SOME ANTENNA AND PROPAGATION FUNDAMENTALS

BY

CHRIS ANGOVE

March 2011

Contents

FIGURES

TABLES

Table 1-1 Abbreviations and Acronyms

Accronym	Meaning
2G	2nd. generation mobile communications
3G	3rd. generation mobile communications
4G	4th. generation mobile communications
ADSL	asymmetric digital subscriber line
AM	amplitude modulation
COFDM	coded orthogonal frequency division multiplex
DAB	digital audio broadcast
dB	decibel
dBi	decibel relative to an isotropic radiator
dBW	decibel relative to one watt
DOCSIS	Data Over Cable Service Interface Specification
E	electric field
EIRP	equivalent isotropic radiated power
ERP	equivalent radiated power
FCC	Federal Communications Commission
FSPL	free space path loss
GHz	gigahertz
H	magnetic field
HF	high frequency
ISM	industrial, scientific and medical
K-band	frequency band designation covering 18 GHz to 27 GHz
kHz	kilohertz
LF	low frequency (30 kHz to 300 kHz)
LOS	line of sight
LW	long wave
m	metre
MF	medium frequency
MHz	megahertz
OFCOM	Office for Communications
O-H	oxygen to hydrogen molecular bond
O-O	oxygen to oxygen molecular bond
PFD	power flux density
PFD	power flux density
PSTN	public switched telephone network
RF	radio frequency
RMS	root mean square
SFH	super high frequency (3 GHz to 30 GHz)
SFN	single frequency network
SI	System International
TEM	transverse electric and magnetic
TV	television
UHF	ultra high frequency (300 MHz to 3 GHz)
UK	United Kingdom
USA	United States of America
VHF	very high frequency
VLF	very low frequency
VSWR	voltage standing wave ratio
W/m	watts per metre squared

1 INTRODUCTION

When designing a new antenna system, some of the non-electrical requirements of the system should be addressed first such as:

- licensing/regulatory approval
- planning approval
- physical dimensions and mass
- cost.

1.1 Regulatory Approval for the Frequency Band(s) of Operation

With few exceptions, such as the industrial, scientific and medical (ISM) frequency bands, virtually all emissions of non-ionising radiation require the regulatory authorities' approval. Most countries have a government department which deals with the applications for licences to radiate such emissions and sometimes to receive them as well. Typically, the licence conditions would include details of the mandatory operating conditions for the equipment used including the geographical area, frequency band(s), maximum power levels, effective radiated power (ERP) and types of modulation. Other requirements might specify the frequency accuracy and stability of the emissions, maximum levels of spurious emissions, mandatory test measurements and rights of access by the authorities for auditing purposes.

1.2 Hazards from Non-Ionising Radiation

Non ionising radiation is another name for radio frequency (RF) radiation, not to be mistaken for its higher energy and more hazardous counterpart, ionising radiation. Also, in recent years the adjective 'wireless' has been revived from the early days of radio to describe forms of communications which used non ionising radiation. Ionising radiation is associated with much higher frequency waves or streams of particles capable of actually ionising atoms and molecules.

The general consensus is that the greatest hazard affecting the human body from non-ionising radiation results from its heating effect. Serious injuries known as RF burns are possible which can be fatal. This is the same phenomenon that is exploited in a microwave oven to heat food.

Mandatory upper limits for exposure to non-ionising radiation are described by levels of RF power flux density (PFD) that may be experienced by the human body from a radiator, sometimes referred to as *an emitter*. These are often expressed as functions of frequency. PFD is usually specified in the System International (SI) unit of Watts per square metre (W/m^2), in a particular direction.

In the vicinity known as *Fraunhofer region* or, more commonly*, the far field*, the waves originating from an antenna can be considered to be *plane* or transverse electric-magnetic (TEM). TEM waves will be considered again in Section 2.8, but here we must assume that the waves are in *free space*, without any interference caused by reflection, refraction or diffraction. TEM means that the electric, magnetic fields and the direction of propagation are mutually orthogonal (at right angles). Free space is another name for a vacuum and, although we know that the atmosphere is not a vacuum, a clean and homogeneous atmosphere free from pollution is electrically a close approximation to one.

In free space there is a square law relationship between the PFD and the magnitude of the electric field strength $\boldsymbol{E}$ or the magnitude of the magnetic field strength $\boldsymbol{H}$. Notice that $\boldsymbol{E}$ and $\boldsymbol{H}$ are in bold type which means that they are both vector quantities, each has a magnitude and direction. We can express the magnitude of each by inserting it between vertical bars, $|\boldsymbol{E}|$ and $|\boldsymbol{H}|$ respectively or by simply using regular type E or H.

We know that, in plane or TEM waves, the $\boldsymbol{E}$ and $\boldsymbol{H}$ field components and the direction of propagation are mutually at right angles. We may assume a rectangular (x, y, z) co-ordinate system for the electric field (x), the magnetic field (y) and the direction of propagation (z). From the TEM wave definition therefore, neither the electric nor the magnetic fields would exist in any other planes. The *time average* PFD P_{zAV} of a plane wave is given by [3]

$$P_{zAV} = \frac{E_{xRMS}{}^2}{\eta_0} = H_{yRMS}{}^2 \eta_0 \quad W/m^2 \tag{1.1}$$

where

E_{xRMS} is the root mean squared (RMS) electric field in volts per metre (V/m);

H_{yRMS} is the RMS magnetic field strength in amperes per metre A/m;

η_0 is the intrinsic impedance of free space ($120\pi\ \Omega$ or approximately 377 Ω).

At the risk of digressing too much, the time average PFD perhaps requires a little more explanation. Time average refers to the power averaged out over a time that is long relative to the oscillation period of the $\boldsymbol{E}$ and $\boldsymbol{H}$ fields. For example, if the frequency of the RF emission was 100 MHz, the period of the $\boldsymbol{E}$ or $\boldsymbol{H}$ wave would be 10 ns, but the period of the power wave oscillation would be half this value at 5 ns on account of the square law relationship we saw in (1.1). To make a reliable measurement of the time average PFD, the averaging time of the measuring instrument used would require to be *many times* this value. In engineering we often need to make the *many times* assumption and this is often the subject of debate amongst engineers but most agree that a lower limit of about 10 works well for most estimations.

1.2.1 Non-Ionising Radiation Hazards in the Vicinity of Transmitting Antennas

There may be significant risks of RF burns relatively close to antennas, not necessarily in the far field, which are transmitting even quite modest powers. For example most broadcast transmitters, which often transmit quite significant powers, operate at frequencies in the MF, HF and VHF frequency bands. (The designations for these and other frequency bands are defined in Table 1-1). The behaviour of the fields themselves and the associated PFD in the *near field* region makes them difficult to predict. Regions on the transmitter site accessible to authorised personnel and close to the antenna are often actually in the near field. Most of the propagation theory related to the fields produced by antennas assumes the far field only and does not work well, if at all in the near field. It is not useful to get too involved with any near field theory yet because it is not possible using this to accurately predict what might happen on a practical site. However the following are some suggested precautions that should be followed in areas around transmit antennas, irrespective of whether they are in the near or far field.

Hazards beyond the site perimeter

Personnel approaching the site should routinely wear a calibrated non-ionising radiation exposure meter (such as the Narda-Alert [35]) as well as always when on site. Since the last inspection, storm damage might have dislodged feeders or antenna elements still connected to the transmitter and left them in accessible areas. VSWR trips and telemetry feedback alarms may have failed or not operated reliably. The transmitter may still be operating and the feeder may actually be radiating dangerous levels of RF.

How reliable are the records of what is operating at the site?

Many sites today are rented out to several different operators and the records of what active services are present at the site might be out of date, unreliable or misleading. There might be a risk of unexpected emissions causing a hazard. The non-ionising radiation exposure meter should help identify such unexpected hazards.

Radiating structures

In particular at LF and MF, antenna masts often form radiating elements which are part of the antenna itself. They may be mistaken simply for harmless mechanical supporting structures, but can produce hazardous fields.

High levels of coupled fields

Any odd length of conducting material such as a mains extension cable, an overhead telephone cable or a length of metal fencing can act as a receiving antenna and collect significant coupled voltages. Either may present a hazard to anybody who happens to be nearby or touching it.

Transmit antennas are also good receiving antennas

Even if a transmitter feeding an antenna is switched off, the antenna itself can still operate as a very effective receive antenna and collect coupled field from other nearby emitting antennas, *even if they operate in different frequency bands*. The antennas might only be separated by a few metres so the field strengths can still be very high. On multi-operator sites there might be sufficiently high level fields nearby to couple hazardous levels of RF voltages back down the feeder cable. This might be a hazard to personnel working on or near the (de-powered) transmitter.

Static build up and lightning

Although not strictly an RF hazard, antenna structures and elements that are insulated from ground may still have significant capacitance relative to ground and allow the build up of electrostatic charges of thousands of volts. These originate from passing charged clouds and air molecules and there may be a risk of lightning discharges through parts of the antenna structure. Any of these might find an easier route to ground through a nearby human body than through any insulators or installed lightning protection equipment. If it is necessary to work near any such antenna structure it should be securely connected to ground.

Other hazards

Of course there are many non-electrical hazards on such sites but these will not be addressed as they are beyond the scope of this paper.

1.2.2 Personal Exposure Limits

In recent years there has been a vast proliferation of all forms of radio (or wireless) communications and therefore also of antenna sites associated with these. High profile, multi-user, transmitting sites, a less tolerant general public and higher levels of accountability has encouraged research into the biological effects of non-ionising radiation. Government bodies such as the Office of Communications (Ofcom) in the UK or the Federal Communications Commission (FCC) in the United States have drawn up maximum personal exposure limits (MPEs) for the human body in terms of power flux density.[36] More recently, these have specified differing thresholds for separate frequency bands. For example, the exposure limit from 30 MHz to 300 MHz is 20 dB less than the equivalent figure for below 1 MHz. These have arisen from research which has found that the average human body absorbs differing levels of non-ionising radiation as a function of its wavelength, which is a reciprocal relationship to its frequency. The more up to date MPEs have also removed some of the ambiguity about which operators of multi-user sites are responsible for corrective action should the limits be exceeded.

1.3 Physical Properties of Antenna Systems

Physical aspects to be considered for a new antenna system include:

- acceptance of the proposal by the planning authorities;
- environmental impact;
- wind loading and weight;
- available power supplies.

In most countries a new installation would require to comply with the local planning rules. Sometimes small receive only installations, such as terrestrial TV antennas do not require such permission provided they comply with generic rules on height and size. The environmental impact would usually be included in the assessment by the local planning department.

The ability of a correctly installed antenna to withstand high winds can be assessed using its wind loading factor. Exposure to winds in excess of this could damage the antenna or pose a hazard should it collapse or drop debris.

Transmit antennas must be capable of withstanding high RF power levels without degradation and a harsh environment of rain, wind, ice and snow in may parts of the World. Some antennas are covered with a low loss dielectric enclosure to reduce ingress and heaters to melt ice and snow accumulations which might otherwise absorb and distort the radiation pattern. However, such options increase the cost and complexity of the chosen antenna system. The feed system would often be pressurised at a small positive pressure to discourage ingress should a small leak occur.

1.4 Typical Antenna Configurations

These are some typical radio communication system architectures:

- point to multipoint terrestrial;
- single frequency networks
- point to point terrestrial;
- free space
- mobile/cellular
- mesh networks

1.4.1 Point to Multipoint Terrestrial

Point to multipoint terrestrial radio communication systems may also be known as terrestrial broadcast systems. The mode of communication for broadcast is simplex, which means the propagation is in one direction only. Of course it is from the transmit antenna to, potentially, many receive-only antennas simultaneously. In addition to receivers and receiving antennas at fixed locations, there may be a requirement for fixed mobile or fully mobile reception. Fixed mobile allows for the receive antenna and therefore the receiver itself to be freely moved around the service area but actually stationary during reception. Fully mobile allows for an acceptable quality of service whilst the receiver and receiving antenna may be either stationary or moving.

Broadcast systems cover many frequency ranges from the low frequency (LF) band to the super high frequency (SHF) band. In the last century or so during the growth of radio communications many different types of antennas have been developed for use across these bands. The frequency band names covering 3 kHz to 30 GHz, including LF and SHF, are given in Table 1-1.

Frequency Band		
Name	**Acronym**	**Range**
Very low frequency	VLF	3 kHz to 30 kHz
Low frequency	LF	30 kHz to 300 kHz
Medium frequency	MF	300 kHz to 3 MHz
High frequency	HF	3 MHz to 30 MHz
Very high frequency	VHF	30 MHz to 300 MHz
Ultra high frequency	UHF	300 MHz to 3 GHz
Super high frequency	SHF	3 GHz to 30 GHz

Table 1-1 Frequency band naming conventions

The bulk of terrestrial broadcast systems operate in the MF, HF, VHF and UHF frequency bands. Although there is no formal definition, microwave frequencies are generally considered to be those above approximately 1 GHz.

A high investment is usually put into the transmit antenna itself, its location and support equipment because it must provide an adequate service to many low cost consumer receivers. For most topographies the optimum position for the transmit antenna is at some reasonably central location in the service area. Such a position would require a reasonably omnidirectional radiation pattern in the horizontal plane.

However, with the scarceness and high cost of land in city areas, finding such a site is often difficult so transmitting antennas must often be located at non-optimum positions sometimes on the edge of the service area. Fortunately there are many techniques which may be used to shape their radiation patterns accordingly to create *directional* antennas. These types of antenna would usually not comprise a single element but instead an array of elements and reflectors to achieve an effective signal distribution to cover the service area. For critical services and those with high audiences, there may also be one or more redundant antenna configurations to allow continuous service during maintenance or repair activities. Antenna configurations for providing efficient service areas are addressed in Section 3.2.3.

1.4.2 Single Frequency Networks

A single frequency network (SFN) is made up from several relatively small point to multipoint service areas. The transmitter powers, antenna characteristics and locations are deliberately chosen to create some significant areas of coverage overlap and the same service will be provided from each transmitter using the same channel allocation. In the overlap areas it will be possible to receive appreciable signals simultaneously from more than one transmitter. The transmitter antenna installations are normally on a small scale, perhaps located at existing antenna sites or on convenient tall buildings.

If an SFN used analog modulation, it could create areas of significant interference in the overlap areas due to out of phase signals cancelling out. Sometimes it is possible to minimise overlap areas for example, by using natural obstructions, directional antennas and adequately low power levels such as might be used in a tunnel re-broadcast system. Whilst this is still a SFN, the rebroadcast system is a very small localised part of the network and the overlap area, if any, would be extremely small and the interference insignificant.

The normal understanding of a SFNs today is a network comprising transmitters of broadly similar coverage areas and carrying a complex multi-carrier digital modulation such as coded orthogonal frequency division multiplex (COFDM). These are typically used for digital terrestrial radio and television broadcasting and include built in delay processing algorithms to intelligently cater for simultaneous signals at the receiver coming from more than one source. To allow for reliable and fully mobile operation, handovers when moving from one transmitter service area into another must be transparent to the user. SFNs are employed widely for the broadcast of audio services intended for reception on mobile receivers.

1.4.3 Point to Point Terrestrial

With point to point terrestrial radio systems a service is provided from one geographically small location to another similar location on the Earth’s surface. Communications may be either simplex, duplex or half duplex. The description of these and other modes of operation of a communications link are given in Table 1-2.

Mode	Description	Example
Simplex	Communications is possible in one direction only.	A broadcast system. To allow mass reception of a service with low cost receivers.
Duplex	Communications possible in both directions simultaneously	A microwave telecommunications link between the mainland and an island. A low capacity service where investment in satellite or cable is not justified.
Half-duplex	Communications are possible in both directions but not simultaneously. Requires strict rules of operation.	An emergency rescue team.
Mesh	Communications are possible in either direction with any other station with which a connection is established.	A wireless network to provide high speed data communications replace the ‘last mile’ analog telephone system.

Table 1-2 Modes of radio communications with examples

A point to point service would normally include highly directional transmit and receive antennas for the following reasons:

- to avoid wasting power by radiating only in the direction required;
- to reduce the risk of interference to other services and to allow frequency reuse elsewhere off the beam;
- to minimise received interference from sources off beam;
- to reduce the risk of multipath reflections;
- to reduce the risk of interception.

For the same type and size of antenna, the gain increases in proportion to the square of the frequency and the beamwidth reduces in direct proportion to the frequency. Therefore with line of sight (LOS) links it is advantageous to use the highest possible frequency that the licensing and propagation conditions will allow. This is one reason why many LOS links operate at the relatively high microwave frequencies for which relatively high gain, narrow beamwidth antennas are available. The higher operating frequencies also generally have proportionally wider bandwidth allocations and therefore increased capacity. This must be balanced against the higher cost of the hardware and installation required for microwave installations.

1.4.4 Communications Through Free Space

In free space communications, the full extent of the medium between the transmit and the receive antenna is free space and not subjected to the influence of any surfaces or objects which might cause reflections, refraction or diffraction.

The free space criteria may be applied to wave propagation through the Earth's atmosphere, even though it is not a vacuum, provided that:

- no reflection is possible from the Earth's surface or anywhere else;
- the air in the propagation path is uniform or well 'mixed up' with no significant temperature gradients or pollution;

However, there are several absorption mechanisms caused by atmospheric gases, in particular the molecular covalent bonds of the oxygen molecule (O-O) and the oxygen-hydrogen molecule (O-H) as present in water vapour.

An example of free space communications would be that of a space communication link between an Earth station and an Earth orbiting satellite or a deep space platform

The frequency bands used for free space communications extend most commonly from the UHF into the SHF spectra. Extending frequencies higher into the microwave region allows increasing antenna gains and reducing beamwidths to be achieved with antennas of manageable dimensions that can be easily deployed on space platforms. Narrow antenna beamwidths reduce the impact of propagation via reflections but require very accurate control of their positioning, or 'pointing', in three dimensions. Also a higher operating frequency allows a greater bandwidth allocation and increased capacity.

The use of highly directional antennas at both ends of a free space link will work if the pointing direction of each is known to a high precision and can be controlled accurately. Such precision might be a reasonable requirement on the Earth's surface but in an Earth satellite link special provision has to be made for maintaining communication over the link under the following conditions:

- when the satellite is being initially deployed to its intended geostationary position;
- if the satellite should go off track.

Therefore communication with the satellite for control and telemetry purposes normally uses antennas which are as close to omnidirectional as possible. Many satellites use three or more simple quarter wave antennas whose outputs are combined to achieve this. Then, whatever the orientation of the satellite, it should be possible to maintain a reliable control channel to it to hopefully bring it back to the correct position and orientation.

1.4.5 Mobile/Cellular Networks

Mobile (or cellular) communication systems require to provide total coverage to specific geographic areas with traffic capacities tailored to the population densities of the areas served [2]. This is known as population penetration.

The reference to cellular networks was coined from the earlier implementations of mobile telephone networks which were designed to cover the total service area using individual hexagon shaped small service areas, each with a dedicated base station, as shown in Figure 1-1. The hexagonal cell shape was chosen from the family of regular polygons for two reasons:

- the service area can be covered without gaps;
- the hexagon allows a convenient antenna beam geometry to be used at the base stations comprising 3 antennas mutually angled at 120° in azimuth and each with 120° azimuth beamwidth.

In fact, the base-station may be located either at the center of each hexagon, or at each of the intersections of the hexagon sides, the resultant coverage over several cells being unchanged.

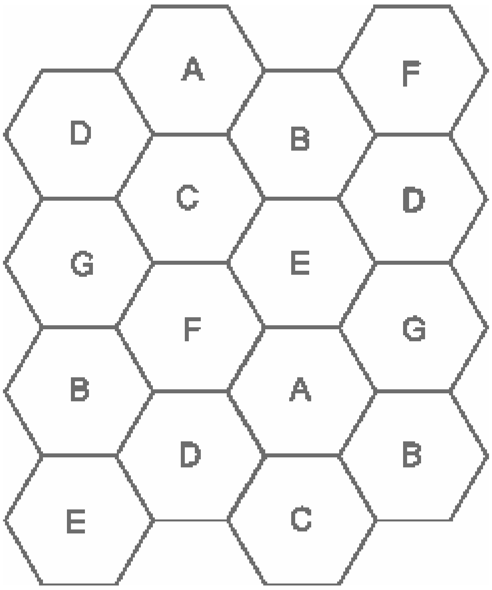

Figure 1-1 The original concept of the cellular network with hexagonal cells. Each letter represents a unique frequency group used by that cell

A new cellular network requires a large infrastructure investment so it must be capable of a rapid roll out to bring in revenue early yet to allow the network to evolve with little or no need to retrofit the earlier installations. The mobile network is also required to connect to the public switched telephone network (PSTN) to provide a comparable level of service within the normal constraints of a radio system. PSTN services within national boundaries operate in full duplex (communications being possible in both directions simultaneously), so also must the mobile communication systems. In the earlier first generation (1G) networks, duplex operation was provided in the frequency domain but in later generations (2G, 3G and 4G) it has been implemented in the time domain.

Normally the areas of high population density and potential high density such as busy roads, airports and stadia are served first. Capacity can be increased by subdividing cells into smaller ones to fill in shadow areas and to further increase capacity. As they become smaller and especially in urban areas the hexagonal structure tends to break down and the smaller cells tend to take on the shapes of typical urban areas such as streets and squares. Frequency planning for the 1G and 2G networks ensures that no adjacent cells operate using the same frequency, thus minimising the risk of interference. For small cells the power levels are reduced accordingly to reduce the risk of RF penetrating beyond its service area.

The cellular concept allows:

- a power limited coverage by each cell allowing the same frequencies to be used again at a distant cell thus allowing efficient use of the relatively small bandwidth allocated to the service.
- flexibility to subdivide cells into smaller ones whilst still allowing frequency reuse and thus increasing capacity.[27]

Choice of cell size, and therefore the antenna systems required have two conflicting requirements:

- A large cell will cover a large geographic area economically but will offer limited capacity. This might prove insufficient as the network develops.

- Small cells allow widespread frequency reuse and therefore increased capacity but at extra infrastructure expense. However the antenna structures will be simpler, easier and cheaper to install but in greater quantity.

Unfortunately it is not possible to increase the cell size to huge proportions, for example by using high altitude high gain base station antennas, because the maximum RF output power of mobile handsets is limited. This limitation is due to the available battery technology and public concerns of a perceived non-ionising radiation hazard to the user. The base station receiver will have a noise threshold below which signals will not be detectable. A mobile transceiver (handset) in use at the edge of the cell service area might literally not have enough output power to provide an acceptable signal to noise ratio when received at the base station.

1.4.6 Cell Types

The coverage for the second generation (2G) of digital mobile communications systems has evolved to include the following cell types:

- macro cells
- micro cells
- pico cells
- umbrella cells

Macro Cells

Macro cells are the largest in which the base station antennas are normally fixed to buildings or structures above the normal roof level and the structure we looked at in Figure 1-1 is typical. In many cases, the base station antennas would be erected on sites shared with other wireless users. The typical radius of a macro cell might range up to 35 km. This arrangement ensured that there was very little leakage outside of the cell and therefore allowed frequency reuse at a nearby non-adjacent cell. If each antenna has a beamwidth of approximately 120° this represents an economic and efficient way of providing coverage. The hexagonal cell shape however tends to break down in mountainous, hilly or urban areas. The antennas themselves must be sufficiently durable to withstand adverse weather for long periods without maintenance.

Macro cell base station structures are designed to allow some adjustment of the antennas in both azimuth and elevation. Azimuth adjustment allows the antenna beam direction to be refined for the best service. Elevation might be adjusted from anywhere between just above to just below the horizontal. There are two reasons for this:

- to accommodate the local topography
- to exploit the profile of atmospheric refractivity with height.

Refractivity is a measure of the refractive index of a medium such as air which is very close to that of free space (unity) and is scaled to allow small changes to be easily identified.

The profile of the land across the service area may require the antenna to be angled slightly up or down to get optimum coverage. For example if the base station was at the top of the hill to provide service to a nearby town in a valley, a slight downward tilt might be required.

Except for the relatively rare *anomalous propagation* conditions, the average refractivity profile of the atmosphere with increasing altitudes decreases on account of the reducing air molecule density. The refractivity profile may be modelled as several thin atmospheric layers with gradually reducing refractive indices as the altitude increases. These tend to refract (bend) the direction of propagation of plane waves from base station back towards the surface of the Earth. This property is an analogy of Snell's Law as applied optical beams which of course are also electromagnetic waves but at a much higher frequency. This small degree of bending can be exploited to maximise the coverage area for larger cells which are not limited by capacity, such as might be required in rural areas. A small positive (upwards) elevation from the horizontal might therefore achieve slightly improved coverage beyond the optical horizon.

Later developments with macro cell antennas include smart antennas. The beam of a smart antenna may be adjusted electronically in azimuth, elevation and even beamwidth. This 'beam steering' property may be used to increase the capacity of certain areas at short notice, though normally at some expense to coverage elsewhere. For example, if the base station includes a busy highway, this may become congested within a matter of minutes. A smart antenna commanded to move the beam in line with the road might provide the short term and temporary increase in capacity required when this occurs.

Micro Cells

Micro cells are typically deployed in local urban areas, or even to individual streets. The base station antennas are normally located below the average roof level, perhaps on utility poles or on the common street furniture found in urban areas. The antennas would often be a small dipole or phased array types adjusted to beam towards the service area.

Pico Cells

Pico cells are even smaller than micro cells, often covering indoor areas of large buildings where a good service is required such as railway stations and airports. The antennas would be small and typically with some limited facility of controlling the radiation pattern. The unit being indoors does provide some protection so the antenna design can be of cheaper construction.

Umbrella Cells

Umbrella cells are tailored to filling in service areas where there was previously poor or no coverage. Inevitably some poor coverage areas occur in the original (macro) cells due to propagation conditions, obstructions by buildings, hills etc. Umbrella cells can be installed on a one by one basis to cover these. Each umbrella would of course use a differing set of frequencies from the original macro cell.

1.4.7 Mesh or Ad Hoc Networks

Mesh networks allow the development of communication services cheaply and quickly, typically around two types of environments:

- urban areas where there may be significant obstructions to the possible radio paths caused by buildings;
- the typical battlefield environment where redundant paths are required and the network is resilient to equipment failure;

Each station, known as a node, can act as either a base station or a consumer station and will communicate with all others that are 'visible' from it. These types of networks may be used to provide broadband communications services to individual homes and businesses. They may be introduced to compete with the 'last mile' digital services such as asymmetric digital subscriber lines (ADSL) and broadband cable, also known as data over cable service interface specification (DOCSIS). Alternatively, such a system may be considered economic in areas where the existing ‘last mile’ analog infrastructure is either not developed or is unreliable. The antenna radiation patterns for such equipment are usually near omnidirectional as this provides the greatest flexibility of a new link being irrespective of direction.

A section of a typical mesh network in an urban in shown in Figure 1-2. Blocks A to G represent the plans of buildings, each of which is prepared to act as a node and B and E are taller than the others. The lines indicate bi-directional high capacity line of sight (LOS) digital radio links and the dotted line connected to H is the connection to the telecommunications infrastructure. If B and E are sufficiently tall to obstruct the LOS path between H and C, H and D, F and C and A and D then LOS links can be set up using the routes shown to bypass B. The links also form a ring and therefore some redundancy: should one of them fail communications can be maintained using the alternative route. In a well developed mesh network, all nodes are connected to all other nodes, ideally with as few as possible single hop paths.

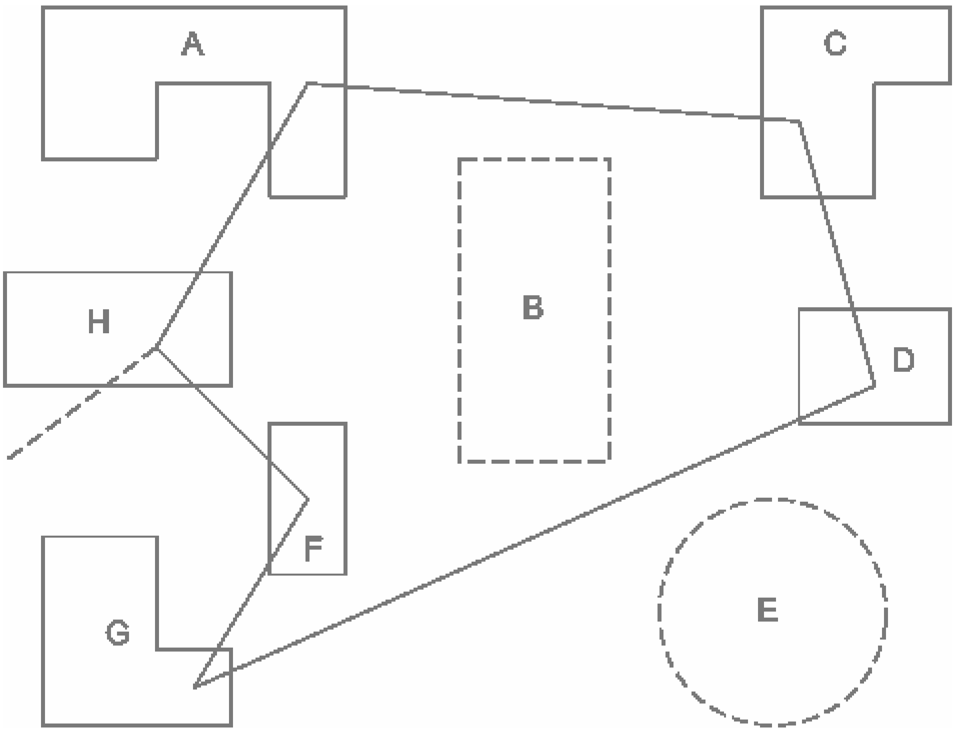

Figure 1-2 Connections established by a typical mesh (or ad hoc) network. Base station (node) antennas are omnidirectional as they might be required to communicate in almost any direction.

2 RADIATION CHARACTERISTICS OF THE HERTZIAN DIPOLE IN FREE SPACE

2.1 The Isotropic Radiator

An isotropic radiator is a theoretical antenna that is of negligibly small size and radiates equally in all directions in 3 dimensions, so its far field radiation pattern is therefore a sphere. It is assumed to be loss-free and so is 100% efficient and has a linear gain of unity. The isotropic radiator is a convenient reference antenna to use for defining antenna gains, especially for deployment in conditions equivalent to free space. An antenna gain measured using the logarithmic decibel (dB) unit relative to an isotropic radiator will have its value expressed in dBi, so the isotropic radiator itself has a gain of $0\ dBi$. In developing the antenna theory it is often more convenient to use linear units so if the logarithmic symbol dBi is not included at the end of the line it is understood that the result is linear.

2.2 Development of the Field Equations

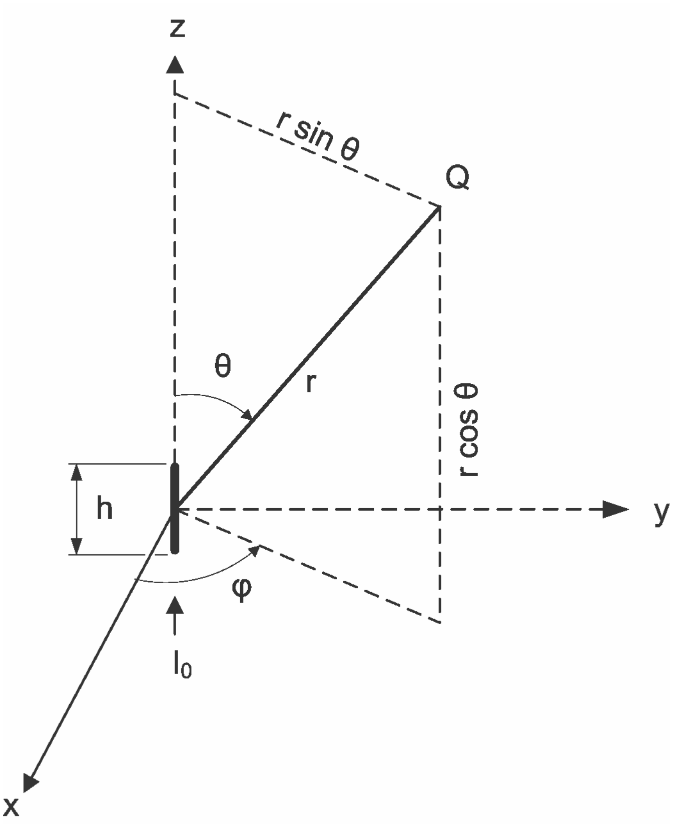

Figure 2-1 Spherical co-ordinate system used for the Hertzian dipole

A linear single element wire antenna may be considered as a large number of very short dipoles connected in series. Each of these is known as a Hertzian dipole.[1] The theory of the Hertzian dipole is fundamental to understanding the behaviour of practically all antennas which are formed from wire elements. To be considered short, the actual length l of the short dipole must be much less than the free space wavelength of the (peak) current I_0 which is exciting it ($l \ll \lambda$), and so short that the current distribution along it is effectively uniform. On this scale it behaves like a half wave dipole being fed by a perfectly balanced transmission line which does not radiate and therefore has no effect on the radiation characteristics. The current variation with respect to time is assumed to be sinusoidal of the form $i = I_0 e^{j\omega t}$ and it is understood that the coefficient $e^{j\omega t}$ is present but not shown.

In order to study the fields produced by the Hertzian dipole, it may be positioned at the origin of a spherical (r,ϕ,θ) co-ordinate system, with its axis is aligned with the z axis of a rectangular (x,y,z) coordinate system [8]. Figure 2-1 shows the how the dipole is positioned related to the two co-ordinate systems. The angular values of ϕ and θ are used to describe the antenna's azimuth and elevation performance respectively.

When the radiation pattern of an antenna is examined in the far field, the position of the antenna itself approximates to a point coincident with the origin of the spherical (and rectangular) co-ordinate systems. The heavy dependence of parameters such as PFD and field strengths on the distance from the antenna accounts for the popularity of the spherical co-ordinate system.

The antenna is assumed to be in free space and not subject to any influence of the surrounding media by reflections, refraction or diffraction. The analysis starts with Maxwell's equations and uses the concept of the *retarded potential* [7]. The expressions for the *peak* components of the $\boldsymbol{H}$ (H_ϕ) and $\boldsymbol{E}$ (E_r and E_θ) are given by the following [8]:

$$H_\phi = \frac{I_0 h}{4\pi} e^{-j\beta r}\left(\frac{j\beta}{r}+\frac{1}{r^2}\right)\sin\theta \qquad (2.1)$$

$$E_r = \frac{I_0 h}{4\pi} e^{-j\beta r}\left(\frac{2\eta}{r^2}+\frac{2}{j\omega\varepsilon r^3}\right)\cos\theta \qquad (2.2)$$

$$E_\theta = \frac{I_0 h}{4\pi} e^{-j\beta r}\left(\frac{j\omega\mu}{r}+\frac{1}{j\omega\varepsilon r^3}+\frac{\eta}{r^2}\right)\sin\theta \qquad (2.3)$$

where

h is the length of the dipole;

I_0 is the peak (temporal) current through the dipole;

η is the intrinsic impedance of the medium;

β is the phase constant;

ε is the permittivity of the medium;

μ is the permeability of the medium;

$\omega = 2\pi f$ is the angular frequency of the current.

There are no r nor θ components of $\boldsymbol{H}$ nor ϕ components of $\boldsymbol{E}$ due to the symmetry of the structure around the z axis. In general the field behaviour for the three vector coefficients in (2.1), (2.2) and (2.3) may be separated into three terms: those varying as $1/r$, $1/r^2$ and $1/r^3$ respectively.

For fields *very* close to the dipole, when r is very small, the $1/r^3$ terms in E_r and E_θ become significant compared to H_ϕ. This region is sometimes known as the electrostatic region and originates from the charges on the antenna element itself.

For fields that are slightly more distant, the $1/r^2$ term becomes significant in H_ϕ, E_r and E_θ. This is known as the inductive region.

In most *traditional 'open air'* communication systems the transmit and receive antennas are displaced by distances of many wavelengths. In this region the important terms in the components of $\boldsymbol{E}$ and $\boldsymbol{H}$ are those varying as $1/r$ and the expressions for H_ϕ and E_θ are:

$$H_\phi = \frac{j\beta I_0 h}{4\pi r}\sin\theta\, e^{-j\beta r} \tag{2.4}$$

$$E_\theta = \frac{j\omega\mu_0 I_0 h}{4\pi r}\sin\theta\, e^{-j\beta r} = \eta_0 H_\phi \tag{2.5}$$

For normal atmospheric (terrestrial) communications, the transmission medium would be air. For a space link it would of course substantially comprise free space. Fortunately is does not make much difference to the result, whichever is chosen. In (2.5) therefore, the intrinsic impedance term η is actually that of free space so the zero suffix is added, with a value of $120\pi \quad \Omega$ [12], so.

$$\eta_0 = 120\pi \quad \Omega \tag{2.6}$$

From plane wave theory,

$$\eta_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} = c\mu_0 = \frac{\omega\mu_0}{\beta} \approx 120\pi \quad \Omega \tag{2.7}$$

It can be seen from (2.4) and (2.5) that H_ϕ and E_θ are in phase and, at great distances from the dipole as r increases, they approximate to the $\boldsymbol{E}$ and $\boldsymbol{H}$ components of a plane wave. A plane wave is equivalent to a transverse electric magnetic (TEM) wave in which there is an orthogonal relationship between the directions of the $\boldsymbol{E}$ field, the $\boldsymbol{H}$ field and propagation. For a TEM wave the amplitudes of $\boldsymbol{E}$ and $\boldsymbol{H}$ are also related to each other by η_0, such that:

$$\eta_0 = \frac{E}{H} \tag{2.8}$$

The $1/r$ region is known as the far field, Fraunhofer region or radiation field [5]. Generally the closer ($1/r^2$ and $1/r^3$) regions are called the near field or Fresnel region [5].

2.3 Total Power Flow and the Far Field Radiation Pattern

The general expression for power flow in regions where r is large, may be obtained by applying the Poynting Vector to what are effectively the orthogonal $\boldsymbol{E}$ and $\boldsymbol{H}$ components, E_θ and H_ϕ respectively [6]. The Poynting Vector is described in more detail in Section 3.5. We noted above that, as r is large, the surface of the sphere approaches a plane flat surface and the associated wave may be considered a plane wave or TEM wave. Again considering the region where r is large, the component of the Poynting vector which gives the time average power flow per unit area P_r , in this case radially away from the dipole, is given by the following equation.

$$P_r = \frac{1}{2}\frac{E_\theta^{\ 2}}{\eta_0} = \frac{1}{2\eta_0}\left(\frac{\omega\mu_0 I_0 h}{4\pi r}\right)^2 \sin^2\theta$$

$$= \frac{1}{2\eta_0}\left(\frac{\beta\eta_0 I_0 h}{4\pi r}\right)^2 \sin^2\theta \qquad (2.9)$$

$$= \frac{\eta_0}{32\pi^2}\left(\frac{\beta I_0 h}{r}\right)^2 \sin^2\theta \quad W/m^2$$

The coefficient of $1/2$ was included in the first line of (2.9) to allow for E_θ being a peak and not RMS value, having been derived from the peak current I_0 in the antenna element. This expression shows that the power flux density (PFD) has maxima at $\theta = 90°$ and $\theta = 270°$ and has minima at $\theta = 0°$ and $\theta = 180°$. For a fixed value of θ, the PFD is constant for a full revolution about the z axis from $\phi = 0°$ to $\phi = 360°$. The radiation pattern may therefore be expressed as being omni-directional in two dimensions (the x-y plane).

This analysis for the PFD originating from the Hertzian dipole *in the far field*, provides a definition of the far field radiation pattern for this antenna. In fact it is common to simply refer to the 'the radiation pattern of an antenna' as actually meaning its far field radiation pattern.

The PFD expression in (2.9) implies that, in 3 dimensions, the radiation pattern may be expressed as the surface of a toroid-like solid with a central 'hole' of zero radius and whose (major) axis is coincident with the z axis. The concept of this is shown in Figure 2-2 in which the Hertzian dipole is aligned with the z axis and is positioned with its mid-point at $z = 0$.

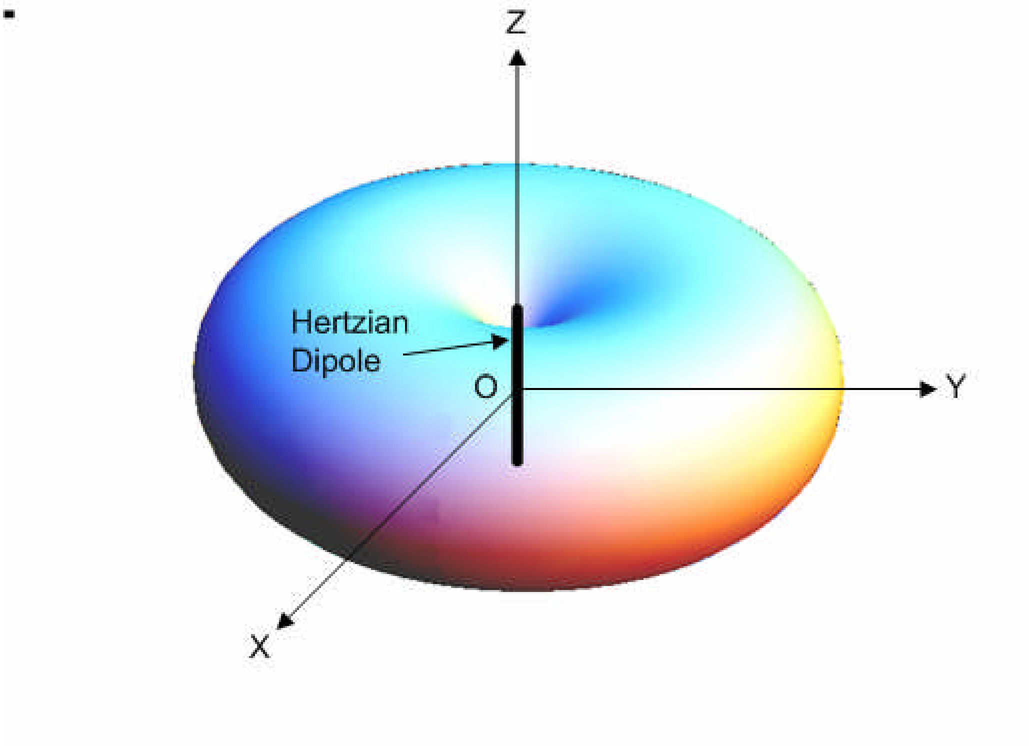

Figure 2-2 The three dimensional shape of the radiation pattern of a Hertzian dipole which is aligned with the z axis, positioned with its mid-point at z = 0

Using the spherical co-ordinate system shown in Figure 2-1, the spherical radius of a point on the surface is directly proportional to the PFD that will be measured in the far field in that same direction. An example of such a point $P(R,\alpha,\beta)$ is shown in Figure 2-3. From the definition of the chosen co-ordinate system shown in Figure 2-1, in this case $r = R$, $\phi = \alpha$ and $\theta = \beta$.

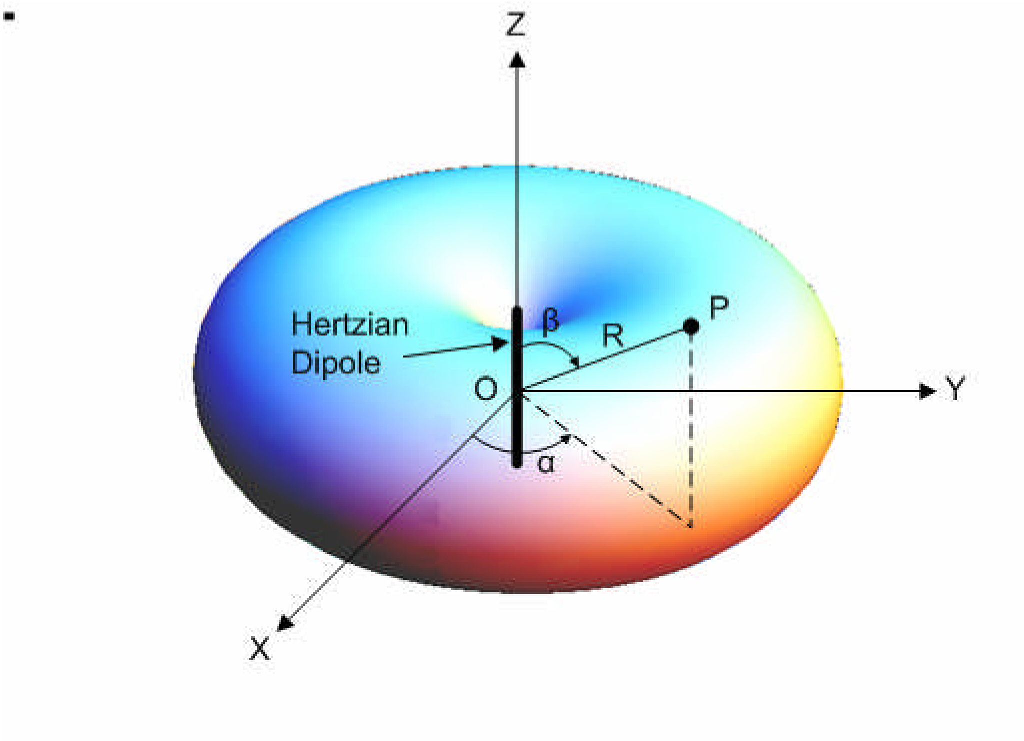

Figure 2-3 An arbitrary point P on the surface of the radiation pattern which is represented in spherical co-ordinates by (R,α,β)

The omnidirectional nature of the radiation pattern helps to simplify its consideration in two dimensions. The expression shown in (2.9) for example is a function of the elevation angle θ only and is independent of the azimuthal angle ϕ.

The total power radiated in all directions from the Hertzian dipole W_{av} *Watts* may be obtained by integrating the PFD over the surface area of a sphere surrounding it. This yields the expression shown in (2.10). (The standard integral for a sine cubed function might be useful, which is repeated in (2.11) [13]).

$$\begin{aligned} W_{av} &= \int_{\theta=0}^{\theta=\pi} 2\pi r \sin\theta P_r d\theta = \frac{\eta_0 \beta^2 I_0{}^2 h^2}{16\pi} \int_0^{\pi} \sin^3\theta d\theta \\ &= \frac{\eta_0 \pi I_0^2}{3}\left(\frac{h}{\lambda}\right)^2 = 40\pi^2 I_0{}^2\left(\frac{h}{\lambda}\right)^2 \end{aligned} \tag{2.10}$$

$$\int_{\phi=0}^{\pi} \sin^3\phi \quad d\phi = \frac{4}{3} \tag{2.11}$$

2.4 The Concept of Antenna Gain Relative to an Isotropic Radiator

The standard definition of (linear) antenna gain *relative to an isotropic radiator* is the ratio of the PFD in the direction of maximum radiation to the PFD that would have been generated if the same power was radiated equally in all directions. So in one sense it may be described as a measure of its ‘focussing’ capability. Remember that we are still considering the behaviour in the far field of the antenna: the properties in the intermediate and near field regions will probably be very different. An isotropic radiator is an antenna which radiates equally in all directions, in 3 dimensions. This is a theoretical concept as such an antenna which is perfectly isotropic could not be designed in practice. This is discussed further in Section 2.1.

From (2.9), the general expression in terms of θ for the radial PFD of the Hertzian dipole is

$$\begin{aligned} P_r &= \frac{1}{2}\frac{E_\theta^{\,2}}{\eta_0} = \frac{1}{2\eta_0}\left(\frac{\omega\mu_0 I_0 h}{4\pi r}\right)^2 \sin^2\theta \\ &= \frac{1}{2\eta_0}\left(\frac{\beta\eta_0 I_0 h}{4\pi r}\right)^2 \sin^2\theta \\ &= \frac{\eta_0}{32\pi^2}\left(\frac{\beta I_0 h}{r}\right)^2 \sin^2\theta \quad W/m^2 \end{aligned} \tag{2.12}$$

noting that ϕ may be ignored as the radiation pattern is axially symmetric about the z axis. The maximum PFD, say P_{r0} is therefore at $\theta = 90°$ or $\theta = 270°$, the direction of maximum radiation, when

$$\begin{aligned} P_{r0} &= \frac{\eta_0}{32\pi^2}\left(\frac{\beta I_0 h}{r}\right)^2 \\ &= \frac{120\pi\, 4\pi^2 I_0^{\,2} h^2}{32\pi^2\lambda^2 r^2} \\ &= 15\pi\left(\frac{I_0}{r}\right)^2\left(\frac{h}{\lambda}\right)^2 \quad W/m^2 \end{aligned} \tag{2.13}$$

The final expression in (2.13) used substitutions for the intrinsic impedance of free space $\eta_0 = 120\pi \quad \Omega$ and the spatial phase constant β given by:

$$\beta = \frac{2\pi}{\lambda} \quad rad/m \tag{2.14}$$

If the same total power W_{av} had been fed into an isotropic radiator then, by definition, it would have been radiated equally in all directions so the power flux density P_{ri} at a distance r from the radiator would have been given by

$$\begin{aligned} P_{ri} &= \frac{W_{av}}{4\pi r^2} = \frac{40\pi^2}{4\pi}\left(\frac{I_0}{r}\right)^2\left(\frac{h}{\lambda}\right)^2 \\ &= 10\pi\left(\frac{I_0}{r}\right)^2\left(\frac{h}{\lambda}\right)^2 \quad W/m^2 \end{aligned} \tag{2.15}$$

The gain of the Hertzian dipole G_{HD} relative to an isotropic radiator is therefore

$$G_{HD} = \frac{P_r}{P_{ri}} = \frac{3}{2} \tag{2.16}$$

2.4.1 Logarithmic Antenna Gain

There are several ways in which the *logarithmic* antenna gain may be expressed. One of the most common is in decibels relative to an isotropic radiator (dBi). If g_{HD} represents the logarithmic gain of the Hertzian dipole relative to an isotropic radiator, then

$$\begin{aligned} g_{HD} &= 10\log_{10} G_{HD} \quad dBi \\ &= 10\log_{10}\left(\frac{3}{2}\right) \\ &= 1.76 \quad dBi \end{aligned} \tag{2.17}$$

2.5 An Example

It might be interesting to substitute some realistic values into (2.12) and examine the results. If the operating frequency was 100 MHz ($f = 100\ MHz$) the wavelength λ would be given by:

$$\lambda = \frac{c}{f} = \frac{3\times10^8}{100\times10^6} = 3\,m \tag{2.18}$$

The length of the Hertzian dipole is, by definition, small compared to a wavelength. The speed of electromagnetic radiation in free space is very close to $c = 3\times10^8\ m/s$ which is used here. If we choose the length of the Hertzian dipole to be $h = 0.01\,m$, then $h/\lambda \approx 0.3\%$ should make it sufficiently short to meet the criterion. At a distance of 10 km from the antenna ($r = 10000\ m$) we should be comfortably in the far field. Supposing the peak value of the temporal current in the dipole was 1 A, then $I_0 = 1A$. From (2.7) we have the intrinsic impedance of free space ($\eta_0 = 120\pi\ \Omega$). The spatial wave constant β from (2.14) is given by $\beta = 2\pi/\lambda = 2\pi/3 \quad rad/m$ To determine the PFD in the far field, we need to substitute these values into (2.12). We have chosen to keep with System International (SI) units so the PFD result will also be in SI derived units, in this case *watts per metre squared* (W/m^2). Taking the coefficient of the sine squared term in the bottom line of (2.12) and substituting the values mentioned in the last few sentences, we have

$$\begin{aligned} \frac{\eta_0}{32\pi^2}\left(\frac{\beta I_0 h}{r}\right)^2 &= \frac{120\pi}{32\pi^2}\left(\frac{2\pi\times1\times0.01}{3\times10000}\right)^2 \quad W/m^2 \\ &= 5.236\times10^{-12} \quad W/m^2 \\ &= 5.236 \quad pW/m^2 \end{aligned} \tag{2.19}$$

This figure of 5.236 pW/m^2 will be the peak value of the sine squared waveform in terms of the *spatial* angular variable θ. That is the PFD that would have been measured by a receiving antenna in the direction defined at $\theta = 90°$ or $\theta = 270°$, one that is perfectly aligned in the horizontal plane. Do not forget that here θ is not a time (*or temporal*) variable as we may often have used it previously. It was defined in the elevation or vertical plane in our co-ordinate system shown in Figure 2-1. We have already defined I_0 as the peak value of the temporal current in the antenna element which is of course sinusoidal but *with respect to time*.

The PFD which would have been received from the isotropic radiator instead of the Hertzian dipole, with all the other parameters unchanged and using (2.16), would therefore be

$$P_{ri} = \frac{2}{3}P_{r0} = \frac{2}{3} \times 5.236\ pW/m^2$$
$$= 3.491 \quad pW/m^2 \qquad (2.20)$$

Now imagine that we examine combined plots of the radiation patterns of the Hertzian dipole and the isotropic radiator. This shown in 2 dimensions in Figure 2-4 and in 3 dimensions in Figure 2-5.

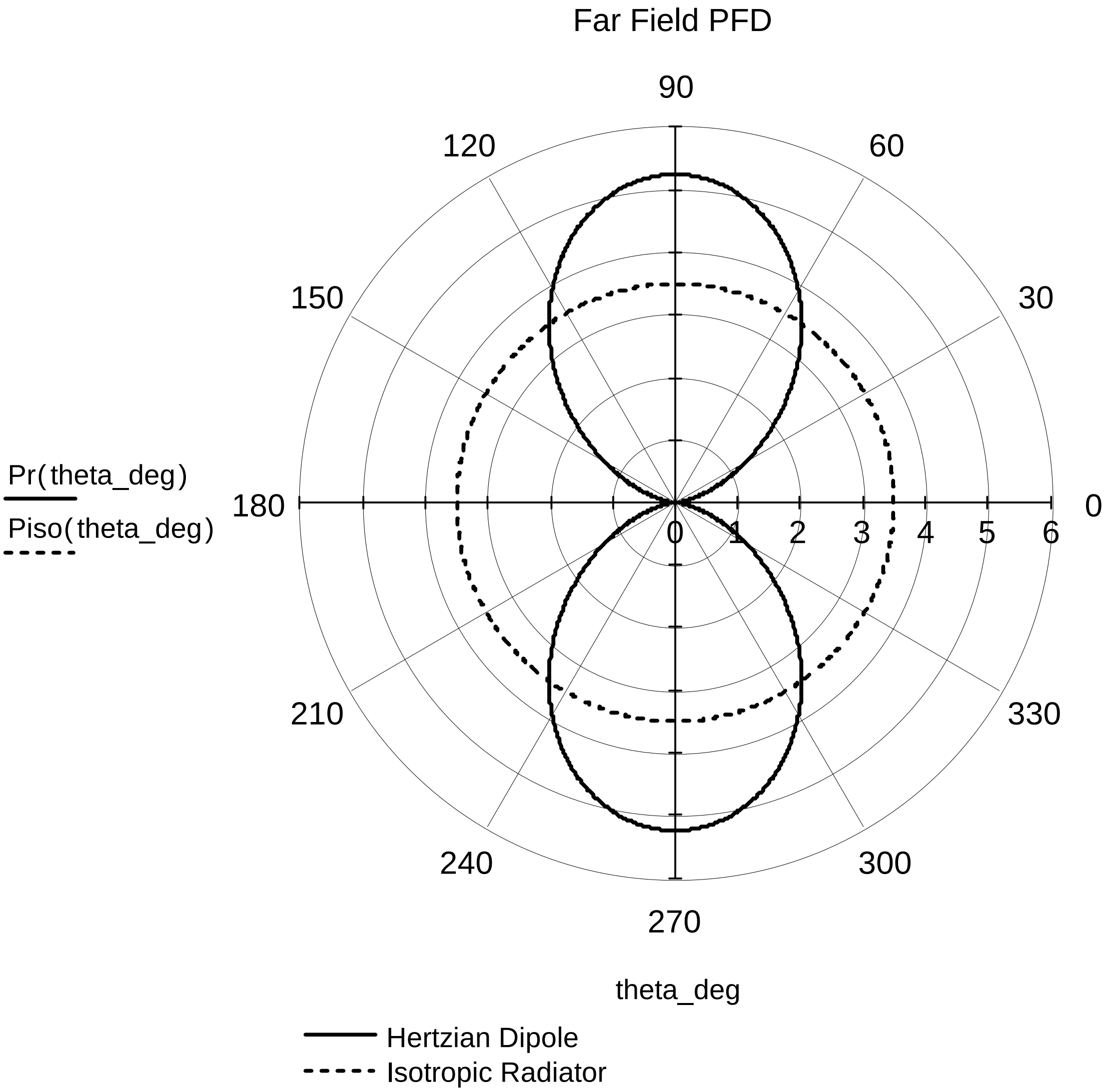

Figure 2-4 Combined far field radiation pattern plots in two dimensions for the Hertzian dipole and the isotropic radiator for the example in Section 2.5. The radial scaling is the PFD expressed in picowatts per square metre (pW/m^2)

In Figure 2-4 the radiation pattern for the isotropic radiator is perfectly circular since it is a section from the sphere which represents its three dimensional radiation pattern. For the Hertzian dipole it is more egg-shaped, in fact this is a sine-squared waveform, but plotted on polar scales.

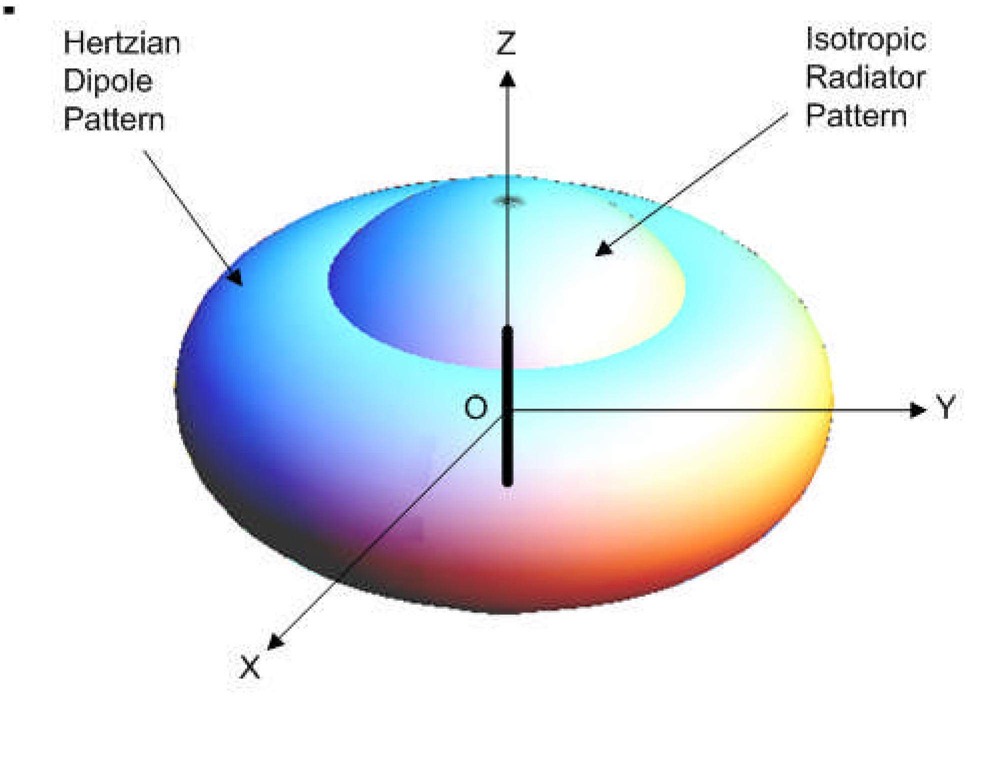

Figure 2-5 The three dimensional radiation patterns for a Hertzian dipole and an isotropic radiator to the same scales

2.6 Radiation Resistance

The radiation resistance of the Hertzian dipole R_r may be defined as the equivalent resistance required to dissipate the same power that is actually radiated [14]. We defined at the beginning of Chapter 2 that I_0 is the *peak* temporal current. Then, for a sinusoidal waveform, the average power W_{av} is given by:

$$W_{av} = \frac{1}{2} I_0{}^2 R_r \qquad (2.21)$$

By equating (2.21) with the expression for the total average power radiated from the Hertzian dipole from (2.10), the radiation resistance of a Hertzian dipole is [15].

$$R_r = 80\pi^2 \left(\frac{h}{\lambda}\right)^2 \quad \Omega \qquad (2.22)$$

Remember that the definition of the Hertzian dipole assumed $h \ll \lambda$ so the result for radiation resistance would be wildly inaccurate if we attempted to calculate it for example with $h = \lambda/2$.

2.7 Gain and Aperture Relationships

The effective aperture of an antenna is the equivalent area through which the normally incident power flux would have to pass. Assuming that the Hertzian dipole has an effective aperture A_{HD} , then *the actual* received power W_{rHD} is related to the incident PFD at the same position P_{dHD} by

$$W_{rHD} = P_{dHD} A_{HD} \qquad (2.23)$$

As the length of the Hertzian dipole is small, we can assume that the electric field across it is uniform and there is a linear relationship between the (RMS) voltage V and length h, therefore

$$V = Eh \tag{2.24}$$

where E is the RMS electric field strength at the Hertzian dipole [16].

If the Hertzian dipole was connected to a perfectly matched load equal to its radiation resistance R_r, then the mean power dissipated in the load would be W_{rHD}, and

$$W_{rHD} = \frac{V^2}{4R_r} = P_{dHD}A_{HD} \tag{2.25}$$

In the far field the waves from the antenna are assumed to be transverse electric magnetic (TEM) or 'plane'. By applying the Poynting Vector to a plane wave the PFD in the direction of propagation P_{dHD} is given by

$$P_{dHD} = \frac{E^2}{\eta_0} = \frac{E^2}{120\pi} \tag{2.26}$$

In this case the coefficient of $1/2$ is omitted from the Poynting vector because the electric field is an RMS quantity. Therefore

$$\frac{V^2}{4R_r} = P_{dHD}A_{HD} = \frac{E^2 A_{HD}}{120\pi} \tag{2.27}$$

and

$$A_{HD} = \frac{30\pi h^2}{R_r} \tag{2.28}$$

Since, from (2.22), for the Hertzian dipole

$$R_r = 80\pi^2 \left(\frac{h}{\lambda}\right)^2 \tag{2.29}$$

then

$$A_{HD} = \frac{30\pi h^2 \lambda^2}{80\pi^2 h^2} = \frac{3\lambda^2}{8\pi} \tag{2.30}$$

We therefore have both the gain of a Hertzian dipole relative to an isotropic radiator (2.16) and its effective aperture (2.30).

In general, if we have two antennas with gains G_1 and G_2 and respective apertures A_1 and A_2, if the apertures are coincident, they are related in the following way:

$$\frac{G_1}{A_1} = \frac{G_2}{A_2} \tag{2.31}$$

For the apertures to be coincident, the antennas must be co-polar and aligned mutually in their far fields for maximum transmission.

Therefore for a Hertzian dipole, assuming G_2 and A_2 are its gain and effective aperture respectively in terms of an isotropic radiator (for which $G_1 = 1$) we have the aperture of the isotropic radiator A_1 given by

$$A_1 = \frac{G_1 A_2}{G_2} = \frac{\lambda^2}{4\pi} \tag{2.32}$$

For a generic antenna of gain G and effective aperture A relative to an isotropic radiator therefore [17]:

$$\frac{G_1}{A_1} = \frac{G}{A} \tag{2.33}$$

giving

$$G = \frac{AG_1}{A_1} = \frac{4\pi A}{\lambda^2} \tag{2.34}$$

(2.34) relates antenna (isotropic) gain and effective aperture for any antenna at a particular frequency (or actually wavelength λ). Remember again that this is another *far field relationship*. It will not be valid for regions closer to the antenna. Furthermore, this equation is not limited to Hertzian dipoles. It is equally valid for any arbitrary antenna provided it is in the far field.

2.8 Regions of the Near Field and Far Field

The full expressions given in Section 2 for the $\boldsymbol{E}$ and $\boldsymbol{H}$ field components generated by a Hertzian dipole apply to all regions: close to and far from the antenna. As the distance from the Hertzian dipole increases, the source more closely resembles a point source to distant antennas and the waves radiated become closer approximations to plane (or TEM) waves. In this (far field) region a number of assumptions in antenna theory are quite valid, reliable and their use considerably simplifies system calculations and performance predictions. Provided the distance between the transmitting and receiving antennas is sufficiently great, each may be considered to be in the far field of the other and such simplifications ensure that the *shape* of the far field radiation pattern is actually no longer a function of distance. If we wish to calculate the absolute levels of PFD, as we saw in the example in Section 2.5, the distance must be included as the PFD is an inverse square law function. Fortunately most communications systems which include antennas and utilise radio propagation do indeed operate in the far field.

Other technologies are actually designed to operate in the *near* field where the $1/r^2$ and $1/r^3$ terms become significant. The measured field patterns are dependent on both the orientations and the distance from the antenna. The antennas become closer approximations to inductively coupled coils which account for the alternative description of the near field region.

Various criteria have been suggested for determining the position at which the near field becomes the far field. These are important in deciding the smallest size anechoic chamber that is necessary to perform far field measurements for particular antennas. Such definitions are generally functions of the wavelength and the largest physical aperture dimension of either the transmit antenna, the receive antenna or both. They are based on studies of how closely the progressively advancing spherical wavefront from an antenna approximates to a plane wave. One widely used definition, derived from empirical measurements, comes from MIL-STD-462D and defines two criteria according to whether the operating frequency is above or below 1.24 GHz. This is described in the following paragraphs [11]:

1. For frequencies less than or equal to 1.24 GHz the far field is defined as the greater value of (2.35), the distance between the transmit antenna aperture and the receive antenna aperture, calculated according to the following equations:

$$R = \frac{2D^2}{\lambda} \quad or \quad R = 3\lambda \tag{2.35}$$

where λ is the wavelength and D is the maximum physical dimension of the transmit antenna.

2. For frequencies greater than 1.24 GHz, the far field is defined as the value of R according to either of the following equations, dependent on how similar D and d are.

For $2.5D \prec d$

$$R = 2\frac{D^2}{\lambda} \tag{2.36}$$

For $2.5D \geq d$

$$R = \frac{(D+d)^2}{\lambda} \tag{2.37}$$

For both (2.36) and (2.37) D is the maximum physical dimension of the transmit antenna and d is the maximum physical dimension of the receive antenna.

3 COMMON ANTENNA PARAMETERS

3.1 Solid Angles

The concept of solid angles used in three dimensional geometry is useful for studying antenna properties such as gain, directivity and radiation patterns. This arises because, as we saw with the Hertzian dipole in Section 2.8, regions in the far field are sufficiently distant from the antenna that the waves originating from it may be considered: (a) to originate from a point source at the antenna location and (b) to propagate from the source along straight lines. The definition is presented with the help of Figure 3-1.

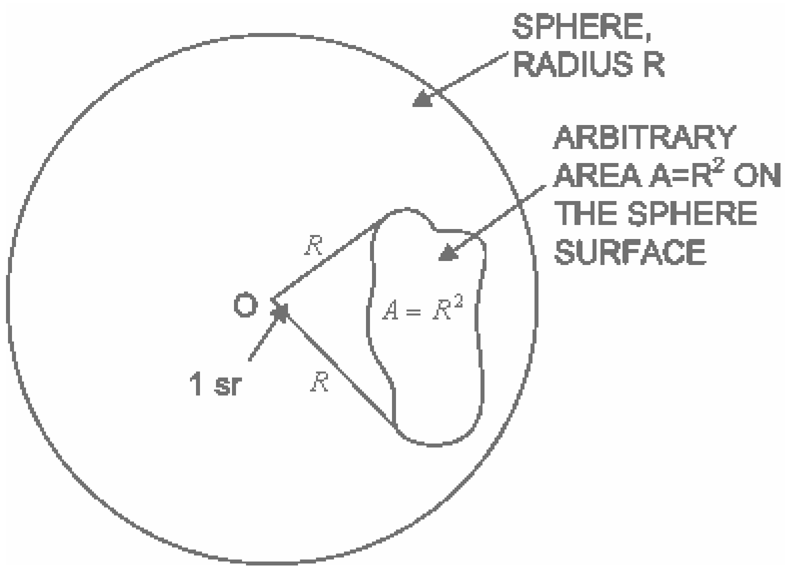

Figure 3-1 An illustration of the definition of one Steredian (1 sr)

If the source is a directional antenna it may be described in two or three dimensions by its radiation pattern. Alternatively, in three dimensions, the main beam may be bound by a specific value of solid angle. The solid angle of a source at a point Ω is proportional to the surface area A of a projection of the source onto a sphere of radius R centered at the same point, divided by the square of the sphere's radius. Therefore,

$$\Omega = \frac{A}{R^2} \qquad (3.1)$$

The unit of solid angle is the Steredian (sr). Figure 3-1 shows a projection of a solid angle of $1 \quad sr$ on to the surface of a sphere of radius R. In this case, the area of the projection will be exactly R^2. Note that the shape of the projected area is not relevant to the definition, only its area. We know that the surface area of a sphere is $4\pi R^2$ so the total solid angle of a sphere is therefore $4\pi \ sr$.

In other words, a solid angle is related to the surface area of a sphere in an analogous way that a coplanar angle is related to the circumference of a circle. A solid angle does not require to have a spherical surface projection and a coplanar angle does not require to have a circle arc projection.

Consider a cube. This may be split up into 6 identical pyramids-like shapes, each of which has a solid angle equal to $1/6$ that of a sphere. Therefore each shape would comprise a solid angle of size $4\pi/6 = 2\pi/3 \, sr$. In this case we know that the 6 pyramids fit together perfectly with no gaps so the solid angle subtended at the vertex of each is easy to calculate.

Now consider the interior of a well constructed room which has perfectly flat and vertical walls and a perfectly horizontal wall and ceiling. All of the walls, ceiling and floor join mutually at angles of exactly $90°$. What would be the solid angle of each of the corners?

We know that a sphere has a total solid angle of $4\pi\ sr$. Now suppose that the sphere is cut into 2 identical hemispheres. Then each hemisphere is cut into 2 identical halfs, giving us 4 identical *spherical wedges.* Each of these is then cut into two further identical pieces. The result will be a total of 8 identical sections. Each section will comprise 3 flat faces, mutually at right angles, the fourth face being a one eighth section of the spherical surface. The 3 flat faces of each piece will fit exactly into each of the corners of the room under consideration. With 8 identical sections and $4\pi\ sr$ in a sphere, the solid angle of each corner will be $\pi/2\ sr$.

3.1.1 Beam Solid Angle

The beam solid angle of an antenna (Ω_A) is the solid angle through which all of the power radiating from it would stream if it were concentrated in one direction with a constant PFD equal to that in the direction of maximum radiation [37]. Note that this definition includes all of the power radiated from the antenna, the main beam and all sidelobes in all directions.

To calculate the beam solid angle for a given antenna the *total* power radiated from it must be determined by integrating the power flux density over the surface area of a sphere in the far field which is centred at the antenna. One may then imagine this total power to be radiated in the direction of maximum radiation with axial symmetry and a PFD equal to that of maximum radiation. The solid angle formed in this way would be a conical-like section from the sphere with which the corresponding flux density is identical to that of maximum radiation. The section would have a curved base, being part of the sphere's surface. For many practical directional antennas this calculation results in a value very close to the solid angle at the half power (or - 3 dB) points of its main beam.

3.1.2 Major and Minor Lobe Solid Angles

The main lobe solid angle Ω_M (noting the upper case subscript) is defined in a similar way to Ω_A, described in Section 3.1.1. However, this definition relates to the power radiated from the antenna in the main lobe only so the integration area must only cover this part. Sometimes it is difficult to determine where the main lobe ends and other lobes start so it is usually adequate to monitor how the received signal level drops away from the value at maximum intensity and take an arbitrary relative level such as - 20 dB to indicate the extent of the main lobe.

The minor lobe solid angle Ω_m (with a lower case subscript) is also defined similarly to Ω_A but this applies to all other lobes except the main lobe.

From the solid angle definitions therefore the beam solid angle is the sum of the major and minor lobe solid angles:

$$\Omega_A = \Omega_M + \Omega_m \qquad (3.2)$$

3.1.3 Antenna Directivity and Gain

The directivity of an antenna is defined by the ratio of the maximum PFD intensity $P_{\max}$ to the average radiation intensity P_{av} [37]. Using solid angles therefore, if W is the total power radiated from the antenna then the directivity D is given by:

$$D = \frac{P_{\max}}{P_{av}} = \frac{W/\Omega_A}{W/4\pi} = \frac{4\pi}{\Omega_A} \tag{3.3}$$

The directivity of an antenna, a linear dimensionless quantity, starts at unity, for an isotropic radiator, and increases as the antenna becomes more directional.

The gain G of an antenna is defined in terms of the far field PFD in the direction of maximum radiation relative to the PFD at the same position from a known (reference) antenna as follows [37]:

$$G = \frac{PFD\ in\ the\ direction\ of\ maximum\ radiation}{PFD\ of\ a\ reference\ antenna\ in\ the\ same\ direction\ with\ the\ same\ input\ power} \tag{3.4}$$

(3.4) is of course a linear equation but very often in practice the antenna gain derived from (3.4) is expressed in the logarithmic form of decibels relative to a standard antenna. If the standard antenna was an isotropic radiator the logarithmic unit symbol would be dBi. Another example of reference antenna is the loss free half wave dipole for which the logarithmic unit symbol is dBd. The definition of antenna gain is very similar to directivity except that the former is defined in terms of the power leaving the antenna. A practical antenna would typically have some loss which would be taken into account in its gain definition but not in its directivity definition. Therefore the gain and directivity are identical for a loss free antenna. In many antenna applications where the absolute antenna gain is not as important as the angular resolution or its focussing capability, the directivity is a significant parameter.

3.2 Free Space Path Loss

Free space is another name for a vacuum, so within one there are no molecules or anything else that might cause absorption, reflections or diffraction of waves propagating through it.

Consider two loss free antennas, one transmit and one receive, which are aligned to be perfectly co-polar in free space and each is in the far field of the other. There is an apparent loss, known as the free space path loss (FSPL), measured from the input of the transmitting antenna to the output of the receiving antenna. FSPL is actually caused by the way in which the propagating RF power spreads out as it moves away from the transmit antenna. In fact, another name for FSPL is spreading loss. The degree of spreading follows an inverse square law as it affects the PFD and reduces as the antenna aperture becomes larger, assuming the wavelength is unchanged. In fact, at an infinitely large aperture, the beam could be considered parallel. This effect of spreading is consistent with the definition of gain arrived at in (2.34) and analogous to the experiments on the wave properties of light by Huygens and Young [38].

In the far field, each antenna appears to the other like a point source. This is entirely consistent with the inverse *linear* law as a function of distance, applied to the $\boldsymbol{E}$ and $\boldsymbol{H}$ field component magnitudes separately from equations (2.4) and (2.5). The Poynting Vector relates either of these to power flux density which will be addressed in Section 3.5.

If P_d is the PFD at a distance d from a transmitting antenna assumed to be in the far field then, by the inverse square law

$$P_d = \frac{G_T P_T}{4\pi d^2} \quad W/m^2 \tag{3.5}$$

where

G_T is the gain of the transmit antenna relative to an isotropic radiator;

P_T is the power input to the transmit antenna input.

From Section 2.7 the effective aperture of an antenna A is related to the isotropic gain of the same antenna G by

$$A = \frac{G\lambda^2}{4\pi} \tag{3.6}$$

where λ is the wavelength.

The power output at the receiving antenna P_r with an effective aperture A_r is therefore given by

$$P_r = P_d A_R = \frac{G_T P_T}{4\pi d^2}\frac{G_R \lambda^2}{4\pi} = G_T G_R P_T \left(\frac{\lambda}{4\pi d}\right)^2 \tag{3.7}$$

The relationship between the transmit power and the received power is given by

$$\frac{P_r}{P_T} = G_T G_R \left(\frac{\lambda}{4\pi d}\right)^2 \tag{3.8}$$

where G_R is the gain of the receive antenna relative to an isotropic radiator. The FSPL is defined for antennas which are isotropic radiators, so

$$G_T = G_R = 1 \tag{3.9}$$

and (3.8) becomes

$$\frac{P_r}{P_T} = \left(\frac{\lambda}{4\pi d}\right)^2 \tag{3.10}$$

Since $P_r \prec P_T$, this is actually a transmission ratio, and FSPL would be its reciprocal P_T / P_R. However, FSPL is most frequently expressed in the logarithmic unit dB, therefore.

$$FSPL = \left|10\log_{10}\left(\frac{\lambda}{4\pi d}\right)^2\right| = \left|20\log_{10}\left(\frac{\lambda}{4\pi d}\right)\right| \quad dB \tag{3.11}$$

The magnitude symbols are used in (3.11) because, by definition, a loss is expressed as a positive quantity and of course the logarithm of a value less than unity is negative.

FSPL calculations sometimes achieve unusually large values, not often encountered with decibels. For example, a satellite ground link operating at the K-band frequency of 14 GHz ($\lambda = 0.021 \quad m$) in geosynchronous orbit ($d = 3.5\times10^7 \quad m$), using (3.11) will have a FSPL of 206 dB. Fortunately such magnitudes are just an intermediate part of most link system calculations. Isotropic antennas would not be suitable for space communications as so much power would be wasted into unwanted directions. Normally a substantial part of the FSPL can be offset by the relatively high gains of the transmit and receive antennas.

3.2.1 Free Space Path Loss Related to ERP and EIRP

Equivalent isotropic radiated power (EIRP) is the product of the power fed into a transmitting antenna P_T and the gain of the same antenna expressed relative to an isotropic radiator G_T, or:

$$EIRP = P_T G_T \qquad (3.12)$$

Re-arranging (3.8) and substituting (3.12) gives the following expression for the received signal power

$$P_r = EIRP * G_R \left(\frac{\lambda}{4\pi d} \right)^2 \qquad (3.13)$$

The most convenient receive antenna to use is of course the isotropic radiator as then we do not have to worry about how it is orientated. For this case therefore G_R is unity and (3.13) becomes:

$$P_r = EIRP * \left(\frac{\lambda}{4\pi d} \right)^2 \qquad (3.14)$$

If the transmit and receive antennas are fixed, then $EIRP$ is constant and, for a fixed frequency (or wavelength) of transmission P_r is inversely proportional to the square of distance between them.

3.2.2 Effects of the Ground Reflected Rays

Previously we assumed free space conditions which implied the absence of any reflections between the transmit and receive antennas. In many cases such as in the design of terrestrial communication systems, it is inadequate to assume solely the free space conditions described in Section 3.2.1 and the inverse square law relationship of (3.14) [1]. *If the service area is reasonably flat, low loss and unobstructed*, the influence of ground reflections between the transmit and receive antennas becomes significant. Here, the term low-loss, in terms of radio wave propagation, does not necessarily imply metallic surfaces. Ordinary ground, especially if it is moisture laden will be reasonably conductive and act as a good reflector of plane waves across a range of frequencies, particularly at small angles of incidence. Although less common, the ground could also still be low loss but a good insulator and reflector such as might apply in some very dry areas.

There are essentially two routes that the wave propagation path (or ray) can take in passing from the transmit antenna to the receive antenna, the direct route and the reflected route via ground reflection. Examples of these rays are shown in Figure 3-2.

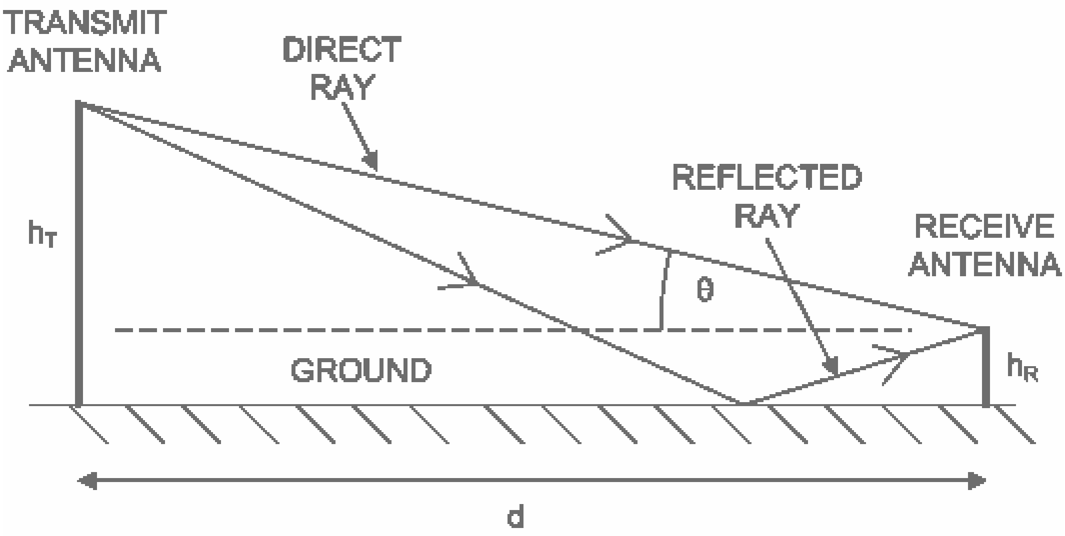

Figure 3-2 Ray diagram for propagation over a flat reflective surface

There are two slightly different results according to whether the waves are vertically polarised or horizontally polarised. These are summarised in the following paragraphs.

Vertical Polarisation

By the rules of right angle geometry and Pythagoras, the length of the route taken by the direct ray d_d is given by:

$$\begin{aligned} d_d &= \sqrt{d^2 + (h_T - h_R)^2} \\ &= d\sqrt{\left(1 + \left(\frac{h_T - h_R}{d}\right)^2\right)} \end{aligned} \qquad (3.15)$$

where

h_T is the height of the transmit antenna;

h_R is the height of the receive antenna;

d is the horizontal distance between the transmit and receive antennas.

Normally however $h_T - h_R \ll d$, so

$$\left(\frac{h_T - h_R}{d}\right)^2 \ll 1 \qquad (3.16)$$

We can then use the Maclaurin series approximation, for $x \ll 1$ [42], or

$$\sqrt{1+x} \approx 1 + \frac{x}{2} \qquad (3.17)$$

This gives the result from (3.15) of

$$d_d \approx d\left[1 + \frac{1}{2}\left(\frac{h_T - h_R}{d}\right)^2\right] \qquad (3.18)$$

For the reflected ray, using the geometric construction shown and applying the same approximation gives the result for the reflected wave of

$$d_r \approx d\left[1+\frac{1}{2}\left(\frac{h_T+h_R}{d}\right)^2\right] \quad (3.19)$$

The path difference between the direct and reflected rays Δd is therefore

$$\begin{aligned}\Delta d = d_r - d_d &= \frac{1}{2d}\left(h_T^{\ 2}+2h_T h_R + h_R^{\ 2}\right) - \frac{1}{2d}\left(h_T^{\ 2}-2h_T h_R + h_R^{\ 2}\right) \\ &= \frac{2h_T h_R}{d}\end{aligned} \quad (3.20)$$

The corresponding (spatial) phase difference, due only to path differences is the product of the phase constant in free space β_0 and the path difference.

$$\Delta\phi = \beta_0 \Delta d = \frac{2\pi}{\lambda}\Delta d = \frac{4\pi h_T h_R}{\lambda d} \quad (3.21)$$

For the VHF and UHF bands, where this type of ground reflection is most significant, and using vertical polarisation there is an additional phase shift of 180° caused by the highly acute angle between the surface and the incident ray. The total phase shift between the direct and reflected rays $\Delta\phi_T$ is therefore given by

$$\Delta\phi_T = \pi \pm \frac{4\pi h_R h_T}{\lambda_d} \quad rad \quad (3.22)$$

For TEM waves in free space, an inverse square law applies to the power flux density and a reciprocal law to the E or H fields. If the electric field is E_0 near to the transmit antenna and, at a distance d from the same antenna, it is E_0/d .

Therefore the resultant field E_R is

$$\begin{aligned}E_R &= 2\frac{E_0}{d}\cos\left(\frac{\pi \pm \frac{4\pi h_R h_T}{\lambda d}}{2}\right) \\ &= 2\frac{E_0}{d}\sin\left(\frac{2\pi h_R h_T}{\lambda d}\right)\end{aligned} \quad (3.23)$$

Normalising with respect to the maximum possible field strength at the receiving antenna,

$$\begin{aligned}\frac{E_R}{E_0/d} &= 2\sin\left(\frac{2\pi h_T h_R}{\lambda d}\right) \\ &\approx \frac{4\pi h_R h_T}{\lambda d}\end{aligned} \quad (3.24)$$

We can see from (3.24) that, for a typical terrestrial broadcast service to a fixed position, h_T and d would be fixed, so that the electric field as the height of the receive antenna h_R is adjusted would be sinusoidal with a maximum value of 2. This relationship is shown in part (a) of Figure 3-3. If the receive antenna had such freedom of movement, its height might be adjusted accordingly to receive the greatest possible field strength.

Horizontal Polarisation

A similar consideration for horizontal polarisation gives the result

$$\frac{E_R}{E_0/d} = 2\sin\left(\frac{2\pi h_R \sin\theta}{\lambda}\right) \approx \frac{4\pi h_R \sin\theta}{\lambda} \qquad (3.25)$$

This result is shown in part (b) of Figure 3-3.

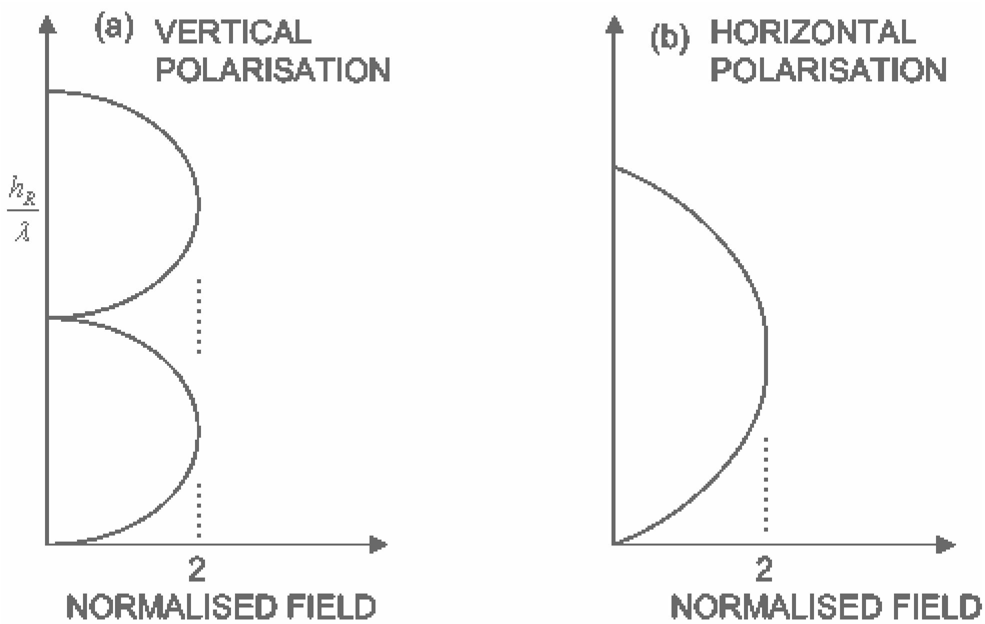

Figure 3-3 Received field strength against receiving antenna height

3.2.3 Approximating an Antenna Radiation Pattern to a Terrestrial Coverage Area

From (3.14) it was noted that, for propagation through free space, in the far field the received PFD is inversely proportional to the square of the distance from the transmit antenna to the receive antenna. A terrestrial service area would of course depart from the free space approximation and there may indeed be ground reflections to consider as we did in Section 3.2.2. Often however, the surface of the ground between the transmit and receive antennas is broken up by low rise buildings such as might be encountered in many suburban areas. Under these conditions there are few predictable ground reflections and free space propagation assumptions hold up well. If we could shape the radiation pattern of the transmit antenna in such a way that it was directly proportional to the square of the distance between it and the receive antenna, then we could estimate a contour over which the received signal level would be reasonably constant. We would then need to adjust the transmit power, and therefore the EIRP, to provide the minimum acceptable signal level at the edge of the service area. Areas inside the service area would then receive somewhat greater levels. This approach is useful, to a first approximation, when designing a new broadcast service.

3.3 Reciprocity

The property of reciprocity means that the characteristics of an antenna such as far field radiation pattern, directivity, aperture and terminal impedance are unchanged whether it is used for transmitting or receiving [28][32]. Reciprocity applies to passive antennas that are used in media that are everywhere linear, passive and isotropic.

This property is useful for simplifying transmitting and receiving antenna systems. However the physical designs of the two antenna types may be very different. For example, in point to multi-point broadcast services transmit antennas have to be designed to handle high powers, they must offer reasonable protection from the environment and they must run for extended periods without maintenance. Receive antennas are more typically low cost high production volume types.

3.4 Polarisation

An unpolarised wave is one in which the polarisation is not fixed. Sometimes a wave which originates with a fixed polarisation undergoes reflections, refraction, diffraction or any combination of these. Each of these phenomena may change the polarisation and the resulting wave(s) may comprise any particular mixture of polarisations. Such a resultant would be considered unpolarised. A polarised wave may in general be either of two types of polarisation: plane or elliptical.

3.4.1 Plane Polarisation

A plane (or TEM) wave, travelling in free space will have only one plane of polarisation which, by convention, is usually taken as the plane of the electric field [29]. Typically, for terrestrial communications, this may be horizontal, vertical or slant. Slant means an intermediate angle of polarisation between horizontal and vertical, often 45°. As free space contains no molecules, it is an isotropic dielectric medium, meaning that the dielectric properties are unchanged, irrespective of the polarisation of any wave passing through the medium. The polarisation of the wave will therefore be unchanged as it propagates through the medium and the plane of polarisation is determined by the transmit antenna. To receive the maximum signal, the receive antenna must be perfectly aligned with the transmit antenna in azimuth and elevation and its plane of polarisation identical with that being propagated. Any polarisation misalignment will result in extra loss in the link which will degrade the link performance. Maximum misalignment would occur when the planes of polarisation of the transmit and receive antennas are at exactly 90°. If the antennas are misaligned, the propagating electric field may be resolved into two components, one horizontal and the other vertical. The component which is in the same plane as the other antenna is that which will couple with it. For example, if the mis-alignment is 45°, this may be split into equal right angle components of which one will be in the same plane as the receive antenna polarisation. The magnitude of each will be $1/\sqrt{2}$ times the magnitude of the electric (or magnetic) field vector. The PFD is proportional to the square of the (electric or magnetic) field which will therefore be reduced by 3 dB compared to the perfectly aligned (co-polar) case.

3.4.2 Elliptic Polarisation

The elliptic polarisation of a wave occurs when the instantaneous plane of polarisation rotates about the axis aligned with the direction of propagation. As with plane polarisation, this is usually defined with respect to the electric vector. Right hand or left hand elliptic polarisation can be generated which rotates at a rate of 2π radians spatially for 2π radians temporally. The 'hand' of polarisation is determined by the phase relationship between them.

Elliptic polarisation implies arbitrary magnitudes of the orthogonal components that make up the electric vector. Circular polarisation, a subset of elliptical polarisation, is also the most common form, occurring when the orthogonal component magnitudes are equal. This may be set up by simple phasing of a pair of orthogonal elements at the transmit antenna.

At microwave frequencies, the shorter wavelengths and more manageable antenna dimensions allow elliptic polarisation to be generated from plane polarisation by placing a suitable anisotropic dielectric sheet combined with a grid in the far field normal to the direction of propagation of the plane polarised wave. When correctly orientated, the dielectric sheet is designed to ensure that the emerging orthogonal components are in quadrature, and therefore form and elliptically polarised wave. This component is also reciprocal meaning that a plane polarised wave can also be produced from an elliptically polarised one. An alternative technique used the phenomenon of Faraday rotation, a similar principle but using a ferrite material [33], [34].

3.4.3 Polarisation and Propagation

Practical terrestrial propagation conditions generally depart significantly from those of free space. On its route from the transmit antenna to the receive antenna, a plane polarised wave will usually undergo reflection, diffraction, refraction and various combinations of these. These are prevalent at VHF and UHF, used extensively for terrestrial broadcasting, where the obstacles encountered are relatively large in terms electrical wavelengths. In many countries these frequency bands are used for the analog services of frequency modulation (FM) sound broadcasts and analog television. Each mechanism will contribute to changes in the polarisation of the transmitted wave, whether it was transmitted in plane or elliptical form [31],[30]. Therefore, there are likely to be components of polarisation at the receive antenna that will differ from the polarisation which was transmitted. In extreme cases, for example where the wave has undergone many reflections, the amplitude components at other polarisations may be stronger than the original polarisation. Sometimes a small improvement in received signal level can be achieved by rotating the receive antenna slightly to better align it with the actual received polarisation.

At VHF in particular it has been found that by using elliptical (usually circular) polarisation, a more reliable signal can be received in urban areas where there are significant reflections from buildings and other objects, by using plane polarised receive antennas. This results from the more uniform way in which circularly polarised waves are affected by propagation conditions described, particularly reflections. In free space, using a linear antenna to receive a circularly polarised wave results in a 3 dB loss compared to if a circularly polarised antenna of the correct hand had been used. Therefore, in general circularly polarised waves require transmitting at levels 3 dB higher than the equivalent using linearly polarised waves.

3.5 The Poynting Vector, Power Flux Density and the Intrinsic Impedance of Free Space

The instantaneous Poynting vector $\boldsymbol{S}$ is defined as the cross product of the vectors representing the electric field $\boldsymbol{E}$ and the magnetic field $\boldsymbol{H}$ [26], therefore

$$\boldsymbol{S} = \boldsymbol{E} \times \boldsymbol{H} \qquad (3.26)$$

It may be applied to time varying fields, such as those generated by an antenna. One component of the Poynting vector result represents an instantaneous PFD, using SI units, measured in watts per square metre W/m^2 in the direction of propagation. Integration of the Poynting vector over a closed surface in three dimensions around a source such as an antenna yields the total power emerging from the source. The direction of the Poynting vector is also the direction of the instantaneous power flow so it is sometimes known colloquially, with some justification, also as the ‘pointing’ vector.

In the far field or plane wave region of an antenna the E and H fields are perpendicular. Using a rectangular (x, y, z) co-ordinate system with respective unit vectors $\boldsymbol{a_x}$, $\boldsymbol{a_y}$ and $\boldsymbol{a_z}$, the Poynting vector evaluates as follows:

$$\begin{aligned}
\boldsymbol{S} &= \boldsymbol{E} \times \boldsymbol{H} \\
&= \begin{vmatrix} \boldsymbol{a_x} & \boldsymbol{a_y} & \boldsymbol{a_z} \\ E_x & E_y & E_z \\ H_x & H_y & H_z \end{vmatrix} \\
&= \boldsymbol{a_x} \begin{vmatrix} E_y & E_z \\ H_y & H_z \end{vmatrix} - \boldsymbol{a_y} \begin{vmatrix} E_x & E_z \\ H_x & H_z \end{vmatrix} + \boldsymbol{a_z} \begin{vmatrix} E_x & E_y \\ H_x & H_y \end{vmatrix} \\
&= \boldsymbol{a_x} \left(E_y H_z - E_z H_y \right) - \boldsymbol{a_y} \left(E_x H_z - E_z H_x \right) + \boldsymbol{a_z} \left(E_x H_y - E_y H_x \right)
\end{aligned} \tag{3.27}$$

The subscript of each field symbol indicates the component. For a plane wave, as would be present in the far field, assuming that the E field is in the x direction and that the H field is in the y direction, then all components of E and H except E_x and H_y would be zero, so the Poynting vector for this case evaluates to:

$$P_z \boldsymbol{a_z} = E_x H_y \boldsymbol{a_z} \tag{3.28}$$

That is to say that there is a component of PFD in the z direction whose magnitude is the product of the E field component and the H field component of the plane wave (E_x and H_y respectively).

Let the combined spatial and temporal representations of the E_x and H_y fields in exponential form be:

$$E_x = E_{x0} e^{j(\omega t - \beta z)} \tag{3.29}$$

and

$$H_y = H_{y0} e^{j(\omega t - \beta z)} \tag{3.30}$$

where E_{x0} and H_{y0} are the peak values of the electric and magnetic fields respectively, ω is the angular frequency and β is the spatial phase constant, then using (3.28) and the expression for the intrinsic impedance of free space η_0 that we met in (2.8)

$$P_z = \frac{E_x^{\ 2}}{\eta_0} = \frac{E_{x0}^{\ 2}}{\eta_0} e^{j2(\omega t - \beta z)} \tag{3.31}$$

The time averaged power flux density P_{zAV} W/m^2 is obtained by integrating the expression for P_z in (3.31) over one cycle and dividing the result by the period T, so

$$\begin{aligned}
P_{zAV} &= \frac{1}{T} \int_{t=0}^{T} P_z \\
&= \frac{1}{T} \int_{t=0}^{T} \frac{E_{x0}}{\eta_0} e^{j2(\omega t - \beta z)} \\
&= \frac{1}{2} \frac{E_{x0}^{\ 2}}{\eta_0} \quad W/m^2
\end{aligned} \tag{3.32}$$

The ‘time average’ of a time varying electromagnetic field is generally understood to mean the power averaged over a ‘long’ time relative to the temporal oscillation period of the field. Note that result of (3.32) is in terms of the *peak* electric field E_{x0}, so, for example, if the root mean square (RMS) electric field E_{xRMS} was used instead, the expression becomes

$$P_{zAV} = \frac{{E_{xRMS}}^2}{\eta_0} \quad W/m^2 \qquad (3.33)$$

Other analogous expressions are:

$$P_{zAV} = {H_{yRMS}}^2 \eta_0 \quad W/m^2 \qquad (3.34)$$

$$P_{zAV} = E_{xRMS} H_{yRMS} \quad W/m^2 \qquad (3.35)$$

It is interesting to note the analogies of (3.32), (3.33), (3.34) and (3.35) with the similar expressions for circuit currents, voltages and powers which are shown in Table 3-1.

Equation Reference	Plane Wave	Unit	Analogous Circuit Expression	Unit
(3.32)	$P_{zAV} = \frac{1}{2}\frac{{E_{x0}}^2}{\eta_0}$	W/m^2	$P_{AV} = \frac{1}{2}\frac{{V_0}^2}{R}$	W
(3.33)	$P_{zAV} = \frac{{E_{xRMS}}^2}{\eta_0}$	W/m^2	$P_{AV} = \frac{{V_{RMS}}^2}{R}$	W
(3.34)	$P_{zAV} = {H_{yRMS}}^2 \eta_0$	W/m^2	$P_{AV} = {I_{RMS}}^2 R$	W
(3.35)	$P_{zAV} = E_{xRMS} H_{yRMS}$	W/m^2	$P_{AV} = V_{RMS} I_{RMS}$	W

Table 3-1 Expressions for mean PFD and corresponding circuit expressions for mean power

3.6 Antenna Radiation Patterns

Fully detailed (far field) antenna radiation patterns, based on field magnitude or power flux density (PFD), are described by one or more surfaces in three dimensions. An example of one, studied in Section 2 for the Hertzian dipole, was found to be the surface of a torus-like solid. This was a regular shape with axial symmetry, but many real antennas will have radiation patterns with far more complex shapes. Any three dimensional radiation pattern may be simplified by converting it into a set of two dimensional radiation patterns most suited to the intended application. Two dimensional radiation patterns are often plotted on magnitude-angle (or polar) scales. Polar plots give a more realistic representation of the actual radiation pattern.

Remember that the reference simply to a ‘radiation pattern’ is normally understood to imply the *far field* radiation pattern and the waves under consideration are plane or transverse electric magnetic (TEM). The radiation pattern indicates how either the (electric or magnetic) field magnitude or the PFD is related to the angular orientation of the antenna, again in three dimensions. Assuming a polar plot, the radius of the surface representing the radiation pattern is proportional to the PFD received in the far field. The radial scaling of the radiation pattern may be either in linear or logarithmic units. The most common form of logarithmic unit is the decibel (dB), often used in this context due to the wide range of signal levels that may be encountered in a typical radio communications system.

Antennas designed for point to point communications will normally have one direction of maximum radiation known sometimes as the main lobe. Most practical antennas however have smaller lobes in other directions, known as sidelobes. These correspond to the leakage of radiation in unwanted directions. Sidelobes usually represent wasted power in transmit antennas and can potentially cause radiation outside of the service area and therefore interference. In receive antennas they can present a path for interference into the receiver perhaps from multipath reflections or interfering sources.

3.6.1 Representing Far Field Radiation Patterns in Two Dimensions

It is difficult to accurately represent an antenna radiation pattern in three dimensions so it is common to simplify this into a set of two dimensional radiation patterns. For a terrestrial antenna system with plane polarisation, these might be in the horizontal and vertical planes or perhaps in the electric (E) and magnetic (H) fields.

Radiation patterns in two dimensions may be represented in rectangular or polar forms. Examples of these are given in Figure 3-4 and Figure 3-5 respectively.

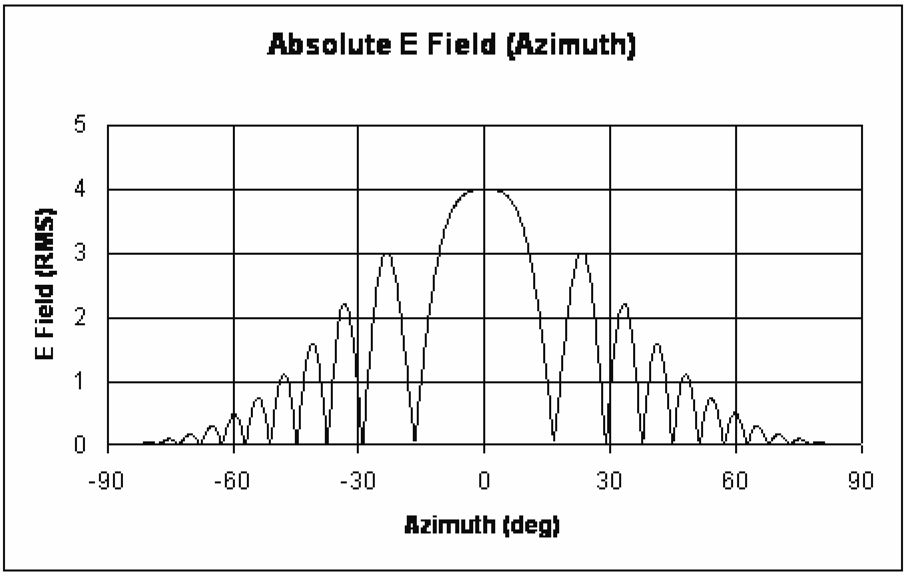

Figure 3-4 An example of an un-normalised two dimensional far field radiation pattern

The pattern depicted in Figure 3-4 shows how the magnitude of the electric field varies with respect to the azimuth position. The direction of maximum radiation or boresight is scaled to zero degrees. This may be used to derive the ERP in terms of PFD using (3.33). In this case the field is horizontally polarised (the E field is in the horizontal, or azimuth plane). In this case the field strength at boresight is 4 V RMS.

Actually a plot like Figure 3-4 is not particularly useful because the *absolute* magnitude of the field is a function of the power being transmitted into the antenna and the distance from the transmit antenna to the receive antenna. In most cases the engineer is more interested in the relative or normalised antenna radiation pattern. Furthermore, the pattern can either be normalised linearly or logarithmically. A pattern that is normalised linearly has a relative magnitude of unity (usually at boresight) and one that is normalised logarithmically has a magnitude of 0 dB at boresight.

If the general expression for the un-normalised RMS electric field is e, then this is a function of the azimuth angle θ, or

$$e = f(\theta) \quad (3.36)$$

If the peak RMS electric field, with respect to θ is e_0, then the linearly normalised electric field e_N is given by:

$$e_N = \frac{f(\theta)}{e_0} \quad (3.37)$$

The logarithmically normalised field e_{LN} would most commonly be expressed in decibels (dB), again relative to the peak value and therefore given by

$$\begin{aligned} e_{LN} &= 20\log_{10}(e_N) \\ &= 20\log_{10}\left(\frac{f(\theta)}{e_0}\right) \quad dB \end{aligned} \quad (3.38)$$

The multiplication factor of 20 in (3.38) arises because we chose originally to measure RMS electric field magnitude as opposed to PFD. Examples of log normalised radiation patterns in polar and rectangular forms are shown in Figure 3-5 and Figure 3-6 respectively.

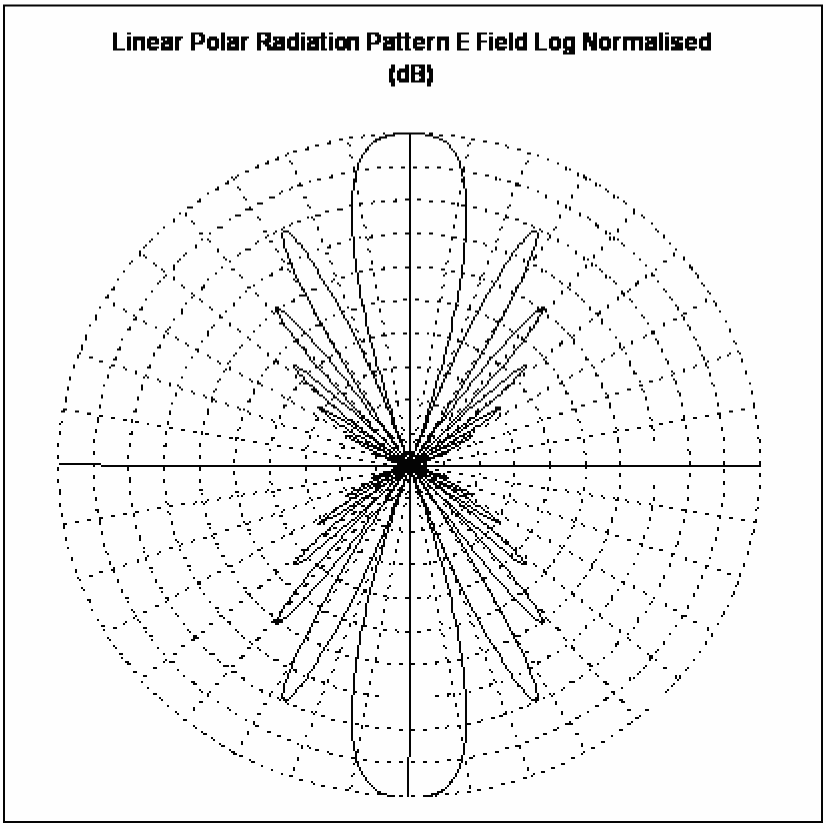

Figure 3-5 An example of a log normalised polar far field radiation pattern

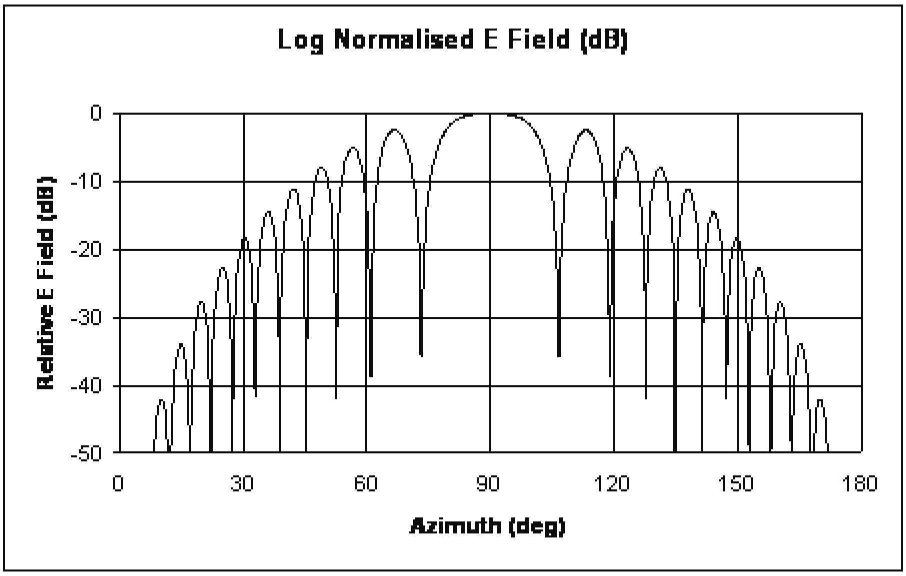

Figure 3-6 An example of a log normalised rectangular far field radiation pattern

3.7 Effective Radiated Power and Effective Isotropically Radiated Power

The effective radiated power (ERP) of a transmit antenna system is determined from the transmitter output power, losses between the transmitter and antenna and the gain of the antenna. The latter parameter means that for all practical antennas the ERP will vary according to the orientation (in three dimensions) at which it is measured. The ERP in a particular direction in linear units is the product of the power fed into the antenna and the linear gain of the antenna in the same direction. However, it is often expressed in logarithmic units such as decibels relative to a unit of power such as the Watt (dBW).

A similar definition of radiated power, the effective *isotropically* radiated power (EIRP). This is a more commonly used at microwave frequencies, where propagation is substantially through the atmosphere, which approximates to free space, well clear of reflective or absorptive objects. EIRP is measured similarly to ERP but in this case it is defined relative the gain of an isotropic radiator instead of that of a half wave dipole. The isotropic radiator was described in Section 2.1. Microwave antennas are more usually aperture types as opposed to dipole derived (wire) types and detailed three dimensional propagation data are more relevant. Again, EIRP is commonly measured in dBW. As shown in Section 3.10, the gain of a half wave dipole is 0 dBd or 2.15 dBi so it is possible to convert between ERP and EIRP by adding or subtracting 2.15 dB as required.

The vertical component of the radiation pattern must not be neglected however, even in a terrestrial system. For a typical broadcast antenna we do not wish to waste power radiating into space, into the ground or into the sides of mountains. Fortunately there are many ways antenna radiation patterns can be shaped to ensure such wasted power is minimised.

The antenna gain figures used in ERP and EIRP calculations must be chosen with caution. With a practical antenna, the true antenna gain varies with the direction in which it is measured. However, it is common for a single figure to be quoted in datasheets for antenna gain, normally understood to be that in the direction of maximum radiation. Such information might be adequate if the antenna was for receive only, requiring alignment with a fixed distant transmitting antenna. Normally the data package for an antenna suitable for transmission will include a polar or rectangular graph of antenna gain, expressed in dBd or dBi, in one or more planes.

3.8 Beamwidth

The beamwidth of an antenna is a measure of its angular resolution derived from its far field radiation pattern. Again, this is usually simplified into two dimensional plots, for example two radiation pattern plots, one for the E field magnitude and one for the H field magnitude. In most cases the E or H field beamwidth is measured between the half power (or - 3dB) points on the main lobe of the radiation pattern. These translate to the $1/\sqrt{2}$ or 0.707 points for linear (E or H field) plots. An example of a normalised E field pattern is shown in Figure 3-7, indicating the - 3 dB beamwidth.

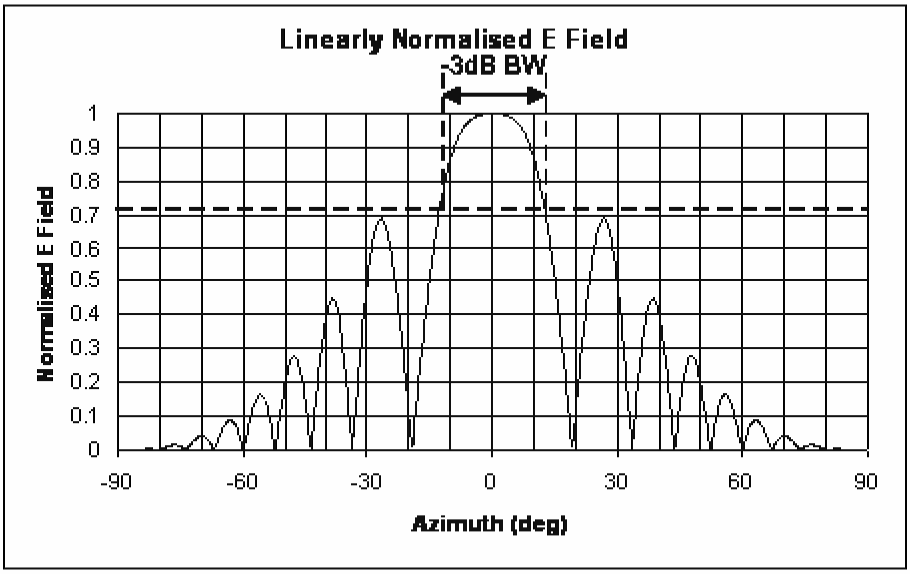

Figure 3-7 A linear normalised radiation pattern indicating the - 3 dB beamwidth

Although antenna beam radiation patterns are actually shapes in three dimensions, they do not in general have axial symmetry about the propagation axis, meaning that the E field and H field beamwidths are not necessarily equal. An exception is the Gaussian beam which is normally designed specifically to be axially symmetric. Advanced techniques exist to shape antenna beams according to specific requirements. For example, a satellite antenna may be required to illuminate a country on the Earth's surface which does not have a circular 'footprint'. Italy springs to mind as being a 'long thin' country which would not be efficiently illuminated with an axially symmetric beam and for which a shaped beam would be more appropriate.

3.9 Antenna Impedance and Radiation Resistance

The impedance of a particular antenna at the frequency of interest may be represented by a one port network comprising a lumped impedance circuit connected at the antenna feed point in place of the antenna itself [23]. This is also known as the driving point impedance and is shown schematically in Figure 3-8, where the impedance of the lumped network is Z_T. For this case it is assumed that the antenna is loss free and free from the influences caused by reflections from nearby objects. The actual impedance measured is affected by the test frequency, the antenna design and construction.

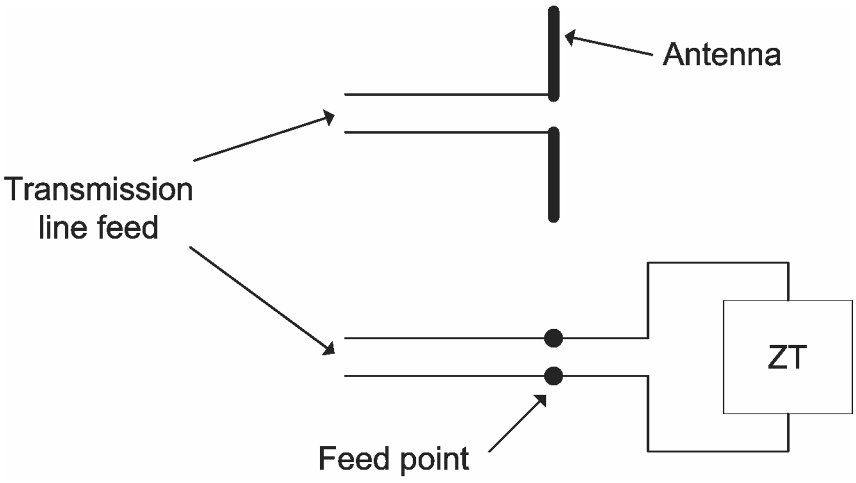

Figure 3-8 An antenna showing connection of the equivalent feed point impedance ZT

At the test frequency used, the equivalent driving point impedance will comprise a real part R_r known as the radiation resistance and a reactive part X_r known as the self reactance. All of the power absorbed in the radiation resistance is identical to that which will be transmitted.

3.10 The Half Wave Dipole

The half wave dipole antenna comprises two axially aligned elements, each one quarter of a wavelength long and fed at the mid point as shown schematically in Figure 3-9. Unlike the isotropic radiator or Hertzian dipole, an antenna approximating to a half wave dipole can actually be constructed and used. The antenna is normally considered in free space well clear of any grounded objects and ideally therefore may be fed by a balanced transmission line as shown. However, balanced transmission lines are often difficult to use and it is common practice to feed the antenna using a coaxial (unbalanced) transmission line and insert a balanced to unbalanced transformer (balun) at the antenna.

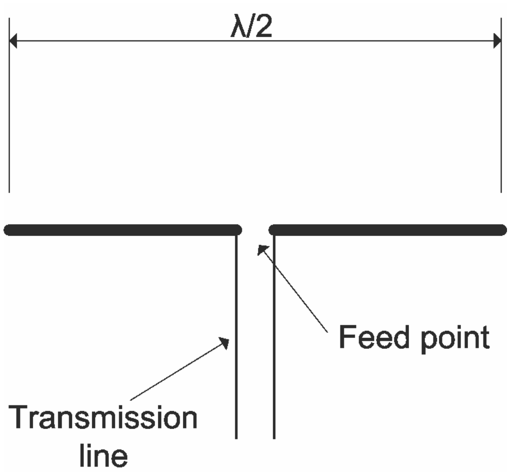

Figure 3-9 The half wave dipole antenna showing a balanced transmission line feed at the mid-point

The far field radiation pattern of a half wave dipole is similar to that for the Hertzian dipole shown in Figure 2-2. Viewed in three dimensions it is axially symmetric like that of a slightly squashed torus with a hole of just zero diameter. The torus axis is aligned with that of the antenna and the feed point is positioned in the central plane of the torus. The two dimensional linear radiation patterns for both the half wave dipole and the isotropic radiator are shown in Figure 3-10. Because of the axial symmetry, these are shown in two dimensions using rectangular co-ordinates plane assuming that the axis of the antenna is aligned with the x axis and positioned at the origin. The origin is also coincident with the centre of the isotropic radiator.

The shape of the radiation pattern for the half wave dipole is described by the equation describing the electric field strength in terms of θ, $E(\theta)$ as follows:

$$E(\theta) = D\frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta} \qquad (3.39)$$

In this case θ is the angle between the direction of the free space ray and the axis on the antenna and D is known as the array factor. The array factor is 1.28, which will become clear when we consider the gain of the half wave dipole relative to that of the isotropic radiator. By definition, the gain of an isotropic radiator is unity so it is represented by a circle with radius unity. The isotropic radiator radiates equally in all directions, so in three dimensions it may be represented by a sphere, also of radius unity, centred at the origin. The half wave dipole does not radiate at all along the axis containing its elements and then progressively radiates more up to a maximum at $90°$ from the axis.

The definitions for antenna directivity and gain are given in Section 3.1.3. For the perfect (loss free) antennas that we are considering, the directivity and gain relative to an isotropic radiator are identical. Furthermore, we have defined the gain of an isotropic radiator to be unity. Although the field radiation pattern of the half wave diople varies significantly according to the value of θ , the definition of its gain is taken in line with the direction of maximum radiation, $\theta = 90°$ in this case. Performing the directivity calculation arrives at gain (or directivity) result of 1.28. That is to say that the gain of a half wave dipole is 1.28 relative to an isotropic radiator.

Remembering that we are still using linear units, the logarithmic of gain expressed in dB relative to an isotropic radiator (dBi) is more common, given by:

$$g_{HWD} = 20\log_{10}(1.28) = 2.15 \quad dBi \qquad (3.40)$$

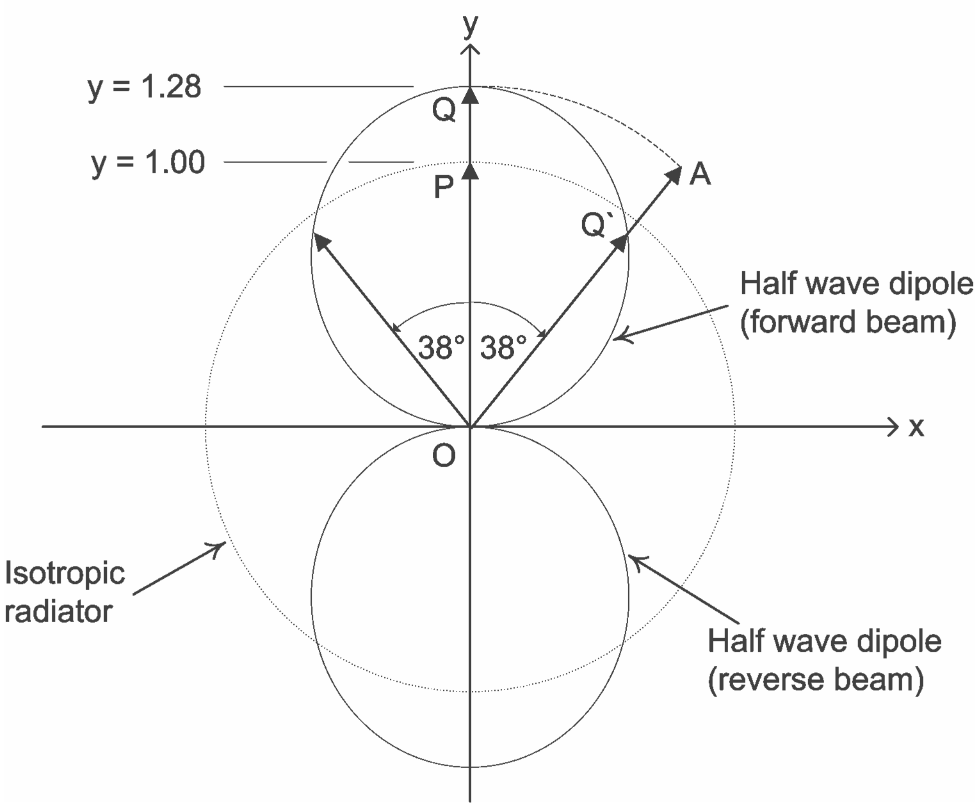

Figure 3-10 Radiation patterns for the half wave dipole related to that for the isotropic radiator in two dimensions. The antenna is located on the x-axis at the origin.

Over the years, due to its simplicity and practicality, the half wave dipole has become the generic reference for antennas designed for use in terrestrial communications instead of the isotropic radiator. Specified gains using this definition have adopted the logartihmic symbol dBd meaning gain relative to the gain of a half wave dipole. Therefore the gain of an antenna expressed in dBd is 2.15 dB less than its gain expressed in dBi.

It is important to note that we have assumed the half wave dipole to be loss free. It is impossible to design such an antenna in practice. Whilst the isotropic radiator is a loss free and hypothetical antenna, the finite conductivity of the antenna elements comprising a half wave dipole, their diameter and general geometry means that it is difficult to achieve a practical gain of more than approximately 1.5 dBi to 2 dBi for a typical example.

3.11 Fourier Transforms applied to Aperture Illuminations

One of the many applications of the Fourier transform is in the field of antenna radiation patterns to predict the far field radiation pattern of an antenna starting with the aperture illumination. The following equation is the form of Fourier transform used to convert a one dimensional aperture illumination (represented as a function of x, $F(x)$ along the x axis) into a far field radiation pattern in the same plane.

$$G(u) = \int_{x=-\infty}^{\infty} F(x) e^{-j2\pi ux} dx \qquad (3.41)$$

(3.41) uses the following substitution:

$$u = \frac{-1}{\lambda} \sin\theta \qquad (3.42)$$

λ is the wavelength of the radiation and θ is the angle from boresight.

$F(x)$ is the electric field at the aperture, as a function of the distance along the considered dimension (x) and $G(\theta)$ is the resulting electric field measured in the far field (or Fraunhofer) region as a function of θ.

3.11.1 Uniform Truncated Aperture Illumination

An example of a uniform (truncated) aperture illumination of amplitude 1 V/m, is shown in Figure 3-11, meets the following relationships:

$$\begin{aligned} F(x) &= 1 \quad \frac{-a}{2} \le x \le \frac{a}{2} \\ F(x) &= 0 \quad elsewhere \end{aligned} \tag{3.43}$$

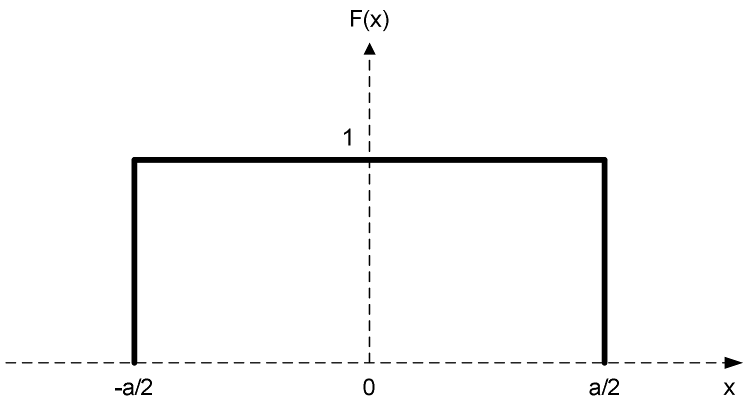

Figure 3-11 A uniform aperture illumination of 1 V/m in one dimension

The resulting far field radiation pattern resulting from this illumination may be obtained mathematically by applying the Fourier transform in the style shown in(3.41) between the limits $x = -a/2$ and $x = a/2$. The finite integral equation becomes:

$$G(u) = \int_{x=-a/2}^{x=a/2} 1 e^{-j2\pi ux} dx \tag{3.44}$$

This may be evaluated with the help of the substitution suggested in (3.42) together with the following Euler's identity:

$$\sin\phi = \frac{1}{2j}\left[e^{j\phi} - e^{-j\phi}\right] \tag{3.45}$$

The result in terms of θ is:

$$G(\theta) = \frac{\lambda}{\pi \sin\theta} \sin\left[\frac{\pi a}{\lambda} \sin\theta\right] \tag{3.46}$$

The result of (3.46) is plotted for the three values of a/λ : 2, 10 and 20 in Figure 3-12. The resulting three radiation patterns are linear versions and therefore include regions extending into both sides of the angle axis. These represent the components of spatial phase of the field in the far field region which are in phase (positive) and in anti-phase (negative).

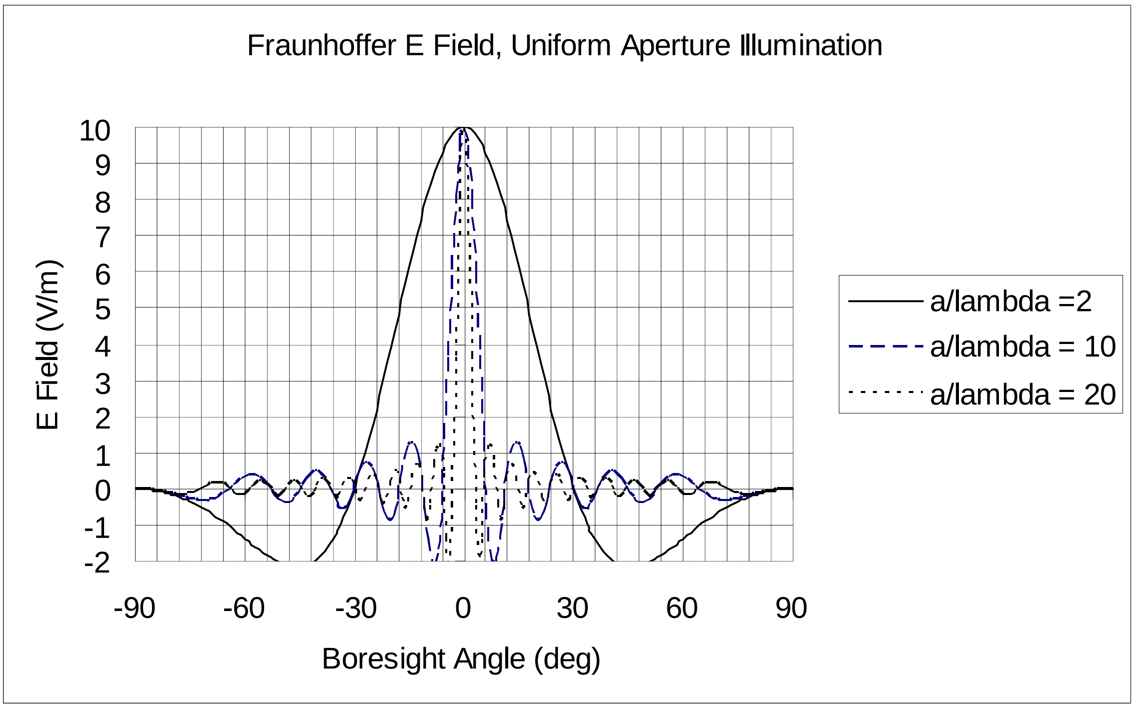

Figure 3-12 One dimensional Fraunhofer radiation pattern for uniform aperture illumination, for three values of a/λ

It is often more useful to illustrate the normalised logarithmic field strength in dB against angle. The graphs in Figure 3-13 are the equivalents of Figure 3-12 in this form. The logarithmic amplitude of the field strength N_{dB} is independent of phase, being obtained using the following expression:

$$N_{dB} = 20\log_{10}\left|G(\theta)\right| \quad dB \tag{3.47}$$

The logarithm coefficient is 20 as opposed to 10 to give a figure which is proportional to PFD in watts per square metre (W/m^2). This is necessary because electric field intensity is expressed as a linear quantity in volts per metre (V/m). Normalisation is usually taken with respect to the PFD in the main lobe as this is used to both define the antenna gain and to use for normal communications. Therefore the (normalised) PFD in the main lobe would rise to a peak at 0 dB.

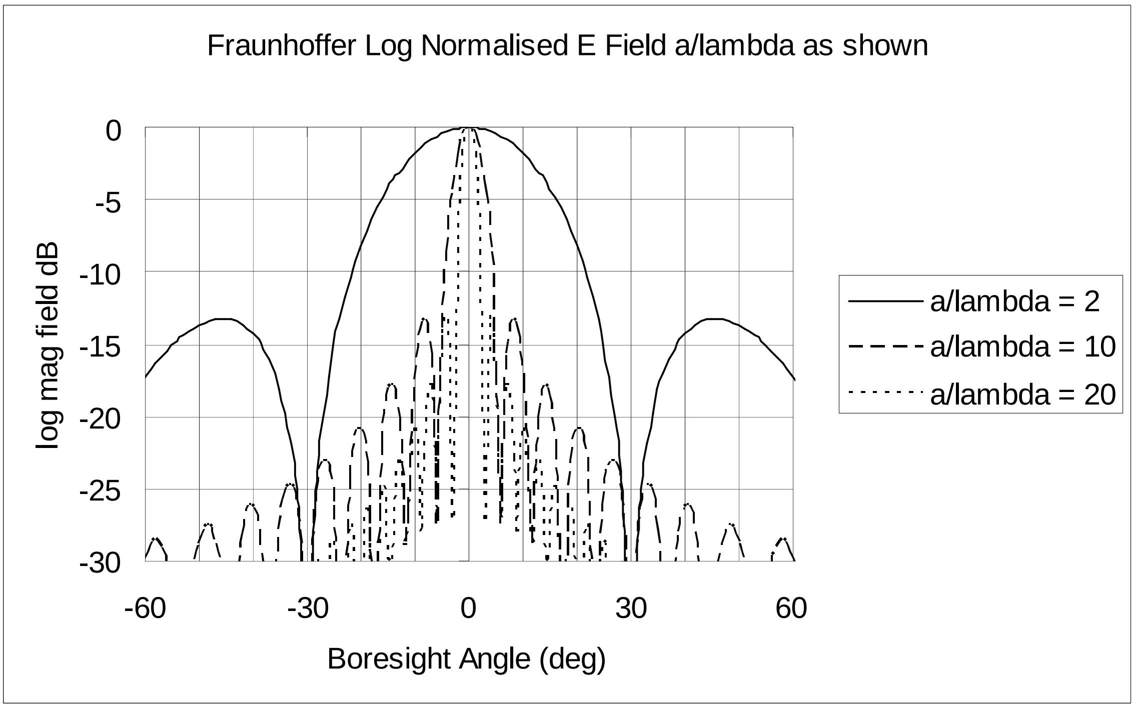

Figure 3-13 The one dimensional Fraunhofer radiation pattern represented in normalised logarithmic form for uniform aperture illumination

The radiation patterns in Figure 3-13 show some interesting properties in terms of the length of the aperture expressed in wavelengths and the ratio a/λ. For a fixed wavelength λ, the main lobe beamwidth becomes narrower, the larger the aperture becomes. Sidelobes are always present: each order of sidelobes having the same level relative to the main lobe, irrespective of the ratio a/λ.

3.11.2 Arbitrary Illumination Profiles

In most antenna applications, any significant sidelobe levels are undesireable. In transmit antennas they represent wasted energy transmitted in an unwanted direction and in receive antennas allow a path for interfering signals or multipath to enter the receiver.

4 PASSIVE ANTENNAS

4.1 Slot Antennas and Booker's Theorem

An antenna can be formed from a slot in a conducting surface as an alternative to elements comprised of wires carrying currents. Slot antenna designs are generally more expensive to produce than the equivalent element types but offer some advantages for use in high power transmit (broadcast) antennas at VHF and UHF. They may be fed more easily by waveguides and do not require as much maintenance.

Bookers theorem is useful in designing slot antennas [19]. This states that a given element antenna may be transformed to the equivalent slot antenna by replacing the free space area around the elements with conductive plane and making the element a void area. The radiation patterns of both types will be identical except that polarisations will be reversed (vertical will become horizontal and vice versa)

The impedances of the element and slot antennas will be related by the following equation

$$\eta_0^{\,2} = 4 Z_d Z_s \qquad (4.1)$$

where

η_0 is the intrinsic impedance of free space

Z_d is the impedance of the dipole (element antenna)

Z_s is the impedance of the slot antenna

4.2 Telescopic Rod Antenna

The ‘telescopic rod’ type of antenna found in many low cost portable receivers is normally provided for operation in the VHF broadcast frequency bands. The most common of these, covering the frequency range 88 MHz to 108 MHz, has been used for many years throughout the world for broadcasting analog frequency modulation (FM) services. More recent developments in digital broadcasting, such as digital analog broadcasting (DAB) use higher frequency portions of the VHF spectrum such as band III (174 MHz to 230 MHz).

The antenna itself is not only telescopic but usually hinged at its base to allow for various orientations as well as lengths. One’s initial thoughts might be that this, combined with the radio body itself, is an approximation to a quarter wave antenna over ‘infinite’ (or at least extensive) ground plane. Unfortunately electrically it only has the vaguest resemblance to this, as the best ground plane available would normally be electrically very small, comprising a few components inside the receiver. The range of adjustment provided by the telescopic rod antenna itself and by the movement of the radio receiver does sometimes provide some small protection against multi-path reflections which are very significant in most radio envoironments today. Any improved reception achieved by adjustment of the rod length and/or orientation is not usually a better approximation to a particular theoretical antenna but more a case of rejecting a particular propagated multi-path component in favour of another. Today’s radio environment is very different from what it was 40 or 50 years ago when band II (88 MHz to 108 MHz) broadcasts were in their infancy. Then signals were very weak in many areas and portable radios were really not very effective in anywhere but relatively close to the transmit antenna, generally considered to be within ‘line-of-sight’ of it. Today we find that PFDs are much greater almost everywhere and propagation predictions actually assume that significant multipath components will reach the receiver antenna. This is somewhat of a brute force method which might not be the best way if we had the opportunity to design a new system from scratch today. The approximate levels of received multipath can be estimated by checking whether the transmit antenna is actually visible from the receiver antenna. Very often it will not be meaning that the received signals are totally by multipath. The arguments about multipath apply equally to band III (174 MHz to 230 MHz) carrying DAB transmissions. However, DAB was in fact developed largely to tolerate significant multipath propagation.

One advantage of having a small groundplane is that several new propagation mechanisms arise: for example diffraction around the edge of the groundplane. Another is the tendancy for the whole chassis to behave as an antenna element separate from the rod. These phenomena allow the reception from a number of unexpected directions as explained by Kraus et. al. [22]. Although this reference considers transmit antennas, the principle of reciprocity make it equally applicable to receive antennas.

4.3 Planar Air Loop and Ferrite Rod Antennas

The ferrite rod antenna is a generic name for what is in fact a small loop antenna wound on a ferrite rod found in low cost receivers covering the AM broadcast bands within the LF and MF frequency bands. Although theoretically suitable for transmitting low powers, in practise they are used almost universally in receivers. Rather than one loop (turn) there would normally be several, the number dependent on the frequency that is being used and the electrical characteristics of the antenna materials.

There are two electrical criteria to be considered with ferrite rod antennas:

- the circuit (predominantly inductive) properties
- the (loop) antenna properties

4.3.1 The Circuit Inductive Properties

Ferrites are a type of ferrimagnetic material derived from compounds of iron and oxygen. A ferrite has an appreciable relative permeability μ_r compared to that of air which is very close to unity. The overall (total) permeability μ is given by:

$$\mu = \mu_0 \mu_r \qquad (4.2)$$

where μ_0 $(= 4\pi \times 10^{-7}\ H/m)$ is the absolute permeability.

Ferrites have been developed to work with relatively low loss into the MF range and higher for various applications so they are used extensively in the antennas of portable radio receivers covering approximately 100 kHz to 1.5 MHz.

The following expression for the inductance L of a solenoid is also a good approximation also for a ferrite rod antenna [20].

$$L = \frac{\mu_0 \mu_r N^2 A}{\sqrt{4R^2 + l^2}} \qquad (4.3)$$

where:

N is the number of turns

A is the cross sectional area (m^2)

R is the internal radius of the solenoid (m)

l is the length (m)

The appreciable relative permeability of the ferrite material compared to that of air (close to unity) means that the magnetic flux associated with the loops of wire is concentrated in the core when the rod is present compared to an equivalent air core. This means that the same level of inductance can be created with far fewer loops (turns) from a shorter length of wire. As the DC resistance of the turns of wire contributes directly to the antenna loss resistance, this shorter length of wire allows a heavier gauge to be used thus reducing the loss resistance.

Within the front end of the receiver circuit, the inductance of the solenoid is normally resonated with a capacitor to give a degree of frequency selectivity centred on the band of interest. For receivers covering the AM broadcast band (approximately 500 kHz to 1.5 MHz), it is common practice to used two separate coils fixed to the ferrite rod, each switchable according to the selected band. The coil with fewer turns is used to resonate the higher frequency (MW) band and the other for the LW band.

4.3.2 The Loop Antenna Properties

The theory of the small loop antenna may be applied to one comprising the coil of wire wound on to a ferrite rod [21]. The loop size is electrically small, meaning that its dimensions are much smaller than a wavelength. This holds true with ease as the shortest wavelength tuneable on a AM receiver is typically about 200 m. The electric field produced in the azimuth plane $E(\theta)$ at an angle θ and distance d , is related to the current i in the loop and its cross sectional area A, as shown in (4.4). Ultimately, the gain of this type of antenna is limited by the cross sectional area of the loop itself, and is independent of whether it might have a ferrite core. The only reason for using the ferrite rod is to enable sufficient inductance in the limited space available as described in the last section.

$$E(\theta) = \frac{120\pi^2 iA}{d\lambda^2}\sin\theta \qquad (4.4)$$

It might be thought that more signal could be 'sucked out' of the antenna just by adding extra turns. Unfortunately this does not work if the coil turns (loops) are close together as they would be within the confined space available in a small plastic box. Each loop has an equivalent aperture area associated with it and each aperture can only be used once. The only way to increase the signal level received from a weak station is to increase the cross sectional area but this is limited due to the small size of the product. In the early days of portable radios, before the development of ferrite materials, antennas were often air loops but of much larger cross sectional areas.

There is a reciprocal relationship between the loss of the ferrite rod antenna and its Q factor. Contributions to the loss come from the resistance of the coil at the operating frequency and the loss factor of the ferrite material itself (known as $\tan\delta$).

As (4.4) shows, the radiation pattern of a loop antenna, considered in the plane containing the axis of the loop, is sinusoidal. As this represents the electric field, the equivalent expression for PFD would include a sine squared term. This describes one very useful property of loop antennas: the ability to null out unwanted signals by rotating the plane of the loop, with respect to the direction that the signal is coming from. In the early days of radio this was exploited in larger (air cored) versions, in direction finding equipment. Due to the sine squared property, the position of the null (or minimum) can be determined with greater accuracy than the position of the maximum.

5 ACTIVE ANTENNAS

5.1 The Linear Array Fed in Phase

The linear array antenna comprises a series of identical antenna elements which are arranged in a straight line in the same plane. A schematic diagram of a linear array comprising four elements numbered 1 thro 4 fed in phase, is shown in Figure 5-1 [24]. It is assumed that the radiation pattern of each individual element is axially symmeric about its centre. The elements could either be considered as isotropic radiators or, to take a more practical example, half wave dipoles with their axes normal to the plane of the diagram. The spacing between adjacent elements of the array (d) is fixed. The direction of the wavefront will be considered at an angle θ to the line containing the array elements as shown. Although the wavefronts shown appear to be close to the array in what we would expect to be the near field, we are making assumptions based on the far field, considered to be many wavelengths from the array. In this region it is accurate to consider the antenna as a point source.

On propagating through space, the component of the wavefront that originates from element 1 is retarded from that originating in the same direction from element 2 by a phase of ϕ where:

$$\phi = \frac{2\pi d}{\lambda} \cos\theta \quad radians \qquad (4.5)$$

This is the product of the spatial phase constant $\beta = \frac{2\pi}{\lambda}$ and the distance between adjacent wavefronts measured along the direction of propagation. The direction of propagation is at $90°$ to the wavefronts.

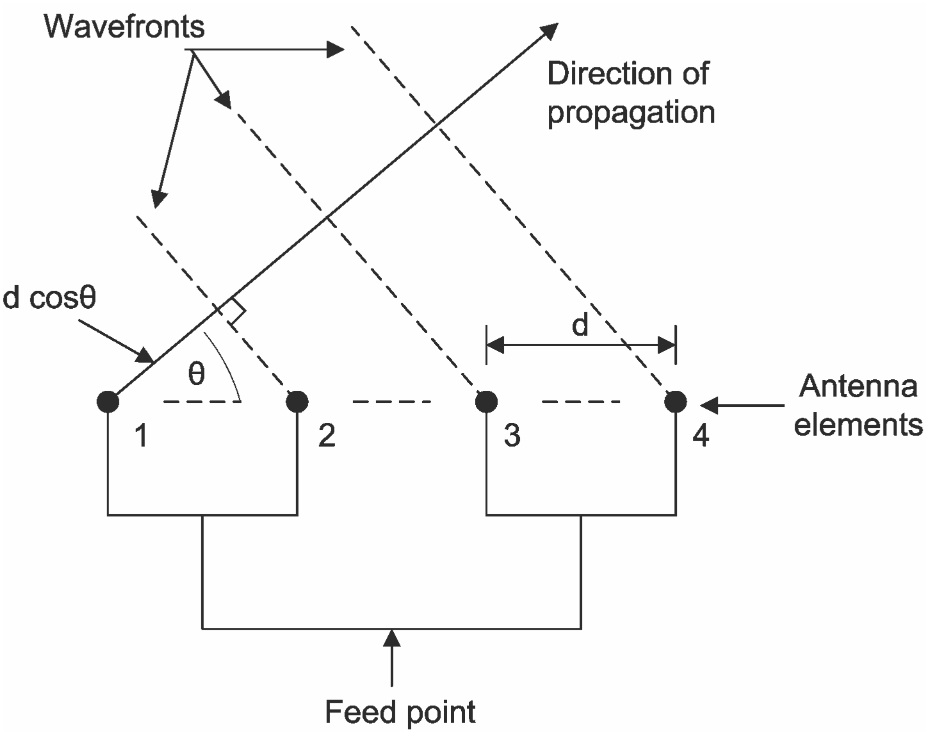

Figure 5-1 A schematic of the broadside linear array with all elements fed in phase

The equation describing the total electric field $E(t, z)$ of a plane wave progressing through space in the direction of propagation z is given by:

$$E(t,z) = \mathrm{Re}\left(E_0 e^{j(\omega t - \beta z)}\right) \tag{4.6}$$

where

E_0 is the peak value of the electric field originating from one element in isolation (V/m).
$\omega = 2\pi f$ is the angular frequency in rad/s, f is the frequency in Hertz.
$\beta = \frac{2\pi}{\lambda}$ is the spatial phase constant in rad/m.

By convention, the real operator if often omitted, but it is understood to be present. In the case being considered the direction of propagation is that shown in Figure 5-1. Therefore the equation of the wavefront which includes contributions from each of the elements in terms of ϕ is given by:

$$E = E_0 e^{j\omega t}\left(1 + e^{-j\phi} + e^{-j2\phi} + \ldots + e^{-j(N-1)\phi}\right) \tag{4.7}$$

where N is the total number of elements. Using the theory of the Taylor series and the Euler identity to convert between the complex and trigonometric versions of the waveform yields the following result [59]:

$$E = E_0 \frac{1 - e^{jN\phi}}{1 - e^{j\phi}} = E_0 \left(\frac{\sin\left(\frac{N\pi d}{\lambda}\cos\theta\right)}{\sin\left(\frac{\pi d}{\lambda}\cos\theta\right)} \right) \tag{4.8}$$

For $\theta = 90°$ this reduces to:

$$E = NE_0 \tag{4.9}$$

$\theta = 90°$ is the direction of maximum radiation, at right angles to the line containing the array elements. This is the reason why a linear antenna fed in this way is often known as a ‘broadside array’ It also shows that the electric field for $\theta = 90°$ is the product of the field due to one element in isolation and the number of elements. The antenna gain is therefore directly proportional to the number of elements. The ratio of the element spacing to the wavelength d/λ influences the radiation pattern beamwidth. The smaller the ratio the wider the beamwidth as the examples in Figure 5-2 show for d/λ values of 0.1, 0.25, 0.5 and 0.9.

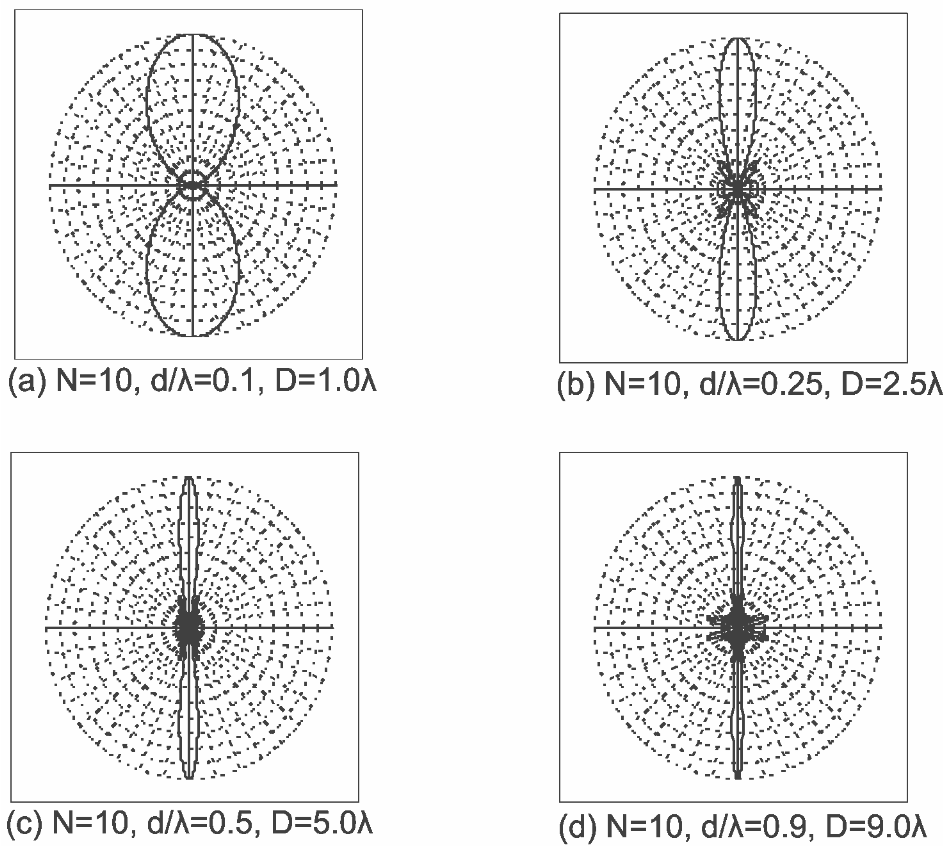

Figure 5-2 Some examples radiation patterns for the values of N and d/λ shown. D is the length of the entire array

The symmetry of the array ensures that there are always two identical beams orientated at $180°$. When d/λ is increased the sidelobes increase to very high levels and occupy beamwidths in excess of the main beam. Two examples of these are shown in Figure 5-3 for d/λ values of 1.0 and 2.0.

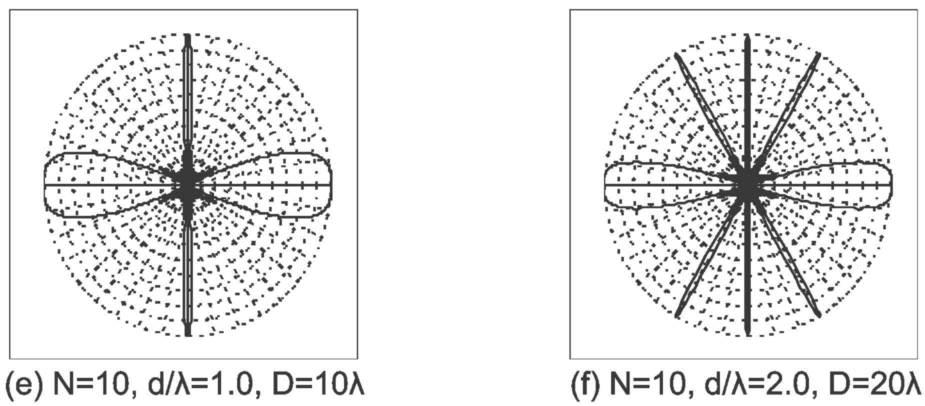

Figure 5-3 Further examples of radiation patterns for d/λ

5.1.1 Scanning the Radiation Pattern from a Linear Array

By properly controlling the relative phase of the signals arriving at each of the elements, the direction of the main beam can be changed. One form of phase control which achieves this is known as linearly progressive. In this the phase shift is applied in proportion to the element position. So for example, if the phase shift applied at element 1 was ϕ that at element 2 would be 2ϕ, at element 3 it would be 3ϕ and so on. A schematic of the array with linearly progressive phase is shown in Figure 5-4.

If the progressive phase shift is δ, this must be added to the phase of the retarded potential expression (4.5). Supposing the new retarded phase of the wavefront is φ, then

$$\varphi = \frac{2\pi d}{\lambda}\cos\theta + \delta \qquad (4.10)$$

φ replaces ϕ in (4.8) to give:

$$E = E_0 \frac{1-e^{jN\varphi}}{1-e^{j\varphi}} = E_0 \left(\frac{\sin\left(\frac{N\pi d}{\lambda}\cos\theta\right)}{\sin\left(\frac{\pi d}{\lambda}\cos\theta\right)} \right) \qquad (4.11)$$

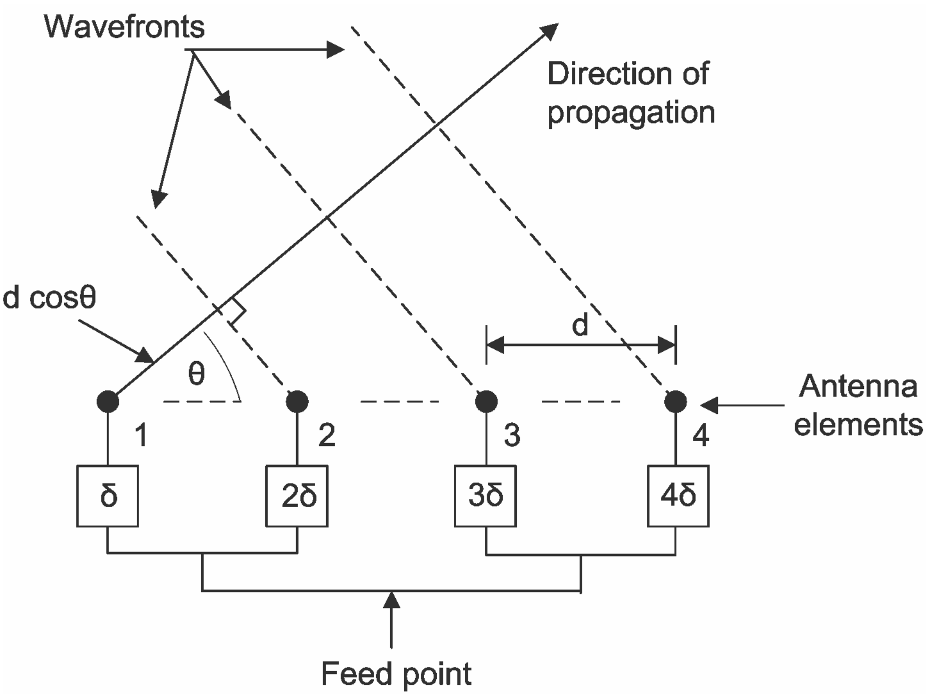

Figure 5-4 Linear array showing the addition of progressive phase shifts

Changing the value of δ allows the main lobe of the radiation pattern beam to be scanned. This is shown in Figure 5-5 for δ values of $0°$, $-45°$, $-70°$ and $-90°$. As the beams move towards the endfire direction, the beams merge to form a single beam of somewhat greater beamwidth.

Figure 5-6 shows an alternative way of achieving linear progressive phase shift, by serially feeding the elements with a shift of δ between each adjacent elements. The linear array antenna with a means of adjusting the phase shift in each of the element feeds is the basis of the scanning antenna.

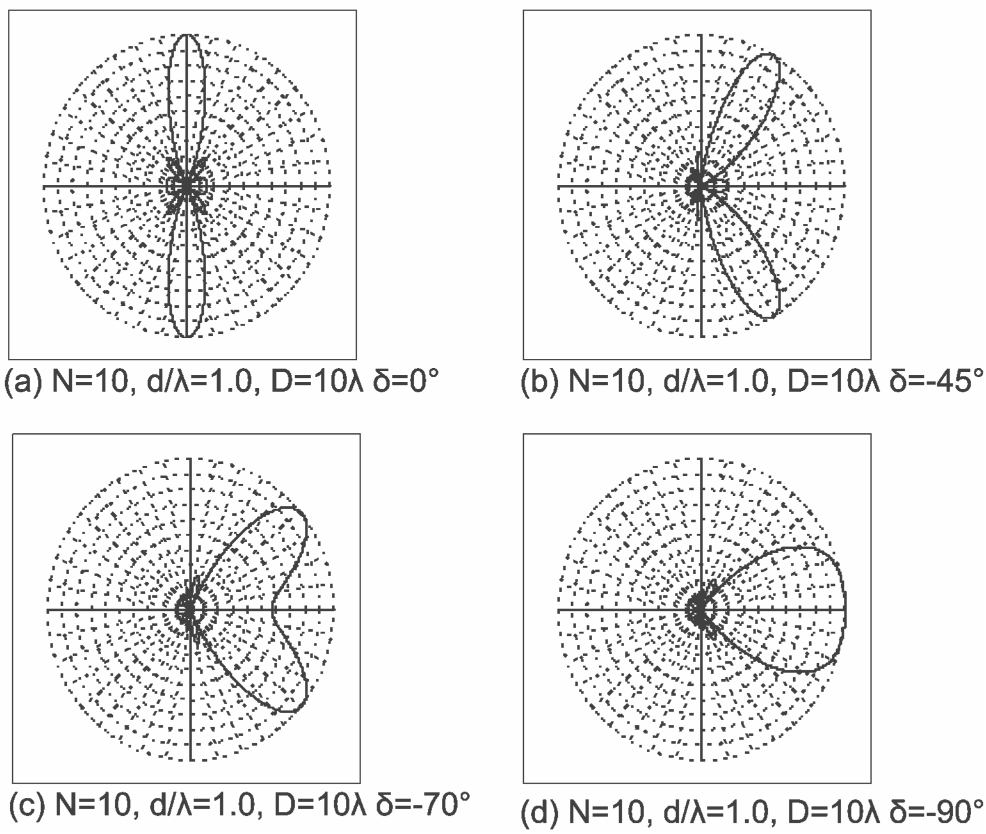

Figure 5-5 Provision of scanning the main lobe of the radiation pattern by changing the value of δ

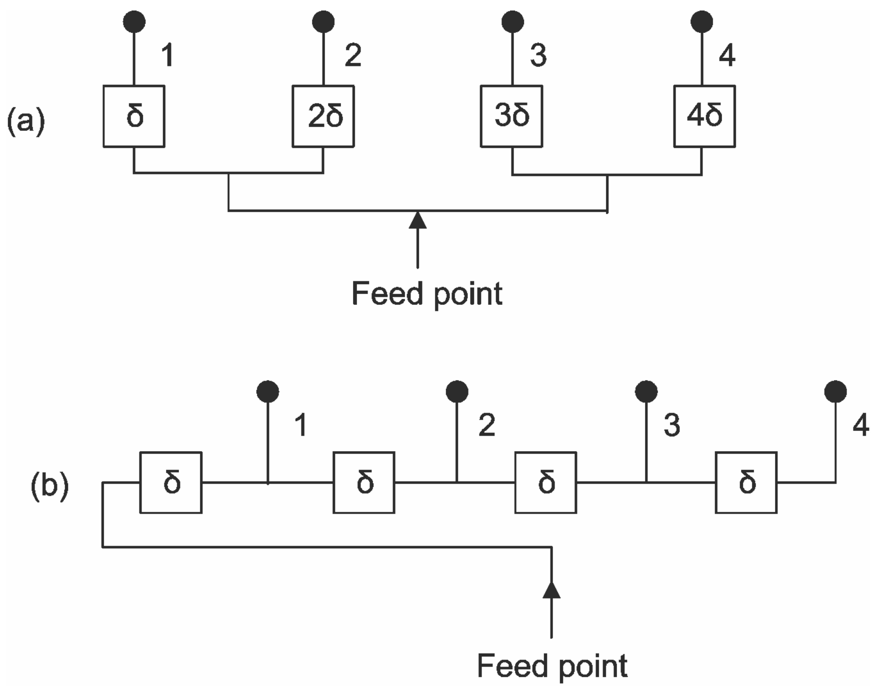

Figure 5-6 Two different techniques for feeding an array: parallel feed (a) and the easier to implement serial feed (b)

6 REFERENCES

1. Connor, F. R.; Antennas; Edward Arnold; pp 10 - 11; ISBN 0 7131 3262 0.T
2. Mouly, Michel & Pautet, Marie; The GSM System for Mobile Communications; (published by the authors); pp 593 - 601; ISBN 2-9507190-0-7
3. Connor (op. cit.); pp 6 - 7.
4. Kraus, John D. & Carver, Keith R.; Electromagnetics - Second Edition; McGraw-Hill Kogakusha Ltd.; pp 606 - 607; ISBN 0-07-035396-4 (1973).
5. Ramo, Simon, Whinnery, John R. & Van Duzer, Theodore; Fields and Waves in Communications Electronics; John Wiley & Sons; pp659 - 661
6. Ramo et. al. (op. cit.); pp 242 - 243.
7. Ramo et. al. (op. cit.); pp 262 – 265.
8. Ramo et. al. (op. cit.); pp 642 - 645.
9. Chipman, Robert A.; Transmission Lines; Schaum's Outline Series in Engineering; McGraw Hill Book Company;
10. Pozar, David M.; Microwave Engineering - Third Edition; John Wiley & Sons Inc; ISBN 0-471-44878-8 (2005); pp 106 - 117.
11. Test Method Standard for the Measurement of Electromagnetic Interference Characteristics; MIL-STD-462D United States Department of Defense; 5 February 1996; Test Method RE103 pp 97 - 102.
12. Ramo et. al. (op. cit.); p 324.
13. Ramo et. al. (op. cit.); p 645.
14. Connor F. R.; Antennas; Edward Arnold (Publishers) Ltd; p 11; ISBN 0-7131-3262-0.
15. Connor F. R. (op. cit.); p12.
16. Connor F. R. (op. cit.); Appendix A.
17. Connor F. R. (op. cit.); p90.
18. Hannan, P. W.; Microwave Antennas Derived from the Cassegrain Telescope; IRE Transactions on Antennas and Propagation March 1961; pp 140 - 153.
19. Kraus et. al. (op. cit); p 687.
20. Kraus et. al. (op. cit); pp 155 - 160.
21. Kraus et. al. (op. cit); pp 654 - 655.
22. Kraus et. al (op.cit); pp679 - 680.
23. Kraus et. al (op.cit); pp670 - 672.
24. Kraus et. al (op.cit); pp631 - 635
25. Kraus et. al (op.cit); pp 390 - 391.
26. Hayt, William H.; Engineering Electromagnetics - 4th Ed.; pp 396 - 398.
27. Sklar, Bernard; Digital Communications, Fundamentals and Applications - Second Edition; Prentice Hall PTR, New Jersey; pp 779 -781. ISBN 0130847887.
28. Kraus et. al (op.cit); pp668 - 669.
29. Kraus et. al (op.cit); pp 363 -364.
30. Kraus et. al (op.cit); pp 452 - 460.
31. Kraus et. al (op.cit); pp 460 - 464.
32. Kennedy, George, Electronic Communication Systems - ISE; McGraw Hill Kogakusha Ltd.; pp253 - 254. ISBN 0-07-034052-8.
33. Kennedy (op. cit); p 279.
34. Baden Fuller, A. J.; Microwaves - Second Edition; Pergammon Press; pp 179 - 181. ISBN 0-08-024227-8.
35. http://www.narda-sts.us/products_personal_main.php; Narda Alert non ionising radiation exposure meter.
36. http://www.rfcafe.com/references/electrical/fcc-maximum-permissible-exposure.htm; FCC (USA) Maximum permitted exposure levels.

37. Kraus et. al (op.cit); pp 620 - 624.
38. Halliday, David and Resnick, Robert; Physics, Parts 1 and 2 Combined, Third Edition; John Wiley & Sons New York; pp 994 - 997. ISBN 03668-4.
39. Protter, Murray H and Morrey, Charles B.; College Calculus and Analitic Geometry, Third Edition; Addison-Wesley Publishing Company; pp 257 - 263.
40. Protter & Morrey (op. cit.); pp 271 - 276.
41. Hannan, Peter W.; Microwave Antennas Derived from the Cassegrain Telescope; IRE Transactions on Antennas and Propagation March TBA; pp 140 -153.
42. Stephenson, G,; Mathematical Methods for Science Students, Second Edition; Longman Group Limited, London; p 104. ISBN 0582444160.