \arcsin\left(\frac{n\lambda}{L_y}\right), \text{ rad.} \quad (17.59)$$

The first null occurs at $n = 1$.

$$\Rightarrow FNBW_E = 2\theta_n = 2 \arcsin\left(\frac{\lambda}{L_y}\right), \text{ rad.} \quad (17.60)$$

In a similar fashion, $FNBW_H$ is determined to be

$$FNBW_H = 2 \arcsin\left(\frac{\lambda}{L_x}\right), \text{ rad.} \quad (17.61)$$

It is apparent that larger aperture widths lead to narrower beams.

(b) Half-power beamwidth (HPBW)

The half-power point in the E -plane occurs when

$$\frac{\sin(0.5\beta L_y \sin \theta)}{(0.5\beta L_y \sin \theta)} = \frac{1}{\sqrt{2}}, \quad (17.62)$$

or

$$0.5\beta L_y \sin \theta_{|\theta=\theta_h} \approx 1.391, \quad (17.63)$$

$$\Rightarrow \theta_h \approx \arcsin\left(\frac{0.443\lambda}{L_y}\right), \text{ rad.} \quad (17.64)$$

$$HPBW_E \approx 2 \arcsin\left(\frac{0.443\lambda}{L_y}\right). \quad (17.65)$$

A first-order approximation is possible for very small arguments in (17.65), i.e., when $L_y \gg 0.443\lambda$ (large aperture):

$$HPBW_E \approx 0.886 \frac{\lambda}{L_y}. \quad (17.66)$$

The half-power beamwidth in the H -plane is analogous:

$$HPBW_H \approx 2 \arcsin \left(\frac{0.443\lambda}{L_x} \right). \quad (17.67)$$

Side-lobe level

It is obvious from the properties of the $|\sin x / x|$ function that the first side lobe has the largest maximum of all side lobes, and it is

$$|E_\theta(\theta = \theta_s)| \approx \left| \frac{\sin 4.494}{4.494} \right| \approx 0.217 \approx -13.26, \text{ dB}. \quad (17.68)$$

When evaluating side-lobe levels and beamwidths in the H -plane, one has to include the $\cos \theta$ factor. The larger the aperture, the less important this factor is.

Directivity

The antenna solid angle Ω_A is needed to calculate the directivity from

$$D_0 = 4\pi / \Omega_A. \quad (17.69)$$

The radiation intensity in any direction can be expressed through the normalized field pattern as

$$U(\theta, \varphi) = U_{\max} \cdot |\bar{F}(\theta, \varphi)|^2. \quad (17.70)$$

The far-field pattern $|\bar{F}(\theta, \varphi)|$ is available from (17.49), namely,

$$|\bar{E}| = F(\theta, \varphi) = \sqrt{1 - \sin^2 \theta \cos^2 \varphi} \cdot \left(\frac{\sin u}{u} \right) \left(\frac{\sin v}{v} \right). \quad (17.71)$$

The antenna solid angle is then calculated as

$$\Omega_A = \int_0^{2\pi} \int_0^{\pi/2} |F(\theta, \varphi)|^2 \sin \theta d\theta d\varphi, \quad (17.72)$$

which, in turn, is used to compute the directivity from (17.69).

However, in the case of an aperture illuminated by a TEM wave, we can use a simpler approach. Generally, for all aperture antennas, the assumption of a uniform TEM wave at the aperture ($\mathbf{E} = \hat{\mathbf{y}}E_0$),

$$\mathbf{H}_a = -\hat{\mathbf{x}}E_0 / \eta, \quad (17.73)$$

is quite accurate (although η is not necessarily the intrinsic impedance of the medium; could be the wave impedance of a waveguide). The far-field components in this case were already derived in (17.41) and (17.42). They lead to the following expression for the radiation intensity,

$$U(\theta, \varphi) = \frac{\beta^2}{32\pi^2\eta} (1 + \cos\theta)^2 (|I_x^E(\theta, \varphi)|^2 + |I_y^E(\theta, \varphi)|^2). \quad (17.74)$$

The maximum value of the function in (17.74) is derived after substituting the radiation integrals from (17.26) and (17.27):

$$U_{\max} = \frac{\beta^2}{8\pi^2\eta} \left| \iint_{S_A} \mathbf{E}_a ds' \right|^2. \quad (17.75)$$

The integration of the radiation intensity (17.74) over a closed sphere is not easy. It can be avoided by observing that the total power reaching the far zone must have passed through the aperture in the first place. In an aperture, where the field obeys (17.73), this power is determined as

$$\Pi = \oiint_S \mathbf{P}_{av} \cdot d\mathbf{s} = \frac{1}{2\eta} \iint_{S_A} |\mathbf{E}_a|^2 ds. \quad (17.76)$$

Substituting (17.75) and (17.76) into (17.69) finally yields

$$D_0 = \frac{4\pi}{\lambda^2} \times \frac{\left| \iint_{S_A} \mathbf{E}_a ds' \right|^2}{\iint_{S_A} |\mathbf{E}_a|^2 ds'}. \quad (17.77)$$

← A_{eff}

In the case of a *uniform rectangular aperture*,

$$\Pi = L_x L_y \frac{|E_0|^2}{2\eta}, \quad (17.78)$$

$$U_{\max} = \left(\frac{L_x L_y}{\lambda} \right)^2 \frac{|E_0|^2}{2\eta}. \quad (17.79)$$

Thus, the directivity is found to be

$$D_0 = 4\pi \frac{U_{\max}}{\Pi} = \frac{4\pi}{\lambda^2} L_x L_y = \frac{4\pi}{\lambda^2} A_p = \frac{4\pi}{\lambda^2} A_{\text{eff}}. \quad (17.80)$$

Note that *the physical and effective areas of a uniform aperture are equal*.

6. The Uniform Rectangular Aperture in Open Space

Now the rectangular aperture is *not* mounted on a ground plane. The field distribution is the same as in (17.43) but now the \mathbf{H} field must be defined, too, in order to apply the equivalence principle with both types of surface currents,

$$\left. \begin{aligned} \mathbf{E}_a &= \hat{\mathbf{y}} E_0 \\ \mathbf{H}_a &= -\hat{\mathbf{x}} E_0 / \eta \end{aligned} \right\} \begin{aligned} -L_x / 2 \leq x' \leq L_x / 2 \\ -L_y / 2 \leq y' \leq L_y / 2. \end{aligned} \quad (17.81)$$

Again, an assumption was made that there is a simple relation between the electric and the magnetic field components through the impedance η .

To form the equivalent problem, an infinite surface is chosen to extend in the $z = 0$ plane. Over the entire surface, the equivalent \mathbf{J}_s and \mathbf{M}_s surface currents must be defined. Both \mathbf{J}_s and \mathbf{M}_s are not really zero outside the aperture in the $z = 0$ plane because the respective tangential field is not zero. Moreover, the field is *not known a priori* outside the aperture. Thus, an exact equivalent problem cannot be built.

The usual assumption is that \mathbf{E}_a and \mathbf{H}_a are zero outside the aperture in the $z = 0$ plane, and, therefore, so are the equivalent currents \mathbf{J}_s and \mathbf{M}_s ,

$$\left. \begin{aligned} \mathbf{M}_s &= -\hat{\mathbf{n}} \times \mathbf{E}_a = \underbrace{-\hat{\mathbf{z}} \times \hat{\mathbf{y}}}_{\hat{\mathbf{x}}} E_0 \\ \mathbf{J}_s &= \hat{\mathbf{n}} \times \mathbf{H}_a = \underbrace{\hat{\mathbf{z}} \times (-\hat{\mathbf{x}})}_{-\hat{\mathbf{y}}} \frac{E_0}{\eta} \end{aligned} \right\} \text{for } \begin{cases} -L_x / 2 \leq x' \leq L_x / 2 \\ -L_y / 2 \leq y' \leq L_y / 2. \end{cases} \quad (17.82)$$

$$\mathbf{J}_s = \mathbf{M}_s = 0 \text{ for } |x'| > L_x / 2, \quad |y'| > L_y$$

Since the equivalent currents are related via the impedance assumption (17.81), only the integral $I_y^E(\theta, \varphi)$ is needed for substitution in the far-field expressions (17.41)-(17.42). $I_y^E(\theta, \varphi)$ is the same as in (17.46), i.e.,

$$I_y^E(\theta, \varphi) = 2E_0L_xL_y \frac{\sin(0.5\beta L_x \sin \theta \cos \varphi)}{(0.5\beta L_x \sin \theta \cos \varphi)} \cdot \frac{\sin(0.5\beta L_y \sin \theta \sin \varphi)}{(0.5\beta L_y \sin \theta \sin \varphi)}. \quad (17.83)$$

The far-field components are obtained by substituting (17.83) into (17.41) and (17.42):

$$\begin{aligned} E_\theta &= C \sin \varphi \frac{(1 + \cos \theta)}{2} \frac{\sin u}{u} \frac{\sin v}{v}, \\ E_\varphi &= C \cos \varphi \frac{(1 + \cos \theta)}{2} \frac{\sin u}{u} \frac{\sin v}{v}, \end{aligned} \quad (17.84)$$

where

$$\begin{aligned} C &= j\beta L_x L_y E_0 \frac{e^{-j\beta r}}{2\pi r}, \\ u &= 0.5\beta L_x \sin \theta \cos \varphi, \\ v &= 0.5\beta L_y \sin \theta \sin \varphi. \end{aligned}$$

The far-field expressions in (17.84) are very similar to those of the aperture mounted on a ground plane, see (17.48). For small values of θ , the patterns of both apertures are practically identical.

An exact analytical evaluation of the directivity is difficult. However, according to the approximations made, the directivity formula derived in (17.77) should provide accurate enough value. According to (17.77), the directivity is the same as in the case of the aperture mounted on a ground plane.

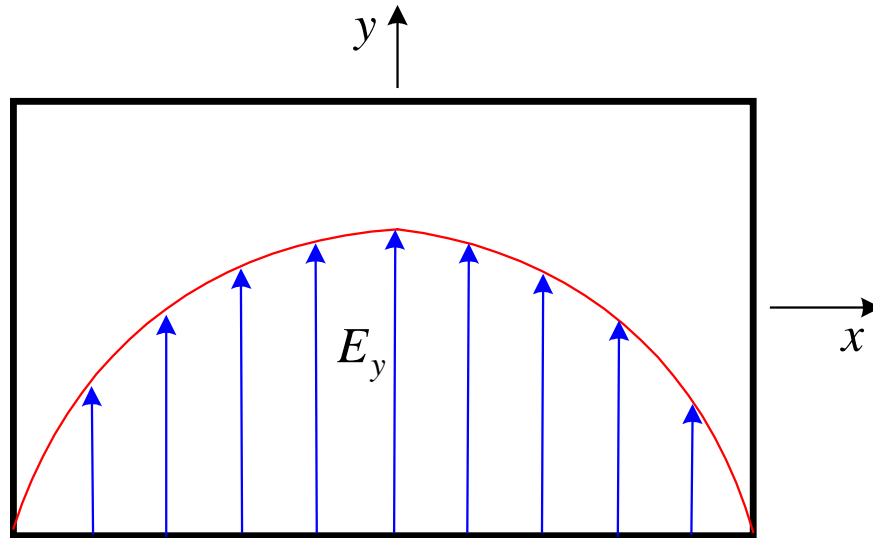
7. The Tapered Rectangular Aperture on a Ground Plane

The uniform rectangular aperture has the maximum possible effective area (for an aperture-type antenna) equal to its physical area. This also implies that it has the highest possible directivity for all constant-phase excitations of a rectangular aperture. However, the directivity is not the only important factor in the design of an antenna. A factor that often is in conflict with the directivity is the side-lobe level (SLL). The uniform distribution excitation produces the highest SLL of all constant-phase excitations of a rectangular aperture. It is shown below that a reduction of the SLL can be achieved by tapering the

equivalent sources distribution from a maximum at the aperture's center to zero values at its edges.

One practical aperture of tapered source distribution is the open rectangular waveguide. The dominant TE₁₀ mode has the following distribution:

$$\mathbf{E}_a = \hat{\mathbf{y}}E_0 \cos\left(\frac{\pi}{L_x}x'\right), \quad \begin{cases} -L_x/2 \leq x' \leq L_x/2 \\ -L_y/2 \leq y' \leq L_y/2 \end{cases} \quad (17.85)$$



The general procedure for the far-field analysis is the same as before (Sections 5 and 6). The only difference is in the field distribution. Again, only the integral $I_y^E(\theta, \varphi)$ is evaluated:

$$I_y^E(\theta, \varphi) = 2E_0 \int_{-L_x/2}^{L_x/2} \cos\left(\frac{\pi}{L_x}x'\right) e^{j\beta x' \sin \theta \cos \varphi} dx' \cdot \int_{-L_y/2}^{L_y/2} e^{j\beta y' \sin \theta \sin \varphi} dy'. \quad (17.86)$$

The integral of the y' variable was already encountered in (17.45)-(17.46):

$$I_y(\theta, \varphi) = \int_{-L_y/2}^{L_y/2} e^{j\beta y' \sin \theta \sin \varphi} dy' = L_y \frac{\sin(0.5\beta L_y \sin \theta \sin \varphi)}{(0.5\beta L_y \sin \theta \sin \varphi)}. \quad (17.87)$$

The integral of the x' variable is also easily solved:

$$\begin{aligned}
I_x(\theta, \varphi) &= \int_{-L_x/2}^{L_x/2} \cos\left(\frac{\pi}{L_x} x'\right) e^{j\beta x' \sin \theta \cos \varphi} dx' = \\
&= \int_{-L_x/2}^{L_x/2} \cos\left(\frac{\pi}{L_x} x'\right) \left[\cos(\beta x' \sin \theta \cos \varphi) + j \sin(\beta x' \sin \theta \cos \varphi) \right] dx' = \\
&= \frac{1}{2} \int_{-L_x/2}^{L_x/2} \left\{ \cos\left[\left(\frac{\pi}{L_x} - \beta \sin \theta \cos \varphi\right) x'\right] + \cos\left[\left(\frac{\pi}{L_x} + \beta \sin \theta \cos \varphi\right) x'\right] \right\} dx' + \\
&+ \frac{j}{2} \int_{-L_x/2}^{L_x/2} \left\{ \sin\left[\left(\beta \sin \theta \cos \varphi - \frac{\pi}{L_x}\right) x'\right] + \sin\left[\left(\beta \sin \theta \cos \varphi + \frac{\pi}{L_x}\right) x'\right] \right\} dx' \\
\Rightarrow I_x(\theta, \varphi) &= \frac{\pi L_x}{2} \frac{\cos\left(\frac{\beta L_x}{2} \sin \theta \cos \varphi\right)}{\left[\left(\frac{\pi}{2}\right)^2 - \frac{\beta L_x}{2} \sin \theta \cos \varphi\right]} \quad (17.88)
\end{aligned}$$

The substitution of (17.87) and (17.88) in (17.86) leads to

$$I_y^E(\theta, \varphi) = \pi E_0 L_x L_y \frac{\cos\left(\frac{\beta L_x}{2} \sin \theta \cos \varphi\right)}{\left[\left(\frac{\pi}{2}\right)^2 - \underbrace{\frac{\beta L_x}{2} \sin \theta \cos \varphi}_{u(\theta, \varphi)}\right]} \times \frac{\sin\left[\frac{\beta L_y}{2} \sin \theta \sin \varphi\right]}{\underbrace{\left(\frac{\beta L_y}{2} \sin \theta \sin \varphi\right)}_{v(\theta, \varphi)}} \quad (17.89)$$

To derive the far-field components, (17.89) is substituted in (17.34)-(17.35):

$$\begin{aligned}
E_\theta(\theta, \varphi) &= -\frac{\pi}{2} C \sin \varphi \cdot \frac{\cos u}{\left[u^2 - \left(\frac{\pi}{2}\right)^2\right]} \cdot \frac{\sin v}{v} \\
E_\varphi(\theta, \varphi) &= -\frac{\pi}{2} C \cos \theta \cos \varphi \cdot \frac{\cos u}{\left[u^2 - \left(\frac{\pi}{2}\right)^2\right]} \cdot \frac{\sin v}{v}
\end{aligned} \quad (17.90)$$

where

$$C = j\beta L_x L_y E_0 \frac{e^{-j\beta r}}{2\pi r},$$
$$u = 0.5\beta L_x \sin \theta \cos \varphi,$$
$$v = 0.5\beta L_y \sin \theta \sin \varphi.$$

Principle plane patterns

In the E -plane, the aperture is not tapered. As expected, the E -plane principal pattern is the same as that of a uniform aperture.

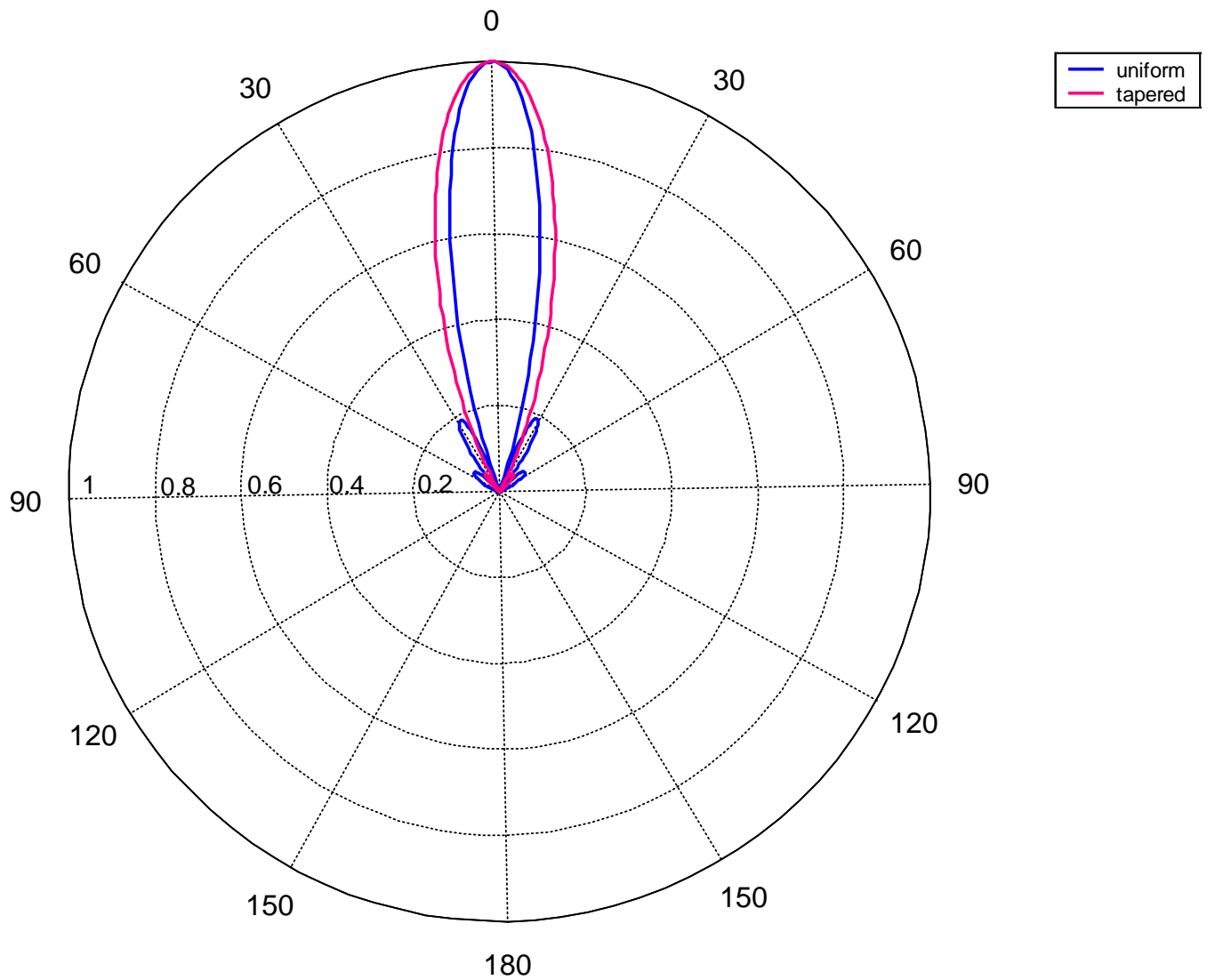
E -plane ($\varphi = 90^\circ$):

$$\bar{E}_\theta = \frac{\sin\left(\frac{\beta L_y}{2} \sin \theta\right)}{\left(\frac{\beta L_y}{2} \sin \theta\right)} \quad (17.91)$$

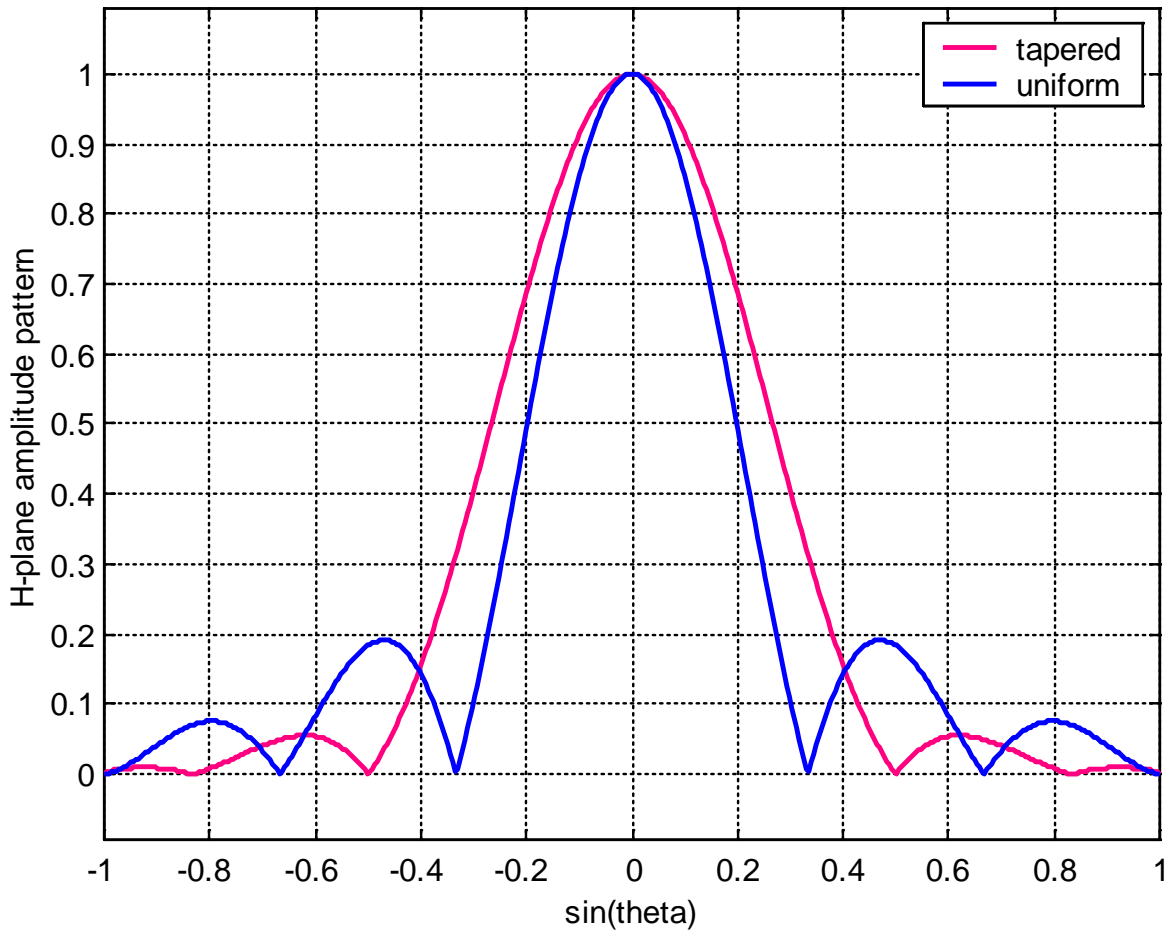
H -plane ($\varphi = 0^\circ$):

$$\bar{E}_\varphi = \cos \theta \frac{\cos\left(\frac{\beta L_x}{2} \sin \theta\right)}{\left(\frac{\beta L_x}{2} \sin \theta\right)^2 - \left(\frac{\pi}{2}\right)^2} \quad (17.92)$$

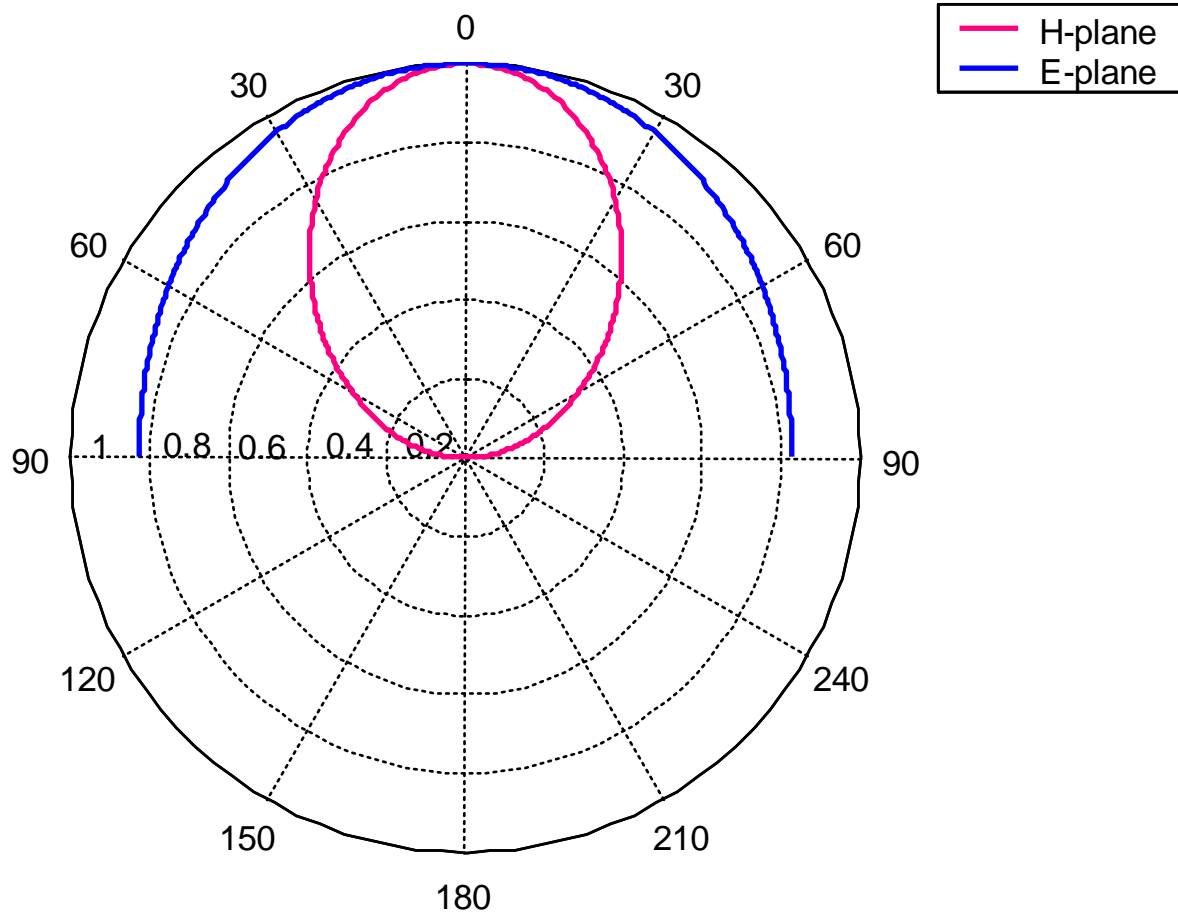
H-PLANE PATTERN – UNIFORM VS. TAPERED ILLUMINATION ($L_x = 3\lambda$):



The lower SLL of the tapered-source pattern is obvious. It is better seen in the rectangular plot given below. The price to pay for the lower SLL is the decrease in directivity (the beamwidth of the major lobe increases).



The above example of $L_x = 3\lambda$ illustrates well the effect of source distribution on the far-field pattern. However, a more practical example is the rectangular-waveguide open-end aperture, where the waveguide operates in a dominant mode, i.e. $\lambda_0 / 2 < L_x < \lambda_0$. Here, λ_0 is the wavelength in open space. Consider the case $L_x = 0.75\lambda_0$. The principal-plane patterns for an aperture on a ground plane look like this:



In the above example, a practical X-band waveguide was considered whose cross-section has the following size: $L_x = 2.286$ cm, $L_y = 1.016$ cm. ($\lambda_0 = 3.048$ cm and $f_0 = 9.84$ GHz)

The case of a dominant-mode open-end waveguide radiating in free space can be analyzed following the approaches outlined in this Section and in Section 6.

The calculation of the beamwidths and the directivity is analogous to the previous cases. Only the final results will be given here for the case of the x -tapered (cosine taper) aperture on a ground plane.

$$\text{Directivity: } D_0 = \frac{4\pi}{\lambda^2} \cdot \frac{8}{\pi^2} L_x L_y \quad (17.93)$$

$$\text{Effective area: } A_{eff} = \frac{8}{\pi^2} L_x L_y \approx 0.81 A_p \quad (17.94)$$

Note the decrease in the effective area compared to the uniform-aperture case.

Half-power beamwidths:

$$HPBW_E = \frac{50.6}{L_y / \lambda}, \text{ deg. } (= HPBW_E \text{ of the uniform aperture}) \quad (17.95)$$

$$HPBW_H = \frac{68.8}{L_x / \lambda}, \text{ deg. } (> HPBW_H \text{ of the uniform aperture}) \quad (17.96)$$

The above results are approximate. Better results are obtained if the following factors are taken into account:

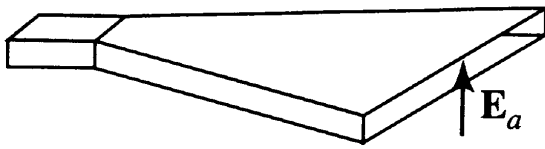
- the phase constant of the waveguide β_g and its wave impedance Z_g are not equal to the free-space phase constant $\beta_0 = \omega\sqrt{\mu_0\epsilon_0}$ and intrinsic impedance $Z_0 = \sqrt{\mu_0 / \epsilon_0}$; they are dispersive;
- the abrupt termination at the waveguide open end introduces reflection, which affects the field at the aperture;
- there are strong fringe currents at the waveguide walls, which contribute to the overall radiation.

LECTURE 18: Horn Antennas

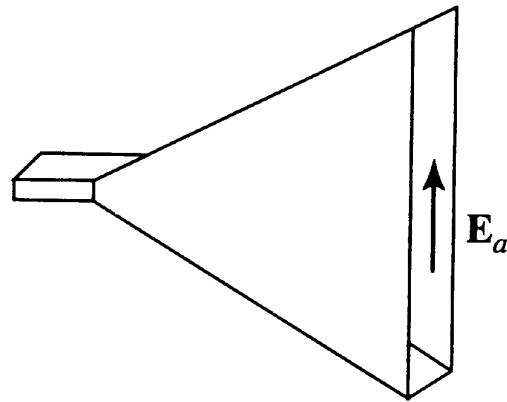
(Rectangular horn antennas. Circular apertures.)

1 Rectangular Horn Antennas

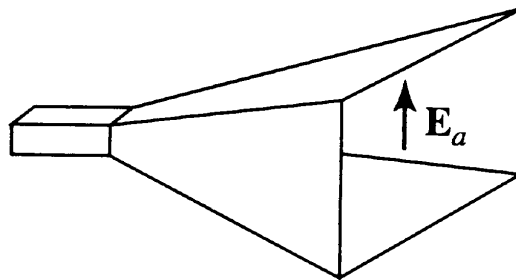
Horn antennas are popular in the microwave bands (above 1 GHz). Horns provide high gain, low VSWR (with waveguide feeds), relatively wide bandwidth, and they are not difficult to make. There are three basic types of rectangular horns.



(a) *H*-plane sectoral horn.



(b) *E*-plane sectoral horn.



(c) Pyramidal horn.

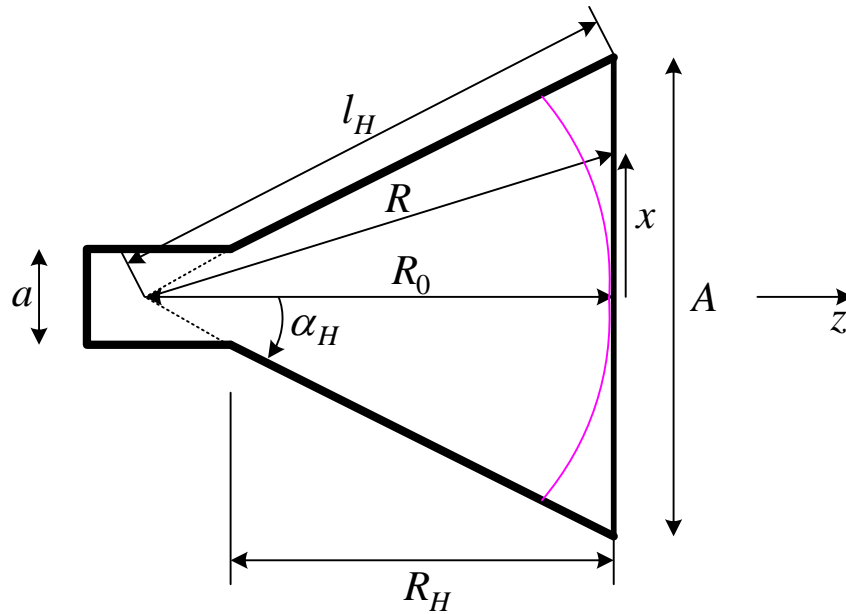
The horns can be also flared exponentially. This provides better impedance match in a broader frequency band. Such horns are more difficult to make, which means higher cost.

The rectangular horns are ideally suited for rectangular waveguide feeds. The horn acts as a gradual transition from a waveguide mode to a free-space mode of the EM wave. When the feed is a cylindrical waveguide, the antenna is usually a *conical horn*.

Why is it necessary to consider the horns separately instead of applying the theory of waveguide aperture antennas directly? It is because the so-called *phase error* occurs due to the difference between the lengths from the center of the feed to the center of the horn aperture and the horn edge. This makes the uniform-phase aperture results invalid for the horn apertures.

1.1 The H -plane sectoral horn

The geometry and the respective parameters shown in the figure below are used in the subsequent analysis.



H -plane (x - z) cut of an H -plane sectoral horn

$$l_H^2 = R_0^2 + \left(\frac{A}{2}\right)^2, \quad (18.1)$$

$$\alpha_H = \arctan\left(\frac{A}{2R_0}\right), \quad (18.2)$$

$$R_H = (A - a) \sqrt{\left(\frac{l_H}{A}\right)^2 - \frac{1}{4}}. \quad (18.3)$$

The two required dimensions for the construction of the horn are A and R_H .

The tangential field arriving at the input of the horn is composed of the transverse field components of the waveguide dominant mode TE_{10} :

$$\begin{cases} E_y(x) = E_0 \cos\left(\frac{\pi}{a}x\right) e^{-j\beta_g z} \\ H_x(x) = -E_y(x) / Z_g \end{cases} \quad (18.4)$$

where

$$Z_g = \frac{\eta}{\sqrt{1 - \left(\frac{\lambda_0}{2a}\right)^2}} \text{ is the wave impedance of the } TE_{10} \text{ waveguide mode;}$$

$$\beta_g = \beta_0 \sqrt{1 - \left(\frac{\lambda_0}{2a}\right)^2} \text{ is the propagation constant of the } TE_{10} \text{ mode.}$$

Here, $\beta_0 = \omega\sqrt{\mu\varepsilon} = 2\pi / \lambda_0$, and λ_0 is the free-space wavelength. The field that is illuminating the aperture of the horn is essentially a spatially expanded version of the waveguide field. Note that the wave impedance of the flared waveguide (the horn) gradually approaches the intrinsic impedance of open space η , as A (the H -plane width) increases.

The complication in the analysis arises from the fact that the waves arriving at the horn aperture are **not in phase** due to the different path lengths from the horn apex. The aperture phase variation is given by

$$e^{-j\beta(R-R_0)}. \quad (18.5)$$

Since the aperture is not flared in the y -direction, the phase is uniform along y . We first approximate the path of the wave in the horn:

$$R = \sqrt{R_0^2 + x^2} = R_0 \sqrt{1 + \left(\frac{x}{R_0}\right)^2} \approx R_0 \left[1 + \frac{1}{2} \left(\frac{x}{R_0}\right)^2\right]. \quad (18.6)$$

The last approximation holds if $x \ll R_0$, or $A/2 \ll R_0$. Then, we can assume that

$$R - R_0 \approx \frac{1}{2} \frac{x^2}{R_0}. \quad (18.7)$$

Using (18.7), the field at the aperture is approximated as

$$E_{a_y}(x) \approx E_0 \cos\left(\frac{\pi}{A}x\right) e^{-j\frac{\beta}{2R_0}x^2}. \quad (18.8)$$

The field at the aperture plane outside the aperture is assumed equal to zero. The field expression (18.8) is substituted in the integral for I_y^E (see Lecture 17):

$$I_y^E(\theta, \varphi) = \iint_{S_A} E_{a_y}(x', y') e^{j\beta(x' \sin \theta \cos \varphi + y' \sin \theta \sin \varphi)} dx' dy', \quad (18.9)$$

$$I_y^E(\theta, \varphi) = E_0 \underbrace{\int_{-A/2}^{+A/2} \cos\left(\frac{\pi}{A}x'\right) e^{-j\frac{\beta}{2R_0}x'^2} e^{j\beta x' \sin \theta \cos \varphi} dx'}_{\sim I(\theta, \varphi)} \times \int_{-b/2}^{+b/2} e^{j\beta y' \sin \theta \sin \varphi} dy'. \quad (18.10)$$

The second integral has been already encountered. The first integral is cumbersome and the final result only is given below:

$$I_y^E(\theta, \varphi) = E_0 \left[\frac{1}{2} \sqrt{\frac{\pi R_0}{\beta}} \cdot I(\theta, \varphi) \right] \times \left[b \frac{\sin(0.5\beta b \cdot \sin \theta \cdot \sin \varphi)}{(0.5\beta b \cdot \sin \theta \cdot \sin \varphi)} \right], \quad (18.11)$$

where

$$I(\theta, \varphi) = e^{j\frac{R_0}{2\beta} \left(\beta \sin \theta \cos \varphi + \frac{\pi}{A} \right)^2} \cdot [C(s'_2) - jS(s'_2) - C(s'_1) + jS(s'_1)] \\ + e^{j\frac{R_0}{2\beta} \left(\beta \sin \theta \cos \varphi - \frac{\pi}{A} \right)^2} \cdot [C(t'_2) - jS(t'_2) - C(t'_1) + jS(t'_1)] \quad (18.12)$$

and

$$s'_1 = \sqrt{\frac{1}{\pi\beta R_0}} \left(-\frac{\beta A}{2} - R_0\beta u - \frac{\pi R_0}{A} \right);$$

$$s'_2 = \sqrt{\frac{1}{\pi\beta R_0}} \left(+\frac{\beta A}{2} - R_0\beta u - \frac{\pi R_0}{A} \right);$$

$$t'_1 = \sqrt{\frac{1}{\pi\beta R_0}} \left(-\frac{\beta A}{2} - R_0\beta u + \frac{\pi R_0}{A} \right);$$

$$t'_2 = \sqrt{\frac{1}{\pi\beta R_0}} \left(+\frac{\beta A}{2} - R_0\beta u + \frac{\pi R_0}{A} \right);$$

$$u = \sin \theta \cos \varphi .$$

$C(x)$ and $S(x)$ are Fresnel integrals, which are defined as

$$\begin{aligned} C(x) &= \int_0^x \cos\left(\frac{\pi}{2}\tau^2\right) d\tau; \quad C(-x) = -C(x), \\ S(x) &= \int_0^x \sin\left(\frac{\pi}{2}\tau^2\right) d\tau; \quad S(-x) = -S(x). \end{aligned} \quad (18.13)$$

We note that more accurate evaluation of $I_y^E(\theta, \varphi)$ can be obtained if the approximation in (18.6) is not made, and E_{a_y} is substituted in (18.9) as

$$E_{a_y}(x) = E_0 \cos\left(\frac{\pi}{A}x\right) e^{-j\beta(\sqrt{R_0^2+x^2}-R_0)} = E_0 e^{+j\beta R_0} \cos\left(\frac{\pi}{A}x\right) e^{-j\beta\sqrt{R_0^2+x^2}}. \quad (18.14)$$

The far field can be calculated from $I_y^E(\theta, \varphi)$ as (see Lecture 17):

$$\begin{aligned} E_\theta &= j\beta \frac{e^{-j\beta r}}{4\pi r} (1 + \cos \theta) \sin \varphi \cdot I_y^E(\theta, \varphi), \\ E_\varphi &= j\beta \frac{e^{-j\beta r}}{4\pi r} (1 + \cos \theta) \cos \varphi \cdot I_y^E(\theta, \varphi), \end{aligned} \quad (18.15)$$

or

$$\begin{aligned} \mathbf{E} &= j\beta E_0 b \sqrt{\frac{\pi R_0}{\beta}} \frac{e^{-j\beta r}}{4\pi r} \left(\frac{1 + \cos \theta}{2}\right) \left[\frac{\sin(0.5\beta b \cdot \sin \theta \cdot \sin \varphi)}{(0.5\beta b \cdot \sin \theta \cdot \sin \varphi)} \right] \times \\ &I(\theta, \varphi) (\hat{\boldsymbol{\theta}} \sin \varphi + \hat{\boldsymbol{\phi}} \cos \varphi). \end{aligned} \quad (18.16)$$

The amplitude pattern of the H -plane sectoral horn is obtained as

$$\bar{E}(\theta, \varphi) = \left(\frac{1 + \cos \theta}{2}\right) \cdot \left[\frac{\sin(0.5\beta b \cdot \sin \theta \cdot \sin \varphi)}{(0.5\beta b \cdot \sin \theta \cdot \sin \varphi)} \right] \cdot I(\theta, \varphi). \quad (18.17)$$

Principal-plane patterns

$$\mathbf{E}\text{-plane } (\varphi = 90^\circ): F_E(\theta) = \left(\frac{1 + \cos \theta}{2} \right) \left[\frac{\sin(0.5\beta b \cdot \sin \theta \cdot \sin \varphi)}{(0.5\beta b \cdot \sin \theta \cdot \sin \varphi)} \right] \quad (18.18)$$

The second factor in (18.18) is exactly the pattern of a uniform line source of length b along the y -axis.

H-plane ($\varphi = 0^\circ$):

$$\begin{aligned} F_H(\theta) &= \left(\frac{1 + \cos \theta}{2} \right) \cdot f_H(\theta) = \\ &= \left(\frac{1 + \cos \theta}{2} \right) \cdot \frac{I(\theta, \varphi = 0^\circ)}{I(\theta = 0^\circ, \varphi = 0^\circ)} \end{aligned} \quad (18.19)$$

The H -plane pattern in terms of the $I(\theta, \varphi)$ integral is an approximation, which is a consequence of the phase approximation made in (18.7). Accurate value for $f_H(\theta)$ is found by integrating numerically the field as given in (18.14), i.e.,

$$f_H(\theta) \propto \int_{-A/2}^{+A/2} \cos\left(\frac{\pi x'}{A}\right) e^{-j\beta\sqrt{R_0^2+x'^2}} e^{j\beta x' \sin \theta} dx'. \quad (18.20)$$

E- AND H-PLANE PATTERN OF H-PLANE SECTORAL HORN

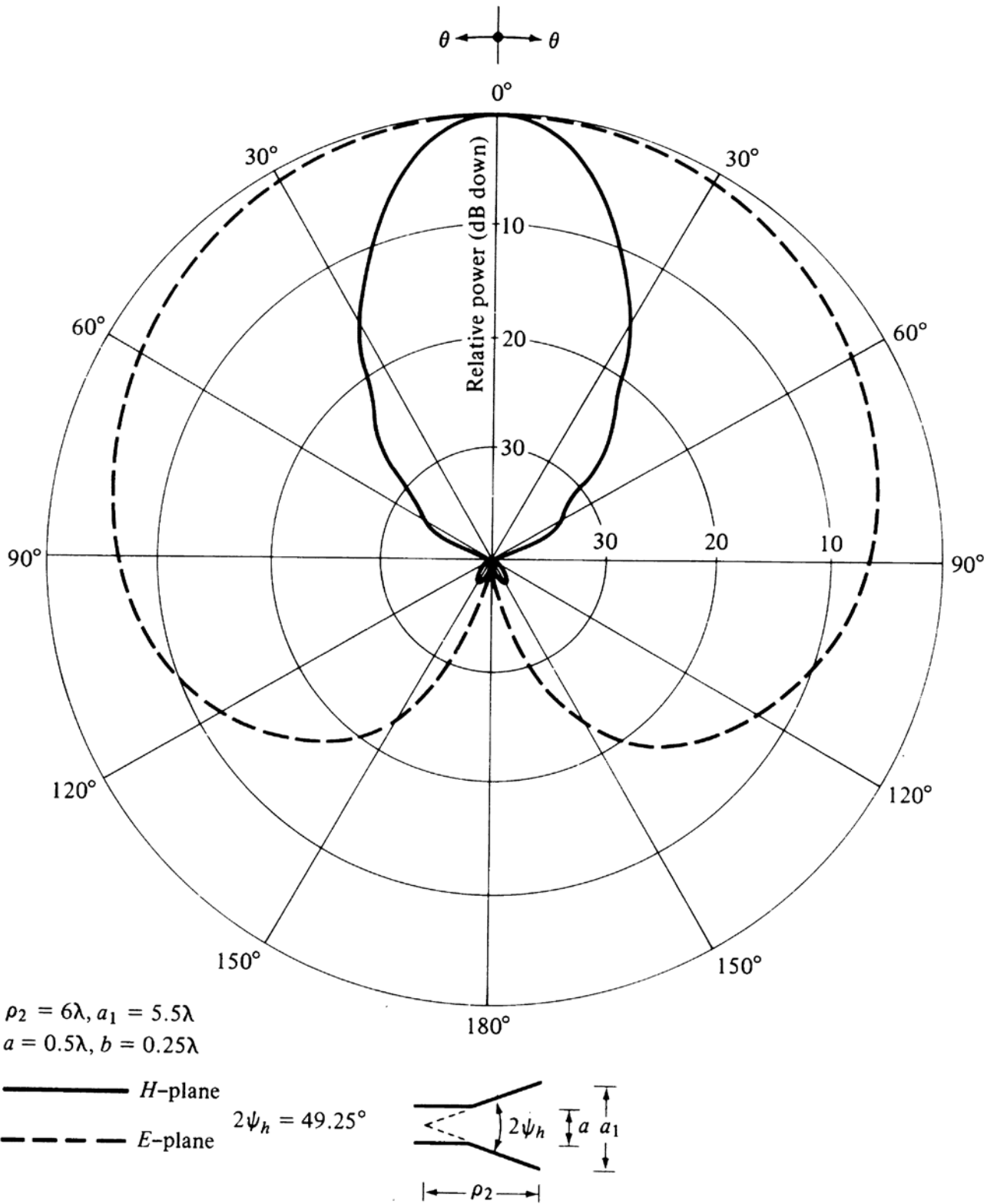


Fig. 13-12, Balanis, p. 674

The directivity of the H -plane sectoral horn is calculated by the general directivity expression for apertures (for derivation, see Lecture 17):

$$D_0 = \frac{4\pi}{\lambda^2} \cdot \frac{\left| \iint_{S_A} \mathbf{E}_a ds' \right|^2}{\iint_{S_A} |\mathbf{E}_a|^2 ds'}. \quad (18.21)$$

The integral in the denominator is proportional to the total radiated power,

$$2\eta\Pi_{rad} = \iint_{S_A} |\mathbf{E}_a|^2 ds' = \int_{-b/2-A/2}^{+b/2+A/2} \int |E_0|^2 \cos^2\left(\frac{\pi}{A}x'\right) dx'dy' = |E_0|^2 \frac{Ab}{2}. \quad (18.22)$$

In the solution of the integral in the numerator of (18.21), the field is substituted with its phase approximated as in (18.8). The final result is

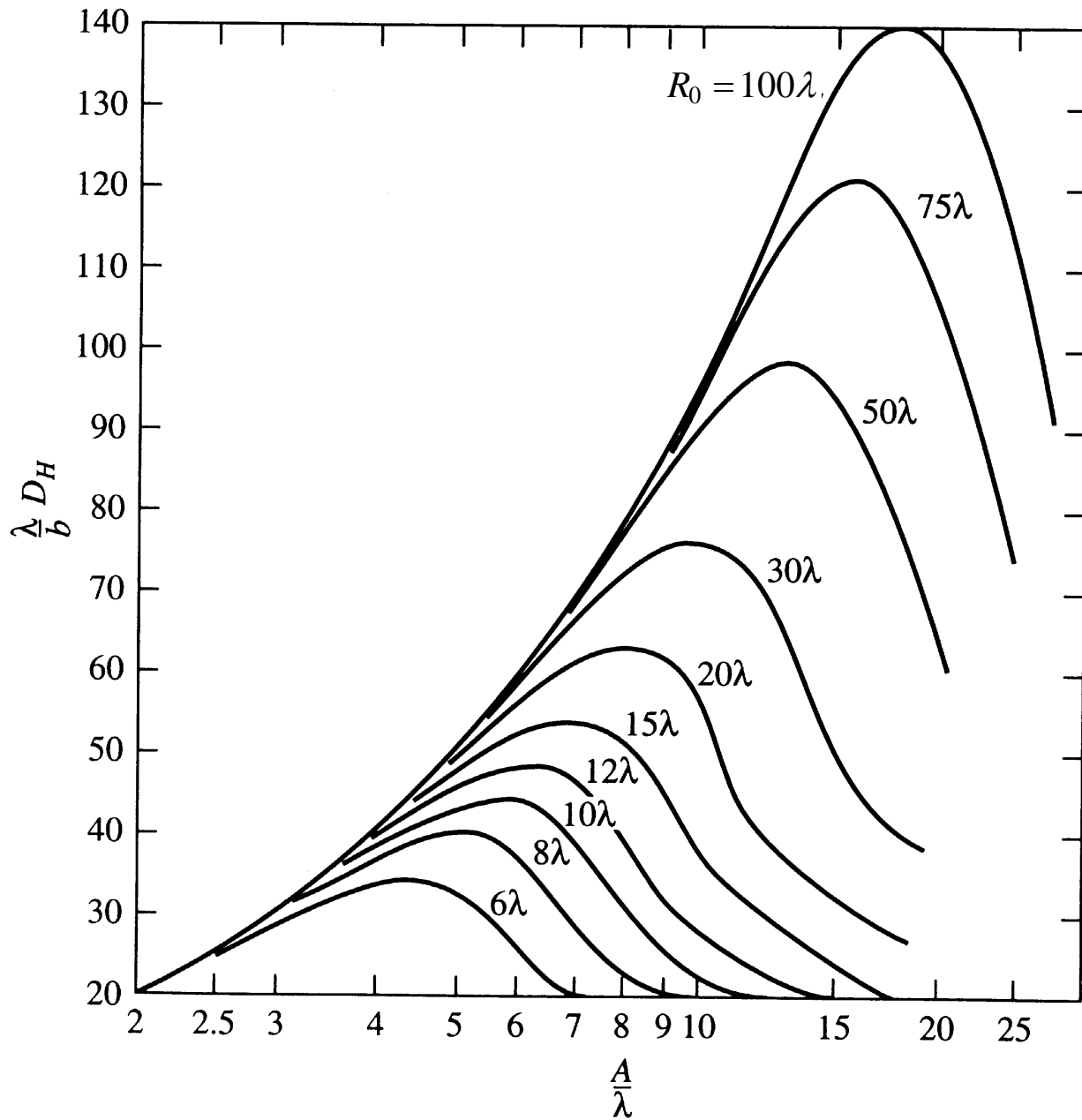
$$D_H = \frac{b}{\lambda} \frac{32}{\pi} \left(\frac{A}{\lambda}\right) \varepsilon_{ph}^H = \frac{4\pi}{\lambda^2} \varepsilon_t \varepsilon_{ph}^H (Ab), \quad (18.23)$$

where

$$\begin{aligned} \varepsilon_t &= \frac{8}{\pi^2}; \\ \varepsilon_{ph}^H &= \frac{\pi^2}{64t} \left\{ [C(p_1) - C(p_2)]^2 + [S(p_1) - S(p_2)]^2 \right\}; \\ p_1 &= 2\sqrt{t} \left[1 + \frac{1}{8t} \right], \quad p_2 = 2\sqrt{t} \left[-1 + \frac{1}{8t} \right]; \\ t &= \frac{1}{8} \left(\frac{A}{\lambda}\right)^2 \frac{1}{R_0 / \lambda}. \end{aligned}$$

The factor ε_t explicitly shows the aperture efficiency associated with the aperture cosine taper. The factor ε_{ph}^H is the aperture efficiency associated with the aperture phase distribution.

A family of universal directivity curves is given below. From these curves, it is obvious that for a given axial length R_0 and at a given wavelength, there is an optimal aperture width A corresponding to the maximum directivity.



[Stutzman&Thiele, *Antenna Theory and Design*]

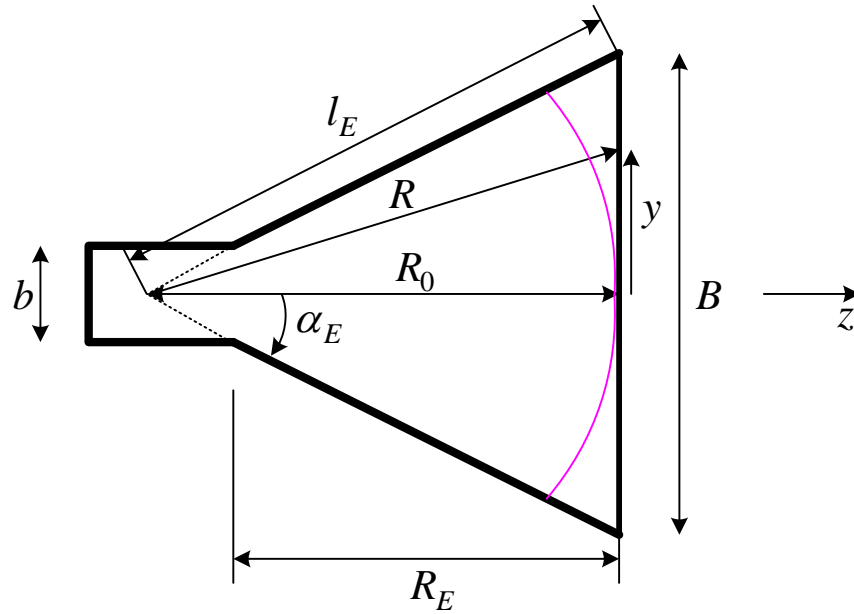
It can be shown that the optimal directivity is obtained if the relation between A and R_0 is

$$A = \sqrt{3\lambda R_0}, \quad (18.24)$$

or

$$\frac{A}{\lambda} = \sqrt{3 \frac{R_0}{\lambda}}. \quad (18.25)$$

1.2 The E -plane sectoral horn



E -plane (y - z) cut of an E -plane sectoral horn

The geometry of the E -plane sectoral horn in the E -plane (y - z plane) is analogous to that of the H -plane sectoral horn in the H -plane. The analysis is following the same steps as in the previous section. The field at the aperture is approximated by [compare with (18.8)]

$$E_{a_y} = E_0 \cos\left(\frac{\pi}{a}x\right) e^{-j\frac{\beta}{2R_0}y^2}. \quad (18.26)$$

Here, the approximations

$$R = \sqrt{R_0^2 + y^2} = R_0 \sqrt{1 + \left(\frac{y}{R_0}\right)^2} \approx R_0 \left[1 + \frac{1}{2} \left(\frac{y}{R_0}\right)^2\right] \quad (18.27)$$

and

$$R - R_0 \approx \frac{1}{2} \frac{y^2}{R_0} \quad (18.28)$$

are made, which are analogous to (18.6) and (18.7).

The radiation field is obtained as

$$\mathbf{E} = j\beta E_0 \frac{4a}{\pi} \sqrt{\frac{\pi R_0}{\beta}} \frac{e^{-j\beta r}}{4\pi r} e^{j\left(\frac{\beta R_0}{2}\right)\left(\frac{\beta B}{2}\sin\theta\sin\varphi\right)^2} \cdot (\hat{\boldsymbol{\theta}}\sin\varphi + \hat{\boldsymbol{\phi}}\cos\varphi) \times \frac{(1+\cos\theta)}{2} \frac{\cos\left(\frac{\beta a}{2}\sin\theta\cos\varphi\right)}{\left[1-\left(\frac{\beta a}{2}\sin\theta\cos\varphi\right)^2\right]} [C(r_2) - jS(r_2) - C(r_1) + jS(r_1)]. \quad (18.29)$$

The arguments of the Fresnel integrals used in (18.29) are

$$r_1 = \sqrt{\frac{\beta}{\pi R_0}} \left(-\frac{B}{2} - R_0 \frac{\beta B}{2} \sin\theta\sin\varphi \right), \quad (18.30)$$

$$r_2 = \sqrt{\frac{\beta}{\pi R_0}} \left(+\frac{B}{2} - R_0 \frac{\beta B}{2} \sin\theta\sin\varphi \right).$$

Principal-plane patterns

The **normalized H -plane pattern** is found by substituting $\varphi = 0$ in (18.29):

$$\bar{H}(\theta) = \left(\frac{1+\cos\theta}{2} \right) \times \frac{\cos\left(\frac{\beta a}{2}\sin\theta\right)}{1-\left(\frac{\beta a}{2}\sin\theta\right)^2}. \quad (18.31)$$

The second factor in this expression is the pattern of a uniform-phase cosine-amplitude tapered line source.

The **normalized E -plane pattern** is found by substituting $\varphi = 90^\circ$ in (18.29):

:

$$\bar{E}(\theta) = \frac{(1+\cos\theta)}{2} |f_E(\theta)| = \frac{(1+\cos\theta)}{2} \sqrt{\frac{[C(r_2)-C(r_1)]^2 + [S(r_2)-S(r_1)]^2}{4[C^2(r_{\theta=0})+S^2(r_{\theta=0})]}}. \quad (18.32)$$

Here, the arguments of the Fresnel integrals are calculated for $\varphi = 90^\circ$:

$$\begin{aligned}
r_1 &= \sqrt{\frac{\beta}{\pi R_0}} \left(-\frac{B}{2} - R_0 \frac{\beta B}{2} \sin \theta \right), \\
r_2 &= \sqrt{\frac{\beta}{\pi R_0}} \left(+\frac{B}{2} - R_0 \frac{\beta B}{2} \sin \theta \right),
\end{aligned} \tag{18.33}$$

and

$$r_{\theta=0} = r_2(\theta = 0) = \frac{B}{2} \sqrt{\frac{\beta}{\pi R_0}}. \tag{18.34}$$

Similar to the H -plane sectoral horn, the principal E -plane pattern can be accurately calculated if no approximation of the phase distribution is made. Then, the function $f_E(\theta)$ has to be calculated by numerical integration of (compare with (18.20))

$$f_E(\theta) \propto \int_{-B/2}^{B/2} e^{-j\beta\sqrt{R_0^2+y'^2}} e^{j\beta\sin\theta\cdot y'} dy'. \tag{18.35}$$

E- AND H-PLANE PATTERN OF E-PLANE SECTORAL HORN

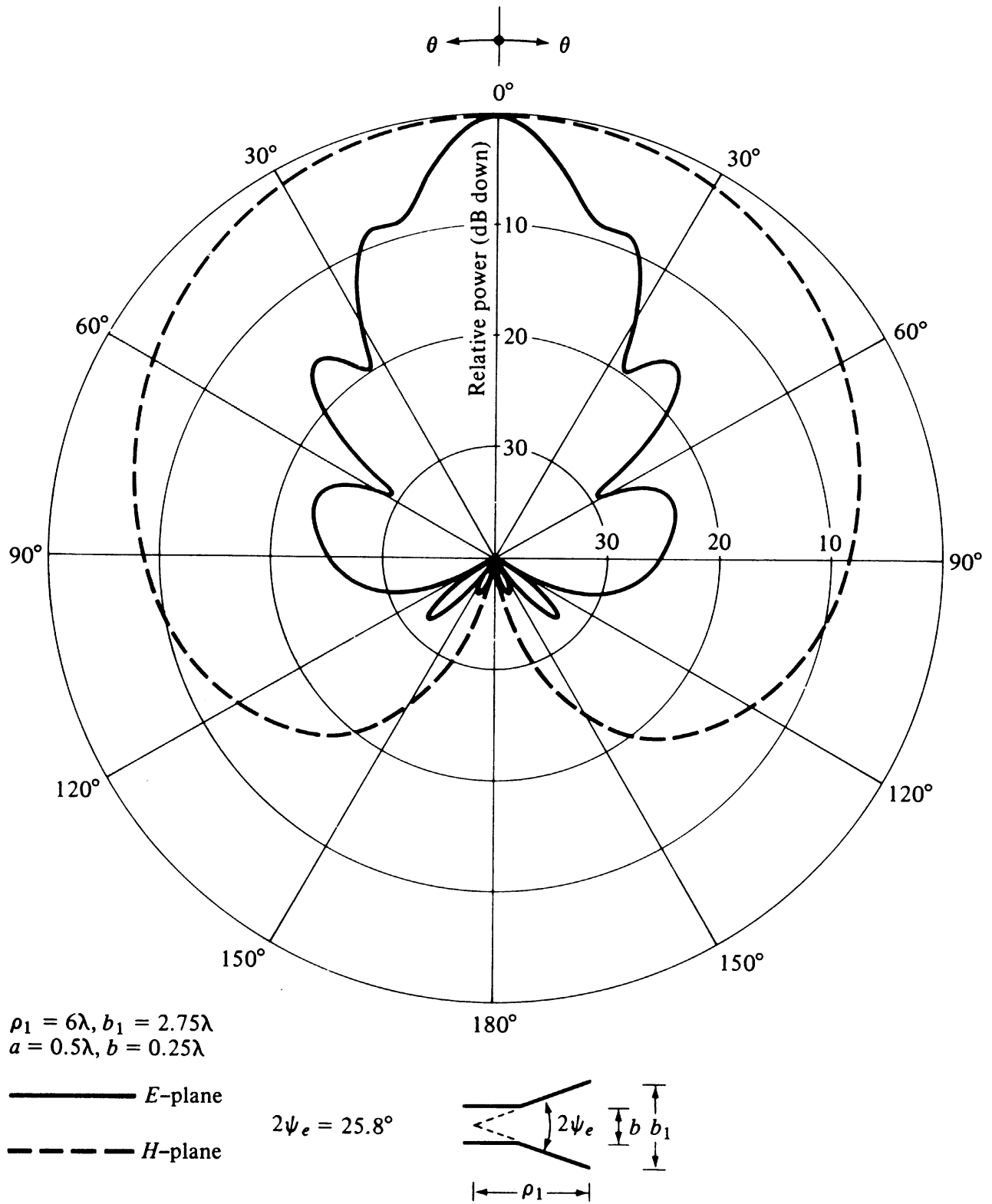


Fig. 13.4, Balanis, p. 660

Directivity

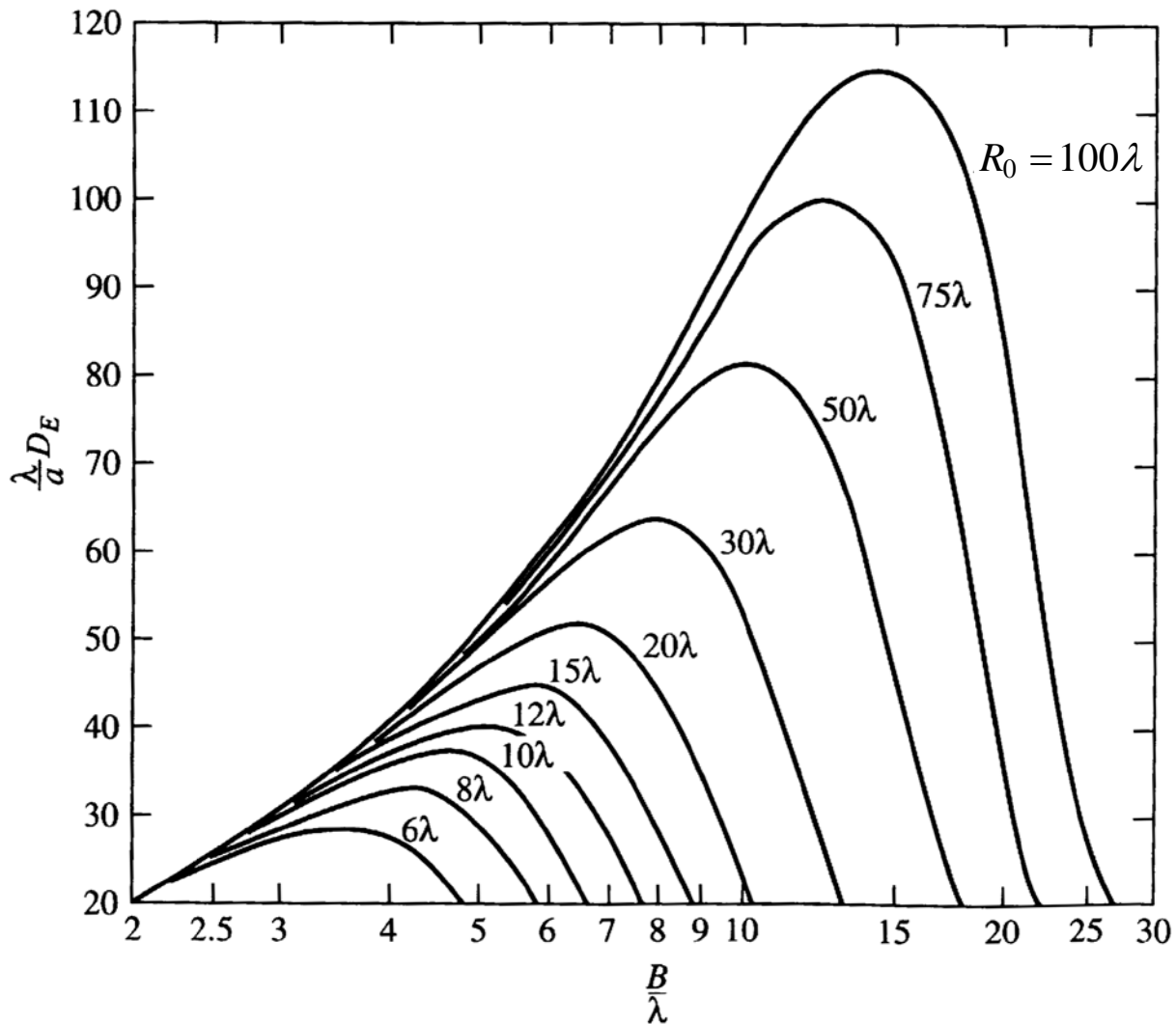
The directivity of the E -plane sectoral horn is found in a manner analogous to the H -plane sectoral horn:

$$D_E = \frac{a}{\lambda} \frac{32}{\pi} \frac{B}{\lambda} \varepsilon_{ph}^E = \frac{4\pi}{\lambda^2} \varepsilon_t \varepsilon_{ph}^E aB, \quad (18.36)$$

where

$$\varepsilon_t = \frac{8}{\pi^2}, \quad \varepsilon_{ph}^E = \frac{C^2(q) + S^2(q)}{q^2}, \quad q = \frac{B}{\sqrt{2\lambda R_0}}.$$

A family of universal directivity curves $\lambda D_E / a$ vs. B / λ with R_0 being a parameter is given below.



[Stutzman&Thiele, *Antenna Theory and Design*]

The optimal relation between the flared height B and the horn apex length R_0 that produces the maximum possible directivity is

$$B = \sqrt{2\lambda R_0}. \quad (18.37)$$

1.3 The pyramidal horn

The pyramidal horn is probably the most popular antenna in the microwave frequency ranges (from ≈ 1 GHz up to ≈ 18 GHz). The feeding waveguide is flared in both directions, the E -plane and the H -plane. All results are combinations of the E -plane sectoral horn and the H -plane sectoral horn analyses. The field distribution at the aperture is approximated as

$$E_{a_y} \approx E_0 \cos\left(\frac{\pi}{A}x\right) e^{-j\frac{\beta}{2}\left(\frac{x^2}{R_0^E} + \frac{y^2}{R_0^H}\right)}. \quad (18.38)$$

The E -plane principal pattern of the pyramidal horn is the same as the E -plane principal pattern of the E -plane sectoral horn. The same holds for the H -plane patterns of the pyramidal horn and the H -plane sectoral horn.

The directivity of the pyramidal horn can be found by introducing the phase efficiency factors of both planes and the taper efficiency factor of the H -plane:

$$D_P = \frac{4\pi}{\lambda^2} \varepsilon_t \varepsilon_{ph}^E \varepsilon_{ph}^H (AB), \quad (18.39)$$

where

$$\varepsilon_t = \frac{8}{\pi^2};$$

$$\varepsilon_{ph}^H = \frac{\pi^2}{64t} \left\{ [C(p_1) - C(p_2)]^2 + [S(p_1) - S(p_2)]^2 \right\};$$

$$p_1 = 2\sqrt{t} \left[1 + \frac{1}{8t} \right], \quad p_2 = 2\sqrt{t} \left[-1 + \frac{1}{8t} \right], \quad t = \frac{1}{8} \left(\frac{A}{\lambda} \right)^2 \frac{1}{R_0^H / \lambda};$$

$$\varepsilon_{ph}^E = \frac{C^2(q) + S^2(q)}{q^2}, \quad q = \frac{B}{\sqrt{2\lambda R_0^E}}.$$

The gain of a horn is usually very close to its directivity because the radiation

efficiency is very good (low losses). The directivity as calculated with (18.39) is very close to measurements. The above expression is a physical optics approximation, and it does not take into account only multiple diffractions, and the diffraction at the edges of the horn arising from reflections from the horn interior. These phenomena, which are unaccounted for, lead to only very minor fluctuations of the measured results about the prediction of (18.39). That is why horns are often used as *gain standards* in antenna measurements.

The optimal directivity of an *E*-plane horn is achieved at $q=1$ [see also (18.37)], $\varepsilon_{ph}^E = 0.8$. The optimal directivity of an *H*-plane horn is achieved at $t = 3/8$ [see also (18.24)], $\varepsilon_{ph}^H = 0.79$. Thus, the optimal horn has a phase aperture efficiency of

$$\varepsilon_{ph}^P = \varepsilon_{ph}^H \varepsilon_{ph}^E = 0.632. \quad (18.40)$$

The total aperture efficiency includes the taper factor, too:

$$\varepsilon_{ph}^P = \varepsilon_t \varepsilon_{ph}^H \varepsilon_{ph}^E = 0.81 \cdot 0.632 = 0.51. \quad (18.41)$$

Therefore, the best achievable directivity for a rectangular waveguide horn is about half that of a uniform rectangular aperture.

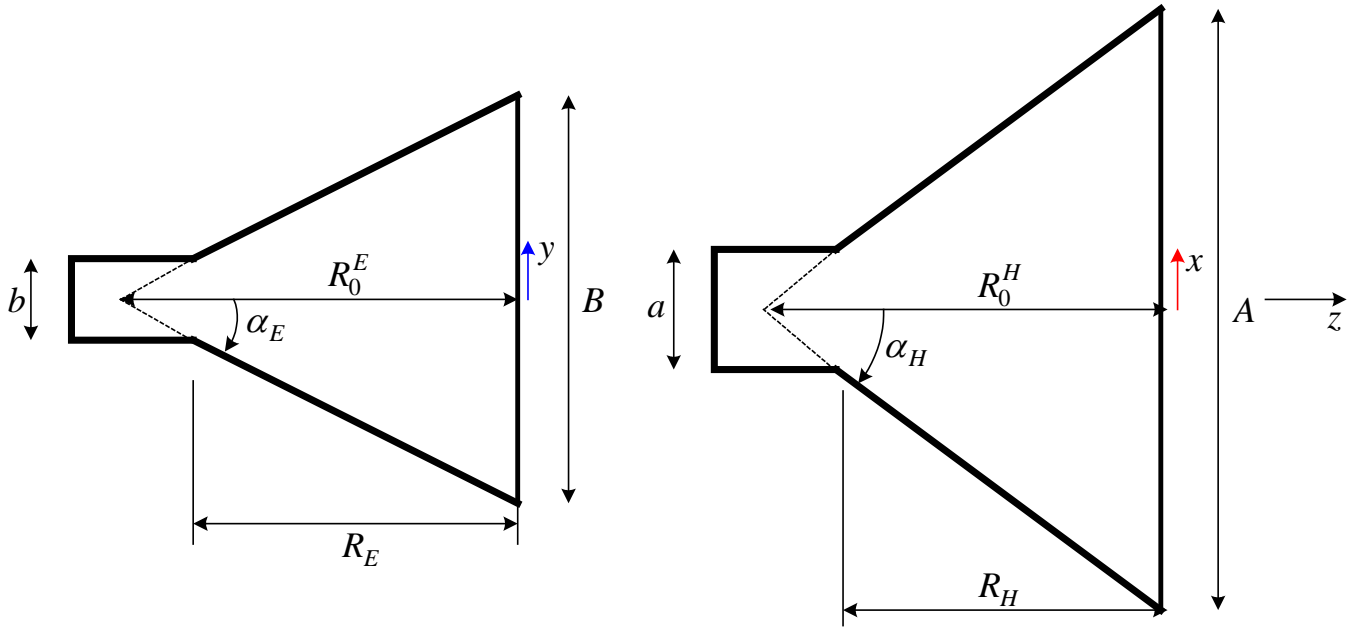
We reiterate that best accuracy is achieved if ε_{ph}^H and ε_{ph}^E are calculated numerically without using the second-order phase approximations in (18.7) and (18.28).

Optimum horn design

Usually, the optimum (from the point of view of maximum gain) design of a horn is desired because it results in the shortest axial length. The whole design can be actually reduced to the solution of a single fourth-order equation. For a horn to be realizable, the following must be true:

$$R_E = R_H = R_P. \quad (18.42)$$

The figures below summarize the notations used in describing the horn's geometry.



It can be shown that

$$\frac{R_0^H}{R_H} = \frac{A/2}{A/2 - a/2} = \frac{A}{A - a}, \quad (18.43)$$

$$\frac{R_0^E}{R_E} = \frac{B/2}{B/2 - b/2} = \frac{B}{B - b}. \quad (18.44)$$

The optimum-gain condition in the E -plane (18.37) is substituted in (18.44) to produce

$$B^2 - bB - 2\lambda R_E = 0. \quad (18.45)$$

There is only one physically meaningful solution to (18.45):

$$B = \frac{1}{2} \left(b + \sqrt{b^2 + 8\lambda R_E} \right). \quad (18.46)$$

Similarly, the maximum-gain condition for the H -plane of (18.24) together with (18.43) yields

$$R_H = \frac{A - a}{A} \left(\frac{A^2}{3\lambda} \right) = A \frac{(A - a)}{3\lambda}. \quad (18.47)$$

Since $R_E = R_H$ must be fulfilled, (18.47) is substituted in (18.46), which gives

$$B = \frac{1}{2} \left(b + \sqrt{b^2 + \frac{8A(A-a)}{3}} \right). \quad (18.48)$$

Substituting in the expression for the horn's gain,

$$G = \frac{4\pi}{\lambda^2} \varepsilon_{ap} AB, \quad (18.49)$$

gives the relation between A , the gain G , and the aperture efficiency ε_{ap} :

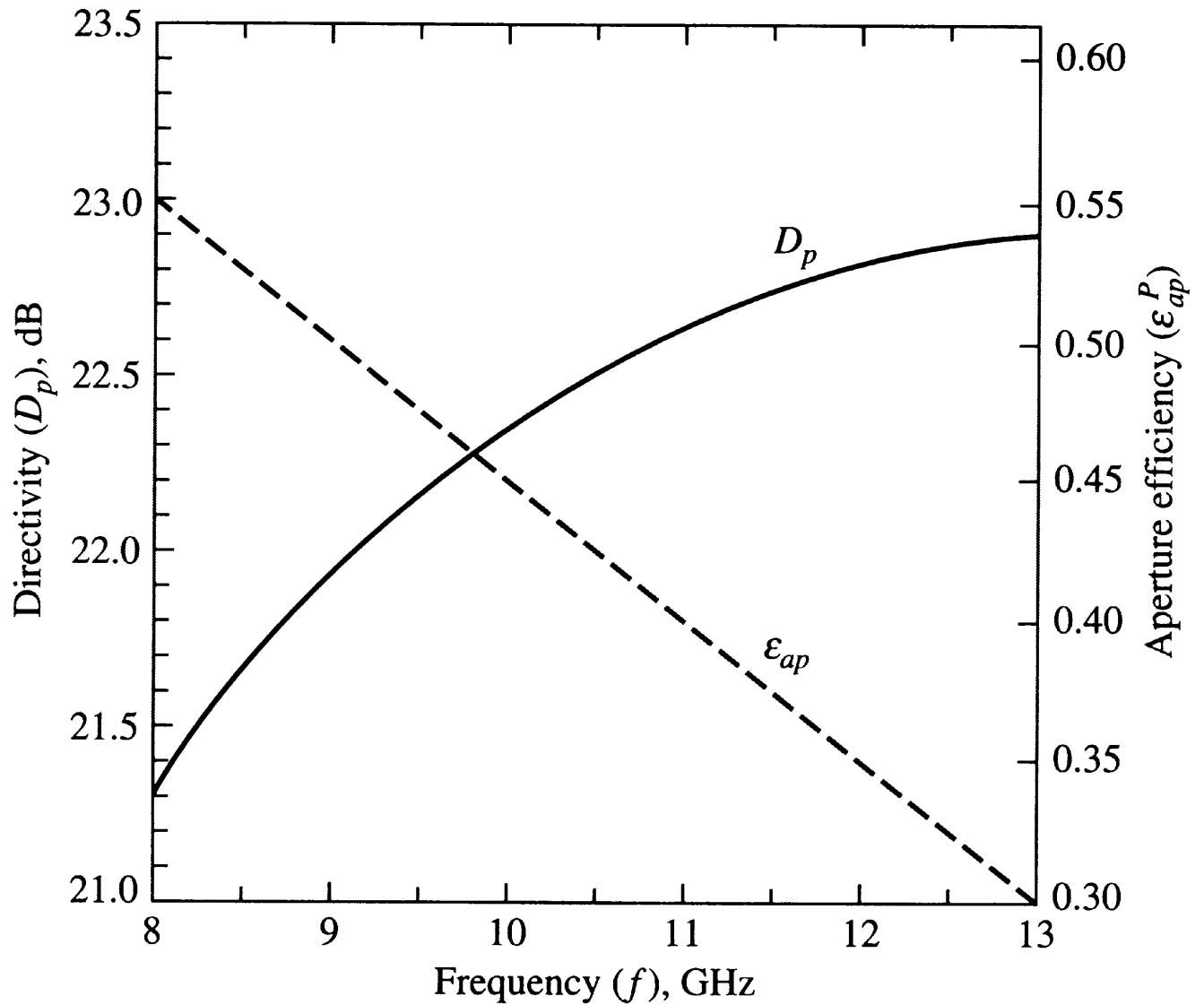
$$G = \frac{4\pi}{\lambda^2} \varepsilon_{ap} A \frac{1}{2} \left(b + \sqrt{b^2 + \frac{8A(a-a)}{3}} \right), \quad (18.50)$$

$$\Rightarrow A^4 - aA^3 + \frac{3bG\lambda^2}{8\pi\varepsilon_{ap}} A - \frac{3G^2\lambda^4}{32\pi^2\varepsilon_{ap}^2} = 0. \quad (18.51)$$

Equation (18.51) is the optimum pyramidal horn design equation. The optimum-gain value of $\varepsilon_{ap} = 0.51$ is usually used, which makes the equation a fourth-order polynomial equation in A . Its roots can be found analytically (which is not particularly easy) and numerically. In a numerical solution, the first guess is usually set at $A^{(0)} = 0.45\lambda\sqrt{G}$. Once A is found, B can be computed from (18.48) and $R_E = R_H$ is computed from (18.47).

Sometimes, an optimal horn is desired for a known axial length R_0 . In this case, there is no need for nonlinear-equation solution. The design procedure follows the steps: (a) find A from (18.24), (b) find B from (18.37), and (c) calculate the gain G using (18.49) where $\varepsilon_{ap} = 0.51$.

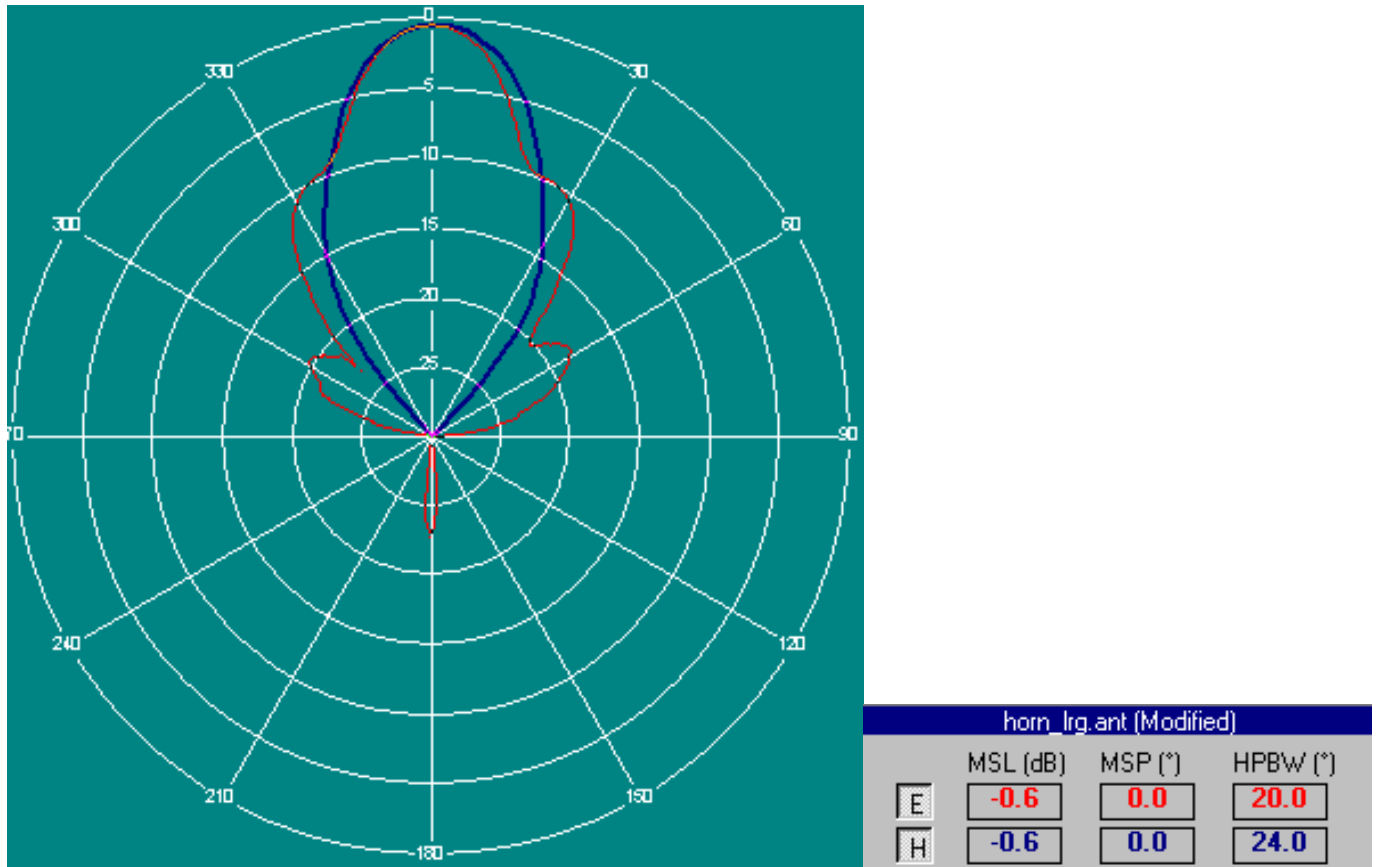
Horn antennas operate well over a bandwidth of 50%. However, gain performance is optimal only at a given frequency. To understand better the frequency dependence of the directivity and the aperture efficiency, the plot of these curves for an X-band (8.2 GHz to 12.4 GHz) horn fed by WR90 waveguide is given below ($a = 0.9$ in. = 2.286 cm and $b = 0.4$ in. = 1.016 cm).



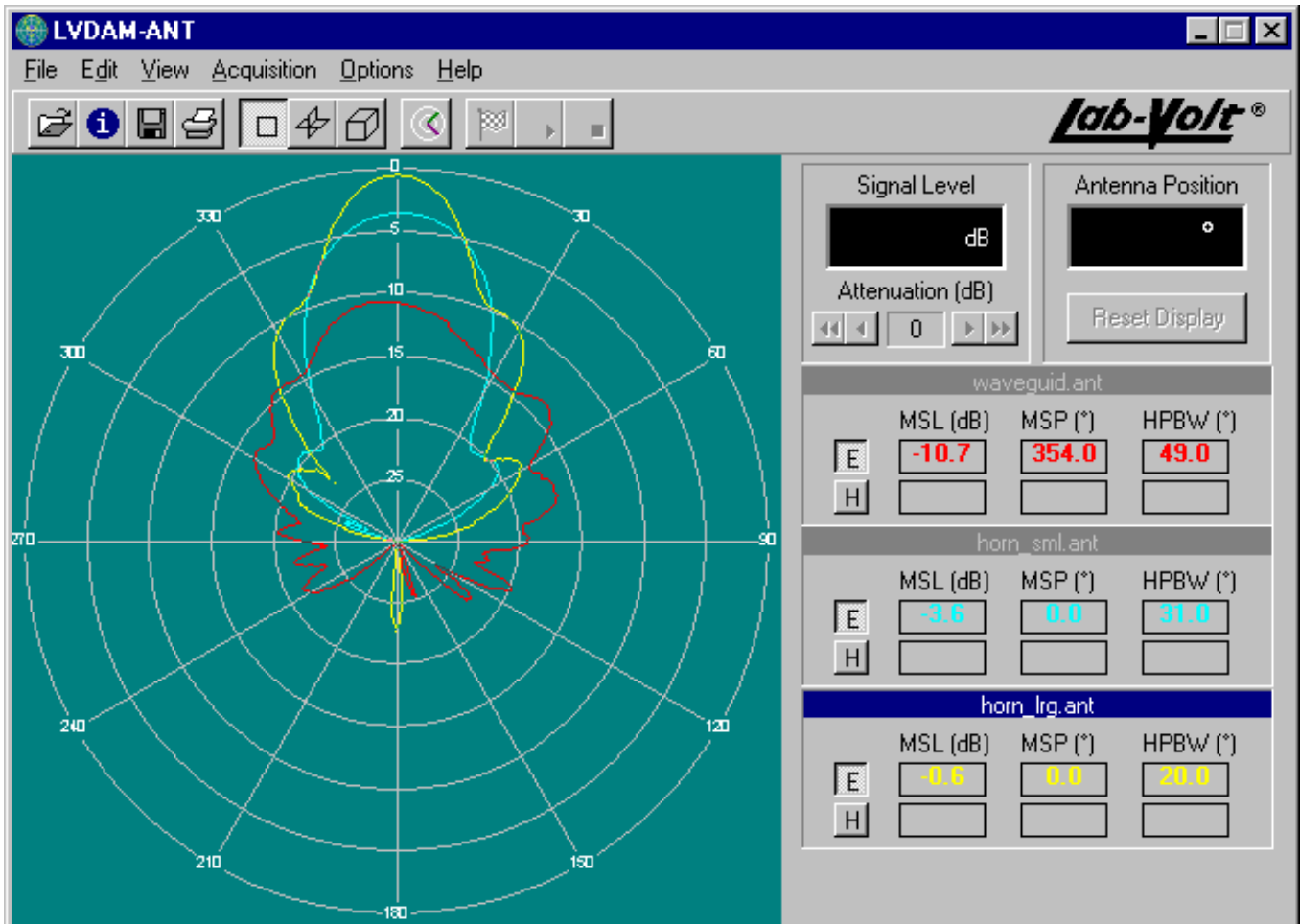
[Stutzman&Thiele, *Antenna Theory and Design*]

The gain increases with frequency, which is typical for aperture antennas. However, the curve shows saturation at higher frequencies. This is due to the decrease of the aperture efficiency, which is a result of an increased phase difference in the field distribution at the aperture.

The pattern of a “large” pyramidal horn ($f = 10.525$ GHz, feed is waveguide WR90):



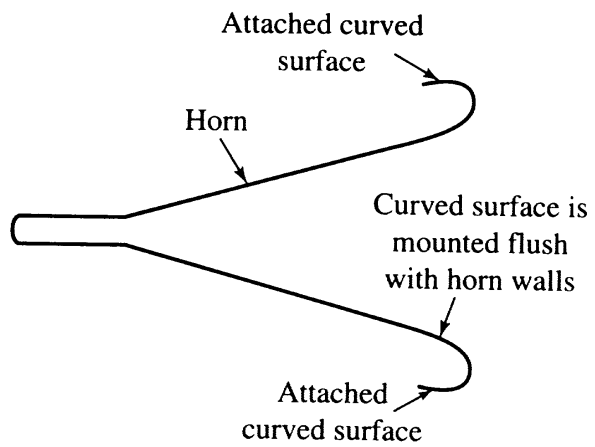
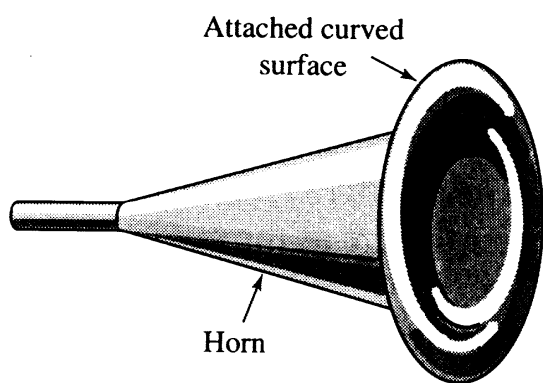
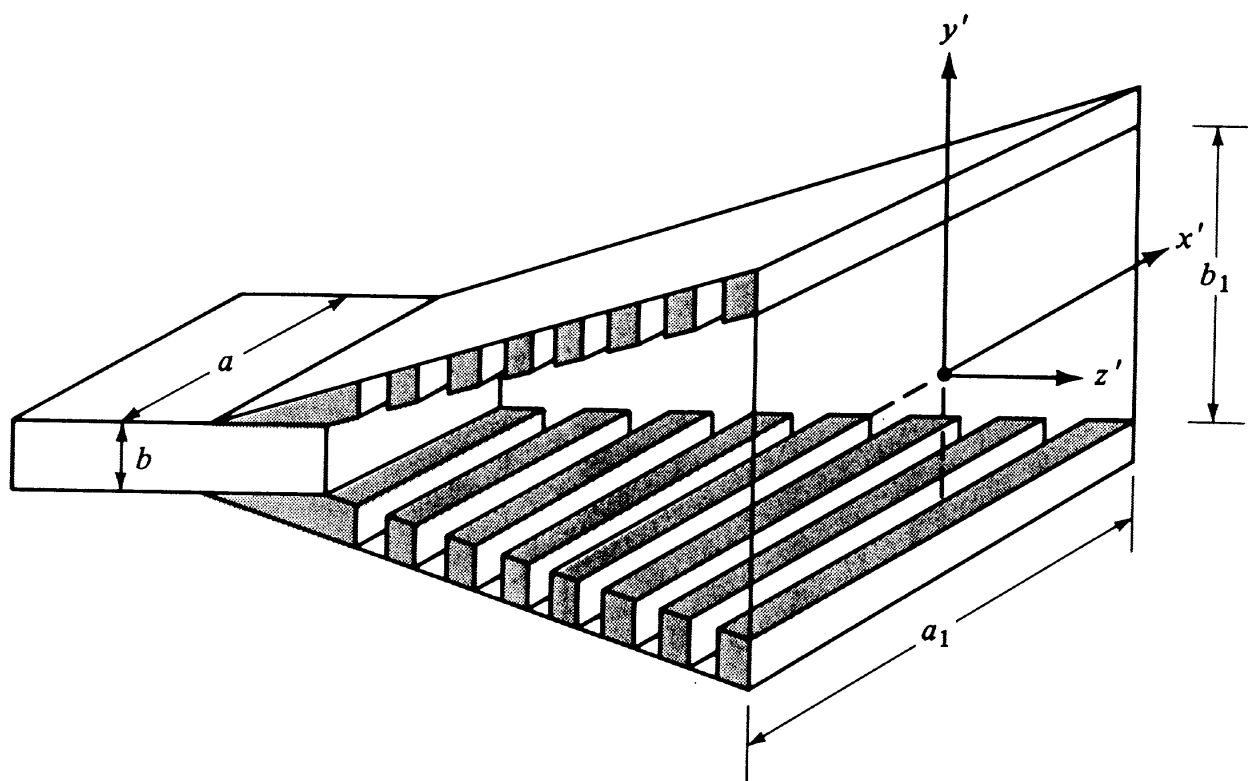
Comparison of the E -plane patterns of a waveguide open end, “small” pyramidal horn and “large” pyramidal horn:



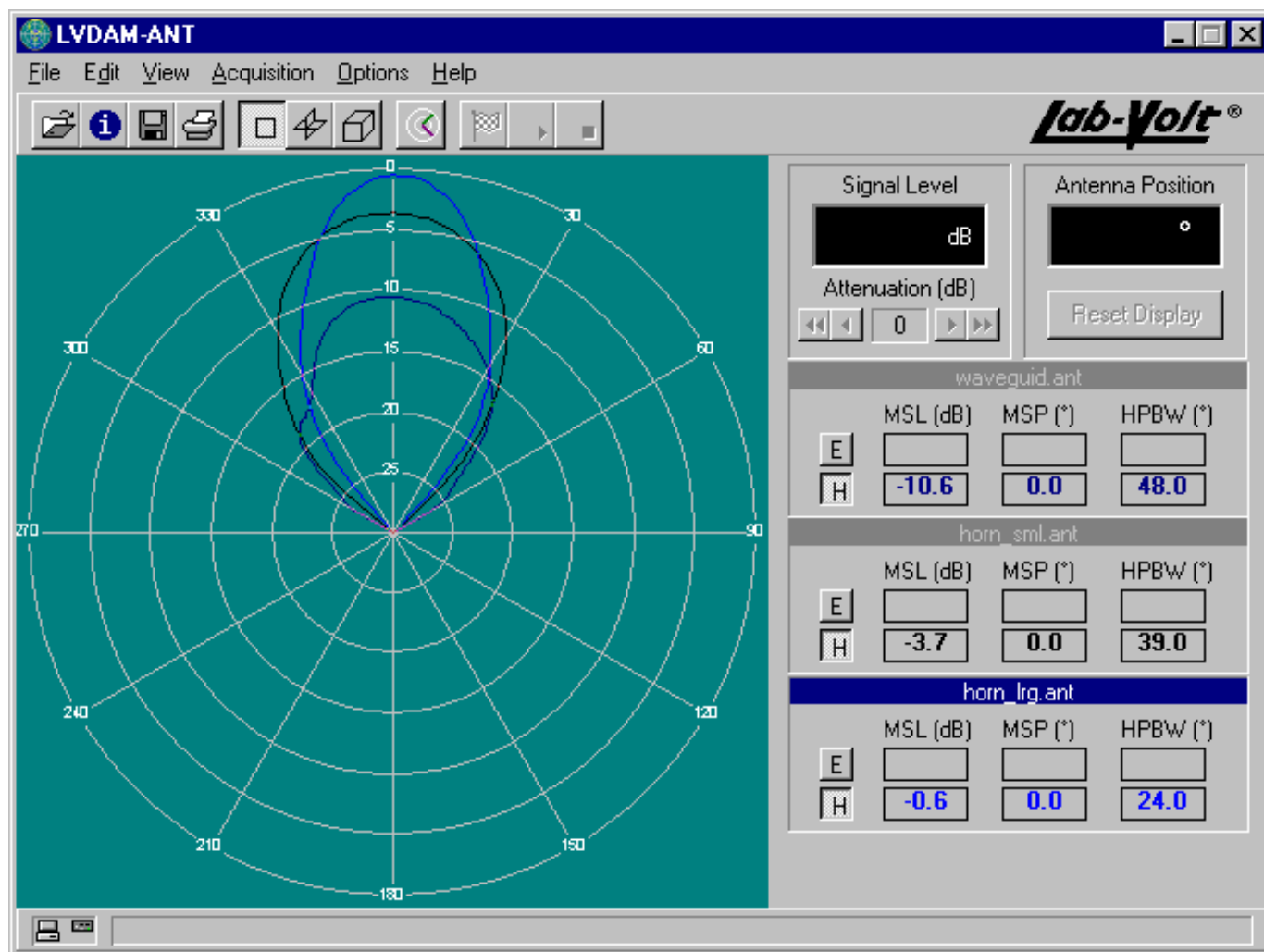
Note the multiple side lobes and the significant back lobe. They are due to diffraction at the horn edges, which are perpendicular to the \mathbf{E} field. To reduce edge diffraction, enhancements are proposed for horn antennas such as

- corrugated horns
- aperture-matched horns

The corrugated horns achieve tapering of the \mathbf{E} field in the vertical direction, thus, reducing the side-lobes and the diffraction from the top and bottom edges. The overall main beam becomes smooth and nearly rotationally symmetrical (esp. for $A \approx B$). This is important when the horn is used as a feed to a reflector antenna.



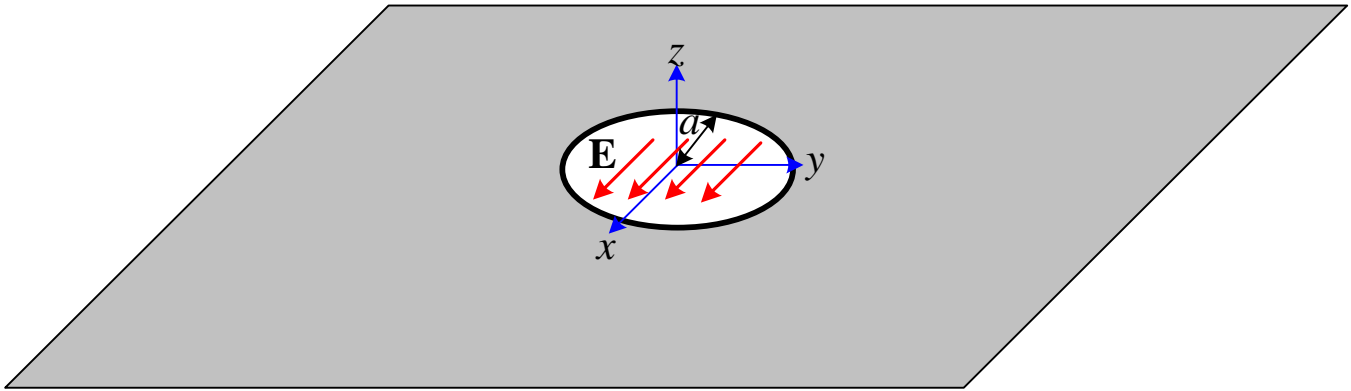
Comparison of the H -plane patterns of a waveguide open end, “small” pyramidal horn and “large” pyramidal horn:



2 Circular apertures

2.1 A uniform circular aperture

The uniform circular aperture is approximated by a circular opening in a ground plane illuminated by a uniform plane wave normally incident from behind.



The field distribution is described as

$$\mathbf{E}_a = \hat{\mathbf{x}}E_0, \quad \rho' \leq a. \quad (18.52)$$

The radiation integral is

$$I_x^E = E_0 \iint_{S_a} e^{j\beta \hat{\mathbf{r}} \cdot \mathbf{r}'} ds'. \quad (18.53)$$

The integration point is at

$$\mathbf{r}' = \hat{\mathbf{x}}\rho' \cos \varphi' + \hat{\mathbf{y}}\rho' \sin \varphi'. \quad (18.54)$$

In (18.54), cylindrical coordinates are used, therefore,

$$\hat{\mathbf{r}} \cdot \mathbf{r}' = \rho' \sin \theta (\cos \varphi \cos \varphi' + \sin \varphi \sin \varphi') = \rho' \sin \theta \cos(\varphi - \varphi'). \quad (18.55)$$

Hence, (18.53) becomes

$$I_x^E = E_0 \int_0^a \left[\int_0^{2\pi} e^{j\beta \rho' \sin \theta \cos(\varphi - \varphi')} d\varphi' \right] \rho' d\rho' = 2\pi E_0 \int_0^a \rho' J_0(\beta \rho' \sin \theta) d\rho'. \quad (18.56)$$

Here, J_0 is the Bessel function of the first kind of order zero. Applying the identity

$$\int xJ_0(x)dx = xJ_1(x) \quad (18.57)$$

to (18.56) leads to

$$I_x^E = 2\pi E_0 \frac{a}{\beta \sin \theta} J_1(\beta a \sin \theta). \quad (18.58)$$

Note that in this case the equivalent magnetic current formulation of the equivalence principle is used [see Lecture 17]. The far field is obtained as

$$\begin{aligned} \mathbf{E} &= \left(\hat{\boldsymbol{\theta}} \cos \varphi - \hat{\boldsymbol{\phi}} \cos \theta \sin \varphi \right) j\beta \frac{e^{j\beta r}}{2\pi r} I_x^E = \\ &= \left(\hat{\boldsymbol{\theta}} \cos \varphi - \hat{\boldsymbol{\phi}} \cos \theta \sin \varphi \right) j\beta E_0 \pi a^2 \frac{e^{j\beta r}}{2\pi r} \frac{2J_1(\beta a \sin \theta)}{\beta a \sin \theta}. \end{aligned} \quad (18.59)$$

Principal-plane patterns

$$\mathbf{E}\text{-plane } (\varphi = 0): E_\theta(\theta) = \frac{2J_1(\beta a \sin \theta)}{\beta a \sin \theta} \quad (18.60)$$

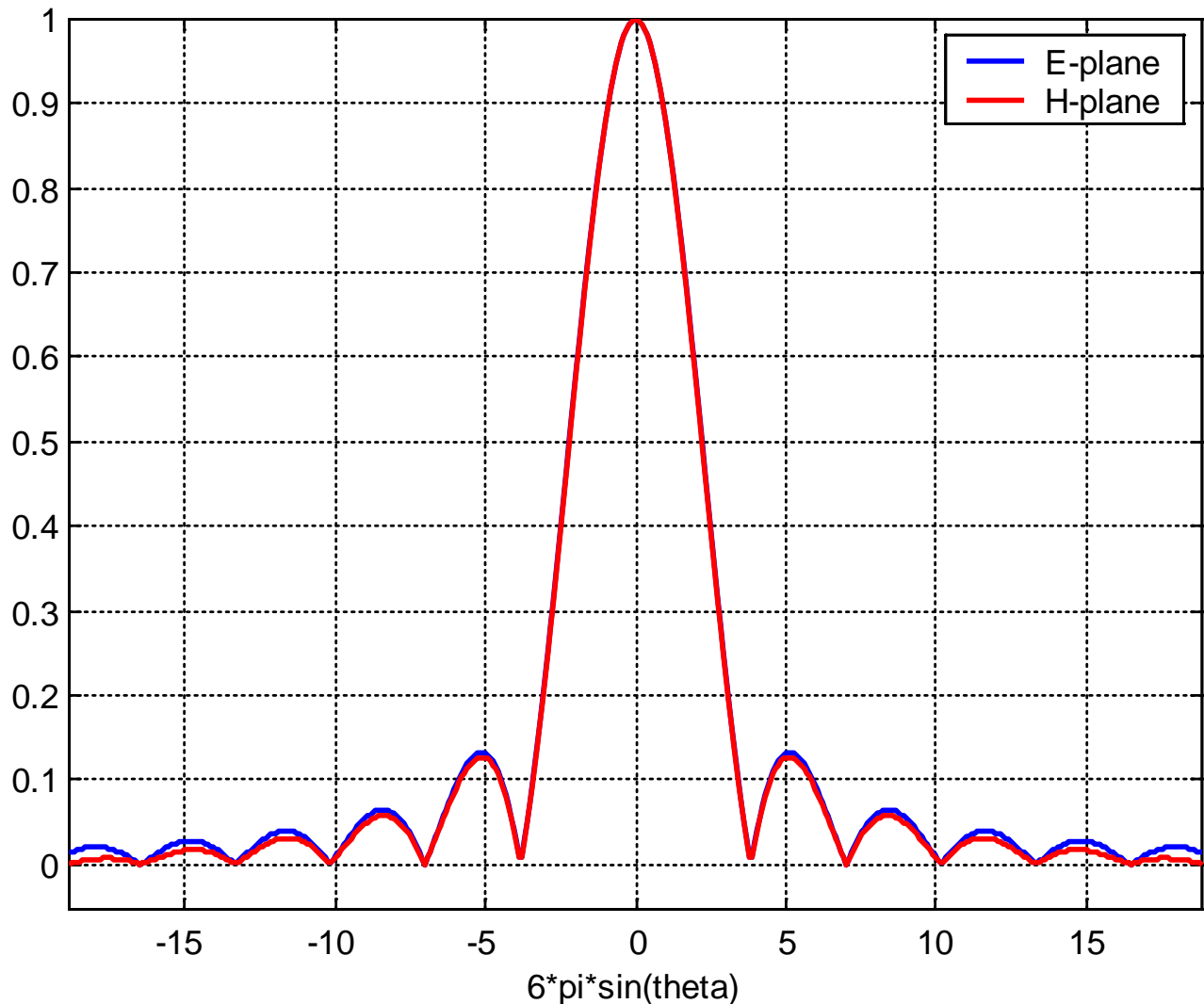
$$\mathbf{H}\text{-plane } (\varphi = 90^\circ): E_\varphi(\theta) = \cos \theta \cdot \frac{2J_1(\beta a \sin \theta)}{\beta a \sin \theta} \quad (18.61)$$

The 3-D amplitude pattern:

$$\bar{E}(\theta, \varphi) = \sqrt{1 - \sin^2 \theta \sin^2 \varphi} \cdot \underbrace{\frac{2J_1(\beta a \sin \theta)}{\beta a \sin \theta}}_{f(\theta)} \quad (18.62)$$

The larger the aperture, the less significant the $\cos \theta$ factor is in (18.61) because the main beam in the $\theta = 0$ direction is very narrow and in this small solid angle $\cos \theta \approx 1$. Thus, the 3-D pattern of a large circular aperture features a fairly symmetrical beam.

Example plot of the principal-plane patterns for $a = 3\lambda$:



The half-power angle for the $f(\theta)$ factor is obtained at $\beta a \sin \theta = 1.6$. So, the HPBW for large apertures ($a \gg \lambda$) is given by

$$HPBW = 2\theta_{1/2} \approx 2 \arcsin\left(\frac{1.6}{\beta a}\right) \approx 2 \frac{1.6}{\beta a} = 58.4 \frac{\lambda}{2a}, \text{ deg.} \quad (18.63)$$

For example, if the diameter of the aperture is $2a = 10\lambda$, then $HPBW = 5.84^\circ$.

The side-lobe level of any uniform circular aperture is 0.1332 (-17.5 dB).

Any *uniform* aperture has unity taper aperture efficiency, and its directivity can be found directly in terms of its physical area,

$$D_u = \frac{4\pi}{\lambda^2} A_p = \frac{4\pi}{\lambda^2} \pi a^2. \quad (18.64)$$

2.2 Tapered circular apertures

Many practical circular aperture antennas can be approximated as radially symmetric apertures with field amplitude distribution, which is tapered from the center toward the aperture edge. Then, the radiation integral (18.56) has a more general form:

$$I_x^E = 2\pi \int_0^a E_0(\rho') \rho' J_0(\beta \rho' \sin \theta) d\rho'. \quad (18.65)$$

In (18.65), we still assume that the field has axial symmetry, i.e., it does not depend on φ' . Often used approximation is the parabolic taper of order n :

$$E_a(\rho') = E_0 \left[1 - \left(\frac{\rho'}{a} \right)^2 \right]^n \quad (18.66)$$

where E_0 is a constant. This is substituted in (18.65) to calculate the respective component of the radiation integral:

$$I_x^E(\theta) = 2\pi E_0 \int_0^a \left[1 - \left(\frac{\rho'}{a} \right)^2 \right]^n \rho' J_0(\beta \rho' \sin \theta) d\rho'. \quad (18.67)$$

The following relation is used to solve (18.67):

$$\int_0^1 (1-x^2)^n x J_0(bx) dx = \frac{2^n n!}{b^{n+1}} J_{n+1}(b). \quad (18.68)$$

In our case, $x = \rho' / a$ and $b = \beta a \sin \theta$. Then, $I_x^E(\theta)$ reduces to

$$I_x^E(\theta) = E_0 \frac{\pi a^2}{n+1} f(\theta, n), \quad (18.69)$$

where

$$f(\theta, n) = \frac{2^{n+1} (n+1)! J_{n+1}(\beta a \sin \theta)}{(\beta a \sin \theta)^{n+1}} \quad (18.70)$$

is the normalized pattern (neglecting the angular factors such as $\cos \varphi$ and $\cos \theta \sin \varphi$).

The aperture taper efficiency is calculated to be

$$\varepsilon_t = \frac{\left[C + \frac{1-C}{n+1} \right]^2}{C^2 + \frac{2C(1-C)}{n+1} + \frac{(1-C)^2}{2n+1}} \quad (18.71)$$

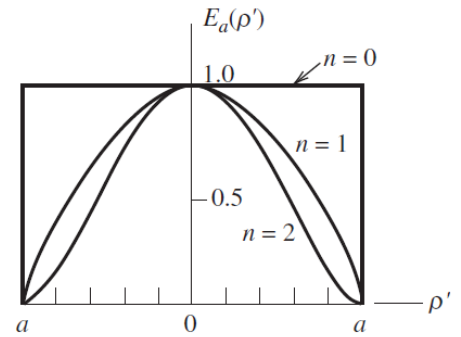
Here, C denotes the *pedestal height*. The pedestal height is the edge field illumination relative to the illumination at the center.

The properties of several common tapers are given in the tables below. The parabolic taper ($n = 1$) provides lower side lobes in comparison with the uniform distribution ($n = 0$) but it has a broader main beam. There is always a trade-off between low side-lobe levels and high directivity (small HPBW). More or less optimal solution is provided by the parabolic-on-pedestal aperture distribution. Moreover, this distribution approximates very closely the real case of circular reflector antennas, where the feed antenna pattern is intercepted by the reflector only out to the reflector rim.

a. Parabolic taper

$$E_a(\rho') = \left[1 - \left(\frac{\rho'}{a} \right)^2 \right]^n$$

$$f(\theta, n) = \frac{2^{n+1}(n+1)!J_{n+1}(\beta a \sin \theta)}{(\beta a \sin \theta)^{n+1}}$$



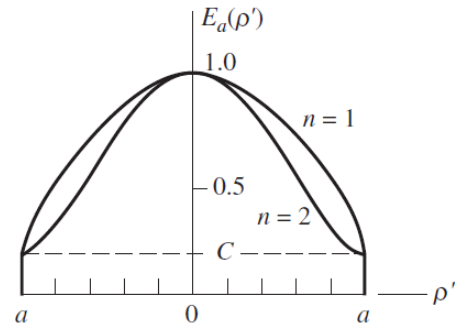
n	HP (rad)	Side Lobe Level (dB)	ε_t	Normalized Pattern $f(\theta, n)$	Distribution
0	$1.02 \frac{\lambda}{2a}$	-17.6	1.00	$\frac{2J_1(\beta a \sin \theta)}{\beta a \sin \theta}$	Uniform
1	$1.27 \frac{\lambda}{2a}$	-24.6	0.75	$\frac{8J_2(\beta a \sin \theta)}{(\beta a \sin \theta)^2}$	Parabolic
2	$1.47 \frac{\lambda}{2a}$	-30.6	0.55	$\frac{48J_3(\beta a \sin \theta)}{(\beta a \sin \theta)^3}$	Parabolic squared

[Stutzman&Thiele]

b. Parabolic taper on a pedestal

$$E_a(\rho') = C + (1 - C) \left[1 - \left(\frac{\rho'}{a} \right)^2 \right]^n$$

$$f(\theta, n, C) = \frac{Cf(\theta, n = 0) + \frac{1 - C}{n + 1}f(\theta, n)}{C + \frac{1 - C}{n + 1}}$$



Edge Illumination		$n = 1$			$n = 2$		
C_{dB}	C	HP (rad)	Side Lobe Level (dB)	ϵ_t	HP (rad)	Side Lobe Level (dB)	ϵ_t
-8	0.398	$1.12 \frac{\lambda}{2a}$	-21.5	0.942	$1.14 \frac{\lambda}{2a}$	-24.7	0.918
-10	0.316	$1.14 \frac{\lambda}{2a}$	-22.3	0.917	$1.17 \frac{\lambda}{2a}$	-27.0	0.877
-12	0.251	$1.16 \frac{\lambda}{2a}$	-22.9	0.893	$1.20 \frac{\lambda}{2a}$	-29.5	0.834
-14	0.200	$1.17 \frac{\lambda}{2a}$	-23.4	0.871	$1.23 \frac{\lambda}{2a}$	-31.7	0.792
-16	0.158	$1.19 \frac{\lambda}{2a}$	-23.8	0.850	$1.26 \frac{\lambda}{2a}$	-33.5	0.754
-18	0.126	$1.20 \frac{\lambda}{2a}$	-24.1	0.833	$1.29 \frac{\lambda}{2a}$	-34.5	0.719
-20	0.100	$1.21 \frac{\lambda}{2a}$	-24.3	0.817	$1.32 \frac{\lambda}{2a}$	-34.7	0.690

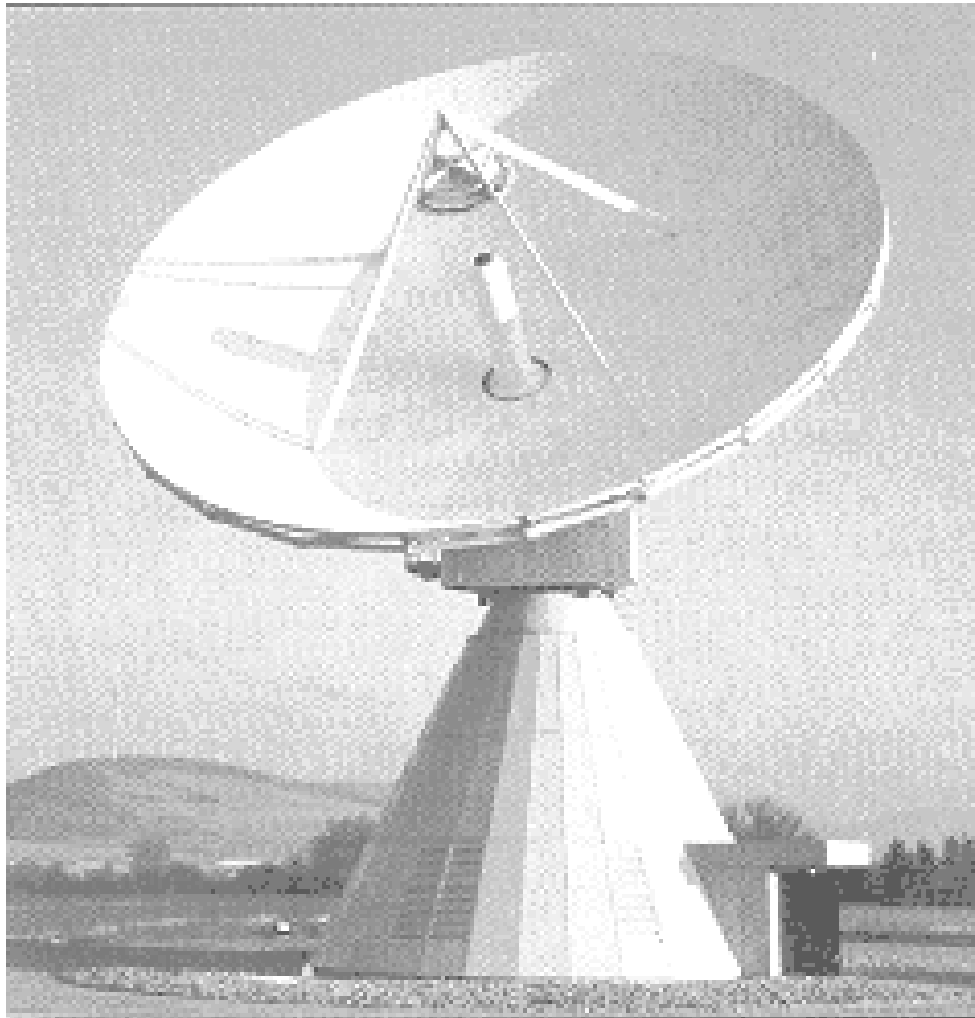
[Stutzman&Thiele]

LECTURE 19: Reflector Antennas

1. Introduction

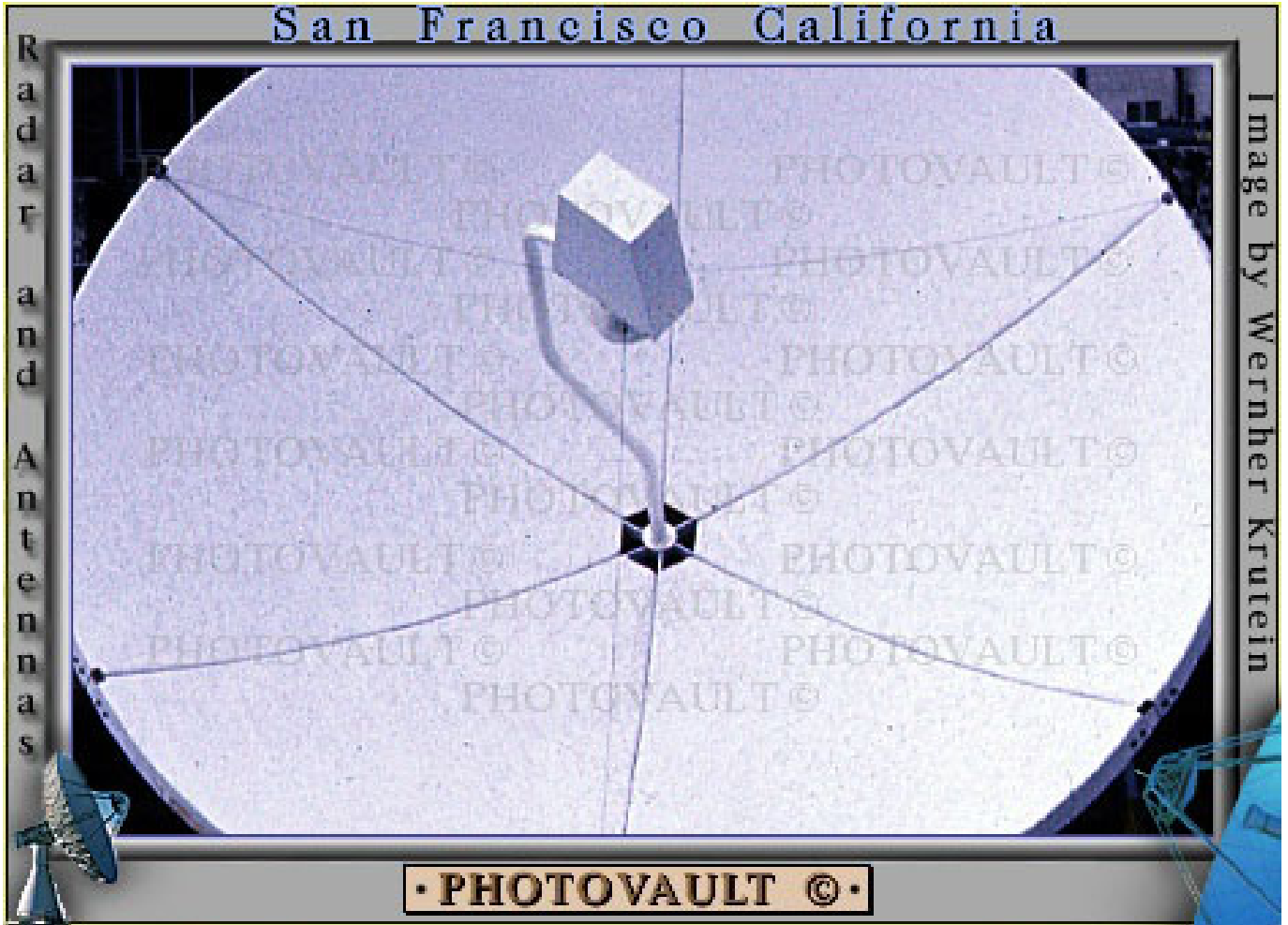
High-gain antennas are required for long-distance radio communications (radio-relay links and satellite links), high-resolution radars, radio-astronomy, etc. Reflector systems are probably the most widely used high-gain antennas. They can easily achieve gains of above 30 dB for microwave and higher frequencies. Reflector antennas operate on principles known long ago from geometrical optics (GO). The first RF reflector system was made by Hertz back in 1888 (a cylindrical reflector fed by a dipole). However, the art of accurately designing such antenna systems was developed mainly during the days of WW2 when numerous radar applications evolved.

18.3 M INTELSAT EARTH STATION (ANT BOSCH TELECOM), DUAL REFLECTOR

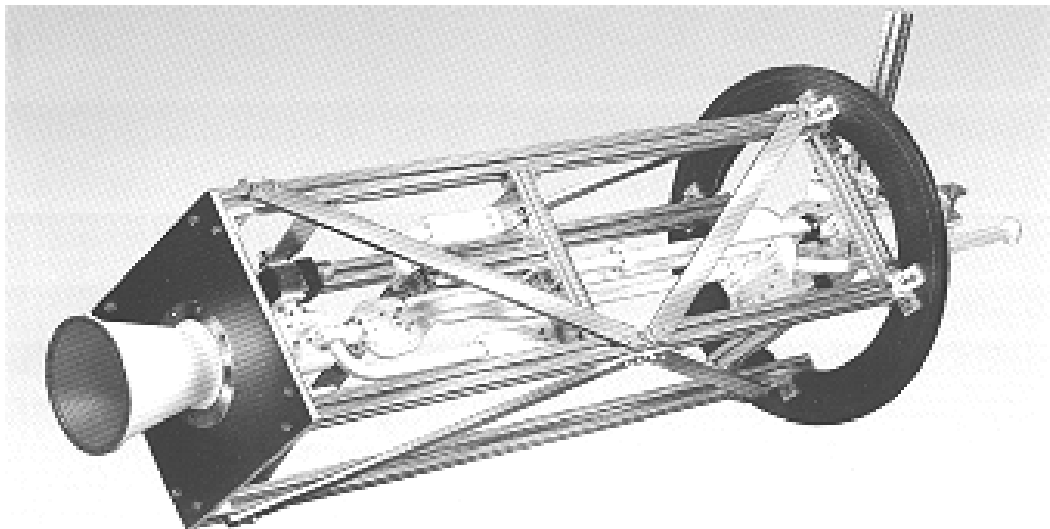


FEED-HORN IS IN FOCAL POINT

San Francisco California

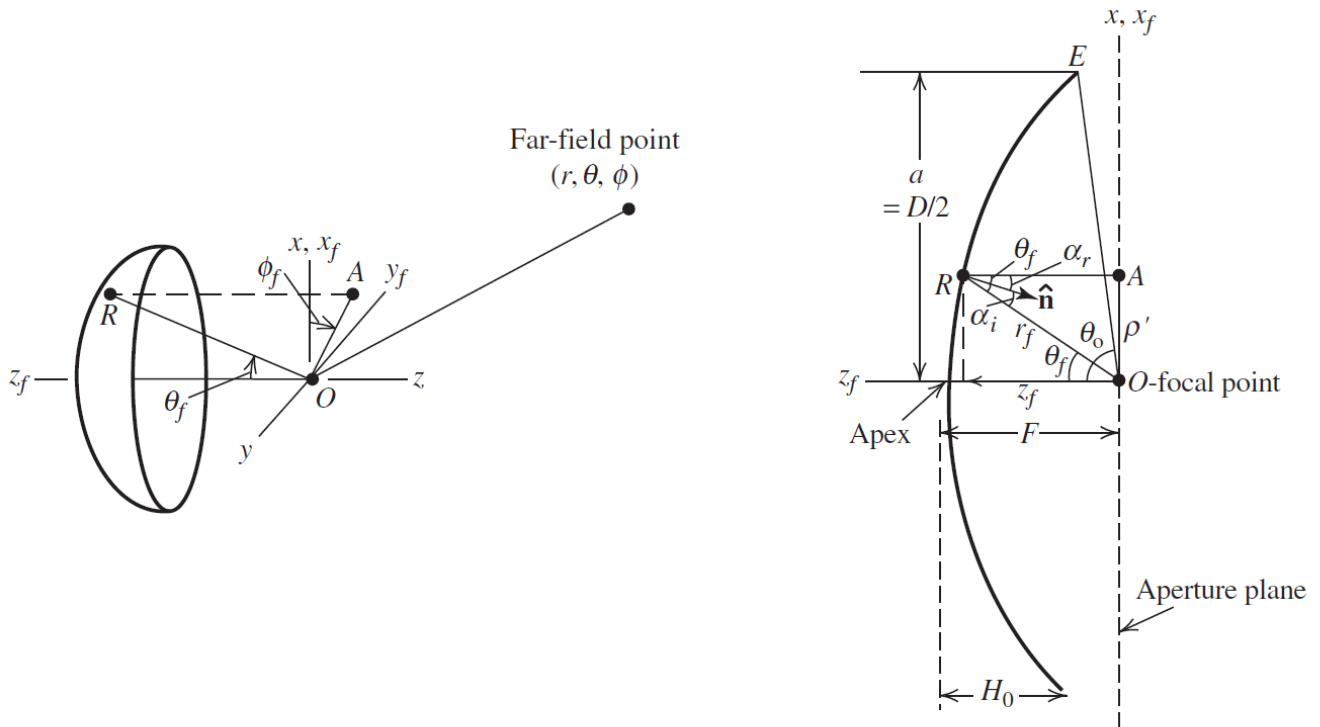


CONICAL HORN PRIMARY FEED



The simplest reflector antenna consists of two components: a reflecting surface and a much smaller feed antenna at the reflector's focal point. Constructions that are more complex involve a secondary reflector (a subreflector) at the focal point, which is illuminated by a primary feed. These are called dual-reflector antennas. The most common main reflector is the parabolic one. Other common reflectors are: cylindrical, corner, and spherical.

2. Principles of parabolic reflectors



(a) Parabolic reflector and coordinate system.

(b) Cross section of the reflector in the xz -plane.

[Stutzman&Thiele]

A paraboloidal surface is described by the equation (see plot *b*)

$$\rho'^2 = 4F(F - z_f), \quad \rho' \leq a. \quad (19.1)$$

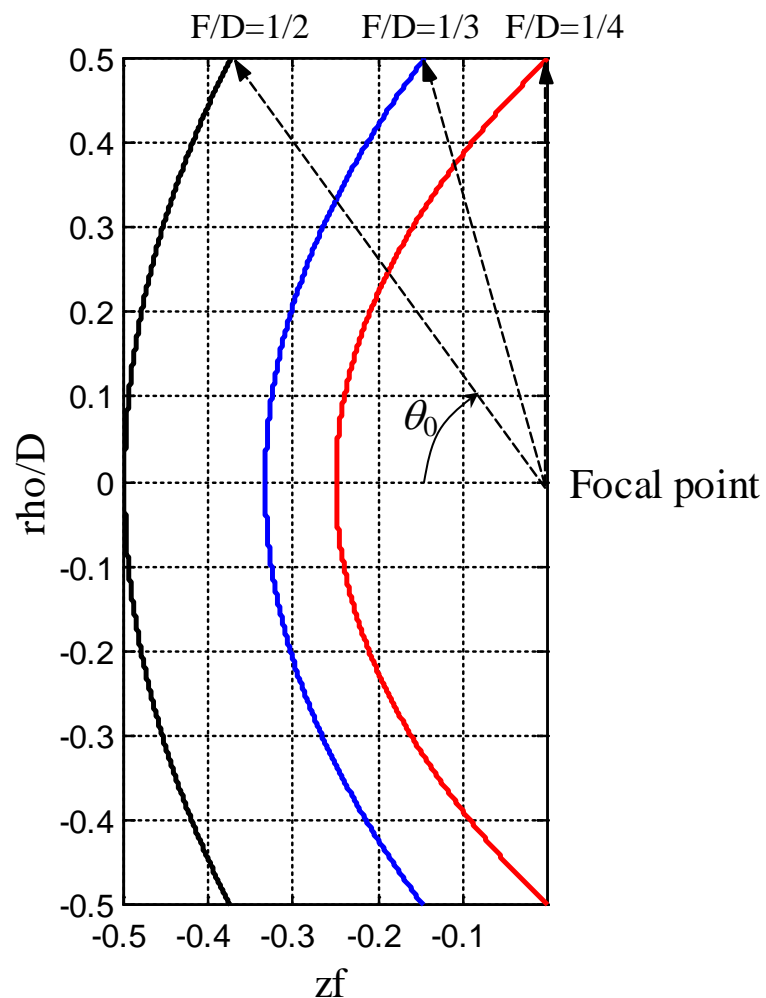
Here, ρ' is the distance from a point A to the focal point O , where A is the projection of the point R on the reflector surface onto the axis-orthogonal plane (the aperture plane) at the focal point. For a given displacement ρ' from the axis of the reflector, the point R on the reflector surface is a distance r_f away from the focal point O . The position of R can be defined either by (ρ', z_f) , which is a rectangular pair of coordinates, or by (r_f, θ_f) , which is a polar pair of coordinates. A relation between (r_f, θ_f) and F is readily found from (19.1):

$$r_f = \frac{2F}{1 + \cos \theta_f} = \frac{F}{\cos^2(\theta_f / 2)}. \quad (19.2)$$

This is the equation of the paraboloidal surface in polar coordinates. Other relations to be used later are:

$$\rho' = r_f \sin \theta_f = \frac{2F \sin \theta_f}{1 + \cos \theta_f} = 2F \tan \left(\frac{\theta_f}{2} \right). \quad (19.3)$$

The axisymmetric (rotationally symmetric) paraboloidal reflector is entirely defined by the respective parabolic line, i.e., by two basic parameters: the diameter D and the focal length F (see plot *b*). Often, the parabola is specified in terms of D and the ratio F/D . When F/D approaches infinity, the reflector becomes flat. Some parabolic curves are shown below. When $F/D = 0.25$, the focal point lies in the plane passing through the reflector's rim.



The angle from the feed (focal) point to the reflector's rim is related to F/D as

$$\theta_0 = 2 \arctan \left[\frac{1}{4(F/D)} \right]. \quad (19.4)$$

The focal distance F of a given reflector can be calculated after measuring its diameter D and its height H_0 (see plot *b*):

$$F = \frac{D^2}{16H_0}. \quad (19.5)$$

Eq. (19.5) is found by solving (19.1) with $\rho' = D/2$ and $z_f = F - H_0$. For example, if $F/D = 1/4$, then $H_0 = D/4 \Rightarrow H_0 = F$, i.e., the focal point is on the reflector's rim plane.

The reflector design problem aims at matching the feed antenna pattern to the reflector. Usually, the feed pattern must be at about a -10 dB level in the direction of the rim, i.e. $\bar{F}_f(\theta = \theta_0) = -10$ dB (0.316 of the normalized amplitude pattern).

The geometry of the paraboloidal reflector has two valuable features:

- All rays leaving the focal point O are collimated along the reflector's axis after reflection.
- All overall ray path lengths (from the focal point to the reflector and on to the aperture plane) are the same and equal to $2F$.

The above properties are proven by the GO methods, therefore, they are true only if the following conditions hold:

- The radius of the curvature of the reflector is large compared to the wavelength and the local region around each reflection point can be treated as planar.
- The radius of the curvature of the incoming wave from the feed is large and can be treated locally at the reflection point as a plane wave.
- The reflector is a perfect conductor, i.e., $\Gamma = -1$.

The collimating property of the parabolic reflector is easily established after finding the unit normal of the parabola,

$$\hat{\mathbf{n}} = \frac{\nabla C_p}{|\nabla C_p|}. \quad (19.6)$$

Here,

$$C_p = F - r_f \cos^2(\theta_f / 2) = 0 \quad (19.7)$$

is the parabolic curve equation [see (19.2)]. After applying the ∇ operator in spherical coordinates, ∇C_p is obtained as

$$\nabla C_p = -\hat{\mathbf{r}}_f \cos^2\left(\frac{\theta_f}{2}\right) + \hat{\boldsymbol{\theta}}_f \cos\left(\frac{\theta_f}{2}\right) \cdot \sin\left(\frac{\theta_f}{2}\right), \quad (19.8)$$

and, therefore,

$$\hat{\mathbf{n}} = -\hat{\mathbf{r}}_f \cos\frac{\theta_f}{2} + \hat{\boldsymbol{\theta}}_f \sin\frac{\theta_f}{2}. \quad (19.9)$$

The angle between $\hat{\mathbf{n}}$ and the incident ray is

$$\cos \alpha_i = -\hat{\mathbf{r}}_f \cdot \hat{\mathbf{n}} = \cos\left(\frac{\theta_f}{2}\right). \quad (19.10)$$

According to Snell's law, $\alpha_i = \alpha_r$. It is easy to show that this is fulfilled only if the ray is reflected in the z -direction:

$$\begin{aligned} \cos \alpha_r = \hat{\mathbf{z}} \cdot \hat{\mathbf{n}} &= \underbrace{(-\hat{\mathbf{r}}_f \cos \theta_f + \hat{\boldsymbol{\theta}}_f \sin \theta_f)}_{\hat{\mathbf{z}}} \cdot \underbrace{\left[-\hat{\mathbf{r}}_f \cos\left(\frac{\theta_f}{2}\right) + \hat{\boldsymbol{\theta}}_f \sin\left(\frac{\theta_f}{2}\right)\right]}_{\hat{\mathbf{n}}} = \\ &= \cos \theta_f \cdot \cos\left(\frac{\theta_f}{2}\right) + \sin \theta_f \cdot \sin\left(\frac{\theta_f}{2}\right) = \cos\left(\frac{\theta_f}{2}\right). \end{aligned} \quad (19.11)$$

Thus, we proved that for any angle of incidence θ_f the reflected wave is z -directed.

The equal-path-length property follows from (19.2). The total path-length L for a ray reflected at the point R is

$$L = \overline{OR} + \overline{RA} = r_f + r_f \cos \theta_f = r_f (1 + \cos \theta_f) = 2F. \quad (19.12)$$

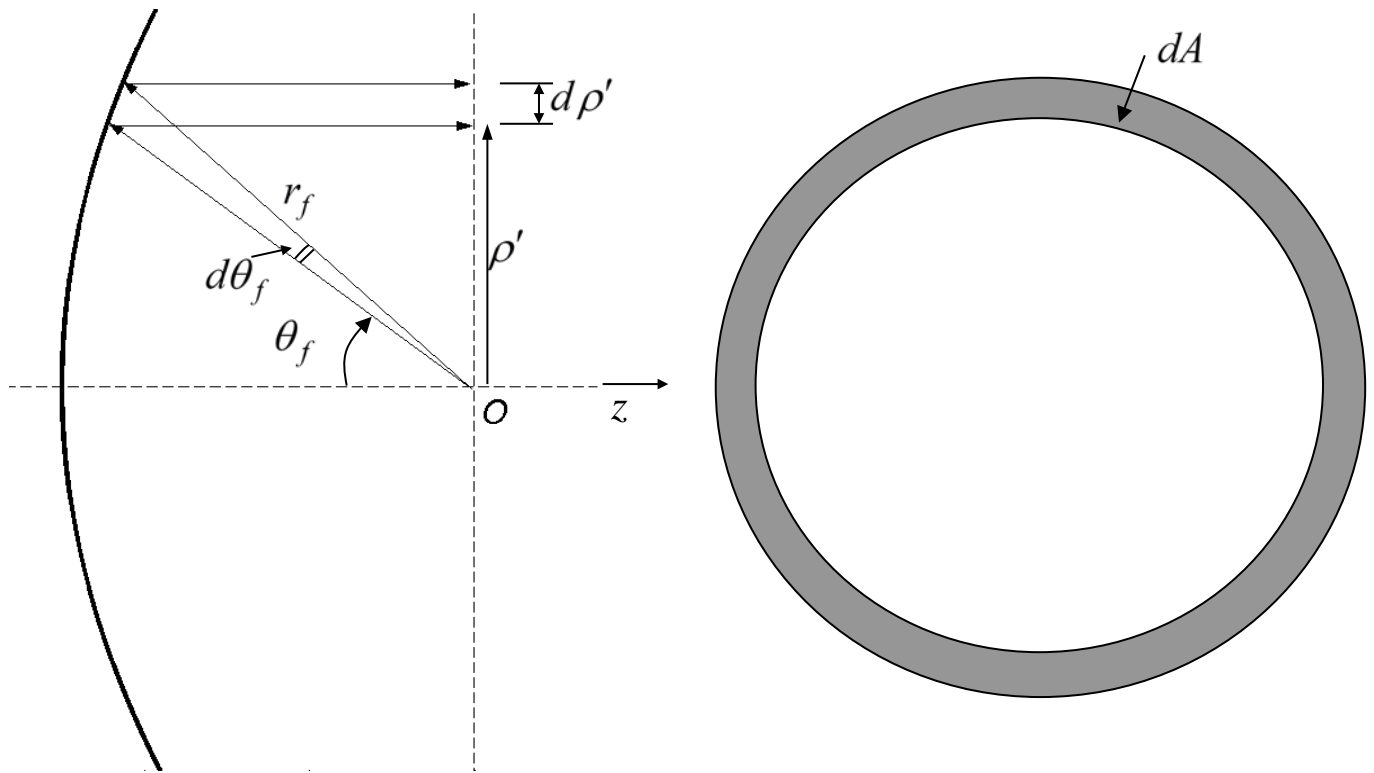
Notice that L is a constant equal to $2F$ regardless of the angle of incidence.

3. Aperture distribution analysis via GO (aperture integration)

There are two basic techniques for the analysis of the radiation characteristics of reflectors. One is called the *current distribution method*, which is a physical optics (PO) approximation. It assumes that the incident field from the feed is known, and that it excites surface currents on the reflector's surface as $\mathbf{J}_s = 2\hat{\mathbf{n}} \times \mathbf{H}^i$. This current density is then integrated to yield the far-zone field. It is obvious that the PO method assumes that the reflector has a perfectly conducting surface and makes use of image theory. Besides, it assumes that the incident wave coming from the primary feed is a locally plane far-zone field.

With the *aperture distribution method*, the field is first found over a plane, which is normal to the reflector's axis, and lies at its focal point (the *antenna aperture*). GO (ray tracing) is used to do that. Equivalent sources are formed over the aperture plane. It is assumed that the equivalent sources are zero outside the reflector's aperture. We first consider this method.

The field distribution at the aperture of the reflector antenna is necessary in order to calculate the far-field pattern, directivity, etc. Since all rays from the feed travel the same physical distance to the aperture, the aperture distribution is of uniform phase. However, there is a non-uniform amplitude distribution. This is because the power density of the rays leaving the feed falls off as $1/r_f^2$. After the reflection, there is practically no spreading loss since the rays are collimated (parallel). The aperture field-amplitude distribution varies as $1/r_f$. This is explained in detail below.



GO assumes that the power density in free space follows straight paths. Applied to the power transmitted by the feed, the power in a conical wedge stays confined within as it progresses along the cone's axis. Consider a conical wedge of solid angle $d\Omega$ whose cross-section angle is $d\theta_f$. It confines power, which after being reflected from the paraboloid, arrives at the aperture plane confined within a cylindrical ring of thickness $d\rho'$ and area $dA = 2\pi\rho'd\rho'$.

Let us assume that the feed is isotropic and it has radiation intensity $U = \Pi_t / 4\pi$, where Π_t is the transmitted power. The power confined in the conical wedge is $d\Pi = Ud\Omega = (\Pi_t / 4\pi)d\Omega$. This power reaches the aperture with a density of

$$P_a(\rho') = \frac{d\Pi}{dA} = \frac{\Pi_t}{4\pi} \frac{d\Omega}{dA}. \quad (19.13)$$

The generic relation between the solid angle increment and the directional-angle increments is

$$d\Omega = \sin\theta d\theta d\varphi, \quad (19.14)$$

(see Lecture 4). In this case, the structure is rotationally symmetric, so we define the solid angle of the conical wedge as

$$d\Omega = \int_0^{2\pi} (\sin \theta_f d\theta_f) d\varphi_f = 2\pi \sin \theta_f d\theta_f. \quad (19.15)$$

The substitution of (19.15) and $dA = 2\pi\rho'd\rho'$ in (19.13) produces

$$P_a(\rho') = \frac{\Pi_t}{4\pi} \frac{2\pi \sin \theta_f d\theta_f}{2\pi\rho'd\rho'} = \frac{\Pi_t}{4\pi} \frac{\sin \theta_f}{\rho'} \frac{d\theta_f}{d\rho'}. \quad (19.16)$$

From (19.3), it is seen that

$$\frac{d\rho'}{d\theta_f} = \frac{F}{\cos^2(\theta_f/2)} = r_f, \quad (19.17)$$

$$\Rightarrow \frac{d\theta_f}{d\rho'} = \frac{1}{r_f}, \quad (19.18)$$

$$\Rightarrow P_a(\rho') = \frac{\Pi_t}{4\pi} \frac{\sin \theta_f}{\underbrace{r_f \sin \theta_f}_{\rho'}} \frac{1}{r_f} = \frac{\Pi_t}{4\pi} \frac{1}{r_f^2}. \quad (19.19)$$

Equation (19.19) shows the spherical nature of the feed radiation, and it is referred to as **spherical spreading loss**. Since $E_a \propto \sqrt{P_a}$,

$$E_a \propto \frac{1}{r_f}. \quad (19.20)$$

If the primary feed is not isotropic, the effect of its normalized field pattern $F_f(\theta_f, \varphi_f)$ is easily incorporated in (19.20) as

$$E_a \propto \frac{F_f(\theta_f, \varphi_f)}{r_f}. \quad (19.21)$$

Thus, we can conclude that the field at the aperture is described as

$$E_a(\theta_f, \varphi_f) = E_m e^{-j\beta 2F} \cdot \frac{F_f(\theta_f, \varphi_f)}{r_f}. \quad (19.22)$$

The coordinates (ρ', φ') are more suitable for the description of the aperture field. Obviously, $\varphi' \equiv \varphi_f$. As for r_f and θ_f , they are transformed as

$$r_f = \frac{4F^2 + \rho'^2}{4F}, \quad (19.23)$$

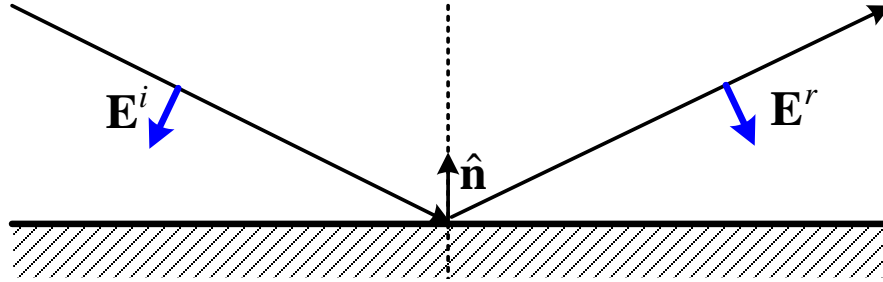
$$\theta_f = 2 \arctan \frac{\rho'}{2F}. \quad (19.24)$$

The last thing to be determined is the polarization of the aperture field provided the polarization of the primary-feed field is known. The law of reflection at a perfectly conducting wall states that $\hat{\mathbf{n}}$ bisects the incident and the reflected rays, and that the total electric field has zero tangential component at the surface, i.e.,

$$\mathbf{E}_\tau^i + \mathbf{E}_\tau^r = 0, \quad (19.25)$$

and

$$\mathbf{E}^r + \mathbf{E}^i = 2(\hat{\mathbf{n}} \cdot \mathbf{E}^i)\hat{\mathbf{n}} \Rightarrow \mathbf{E}^r = 2(\hat{\mathbf{n}} \cdot \mathbf{E}^i)\hat{\mathbf{n}} - \mathbf{E}^i. \quad (19.26)$$



Since $|\mathbf{E}^i| = |\mathbf{E}^r|$, from (19.26), it follows that

$$\hat{\mathbf{e}}_r = 2(\hat{\mathbf{n}} \cdot \hat{\mathbf{e}}_i)\hat{\mathbf{n}} - \hat{\mathbf{e}}_i. \quad (19.27)$$

Here, $\hat{\mathbf{e}}_i$ is the polarization vector of the incident field, and $\hat{\mathbf{e}}_r$ is the polarization vector of the reflected field.

The aperture field distribution is fully defined by (19.22) and (19.27). The radiation integral over the electric field can now be found. For example, a circular paraboloid would have a circular aperture (see Lecture 18), and the radiation integral becomes

$$\mathbf{I}^E = [(\hat{\mathbf{e}}_r \cdot \hat{\mathbf{x}})\hat{\mathbf{x}} + (\hat{\mathbf{e}}_r \cdot \hat{\mathbf{y}})\hat{\mathbf{y}}] E_m \int_0^{2\pi} \int_0^{D/2} \frac{F_f(\rho', \varphi')}{r_f} e^{j\beta\rho' \sin\theta \cos(\varphi - \varphi')} \rho' d\rho' d\varphi'. \quad (19.28)$$

In the above considerations, it was assumed that the aperture field has uniform phase distribution. This is true if the feed is located at the focal point. However, more sophisticated designs often use an offset feed. In such cases, the PO method (i.e., the current distribution method) is preferred.

4. The current distribution (PO) method (surface integration)

The basic description of this approach and its assumptions were already given in the previous section. Once the induced surface currents \mathbf{J}_s are found on the reflector's surface, the magnetic vector potential \mathbf{A} and the far-zone field can be calculated. In practice, the electric far field is calculated directly from \mathbf{J}_s by

$$\mathbf{E}^{far} = -j\omega\mu \frac{e^{-j\beta r}}{4\pi r} \iint_{S_r} \underbrace{[\mathbf{J}_s - (\mathbf{J}_s \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}]}_{\mathbf{J}_{s,\perp\hat{\mathbf{r}}}} e^{j\beta\hat{\mathbf{r}}\cdot\mathbf{r}'} ds'. \quad (19.29)$$

Equation (19.29) follows directly from the relation between the far-zone electric field and the magnetic vector potential \mathbf{A} ,

$$\mathbf{E}^{far} = -j\omega\mathbf{A}_\perp, \quad (19.30)$$

which can be written formally as

$$\mathbf{E}^{far} = -j\omega\mathbf{A} - (-j\omega\mathbf{A} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} = -j\omega(A_\theta\hat{\boldsymbol{\theta}} + A_\varphi\hat{\boldsymbol{\phi}}). \quad (19.31)$$

This approach is also known as Rusch's method. The integral in (19.29) is usually evaluated numerically by computer codes in order to render the approach versatile with respect to any aperture shape and any aperture current distribution.

In conclusion, we note that both the GO and the PO methods produce very accurate results for the main beam and first side lobe. The pattern far out the main beam can be accurately predicted by including diffraction effects (scattering) from the reflector's rim. This is done by augmenting GO with the use of *geometrical theory of diffraction* (GTD) (J.B. Keller, 1962), or by augmenting the PO method with the *physical theory of diffraction* (PTD) (P.I. Ufimtsev, 1957).

5. The focus-fed axisymmetric parabolic reflector antenna

This is a popular reflector antenna, whose analysis is used here to illustrate the general approach to the analysis of any reflector antenna. Consider a linearly polarized feed, with the \mathbf{E} field along the x -axis. As before, the reflector's axis is along z . Let us also assume that the field of the feed is represented by

$$\mathbf{E}_f(\theta_f, \varphi_f) \approx E_m \frac{e^{-j\beta r_f}}{r_f} \left[\hat{\boldsymbol{\theta}}_f C_E(\theta_f) \cos \varphi_f - \hat{\boldsymbol{\phi}}_f C_H(\theta_f) \sin \varphi_f \right]. \quad (19.32)$$

Here, $C_E(\theta_f)$ and $C_H(\theta_f)$ denote the principal-plane patterns. The expression in (19.32) is a common approximation of a 3-D pattern of an x -polarized antenna by knowing only the two principal-plane 2-D patterns. This approximation is actually very accurate for aperture-type antennas because it directly follows from the expression of the far-zone fields in terms of the radiation integrals (see Lecture 17, Section 4):

$$E_\theta = j\beta \frac{e^{-j\beta r}}{4\pi r} [I_x^E \cos \varphi + I_y^E \sin \varphi + \eta \cos \theta (I_y^H \cos \varphi - I_x^H \sin \varphi)], \quad (19.33)$$

$$E_\varphi = j\beta \frac{e^{-j\beta r}}{4\pi r} [-\eta (I_x^H \cos \varphi + I_y^H \sin \varphi) + \cos \theta (I_y^E \cos \varphi - I_x^E \sin \varphi)]. \quad (19.34)$$

The aperture field is now derived in terms of x - and y -components. To do this, the GO method of Section 2 is used. An incident field of $\hat{\mathbf{e}}_i = \hat{\boldsymbol{\theta}}_f$ polarization produces an aperture reflected field of the following polarization [see (19.9) and (19.27)]:

$$\begin{aligned} \hat{\mathbf{e}}_r^\theta &= 2(\hat{\mathbf{n}} \cdot \hat{\boldsymbol{\theta}}_f) \hat{\mathbf{n}} - \hat{\boldsymbol{\theta}}_f = 2 \sin\left(\frac{\theta_f}{2}\right) \hat{\mathbf{n}} - \hat{\boldsymbol{\theta}}_f = 2 \sin\left(\frac{\theta_f}{2}\right) \cdot \left(-\hat{\mathbf{r}}_f \cos \frac{\theta_f}{2} + \hat{\boldsymbol{\theta}}_f \sin \frac{\theta_f}{2}\right) - \hat{\boldsymbol{\theta}}_f \\ \Rightarrow \hat{\mathbf{e}}_r^\theta &= -\hat{\mathbf{r}}_f \left(2 \sin \frac{\theta_f}{2} \cdot \cos \frac{\theta_f}{2}\right) - \hat{\boldsymbol{\theta}}_f \left(1 - 2 \sin^2 \frac{\theta_f}{2}\right) = -\hat{\mathbf{r}}_f \sin \theta_f - \hat{\boldsymbol{\theta}}_f \cos \theta_f. \end{aligned} \quad (19.35)$$

Similarly, an incident field of $\hat{\mathbf{e}}_i = \hat{\boldsymbol{\phi}}_f$ polarization produces an aperture reflected field of the following polarization:

$$\hat{\mathbf{e}}_r^\phi = -\hat{\boldsymbol{\phi}}_f. \quad (19.36)$$

Transforming (19.35) and (19.36) to rectangular (x and y) coordinates at the aperture plane gives:

$$\begin{aligned} \hat{\mathbf{e}}_r^\theta &= -\hat{\mathbf{x}} \cos \varphi_f - \hat{\mathbf{y}} \sin \varphi_f, \\ \hat{\mathbf{e}}_r^\phi &= +\hat{\mathbf{x}} \sin \varphi_f - \hat{\mathbf{y}} \cos \varphi_f. \end{aligned} \quad (19.37)$$

Superimposing the contributions of the $\hat{\boldsymbol{\theta}}_f$ and $\hat{\boldsymbol{\phi}}_f$ components of the field in (19.32) to the aperture field x and y components produces

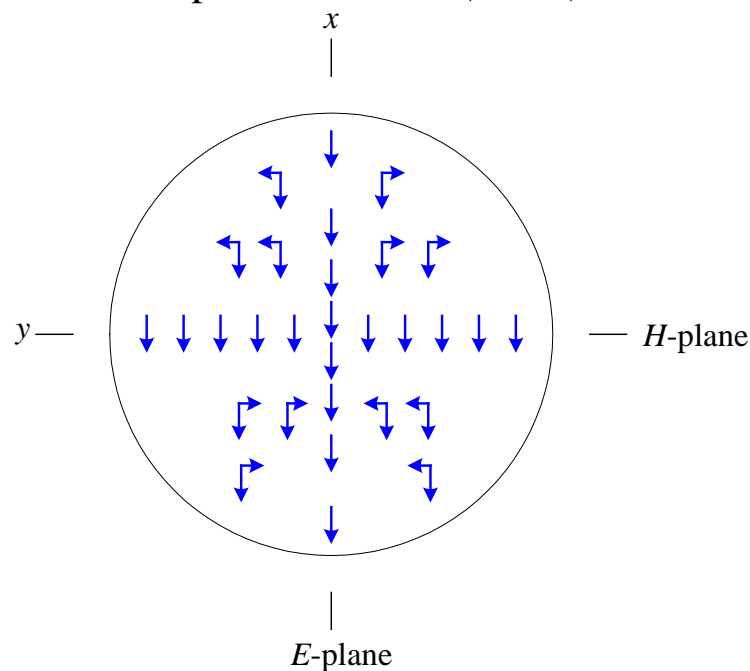
$$\mathbf{E}_a(\theta_f, \varphi_f) = E_m \frac{e^{-j\beta 2F}}{r_f} \times \left\{ -\hat{\mathbf{x}} \left[C_E(\theta_f) \cos^2 \varphi_f + C_H(\theta_f) \sin^2 \varphi_f \right] - \hat{\mathbf{y}} \left[C_E(\theta_f) - C_H(\theta_f) \right] \sin \varphi_f \cdot \cos \varphi_f \right\}. \quad (19.38)$$

In (19.38), the magnitude and phase of the vector are expressed as in (19.22). Note that a y -component appears in the aperture field, despite the fact that the feed generates only E_x field. This is called **cross-polarization**. If the feed has rotationally symmetric pattern, i.e. $C_E(\theta_f) = C_H(\theta_f)$, there is no cross-polarization. From equation (19.38), it is also obvious that cross-polarization is zero at $\varphi_f = 0^\circ$ (E -plane) and at $\varphi_f = 90^\circ$ (H -plane). Cross-polarization is maximum at $\varphi_f = 45^\circ, 135^\circ$. Cross-polarization in the aperture means cross-polarization of the far field, too. Cross-polarization is usually unwanted because it leads to polarization losses depending on the transmitting and receiving antennas.

It is instructive to examine (19.38) for a specific simple example: reflector antenna fed by a very short x -polarized electric dipole. Its principal-plane patterns are $C_E(\theta_f) = \cos \theta_f$ and $C_H(\theta_f) = 1$. Therefore, it generates the following aperture field:

$$\mathbf{E}_a = E_m \frac{e^{-j\beta 2F}}{r_f} \left[-\hat{\mathbf{x}}(\cos \theta_f \cos^2 \varphi_f + \sin^2 \varphi_f) - \hat{\mathbf{y}}(\cos \theta_f - 1) \sin \varphi_f \cos \varphi_f \right]. \quad (19.39)$$

An approximate plot of the aperture field of (19.39) is shown below.



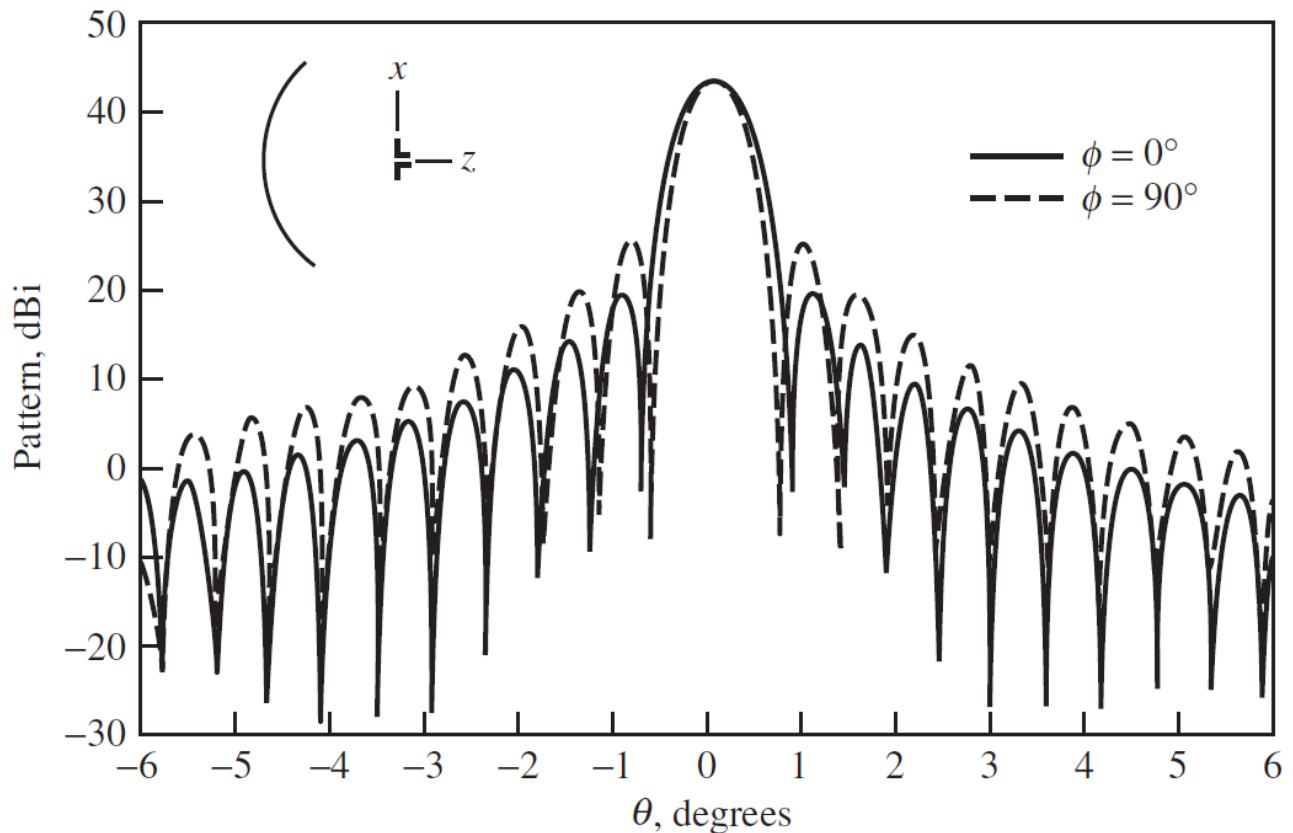
We also note that cross-polarization decreases as the ratio F / D increases. This follows from (19.4), which gives the largest feed angle $(\theta_f)_{\max} = \theta_0$. As F / D increases, θ_0 decreases, which makes the cross-polarization term in (19.39) smaller. Unfortunately, large F / D ratios are not very practical.

Finally, we add that a similar analysis for a y -polarized small dipole feed leads to an expression for the aperture field similar to the one in (19.39) but with a polarization vector

$$\hat{\mathbf{e}}_a = \frac{\hat{\mathbf{x}} \sin \varphi_f \cos \varphi_f (1 - \cos \theta_f) - \hat{\mathbf{y}} (\cos \theta_f \sin^2 \varphi_f + \cos^2 \varphi_f)}{\sqrt{1 - \sin^2 \theta_f \sin^2 \varphi_f}}. \quad (19.40)$$

An example is presented in W.L. Stutzman, G. Thiele, *Antenna Theory and Design*, of an axisymmetric parabolic reflector with diameter $D = 100\lambda$ and $F / D = 0.5$, fed by a half-wavelength dipole located at the focus.

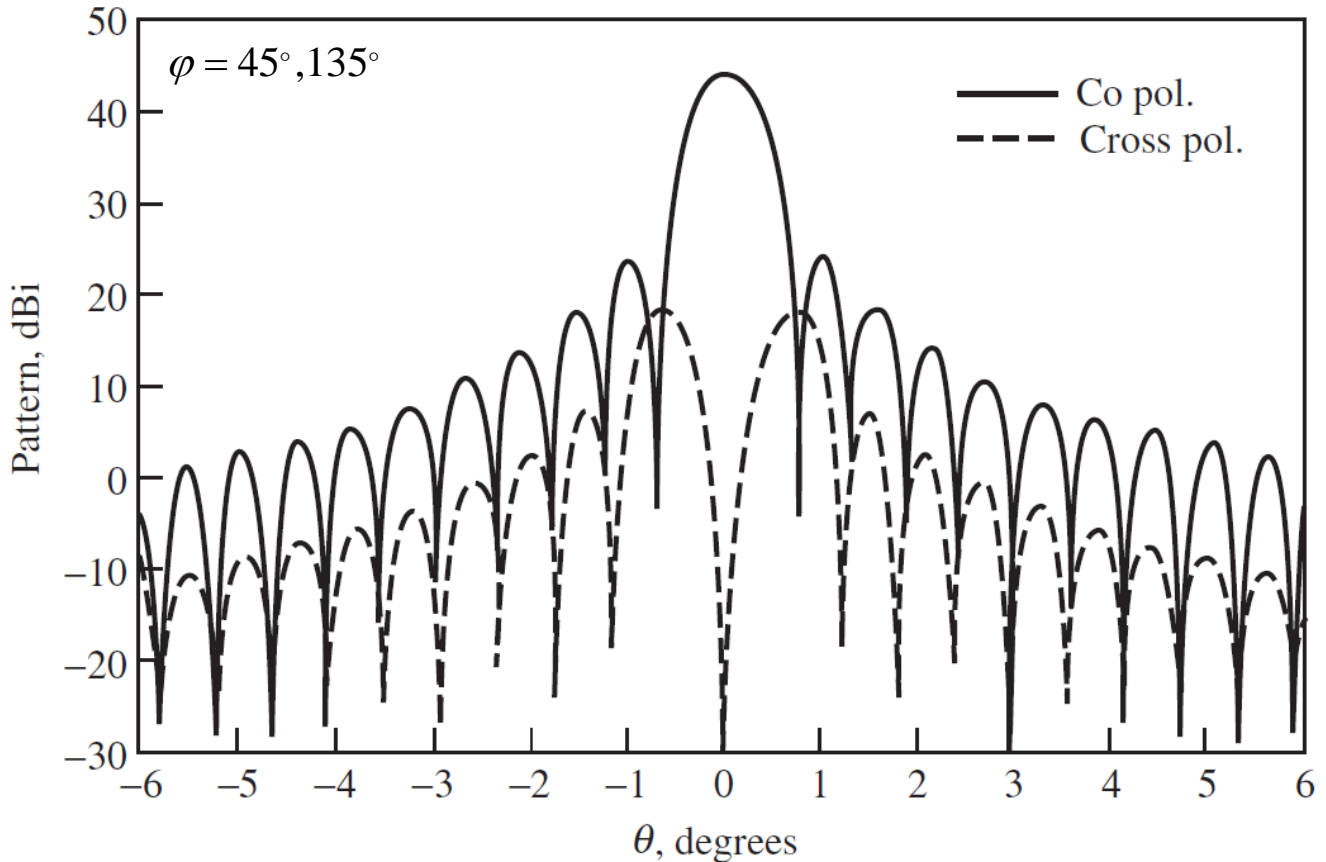
CO-POLARIZATION



(a) Principal plane patterns.

[Stutzman&Thiele]

CROSS-POLARIZATION



[Stutzman&Thiele]

The results above are obtained using commercial software (GRASP) using PO methods (surface current integration).

Cross-polarization of reflectors is measured as the ratio of the peak cross-polarization far-field component to the peak co-polarization far field. For example, the above graph shows a cross-polarization level of $XPOL = -26.3$ dB.

6. Offset parabolic reflectors

One disadvantage of the focus-fed reflector antennas is that part of the aperture is blocked by the feed. To avoid this, offset-feed reflectors are developed, where the feed antenna is away from the reflector's aperture. The reflectors are made as *a portion of* the so-called *parent reflector* surface. The price to pay is the increase of cross-polarization. That is why such reflectors are usually fed with primary feeds of rotationally symmetrical patterns, where $C_E \approx C_H$. This effectively eliminates cross-polarization.

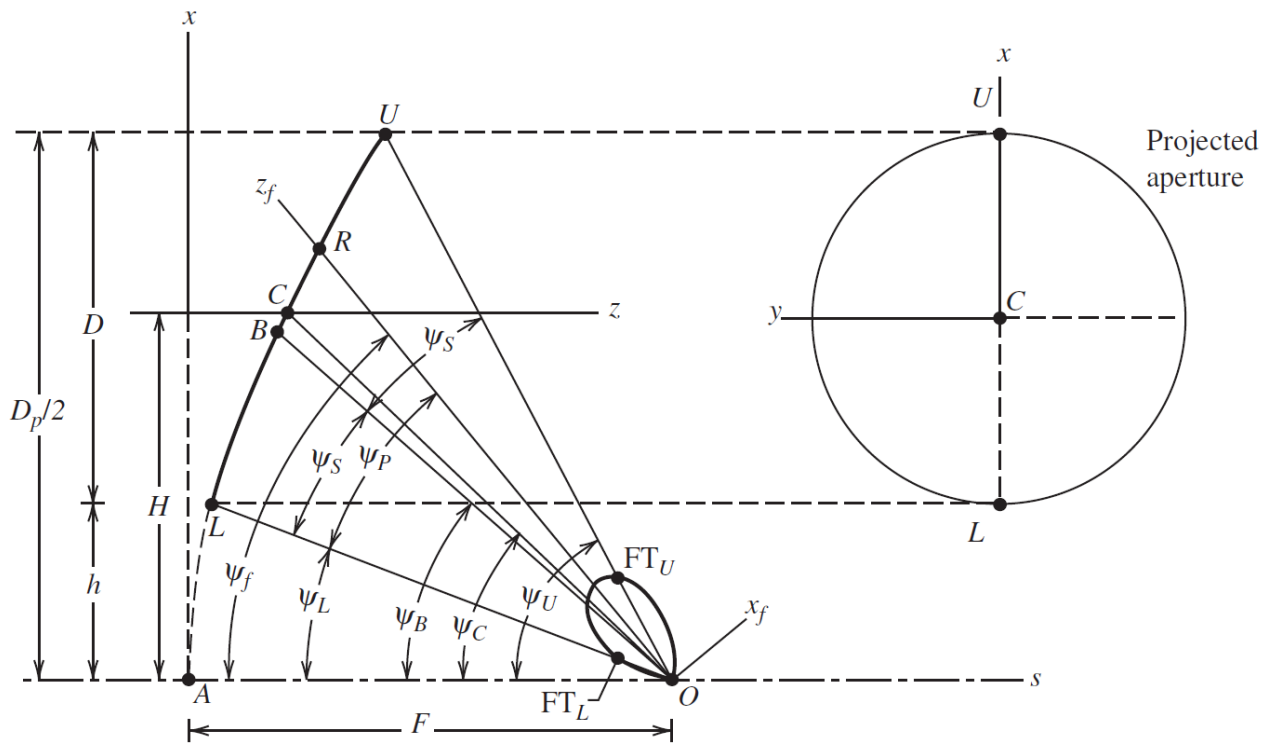
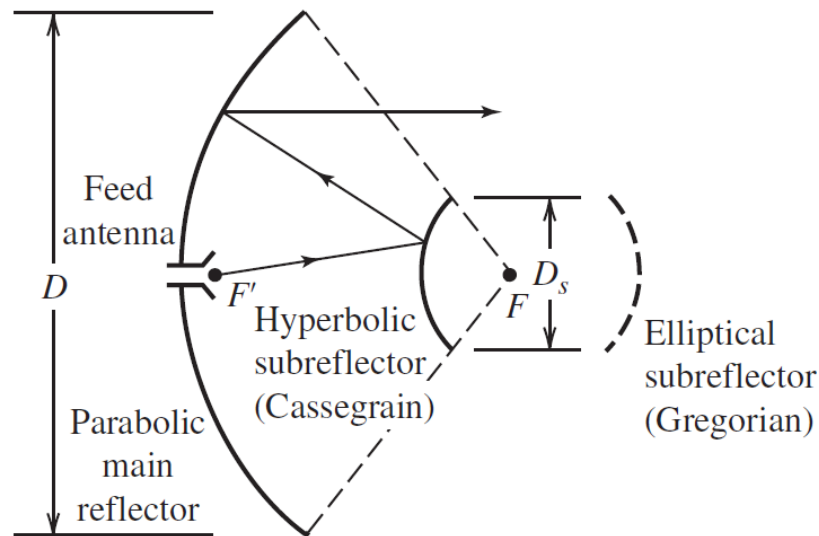


Figure 9-31 Geometry of the offset parabolic reflector of diameter D and focal length F . The axis of symmetry s bisects the parent parabolic curve of diameter D_p . Note that the axisymmetric case occurs when $H = 0$.

[Stutzman&Thiele]

The analysis techniques given in the previous sections are general and can be applied to these reflectors, too. Generally, the PO method (surface currents integration) is believed to yield better accuracy. Both, the PO and the GO methods are accurate only at the main beam and the first couple of side-lobes.

Offset reflectors are popular for antenna systems producing *contoured beams*. A contoured beam is a beam, the cross-section (or footprint) of which has a shape that conforms to a desired Earth region such as a country. To obtain such beams, multiple primary feeds (usually horns) are needed to illuminate the reflector at different angles. Such multiple-antenna feeds may constitute a significant obstacle at the antenna aperture and offset reflectors are indeed necessary.



[Stutzman&Thiele]

7. Dual-reflector antennas

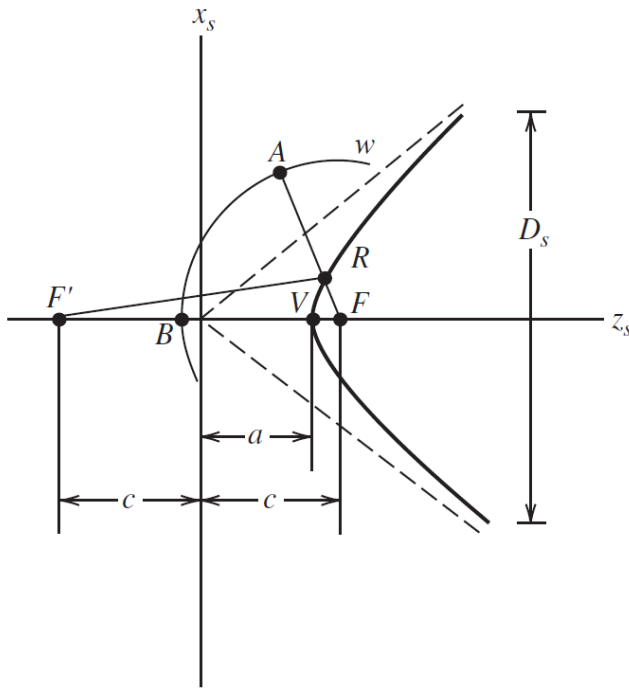
The dual-reflector antenna consists of two reflectors and a feed antenna. The feed is conveniently located at the apex of the main reflector. This makes the system mechanically robust, the transmission lines are shorter and easier to construct (especially in the case of waveguides). The virtual focal point F is the point from which the rays transmitted toward the reflector appear to emanate after reflection from the subreflector.

The most popular dual reflector is the axisymmetric *Cassegrain* antenna. The main reflector is parabolic and the *subreflector is hyperbolic* (convex).

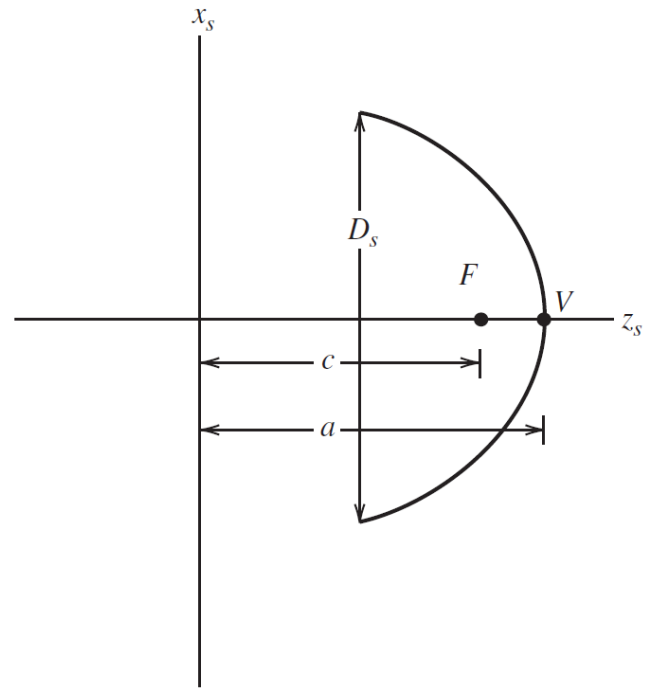
A second form of the dual reflector is the *Gregorian* reflector. It has a concave *elliptic subreflector*. The Gregorian subreflector is more distant from the main reflector and, thus, it requires more support.

Dual-reflector antennas for Earth terminals have another important advantage beside the location of the main feed. They have almost no spillover toward the noisy ground, as do the single-feed reflector antennas. Their spillover (if any) is directed toward the much less noisy sky region. Both, the Cassegrain and the Gregorian reflector systems have their origins in optical telescopes and are named after their inventors.

The subreflectors are rotationally symmetric surfaces obtained from the curves shown below (a hyperbola and an ellipse).



(a) Hyperbolic subreflector.



(b) Elliptical subreflector.

[Stutzman&Thiele]

The subreflector is defined by its diameter D_s and its eccentricity e . The shape (or curvature) is controlled by the eccentricity:

$$e = \frac{c}{a} \begin{cases} > 1, & \text{hyperbola} \\ < 1, & \text{ellipse} \end{cases} \quad (19.41)$$

Other special cases include:

- $e = \infty$, straight line (plane)
- $e = 0$, circle (sphere)
- $e = 1$, parabola

Both, the ellipse and the hyperbola, are described by the equation

$$\frac{z_s^2}{a^2} - \frac{x_s^2}{c^2 - a^2} = 1. \quad (19.42)$$

The function of a hyperbolic subreflector is to convert the incoming wave from a feed antenna located at the focal point F' to a spherical wave front w that appears to originate from the virtual focal point F . This means that the optical path from F' to w must be constant with respect to the angle of incidence:

$$\overline{F'R} + \overline{RA} = \overline{F'V} + \overline{VB} = \underbrace{c + a}_{\overline{F'V}} + \overline{VB}. \quad (19.43)$$

Since

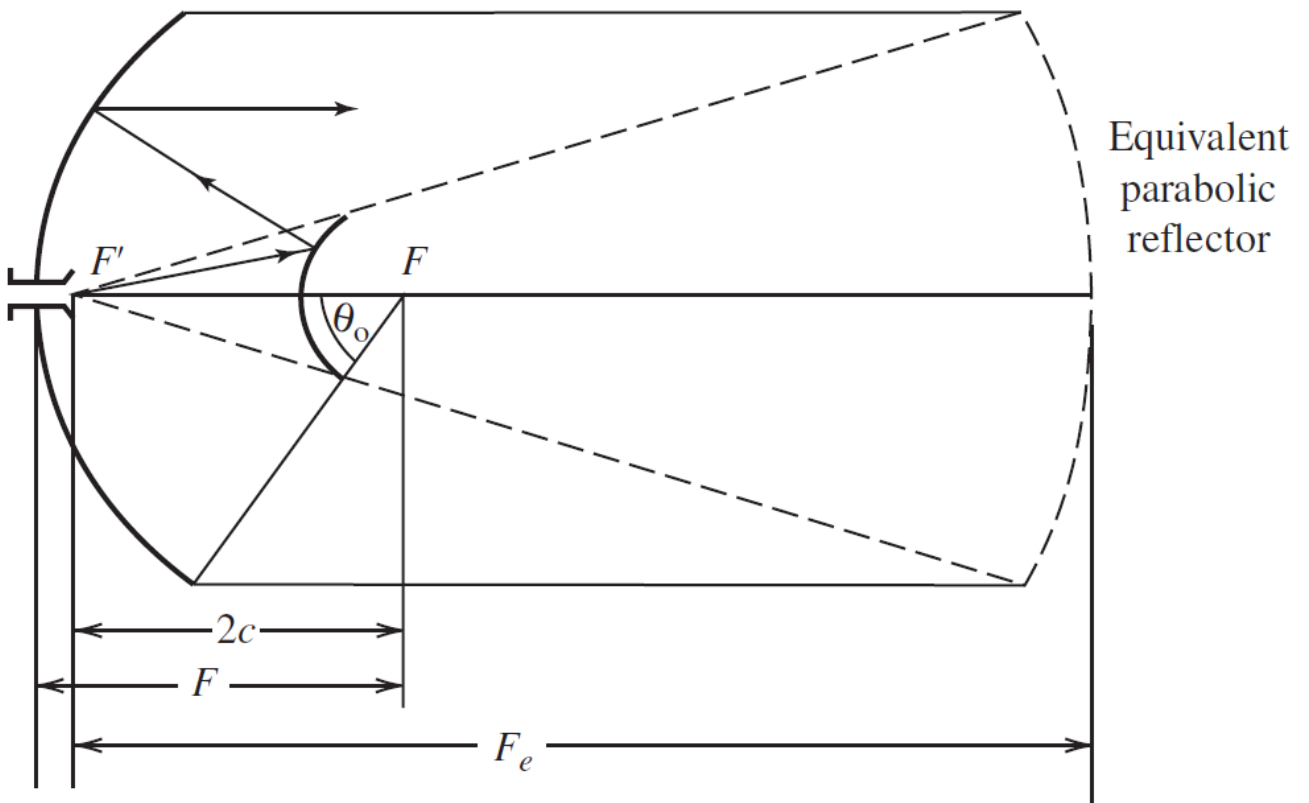
$$\overline{RA} = \overline{FA} - \overline{FR} = \overline{FB} - \overline{FR}, \quad (19.44)$$

($\overline{FA} = \overline{FB}$ because the reflected wave must be spherical)

$$\Rightarrow \overline{F'R} - \overline{FR} = c + a - (\overline{FB} - \overline{VR}) = c + a - (c - a) = 2a. \quad (19.45)$$

Note: Another definition of a hyperbola is: a hyperbola is the locus of a point that moves so that the difference of the distances from its two focal points, $\overline{F'R} - \overline{FR}$, is equal to a constant, $2a$. In contrast, an ellipse is defined as the locus traced by a point moving in a plane so that the sum of its distances from its two foci is constant.

The dual axisymmetric Cassegrain reflector can be modeled as a single equivalent parabolic reflector as shown below.



[Stutzman&Thiele]

The equivalent parabola has the same diameter, $D_e = D$, but its focal length is longer than that of the main reflector:

$$F_e = \left(\frac{e+1}{e-1} \right) \cdot F = M \cdot F. \quad (19.46)$$

Here, $M = (e+1)/(e-1)$ is called *magnification*.

The increased equivalent focal length has several advantages:

- less cross-polarization;
- less spherical-spread loss at the reflector's rim, and therefore, improved aperture efficiency.

The synthesis of dual-reflector systems is an advanced topic. Many factors are taken into account when *shaped* reflectors are designed for improved aperture efficiency. These are: minimized spillover, less phase error, improved amplitude distribution in the reflector's aperture.

8. Gain of reflector antennas

The maximum achievable gain for an aperture antenna is

$$G_{\max} = D_u = \frac{4\pi}{\lambda^2} A_p. \quad (19.47)$$

This gain is possible only if the following is true: uniform amplitude and phase distribution, no spillover, no ohmic losses. In practice, these conditions are not achievable, and the effective antenna aperture is less than its physical aperture:

$$G = \varepsilon_{ap} D_u = \frac{4\pi}{\lambda^2} \varepsilon_{ap} A_p, \quad (19.48)$$

where $\varepsilon_{ap} \leq 1$ is the aperture efficiency. The aperture efficiency is expressed as a product of sub-efficiencies:

$$\varepsilon_{ap} = e_r \varepsilon_t \varepsilon_s \varepsilon_a, \quad (19.49)$$

where:

- e_r is the radiation efficiency (loss),
- ε_t is the aperture taper efficiency,
- ε_s is the spillover efficiency, and
- ε_a is the achievement efficiency.

The taper efficiency can be found using the directivity expression for aperture antennas (see Lecture 17, Section 5):

$$D_0 = \frac{4\pi}{\lambda^2} \frac{\left| \iint_{S_A} \mathbf{E}_a ds' \right|^2}{\iint_{S_A} |\mathbf{E}_a|^2 ds'}. \quad (19.50)$$

$$\Rightarrow A_{eff} = \frac{\left| \iint_{S_A} \mathbf{E}_a ds' \right|^2}{\iint_{S_A} |\mathbf{E}_a|^2 ds'}. \quad (19.51)$$

$$\Rightarrow \varepsilon_t = \frac{A_{eff}}{A_p} = \frac{1}{A_p} \frac{\left| \iint_{S_A} \mathbf{E}_a ds' \right|^2}{\iint_{S_A} |\mathbf{E}_a|^2 ds'}. \quad (19.52)$$

Expression (19.52) can be written directly in terms of the known feed antenna pattern. If the aperture is circular, then

$$\varepsilon_t = \frac{1}{\pi a^2} \frac{\left| \int_0^{2\pi} \int_0^a \mathbf{E}_a(\rho', \varphi') \rho' d\rho' d\varphi' \right|^2}{\int_0^{2\pi} \int_0^a |\mathbf{E}_a(\rho', \varphi')|^2 \rho' d\rho' d\varphi'}. \quad (19.53)$$

Substituting $\rho' = r_f \sin \theta_f = 2F \tan(\theta_f / 2)$ and $d\rho' / d\theta_f = r_f$ in (19.53) yields

$$\varepsilon_t = \frac{4F^2}{\pi a^2} \frac{\left| \int_0^{2\pi} \int_0^{\theta_o} \mathbf{F}_f(\theta_f, \varphi') \tan\left(\frac{\theta_f}{2}\right) d\theta_f d\varphi' \right|^2}{\int_0^{2\pi} \int_0^{\theta_o} |\mathbf{F}_f(\theta_f, \varphi')|^2 \sin \theta_f d\theta_f d\varphi'}. \quad (19.54)$$

All that is needed to calculate the taper efficiency is the feed pattern $\mathbf{F}_f(\theta_f, \varphi')$.

If the feed pattern extends beyond the reflector's rim, certain amount of power is not redirected by the reflector, i.e., it is lost. This power-loss is referred to as *spillover*. The spillover efficiency measures that portion of the feed pattern, which is intercepted by the reflector relative to the total feed power:

$$\varepsilon_s = \frac{\int_0^{2\pi} \int_0^{\theta_0} |\mathbf{F}_f(\theta_f, \varphi')|^2 \sin \theta_f d\theta_f d\varphi'}{\int_0^{2\pi} \int_0^{\pi} |\mathbf{F}_f(\theta_f, \varphi')|^2 \sin \theta_f d\theta_f d\varphi'}. \quad (19.55)$$

The reflector design problem includes a trade-off between aperture taper and spillover when the feed antenna is chosen. Taper and spillover efficiencies are combined to form the so-called *illumination efficiency* $\varepsilon_i = \varepsilon_t \varepsilon_s$. Multiplying (19.54) and (19.55), and using $a = 2F \tan(\theta_0 / 2)$ yields

$$\varepsilon_i = \frac{D_f}{4\pi^2} \cot^2 \frac{\theta_0}{2} \left| \int_0^{2\pi} \int_0^{\theta_0} \mathbf{F}_f(\theta_f, \varphi') \tan \frac{\theta_f}{2} d\theta_f d\varphi' \right|^2. \quad (19.56)$$

Here,

$$D_f = \frac{4\pi}{\int_0^{2\pi} \int_0^{\pi} |\mathbf{F}_f(\theta_f, \varphi')|^2 \sin \theta_f d\theta_f d\varphi'}, \quad (19.57)$$

is the directivity of the feed antenna. An ideal feed antenna pattern would compensate for the spherical spreading loss by increasing the field strength as θ_f increases, and then would abruptly fall to zero in the direction of the reflector's rim in order to avoid spillover:

$$F_f(\theta_f, \varphi') = \begin{cases} \frac{\cos^2(\theta_0 / 2)}{\cos^2(\theta_f / 2)}, & \theta_f \leq \theta_0 \\ 0, & \theta_f > \theta_0 \end{cases} \quad (19.58)$$

This ideal feed is not realizable. For practical purposes, (19.56) has to be optimized with respect to the edge-illumination level. The function specified by (19.56) is well-behaved with a single maximum with respect to the edge-illumination.

The achievement efficiency ε_a is an integral factor including losses due to: random surface error, cross-polarization loss, aperture blockage, reflector phase error (profile accuracy), feed phase error.

A well-designed and well-made aperture antenna should have an overall

aperture efficiency of $\varepsilon_{ap} \approx 0.65$ or more, where “more” is less likely.

The gain of a reflector antenna also depends on *phase errors*, which theoretically should not exist but are often present in practice. Any departure of the phase over the virtual aperture from the uniform distribution leads to a significant decrease of the directivity. For paraboloidal antennas, phase errors result from:

- displacement of the feed phase centre from the focal point;
- deviation of the reflector surface from the paraboloidal shape, including surface roughness and other random deviations;
- feed wave fronts are not exactly spherical.

Simple expression has been derived¹ to predict with reasonable accuracy the loss in directivity for rectangular and circular apertures when the peak value of the aperture phase deviations is known. Assuming that the maximum radiation is along the reflector’s axis, and assuming a *maximum aperture phase deviation* m , the ratio of the directivity without phase errors D_0 and the directivity with phase errors D is given by

$$\frac{D}{D_0} \approx \left(1 - \frac{m^2}{2}\right)^2. \quad (19.59)$$

The maximum phase deviation m is defined as

$$|\Delta\phi| = |\phi - \bar{\phi}| \leq m, \quad (19.60)$$

where ϕ is the aperture’s phase function, and $\bar{\phi}$ is its average value. The aperture phase deviation should be kept below $\pi/8$ if the gain is not to be affected much. Roughly, this translates into surface profile deviation from the ideal shape (e.g. paraboloid) of no more than $\lambda/16$.

¹ D.K. Cheng, “Effects of arbitrary phase errors on the gain and beamwidth characteristics of radiation pattern,” *IRE Trans. AP*, vol. AP-3, No. 3, pp. 145-147, July 1955.

LECTURE 20: MICROSTRIP ANTENNAS – PART I

(Introduction. Construction and geometry. Feeding techniques. Substrate properties. Loss calculation.)

1. Introduction

Microstrip antennas (MSA) received considerable attention in the 1970's, although the first designs and theoretical models appeared in the 1950's. They are suitable for many mobile applications: handheld devices, aircraft, satellite, missile, etc. The MSA are low profile, mechanically robust, inexpensive to manufacture, compatible with MMIC designs and relatively light and compact. They are quite versatile in terms of resonant frequencies, polarization, pattern and impedance. They allow for additional tuning elements like pins or varactor diodes between the patch and the ground plane.

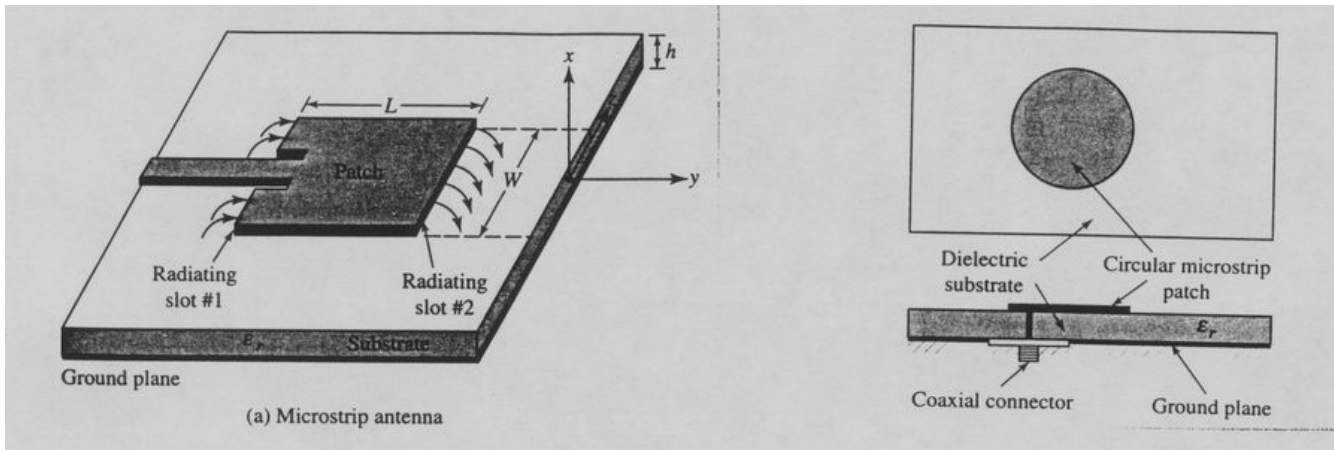
Some of the limitations and disadvantages of the MSA are:

- relatively low efficiency (due to dielectric and conductor losses)
- low power
- spurious feed radiation (surface waves, strips, etc.)
- narrow frequency bandwidth (at most a couple of percent)
- relatively high level of cross polarization radiation

MSA are applicable in the GHz range ($f > 0.5$ GHz). For lower frequencies their dimensions become too large.

2. Construction and Geometry

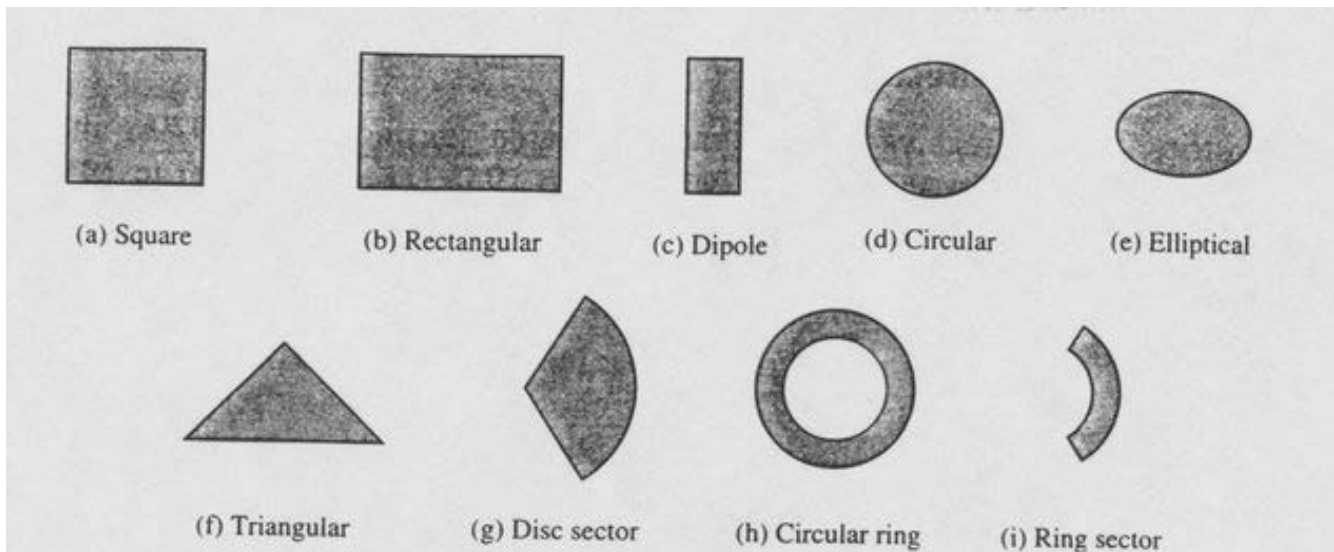
Generally the MSA are thin metallic patches of various shapes etched on dielectric substrates of thickness h , which usually is from $0.003\lambda_0$ to $0.05\lambda_0$. The substrate is usually grounded at the opposite side.



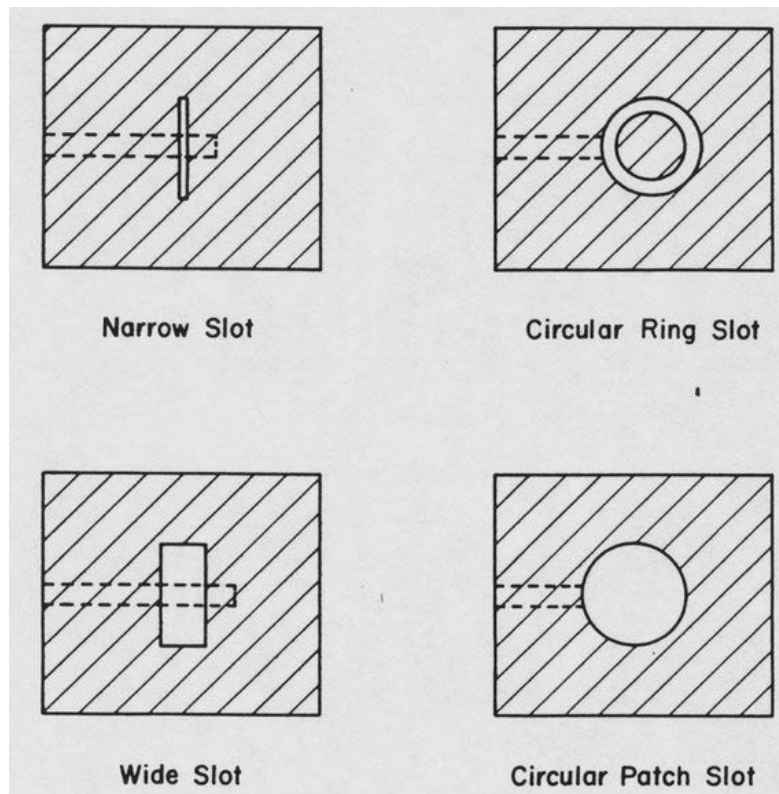
The dimensions of the patch are usually in the range from $\lambda_0/3$ to $\lambda_0/2$. The dielectric constant of the substrate ϵ_r is usually in the range from 2.2 to 12. The most common designs use relatively thick substrates with lower ϵ_r because they provide better efficiency and larger bandwidth. On the other hand, this implies larger dimensions of the antennas. The choice of the substrate is limited by the RF or microwave circuit coupled to the antenna, which has to be built on the same board. The microwave circuit together with the antenna is usually manufactured by photo-etching technology.

Types of microstrip radiators:

(a) single radiating patches

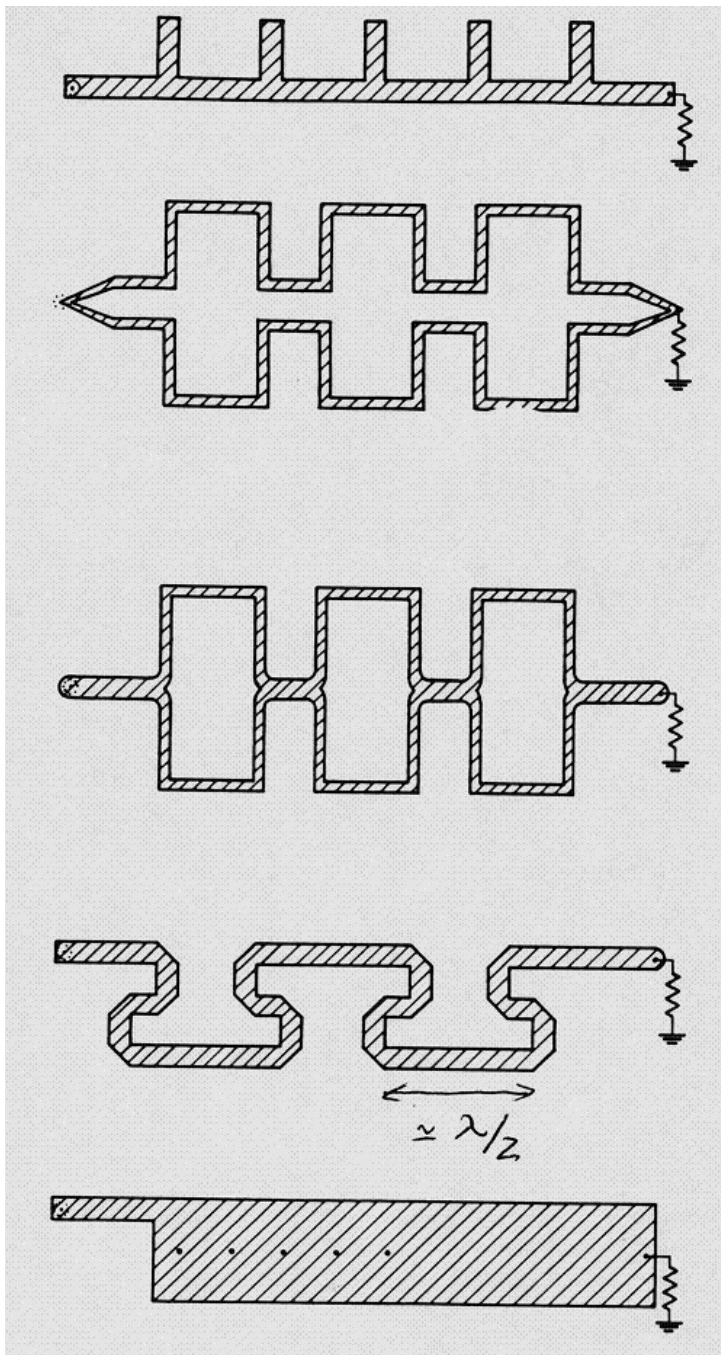


(b) single slot radiator



The feeding microstrip line is beneath (etched on the other side of the substrate) – see dash-line.

(c) microstrip traveling wave antennas (MTWA)



Comb MTWA

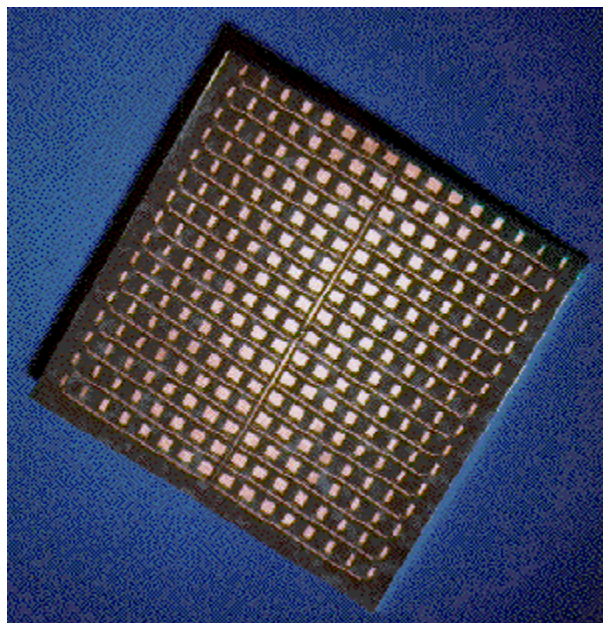
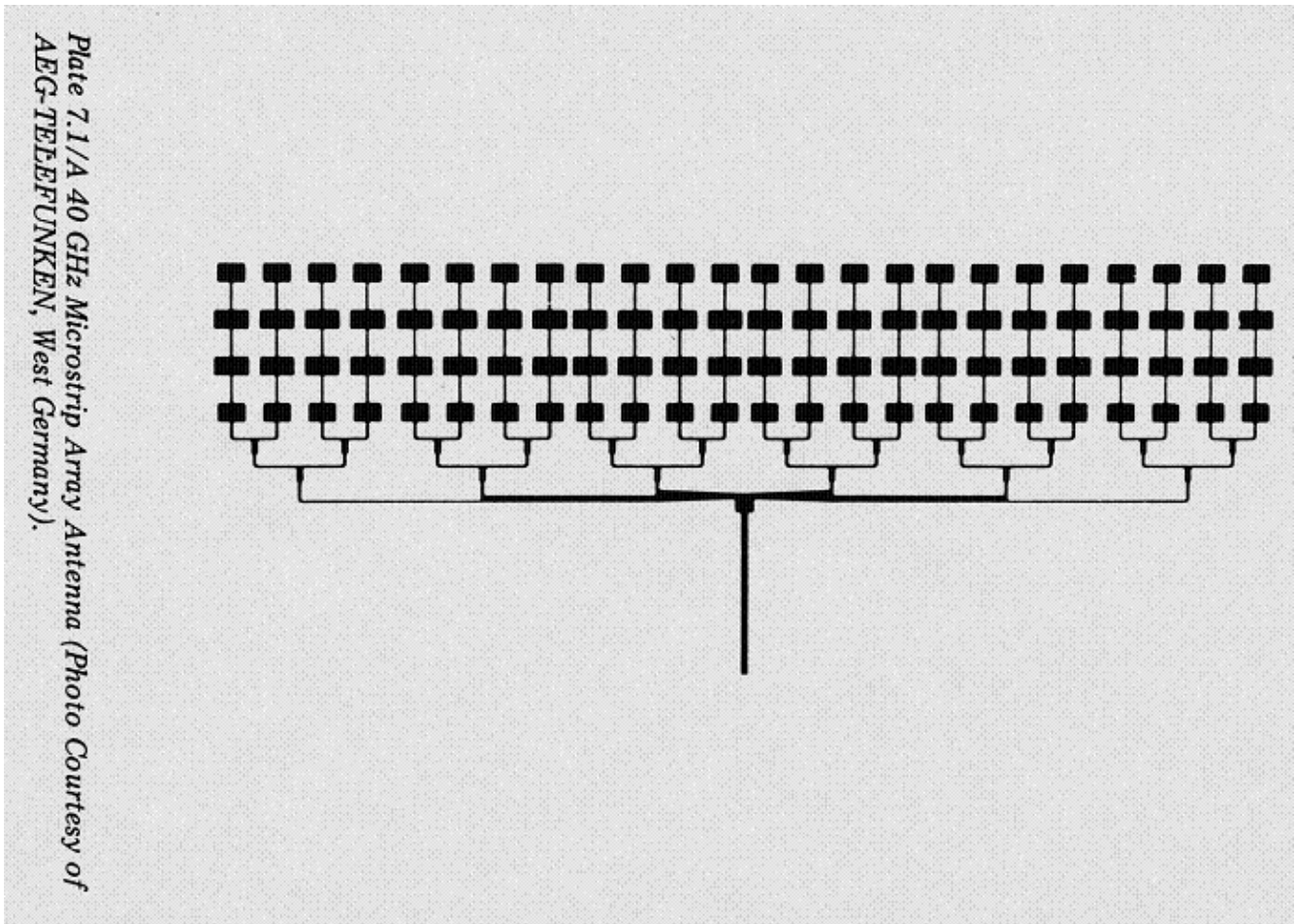
Meander Line Type MTWA

Rectangular Loop Type MTWA

Franklin – Type MTWA

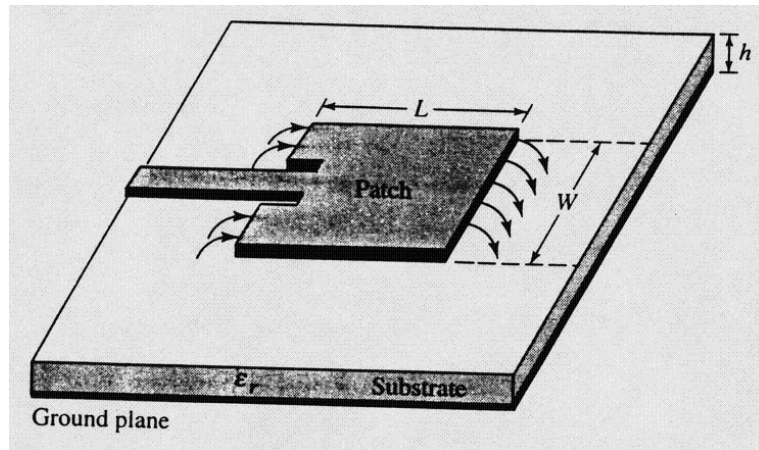
The open end of the long TEM line is terminated in a matched resistive load.

(d) microstrip antenna arrays

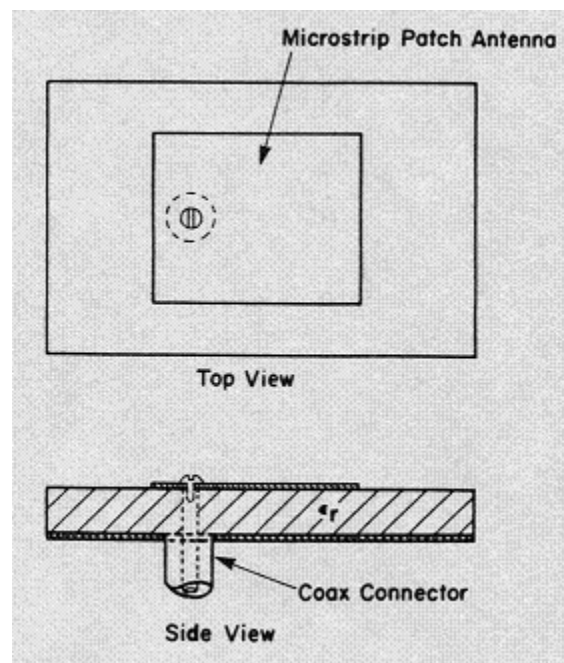


3. Feeding Methods

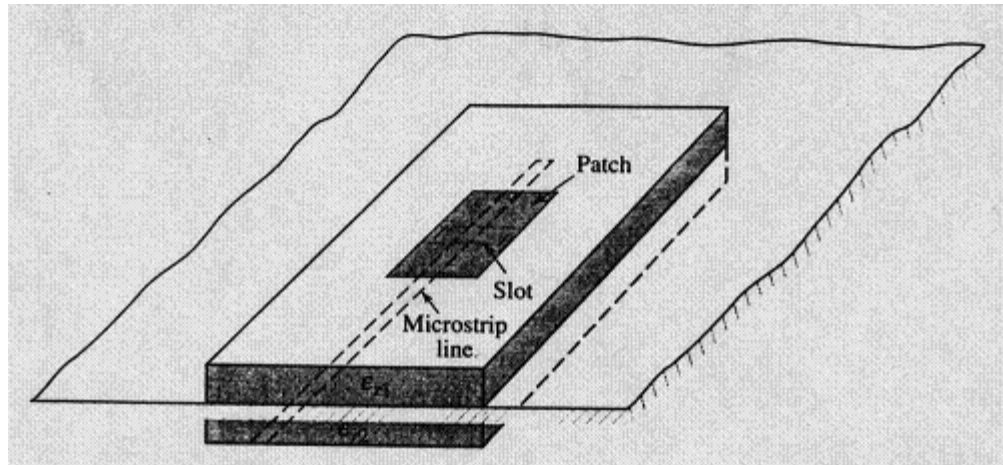
- 1) Microstrip feed – easy to fabricate, simple to match by controlling the inset position and relatively simple to model. However, as the substrate thickness increases, surface waves and spurious feed radiation increase.



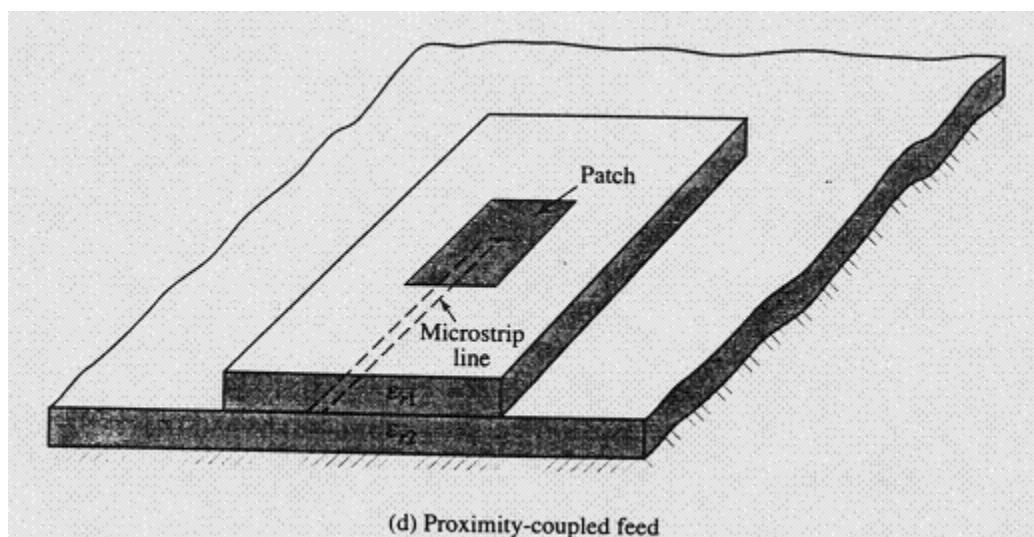
- 2) Coaxial probe feed – easy to fabricate, low spurious radiation; difficult to model accurately; narrow bandwidth of impedance matching.



- 3) Aperture coupling (no contact), microstrip feed line and radiating patch are on both sides of the ground plane, the coupling aperture is in the ground plane – low spurious radiation, easy to model; difficult to match, narrow bandwidth.

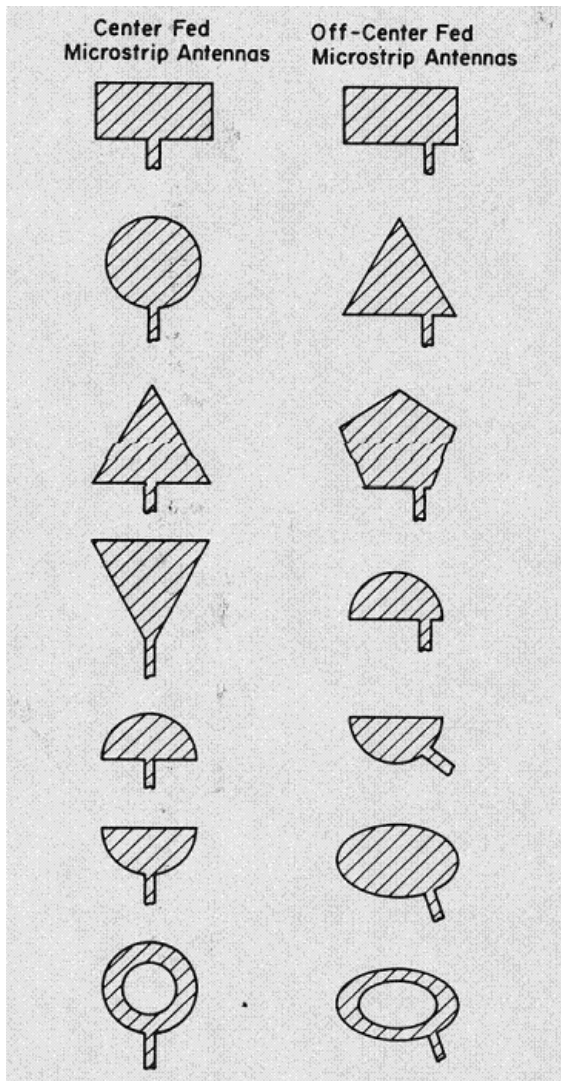


- 4) Proximity coupling (no contact), microstrip feed line and radiating patch are on the same side of the ground plane – largest bandwidth (up to 13%), relatively simple to model, has low spurious radiation.

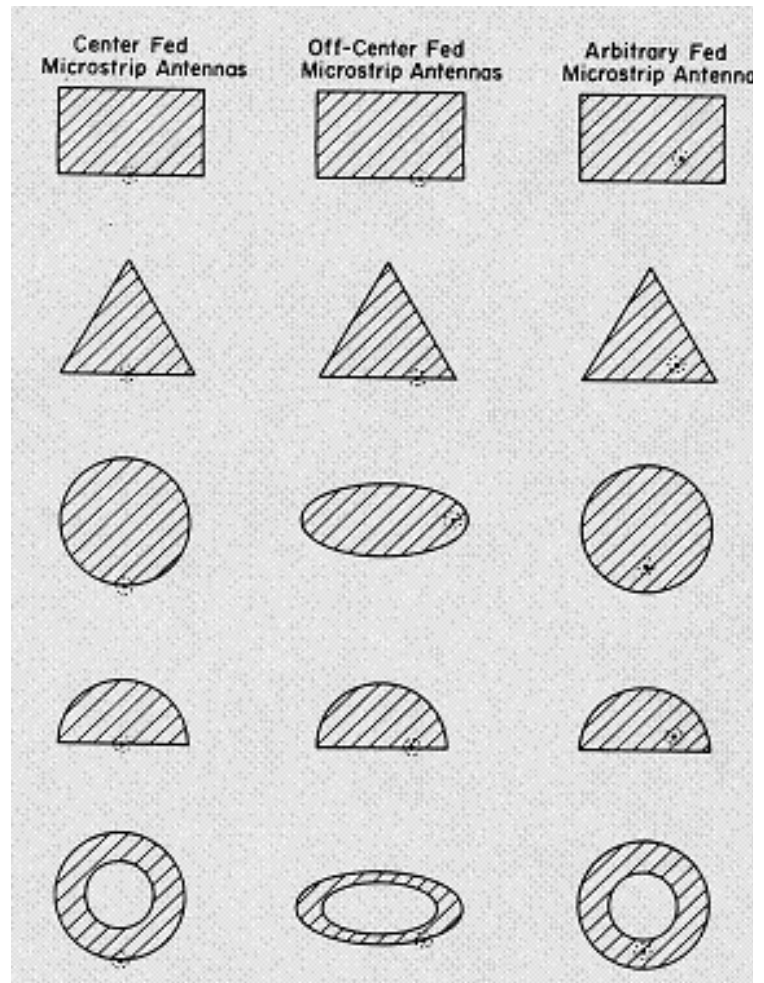


More examples of microstrip and coaxial probe feeds:

STRIP FEEDS



COAX FEEDS



4. Surface Waves

Surface waves can be excited at the dielectric-to-air interface. They give rise to end-fire radiation. In addition, they can lead to unwanted coupling between array elements. The phase velocity of the surface waves is strongly dependent on the dielectric constant ϵ_r and the thickness h of the substrate. The excitation of surface waves in a dielectric slab backed by a ground plane has been well studied (Collin, *Field Theory of Guided Waves*). The lowest-order TM mode, TM_0 , has no cut-off frequency. The cut-off frequencies for the higher-order modes (TM_n and TE_n) are given by

$$f_c^{(n)} = \frac{n \cdot c}{4h\sqrt{\epsilon_r - 1}}, \quad n = 1, 2, \dots, \quad (1)$$

where c is the speed of light in vacuum. The cut-off frequencies for the TE_n modes are given by the odd $n = 1, 3, 5, \dots$, and the cut-off frequencies for the TM_n modes are given by the even n . For the TE_1 mode, the calculated values of $h / \lambda_c^{(1)}$ are [$\lambda_c^{(1)} = c / f_c^{(1)}$, $h / \lambda_c^{(1)} = n / (4\sqrt{\epsilon_r - 1})$]:

- a) 0.217 for duroid ($\epsilon_r = 2.32$),
- b) 0.0833 for alumina ($\epsilon_r = 10$).

Thus, the lowest-order TE_1 mode is excited at 41 GHz for 1.6 mm thick duroid substrate, and at about 39 GHz for 0.635 mm thick alumina substrate. The substrate thickness is chosen so that the ratio h / λ_0 is well below $h / \lambda_c^{(1)}$ (λ_0 is the free-space wavelength at the operating frequency), i.e., [3]

$$h < \frac{c}{4f_u\sqrt{\epsilon_r - 1}}, \quad (2)$$

where f_u is the highest frequency in the band of operation. Note that h should be chosen as high as possible under the constraint of (2), so that maximum radiation efficiency is achieved. Also, h has to conform to the commercially available substrates. Another practical formula for h is given in [2]:

$$h \leq \frac{0.3c}{2\pi f_u\sqrt{\epsilon_r}}. \quad (3)$$

The TM_0 mode has no cut-off frequency and is always present to some extent. The surface TM_0 wave excitation becomes appreciable when $h/\lambda > 0.09$ ($\epsilon_r \cong 2.3$) and when $h/\lambda > 0.03$ ($\epsilon_r \cong 10$). Generally, to suppress the TM_0 mode,

the dielectric constant should be lower and the substrate height should be smaller. Unfortunately, decreasing ϵ_r increases the antenna size, while decreasing h leads to smaller antenna efficiency and narrower frequency band.

5. Criteria for Substrate Selection

- 1) surface-wave excitation
- 2) dispersion of the dielectric constant and loss tangent of the substrate
- 3) copper loss
- 4) anisotropy in the substrate
- 5) effects of temperature, humidity, and aging
- 6) mechanical requirements: conformability, machinability, solderability, weight, elasticity, etc.
- 7) cost

The first 3 factors are of special concern in the millimeter-wave range ($f \geq 30$ GHz).

ELECTRICAL PROPERTIES OF COMMONLY USED SUBSTRATE MATERIALS
FOR MICROSTRIP ANTENNAS

Material	Dielectric Constant	Loss Tangent
Unreinforced PTFE, Cuflon	2.1	0.0004
Reinforced PTFE, RT Duroid 5880	2.20 (1.5%)	0.0009
Fused Quartz	3.78	0.0001
96% Alumina	9.40 (5%)	0.0010
99.5% Alumina	9.80 (5%)	0.0001
Sapphire	9.4, 1.6	0.0001
Semi-Insulating GaAs	12.9	0.0020

**NON-ELECTRICAL PROPERTIES OF COMMONLY USED SUBSTRATE MATERIALS
FOR MICROSTRIP ANTENNAS**

Properties	PTFE	Fused Quartz	Alumina	Sapphire	GaAs
temperature range (°C)	-55 – 260	< +1100	< +1600	-24 – 370	-55 – 260
Thermal conductivity (W/cm·K)	0.0026	0.017	0.35 to 0.37	0.42	0.46
coefficient of thermal expansion (ppm/K)	16.0 to 108.0	0.55	6.30 to 6.40	6.00	5.70
Temperature coefficient of dielectric constant (ppm/K)	+350.0 to +480.0	+13.0	+136.0	+110 to +140	-
minimum thickness (mil)	4	2	5	4	4
Machinability	Good	Very poor	Very poor	Poor	Poor
Solderability	Good	Good	Good	Good	Good
Dimensional Stability	Poor for unreinforced, very good for others	Good	Excellent	Good	Good
Cost	Very low	High	Low	-	Very high

6. Dispersion Effects in the Substrate

The dependence of the dielectric constant ϵ_r and the loss tangent on the frequency is referred to as frequency dispersion. For frequencies up to 100 GHz (the typical range for printed antennas is < 30 GHz), the dispersion of ϵ_r is practically negligible. The losses, however, display noticeable changes with frequency. In general, the loss increases with frequency.

7. Dielectric Loss and Copper Loss

The loss in the feed lines and the patches themselves are usually computed with formulas, which were first derived for microstrip transmission lines, i.e., the patch is treated as a wide piece of a microstrip line.

a) Dielectric loss (in dB per unit length, length is in the units used for λ_0)

$$\alpha_d = 27.3 \cdot \frac{\epsilon_r}{\sqrt{\epsilon_{r_{\text{eff}}}(f)}} \cdot \frac{[\epsilon_{r_{\text{eff}}}(f) - 1]}{(\epsilon_r - 1)} \cdot \frac{\tan \delta}{\lambda_0} \quad (4)$$

b) Copper loss (in dB per unit length)

$$\alpha_c = \begin{cases} 1.38 \cdot \frac{R'_s}{hZ_0} \cdot \left[\frac{32 - \left(\frac{W'}{h}\right)^2}{32 + \left(\frac{W'}{h}\right)^2} \right] \Lambda, & \text{for } \frac{W}{h} \leq 1 \\ 6.1 \times 10^{-5} \cdot \frac{R'_s Z_0 \epsilon_{r_{\text{eff}}}(f)}{h} \cdot \left[\frac{W'}{h} + \frac{0.667 \frac{W'}{h}}{\frac{W'}{h} + 1.444} \right] \Lambda, & \text{for } \frac{W}{h} \geq 1 \end{cases} \quad (5)$$

In the above equations:

- $\varepsilon_{r_{eff}}(f)$ is the effective dielectric constant (generally, dispersive). Its quasi-static (low frequency) expression [2] is

$$\varepsilon_{r_{eff}}(0) = \begin{cases} \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2} \cdot \left(1 + 12 \frac{h}{W}\right)^{-1/2}, & \text{for } W/h > 1 \\ \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2} \cdot \left[\left(1 + 12 \frac{h}{W}\right)^{-1/2} + 0.04 \left(1 - \frac{W}{h}\right)^2 \right], & \text{for } \frac{W}{h} \leq 1 \end{cases} \quad (6)$$

Alternative expression for the quasi-static approximation of $\varepsilon_{r_{eff}}$ can be found in [5].

The quasi-static expressions need a dispersion correction for frequencies higher than 8 GHz. One possible correction is based on an empirical formula for the dispersive phase velocity in a microstrip line [5]. We first compute a normalized frequency (normalized with respect to the cut-off of the TE₁ mode):

$$\bar{f} = \frac{f}{f_c^{(1)}} = \frac{4h\sqrt{\varepsilon_r - 1}}{\lambda_0}. \quad (7)$$

Then, the dispersive phase velocity is calculated as

$$v_p = \frac{1}{\sqrt{\varepsilon_0 \varepsilon_{r_{eff}}(0)}} \cdot \frac{\bar{f}^2 \sqrt{\varepsilon_{r_{eff}}(0)} + \sqrt{\varepsilon_r}}{\bar{f}^2 + 1}. \quad (8)$$

Finally,

$$\varepsilon_{r_{eff}}(f) = (c / v_p)^2. \quad (9)$$

For alternative formulas, refer to [5].

- Z_0 is the characteristic impedance of the microstrip line (generally, dispersive):

$$Z_0 = \begin{cases} \frac{120\pi\sqrt{\varepsilon_{r_{eff}}}}{\frac{W}{h} + 1.393 + 0.667 \ln\left(\frac{W}{h} + 1.444\right)}, & \text{for } \frac{W}{h} \geq 1 \\ \frac{60}{\sqrt{\varepsilon_{r_{eff}}}} \cdot \ln\left(\frac{8h}{W} + 0.25\frac{W}{h}\right), & \text{for } \frac{W}{h} \leq 1 \end{cases} \quad (10)$$

- Λ is a constant dependent on the strip thickness t :

$$\Lambda = \begin{cases} 1 + \frac{h}{W'} \left[1 + \frac{1.25t}{\pi W} + \frac{1.25}{\pi} \ln\left(\frac{4\pi W}{t}\right) \right], & \text{for } \frac{W}{h} \leq \frac{1}{2\pi} \\ 1 + \frac{h}{W'} \left[1 - \frac{1.25t}{\pi W} + \frac{1.25}{\pi} \ln\left(\frac{2t}{t}\right) \right], & \text{for } \frac{W}{h} \geq \frac{1}{2\pi} \end{cases} \quad (11)$$

- W' is the effective strip width:

$$\frac{W'}{h} = \begin{cases} \frac{W}{h} + \frac{1.25t}{\pi h} \left[1 + \ln\left(\frac{4\pi W}{t}\right) \right], & \text{for } \frac{W'}{h} \leq \frac{1}{2\pi} \\ \frac{W}{h} + \frac{1.25t}{\pi h} \left[1 + \ln\left(\frac{2h}{t}\right) \right], & \text{for } \frac{W'}{h} \geq \frac{1}{2\pi} \end{cases} \quad (12)$$

- R'_s is the effective surface resistance of the conductor:

$$R'_s = R_s \left\{ 1 + \frac{2}{\pi} \arctan \left[1.4 \left(\frac{\Lambda}{\delta} \right)^2 \right] \right\}, \Omega \quad (13)$$

where $R_s = \sqrt{\pi f \mu / \sigma}$ is the high-frequency surface resistance of the conductor. R_s relates to the skin-depth δ as $R_s = (\delta \sigma)^{-1}$. For a uniform surface current distribution over a conducting rod of length l and perimeter of its cross-section P , the resultant resistance is

$$R_{hf} = R_s \cdot l / P, \Omega. \quad (14)$$

Finally, the total loss is the sum of the conduction and dielectric losses:

$$\alpha_t = \alpha_d + \alpha_c. \quad (15)$$

SURFACE RESISTANCE AND SKIN-DEPTH OF COMMONLY USED CONDUCTORS

Metal		R_s [Ohm/square $\times 10^{-7}f$]		Skin-depth at 2 GHz [μm]
Silver	Ag	2.5	$\sigma = 6.1 \times 10^7 \text{S/m}$	1.4
Copper	Cu	2.6	$\sigma = 5.8 \times 10^7 \text{S/m}$	1.5
Gold	Au	3.0	$\sigma = 4.1 \times 10^7 \text{S/m}$	1.7
Aluminum	Al	3.3	$\sigma = 3.5 \times 10^7 \text{S/m}$	1.9

Some References

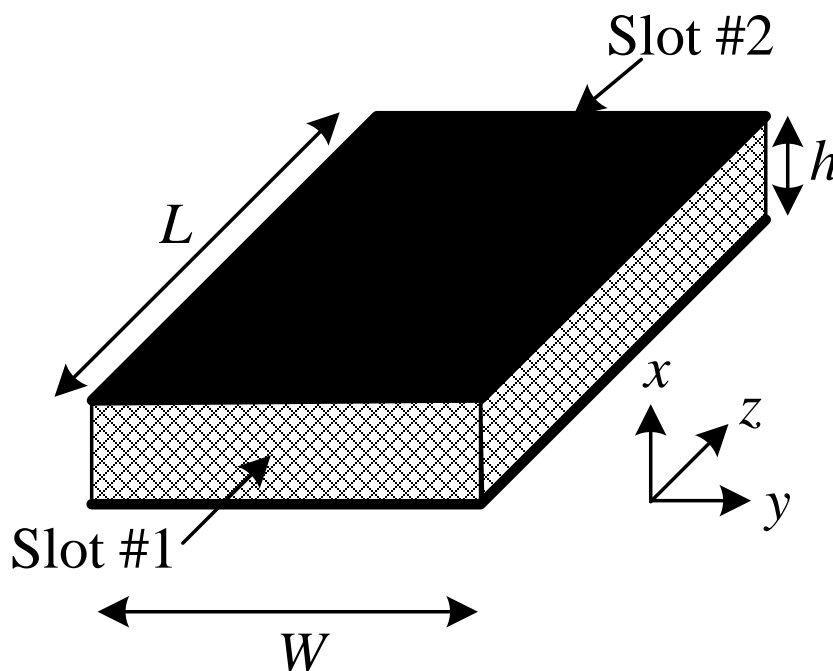
- [1] D. M. Pozar and D. H. Schaubert, eds., *Microstrip Antennas*, IEEE Press, 1995 (a collection of significant manuscripts on microstrip antennas).
- [2] R. A. Sainati, *CAD of Microstrip Antennas for Wireless Applications*, Artech, 1996 (comes with some CAD freeware).
- [3] P. Bhartia, K. V. S. Rao, and R. S. Tomar, *Millimeter-wave Microstrip and Printed Circuit Antennas*, Artech, 1991.
- [4] J.-F. Zürcher and F. E. Gardiol, *Broadband Patch Antennas*, Artech, 1995.
- [5] K. C. Gupta *et al.*, *Microstrip Lines and Slotlines*, 2nd ed., Artech, 1996.

LECTURE 21: MICROSTRIP ANTENNAS – PART II

(Transmission-line model. Design procedure for a rectangular patch. Cavity model for a rectangular patch.)

1. Transmission Line Model – The Rectangular Patch

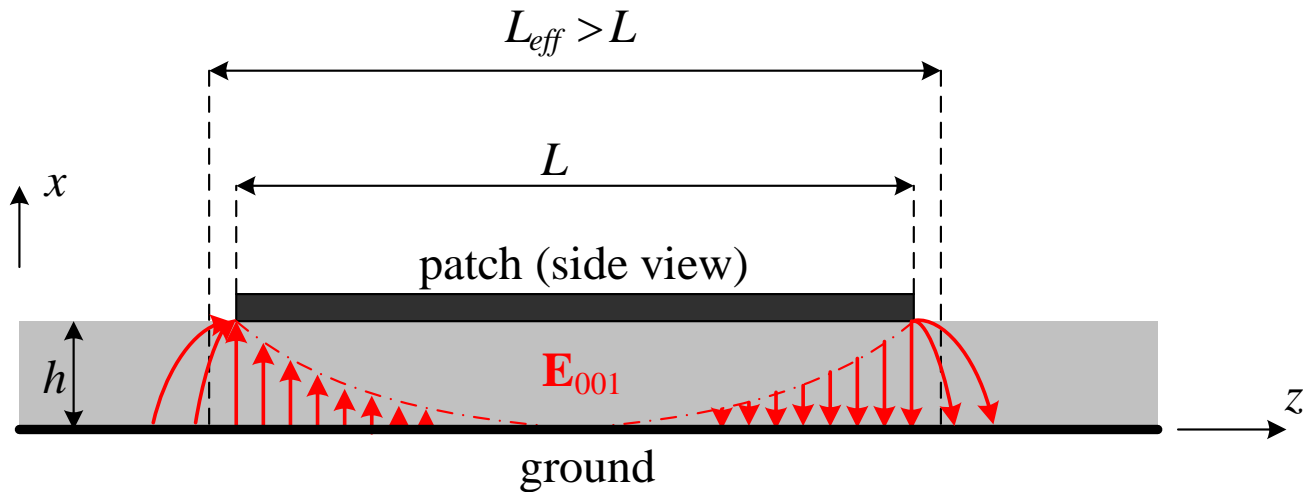
The TL model is the simplest of all, representing the rectangular patch as a parallel-plate transmission line connecting two radiating slots (apertures), each of width W and height h . In the figure below, z is the direction of propagation of the transmission line.



The TL model is not accurate and lacks versatility. However, it gives a relatively good physical insight into the nature of the patch antenna and the field distribution for all TM_{00n} modes.

The slots represent very high-impedance terminations on both sides of the transmission line (almost an open circuit). Thus, the patch has highly resonant characteristics depending crucially on its length L along z . The resonant length of the patch, however, is not exactly equal to the physical length due to the fringing effect. The fringing effect makes the effective electrical length of the patch longer than its physical length, $L_{eff} > L$. Thus, the resonance condition

$\beta^{(n)}L_{eff} = n \cdot \pi / 2$, $n = 1, 2, \dots$, depends on L_{eff} , not L . A sketch of the **E**-field distribution for the first (dominant) resonant mode, $n = 1$, is shown in the figure below.

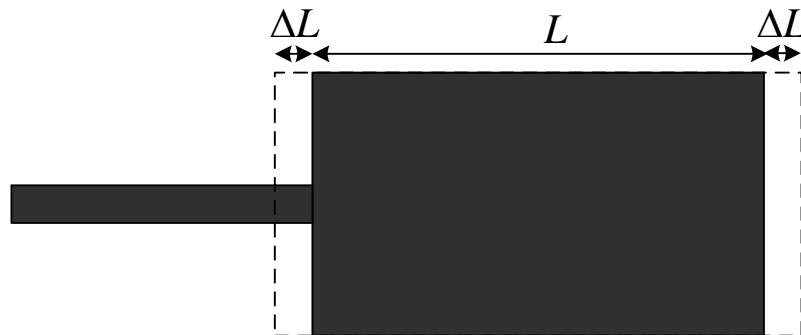


A. Computing the effective patch length

$$\frac{\Delta L}{h} = 0.412 \frac{(\epsilon_{r_{eff}} + 0.3) \left(\frac{W}{h} + 0.264 \right)}{(\epsilon_{r_{eff}} - 0.258) \left(\frac{W}{h} + 0.8 \right)}. \quad (21.1)$$

For the computation of $\epsilon_{r_{eff}}$, see previous Lecture. The effective length is

$$L_{eff} = L + 2\Delta L. \quad (21.2)$$



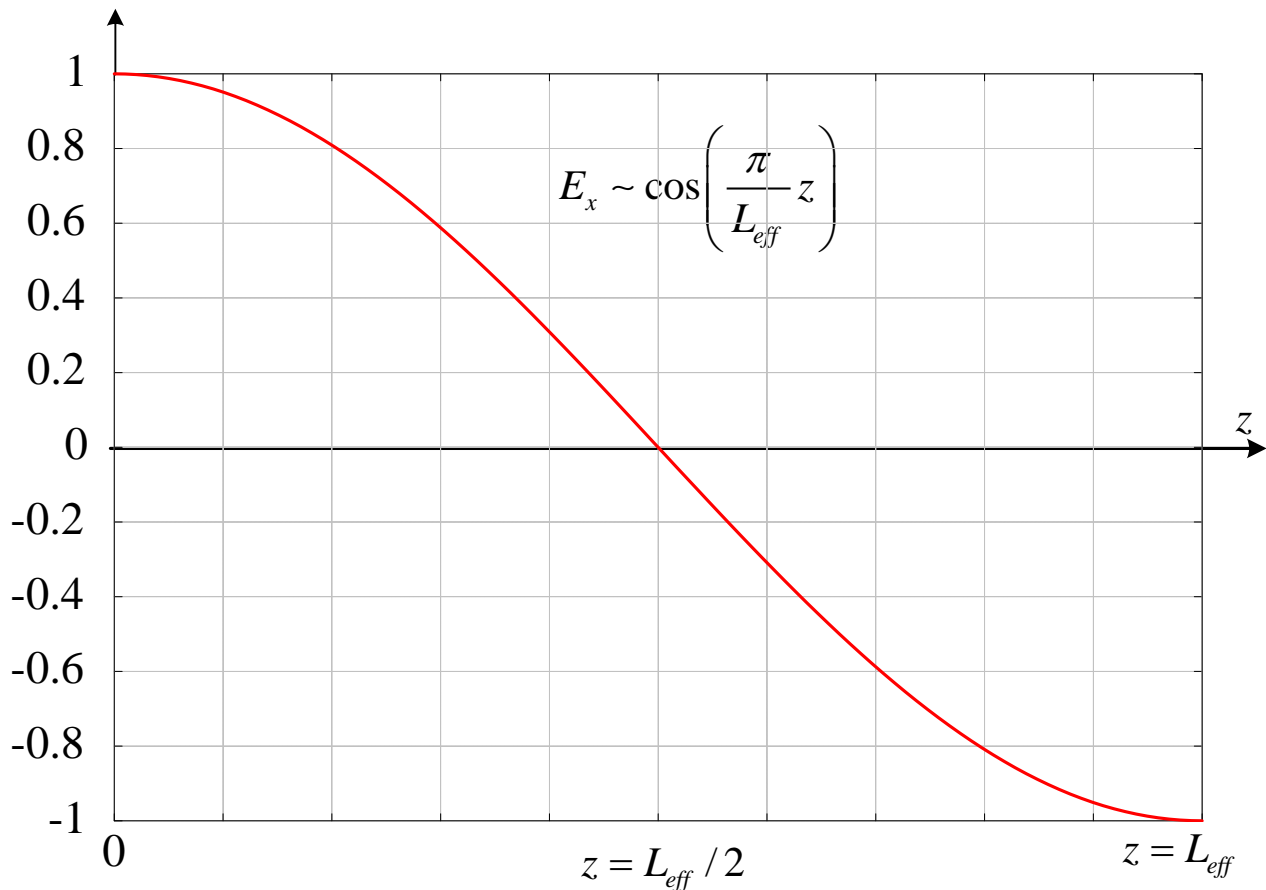
B. Resonant frequency of the dominant TM₀₀₁ mode

$$L_{eff} = \frac{\lambda_0}{2} = \frac{v}{2f^{(001)}} = \frac{c}{2\sqrt{\epsilon_{r_{eff}}} f^{(001)}} \Rightarrow f_r^{(001)} = \frac{c}{2L_{eff} \sqrt{\epsilon_{r_{eff}}}} \quad (21.3)$$

The resonant frequency of a patch depends strongly on L , therefore, the exact calculation of L_{eff} is necessary to predict the antenna resonance:

$$\boxed{f_r^{(001)} = \frac{c}{2\sqrt{\epsilon_{r_{eff}}} (L + 2\Delta L)}} \quad (21.4)$$

The field of the TM₀₀₁ mode does not depend on the x and y coordinates but it strongly depends on the z coordinate, along which a standing wave is formed. The figure below shows the vertical \mathbf{E} -field distribution along z when the patch is in resonance.



C. The patch width W

$$W = \frac{1}{2f_r \sqrt{\mu_0 \epsilon_0}} \sqrt{\frac{2}{\epsilon_r + 1}} = \frac{c}{2f_r} \sqrt{\frac{2}{\epsilon_r + 1}} \quad (21.5)$$

Expression (21.5) makes the width W equal to about half a wavelength. It leads to good radiation efficiencies and acceptable dimensions.

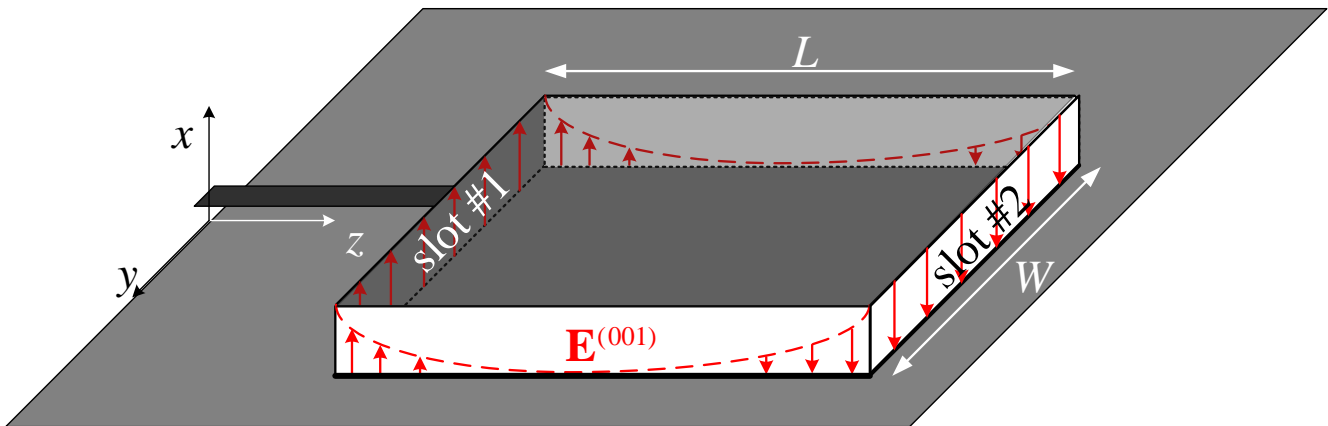
D. Equivalent circuit of the patch

The dominant TM_{001} mode has a uniform field distribution along the y -axis at the slots formed at the front and end edges of the patch. The equivalent conductance G is obtained from the theory of uniform apertures while B is related to the fringe capacitance:

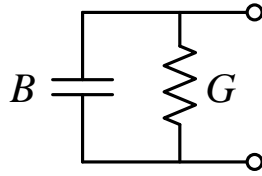
$$G = \frac{W}{120\lambda_0} \left[1 - \frac{1}{24} \left(\frac{2\pi h}{\lambda_0} \right)^2 \right], \quad \text{for } \frac{h}{\lambda_0} < \frac{1}{10}, \quad (21.6)$$

$$B = \frac{W}{120\lambda_0} \left[1 - 0.636 \ln \left(\frac{2\pi h}{\lambda_0} \right)^2 \right], \quad \text{for } \frac{h}{\lambda_0} < \frac{1}{10}. \quad (21.7)$$

The limitation $(h/\lambda_0) < 0.1$ is necessary since a uniform field distribution along the x -axis is assumed. The patch has two radiating slots (see the figure below).



The equivalent circuit of a slot is constructed as a parallel R - C circuit, using the values computed by (21.6) and (21.7):



$G = 1/R$ represents the radiation loss, while $B = j\omega C$ is the equivalent susceptance, which represents the capacitance of the slot.

More accurate values for the conductance G can be obtained through the cavity model:

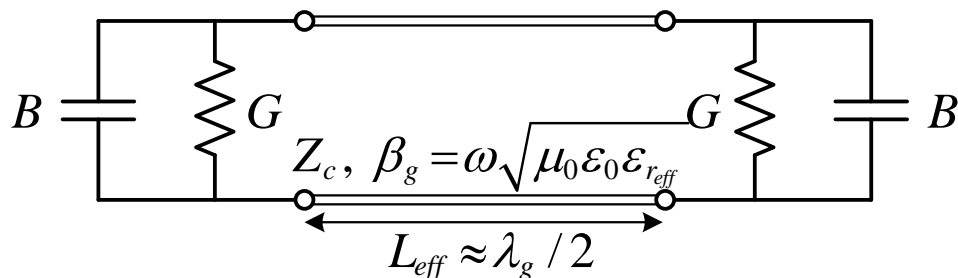
$$G = I / (120\pi^2), \quad (21.8)$$

where

$$I = \int_0^\pi \left[\frac{\sin(0.5k_0W \cos \theta)}{\cos \theta} \right]^2 \sin^3 \theta d\theta = -2 + \cos X + X \cdot S_i(X) + \frac{\sin X}{X}, \quad (21.9)$$

and $X = k_0W$, $k_0 = \omega\sqrt{\mu_0\epsilon_0}$. S_i denotes the sine integral, $S_i(x) = \int_0^x (\sin y) / y dy$.

The equivalent circuit representing the whole patch in the TM_{001} mode includes the two radiating slots as parallel R - C circuits and the patch connecting them as a transmission line, the characteristics of which are computed in the same way as those of a microstrip line.

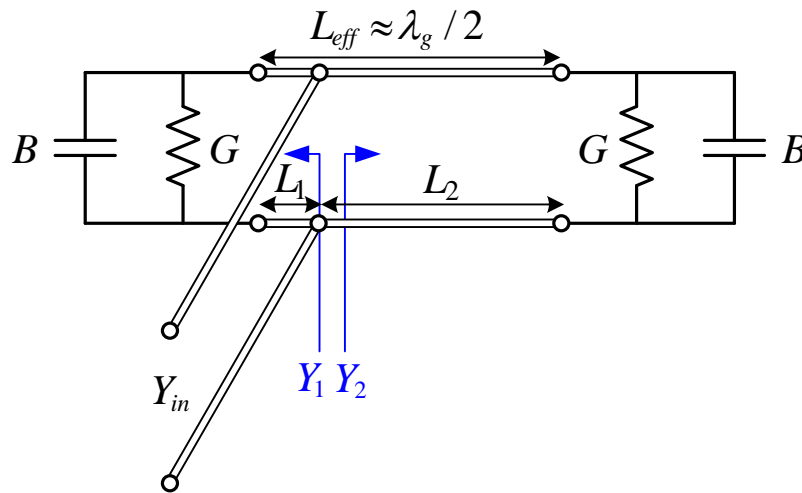


Here, Z_c is the characteristic impedance of the line, and β_g is its phase constant. When the losses are not neglected, we must include also the attenuation constant α (see Lecture 20). For each slot, G represents the radiation loss and $B = \omega C$ represents the capacitance associated with the fringe effect.

E. Resonant input resistance

When the patch is resonant, the susceptances of both slots cancel out at the feed point regardless of the position of the feed along the patch. Thus, the input admittance is always purely real. This real value, however, strongly depends on the feed position along z . This is easily shown through the Smith chart for the admittance transformation through a transmission line.

At the feed point, the impedance of each slot is transformed by the respective transmission line representing a portion of the patch:

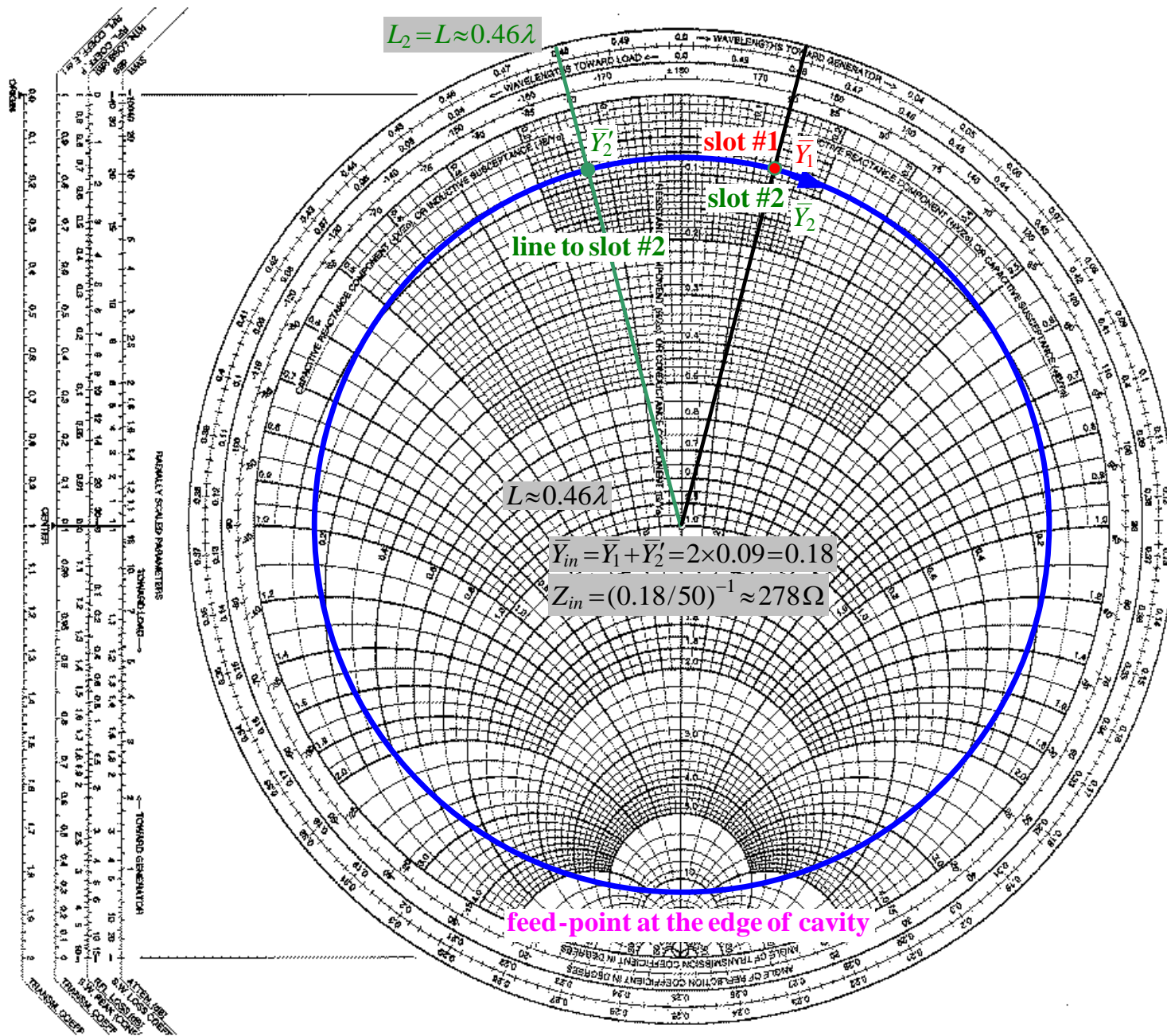


$$Y_{in} = Y_1 + Y_2 \quad (21.10)$$

The admittance transformation is given by

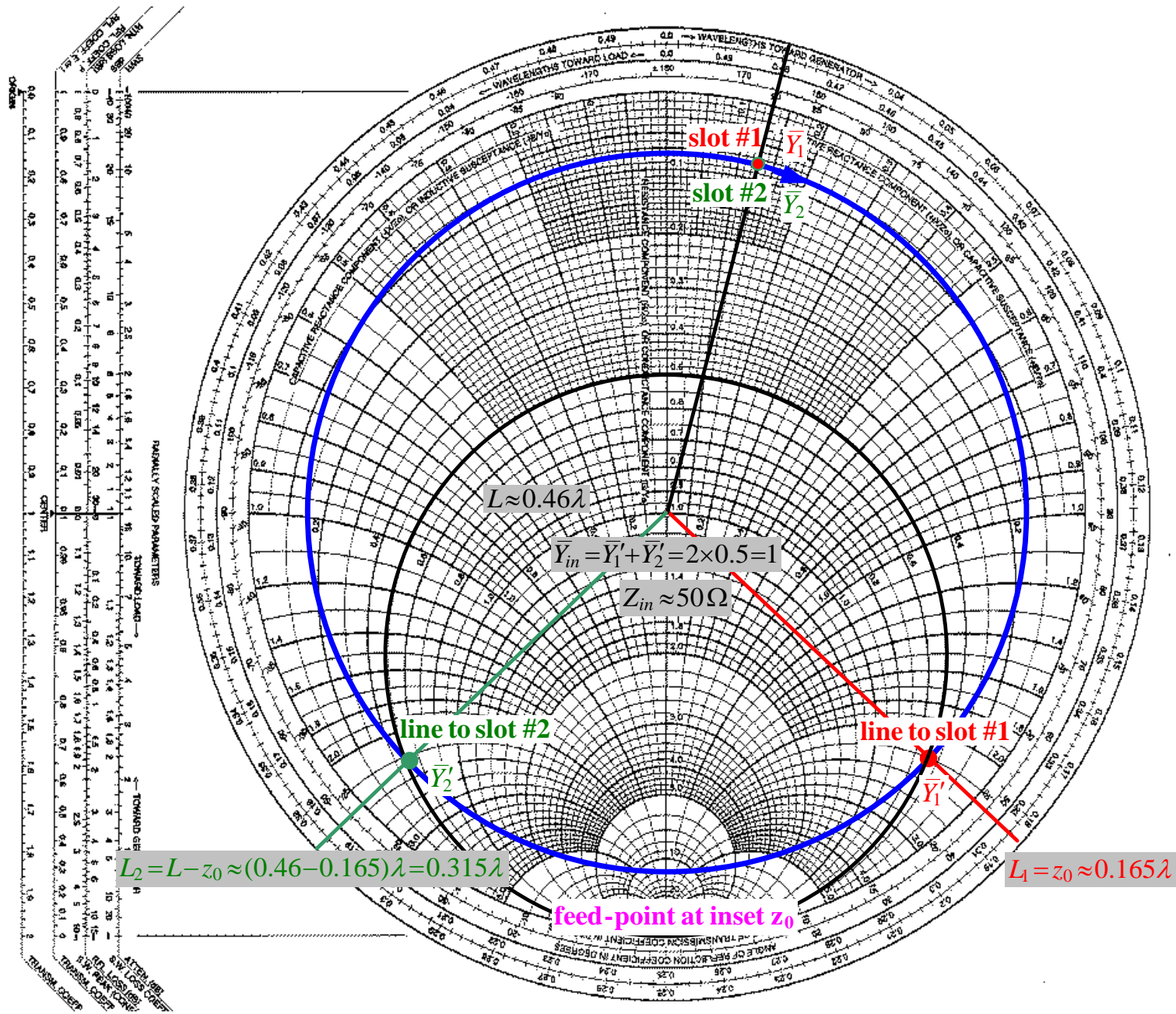
$$Y_{in} = Y_c \left[\frac{Y_L + jY_c \tan(\beta_g L)}{Y_c + jY_L \tan(\beta_g L)} \right] = Y_L \Big|_{\beta_g L = \pi}, \quad Y_c = Z_c^{-1} \quad (21.11)$$

if the line is loss-free. Below, the Smith charts illustrate the slot-impedance transformations and their addition, which produces a real normalized admittance, in three cases: (1) the patch is fed at one edge ($L_1 = 0$, $L_2 = L$), (2) the patch is fed at the center ($L_1 = L_2 = L/2$), and (3) the patch is fed at a distance (feed inset) $z_0 = 0.165\lambda$.

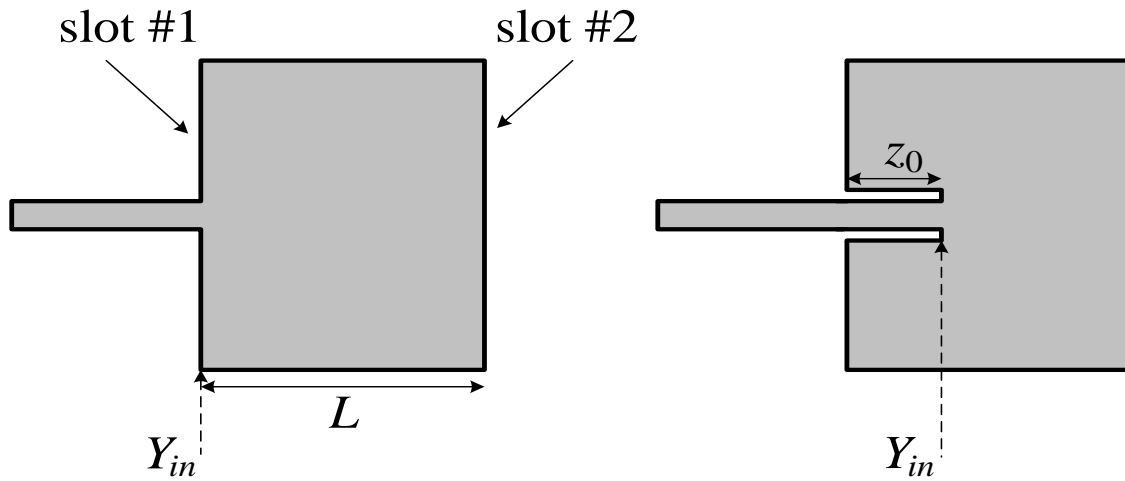


Smith Chart

Smith Chart



The edge feed and the inset feed are illustrated below.



The two slots are separated by an electrical distance of 180° . However, because of the fringe effect the physical length L is slightly less than $\lambda/2$. The reduction of the length is not much. Typically, it is $0.48\lambda \leq L \leq 0.49\lambda$.

Ideally, the resonant input impedance of the patch for the dominant TM_{001} mode is entirely resistive and equal to half the transformed resistance of each slot:

$$Z_{in} = \frac{1}{Y_{in}} = \frac{1}{2G'_1} = R_{in}. \quad (21.12)$$

In reality, there is some mutual influence between the two slots, described by a mutual conductance and it should be included for more accurate calculations:

$$R_{in} = \frac{1}{2(G'_1 \pm G_{12})}, \quad (21.13)$$

where the “+” sign relates to the odd modes, while the “-” sign relates to the even modes. Normally, $G_{12} \ll G'_1$.

For most patch antennas fed at the edge, R_{in} is greater than the characteristic impedance Z_c of the microstrip feed line (typically $Z_c = 50$ to 75Ω). That is why, the inset-feed technique is widely used to achieve impedance match.

The figure below illustrates the normalized input impedance of a 1-D (along the y axis) loss-free open-ended transmission-line, the behavior of which is very close to that of the dominant mode of the patch.

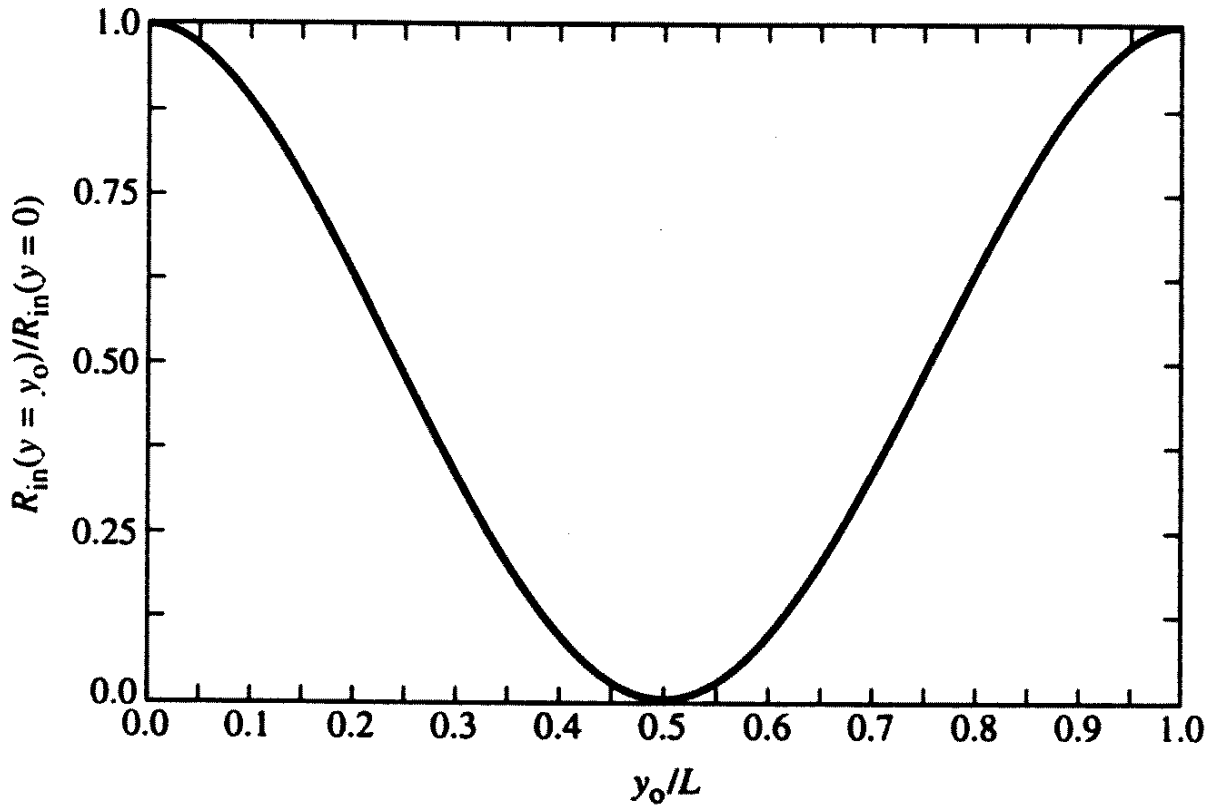


Fig 14.14, pp 735, C. Balanis

Using modal expansion, the input resistance for the inset-feed at $z = z_0$ is given approximately by

$$R_{in} = \frac{1}{2(G_1 \pm G_{12})} \left[\cos^2 \left(\frac{\pi}{L} z_0 \right) + \frac{G_1^2 + B_1^2}{Y_c^2} \sin^2 \left(\frac{\pi}{L} z_0 \right) - \frac{B_1}{Y_c} \sin \left(\frac{2\pi}{L} z_0 \right) \right]. \quad (21.14)$$

Here, G_1 and B_1 are calculated using (21.6) and (21.7). For most feeding microstrips, $G_1 / Y_c \ll 1$ and $B_1 / Y_c \ll 1$. Then,

$$R_{in} = \frac{1}{2(G_1 \pm G_{12})} \cos^2 \left(\frac{\pi}{L} y_0 \right) = R_{in(z=0)} \cos^2 \left(\frac{\pi}{L} y_0 \right). \quad (21.15)$$

Notice that the inset feeding technique for impedance match of the microstrip antennas is conceptually analogous to the off-center or asymmetrical feeding techniques for dipoles. In both cases, a position is sought along a resonant structure, where the current magnitude has the desired value.

2. Designing a Rectangular Patch Using the Transmission Line Model

Input data: ϵ_r, h, f_r

- 1) Calculate W using (21.5).
- 2) Calculate ϵ_{reff} using (21.5) and equation (6) from Lecture 20.
- 3) Calculate the extension ΔL due to the fringing effect using (21.1).
- 4) Calculate the actual (physical) length of the patch using

$$L = \frac{\lambda_0}{2} - 2\Delta L \text{ or } L = \frac{1}{2f_r \sqrt{\epsilon_{reff}} \sqrt{\mu_0 \epsilon_0}} - 2\Delta L. \quad (21.16)$$

- 5) Calculate radiating slot admittance using (21.6) and (21.7).
- 6) Calculate resonant input resistance at patch edge using (21.12) or (21.13) with $G'_1 = G$ from (21.6).
- 7) If R_{in} calculated in step 6 is too large, calculate the inset distance z_0 using (21.14) or (21.15).

3. Cavity Model for the Rectangular Patch

The TL model is very limited in its description of the real processes taking place when a patch is excited. It takes into account only the TM_{00n}^x modes where the energy propagates only in the longitudinal z direction. The field distribution along the x and y axes is assumed uniform. It is true that the dominant TM_{001}^x is prevalent but the performance of the patch is also affected by higher-order modes.

The cavity model is a more general model of the patch which imposes open-end conditions at the side edges of the patch. It represents the patch as a dielectric-loaded cavity with:

- electrical walls (above and below), and
- magnetic walls (around the perimeter of the patch).

The magnetic wall is a wall at which

$$\hat{\mathbf{n}} \times \mathbf{H} = 0 \text{ (the } \mathbf{H}\text{-field is purely normal)}$$

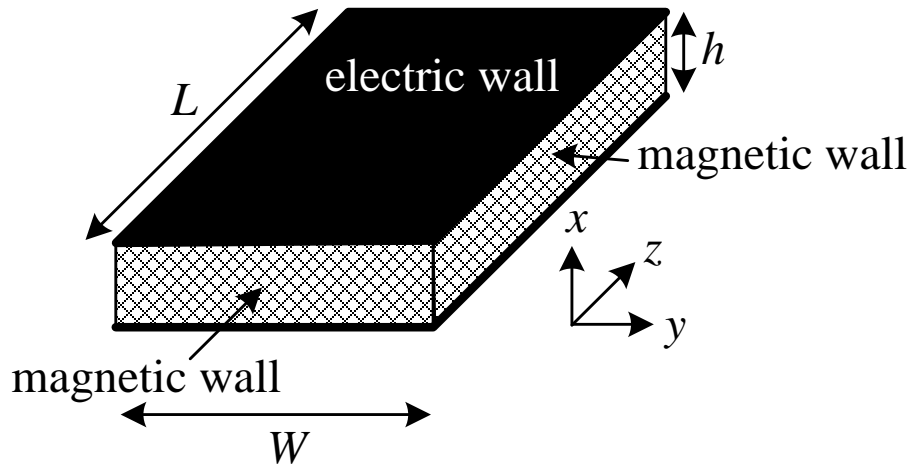
$$\hat{\mathbf{n}} \cdot \mathbf{E} = 0 \text{ (the } \mathbf{E}\text{-field is purely tangential)}$$

It is analogous to the open end termination in the theory of transmission lines.

If we treat the microstrip antenna only as a cavity, we can not represent

radiation because an ideal loss-free cavity does not radiate and its input impedance is purely reactive. To account for the radiation, a loss mechanism is introduced. This is done by introducing an effective loss tangent, δ_{eff} .

The thickness of the substrate is very small. The waves generated and propagating beneath the patch undergo considerable reflection at the edges of the patch. Only a very small fraction of them is being radiated. Thus, the antenna is quite inefficient. The cavity model assumes that the \mathbf{E} field is purely tangential to the slots formed between the ground plane and the patch edges (magnetic walls). Moreover, it considers only TM^x modes, i.e., modes with no H_x component. These assumptions are, basically, very much true.



The TM^x modes are fully described by a single scalar function A_x – the x -component of the magnetic vector potential:

$$\mathbf{A} = A_x \hat{\mathbf{x}}. \quad (21.17)$$

In a homogeneous source-free medium, A_x satisfies the wave equation:

$$\nabla^2 A_x + k^2 A_x = 0. \quad (21.18)$$

For regular shapes (like the rectangular cavity), it is advantageous to use the separation of variables:

$$\frac{\partial^2 A_x}{\partial x^2} + \frac{\partial^2 A_x}{\partial y^2} + \frac{\partial^2 A_x}{\partial z^2} + k^2 A_x = 0 \quad (21.19)$$

$$A_x = X(x)Y(y)Z(z) \quad (21.20)$$

$$YZ \frac{\partial^2 X}{\partial x^2} + XZ \frac{\partial^2 Y}{\partial x^2} + XY \frac{\partial^2 Z}{\partial x^2} = -k^2 XYZ$$

$$\frac{1}{X} \frac{\partial^2 X}{\partial x^2} + \frac{1}{Y} \frac{\partial^2 Y}{\partial y^2} + \frac{1}{Z} \frac{\partial^2 Z}{\partial z^2} = -k^2 \quad (21.21)$$

$$\frac{d^2 X}{dx^2} + k_x^2 X = 0, \quad \frac{d^2 Y}{dy^2} + k_y^2 Y = 0, \quad \frac{d^2 Z}{dz^2} + k_z^2 Z = 0 \quad (21.22)$$

The eigenvalue equation is

$$k_x^2 + k_y^2 + k_z^2 = k^2. \quad (21.23)$$

The solutions of (21.22) are harmonic functions:

$$X(x) = \sum_n A_n^c \cos(k_{xn}x) + A_n^s \sin(k_{xn}x),$$

$$Y(y) = \sum_n B_n^c \cos(k_{yn}y) + B_n^s \sin(k_{yn}y), \quad (21.24)$$

$$Z(z) = \sum_n C_n^c \cos(k_{zn}z) + C_n^s \sin(k_{zn}z).$$

When the functions in (21.24) are substituted in (21.20), they give the general solution of (21.18). The particular solution of (21.18) depends on the boundary conditions.

In our case, there are electric walls at $x = 0$ and $x = h$. There, the tangential **E**-field components must vanish, i.e., $E_y = E_z = 0|_{x=0,h}$. Having in mind that

$$E_x = \frac{1}{j\omega\mu\varepsilon} \left(\frac{\partial^2 A_x}{\partial x^2} + k^2 A_x \right), \quad E_y = \frac{1}{j\omega\mu\varepsilon} \left(\frac{\partial^2 A_x}{\partial x \partial y} \right), \quad E_z = \frac{1}{j\omega\mu\varepsilon} \left(\frac{\partial^2 A_x}{\partial x \partial z} \right), \quad (21.25)$$

we set A_x at the top and bottom walls as

$$\left. \frac{\partial A_x}{\partial x} \right|_{x=0,h} = 0. \quad (21.26)$$

At all side walls, we set a vanishing normal derivative for A_x :

$$\left. \frac{\partial A_x}{\partial z} \right|_{z=0,L} = 0, \quad \left. \frac{\partial A_x}{\partial y} \right|_{y=0,W} = 0. \quad (21.27)$$

This ensures vanishing H_x and H_y at $z=0$ and $z=L$, as well as vanishing H_x and H_z at $y=0$ and $y=W$ (magnetic walls), as follows from the relation between the \mathbf{H} -field and A_x ,

$$H_x = 0, H_y = \frac{1}{\mu} \left(\frac{\partial A_x}{\partial z} \right), H_z = \frac{1}{\mu} \left(\frac{\partial A_x}{\partial y} \right). \quad (21.28)$$

It is now obvious that the solution must appear in terms of the functions

$$\begin{aligned} X(x) &= \sum_n A_n^c \cos(k_{xn}x), \quad k_{xn} = n \frac{\pi}{h}, \\ Y(y) &= \sum_n B_n^c \cos(k_{yn}y), \quad k_{yn} = n \frac{\pi}{W}, \\ Z(z) &= \sum_n C_n^c \cos(k_{zn}z), \quad k_{zn} = n \frac{\pi}{L}. \end{aligned} \quad (21.29)$$

The spectrum of the eigenmodes in the cavity is discrete. The frequencies of those modes (the resonant frequencies) can be calculated from (21.23) as

$$\left(\frac{m\pi}{h} \right)^2 + \left(\frac{n\pi}{W} \right)^2 + \left(\frac{p\pi}{L} \right)^2 = \left(\omega_r^{(mnp)} \right)^2 \mu \epsilon, \quad (21.30)$$

$$\boxed{f_r^{(mnp)} = \frac{1}{2\pi\sqrt{\mu\epsilon}} \sqrt{\left(\frac{m\pi}{h} \right)^2 + \left(\frac{n\pi}{W} \right)^2 + \left(\frac{p\pi}{L} \right)^2}}. \quad (21.31)$$

The mode with the lowest resonant frequency is the **dominant mode**. Since usually $L > W$, the lowest-frequency mode is the TM_{001}^x mode, for which

$$f_r^{(001)} = \frac{1}{2\pi\sqrt{\mu\epsilon}} \frac{\pi}{L} = \frac{c}{2L\sqrt{\epsilon_r}}. \quad (21.32)$$

The dominant TM_{001}^x mode is exactly the mode considered by the transmission-line model (see previous sections). The field distribution of some low-order modes is given in the following figure.

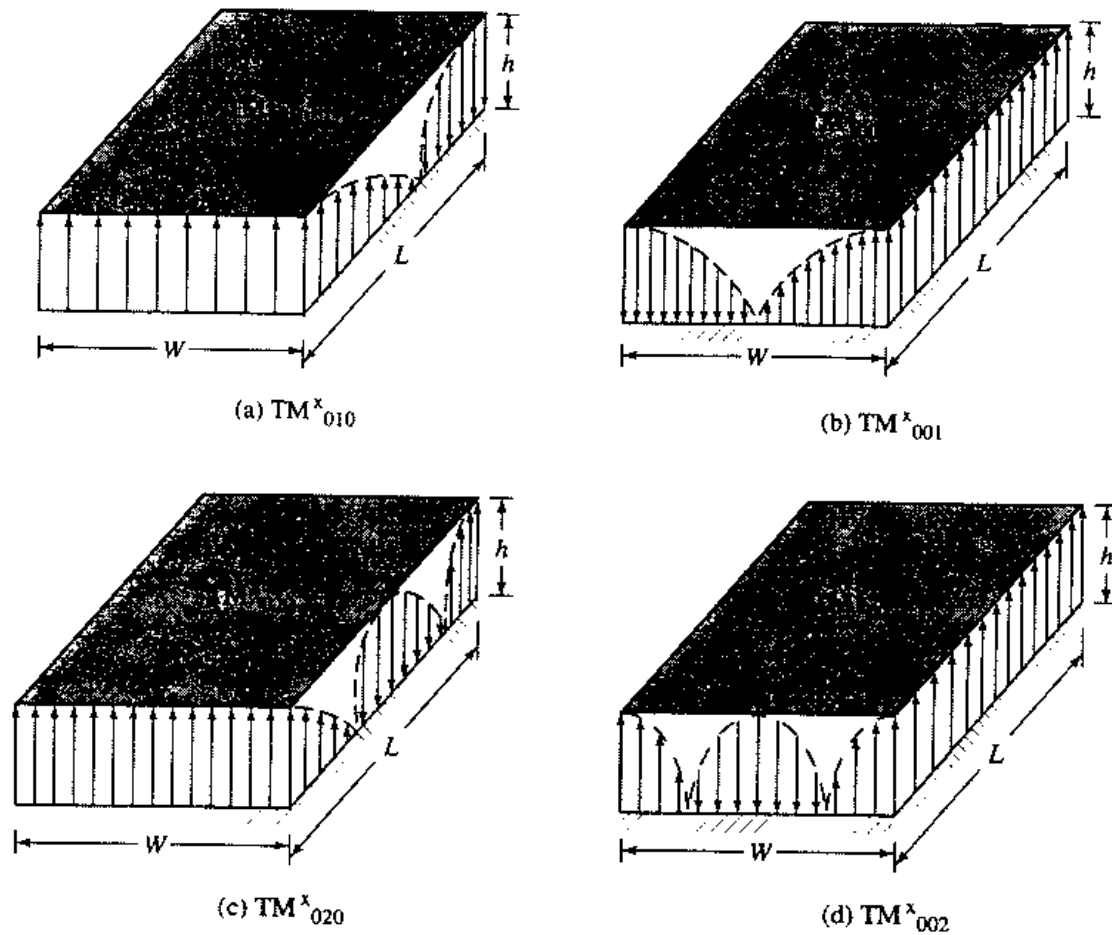


Fig. 14.13, pp. 741, Balanis

The general solution for the $A_x^{(mnp)}$ [see (21.20) and (21.24)] is

$$A_x^{(mnp)} = \left[A_x^c \cos\left(m \frac{\pi}{h} x\right) \right] \left[B_x^c \cos\left(n \frac{\pi}{W} y\right) \right] \left[C_x^c \cos\left(p \frac{\pi}{L} z\right) \right], \quad (21.33)$$

or

$$A_x^{(mnp)} = A_{mnp} \cdot \cos\left(m \frac{\pi}{h} x\right) \cdot \cos\left(n \frac{\pi}{L} y\right) \cdot \cos\left(p \frac{\pi}{W} z\right). \quad (21.34)$$

The respective field solution for the (m,n,p) mode is

$$E_x = -j \frac{(k^2 - k_x^2)}{\omega \mu \epsilon} A_{mnp} \cdot \cos(k_x x) \cdot \cos(k_y y) \cdot \cos(k_z z), \quad (21.35)$$

$$E_y = -j \frac{k_x k_y}{\omega \mu \epsilon} A_{mnp} \cdot \sin(k_x x) \cdot \sin(k_y y) \cdot \cos(k_z z), \quad (21.36)$$

$$E_z = -j \frac{k_x k_z}{\omega \mu \epsilon} A_{mnp} \cdot \sin(k_x x) \cdot \cos(k_y y) \cdot \sin(k_z z), \quad (21.37)$$

$$H_x = 0, \quad (21.38)$$

$$H_y = -\frac{k_z}{\mu} A_{mnp} \cdot \cos(k_x x) \cdot \cos(k_y y) \cdot \sin(k_z z), \quad (21.39)$$

$$H_z = \frac{k_y}{\mu} A_{mnp} \cdot \cos(k_x x) \cdot \sin(k_y y) \cdot \cos(k_z z). \quad (21.40)$$

For the dominant TM_{001}^x mode,

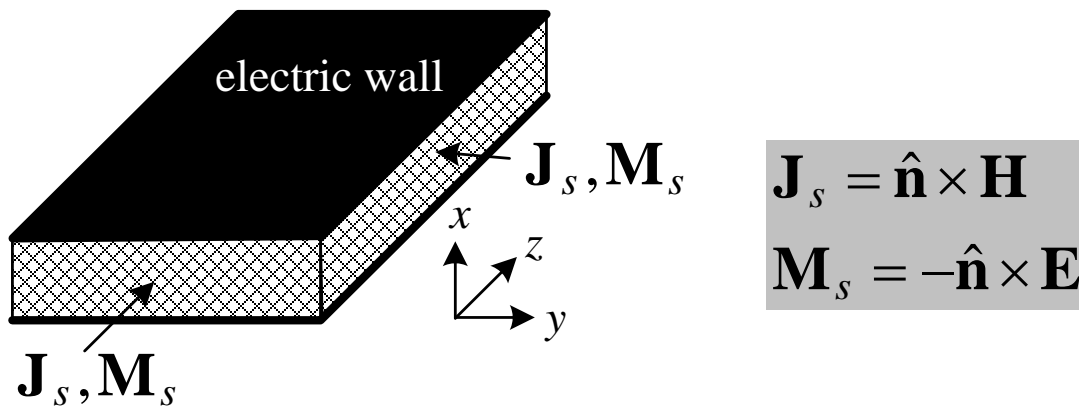
$$E_x = \left[-j(k^2 - \pi^2 / h^2) / (\omega \mu \epsilon) \right] A_{001} \cos(\pi z / L), \quad E_y = E_z = 0, \quad (21.41)$$

$$H_y = -(\pi / \mu L) A_{001} \sin(\pi z / L), \quad H_x = H_z = 0. \quad (21.42)$$

4. Cavity Model for the Radiated Field of a Rectangular Patch

The microstrip patch is represented by the cavity model reasonably well assuming that the material of the substrate is truncated and does not extend beyond the edges of the patch. The four side walls (the magnetic walls) represent four narrow apertures (slots) through which radiation takes place.

The equivalence principle is used to calculate the radiation fields. The field inside the cavity is assumed equal to zero, and its influence on the field in the infinite region outside is represented by the equivalent surface currents on the surface of the cavity.



Because of the very small height h of the substrate, the field is concentrated beneath the patch. There is some actual electrical current at the top metallic plate, however, its contribution to radiation is negligible. That is because: (1) it is backed by a conductor, and (2) it is very weak compared to the equivalent currents at the slots. The actual electrical current density of the top patch is maximum at the edges of the patch.

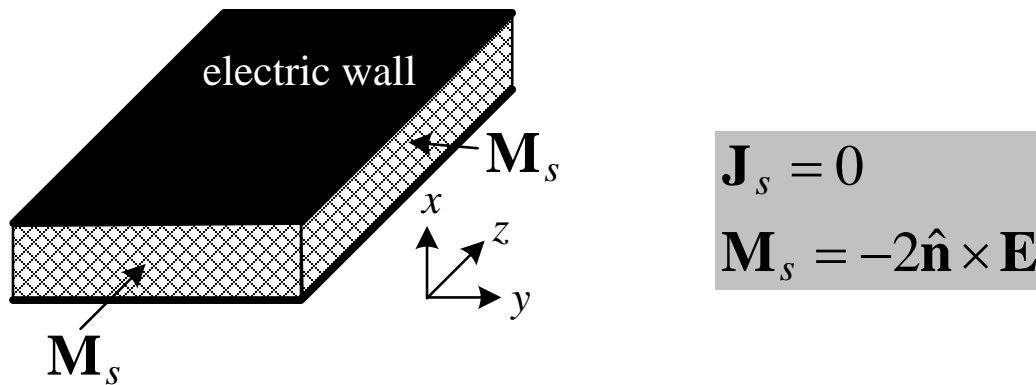
In the cavity model, the side walls employ magnetic-wall boundary condition, which sets the tangential \mathbf{H} components at the slots equal to zero. Therefore,

$$\mathbf{J}_s = \hat{\mathbf{n}} \times \mathbf{H} = 0. \quad (21.43)$$

Only the equivalent magnetic current density

$$\mathbf{M}_s = -\hat{\mathbf{n}} \times \mathbf{E} \quad (21.44)$$

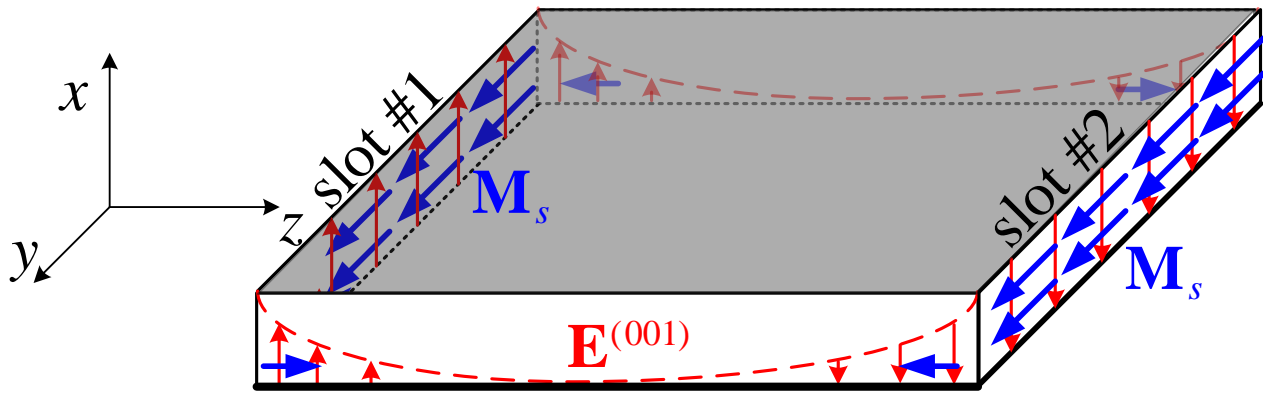
has substantial contribution to the radiated field.



The influence of the infinite ground plane is accounted for by the image theory, according to which the currents \mathbf{M}_s in the presence of the infinite plane radiate as if magnetic currents of double strength radiate in free space:

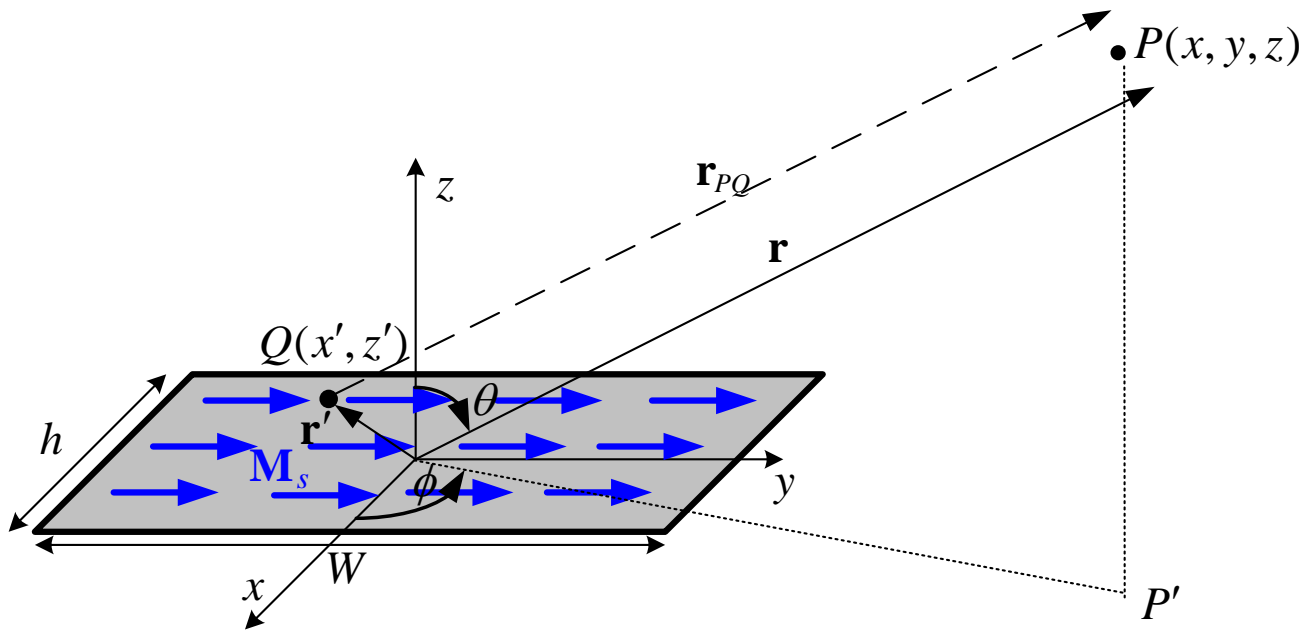
$$\mathbf{M}_s = -2\hat{\mathbf{n}} \times \mathbf{E}. \quad (21.45)$$

Note that an E_x field at the slots corresponds to \mathbf{M}_s density vector, which is tangential to the ground plane. Thus, its image is of the same direction. The equivalent magnetic current densities for the dominant TM_{001}^x mode are sketched below.



At slots #1 and #2, the equivalent \mathbf{M}_s currents are co-directed and with equal amplitudes. They are constant along x and y .

Radiation from a slot with constant current density



The radiation from an $(x-y)$ slot of constant \mathbf{M}_s currents is found using the electric vector potential \mathbf{F} . Since \mathbf{M}_s has only a y component, so does \mathbf{F} : $\mathbf{F} = F_y \hat{\mathbf{y}}$.

$$F_y(r, \theta, \phi) = \frac{\epsilon}{4\pi} \int_{-h/2}^{h/2} \int_{-W/2}^{W/2} \frac{M_y}{r_{PQ}} e^{-jk_0 r_{PQ}} dx' dy'. \quad (21.46)$$

Here, $M_y = -2E_0$, E_0 being the phasor of the \mathbf{E} -field at the radiating slot, and $r_{PQ} = r - \mathbf{r}' \cdot \hat{\mathbf{r}} = r - x' \sin \theta \cos \phi - y' \sin \theta \sin \phi$.

$$F_y = -2\varepsilon E_0 \frac{e^{-jk_0 r}}{4\pi r} \int_{-h/2}^{h/2} \exp(jk_0 x' \sin \theta \cos \phi) dx' \cdot \int_{-W/2}^{W/2} \exp(jk_0 y' \sin \theta \sin \phi) dy' \quad (21.47)$$

$$\Rightarrow F_y = -\frac{\varepsilon E_0 W h}{2\pi r} \cdot e^{-jk_0 r} \cdot \frac{\sin X}{X} \cdot \frac{\sin Y}{Y} \quad (21.48)$$

where

$$X = \frac{k_0 h}{2} \sin \theta \cos \phi,$$

$$Y = \frac{k_0 W}{2} \sin \theta \sin \phi.$$

According to the relation between the far-zone \mathbf{E} -field and the vector potential,

$$E_r \approx 0, E_\phi = j\omega\eta F_\theta, E_\theta = -j\omega\eta F_\phi, \quad (21.49)$$

where $\eta = \sqrt{\mu_0 / \varepsilon_0}$, $F_\theta = F_y \cos \theta \sin \phi$, and $F_\phi = F_y \cos \phi$.

$$\Rightarrow E_\phi = j\omega\eta\varepsilon_0 \frac{WhE_0}{2\pi r} e^{-jk_0 r} \cos \theta \sin \phi \frac{\sin X}{X} \frac{\sin Y}{Y}, \quad (21.50)$$

$$\Rightarrow E_\theta = -j\omega\eta\varepsilon_0 \frac{WhE_0}{2\pi r} e^{-jk_0 r} \cos \phi \frac{\sin X}{X} \frac{\sin Y}{Y}. \quad (21.51)$$

Since $\omega\eta\varepsilon_0 = k_0$,

$$E_\phi = jk_0 W \frac{V_0}{2\pi r} e^{-jk_0 r} \left(\cos \theta \sin \phi \frac{\sin X}{X} \frac{\sin Y}{Y} \right), \quad (21.52)$$

$$E_\theta = -jk_0 W \frac{V_0}{2\pi r} e^{-jk_0 r} \left(\cos \phi \frac{\sin X}{X} \frac{\sin Y}{Y} \right). \quad (21.53)$$

Here $V_0 = hE_0$ is the voltage between the patch edge and the ground plane.

Slots #1 and #2 form an array of two elements with excitation of equal magnitude and phase, separated by the physical distance L . Their AF is

$$AF_{12} = 2 \cos \left(\frac{k_0 L_{eff}}{2} \cos \theta \right). \quad (21.54)$$

Here $L_{eff} = L + 2\Delta L$ is the effective patch length. Thus, the total radiation field is

$$E_{\phi}^t = j \frac{k_0 W V_0}{\pi r} e^{-jk_0 r} \left(\cos \theta \sin \phi \frac{\sin X}{X} \frac{\sin Y}{Y} \right) \times \left[\cos \left(\frac{k_0 L_{eff}}{2} \cos \theta \right) \right], \quad (21.55)$$

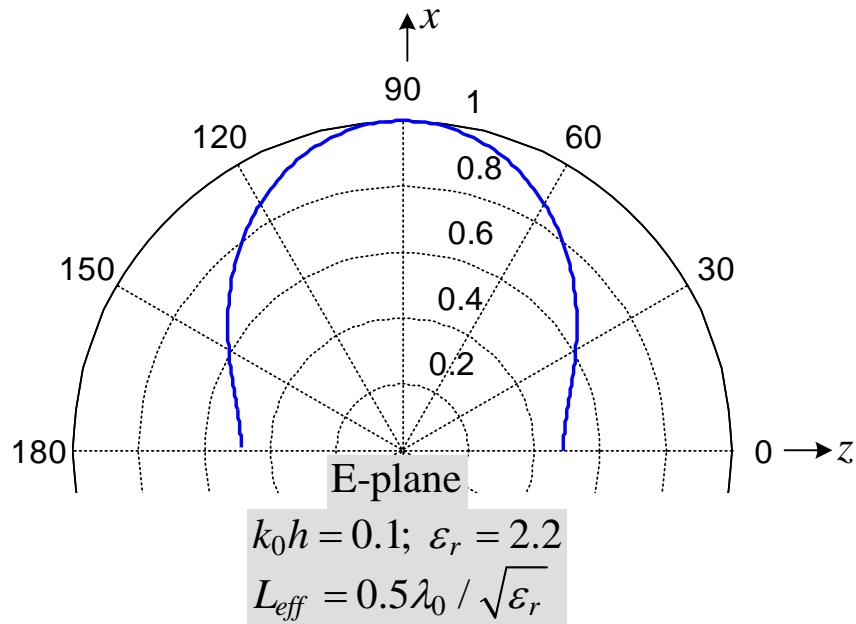
$$E_{\theta}^t = -j \frac{k_0 W V_0}{\pi r} e^{-jk_0 r} \left(\cos \phi \frac{\sin X}{X} \frac{\sin Y}{Y} \right) \times \left[\cos \left(\frac{k_0 L_{eff}}{2} \cos \theta \right) \right]. \quad (21.56)$$

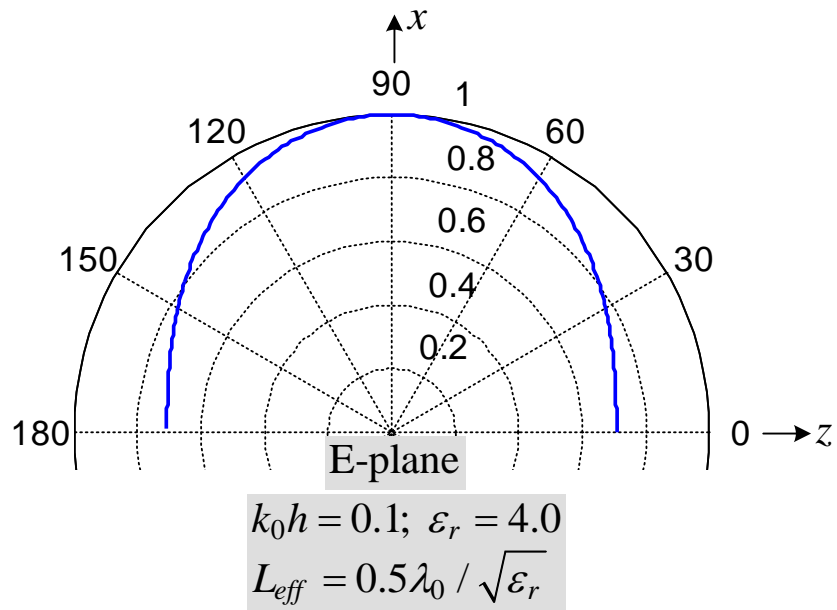
Introducing $Z = (k_0 L_{eff} / 2) \cos \theta$, the pattern of the patch is obtained as

$$f(\theta, \phi) = \sqrt{\bar{E}_{\phi}^2 + \bar{E}_{\theta}^2} = \sqrt{1 - \sin^2 \phi \cdot \sin^2 \theta} \cdot \frac{\sin X}{X} \frac{\sin Y}{Y} \cos Z. \quad (21.57)$$

E-plane pattern (xz plane, $\phi = 0^\circ$, $0^\circ \leq \theta \leq 180^\circ$)

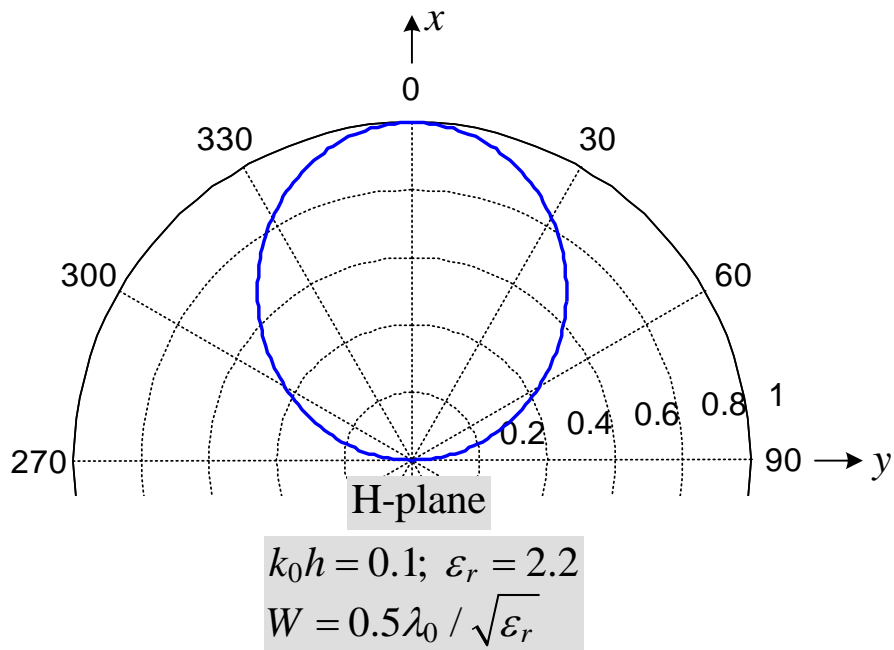
$$f_E(\theta) = \frac{\sin \left(\frac{k_0 h}{2} \sin \theta \right)}{\frac{k_0 h}{2} \sin \theta} \cdot \cos \left(\frac{k_0 L_{eff}}{2} \cos \theta \right). \quad (21.58)$$

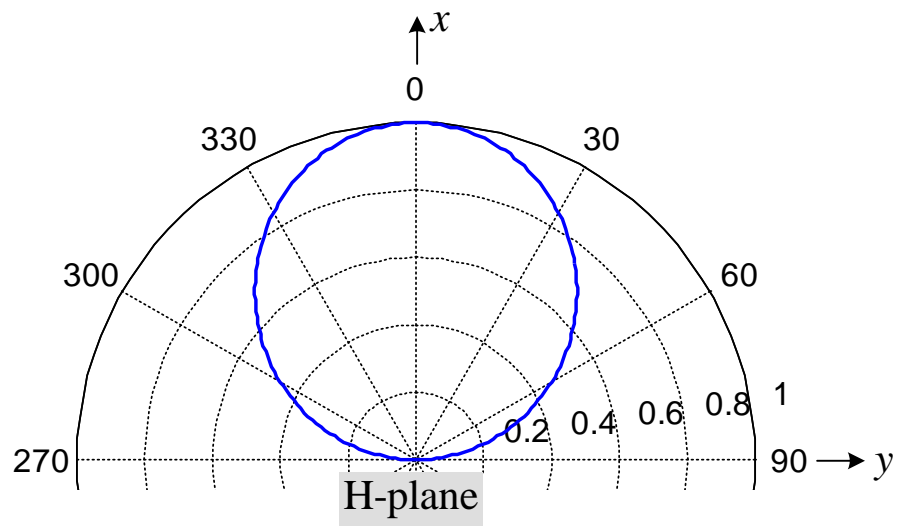




H-plane pattern (xy plane, $\theta = 90^\circ$, $0^\circ \leq \phi \leq 90^\circ$ and $270^\circ \leq \phi \leq 360^\circ$)

$$f_H(\theta) = \cos \phi \cdot \frac{\sin\left(\frac{k_0 h}{2} \cos \phi\right)}{\frac{k_0 h}{2} \cos \phi} \cdot \frac{\sin\left(\frac{k_0 W}{2} \sin \phi\right)}{\frac{k_0 W}{2} \sin \phi} \quad (21.59)$$

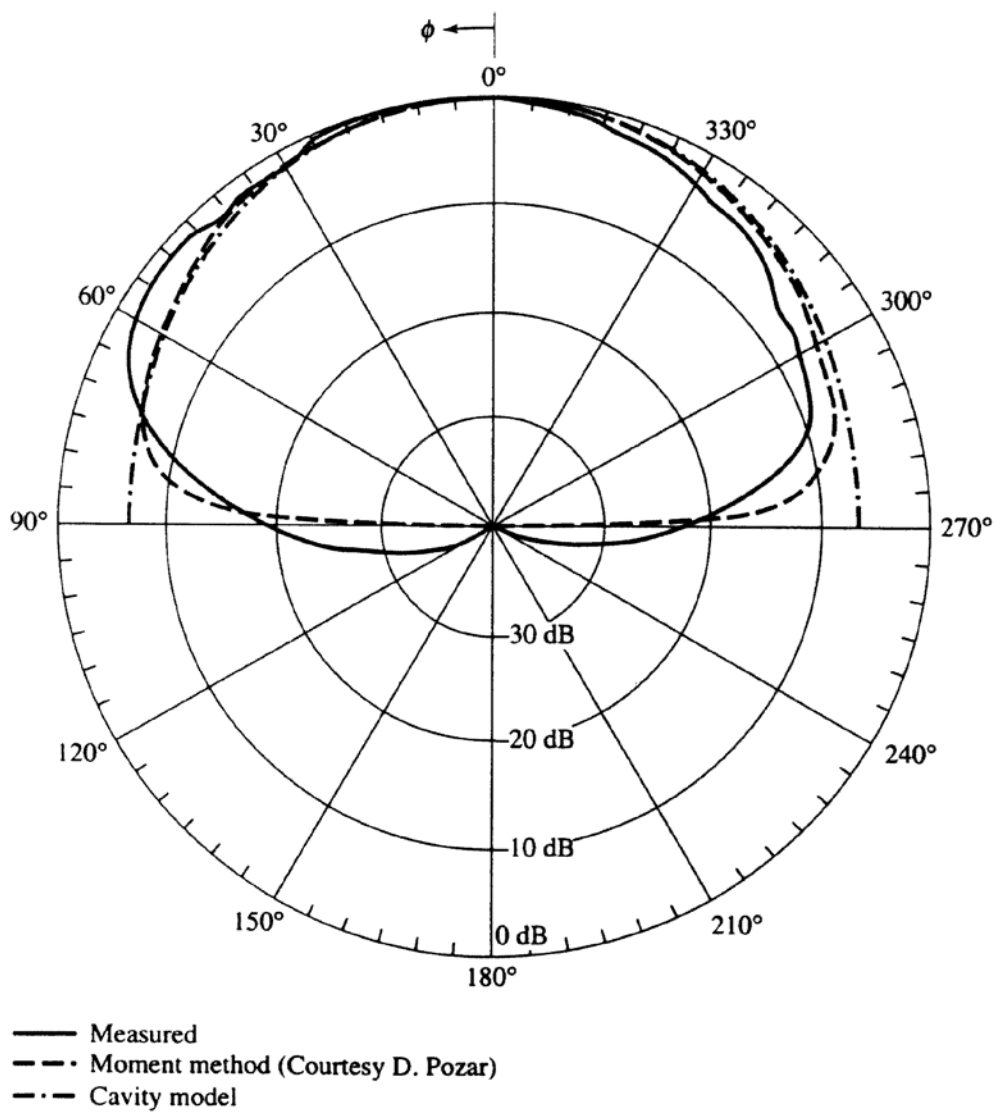




H-plane

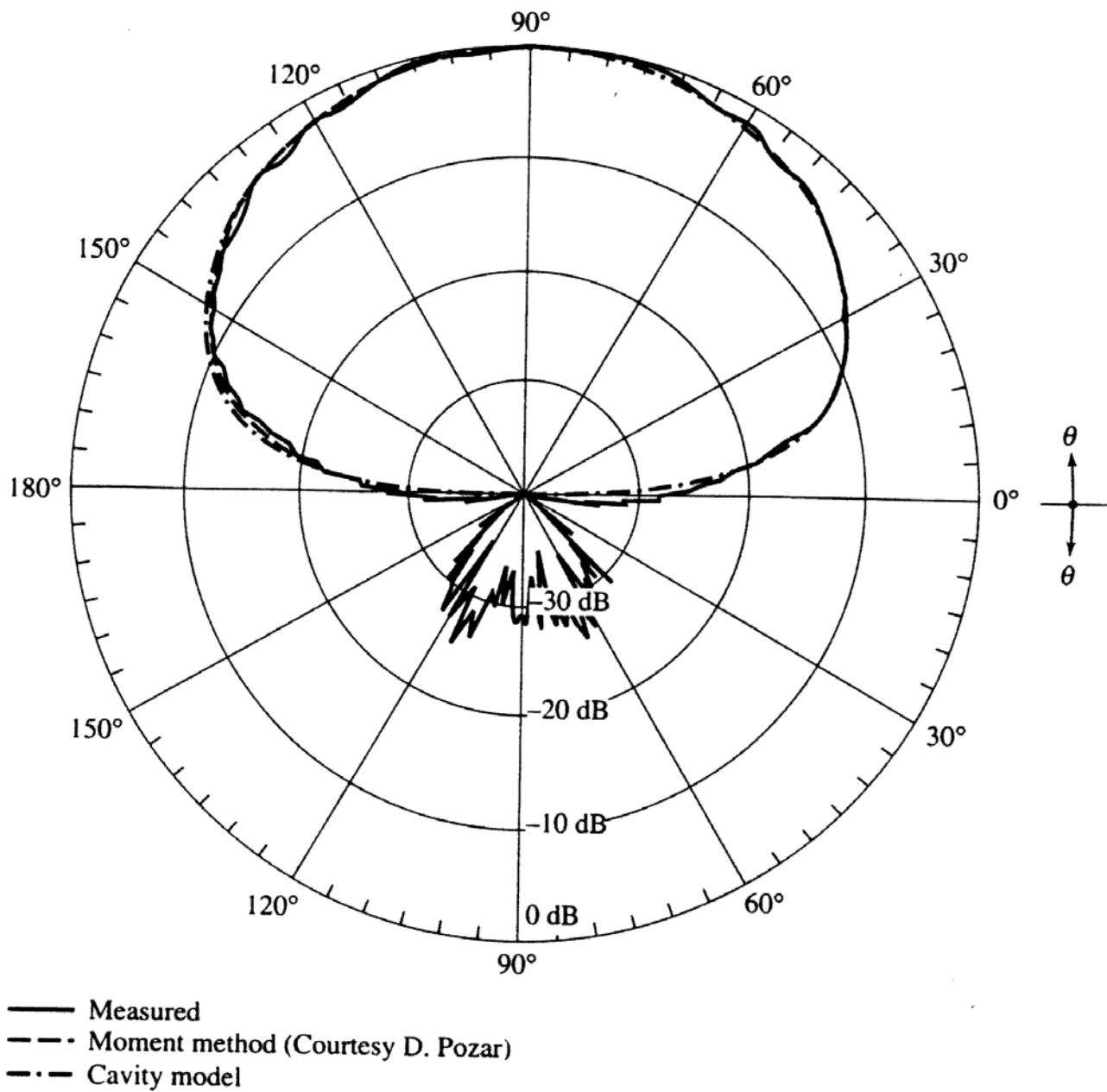
$$k_0 h = 0.1; \epsilon_r = 4.0$$

$$W = 0.5 \lambda_0 / \sqrt{\epsilon_r}$$



(a) *E*-plane ($\theta = 90^\circ$)

Fig. 14.17, p. 746, Balanis



(b) H -plane ($\phi = 0^\circ$)

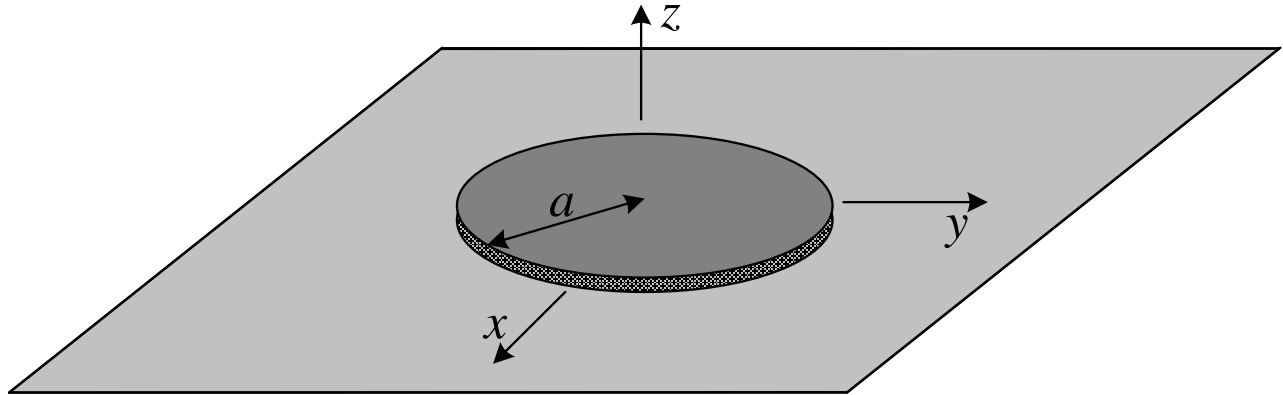
Fig. 14.18, p. 747, Balanis

Non-radiating slots: It can be shown that the slots at $y = -W/2$ and $y = W/2$ do not radiate in the principle \mathbf{E} - and \mathbf{H} -planes. In general, these two slots do radiate away from the principle planes, but their field intensity is everywhere small compared to that radiated by slots #1 and #2.

LECTURE 22: MICROSTRIP ANTENNAS – PART III

(Circular patch antennas: the cavity model. Radiation field of the circular patch. Circularly polarized radiation from patches. Arrays and feed networks.)

1. Circular patch: the cavity model



The circular patch cannot be analyzed using the TL method, but can be accurately described by the cavity method. It is again assumed that only TM_z modes are supported in the cavity. They are fully described by the VP $\mathbf{A} = A_z \hat{\mathbf{z}}$. The A_z VP function satisfies the Helmholtz equation,

$$\nabla^2 A_z + k^2 A_z = 0 \quad (22.1)$$

which now is solved in cylindrical coordinates:

$$\frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial A_z}{\partial \rho} \right) + \frac{1}{\rho^2} \frac{\partial^2 A_z}{\partial \phi^2} + \frac{\partial^2 A_z}{\partial z^2} + k^2 A_z = 0, \quad (22.2)$$

$$\Rightarrow \frac{1}{\rho} \frac{\partial A_z}{\partial \rho} + \frac{\partial^2 A_z}{\partial \rho^2} + \frac{1}{\rho^2} \frac{\partial^2 A_z}{\partial \phi^2} + \frac{\partial^2 A_z}{\partial z^2} + k^2 A_z = 0. \quad (22.3)$$

Using the method of separation of variables,

$$A_z = R(\rho)F(\phi)Z(z), \quad (22.4)$$

$$\Rightarrow \frac{1}{\rho} FZ \frac{\partial R}{\partial \rho} + FZ \frac{\partial^2 R}{\partial \rho^2} + \frac{RZ}{\rho^2} \frac{\partial^2 F}{\partial \phi^2} + RF \frac{\partial^2 Z}{\partial z^2} + k^2 RFZ = 0, \quad (22.5)$$

$$\Rightarrow \frac{1}{\rho R} \frac{\partial R}{\partial \rho} + \frac{1}{R} \frac{\partial^2 R}{\partial \rho^2} + \frac{1}{\rho^2 F} \frac{\partial^2 F}{\partial \phi^2} + \frac{1}{Z} \frac{\partial^2 Z}{\partial z^2} = -k^2. \quad (22.6)$$

The 4th term is independent of ρ and ϕ , and is being separated:

$$\frac{1}{Z} \frac{\partial^2 Z}{\partial z^2} = -k_z^2. \quad (22.7)$$

Then,

$$\frac{1}{R\rho} \frac{\partial R}{\partial \rho} + \frac{1}{R} \frac{\partial^2 R}{\partial \rho^2} + \frac{1}{\rho^2 F} \frac{\partial^2 F}{\partial \phi^2} = -(k^2 - k_z^2) = \text{const.} \quad (22.8)$$

$$\Rightarrow \frac{\rho}{R} \frac{\partial R}{\partial \rho} + \frac{\rho^2}{R} \frac{\partial^2 R}{\partial \rho^2} + \frac{1}{F} \frac{\partial^2 F}{\partial \phi^2} + (k^2 - k_z^2)\rho^2 = 0. \quad (22.9)$$

Now, the 3rd term is independent of ρ , and the other terms are independent of ϕ . Thus, (22.9) is separated into two equations:

$$\frac{1}{F} \frac{\partial^2 F}{\partial \phi^2} = k_\phi^2 \quad (22.10)$$

and

$$\frac{\rho}{R} \frac{\partial R}{\partial \rho} + \frac{\rho^2}{R} \frac{\partial^2 R}{\partial \rho^2} + (k^2 - k_z^2)\rho^2 - k_\phi^2 = 0. \quad (22.11)$$

We define

$$k_\rho^2 = k^2 - k_z^2. \quad (22.12)$$

Then (22.11) can be written as [note that (22.11) depends only on ρ]:

$$\rho \frac{\partial}{\partial \rho} \left(\rho \frac{\partial R}{\partial \rho} \right) + [(k_\rho \rho)^2 - k_\phi^2] \cdot R = 0. \quad (22.13)$$

Thus, equation (22.1) has been separated into three ordinary differential equations — (22.7), (22.10) and (22.13).

A. The Z-equation

Equation (22.7) is complemented by the Neumann BC at the top patch and the grounded plane (electric walls):

$$\frac{\partial A_z}{\partial z} = 0 \Rightarrow \frac{\partial Z}{\partial z} = 0. \quad (22.14)$$

Its solution, therefore, is in the form

$$\boxed{Z(z) = \sum_p c_p \cos\left(p \frac{\pi}{h} z\right)} \quad (22.15)$$

with the eigenvalues are $k_z = p\pi / h$. Here, p is an integer.

B. The F -equation

The solution of (22.10) is also a harmonic function. We are interested in real-valued harmonic functions, i.e.,

$$F(\phi) = \sum_n b_n^c \cos(k_\phi^n \phi) + b_n^s \sin(k_\phi^n \phi). \quad (22.16)$$

Since there are no specific BC's to be imposed at certain angular positions, the only requirement for the eigenvalues k_ϕ^n comes from the condition that the $F(\phi)$ must be periodic in ϕ ,

$$F(\phi) = F(\phi + 2\pi). \quad (22.17)$$

Equation (22.17) is true only if k_ϕ^n are integers. That is why the usual construction of a general solution for $F(\phi)$ for a complete cylindrical region ($\phi = 0$ to 2π) is in the form

$$F(\phi) = \sum_n b_n^c \cos(n\phi) + b_n^s \sin(n\phi), \quad (22.18)$$

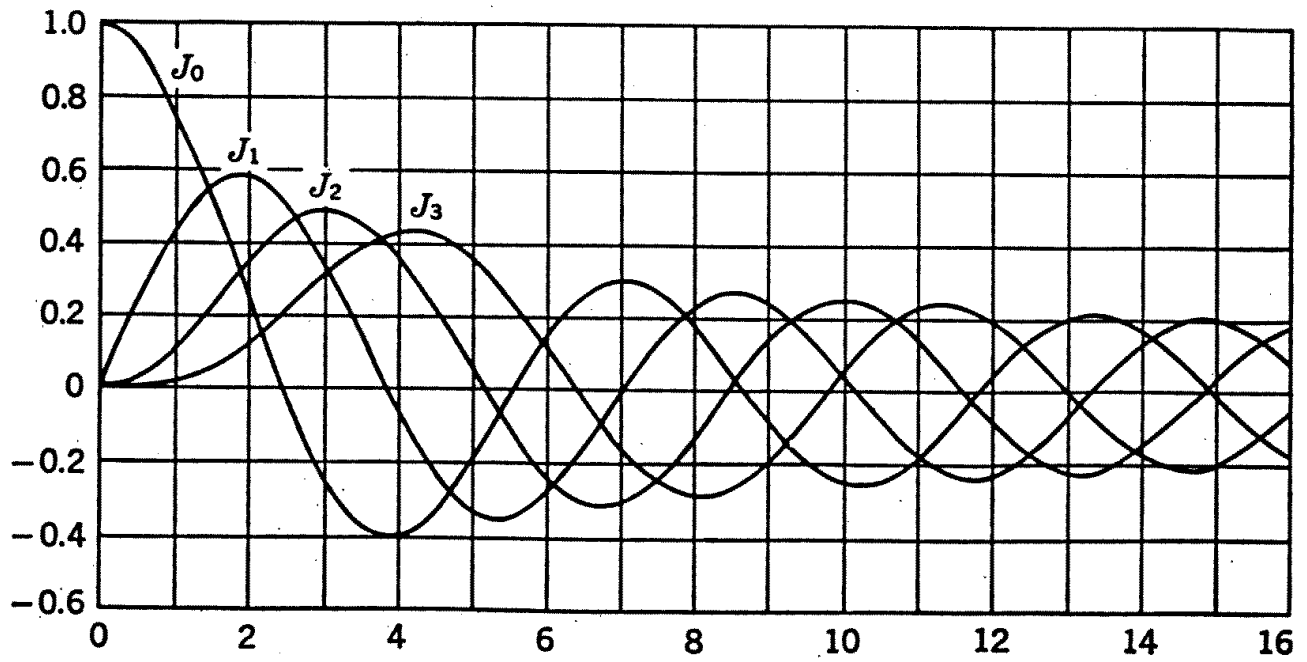
where n is an integer. This is the well-known Fourier-series expansion.

C. The R -equation

Equation (22.13) is a Bessel equation in which k_ϕ is an integer ($k_\phi = n$). Solutions are of the form of the following special functions:

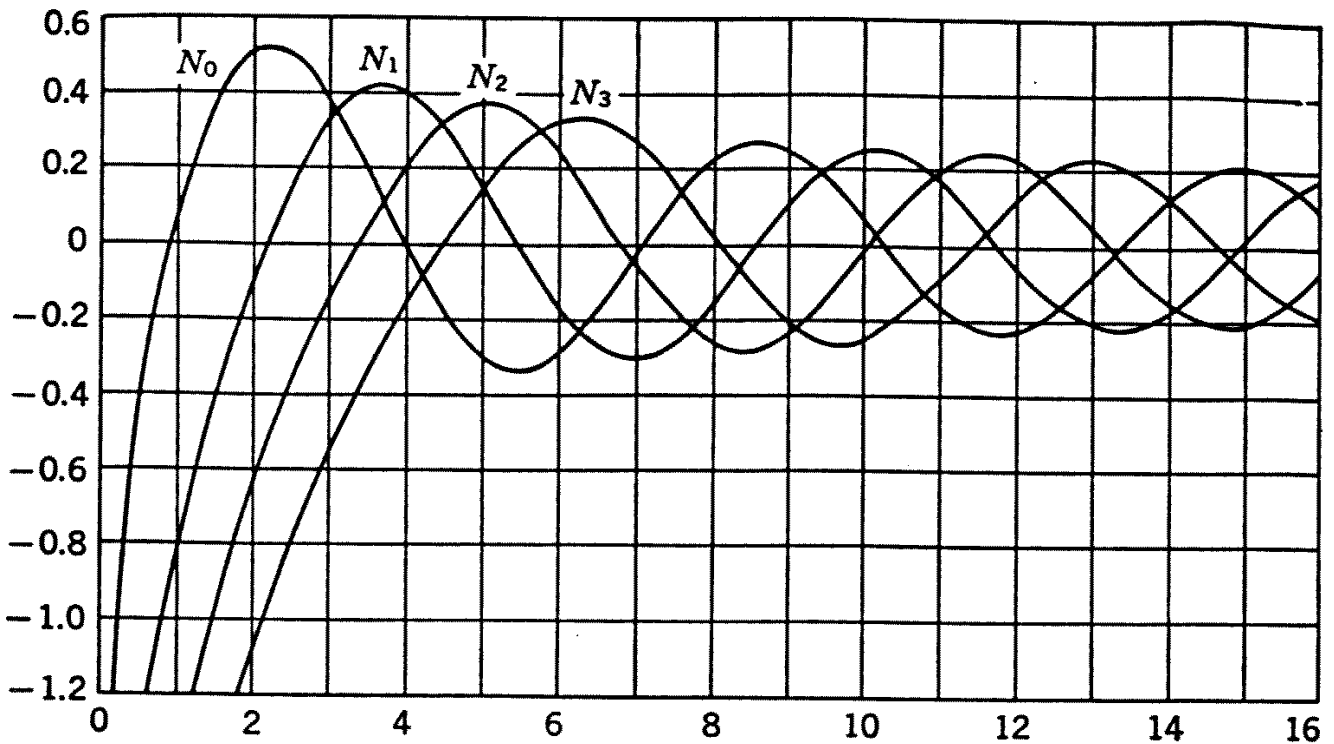
- $J_n(k_\rho \rho)$ — Bessel function of the first kind,
- $N_n(k_\rho \rho)$ — Bessel function of the second kind (Neumann function),
- $H_n^{(1)}(k_\rho \rho)$ — Hankel function of the first kind,
- $H_n^{(2)}(k_\rho \rho)$ — Hankel function of the second kind.

Note: $H_n^{(1)} = J_n + jN_n$; $H_n^{(2)} = J_n - jN_n$.



Bessel functions of the first kind.

Fig. D-1, Harrington, p. 461



Bessel functions of the second kind.

Fig. D-2, Harrington, p. 462

The eigenvalues are determined according to the boundary conditions. In the cavity model, it is required that (magnetic wall)

$$\frac{\partial A_z}{\partial \rho} = 0 \Rightarrow \frac{\partial R}{\partial \rho} = 0 \Big|_{\rho=a}, \quad (22.19)$$

and that the field is finite for $\rho = a$. The Bessel functions of the first kind $J_n(k_\rho \rho)$ are the suitable choice. The eigenvalues k_ρ are determined from (22.19):

$$\frac{\partial J_n(k_\rho \rho)}{\partial \rho} = 0 \Big|_{\rho=a}, \Rightarrow k_\rho^{nm} = \frac{\chi'_{nm}}{a}, \quad (22.20)$$

where χ'_{nm} is the m^{th} null of the derivative of the Bessel function of the n^{th} order J'_n . Thus, the solution of the Helmholtz equation for A_z can be given in a modal form as, see (22.4),

$$A_z^{(mnp)} = M_{mnp} J_m \left(\chi'_{nm} \frac{\rho}{a} \right) \cdot [b_n^c \cos(n\phi) + b_n^s \sin(n\phi)] \cdot \cos \left(p \frac{\pi}{h} z \right). \quad (22.21)$$

The characteristic equation (22.12) is finally obtained as

$$k^2 = \omega^2 \mu \varepsilon = k_\rho^2 + k_z^2. \quad (22.22)$$

From (22.22), the resonant frequencies of the patch can be obtained:

$$\mu \varepsilon \omega_{mnp}^2 = \left(\frac{\chi'_{nm}}{a} \right)^2 + \left(p \frac{\pi}{h} \right)^2, \quad (22.23)$$

$$f_{r(mnp)} = \frac{1}{2\pi \sqrt{\mu \varepsilon}} \sqrt{\left(\frac{\chi'_{nm}}{a} \right)^2 + \left(p \frac{\pi}{h} \right)^2}. \quad (22.24)$$

Equation (22.24) does not take into account the fringing effect of the circular patch. To account for the effective increase of the patch size due to fringing, the actual radius a is replaced by an effective one,

$$a_e = a \left\{ 1 + \frac{2h}{\pi a \varepsilon_r} \left[\ln \left(\frac{\pi a}{2h} \right) + 1.7726 \right] \right\}^{1/2}. \quad (22.25)$$

The first four modes in ascending order are TM_{z110} , TM_{z210} , TM_{z010} ,

TM_{z310} where the respective nulls χ'_{nm} are

$$\begin{aligned}\chi'_{11} &= 1.8412 & \chi'_{01} &= 3.8318 \\ \chi'_{21} &= 3.0542 & \chi'_{31} &= 4.2012\end{aligned}$$

The resonant frequency of the dominant TM_{z110} mode can be determined from (22.25) as

$$f_{r(110)} = \frac{1.8412 \cdot c}{2\pi a_e \sqrt{\epsilon_r}} \quad (22.26)$$

where c is the speed of light in vacuum.

The VP of the dominant TM_{z110} mode is

$$A_z^{(110)} = M_{110} J_1 \left(\chi'_{11} \frac{\rho}{a} \right) \cdot (b_n^c \cos \phi + b_n^s \sin \phi). \quad (22.27)$$

Assuming excitation at $\phi = 0$ (A_z has vanishing angular first derivative), we set $b_n^s = 0$. The field components are computed from A_z according to the field-potential relations

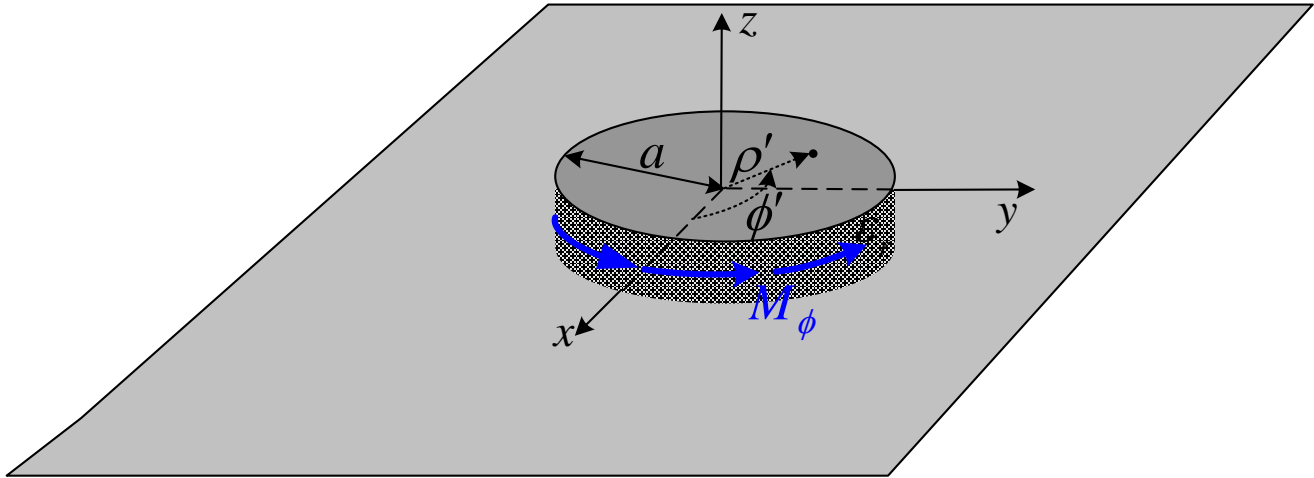
$$\begin{aligned}E_\rho &= -\frac{j}{\omega\mu\epsilon} \frac{\partial^2 A_z}{\partial \rho \partial z} & H_\rho &= \frac{1}{\mu} \frac{1}{\rho} \frac{\partial A_z}{\partial \phi} \\ E_\phi &= -\frac{j}{\omega\mu\epsilon} \frac{1}{\rho} \frac{\partial^2 A_z}{\partial \phi \partial z} & H_\phi &= -\frac{1}{\mu} \frac{\partial A_z}{\partial \rho} \\ E_z &= -\frac{j}{\omega\mu\epsilon} \left(\frac{\partial^2 A_z}{\partial z^2} + k^2 A_z \right) & H_z &= 0\end{aligned} \quad (22.28)$$

For the dominant TM_{z110} mode,

$$\begin{aligned}E_\rho = E_\phi &= 0 & H_\rho &= j \frac{E_0}{\omega\mu_0} \frac{1}{\rho} J_1(\chi'_{11}\rho/a) \sin \phi \\ E_z &= E_0 J_1(\chi'_{11}\rho/a) \cos \phi & H_\phi &= j \frac{E_0}{\omega\mu_0} J_1'(\chi'_{11}\rho/a) \cos \phi\end{aligned} \quad (22.29)$$

From the field components, we can compute the cavity modal impedance for any feed point specified by ρ and ϕ . In view of the closed-wall nature of the BCs, the impedance will be reactive. To obtain the real part of the antenna impedance, the radiated power has to be computed.

2. Radiated fields and equivalent surface currents of the circular patch



As with the rectangular patch, the field radiated by the circular slot is determined using the equivalence principle. The circumferential wall of the cavity is replaced by an equivalent circular sheet of magnetic current density

$$M_{s\phi} = 2E_z|_{\rho=a}, \text{ V/m}, \quad (22.30)$$

radiating in free space. The factor of 2 accounts for the ground plane. Since the height of the slot h is very small and the slot field is independent of z , we can substitute the surface magnetic current density over the slot with a filamentary magnetic current $I_m = M_{s\phi}h$:

$$I_m = \underbrace{2hE_0J_1(\chi'_{11})}_{2V_0} \cos \phi, \text{ V}. \quad (22.31)$$

Here, $V_0 = hE_0J_1(\chi'_{11})$ is the voltage between ground and the top plate of the patch at the feed ($\phi = 0$).

Using the theory for the radiation field of a circular slot, the following expressions are obtained for the far field of the circular patch:

$$E_r = 0, E_\theta = -C(r) \cdot \cos \phi \cdot J'_{02}, E_\phi = C(r) \cdot \cos \theta \sin \phi \cdot J_{02}, \quad (22.32)$$

where

$$C(r) = j \frac{k_0 a_e V_0 e^{-jk_0 r}}{2r},$$

$$J_{02} = J_0(k_0 a_e \sin \theta) + J_2(k_0 a_e \sin \theta),$$

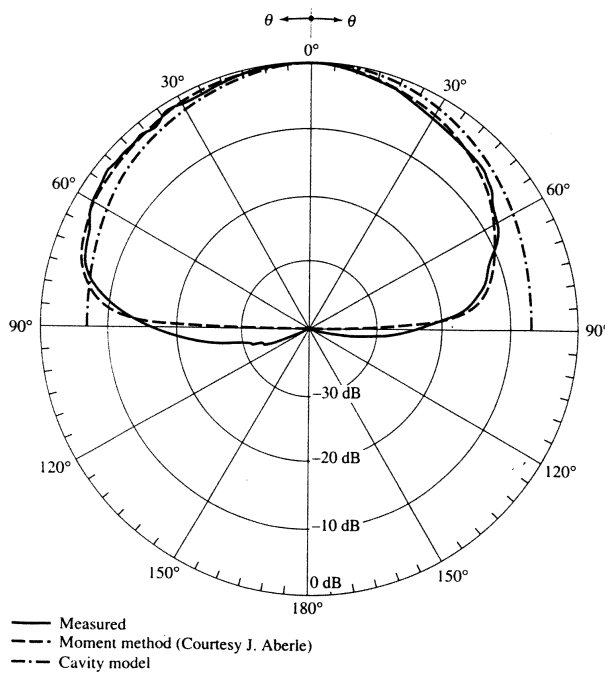
$$J'_{02} = J_0(k_0 a_e \sin \theta) - J_2(k_0 a_e \sin \theta).$$

E-plane amplitude pattern:

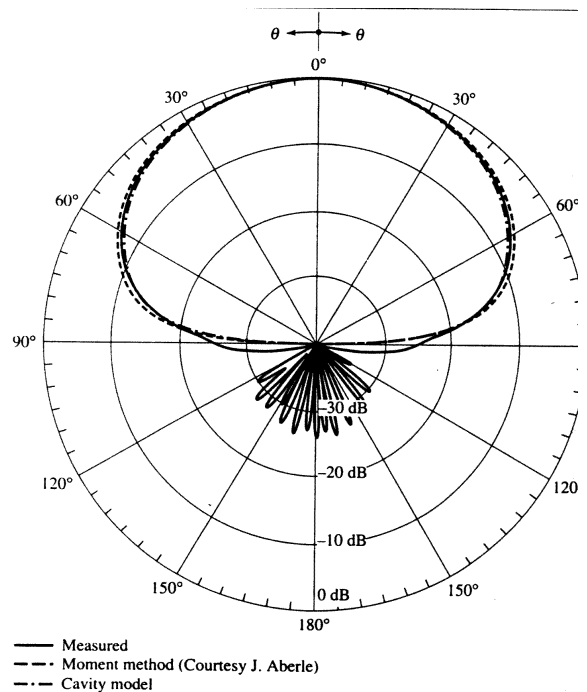
$$\bar{E}_\theta(0^\circ \leq \theta \leq 90^\circ, \varphi = 0^\circ, 180^\circ) = J'_{02}, \quad \bar{E}_\phi = 0$$

H-plane amplitude pattern:

$$\bar{E}_\phi(0^\circ \leq \theta \leq 90^\circ, \varphi = 90^\circ, 270^\circ) = \cos \theta \cdot J_{02}, \quad \bar{E}_\theta = 0$$



(a) E-plane ($\phi = 0^\circ, 180^\circ$)



(a) H-plane ($\phi = 90^\circ, 270^\circ$)

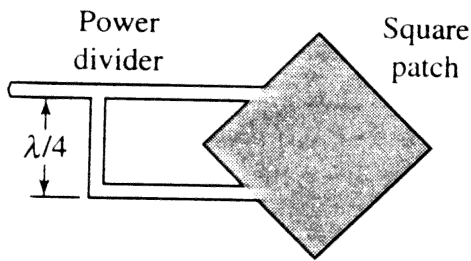
Measured and computed (based on moment method and cavity models)
 E- and H-plane patterns of circular microstrip patch antenna ($a = 0.525$ cm, $a_e = 0.598$ cm,
 $\gamma_f = 0.1$ cm, $\epsilon_r = 2.2$, $h = 0.1588$ cm, $f_0 = 10$ GHz, $\lambda_0 = 3$ cm).

Fig. 14.23, p. 758, Balanis

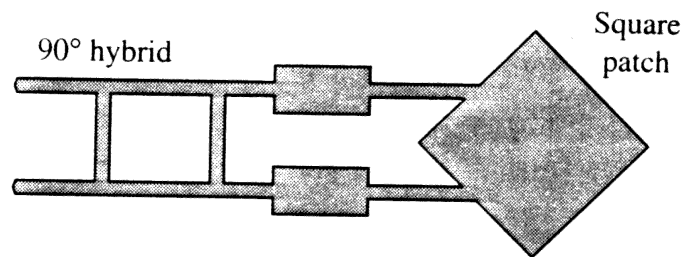
3. Circular polarization with patch antennas

Circular polarization can be obtained if two orthogonal modes are excited with a 90° time-phase difference between them. This can be accomplished by adjusting the physical dimensions of the patch and using either one or two feed points.

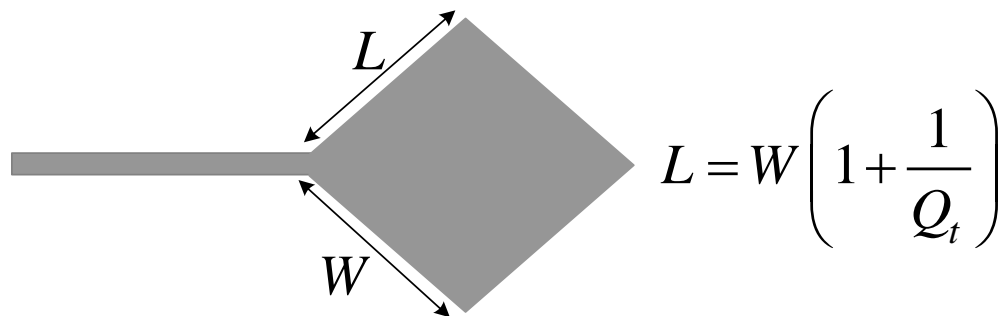
A. Square patch with circularly polarized field



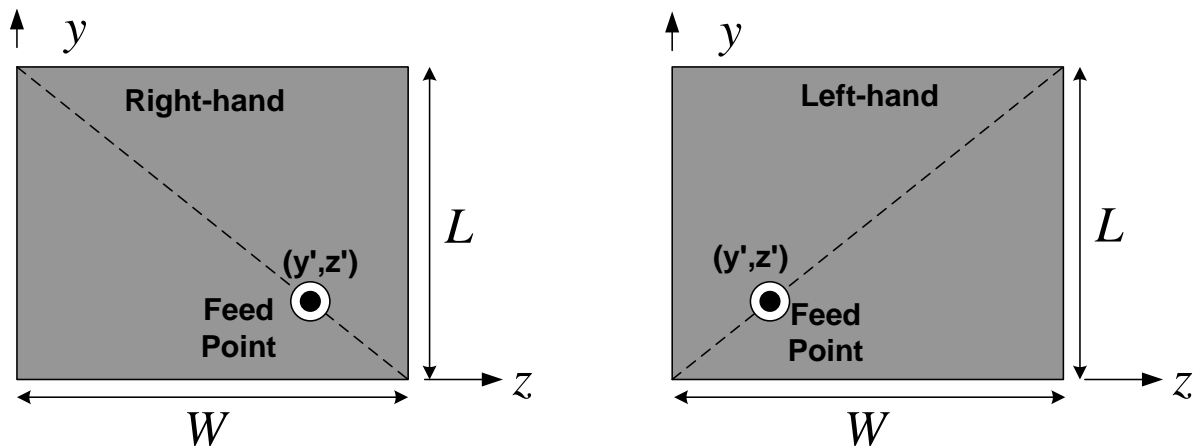
(a) Square patch driven at adjacent sides through a power divider



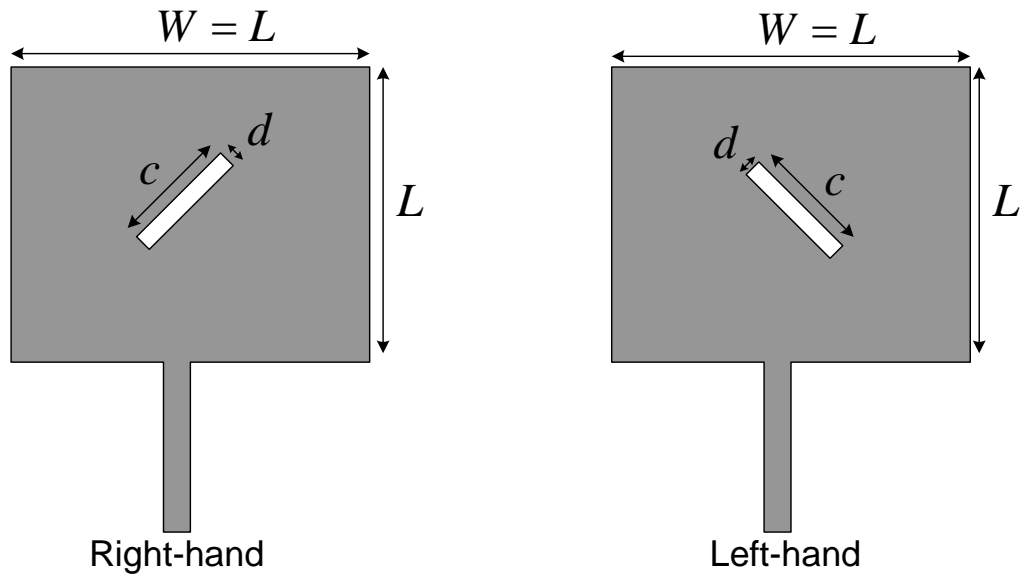
(b) Square patch driven at adjacent sides through a 90° hybrid



(c) Nearly square patch with microstrip-line feed for CP accounting for losses; $Q_t = 1 / \tan \delta_{eff}$

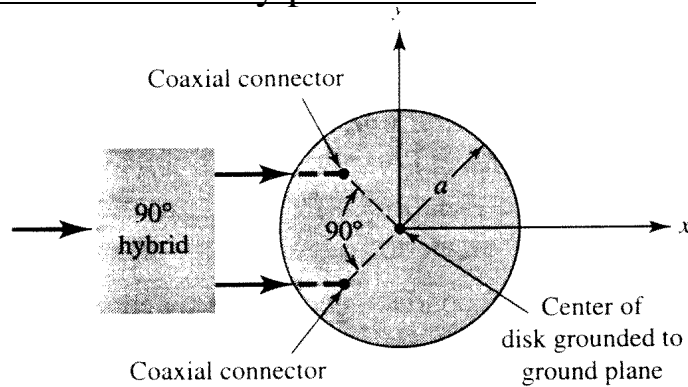


(d) Coax-feeds for CP

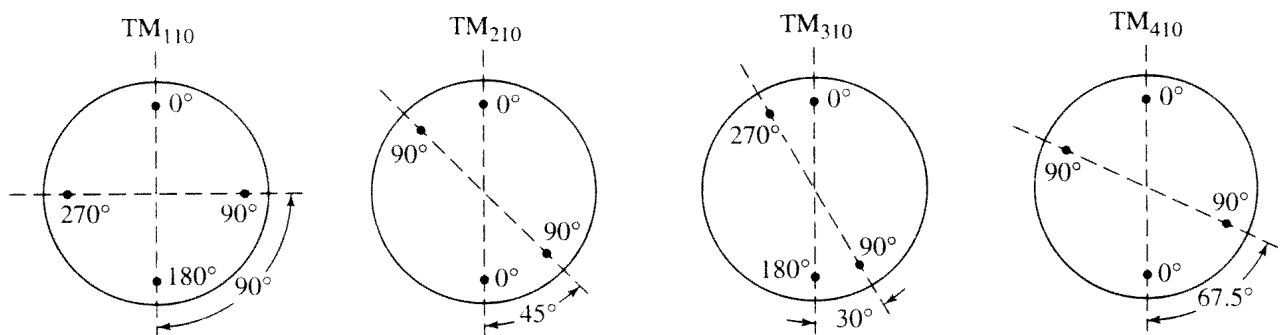


(e) CP for square patches with thin slots: $c = L / 2.72 = W / 2.72$, $d = c / 10$

B. Circular patch with circularly polarized field



(c) Circular patch fed with coax



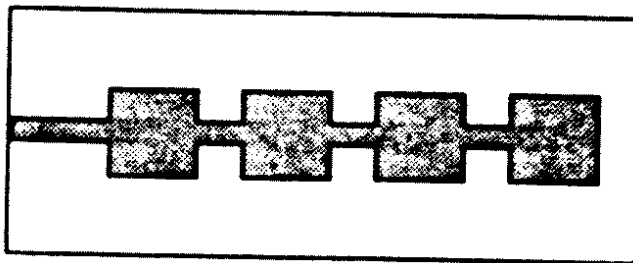
(d) Circular patch feed arrangements for TM_{110}^z and higher order modes

(Source: J. Huang, "Circularly Polarized Conical Patterns from Circular Microstrip Antennas," IEEE Trans. Antennas Propagat., Vol. AP-32, No. 9, Sept. 1984. © 1984 IEEE.)

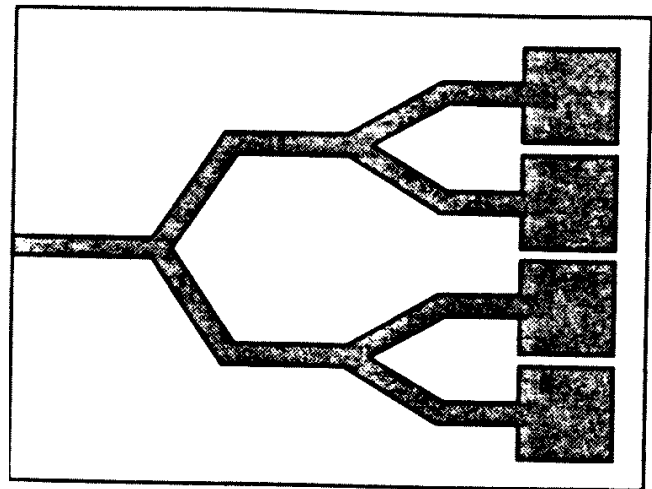
FEED-PROBE ANGULAR SPACING OF DIFFERENT MODES FOR CIRCULAR POLARIZATION

	TM_{110}	TM_{210}	TM_{310}	TM_{410}	TM_{510}	TM_{610}
α	90°	45° or 135°	30° or 90°	22.5° or 67.5°	$18^\circ, 54^\circ$ or 90°	$15^\circ, 45^\circ$ or 75°

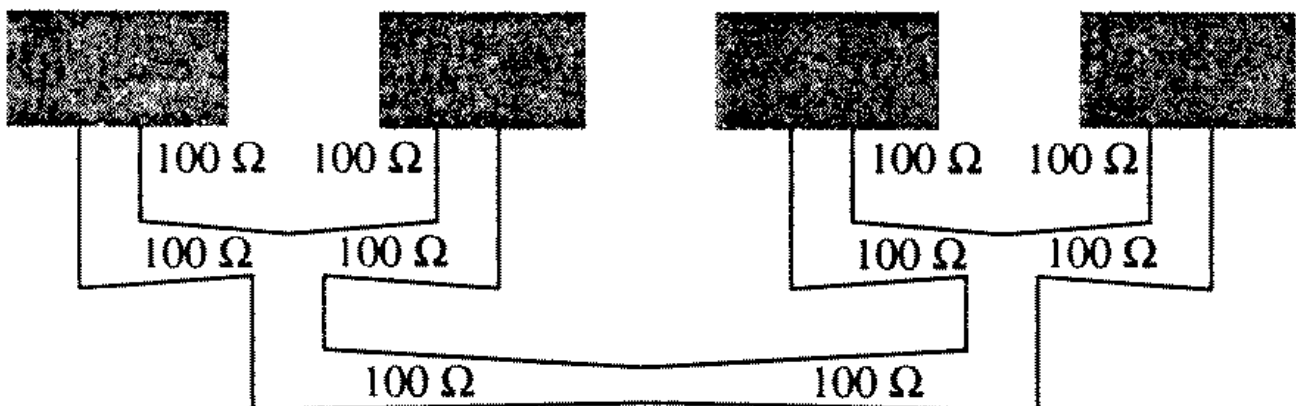
4. Array and feed networks



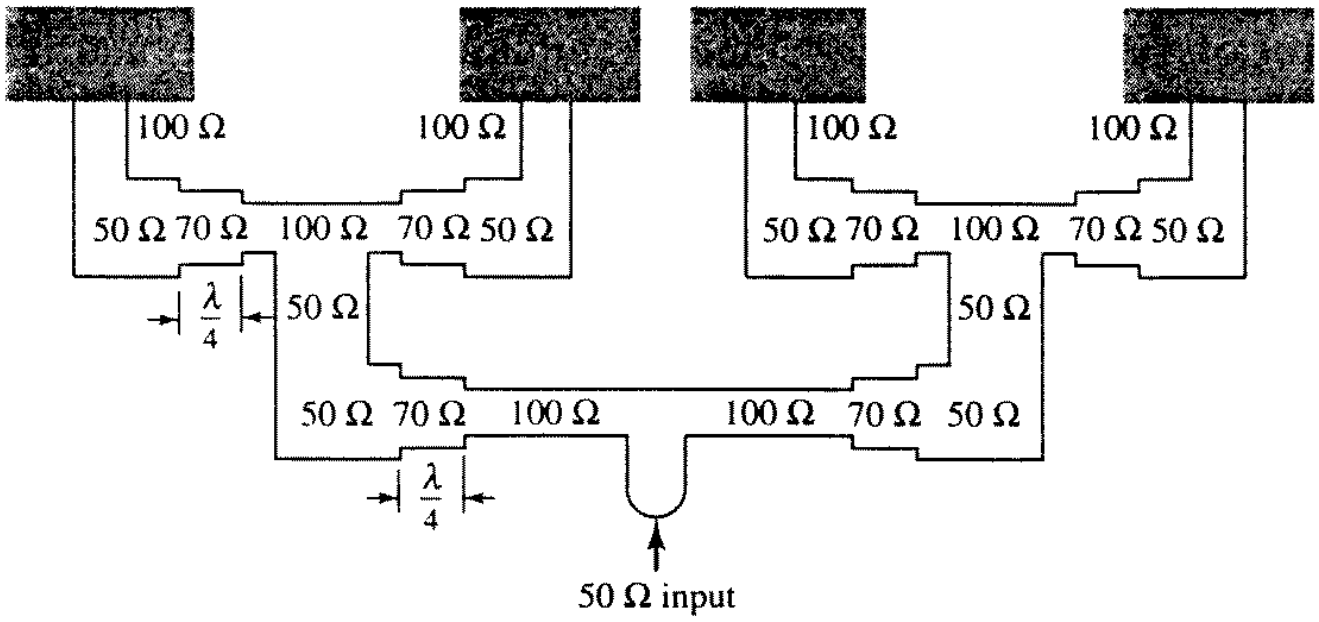
(a) Series feed



(b) Corporate feed



(a) Tapered lines



(b) $\lambda/4$ transformers